Mining Tourist Behavior: A study of Tourist Sequential Activity Pattern through Location Based Social Networks

Masters By Research

VU RESEARCH Institute for Sustainable Industries & Liveable Cities, Victoria University

Submitted By : Anmoila Talpur in 2019

ABSTRACT

Much of the current research in tourism has focused on tourists' behavior analyses in order to help management constructing effective tourism policies and strategic planning to cater for a diverse range of tourists. Insight into tourist movement and activity patterns is deemed beneficial for the tourism sector in many ways, such as designing better travel packages for tourists, maximizing the tourist activity participation and meeting the tourist demands. Existing works in this field have only focused on finding tourists' travel trajectories; however, they have not been able to provide comprehensive and complete information about the actual anticipated activities at visited locations. This is probably due to the limitation of traditional data collection and analysis approaches. This research proposes to adopt mobile social media data for effective capturing of tourist activity information and utilizes advanced data mining techniques for extracting valuable insights into tourist behavior. The proposed methods and findings of the study have the potential to support tourism managers and policy makers in making better decisions in tourism destination management.

Masters by Research Student Declaration

Master by Research Declaration

"I, Anmoila Talpur, declare that the Master by Research thesis entitled "Mining Tourist Behavior: A study of tourist sequential activity patterns through location based social networks" is no more than 60,000 words in length including quotes and exclusive of tables, figures, appendices, bibliography, references and footnotes. This thesis contains no material that has been submitted previously, in whole or in part, for the award of any other academic degree or diploma. Except where otherwise indicated, this thesis is my own work".

Signature : ***A. Talpur*** **Date : 11/12/2019**

Table of Contents

CHAPTER 1: INTRODUCTION 8
1.1 GENERAL 8
1.2 AIM OF THE STUDY 11
1.3 THE SCOPE 16
1.4 RESEARCH QUESTIONS 16
1.5 RESEARCH OBJECTIVES 16
1.6 THE LIMITATIONS 17
1.7 SEQUENTIAL PATTERN MINING 20
1.8 TOURIST SEQUENTIAL ACTIVITY PATTERNS 26
1.9 SUMMARY 30
CHAPTER 2: LITERATURE REVIEW 32
2.1 INTRODUCTION 32
2.2 SOCIAL MEDIA ROLE IN TOURISM 36
2.3 DESTINATION MARKETING IN TOURISM 38
2.4 DESTINATION MANAGEMENT - A KEY PLAYER IN TOURISM 40
2.5 AN OVERVIEW OF TOURISTS 44
2.5.1 ***Tourist Mobility Patterns*** 44
2.5.2 ***Tourist Travel Behavior*** 47
2.5.3 ***Tourist Activity Patterns*** 49
2.6 ECONOMY OF THE TRAVEL INDUSTRY 52
2.7 THEORIES OF TOURIST BEHAVIORAL MINING 53
2.7.1 Compendium Theory 53
2.7.2 Generation Theory 53
2.7.3 Cohen's Theory 54
2.7.4 Butler's Theory 54
2.8 EXPLORING THE GLOBAL TOURIST SEQUENTIAL ACTIVITY PATTERNS 55
CHAPTER 3: METHODOLOGY 90
3.1 AN OVERVIEW 90
3.2 PRIVACY CONCERNS 91
3.3 DATA COLLECTION 92
3.4 FOURSQUARE 93

3.5 DATA PRE-PROCESSING 94
3.6 ALGORITHM DEVELOPMENT 96
3.6.1 ***The Pattern-Growth Method: Prefix span Algorithm*** 97
3.6.2 ***Prefix Span Pseudocode*** 99
CHAPTER 4: DATA ANALYSIS AND FINDINGS 105
4.1 GENERAL DEMOGRAPHICS OF DATA 106
4.2 TOURIST SEQUENTIAL ACTIVITY ANALYSIS 108
4.3 SNAPSHOT OF TOURIST SEQUENTIAL PATTERNS 115
4.3.1 ***SEQUENCE 1: Most Popular Tourist Morning Activities*** 115
4.3.2 ***SEQUENCE 2: Most Popular Tourist Afternoon Activities*** 116
4.3.3 ***SEQUENCE 3: Popular Tourism Months To Visit Singapore*** 117
4.3.4 ***SEQUENCE 4: Most Frequently Visited Tourist Destinations*** 118
4.3.5 ***SEQUENCE 5: Most Popular Tourist Venues In Singapore- In Sequential Order*** 119
4.3.6 ***SEQUENCE 6: Female Tourist Sequential Activity Patterns*** 120
4.3.7 ***SEQUENCE 7: Male Tourist Sequential Activity Patterns*** 121
4.3.8 ***SEQUENCE 8: Places Mostly Visited by Male Tourists*** 122
4.3.9 ***SEQUENCE 9: Tourist Sequential Activity Patterns on Daily Basis*** 123
4.3.10 ***SEQUENCE 10: Tourism Receipts for Visitor's Arrival*** 124
4.4 SUMMARY 126
CHAPTER 5: DISCUSSION, RECOMMENDATIONS AND CONCLUSION 127
5.1 DISCUSSION 127
5.2 RECOMMENDATIONS 129
5.3 CONCLUSION 130
5.4 ACKNOWLEDGEMENTS 132
REFERENCES 133

Paper Submitted in IEEE Conference of Orange Technologies 2018 Bali

"A Study of Tourist Sequential Activity Pattern through Location Based Social Network (LBSN)"

LIST OF TABLES

Table 1.1: International Arrivals in Emerging and Advanced Economies …… 10

Table 1.2: Example of a Sequence Database ……28

Table 1.3: APRIORI Algorithms for SPM……28

Table 1.4: Pattern-Growth Algorithms for SPM……29

Table 2.1: Classification and Nodes of Activities ……65

Table 2.2: General Statistics of Guimaraes sequence……71

Table 2.3: Guimaraes Sequential Patterns ……71

Table 2.4: Berlin Travel Sequence Statistics ……73

Table 2.5: Sequential Patterns, Berlin with <L, W> Specifications ……73

Table 2.6: Data Collection Statistics of the Hong Kong Inbound Tourists ……77

Table 2.7: Areas of Interest by Popularity in Hong Kong ……78

Table 2.8: Spring Activity Hotspots…… ……92

Table 3.1: A Snapshot of Tourist Statistics……95

Table 3.2: Projected Database and Sequential Patterns …… ……102

Table 3.3: An Example of the Sequence Database ……103

Table 3.4: Illustration of a Sequential Database ……107

Table 4.1: Singapore Population Census Chart……110

Table 4.2: Diversification of Visitors in Singapore ……110

Table 4.3: Tourist Classification in Singapore…… ……111

Table 4.4: Food Preference of Singapore Tourists … ……118

Table 4.5: Tourist Morning Activity Patterns …… ……120

Table 4.6: Tourist Afternoon Activity Patterns …… ……121

Table 4.7: Popular Tourism Months …. ……122

Table 4.8: Popular Tourist Destinations … ……123

Table 4.9: Popular Tourist Venues ……124

Table 4.10: Sequential Activity Patterns of Female Tourists ……125

Table 4.11: Sequential Activity Patterns of Male Tourists ……126

Table 4.12: Places Visited by Male Tourists in Sequential Order ……127

Table 4.13: Tourist Daily Activity Patterns ……128

Table 4.14: Global Tourism Receipts ….. …………………………………………………….129

LIST OF FIGURES

Figure 1A: International Tourism Arrivals…………………………………………………...12

Figure 2A: Role of Destination Management ………………………………………………..44

Figure 2B: Travel Behavior Framework Analysis……………………………………………61

Figure 2C: An Illustration of the Multiple Behavioral Groups of the Sequential Activity-Location Patterns ……..………………………….……………………………………………..66

Figure 2D: The Novel Approach Framework ………………………………………………..70

Figure 2E: Movement Pattern Analysis of Asian and Western Tourists ….. …………………..79

Figure 2F: The Peak Tower Visitation Likelihood …………………………………………..80

Figure 2G: The Hong Kong Central Visitation Likelihood ….. ……………………………..80

Figure 2H: Prediction Accuracy of Heuristic Prediction Algorithm & Markov Chain Method 88

Figure 2I: The Seasonal Distribution of the Activity Points at the SNP …. …………………...93

Figure 3A: Data Collection from the Twitter API Interface …………….. ……………………97

Figure 3B: Raw Data Filtering …………………. ……………………………………………..99

Figure 3C: A Sample Tourist Check-in File …. ………………………………………………100

Figure 4A: Bar Plot of Gender of the Tourists……………………………………………….. 111

Figure 4B: Long & Short Trip Check-ins of Different Categories …. ……………………… 112

Figure 4C: Bar Plot of Most Popular Tourist Venues …. …………………………………… 113

Figure 4D: Results Generated through SPM using the Apache Software …………………… 114

Figure 4E: First 100 SPM Patterns in Graphical Format ….. ……………………………….. 115

Figure 4F: Group Matrix for the 100 Sequential Patterns …. ………………………………...116

Figure 4G: Sequential Patterns with Confidence, Lift & Support Values….. ………………...116

Figure 4H: Group Matrix for the 200 Sequential Patterns ….. ………………………………117

Figure 4I: International Tourist Arrival in Singapore … ……………………………………..129

Figure 4J: Tourist Arrival in Singapore - Forecast & Actual …. ……………………………..130

Figure 4K: Annual Tourism Receipts in Singapore …………………………………………..130

Figure 4L: Cumulative Tourist Interests in Singapore ……………………………………… 131

OUTLINE OF THE THESIS

CHAPTER 1: The first chapter provides a highlight of the sequential pattern mining process and how to develop the activity patterns from a given dataset. It also gives a detailed overview on sequences and frequent sequences including the algorithms and ethos used in the mining of tourists' behaviors. The chapter also captures the scope of the study and the limitations of the sequential pattern mining in a wide range of applications including tourism.

CHAPTER 2: Chapter two dwells on the past related studies on tourist mining behaviors across the world. Thereby providing an overview of the methods that have been used previously to develop a sequence of activity patterns in tourism research.

CHAPTER 3: It covers the methodology and the research design that has been used in the study. Primarily, it explores the Pattern-Growth Method and the SPADE Algorithm that was used in the generation of the sequential activity patterns of tourists.

CHAPTER 4: It encompasses the generation of the tourist sequential activity patterns, while highlighting on the breakdown of the data to become suitable for the mining process.

CHAPTER 5: It provides the interpretation of the sequential activity patterns including conclusions and the recommendations about the tourists mining behaviors in Singapore.

CHAPTER 1: INTRODUCTION

1.1 GENERAL

Tourism is considered to be a major contributor to a nation's economic growth and development. As compared to the past century, the tourism industry has expanded significantly not only in Europe but across the globe. Many people have become indulged into the idea of traveling and exploring new places including different cultural and religious practices. There are numerous tourism activities ranging from shopping, dining, sightseeing, hiking, recreation, visiting historical places among others. However, the range of tourist activities is different from one region to another. Moreover, due to the advancement in technology, there has been more creativity and innovations in terms of diversity in the range of activities.

Although tourists tend to make individual choices in terms of the time and place of destination, there ought to be a particular pattern that most of the tourists prefer more as compared to others. In regards, it is vital to understand the activity patterns of the tourists in order to make the right decisions in a bid to better the service delivery among decision makers. There are certain times of the year in which tourists tend to travel more as compared to others preferably in the last quarter of the year. In the first quarter, most of the regions in the country focus on individuals and business development as opposed to leisure activities.

- Transportation offices can traverse all methods of movement— expressways, flying, conduits, open travel, and railways.
- Any of these modes might be pertinent to the travel industry or then again recreational travel. Recreational offices can incorporate parks, arenas, brandishing offices, and shorelines.
- The travel industry destinations are attractions for outcasts and in addition neighborhood inhabitants and may incorporate recreational offices along with social attractions (for example chronicled, melodic or instructive offices). Any kind of recreational or travel industry office can have uncommon transportation needs.

The travel industry and recreational activities present numerous comparative travel contemplations, which regularly contrast from worker travel and business transport issues. The connection between the travel industry and transportation has originated principally from the idea of the travel industry as a generator of travel requests and transportation as the way of getting to real tourist destinations. Transportation can be a basic component of the task of visitor attractions and of supporting activities, for example, entryway networks to national parks. Successful transportation arrangements can likewise create suitable answers for adjusting the activity needs of various voyager bunches amid pinnacle travel industry seasons or unique holidays. These connections give a typical base of enthusiasm for transportation and the travel industry offices are in this manner the inspiration for interagency coordination. The way of tending to these normal interests (and their definitive usage) is the improvement of compelling procedures for coordination between different transportation offices, travel industry offices, other arranging associations, and private-part interests; to an extensive variety of the travel industry and diversion activities as "tourism."

Due to the deepening diversity, the global tourism market has grown significantly both in the developing and developed nations. According to the United Nations World Tourism Organization (UNWTO), the international tourists' arrivals across the globe is projected to reach 1.8 Billion by 2030. Notably, in 1950, the average number of international tourists' arrivals was 25 million which has increased to about 1.2 billion arrivals in 2015 [1]. From the UN conference held in 2012, tourism was poised to be an ingredient towards reconciliation of environmental and economic destination of the global community. By creating jobs and promoting intercultural dominance, the sectors foster economic growth.

INTERNATIONAL ARRIVAL FOR TOURISM

	1990	1995	2000	2005	2010	2014	2015
World	435	527	674	809	950	1,134	1,186
Advanced Economies	299	339	424	470	516	622	653
Emerging Economies	136	188	250	339	434	512	533

Table 1.1: International Arrivals in Emerging and Advanced Economies. (Source: UNWTO)

From table 1.1, it is evident that there is more movement of tourists in the advanced economies than the emerging economies. However, both economies exhibit an upward trend in the annual number of international arrivals from 1990 – 2015. In the future, the emerging economies are expected to receive twice the growth in the number of international tourists as compared to the advanced economies. Subsequently, the number of international arrivals are expected to double in Africa and the Middle East in the next decade. Similarly, an upward trend is exerted in Asia and Pacific countries. This is attributed to the willingness of the countries to invest in the tourism industry for example in the marketing of their attraction sites to the world. Also, people of this generation have become accustomed to the idea of 'traveling and exploring' new places other than their own countries.

Tourists are motivated by various factors towards traveling from one region to another. During the year 2015, more than an average number of international arrivals were motivated by leisure and holiday-related activities with only 14% being for professional and business travel. In regard, it is evident that the common purpose for tourist travel is for recreation and leisure activities. Therefore, it is important to understand their sequence of activities right from arrival to departure to be able to identify the most frequent sequential patterns. By doing so, it helps the relevant organizations including the WTO to better manage and improve service delivery in various destinations. As a result of strong economic ties and favorable exchange rates, China and the United Kingdom are leading in the numbers of outbound tourists between the two countries.

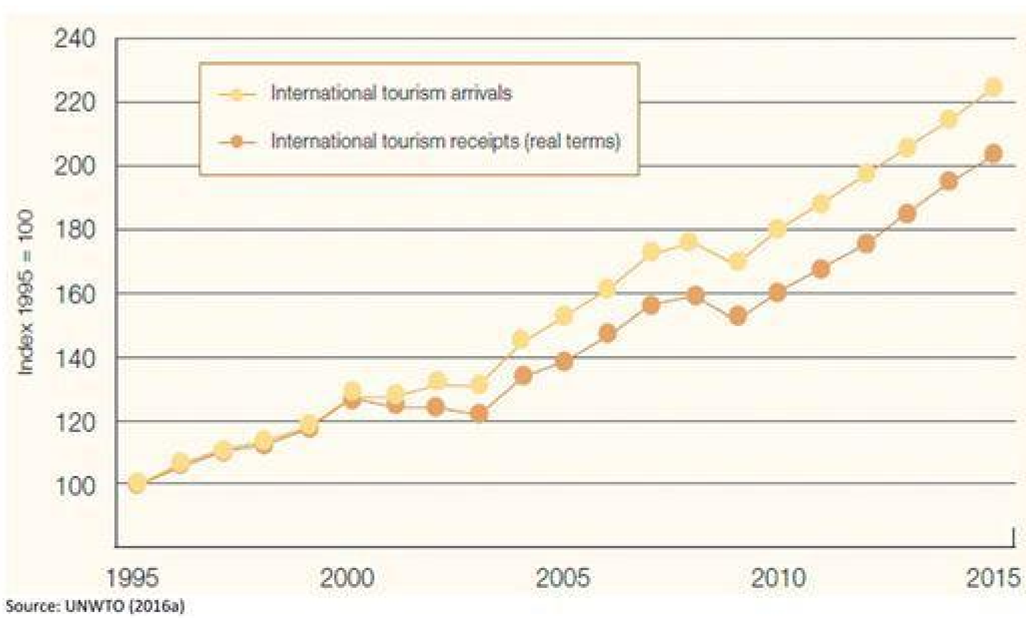


Figure 1A: International Tourism Arrivals

There are various factors that play a key role in determining the tourist movement patterns. The factors range from human factors, physical and trip factors which ultimately influence the shaping of the patterns. Some of these factors include destination terrain, weather, length of stay, motivation, types of individuals, place of origin, and travel party. The Geographic Information System is one of the instruments used in documenting the movement of tourists from one destination to another and is one of the commonly used methods. In the past, Lau has tried to engage in examining the inter-destination movement of tourists, however, this has been limited to confined boundaries [2].

1.2 AIM OF THE STUDY

Understanding tourist movement patterns within a destination has significant implications for tourism development as well as destination marketing. Past researchers, by mainly focusing on the spatial and temporal patterns of tourists with respect to different destinations, have been able to map and explain tourist behavior and movement patterns [3] [4] [5] [6] [7] [8]. There has also been an evaluation of most popular tourist trends and choices [9] [10] [11] [12] [13] [14]. The decision making process as well as the factors affecting tourist choice of destinations have also been studied in detail making significant contributions to the tourism literature [15] [16] [17] [18]. The importance of 'choice of activities' in the decision to make a trip and the role of 'motivational' factors for tourists and visitors has facilitated better Tourism across the globe [4] [19] [20] [21].

Despite the existing efforts, there has been very limited work on identifying the sequential activity patterns of tourists, which is very important to design travel itineraries and digital tour guides. A sequential activity for any tourist for e.g. tourist X, gives a detailed account of all the activities tourist X has been involved in a logical order or sequence. The extraction of sequential activity patterns can assist the tourism companies to market, price, package and promote their products and services to suit the tourist demands of activities. To the best of our knowledge, the prior work in this particular area has not been able to present an effective approach to capture and analyze the sequential activity of tourists. As such tourism managers still face difficulty in designing detailed travel packages which are most suitable for varied tourist groups.

The modern advancement in the internet and mobile techniques have made Location Based Social Networks widely accepted and used by tourists. A Location Based Social Network (LBSN) is a social structure made up of individuals connected by the interdependency derived from their locations in the physical world as well as their location-tagged media content, for example photos, videos and texts [22]. The key features of a LBSN include real-time check-ins and directions, friend finders, real-time tracking, location-based advertising, discovering places, and location-induced searches [23]. One of the main benefits of studying tourist trails from LBSNs is the richness of data, as it contains a lot of descriptive information about the venue, its type, comments and photos rather than the raw geo-coordinates of the location. It will be beneficial to utilize the LBSN data, especially the user check-in information, to capture and identify tourist activities at a particular destination.

Establishing tourist patterns and sequences, helps the management authorities to identify the potential areas of exploitations and focus. However, there is a need for the tourism sector in Singapore to lay focus on all components of the industry to achieve uniform success. In the modern world, the tourism patterns and sequences have changed significantly in the Asian countries, in this case: Singapore. This may be attributed to the market fluctuations and the rising cases of global insecurity. On the other hand, the drift in these patterns have also created newer opportunities in the industry. Ideally, the main aim of the study is to establish the significant insights from the analysis of the tourist sequential activity patterns using a design method on the LBSN data in Singapore. Also, the study will endeavor to evaluate the performance using the design process by exploring the sequential activity patterns of the various tourist destinations in the country.

The insights into tourist mobility and behaviors through various destinations are important for tourism managers to engage in strategic management and planning in a bid of building a sustainable industry. Ideally, both private and public tourism sectors are largely dependent on sustainable development in enhancing the tourist experience as well as maintaining the natural environments [24]. Therefore, the major aim of the study is to try and extract some of the frequent sequential activity patterns of the tourists in Singapore which are helpful for the management in the tourism industry. As a result of the unpredictable nature in the tourism industry, it is vital to understand tourist behavior and mobility. In the recent years, most of the tour managers and other relevant authorities have been seeking the insights of tourist travel behaviors and activity patterns with a view of attracting marketing, development of new products and management of destinations. Singapore is one of the vibrant countries in Asia in the tourist sector; hence it is important to provide an understanding of the tourist behavior by generating their activity patterns. For instance, to be able to plan good transportation systems, one needs to understand tourist preferences, transportation flows from one place to another as well as their daily itineraries. Through the sequential activity patterns, the authorities shall be able to properly draft an effective transportation system in the country.

The potential for the travel industry improvement is founded on an appealing characteristic and social legacy, nature parks, wine streets for social and recorded attractions, for example, mind-blowing parks. The rich and differing regular and social assets and a major tourist showcase in the quick region are additionally immense potentials for the advancement of provincial travel industry in many countries. The travel industry is a social and monetary marvel that vigorously impacts contemporary society. These days, the travel industry can be considered as business conduct since it may impact the improvement of a neighborhood financial. Along these lines, places are going up against each other to advance themselves as merchandise. The mystery for an effective destination is to approach the correct target showcase and to give a suitable blend of nearby travel industry items and administrations. Paresh [25] contends that there will be a clash of destination marketing in future, promoting that the destination is seeming to end up the travel industry's greatest brands.

Destination marketing associations for tourist throughout the world need to execute inventive and fitting methodologies and utilize sufficient devices and strategies so as to enhance their promoting exercises adequacy and effectiveness. The study effectively addresses the unpredictability in the field of destination advertising because of the different partners included and to the idea of the travel industry item/encounter. Along these lines, this study effectively proposes a reasonable system adding to enhance adequacy and effectiveness of exercises of destination advertising associations by embracing an incorporated methodology dependent on well-established theories.

Destination advertising is a sort of marketing for tourists that advances a destination (town, city, area or nation) with a reason to expand the number of guests. As it were, ‘destination promotion’ is the travel industry publicizing for an explicit area. Dissimilar to item marketing, where the items are conveyed to users through circulation channels, in destination promotion purchasers travel to the destinations. To advance a town or a city, focus on tributes administrations, the potential tourists might want to check references concerning settlement offices, exercises, and the area accessibility. Furthermore, the most helpful instrument to advance a district will be web crawler promotions and online life (focus on social and normal attractions, locale quirks and features). The destination is an expansive substance with sets of material and non-material components. Each destination is special since it's assets develop a novel 'distinguishing proof'. This 'recognizable proof' is for the most part perceived as the pictures anticipated to tourists. The pictures offered by an explicit destination might be like others however never the equivalent. Procedures of building 'recognizable proof' impact the destination pictures and in this manner will impact its future planning; as far as possible, they will influence destination improvement. Components of a destination are the establishment of destination; they produce the 'ID' and lastly develop the destination picture. What is important here is the manners by which the destination produces the 'recognizable proof' or as it were the methods for marketing. Destination marketing can be characterized as an approach to impart a destination's one of a kind personality by separating a destination from its rivals. Also, the activity patterns and movement information are useful in identifying the barriers and the hindrances of the flow between accommodation and the attractions sites. Marketing is one of the important aspects of the tourism industry. In regard, the information of activity patterns and behavior of tourists helps in the development of proper advertisement products, redefining the existing attraction sites and being able to present the ideal products to the potential tourists [24]. The undertaking of the actual routes taken by the tourists aid in the redefining of boundaries as well as pointing out some of the appropriate gateways. Subsequently, the information about tourist activities assists in the development of new products and attractions including - providing the contrast between destination and district nodes in the region. The understanding of the space and time characteristics of the frequently visited destinations, helps in preventing product capacity overload hence warding off the negative social and cultural impacts in the tourism industry.

Destination over the world vigorously rival one another, so as to keep up their appeal and aggressiveness in the worldwide visitor industry. As such, it is vital for destination experts to have the capacity to address the distinctive needs of various market sections, and additionally advance their picture and oversee destination in a way that draws in tourists. At the end of the day, they have to successfully actualize 'Destination Marketing', the term referring to advancing traveler with the destination as a method for enhancing their symbolism and notoriety [26]. As indicated by Koutoulas and Zoyganeli [27], destination marketing for tourists happens in two dimensions. At the smaller scale level, free visitor administrators, for example, inns and transportation organizations, which advance the items and administrations they offer in the business. At the full-scale level, governments and other authority experts advance their nations and states as tourist destinations.

Destination Marketing basically includes the idea of Destination Management for tourists, which refers to every one of those endeavors made towards advancing economies, the travel industry and the enthusiasm of partners. For that reason, the accompanying two sections are of 'tourists as the key determinants of destination marketing' and 'destination as a marketing executive' respectively.

The key determinant of destination marketing is an issue which has been comprehensively talked about in the scholastic writing. It has been examined the components which decide the achievement in a visitor destination. For that reason, the creators disseminated polls in an example comprised of 100 tourists who had picked Singapore as a visitor destination [28]. Moreover, the analysts utilized factor investigation and auxiliary condition demonstrating strategies for breaking down their essential information. The after effects of this paper demonstrated that destination marketing proficiency is impacted by four variables, in particular: a) fulfillment of the movement cost; b) the incorporated travel industry item; c) the travel industry item properties and d) the travel industry item board. These elements are dictated by a few characteristics. In any case, the travel industry item is detailed by the fulfillment of the tourists from the ocean, the sun, the shorelines, the mountains, the inns, the commercial centers and the eateries of a destination. The travel industry item executives are controlled by the attractions, the civilities, the gets to and the picture of a destination for tourists. Finally, the fulfillment of the movement cost of tourists is controlled by the carrier cost, the inn and the guesthouse cost, and the aggregate expense of the residential excursion in Singapore.

Moreover, Buhalis [29] recognizes three key bearings that can upgrade destination marketing effectiveness for tourists that can improve the fulfillment of tourists and pleasure the guests and give benefit of the nearby travel industry and of the neighborhood - little and medium-sized travel industry ventures, and to build up the maintainability of the destination for tourists and services of money tourism on quality by focusing. Every one of these three bearings join a few vital targets. All the more special to upgrade the fulfillment of the guests, destination and the travel industry undertakings ought to enhance their administrations, practice their travel industry item and offer esteem - for - cash the travel industry benefits by concentrating on quality.

1.3 THE SCOPE

Being one of the major contributors to the economic growth and development, tourism management is the center of focus in trying to understand the tourist behavior and mobility. In the past, several scholars and researchers have formulated the methods to understand the behavior of tourists in a bid for improving service delivery. In figuring out the sequential patterns while identifying the most frequent patterns from the tourist activities; one is able to make a forecast of the impending situations hence aiding towards the preparations. For example, 'Hiking > Dining > Entertainment', it is predictable from the visit that majority of the tourists would do hiking first then entertainment thereby the host countries through the management authorities will be able to prepare adequately. Past studies have only focused on the frequent sequential patterns of the tourists with regard to their movements from one place to another. However, this research study will be unique since it will do the activity patterns with respect to time constraints. Therefore, the research study is considered to be more specific as compared to the past general studies, since it gives an actual account of tourist activities at the particular destinations. Notably, the focus will be in Singapore, a country in Asia where we will develop tourist sequential activity patterns with regards to check-in times on Foursquare - a social media platform.

1.4 RESEARCH QUESTIONS

1. What are the most common tourist sequential activity patterns?
2. What are the effective methods of sequential activity analysis using the LBSN data?
3. What are the most interesting insights resulting from the tourist sequential activity pattern analysis?

1.5 RESEARCH OBJECTIVES

1. To design a new method to effectively process and analyze LBSN data for sequential activity analysis.
2. To evaluate the performance of the proposed method in exploring sequential activity at several popular tourist destinations.
3. To offer insights about tourist sequential activity patterns to tourism managers and policy makers for designing better packages in tourism marketing and destination management.

4 To use a new approach to capture and analyze tourist activity patterns, thereby giving new insights into tourist behavior at the destination in the form of activity sequences which has not been captured in the previous studies. This can also aid in the discovery of preference differences in terms of anticipated activities at different destinations, by comparing different tourist groups.

5 To provide a quick and efficient way to capture tourist activities at destination without the need of direct contact.

6 The findings and recommendations will be useful for tourism management tasks, such as customized tour packages, tour guiding and making tour recommendations to suit tourist demands.

1.6 THE LIMITATIONS

In as much as there are the benefits of sequential pattern mining in deriving the frequency subsequences from a given database, there are also challenges that are associated with the technique. These obstacles may occur during the mining process at the interpretation stage where the applicability of the results may offer a false representation of the reality. However, the method is chosen or algorithms used to facilitate the mining of the tourist activity patterns, each of them with its own pros and cons. Although the sequential pattern mining problem has been studied for the past two decades, there are still some challenges as well as opportunities that come with the concept. Generally, there are many challenges faced by big data mining especially in the tourism industry which has millions of records of the tourists visiting various places around the world [30]. The study ought to cope with these challenges since it has not been fully addressed. Nonetheless, there are major improvements in most of the methods used in sequential pattern mining. Tourism is known to be the largest contributor to the economic growth and development hence large databases and datasets are highly expected to be involved as in the Singapore case.

Some of the challenges of sequential pattern mining which are the limitation of the research study include: data inconsistency, data incompleteness, timeliness, security, and scalability. Not only are these challenges noticed during the mining process but also in the data analysis and sequential pattern development stage. In tourism research, there are numerous complex sequence data attributed to the influx in the numbers of tourists visiting one place or another. Multi-model data is also another limitation which is a common feature of large dataset [30]. Data fusion and multimodality are potential challenges that ought to be dealt with first before rolling the sequential pattern mining process. Multi-model data are difficult to deduce or rearrange to fit a particular manager to be able to make it easier for the algorithm execution. The sequential data in any given application, in this case, tourism research, is considered to be dynamic rather than static which makes it more complicated. Online progressive mining and incremental mining are some of the proposed ways of dealing with dynamic datasets. The SPM algorithms are easier to implement in small datasets as compared to the bigger ones. The data to be used in the research study is huge having various factors including check-in times and venues categories hence making the mining of the tourist behavior through activity patterns a bit tedious. In order to adequately execute the SPM algorithms, there are also some levels of required scalability that ought to be reset in the data.

Privacy is another major challenge since most of the tourists want their details and information about the places they have visited to be kept confidential. It is a concern that data management and data mining is faced by the risk of leakage of information by unauthorized individuals. The evidence is as the community of global transport which moves further into a digital environment, facilities are increasingly connected to, and dependent on, cyber systems. Security is an important aspect of data mining. As industries worldwide have turned towards greater reliance on cyber-systems, organized crime, state-sponsored hackers, terrorists and other malicious actors have turned towards exploiting weaknesses in cyber-security to gain intelligence, facilitate illegal activities and cause economic and physical damage. In the recent years, the concern about the public privacy has been on the rise with the majority of people being sensitive on what they share through social media which forms the major source of the tourism data. Most of the platforms put emphasis on user protection in terms of policy, making it difficult to access the required information about the tourist movement from one place to another including the time of visitation. The researcher and scholar who engage in the management of data mining process are required to uphold the privacy of the users with technologies including algorithms being developed to ensure presentation of the user privacy [30]. The use of SPM to use personal data requires utmost presentation of privacy. However, the use of parallelism in the sequential data mining makes it challenging to address the privacy concern fully. It is important to keep the patterns simple and short as well as develop hardware and software that seek to address these challenges in the near future.

Another potential limitation to the study is the focus on the international tourists based on the different locations in the dataset provided. The implication is that the local tourists of Singapore may feel neglected hence demoralizing the potential domestic market which is deemed to be the epicenter of the tourism sector in the country [31]. The native population ought to feel proud of their treasures, therefore, they ought to be included in the bracket by the tourist sector in Singapore in their quest in establishing the tourist patterns and behaviors altogether [31] [32]. On the other end, the constant fluctuation of the tourism market and demand for attraction sites makes the created tourism patterns short-lived. In essence, the tourism management may not be able to rely on the patterns obtained from the analysis for quite a long time owing to the periodic changes in the industry.

By breaking down large datasets, sequential pattern mining helps in extracting unknown interesting patterns or sequences of groups of records. However, the use of association rules to develop frequent patterns of the tourist activities has failed to handle the real dataset having timestamps [32]. In regard, this has been one of the setbacks of the use of sentential mining in the study. Most of the algorithms, for example, the SPADE requires strict periodicity in which some of the datasets of the study may not provide hence will undermine the outcomes of the sequences of the tourist activities.

The Sequential Pattern Mining is an essential segment in building up examples and mining patterns of certain exercises for tourist activities. Earlier, this strategy has been utilized in different fields, for example, purchaser watch, making future forecasts and breaking down and deciphering expansive datasets for profoundly installed principles and affiliations. The subjective subtleties of Singapore visitors' Foursquare registration, spoke to in an unthinkable frame, is a case of a successive database. Hence, the Pattern Growth technique which utilizes Prefixspan Algorithm is utilized in this examination to acquire the Tourist Sequential Activity Patterns.

The knowledge into visitor development and movement designs is considered helpful for the travel industry area from multiple points of view, for example, structuring better travel bundles for tourists, augmenting the tourist movement support and taking care of the visitor requests. This examination proposes to receive versatile internet-based life information for powerful catching of visitor movement data in Singapore and uses mining systems of data for separating important experiences into visitor conduct. The proposed techniques and discoveries of the study can possibly bolster the travel industry chiefs and approach producers in settling on better choices in the travel industry destination for tourist activities.

1.7 SEQUENTIAL PATTERN MINING

In essence, in any given data, there is a high possibility of the activities, events or numbers that follow a particular order. However, the challenge is to decipher the frequent subsequences of those events or activities with respect to time which forms the epitome of our research study, a case of Singapore tourists. Different methods and algorithms have been proposed for sequential pattern mining, a field of data mining. Each of the approaches used has their underlying constraints and features which fit respective kinds of datasets. Notably, the datasets used for sequential pattern mining are not only limited to longitudinally maintained datasets. Also, data from weblogs can be used in mining. In sequential pattern mining, the algorithm development has been largely focused on the specific domains of datasets such as biotechnology, spatial and telecommunications. This helps to maximize the outcomes in terms of pattern generation and ensure that the output can be relied upon for decision making. Often, the resulting output of sequential pattern mining is used for decision making.

The sequential activity patterns can be represented in notation as follows;

Let $\mathcal{E} = \{a_1, a_{2,} a_3, a_4, a_5, \ldots\ldots\ldots\ldots a_k\}$ be a set of items or literals which is referred to as an event

Without the loss of generality, it is assumed that items on the set are arranged in lexicographical order. An ordered list of an event is referred to as a sequence. Notably, an E-sequence has E-items.

A sequence is denoted by $\beta = \{\partial_1, \partial_2, \partial_3, \partial_4, \ldots., \partial_n\}$ where ∂_n $is\ an\ event.$

A typical database of sequences contains attributes such as id, event time and the list of items. The support or frequency is a measure of the number of occurrences of the activity pattern in a given database. Based on the algorithm, a different method is used in calculating support of a sequence. In a given super-threshold, support is the number of proportions of occurrence more than the set minimum support in the database.

The algorithm used is dependent on the type of dataset to be used in the generation of the activity patterns. In the past, data mining algorithms have been utilized to generate sequential activity patterns by different scholars [33]. Ideally, these algorithms are categorized into either by the designation of the broad class they belong to or by the type of the dataset they subscribe to. The APRIORI family of algorithms uses transactional databases as their source of data. From the datasets, they derive the inter-transactional associations using the generated rules. In essence, the algorithms work on the datasets by forts generating association rules which are then used in the formulation of the activity patterns. Each of the rules has its own set of constraints and characteristics hence the development of different activity patterns. However, there is the possibility of having repeated patterns from the same dataset.

Using a breadth-first approach, the APRIORI algorithm works in five systematic phases namely:

- Sorting: the introductory phase helps in converting the original database into a customer-oriented sequence database through sorting of fields.
- Large item set: all item sets having met the minimum support criteria are grouped together to enhance the optimization of future comparisons.
- Transforming: in this phase, a sequence that does not have large item sets are dropped from the database. However, the dropped sequences still contribute to the overall count of the number of sequences in the database.
- Sequence phase: It involves the discovery of frequent subsequences. Similarly, those that do not meet the threshold of minimum support are dropped.
- Maximal: Among the large set of sequences, this phase involves the identification of the maximal sequences. The process may be equivocated to the idea of finding all the subsets in a given large item set.

In further categorizing these algorithms, they fall into two forms namely: horizontal database format algorithms which harbor the APRIORI algorithm family and the vertical database format algorithms. An example of the vertical database format algorithms is the Sequential Pattern Discovery using the Equivalence classes (SPADE) algorithm including the constraint SPADE which is the variant of the former. In mining sequence patterns, the vertical database format algorithms employ the depth-first approach then later the Pattern-Growth methods. The shift in the approach calls for the organization of the dataset in an alternate manner where rows having an object-stamped pair are given priority hence making it easier in generating the activity sequence patterns of the events. The shift in the data layout helps in the focus on the constraints in the mining process.

The SPADE algorithm uses the lattice-based technique which allows constraints to be incorporated in the mining process. The vertical id-list is one of the prime feats of the SPADE algorithm in sequential pattern mining. The CSPADE is almost similar to the SPADE algorithm but it has some additional syntactic constraints involved during the mining processes. These constraints include:

- It has width and length limitations of the sequences especially for highly structured data;
- Minimum or maximum gap constraints on every consecutive sequence pattern;
- It applies time window on the allowable set of sequences;
- Incorporates item constraints;
- Identification of sequences having predictive classes.

On the other end, the SPAM (Sequential Pattern Mining using the bitmap representation) includes various pruning mechanisms with the application of the depth-first method of mining. The use of bitmap representation enhances support counting while scanning the database for the first time. As illustrated every algorithm has their own constraints which define the way in which the sequential patterns are generated. Also, each of the applicable algorithms and methods of sequential pattern mining has their respective setbacks and strengths. One of the limitations of the SPAM algorithm is that it insists on all data fitting into the memory before defining the process of mining. Sequence extended is one of the ways in which it ensures that the extra candidates are accommodated in the memory. By using the depth-first method, it ensures that all the nodes are visited during the pruning of the candidate's phase. Each of the proposed algorithms adopts a different approach while executing the sequential pattern mining process.

The Cache-based Constrained Sequence Miner (CCSM) algorithm employs the level-wise approach. It involves the use of k-ways intersections to complete the required candidates for the process. Besides, it engages a cache which is used to store intermediate id-lists for future usage. The algorithm is equivocated to the SPADE because of the ability to visit sequential patterns in a tree database. However, the major difference is that the CCSM algorithm uses a cache to store intermediate lists hence speeding up the process of support counting [31] [33]. In regard, the CCSM algorithm is considered to be more reliable and flexible for mining process as compared to SPADE. Notably, various comparisons have been made on the algorithms but it depends on the individual's objectives coupled with the constraints at hand to be able to make a decision on the algorithm to use in the mining of the sequential patterns of a particular dataset or database. Ideally, the purpose of the algorithm is to scan a given dataset and identify distinct sequences as well as the frequent sequences which are facilitated by the use of support and confidence figures obtained from the computations.

Support and confidence are considered to be important attributes for sequential pattern mining. If ABC is a frequent sequence then activity A happened first then followed by B then C which occurs at a later stage. Therefore, the sequence is represented as A >B > C. Similar to association rule mining both support and confidence are formed and applied in X > Y where Y and X are continuity patterns of windows w1 and w2 respectively. In order to achieve proper sequence patterns, it is important to inquorate all constraints and follows all the required steps systematically during the mining process. The performance of the algorithms can be affected whenever the support is low. In regard, it is vital to use test technique as well as the candidate generation procedures.

The constraints employed by the majority of the algorithms have not only been applied in sequential pattern mining but also in association rule mining. Regardless of the algorithm used, the constraints applied in sequential mining can be played into eight categories namely: item constraint, length, model-based, aggregate, regular expression, regulation, gap, and timing marks. Sequential patterns can not only be contained with timestamps but also with the timing marks for specific tokens of events [31] [32] [33]. The gap constraint is identified by the time difference between the two consecutive events or items. On the other end, the duration constraint is defined by the time difference between the first and the last items or between the first and the last transaction taken. In all the algorithms having a temporal database as its source of data, the duration constraint is executed in a sliding window. The aggregate constraint involves the use of aggregate functions such as average, minimum, maximum, range among other related functions hence the generation of sequence pattern according to the command call. The length constraint provides the specification of the length of the patterns that are intended to be mined eventually while on the other end item constraints indicated the type of items to be dealt with during the mining process.

Single-node techniques have been adopted by most of the traditional algorithms because they are designed to run on a single computer to provide the required sequential activity patterns in the end. In recent decades, data mining has been extensively used across various disciplines with the algorithms having the ability to generate meaningful patterns that can be used to make decisions and choices. Notably, most of the data mining methods and algorithms have limited memory capacity and heavily centered on single node computations hence are unlikely to be efficient on big datasets [31] [32] [33]. However, involving parallel technique in the mining process helps to make use of the process more flexibly and efficiently regardless of the size of the data used. The parallel data mining technique uses a form of computer architecture to execute several commands simultaneously. Ideally, the technique distributes computation evenly over multiple nodes as opposed to the reliance on the single node execution [34]. Due to the advancement in technology, the application of the technique in sequential pattern mining has been realized. Sequential Pattern Mining has largely become an important task in the data mining field having accrued numerous real-world applications. It involves the process of identification of the most interesting, meaningful and unexpected patterns in databases or datasets. To begin with, the APRIORI algorithm was the first one to be proposed by both Srikant and Agrawal in the 1990s. The algorithm was primarily used to identify frequent datasets while extracting association rules used in the development of sequence patterns from the data. In the event that the sequential ordering of the data is required, association rules mining technique helps in analyzing the data giving meaningful results for decision making [44].

The concept of pattern mining led to the formulation of the sequential pattern mining with a view to addressing the ambiguity in the patterns generated from the databases. The sequential pattern mining uses a time constraint to create subsequences from a set of sequences of items or events while providing statistical relevance to the data. A sequence is termed to be frequent if it meets the user-specified threshold of minimum support and it occurs more than the minimum set times of the processed database. Since most of the traditional algorithms were only limited to small datasets, the idea of parallelism assisted in addressing the problem. By doing so, large amounts of data were stored and distributed to multiple sites for processing [30] [32]. Notably, it is quite expensive to accumulate large amounts of data and perform the sequential pattern mining process in a single computer hence the need for the adoption of parallelization concept. The characteristics of parallelism method include the assigning of portions to pace the search of separate processes, thereby distributing the dataset over the available processors in an organized manner.

In the modern world, parallel computing has been utilized to enhance the performances of a computer program, especially in the data mining field. In the data mining field, parallelization is achieved through the following components: horizontal versus vertical data layout, static versus dynamic load balancing, and distributed versus shared memory systems. In recent years, most of the data mining platforms and tools such as grid computing and multicore computing have incorporated parallelism technique to realize high achieve performances. Both frequent item mining and sequential pattern mining has embraced the new innovations hence being able to handle the data efficiently. An example of a sequence database is illustrated in the table below:

SID	Sequence
1	$< \{a,d\},\{c\},\{d,g\},\{g\},\{e\} >$
2	$< \{a\},\{d\},\{c,g\},\{e\} >$
3	$< \{a,b\},\{c\},\{f,g\},\{a,b,g\},\{e\} >$
4	$< \{b\},\{c\},\{d,f\} >$
5	$< \{a,b\},\{c\},\{d,f,g\},\{g\},\{e\} >$

Table 1.2: Example of a Sequence Database

The table has five sequences of different length and width constraints. It means that the outcome of the sequential pattern mining process results in sequences of varied lengths.

In comparison to frequent item mining, sequential pattern mining is considered to be more complex due to the absence of time constraints in frequent item mining. Besides, sequential pattern mining has the potential of accruing huge candidate sequences hence the possibility of a large number of the sequential patterns at the end of the process. In essence, frequent item mining is more suitable for small data while sequential pattern mining is suitable for large or big datasets.

APRIORI algorithms for Sequential Pattern Mining

Name	Description	Year
APRIORI	The first algorithm for sequence pattern mining	1995
GSP	Generalized sequence patterns	1996
PSP	Retrieval optimizations more efficient than GSP	1998
SPADE	Sequential Pattern Discovery using Equivalence classes	2001
SPAM	Sequential pattern mining with bitmap representation	2002
LAPIN	SPM with Last Positions Induction	2004
LAPIN – SPAM	Last Position Induction with sequential pattern mining	2005

Table 1.3: APRIORI Algorithms for Sequential Pattern Mining

Pattern-Growth Algorithms for Sequential Pattern Mining

Name	Description	Year
Free Span	Frequent pattern-projected sequential pattern mining	2000
WAP-Mine	SPM suffix growth	2000
Prefix Span	Prefix-projected sequential pattern mining	2001
LP Miner	Sequential pattern mining with length decreasing support	2001
SLP Miner	Sequential pattern mining with length decreasing support	2002
FS-Miner	SPM with suffix growth	2004
LAPIN-Suffix	SPM with suffix growth	2004
PLWAP	SPM with prefix growth	2005

Table 1.4: Pattern-Growth Algorithms for Sequential Pattern Mining

Based on the analytics, the pattern-growth algorithms perform better as compared to the APRIORI family of algorithms. On the other hand, hybrid algorithms incorporate different technology to improve communication and reduce memory usage hence the hybrid method also have better performances. However, each of the categories has its advantages and disadvantages.

1.8 TOURIST SEQUENTIAL ACTIVITY PATTERNS

Today, the knowledge of the sequential activity pattern analysis is being applied in the tourism industry although initially it was meant for the manufacturing industry. Its relevance in the tourism industry is to understand the activity patterns of tourists with a view of better management and delivery of services. Tourists tend to follow a particular pattern while visiting new places. There are numerous methods that can be used to decipher such activity patterns based on individual movements and engagements with respect to time. In the past, the Euclidean distances were used precisely to provide a measure and comparison among activity patterns [35]. Notably, there are underlying mechanisms that ought to show the correspondence between the observed activity patterns and the predicted activity patterns. In this study, the focus is on the observed activity patterns of the tourists as we seek to unearth the sequential activities of tourists visiting Singapore.

Space-time semantics also play a key role in realizing human mobility patterns inferring what is happening at certain places at a particular time. Apart from time and space, there are other factors that contribute to the formulation of activity patterns of humans. The understanding of the sequential activity patterns has had wide applicability in forecasting, emergency response, urban planning, and also in environmental conservation. Mostly the application is biased towards the decision-making phase where the information about certain human activity patterns helps in making decisions in factor or against a particular situation.

Social media plays a key role in providing the data and information about activities taking place in a particular time. However, there is a need to guarantee the confidentiality of the users' personal and sensitive information. Majority of the tourists tend to post images or reviews on social media about their travel behaviors from one place to another. Due to the advancement in technology, the images capture the exact time and date at which they were taken. This information is frequently used in the formulation of the sequential activity patterns in tourism research. Subsequently, more focus on tourism research has shifted in the analysis of the tourist behavior which is crucial in the decision-making process [36]. The travel behaviors are largely associated with travel motivations as well as social and demographic characteristics of the intended region of visit.

In understanding the tourist activity patterns, different activity-based approaches are involved namely: simulations process, econometrics, mathematical programming, and constraint based. The latter relies on activity programs with respect to certain constraints such as the working hours in a restaurant. The econometrics approach perceives the activity patterns as the outcomes of choices made by the individuals' activity doer. For instance, if a tourist opts to go hiking in the afternoon rather than in the morning, the choice made adopts an econometric approach in determining the eventual activity patterns of that particular tour. In regards, the econometric models are widely used because of their behavioral foundation and friability. On the other end, the simulation process acknowledges the scheduling of activities as the sole determinant of the individual activity patterns. It relies on a stochastic process which follows a periodic time occurrence of events hence the simulation process results in predictive activity patterns. With the simulation process, the individuals involved are required to stick to the stipulated activity schedule. Under the activity-based framework, the mathematical programming approach is dependent on the combined methods to decipher the respective activity patterns. Moreover, space-time factors and constraints are also a figure in the mathematical programming approach. A given program is fed factors of time and space including other constraints which then yields the activity patterns upon execution. In formulating the sequential patterns, time allocation and activity participation are taken into account [37]. These factors play a key role in the sequential pattern mining which forms the basis of the research study.

The adoption of technology in the tourism industry has resulted in e-tourism where tourists rely on technology to explore various destinations using mapping before making a choice of visiting the place. Sequential pattern mining relies on technology to capture the trends and the behaviors of tourists from one location to another hence the possibility of deriving patterns from the data. Today, tourists are able to plan trips, make reservations, and have an overview of the intended place of visit through the web. It is through these engagements that data is gathered of time, location, and activity which are used to generate the activity patterns respectively. However, due to the increase in the cases of online fraud, it is difficult to find the required information in terms of credibility since one may not be able to identify the genuine tourists. Furthermore, issues of cyber insecurity and hacking have greatly undermined the industry and generation of reliable activity patterns of the tourists in the industry. Also, poor network coverage in some regions hinders the tourists from posting the images which capture the required inputs of pattern mining.

In the modern world, the information about the physical activities, time, location, and media consumed by users can be recorded in mobile phone devices or related applications. It is from the collected information that the generation of the tourist activity patterns occurs. Consequently, these patterns are used to make recommendations accordingly and where required. Besides a given tourists activity pattern, one is able to make predictions about the extent to which the person is going to perform. In assumption, people tend to repeat similar patterns from time to time in comparing the first and second visits. The contents of a certain activity will infer the next activity in which the user is going to get involved. For example, visiting the garden, dining, listening to music is a concise tourist activity pattern which can be inferred. The activity pattern is derived from past activities. In this case, an algorithm is used to generate these tourist activity patterns. Several algorithms have been recommended but they all depend on the attribution to time, location, and activity name among other factors to deliver since outcomes.

Prior research has utilized different techniques in a bid for generating and understanding sequential activity patterns. However, these efforts have not been of much help in the understanding of the tourist behavior. Tourist market segmentation is one of the methods or techniques that have been used in the development of tourist sequential activity patterns [24] [25] [27]. Also, the Origin-Destination travel flow model has been used in understanding the travel patterns of inbound tourists in Australia. In essence, various methods have emerged and been utilized in the development of tourist activity patterns across the world. Notably, each of the method or technique has their own bottlenecks and advantages. Similarly, social network analysis using online travel diaries has also been used in the understanding of tourist mobility and travel behavior. Nonetheless, there are some common barriers that hinder the execution of these techniques in the generation of the sequential patterns. As a result, both tourism managers and researchers have found it difficult to capture and decipher the travel behavior of international bound tourists. The ‘travel preferences’ and the ‘information capture’ are the two difficulties affecting the mining process of the frequent sequential patterns. The methods used in the capturing of the data and information is limited in terms of their scale and time consumption. On the other end, it is difficult to have control of the individual preferences of the tourists which tends to change from time to time hence greatly affecting the mining process of the tourist sequential activity patterns. Different tourists from various countries have different tastes and preferences which are in-line with their tourist activities. Most of the techniques do not take these preferences into consideration which affects the ultimate outcome of the tourist activity patterns and behavior.

The advancement in the technology for both mobile and multimedia has enhanced the mining process of the tourist sequential activity patterns. In regard, large volumes of the user-generated data are being captured including travel photos. Most of the mobile applications in the modern world have the ability to capture useful information such as time, location, longitude and latitude thus making the generation of the tourist activity patterns easier. Majority of the photo-capturing devices such as tablets and mobile phones have the built-in Global Positioning System (GPS) which is capable of recording the geographic information to be used in the analysis and understanding of the tourist behavior. The Flickr platform offers the opportunity to extract this geographic information including the time constraint. The application has been effective in the extraction and the generation of the tourist activity patterns. It is considered to be a valuable data mining resource, especially in the tourism industry. Since the introduction of the GPS, the exploration of the mobility patterns of the tourists have been enhanced hence the ability to identify multiple destination travel behavior of the tourists across the globe [24] [26] [27]. Both Global Positioning Systems (GPS) and the Global Information Systems (GIS) have been used together in the analysis of the travel behavior. Also, the technologies have facilitated the ability to compare and contrast the activity patterns of the new and repeat tourists in particular destinations. With the advancement in technology, the future of the tourism industry is bright in terms of the exploration and extraction of the tourist sequential activity patterns. The advances in the information technology have made the capturing of the information and data of tourists easier thereby enhancing the sequential pattern mining process.

1.9 SUMMARY

The growth of tourism sector has been attributed to many factors alongwith the ease of travel, bookings, accomodation and a variety of platforms to assist with individual requirements. Tourism has also been an important factor in determining the global marketing trends and choice destinations. Thus, the section has outlined the importance of tourism in determining the tourist popular trends as well as activities. The research aims to make significant contribution to the tourism literature by highlighting the popular choice of activities and the role of tour managers in fulfilling and maximising tourist growing demands, which will benefit both the consumer (tourists) and supplier (tour managers) at the same time. The purpose of using LBSN dataset is to capture real-time data in a precise, succinct and efficient manner which will in turn give accurate results and findings on tourist activity trajectory and give a complete insight into tourist movement patterns. The Sequential Pattern Mining Algorithms mainly SPADE and SPAM will be applied to the processed dataset in order to draw out interesting activities patterns and tourist trends. The tourist Sequential Activity Patterns will then be discussed and anaylsed in detail giving further recommendations.

CHAPTER 2: LITERATURE REVIEW

2.1 INTRODUCTION

There has been a lot of research in the tourism sector to study tourist behavior in detail along with the factors contributing to their individual and collective choice decisions. Questions pertaining to tourist trends, choices and behaviors have gained much importance in the tourism literature, such as “How do tourists behave?, Which factors affect tourists’ choice of travel?, What are tourists’ travel preferences?, Which activities do they like to indulge in while visiting a particular destination?, and How do they plan their trip?” Over the past three decades, there have been inconsistent and spatial tourist activity patterns across the world and especially the Asia-pacific countries [38]. However, in the last decade, the steady but rapid growth of both inbound and outbound travel was noticed by the Pacific Asia Travel Association. In the assessment of the tourists' spatial patterns and flows in the early 1960s, it was established that these flows happened as a result of the political and economic prosperity among the Asian-Pacific countries. But owing to the rapid growth in the tourism sector, the researchers have become interested in the tourism flows in terms of nature, patterns, and intensity. There are two approaches that influence the state of the touristic patterns and activities in any given region. The political-economic approach is often unidirectional and relies on economic situations to define the sequential patterns [39]. On the other end, the supply-demand interaction approach shapes touristic movement as well as consumption based on individual and collective preferences in a given tourist generation. For example, places with more tourism resources are poised to attract more tourists hence influence the patterns.

With the quick improvements in data and correspondence innovation (ICT) and in the period of globalization, the physical obstructions of reality have been disposed of with the initiation of the Internet and the World Wide Web. This has enabled business associations to completely misuse the potential focal points offered by this new innovation, permitting them to associate with their shoppers on the web and advertising them colossal items on the web. Further improvements in ICT devices have likewise lead to the improvement of online networking systems, websites, RSS channels, smaller scale sites and wikis, which have enabled business associations to straightforwardly draw in with their buyers. Through these correspondence openings, numerous brands have associated with their shoppers straightforwardly to enhance their efficiency, execution, and productivity. This methodology has likewise been received by the travel industry for tourist activities as well. Research has investigated the accomplishment of usage of marketing efforts and demonstrated the estimation of brand creation in an endeavor to make mindfulness and pull in more tourists to a particular region of choice. It additionally discovered the money related resources and the aftereffects of marketing activities taken so as to demonstrate its prosperity and also mindfulness among tourists of the household. The point of the exploration is to examine the travel industry capability of the two regions and its achievement in drawing in tourists, day by day visitors and getting to be known and prevalent residential and local as the travel industry destination. The philosophy utilizing in this exploration comprises of quantitative and subjective research, auxiliary research, strategy for investigation and also the perception from genuine because of creator's involvement and working with the two areas in advertising and advancement.

Auxiliary information investigation is made through broad research of writing audits and articles composed and distributed on the point of destination executives, advertising of the travel industry destination and by investigating records and limited time locales of the two areas in the mainland and rustic travel industry. There is an additional audit of insights that demonstrate the number of visitors, most visited and famous attractions, the number of medium-term stays and everyday visits. Quantitative information investigation was led among the populace of many countries from all regions so as to demonstrate and discover the familiarity with the travel industry attractions and it's usage. The example utilized was haphazardly chosen and the organized study was disseminated online. The subjective technique is directed by top to bottom meetings with delegate specialists and administrators in charge of promoting the travel industry.

In the new century, destination advancement rises as a key issue in nearby and local improvement. Marketing of destination ought to likewise manage the travel industry impacts, advancement and the amplification of benefits for the region [29]. Holidays would in general draw in enormous marketing charges, rule media time, and are evident as the travel industry explore openings. The network held occasion contributes altogether to visitors' consumptions and the fundamental reason for facilitating holidays is financial in nature. Network holidays are relied upon to expand the number of tourists and the measure of usage of the facilities.

Destination board and advancement activities of the travel industry destination are imperative elements for a destination to pull in more tourists and increment their ubiquity among populace and visitors. The fitting arrangement of limited time-blend components and propriety of the travel industry activities in the destination can improve the picture of the destination and make a mark. The brand is especially effective in empowering points of interest for tourists and fulfilling their requirements. The brand is likewise situated in the market and offers a special arrangement of estimations of tourists with regards to the travel industry destination and it moves toward becoming a test for a destination with comparative travel industry contributions and attractions.

As indicated by Kotler & Gertner [40], marketing of the travel industry destination is an essential device for drawing in tourists and visitors and for brand creation and increment of brand familiarity with any travel industry destination. Advancement of destination comprises of target showcase definition, defining special objectives and characterizing proper advancement blend components (promoting, individual moving, marketing, and advertising). It additionally includes mark improvement and situating which is of huge significance for the travel industry destination advancement.

Social travel industry can be characterized in a more extensive and smaller view. It very well may be clarified as a thought process in motion so as to visit social estimation of destination, take an interest in indications of social qualities and other social holidays. Enthusiasm for social travel industry and its ubiquity is taken from various interest factors, for example, increment of enthusiasm for culture, development of social capital, development of populace from created nations, expanded portability and in addition by supply factors comprising of development in supply of social attractions, expanding estimation of non-material culture, picture, and climate. A destination that needs to improve the notoriety of social destination ought to grow new items and advance them as social travel industry attractions so as to end up world known and draw in visitors.

With the time and cost constraints, it creates intervening destinations thus providing an avenue where tourists are able to make comparisons and select suitable destinations. Moreover, the flows and movement of tourists in the Asia-pacific countries depends on other factors such as marketing effectiveness, promotion or offers, destination attributes and demographic characteristics [39]. In establishing tourist activity patterns and flows, different countries have adopted various tools and methods to complete the process and comprehend their tourists better. Among the commonly used tools include the Country Potential Generation Index (CPFI) and the Gross Travel Propensity (GTP). The latter is used to evaluate the capability of a region or country to generate trips taking into account a particular population. In essence, it provides an estimate of the travel trips in a given region of the country. The Asia-pacific region has gained attention from the world in terms of tourism owing to the economic and demographic development including less inbound travel restrictions [40] [41].

Owing to the great potential of tourism growth and development, there is a need for Asia-pacific countries to better comprehend tourist activity patterns and flows to enhance their management. The statistical report of the Pacific Asia Travel Association (PATA), exhibited different travel flows among the member countries meaning the tourist activity patterns are unique to a region. In 1995, the United States was the leading tourist destination followed by Canada, Hong Kong, Singapore, China, and Australia. However, the trends of the travel flow between 1995 and 2004 changed with China and Hong Kong topping the list of the leading destinations [39] [40].

In the past, there have been several design methods that were used to capture and analyze data on travel patterns and behaviors of tourists. However, there has not yet been a design method that has enabled the analysis of the sequential activity patterns of tourists to deeply understand their behaviors in the sector [42]. Social media platforms such as Flickr have helped provide rich data sources in terms of the historical data of the tourists and their individual preferences. With millions and thousands of tourists visiting different places across the globe, the information from that platform assists in planning trips appropriately, especially to those unfamiliar cities. Ideally, the hosting service uses geo-tagged photos to identify tourist trajectories in a bid to explore topological spaces with the adoption of the motifs concept to unearth the tourist mobility patterns.

Modern tourists prefer to travel to different cities or places to spend their holiday hence they require adequate information in terms of tourist trajectories and past trends in order to make substantial decisions. Flickr and Twitter are known to be the contemporary social media platforms that help to provide the most convenient tourist and travel recommendations to the users [43]. However, the use of this method has been undermined by the restrictions imposed on the access of private information. The privacy and the scalability issues have made the use of Flickr to obtain the tourists' activity patterns to be ineffective. Nonetheless, the travel recommendations can either be generic or personalized [42]. The latter highlights individual preferences with regard to the matching of the locations during visits. On the other hand, the generic recommendations follow a specific order which includes; trajectory identification > interesting locations > travel sequences > planning > activity recommendation.

Tourist trajectories comprise sequences of landmarks with semantic, temporal, and spatial information [45]. The use of the trajectory methods in understanding travel activity patterns requires the separation of the native tourist and international tourists to better understand the flows. Flicker works by accumulating a large collection of photos and storing the metadata namely: size, time, and location. The latter is crucial in the analysis of the tourists' activity patterns in a given region of place [44]. These methods will obtain the metadata, analyze and provide recommendations accordingly to the users. The geo-tagged photos help in partially capturing the travel information which can be used to construct travel trajectories and eventually unearth the tourists' activity patterns for a given location. Consequently, the similarity matrix framework obtained after the construction of the respective trajectories help in grouping tourists accordingly. From the analysis of the travel semantic motif of the tourists, it was established that more tourists preferred the natural parks in the sequence as opposed to state buildings i.e. Central Park > Brooklyn Bridge > Rockefeller Center. Different analysis using the trajectory framework from the metadata obtained from Flickr gets unique sequences depending on the tourist activity of that particular location [42] [43]. Notably, tourists with similar interests in terms of travel preferences are normally grouped or clustered together hence the possibility of developing behavior patterns and recommendations.

2.2 SOCIAL MEDIA ROLE IN TOURISM

Web-based life information is progressively utilized as the wellspring of research in an assortment of areas. A run of the mill model is an urban examination, which goes for taking care of urban issues by examining information from various sources including online life. The potential estimation of web-based life information in the travel industry studies, which is one of the key points in urban research, in any case, has been significantly less examined. This research tries to comprehend the connection between web-based life elements and the meeting examples of visitors to touristic areas in genuine cases. By directing a similar examine, we show how online life portrays touristic areas uniquely in contrast to other information sources. Our investigation further shows that web-based life information can give continuous bits of knowledge of tourists' meeting designs in enormous holidays, along these lines adding to the comprehension of social media information utility in the travel industry.

Today, our general public is progressively hyper-associated with the uncommon ascent of web-based life; presently, online life has 2.206 Billion dynamic users with 30% worldwide penetration. Web-based life activities are in this way a vital class of everyday activities performed by individuals worldwide to satisfy their social needs. These internet-based life activities have created an abundance of social information, which can give important and even conceivably, continuous bits of knowledge to an assortment of studies. Web-based life has subsequently been utilized as a major aspect of promoting a system for enterprises, through "latent advertising" (as wellsprings of market knowledge to pick up bits of knowledge of the users). The internet-based life ought to be viewed as an indispensable piece of an association's advertising procedure and ought not be trifled with.

Destination over the world intensely contends with one another, so as to keep up their engaging quality and aggressiveness in the worldwide visitor industry. Hence, it is essential for destination experts to have the capacity to address the distinctive needs of various market groups, and additionally advance their picture and oversee destination in a way that draws in tourists. As such, they have to successfully execute destination marketing, the term referring to advancing tourists' destination as a method for enhancing their symbolism and ubiquity [26]. According to Koutoulas and Zoyganeli [27], destination marketing happens in two dimensions. At the smaller scale level, autonomous visitor administrators, for example, inns and transportation organizations, which advance the items and administrations they offer in the business. At the full-scale level, governments and other authority specialists advance their nations and states as a tourist destination.

In the first place, the travel industry item is detailed by the fulfillment of the tourists from the ocean, the sun, the shorelines, the mountains, the inns, the commercial centers and the eateries of a destination. The travel industry item executives are controlled by the attractions, the pleasantries, the gets to and the picture of a destination. Finally, the fulfillment of the movement cost of tourists is dictated by the carrier cost, the inn and the visitor house cost, and the aggregate expense of the local outing.

2.3 DESTINATION MARKETING IN TOURISM

Destination over the world vigorously rival one another, so as to keep up their engaging quality

aggressiveness in the worldwide tourist industry. As such, it is essential for destination specialists to have the capacity to address the distinctive needs of various market fragments, and

in addition, advance their picture and oversee destination in a way that pulls in tourists. Therefore, they have to successfully actualize 'Destination Marketing', the term referring to advancing visitor destination as a method for enhancing their symbolism and ubiquity for tourists [27] 27] [28]. As indicated by Koutoulas and Zoyganeli [27], destination marketing for tourist activities happens in two dimensions. At the smaller scale level, autonomous tourists' administrators, for example, lodgings and transportation offices, which advance the items and administrations they offer in the business. At the large-scale level, governments and other authority experts advance their nations and states as a tourist destination.

A significant number of analysts and industry experts have concentrated on understanding visitor conduct to enhance their general promoting system, item improvement and the nature of advertised administrations. It is fundamental that travel organizations comprehend the visitors' goal inclinations since it enables them to grow new items and administrations by using the strategies of appropriate marketing. It is additionally fundamental to comprehend the movement patterns of tourists to recognize their conduct and buy aim. It recognizes their rivals and enables them to create items and administrations based on various customer portions, as indicated by their necessities and prerequisites.

The key determinant of destination marketing for tourist activities is an issue which has been extensively talked about in the scholarly writing, specifically the factors which decide the accomplishment in a visitor destination [28] [29]. For that reason, the creators disseminated polls comprised of 100 tourists who had picked their destination country as a visitor destination. Also, the scientists utilized factor investigation and basic condition displaying procedures for breaking down their essential information. The aftereffects of this paper showed that destination promotion is affected by four elements, in particular: a) fulfillment of the movement cost, b) the coordinated travel industry item, c) the travel industry item qualities and d) the travel industry item executives. These factors are controlled by a few traits. In the first place, the travel industry is figured by the fulfillment of the tourists from the ocean, the sun, the shorelines, the mountains, the lodgings, the commercial centers and the restaurants for their destination. The travel industry item board is dictated by the attractions, the conveniences, the gets to and the picture of a destination. Ultimately, the fulfillment of the movement cost of tourists is dictated by the aircraft cost, the inn and the visitor house cost, and the aggregate expense of the residential excursion in their destination country.

Buhalis [29] recognizes three vital headings that can improve destination marketing proficiency: 1) upgrade the fulfillment of tourists or visitors, 2) reinforce the long-term intensity and productivity of the nearby travel industry and of the neighborhood and 3) build up the manageability of the destination and ensure the host populace. Every one of these three bearings joins a few vital destinations. To upgrade the fulfillment of the visitors, destination and the travel industry ventures ought to enhance their administrations, practice their travel industry item and offer esteem - for - cash travel industry benefits by concentrating on quality. The enhancement of the administrations can be accomplished with the accompanying errands: institutionalize the conveyance of the services, reception and advancement of value control frameworks, improve operational administration strategies, give adaptable travel industry administrations, center in the advancement of individual associations with the users, put resources into the preparation of the workforce and enlarge the travel industry product.

Echtner and Ritchie [46] compose that from the travel industry point of view imperative elements which decide the picture of a destination are: the landscape and the characteristic attractions, valuing systems, neighborliness and cordiality, atmosphere, tourists' exercises, nightlife and excitement, sport offices, national parks and exhibition halls, nearby framework and transportation, and settlement offices. The management of destination is to oversee and bolster the integration of various assets, exercises and partners through appropriate arrangements and activities. In this viewpoint destination board has six noteworthy undertakings: 1) to enhance the personal satisfaction of the occupants of the destination, 2) to guarantee the nature of visits of the tourists, 3) to pick up and maintain the aggressiveness on the business sectors, 4) to advance manageable improvement, 5) to save neighborhood assets and 6) to make items for explicit market portions. The advancement of local travel industry, arrangement of top-notch visitor administrations, effective utilization of the accessible limits of a destination, resources expansion in the tourist season, increment of remote tourists' movement and quicken the improvement in the immature regions in the destination, all help in the proficient management of a destination. To enhance the symbolism and fame of a tourist destination; this is executed utilizing numerous procedures, for example, publicizing through media, disseminating limited time materials, and offering special occasion bundles [46] [47]. The real test related with destination marketing for tourist activities is that visitor destinations are multi-dimensional, so it is troublesome for advertisers and the destination marketing people for tourist activities to consolidate the distinctive measurements to target their customers.

The market research is additionally required all the time so that destination marketing keeps on top of changing tourist's pattern and advancements in the future destinations. Staying aware of headways in innovation is additionally a basic achievement factor for destination marketing, which can prefer standpoint of new media and data advances to connect with their intended interest groups. The travel industry movement depends on a considerable measure on transportation and correspondence due to the separation and time that incredibly influences a man's craving to travel. The most vital component is the availability of transport for tourist activities. Other than transportation, foundation identified with openness is covering streets, spans, terminals, stations, and airplane terminals. This framework serves to interface a privilege to other sites and destinations. The presence of transport foundation will influence the rate of transport level itself for tourist activities. Great framework conditions will make an ideal transport rate which every tourist will like. The travel industry offices won't separate the accommodation of hospitality, since the travel industry works parallel to it. The facilities of tourists support the comfort zones of all visitors so that they can visit a tourist destination.

2.4 DESTINATION MANAGEMENT - A KEY PLAYER IN TOURISM

Nezirovic [48] composes that imperative perspectives for an effective destination board incorporate activities, such as, advancement of local travel industry, arrangement of brilliant tourist administrations, proficient utilization of the accessible limits of a destination which put resources into the expansion of the tourist season, sort out business ventures productively, center in the growth of the profitability, increment remote visitor movement and quicken the improvement in the underdeveloped areas of the destination. To the extent the destination board is concerned, one of the principal jobs of 'Destination Management' is to secure and enhance the picture of the destination, and additionally offer occasional bundles to tourists [48] [49]. Aside from that, destination management basically includes the advancement of practical travel industry strategies, which likewise defend the requirements and interests of different partners, for example, those possessed with the travel industry all in all, the occupants of destination, the common assets and the earth of destination, and additionally the general public all in all [50]. Chaitip et al. [28] point out that destination management includes three components, in particular, the intensity of the destination, its execution regarding manageability and tourists' fulfillment, and the accomplishment of maintainable objectives. As it occurs with destination marketing, destination management is likewise perplexing, because of the mind-boggling nature of undertakings to be practiced and the number of partners engaged with the tourist item.

In a similar scenario, Mazilu [51] takes note that the role of destination management is to oversee and facilitate all the tourist groups which are engaged with the nearby travel industry. Destination management cooperates with the destination components which are: the fascination of the destination, the amusement offices, openness, HR techniques, the picture of the destination, and the cost of the touristic administrations. The job of destination management impacts and is affected by the destination components. At that point, the job of destination management is to make the fitting condition for the improvement of the travel industry in the destination by building up strategies, enactments, directions, and charges. The destination management assumes a basic job in creating destination marketing efforts so as to pull in more visitors to the region.

Role of Destination Management

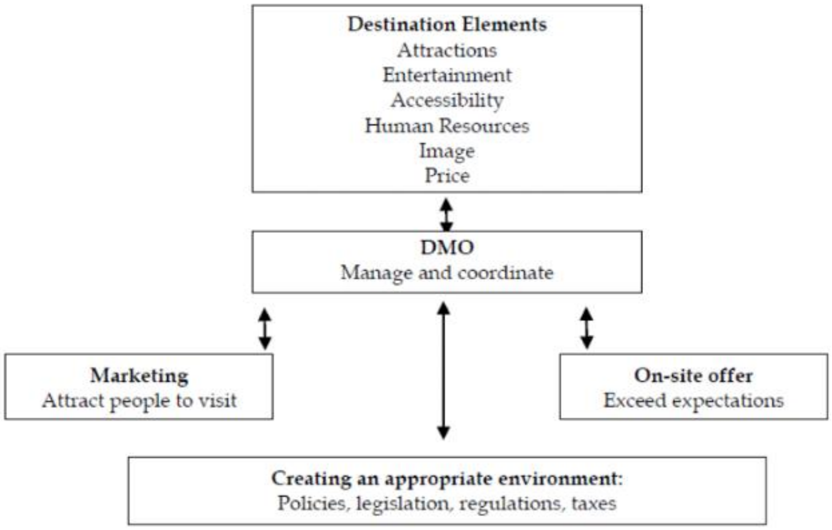


Figure 2A: Role of Destination Management

In this figure, the role of destination management for tourists is defined [51]. Angella & Go [52] claim that destination management and destination firms ought to have cozy associations with explicit commitments and rewards between the two sections. In more subtleties, destination management gets the accompanying commitments from the destination firms: a) reserves, b) accord and authenticity, c) basic leadership capacity with respect to limited time activities to the tourist activities and advancement of the travel industry administrations, and d) capacity to arrange activities, for example, appointments, advancements and holidays, and so on. Then again, destination management provides facilities to the tourism department by the accompanying commitments as:

1) arranging activities, 2) raising the fund, 3) advancement of the travel industry items/administrations, 4) marketing activities, 5) offering activities to have universal congresses, 6) research on the project, 7) preparing to the general population who are working in the nearby travel industry, 8) the executives of the travel industry information and answering to worldwide establishments to position the destination in the worldwide market, 9) know how and involvement in the travel industry executives, 10) principles, data and announcement of the

destination's execution and 11) consistent quality control and assurance of value standards for the tourist activities.

For accomplishing an organized offer, destination management for tourist activities ought to build up an effective advertising framework to offer broadened administrations for achieving more market portions. The decrease of regularity can be accomplished through the utilization of destination advertising activities, the offering of differentiated visitor administrations and the enhanced convenience limits. For advancing the institutional help in the nearby tourist industry, destination management for tourist activities ought to heighten their examination activities, allowing self-rule to visitor associations, set up an unmistakable lawful premise of tourist improvement and create and deal with an incorporated data framework. At last, destination management for tourist activities ought to advance enterprise improvement and in addition convey the significance of the travel industry so as to motivate the nearby community.

Discoveries additionally demonstrate that the area which is advanced as a brand is more fruitful in making mindful travel industry visits. It incorporates proposals for enhancements in traveling potential and recommends explicit activities and utilization of open and web-based life to advance the destinations all the more effectively among residential and remote visitors. Commitment of the exploration is found in the methodology of investigating interests in traveling and giving outcomes so as to underline how less money-related speculation can make better progress and how budgetary interest in advertising isn't adequate, for example, a mix of monetary and other promoting activities and limited time devices so as to pull in more visitors and make attention to the travel industry destination. A destination which chooses to make brands is more effective than those putting a lot of cash in advertising yet not making a solid brand. The discoveries and point of this exploration can add to assist fruitful advancement of the regions and to make consciousness of the present circumstance among open, local and outside tourists and additionally destination supervisors and other intrigued social gatherings [51] [52].

The enhancement of the administrations can be accomplished with the accompanying errands: institutionalize the conveyance of the administrations, appropriation and improvement of value control frameworks, upgrade operational administration techniques, give adaptable travel industry administrations, center in the advancement of individual associations with the users, put resources into the preparation of the workforce and enlarge the travel industry item.

On the other hand, fortifying the long-term completeness and productivity of the nearby travel industry, the official groups of the travel industry should set the accompanying targets: increment incomes, oversee costs effectively, put resources into HR board and advance cooperation among state and open the travel industry associations. In any case, destination and travel industry undertakings can expand their incomes by: expanding tourist volume, focusing on new markets and boosting entrance in existing markets, embracing forceful promoting systems, expanding normal spending per user, choosing high caliber particular visit administrators, putting resources into elective travel industry, propelling the representatives who are working in the travel industry with rewards and finally utilizing elective appropriation administrations for conveying their administrations.

Lopes [53] supports that when tourists pick a visitor destination, it is impacted essentially by the picture of the destination. In this setting the scientist makes reference to the factors which decide the picture in the travel industry destination, to be specific: the view of the visitors, the adequacy of the travel industry advertising activities, the instructive foundation of the visitors, the social and financial attributes of the tourists, the intentions of the visitors, the media (TV, magazines, paper, books, and so forth.), the encounters of the tourists and the mental qualities of the visitors [53]. It ought to be noticed that Lopes [53] recognizes two kinds of destination picture: essential picture and auxiliary picture.

2.5 AN OVERVIEW OF TOURISTS

This section covers the role of tourists in tourism particularly tourists' mobility patterns, choice of destinations, travel behaviour and activity patterns:

2.5.1 *Tourist Mobility Patterns*

The approach of another period of globalization, another period of private enterprise, another time of legacy and another time of the travel industry. While these advancements have been unmistakably distinguished, they are regularly drawn nearer independently. In the event that the progressions identifying with legacy can be clarified by an emergency in our associations with time, or even by another "presents" perspective of history, the new portability examples of the travel industry relate another association with space that is run of the mill of contemporary globalization. The examination venture in which the post-doc hopeful will be included embarks to think about these elements of globalization, legacy and the travel industry conjointly. The point is to investigate the place of tourist mobilities and the travel industry all the more by and large as an aggregate, globalized inside the "creation" of legacy, in a methodology that all the while imagines globalization of our legacy, and globalization by method for our legacy a "legacy connected globalization". This view continues as far as procedures and scaled reconfigurations. It subsequently goes past the thought of various augmentations in the idea of legacy and favors the thought of a subjective jump and an adjustment in rationale - the move from a specific way to deal with chronicled items to a widely inclusive perspective of legacy, and all the more on a very basic level an adjustment in theirs situating in the social field: the "produce" of legacy is meddling to an ever increasing extent, in various structures, in numerous types of globalization. Rescaling can be watched, as a matter of first importance in the rising significance of universal, supranational and transnational players in the generation of legacy – UNESCO, the EU, the transnational tip top, and additionally the tourist business and tourist themselves.

This new stage in globalization is gone to by another association with the domain, which reconfigures tourist mobilities, legacy, and the connections between them. The entry in a period of hypermobility and the presence of a "roaming" society, encouraged by the new transport insurgency and new associations with outskirts, have prompted an expanding multifaceted nature in touristic portability designs on various spatial and transient scales.

Considerable efforts have been spent on finding out tourists' mobility patterns, aiming to identify the tourists' mode and path of travel. Klahn et al. [16] researched the mode and extent of travel of visitors to Munich by carrying out a survey at the Domestic and International Airport Terminal, Central Train Station and Central Bus Stop in Munich. Another study by Yun and Park [54] on the festival visitors in rural areas showed that most tourists visit only the entrance and central spaces, walking along on the same paths, while only a small number of visitors visit the regional commercial area, including the traditional market, main streets in the downtown area, although their tickets allow them to visit these spaces. Some other researchers have also observed the effects of travel preparation on tourists mobility patterns in and around the destination and found that the tourists who prepared their visit before travelling to a destination were the ones covering most of the attractions and activities at a particular destination [12] [17] [21].

Another popular area of research with regards to tourists, is exploring their spatial and temporal patterns when visiting places of interest. Some researchers have confined their research on finding the tourist spatio-temporal patterns while in transit or at a destination [9] [12] [17] [55] [56]. Others have gone deeper and extracted useful information about the tourist trajectories, activity location with respect to time spent at each activity and difference with regards to first time visit and repeat visitation [5] [6] [7] [11] [57]. Birenboim et al. [9] used GPS technology to track and record the time-space trajectories of tourists at a theme park in Spain. The research studied the mass activity patterns of tourists during the 24 hours period while visiting the Theme park location. The findings resulted in the time spent at each activity in the park and the activity pattern of the tourists. 5 main categories of activity were Rides, Shows, Restaurants, Shops, Games. The findings also showed the diurnal and intra-diurnal activity patterns. Another research by Cantis et al. [11] used a segmentation approach based on movement patterns of cruise passengers in the city of Palermo. The results of the study can be very useful to the cruise line companies as well as destination managers and tour operators in helping them devise routes and plan activities to gain maximum satisfactory feedback from their tourists and at the same time maximize profits for themselves. Similarly Orellana et al. [12] used the GPS tracking data in a National Park in Netherlands to study the collective movement patterns of tourists. The focus was on two kinds of movement patterns – MSPs (Movement Suspension Patterns) and GSPs (Generalized Sequential Patterns). The MSPs was used to detect the main places visited by people in a recreational park while the GSPs was used to establish the sequence in which these places were visited. Grinberger, Shoval and McKercher [58] made another interesting addition to the literature by presenting a conceptual framework to describe and understand tourists' spatio-temporal behavior, according to which the effective reach of an individual is defined by time–space constraints and the path taken by the individual. It is evident that the research under these topics has been conducted to draw out useful and real-time information from tourists by collecting data through their hand-held GPS devices. The GPS/GIS device coverage and signals can pose a limitation to the collected data. The GPS device may not be accessible in many remote areas and there is a high chance of disruption in the signals' quality and strength [5] [12]. Furthermore, the past work using this GPS technology was unable to reveal the actual activities undertaken by tourists at each visited location.

2.5.2 ***Tourist Travel Behavior***

Understanding tourist travel behaviour is pivotal for directors, occupied with key planning and basic leadership to make a manageable travel industry. In any case, they keep on confronting huge difficulties in completely capturing and understanding the conduct of global tourists. The difficulties are principal because of the wasteful information gathering approaches. Past research has presented another way to deal with this undertaking by using the socially produced and user-contributed geotagged photographs currently made freely accessible on the Internet. The contextual analysis centres around Hong Kong inbound travel industry utilizing 29,443 photographs gathered from 2100 tourists [24]. The research showed how a dataset built from such geotagged photographs which can help address such difficulties and in addition give valuable pragmatic implications to the development of tourist destination, the arrangement of transportation, and management of impact. This investigation can possibly profit the travel industry specialists worldwide through better understanding of travel conduct and creating the travel industry ventures more economical.

Travel conduct referred to the manner by which tourist act as per their frames of mind previously, furthermore, in the wake of traveling. Knowledge with respect to travel conduct can help with advertising and item planning and improvement which can expand the number of visitors to the tourism items, for example, resorts. Recently, a study was conducted to study and understand tourist behaviour visiting South African hotels, since a gap was discovered in this region of the world. The motivation behind this investigation was to decide the travel conduct and all the more explicitly the movement inspiration of tourist visiting resorts. The outcomes showed that the principle travel inspirations are resting and unwinding, enhancing and knowledge encounters which help in recreating the individual qualities and social encounters for tourism [42]. These outcomes showed similitudes with the discoveries of past research studies as well.

Many experiments have been conducted in the past to establish a link between tourist travel behavior and their socio-demographic as well as psycho-social features. Some of the studies have been based on specific geographical areas and concentrate on finding the type and nature of popular activities chosen by the visitors while visiting those locations while others have contributed by merely exploring the tourist behavior in general. Chen, Wang and Prebensen [10] have explored the activity patterns of tourists based on their demographics and travel companions i.e. travelling alone, with partner or with children. The study conducted in Norway focused around 25 different activities for tourists ranging from short trips, tasting foods, skiing, fishing, theme parks, festivals etc. Yeung et al. [14] in their study, aimed to find out whether the Japanese Tourists differ in their choice preferences, travel behavior & image perception according to their socio-demographic and travel-related factors. The data was collected through self-administered questionnaires which asked tourists input about their preferences of tourist activities, shopping & travel behavior while visiting Hong Kong. The findings of the study recorded differing travel and behavior patterns in Japanese Tourists according to their age, gender, level of education, type and frequency of visit. Kitajima et al. [59] conducted a study at a hot spring resort in Western Japan to understand the tourist behavioral selections based on their chronological development. This study deployed a new methodology for analyzing tourists' in situ behavior, cognitive chrono-ethnography, which incorporates the qualitative understanding of the decision-making process. The findings proved that tourists' activity patterns are dependent on their intrinsic nature and state of mind rather than extrinsic factors. The activities chosen by tourists were dependent on their motivation levels, past experiences and individual preferences. Another research by Dejbakhsh, Arrowsmith and Jackson [15] investigated the spatial behavior pattern of International Tourists travelling to Melbourne City in Australia. It demonstrated that these behavioral patterns are the cause of cultural differences among these tourists. The results showed marked differences between tourists from varied cultural backgrounds with respect to their spatial behavioral patterns. These patterns included choice of accommodation location, mode of transport to travel, length and direction of movement and choice of activity at a particular location.

Several studies have been conducted to investigate the decision process of tourists, their selection of intra-destination activities and factors affecting the choice of a touristic destination. Li, Deng and Moutinho [19] showed how experience activities influenced tourists' impulse buying. Zoltan and McKercher [60] and Bujosa et al. [6] analyzed the tourist intra-destination movements and activity participation by clustering tourist groups based on their spatial and/or activity patterns. This ability to segment tourists based on their dominant movement patterns can help forecast likely future movements and can help in better planning and management of tourist flows. The research also identified a number of factors that influence both the intensity and spatial dispersion of movements. Wang et al. [61] studied the influence of the High-Speed Rail on the spatial distribution of regional tourism in China which is an extrinsic factor forging an increase in tourism within the region. Another study by Ting and Hu [13] was conducted at a Summer Palace in Beijing China, which attempted to cluster tourists based on their spatio-temporal behavior by using cluster analysis. This approach combined time-space data and tourist activity information to classify tourists into the distant groups.

These studies have assisted in the understanding of tourists' behavior in general and further classifying them according to their needs and preferences. The knowledge gained is partial and limited to the locations tourist prefer and the factors responsible for the tourist choice of destination; however, it fails to provide complete information about tourists visit. Another limitation in these above mentioned studies is that the data was sourced mainly from surveys, questionnaires, individual and group interviews, which were conducted with the volunteered visitors. This data could not give very accurate results due to the human error factor – people cannot recall sometimes about the times / locations and most importantly activities carried out – even if they do, there is still a chance of confusion and misinterpretation [6] [60]. Also the data source was limited to only a preset number of volunteered visitors, thus the findings could not be generalized to the entire tourist population at the destination.

Convenience and quality food and a popular destination for tourist with the end destination for significant factors that impact the positioning choices of many tourists and keep motivated to travel more and to take the best destination decision for their tourism. This implies, a fascination with the great and institutionalized settlement and great food for tourist which impact their positive connection to the site/fascination, others indicated want for tranquility and availability reasons, and in addition the remarkable the travel industry items offered at the tourism countries.

2.5.3 ***<u>Tourist Activity Patterns</u>***

There have been some attempts made in the past in studying the touristic flows and predicting tourist future destinations. One of the early contributions done in this respect was from Yang, Fik and Zhang [62] who focused on the decision making process of tourists and tried to model their next destination by using the nested logit model. This model assumed the 'utility maximization' as the tourists demand and targeted on only one subsequent destination, unlike tourists, who can visit multiple subsequent destinations. Another research was done by Zheng, Huang and Li [63] who tried to model the tourist next destination through a survey group who collected data on tourists' intra-attraction spatial-temporal behavior and demographic characteristics using hand held GPS tracking devices and activity diary questionnaires. One of the latest research done in this category is by Vu et al. [24] who explored the travel behaviors of tourists in Hong Kong by using the data from geo-tagged photos uploaded on Flickr. The tourists' movement trajectories were highlighted and patterns were drawn to indicate the most popular tourists' destinations. The location preferences was also categorized with respect to 2 main groups Asian n Westerns. A similar study was carried out while exploring visitors' activities in Hong Kong Parks [64] and Temples [65]. The Twitter 's API has been used by Chua et al. [66] in which the geo-tagged social media data is used to categorize tourists flow in Italy. This research used the geo-tagged social media data from Twitter to characterize spatial, temporal and demographic features of tourists' flow in Cilento, Southern Italy. The study focused on 3 main areas which are: the tourists' profiles, tourists' travel patterns in the region, tourists' attraction in the region and their popularity. The geo-tagged photos on Flickr have been used in many other studies [7] [24] [64]. However, Flickr only gives the geo-coordinates of the photos without giving any information about the actual place, its type, category and user comments associated with it. What the tourist actually did at that particular location, which activities they were involved in, and how much time they spent at that particular activity cannot be assessed via the data obtained from Flickr.

2.5.4 ***Tourist Destination Choice***

The current travel industry which has turned out to be one of the quickest developing segments of the world economy, is generally perceived for its commitment to local and national monetary improvement. There are many factors that directly and indirectly affect the tourist choice of destinations, hence these have raised concern and have been an area of interest for the researchers for a long time. Tourist demographics, culture, race, ethnicity and social class all play a crucial role in their choice of destination and trip patterns. The 'visitor's travel preparation' is ranked as one of the high ranging factor to influence tourist mobility and spatial patterns. On the basis of a sample of 330 questionnaires and 162 GPS tracks, a past study explored mobility paths of same-day visitors in Freiburg, Germany, showing that well-prepared and not well-prepared visitors are two distinct types of tourists each with its own specific mobility patterns [21]. Another research, 'GPS Tracking of Travel Routes of Wanderers and Planners', was carried out by Beeco et al. [17] to determine if the spatial and temporal patterns differed between people with varied travel styles – wanderers and planners. GPS tracking was used to track the tourists' movement, utilization of an area and activity styles. Time spent by tourists on primary roads, secondary roads and stopping locations were the three dependent variables to measure the actual travel pattern differences between these two main categories of travellers. A novel approach was taken by Cantis et al. [11] who used a segmentation approach based on movement patterns of cruise passengers in the city of Palermo. The cruise passengers' mobility within the city and its environs was carefully examined through the use of high resolution tracking data, collected through GPS devices. The aim of this research was to expand the knowledge of the cruise passengers' behavior at the destination by observing their spatial and temporal characteristics. It also identified the main factors affecting the cruise passengers' behavior at the destination which were mostly related to socio-demographic and economic features. The results and findings contradict the widely held idea that the cruise passengers remain confined within a small area near the port. Instead, it proves that the cruise passengers engage in different activities and visit varied number of attractions according to their preferences based on several socio-demographic characteristics. These results can be very useful to the cruise line companies as well as destination managers and tour operators in helping them devise routes and plan activities to gain maximum satisfactory feedback from their tourists and at the same time maximize profits for themselves. Another interesting study carried by Dejbakhsh et al. [15] investigating the spatial behavior pattern of International Tourists travelling to Melbourne City in Australia. It further demonstrates that these behavioral patterns are the cause of cultural differences among these tourists. The data is collected from a total of 278 survey questionnaires filled in by International Tourists. The results showed marked differences between tourists from varied cultural backgrounds with respect to their spatial behavioral patterns. These patterns included choice of accommodation location, mode of transport to travel, length and direction of movement and choice of activity at a particular location. The study further focused on the tourists' movement patterns based at the micro as well as the macro-spatial scale where the different locations around the city center formed the macro scale while the major attractions at these locations formed the micro scale.

Understanding Tourists In Situ Behaviour is another area of study that has also captured researchers attention. A recent study in this particular area aims to understand tourists' behavioral selections in terms of their chronological development. The research was conducted at a hot spring resort in Western Japan [59]. The data was collected through written surveys, questionnaires and interviews from 43 individual groups who opted to take part in the study. The study deploys a new methodology for analysing tourists' in situ behaviour, cognitive chrono-ethnography, which incorporates the qualitative understanding of the decision-making process. It shows and proves that tourists' activity patterns are dependent on their intrinsic nature and state of mind rather than extrinsic factors. Which activities they chose also depends on their motivation, past experience and preferences.

Travel companions and tourist demographics also play a vital role in tourist destination choice and activity preferences. A study conducted by Chen, Wang & Prebensen [10] aims to analyse the activity patterns of tourists based on their demographics and travel companions (travelling alone, with partner, with children etc). It concentrates on 25 tourists' activities in Norway. The activities range from short trips, tasting foods, Skiing, Fishing, Theme Parks, Festivals etc. The data is collected through an online questionnaire survey of residents from UK, Sweden and France. Total number of respondents equalling 6,935.

2.6 ECONOMY OF THE TRAVEL INDUSTRY

The travel industry dramatically effects the world's economy and advancement. It is a hard fact that numerous nations and regions' economies depend on the travel industry income which surpassed $700 billion universally in 2006. For instance, 3.9% of the GDP, 6.2% of Switzerland's GDP, and over 11% of the European Union's GDP are created from the travel industry. The significance of recreation and relaxation in the travel industry reaches out past just being a source of income and gives significant boost to the growth of the economy which impacts territorial foundation, supports nearby industry, adds to activity clog, impacts cargo developments, and supports urban advancement. The travel industry in the African landmass has been viewed as a method for improving monetary development and improvement and additionally propelling the picture of the continent to the outside world. As has been accounted for by UNWTO 2009, there has been huge development (about 4%) in the dimension of worldwide tourists' arrivals in many countries in spite of the worldwide monetary crunch that has influenced tourists entries to Europe and Asia adversely.

2.7 THEORIES OF TOURIST BEHAVIORAL MINING

2.7.1 Compendium Theory

Compendium theory is one of the widely recognized theory in tourism research that helps in understanding the motivations and in predicting travel-decision behavior of tourists. The explanation on the consumption behavior in tourism helps the marketing department of management to allocate scarce resources effectively [67]. Essentially, the theory uses a push-pull factor mechanism to establish the motivation behind the decision making in regards to the tourist destination and patterns [68]. The compendium theory examines some of the ‘push' and ‘pull' factors which motivate tourists to travel to certain places or engage in particular activities. Ideally, the theory states that people are motivated by internal and external forces which cause a push and pull effect [69]. Understandably, the pull factors are what makes one travel to a given destination area such as sunshine and breeze. On the other end, push facts are wants and needs of the tourist which motivate them to travel to a destination.

2.7.2 Generation Theory

The generation theory was first introduced in the tourism sector to establish the travel patterns and market segmentation. There was the need to use resources sparingly to target a certain group of tourists rather than as a whole [70]. As a result, there has been a better understanding of tourism management making decisions in terms of generational evolutions. The theory states that an era in which an individual is born affects the way they perceive the world. The developments and activities of a particular generation are influenced by the economic, social and environmental factors of the time. Consequently, different generations are expected to have unique characteristics [18]. For example, the tastes and preferences of the generation of the nineteenth century are not similar to those of the current generation X. Individuals in the past may be interested in the monumental sites while the modern generation may be interested in art and design as part of their respective tourism destinations [71]. While the elderly or senior groups are conservative, the young people are being accustomed to online travel information putting themselves on the verge of discovering new destination areas. In regards, the generation theory has been used in addressing the different taste and preferences including the contrasting behavior of tourists.

A proper understanding of tourist activities is important for tourism management to ensure they provide the best service, meet tourist expectation and also gain repeat visitation. The past studies have been concentrated around tourists' decision-making process, travel trajectories and movement patterns. The knowledge gained from the past research has been beneficial in understanding tourists' behavior, however, little information is gained about tourist activity at a particular destination which is very important for tourism management and can assist in many ways. This research will fill the gap in the tourism literature by studying tourist activities in the sequential order, thereby providing rich information about tourist choices, preferences and decisions while visiting a particular destination.

*2.7.3 **Cohen's Theory***

To build up a hypothesis which is identified with the conduct of tourists, the typology is categorised into four areas:

1. The tourists of Organized mass – travel in gatherings; purchase a bundled visit which is organized ahead of time by movement specialists or visit administrators.
2. Individual mass tourists – every individual from the gathering has a specific level of authority over his time and schedule and isn't bound to a gathering. He settles on his individual choice about his exercises.
3. The Explorers – such tourists orchestrate their own outing. They connect with the neighborhood inhabitants and endeavor to talk the nearby language; however, they don't totally receive the way of life of the host network or nation.
4. The Drifters – they avoid contact with different tourists. They remain with local people and offer their habits, food, and shelter. They are completely inundated in the host culture. They hold just the most essential of their local traditions. They don't view themselves as tourists.

*2.7.4 **Butler's Theory***

This is a hypothesis based on the thoughts of Christaller, Plog, Cohen, and Doxey. It was just recognized that this hypothesis depended on before speculations yet likewise demonstrated that it depended on the business idea of the item life cycle of tourist activities.

It was demonstrated that regardless of a few reactions, after nearly 20 years, there was much help for his unique model. He recommended that the accompanying key focuses affirmed the legitimacy of his unique theory which is:

1. The key idea is dynamism. Destination change additional time.
2. There is a typical procedure of improvement of the destination of tourist.
3. There are breaking points to development. If the interest for visits is more than the limit of the destination, the quantity of tourist will diminish and in this way decay.
4. There are triggers or factors that achieve change in a destination.
5. Management is an imperative factor. Great administration is expected to maintain a strategic distance from the disappointments recommended by the model.
6. The view of long-term. There is a need to search ahead for a long time, not 5 years, to maintain a strategic distance from the failures recommended by the model.

The item life cycle is a hypothesis in which offers of the new item develops gradually and afterward encounters a time of fast development before balancing out and after that declining. At the point when connected to traveler destination, the model proposes that hotels create and change after some time and that there are four stages specifically: investigation, inclusion, advancement, and solidification. In these stages, the destination has an expanding number of visitors. After the combination organizes, the destination could stagnate; it could decrease or it could rejuvenate as well.

2.8 EXPLORING THE GLOBAL TOURIST SEQUENTIAL ACTIVITY PATTERNS

Time is a critical factor in the establishment of activity patterns in any given scenario. The determination of the activity patterns can be done in numerous ways. In the modern world, the classification of human activity patterns is still challenging although there has been an improvement in the technology capable of capturing and processing large databases. As a result of improved economic opportunities, more international connectivity has been enhanced globally hence the need to understand the travel behaviors and activity patterns. In the tourism industry, the understanding of the travel patterns of the tourists is useful to the transportation agencies by acquiring information on how often tourists travel from one place to another and in what times of the year. The introduction of smart cards which store customers' information has made it possible to analyze and evaluate personal individualized travel patterns monitored over a certain period of time. Nonetheless, there is a need to maintain the confidentiality of the customer information whenever the information is captured at a particular transaction for instance at the airports, hotels and attraction sites [72]. Today, most of the travel operators allow the use of cashless systems thus making it easier for tourists to do transactions efficiently. With the smart card data, one is able to identify repeated patterns of individuals moving from one destination to another. In a given population especially in complex economies, people tend to commute from one place to another after every short time interval.

Smart cards hold detailed information about the users which might sometimes be used in inferring the living standards of the holders. In order to conceptualize the travel behaviors of users, the adoption of the activity-based travel theory assists in the description of the relationship between travel and activity patterns including the life choices made by the respective users. Most people tend to prefer public travel operators as compared to private agencies in moving from one destination to another. It is assumed that people travel from one place to another either for business needs or tourism activities. However, there may be other reasons why people travel from one place to another such as seeking medical help, sporting activities which are deemed to be minimal. The activity-based travel theory (ABTT) lays down a framework that can be used in travel forecasting as well as understanding the travel behaviors and activity patterns of tourists.

The assumption of the activity-based travel theory (ABTT) is that people engage in activities located in different places hence they must move to the respective places to perform those activities. The outcomes of the travel patterns may be associated with certain lifestyles. The social class of people can be drawn from the travel activity patterns. For instance, people who frequently use flights are considered to be of higher social class in society. The ABTT theory helps in outlining the importance of breaking down individual journey with respect to time to form sequences. Subsequently, these sequences can be used by the travel agencies in preparation for the peak travel seasons by understanding the similar patterns as well as the less frequented travel patterns. The activity patterns are not only defined by the order in which they occur but also by the attributes of the respective activities in the permutable sequences. Similarly, the repeated patterns are defined by the number of times the events occurred on a weekly basis. In the past, surveys had been used in the understanding of the individual travel behaviors and patterns from one destination to another. However, this method has been complex hence the capturing of the information may be done over a short period of time. Also, the travel diary surveys are considered to be extensive because more time is required to manually record the information of the tourist while sorting them into respective groups in terms of the activities.

Technological developments in recent years have made the gathering of the travel information easier. Most of the devices such as smart phones, tablets, and computers have incorporated the GPS technology which has the capacity to automatically capture, store, and channel data accordingly. The information may be retrieved and organized into large databases in which the sequential pattern mining (SPM) technique is applied to yield sequential activity patterns. Smart cards are also part of the technological advancement precisely in the tourism industry. Most of the travel operators have implemented the smart card technology hence periodically streaming information about the users to large databases. In the smart cards, the transactions received contain useful information including the time of travel hence the patterns can be derived accordingly. Although the smart cards may provide additional information about the users' lifestyles and other attributes, it is important for the travel agencies to only focus on the travel information and to uphold the confidentiality of the user's information. All of the smart card holders may not be comfortable with sharing too much information about their lifestyles and social class hence it is important to stick to just location, activity, and time transaction was done.

Using the smart card data, the travel behavior of tourists is organized into three different layers namely; transit usage, travel patterns, and lifestyles & attitudes. The latter relates to the living standards of the individuals based on the travel information shared via the smart card technology. For the travel and activity patterns, the assumption of the tourists' lifestyles characteristics may be made although it is not important in understanding travel behaviors of the respective individuals. The transit usage is considered to be an observable layer which involves the capturing of the tourists' explicit information about their destination inferences, and route choices, also the time of the travel is important at this stage. The departure time and transit time including origin and destination are captured in the process which is helpful in the formulation of the travel patterns layer. The breakdown of the travel behavior into three relatable layers helps in the construction and analysis of the sequential patterns of the tourist. Generally, the travel operators have the mandate to instruct on the usage of the captured information. Also, the control of the information sharing with the public to uphold the confidentiality of the tourists.

In instances where the captured information from the smart cards is not adequate, travel surveys are often used as a complement to seek for instance the attributes of the tourist with respect to a particular destination. Based on the analysis of the framework, the activity patterns are organized according to socio-demographic constraints, locations, and time constraints. In the long-term interests, the constraints may be extended to include lifestyles attributes such as housing and employment. Ideally, the smart cards only provide the travel trajectories and the collective information of the users. Afterward, different methods including sequence alignment method with algorithms are applied to result in the respective patterns of travel behaviors.

Travel Behavior Framework Analysis

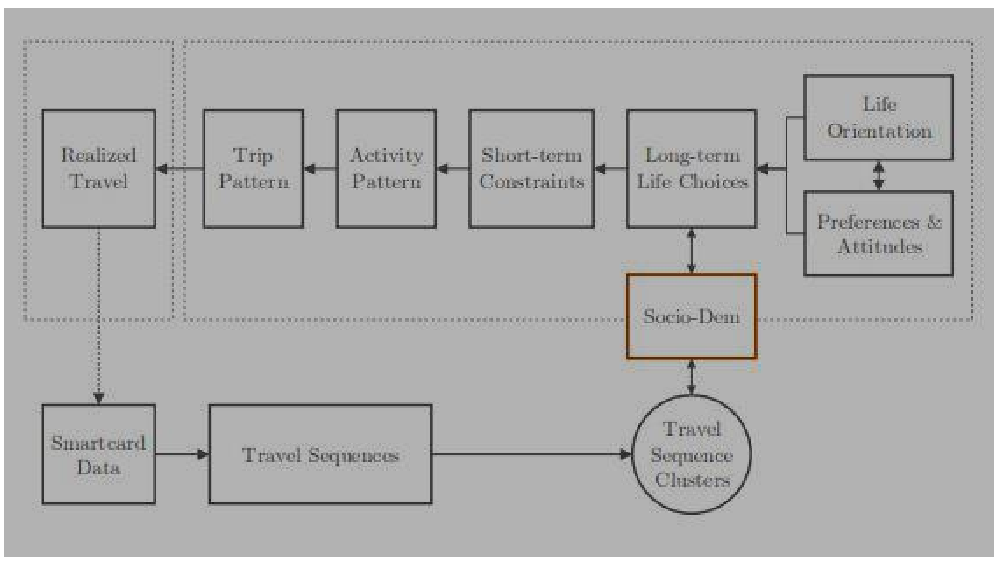


Figure 2B: Travel Behavior Framework Analysis

The figure above shows, the transactions captured via the smart card are organized into travel trajectories which helps in pining together similar activities hence eventually creating the activity patterns. As a result, travel sequences are formulated from each of the travel trajectories. The life choices are dependent on the constraints such as time and social class which define the activity patterns.

There are various methods that have been proposed by researchers for mining sequential activity patterns of events. Nonetheless, some of the methods are restricted and are not applicable in various industries including tourism. Also, some of the proposed sequence modeling strategies in the tourism sector have failed to yield the required activity patterns from the given data or databases [72]. In bio-informatics, dynamic algorithms have been used in identifying both the differences and similarities of the events and travel patterns of individuals with a view of having the sequential activity patterns. The activity time and the transition from one activity to another are some of the important attributes of the sequential activity pattern in any given application. Due to the varied motivations of travel to a certain destination, there is a possibility of having shared patterns among the participants. For example, there is the possibility of tourists engaging in one activity at a particular time. Consequently, it becomes easier to construct the sequential activity patterns using the appropriate framework. Some of the methods and models tend to yield more patterns as compared to others as highlighted in the first chapter.

Today, software's such as the CLUSTAL_TXY algorithm has been implemented in the extraction and classification of the activity patterns from the gathered data. According to Hagerstrand, the incorporation of time to travel-activity trajectory is paramount to ensure a reliable sequence of activity patterns. In the mid-1960s there was the urge of constructing urban models of travel patterns in relation to the daily activities people are engaged in. As a result, it aided in the understanding of the human behavior on a daily basis [73]. Economies, including business investor have utilized these models to make substantial decisions on long-term investments in a particular region of choice. There has been a difficulty in defining household activity patterns as well as measuring the similarity of the respective activity patterns. Notably, most people tend to leave their patterns after a certain period of time to shift into another pattern and try to explore and identify what lies in those other patterns. The human curiosity affects the credibility of the existing patterns, hence the need to conduct a new survey while updating the databases periodically to ensure reliable sequential activity patterns at all times. A travel survey conducted in Upasana in 1971 recorded successive activities, covered by the respondents. Also, the results showed a smooth transition from one activity to the next, especially by the travel locations.

The complexity of the activity patterns came in the attempt to explore the relationship between the extreme travel dimensions and the socio-demographic characteristics. In order to realize it, the principal component analysis was used having over 50 indicators. The indicators played a key role in regressing the components to their adjacent travel characteristics. However, the results of the analysis did not yield the true representation of the daily sequence of activity patterns. Moreover, a similarity matrix was used in unearthing the relationship between travel activities and the sequences at the destination area. The approaches of the similarity matrix and the Hagerstrand trajectory has helped in the understanding of patterns and the relationship between travel and activity. Due to the complexity of the travel-activity relationship analysis, the application of the algorithms and SPM methods has been limited to the development of the sequential activity patterns from the dataset. In the transport industry, a combined analysis has been adopted with the incorporation of spatial dimension for each of the unit analysis.

Man, Wah and Yeung said that the sequence comparison algorithms were first applied in the social science field in the analysis of the career pattern taken by students. Both the comparison and alignment algorithms were developed by computer science and biochemistry professionals for academic purposes. However, these algorithms attracted wider fields such as tourism in an attempt to try and understand the sequential activity patterns of tourists visiting a given region or nation such as Brazil. Subsequently, the mining methods also used in the migration studies, transportation and in the understanding of the tourists' behaviors. In recent years, there have been some emerging algorithms to aid in solving the modern world problems in various sectors of the economy. The alignment algorithms for the evaluation of the sequential activity patterns can either be pairwise or multiple. The latter involves the analysis of more than two sequences of activities. Pairwise alignment involves only two sequences in evaluating both the similarities and differences between them. In essence, pairwise alignment uses appropriate operations in trying to equivocate the source and the target. In the tourism research, the pairwise alignment may be applied in trying to equivocate the origin and the destination of the tourists while trying to establish the possible relations between the two locations. Some of the operations used in the analysis include substitution, deletion, and matching. The latter has been used as a similarity measure to pair together sequences with similar characteristics.

Barenboim said that in order to efficiently establish sequential activity patterns, there is a need to allow an amendment to the algorithm of the adopted methodology to suit the intended dataset. For example, whenever the activities are identical and happen in a similar locality, there is a need to amend the algorithm to accommodate such situations. In that case, when the activities are in the same locality, the travel distance is equal to zero. Typically, a location contains coordinates (x, y) which indicated the presence of a Euclidean space between events. Where the coordinates are required, the CLUSTAL_TXY software package is important in computing the matrix as well as the Euclidean distances between two points in different locations. This complex analysis may not be possible in deciphering the sequential activity patterns with the incorporation of the software packages of a comprehensive algorithm.

The urban survey conducted in 1972 involving Cambridge University students, 450 respondents were engaged which included a recording of their activities over a week-long period for each of the respondents. In order to efficiently implement the algorithm, the size of the sample was reduced to 53 respondents. The time-space trajectory adopted assigned the activity and location a letter since every activity had a location. The week-long dairy was organized such as to be able to complete it, the respondents were required to fill in the two letters representing the activity and location respectively. Subsequently, the letters were fed into the CLUSTAL_TXY which involved a series of computations [38].

Classification and Nodes of Activities

Each of the activity-location letters is recorded in 30 min interval for all the recorded reporting days. In order to improve the accuracy of the representation, shorter episodes were preferred including the contact start time of the activities for all the respondents involved.

Activity	Letter	Shade	Activity	Letter	Shade
Sleep	Z		TV, media use	M	
Personal care	P		Travel, private mode	T	
Work - paid	W		Travel, public	U	
Domestic work	D		Private leisure	R	
Cooking, dishwashing	C		Social activity	V	
Family care	K		Sport, hobbies	A	
Education	I		Organized leisure	L	
Reading, study	J		Religious observance volunteer activity	Y	
Eating	E		Not reported	X	
Shopping	S				

Table 2.1 Classification and Nodes of Activities

The adoption of the CLUSTAL_TXY resulted in the following fragment of the activity-location sequence: Pa-EA-Da-Tb-WBWBWBWB

The sequence is a representation of an individual's typical day which starts with personal grooming (Pa) then Eating (EA). Both activities happen at home. Afterward, the respondent travels to location (Tb) then engages in work (Wb) for an extended time period. The domestic work (Da) was completed first before traveling to location (Tb). The sequence pattern indicates both time-activity and time-space projections. Notably, the activities happening in the locations are aligned together based on the output of the CLUSTAL_TXY. For example, personal grooming and eating happened at home hence they are shaded using the same color. Essentially, the shading for the alignments helps in the readability of the output of the mining process. Each of the records is assigned a letter and a number for profound identification. One of the respondents was a young married woman who was unemployed. Thus, most of the activities were only listed to a home location (a). Only three locations were recorded namely: two stores and home. The inclusion of the locations in the determination of the sequential activity patterns has helped in the deeper understanding of the patterns.

To make it more interesting each of the records and dairies is organized into groups as shown in figure 5 below. The application of multiple alignment algorithms identified the primary activity groups as depicted by the CLUSTAL_TXY software. Activities done in the same locations are shaded together for clarity and understanding. Ideally, the behavioral groups from the results can be identified in different ways as illustrated in the software. One of the important ways is the shading of patterns which tends to assimilate records having some similarity in terms of shading. Each of the activities and the locations is assigned letters which are provided with initials to enable ordinary readers to be able to understand the sequential patterns appropriately. For example, the light grey shading represented events done in the morning such as washing, eating, and personal care. Apart from shading, one may decide to use a clustering software to pair the groups by making the necessary alignments. The purpose of the coloring is to make the sequential patterns more comprehensible and informative to the consumers.

Multiple Behavioral Groups of the Sequential Activity-location Patterns

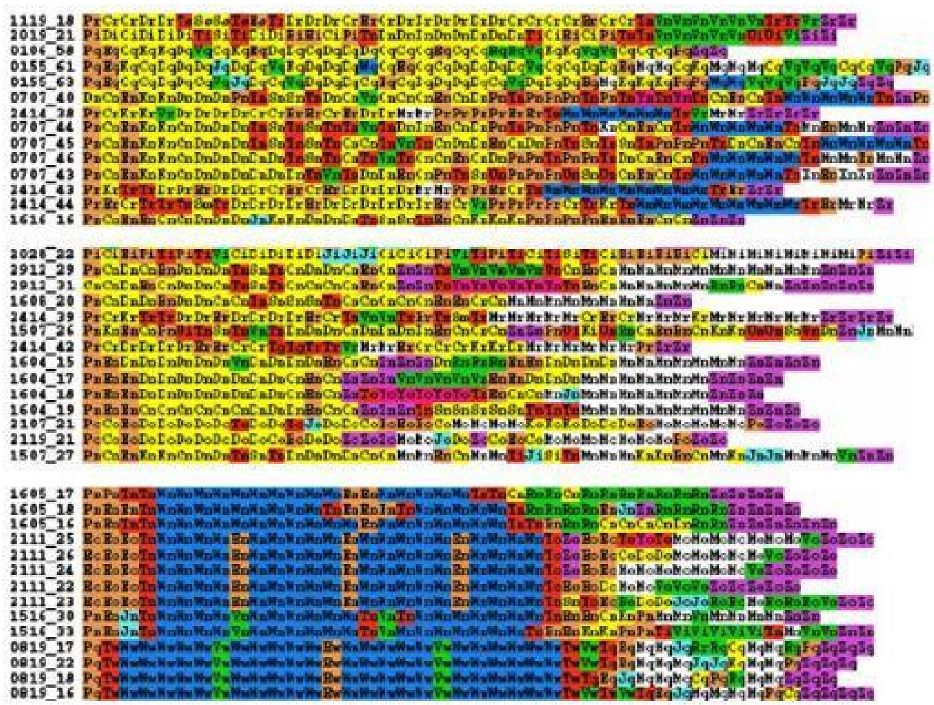


Figure 2C: An Illustration of the Multiple Behavioral Groups of the Sequential Activity-Location Patterns

According to Aggarwal [44], Both the first and the second groups are dominated by the yellow shading indicating more of the domestic work. Also, there are fewer transitions from domestic work to leisure activities. Notably, the white characteristics occur in the afternoon hours indicating resting after completion of the domestic work. The red shading indicates traveling activities. As noted in the first two groups, the bottom group has some white shading at the extreme end also inferring leisure activities such as watching movies after a hard day.

Multivariate methods and descriptive statistics are used to further break down and understand the multiple behavioral groups as provided by the CLUSTAL_TXY. The method helps in the understanding of the associations between economic and demographic factors in the initial dataset. The association helps in linking the activities done by the respondents to the travel demands required to be accomplished in order to reach the respective destinations of the activities. The scientific paradigm helps in the determination of the appropriate method and algorithm used. In this case, the idea of location's coordinates and the CLUSTAL_TXY have enabled the formation of the activity-location patterns which is more informative as compared to having only the sequences of activity patterns. The choice of a particular method is defined by the nature of the dataset and the objective that needs to be accomplished in the research study [73]. In a given sample of a dataset, the use of alignments helps in providing a complete illustration of the activity-location association by showing all the events and the transition involved in one activity to another. Although the activity patterns are meant to be long-term, the trajectory representation has an impact on the social and economic changes resulting in overtime on the existing sequential activity patterns.

Goulet-Langlois [72] said that there has been an increase in the growth of international tourism and travel among the various destinations throughout the world. People of the current generation have developed the urge to explore new areas other than being used to the normal residential places. Japan is one of the potential destinations although it has been known over time as the vibrant industrial country especially in the manufacturing sector. Having reached the historic record of 19 million inbound tourists in 2015 and 2016, the government has endeavored to invest more in the tourism industry to reach over 20 million inbound tourists by the year 2020 and subsequently approximately 60 million tourists by 2030 [74]. Based on the statistics, it means that the large datasets are gathered regarding the account for the inbound tourists visiting the country annually. However, there has been minimal research done on the understanding of the tourists' sequential activity patterns in the country. The understanding of the activity patterns aid in the management of the tourism sector to better manage the institution while improving the delivery of services to the consumers.

Minimal data mining techniques have been used in the analysis of the tourists' behaviors in Japan as one of the destinations for inbound tourists. Despite a large number of tourists visiting the country, the advancement in the technology has made it easier to gather the data and information including storage in databases for future usage and reference. The use of the sequential pattern mining techniques results in a better understanding of the movement patterns of the tourists from one destination to another. As opposed to classical statistics, data mining helps in better understanding of the activity patterns of individuals and is considered to be more advanced. In the application of the sequential pattern mining method and algorithm, data preparation is paramount including data cleaning to ensure better outcomes in terms of the activity patterns. Sequential data mining techniques were applied to the data collected from the Japanese Tourism Board (JTP).

The decision trees, one of the mining techniques, were used to provide an understanding of the satisfaction levels of inbound tourists based on the sample data [74]. Also, the technique was used in the assessment of the intention to return to Japan having accomplished the tourism schedule. Ideally, the analysis was categorized into two groups namely: Asian and Non-Asian tourists. From the result, it was established that nationality played a significant role in the odd ratio of satisfaction and the intention to return back. Primarily, the results revealed that more Asian were satisfied as compare to the non-Asians on the odds ratio of 3:9.

According to Kim et al. [75], The novel approach is one of the data mining methods used in determining the sequence of activity patterns. The methods have been proposed and utilized by some of the researchers in the field of sequential pattern mining. Ideally, the approach is mainly used in analyzing traveling trajectories of both people and objects from one place to another. It defines annotated sequences based on a framework which is applicable on any given source of datasets. Also, the sequences obtained owing the novel approach are of a varied length which makes it appropriate for large databases especially in tourism research [76]. In essence, the approach involved a series of steps the first one being the assigning of the points of interest to every photo location collected. The second step included the application of the density-based algorithm to the organized data. Subsequently, the third step involved the mining of the individual travel sequences. Ultimately, the travel sequential patterns are developed using the characteristics obtained from the first two steps. The novel approach is suitable for geotagged photos kind of data and has been implemented in Portugal and Germany in understanding the tourists' behaviors and movements from one destination to another.

The emerging of the location acquisition technologies has made it possible to analyze human mobility based on the trajectories collected from the respective devices using the GPS technology. Due to the challenge of data acquisition problems, there has been some of the missing information on the data collected while some of the large datasets have become unavailable. In recent years, there has been a shift to the reliance of the photo sharing sites such as Flickr in the analysis of the travel behaviors and human activities. The geo-tagged photos posted to the sites contain information of locations including the time in which the photos were taken. As a result, these are some of the important attributes in the mining of sequential activity and travel patterns. Users also indicate the type of activity they are engaged in before posting the photos. In a situation where the activities have not been indicated, one is able to deduce the activity based on the happenings in the photo. The information affiliated to the photos posted to the sites is helpful in discovering the travel patterns and ultimately in the understanding of the order in which people visit places [70]. The act of sharing photos among the GPS enabled users has not only been fun but has helped in the scientific study of the human activities and travel behaviors.

The concept used by the novel approach has also been used in another field such as artificial intelligence and bio-informatics [76]. The challenge of every sequential pattern mining method is the data preprocessing stage which includes the breaking down of raw data, data cleaning, and data adjustment. The data preprocessing phase is a very important stage in the process of mining sequential patterns. It helps to ensure that reliable patterns are developed eventually which can be used by the management and relevant authorities in the management of the various attraction sites and in the improvement of production and service delivery [69] [70].

The Novel Approach Framework

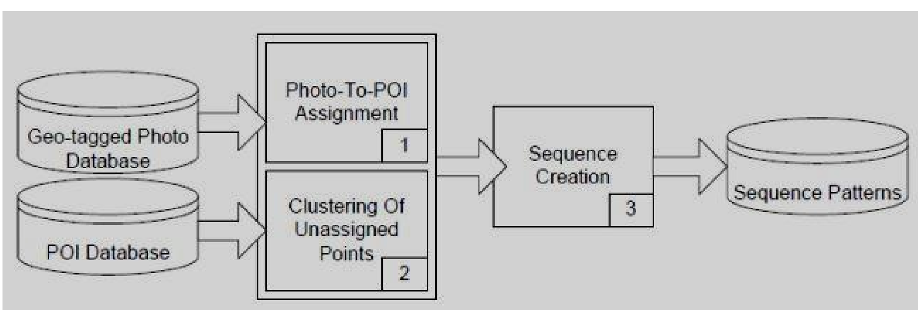


Figure 2D: The Novel Approach Framework

The POI are Point of Interest which is initial steps of the approach in the procedure. Basically, the ultimate mining of the sequential patterns is accomplished in there systematic steps as shown in the above figure. Web crawling technique is required in the extraction of information of photos from the Flickr which is one of the commonly used photo sharing site in the world. The metadata from the photos is downloaded into databases and organized to be utilized in the mining of sequential patterns either for activity or travel.

Although Guimaraes was known to be the first capital city of Portugal, a research study on mining travel sequences revealed that few people shared photos on Flickr as compared to other less prominent cities in the country. Therefore, it was difficult to conduct the data collection procedure from the site. Essentially, large datasets provide a true reflection of the travel behavior of people owing to the sequences obtained from the process. After the area assessment and the definition of the city by use of longitudes and latitudes, approximately 391 photos from 152 users were obtained. The data obtained was appropriate for the mining of the sequential activity patterns. In order to obtain the patterns, a distance of between 200 and 400 meters was taken in the assignment of photos to the nearby point of interest (POI). As a result, some of the defined points of interest included: Guimaraes Castle, Guimaraes Historical Center, São Paulo, Oliveira do Castillo, Azurem University, and the Palace of the Duke of Braganza. Using DBSCAN, the unassigned photos were identified and allocated the remaining unknown POIs having in mind the distance threshold.

The Teiresias algorithm was applied in the second step of the process to yield the sequential travel patterns. The statistics of the patterns and the sequences from the Guimaraes locality has been summarized in the table below. From the 400 meters distance threshold, only 24 were considered to be valid sequences out of the 138 sequences. Similarly, the 200 meters distance threshold provided 18 valid sequences out of the possible 124 sequence patterns. This means that the larger the distance covered between the photos taken and the points of interest, the higher the number of valid sequential patterns obtained from the process.

General Statistics of Guimaraes Sequences

Photo-to-POI threshold	<L,W>	# of people in sequences	# of valid sequences	# of sequence patterns
200	<2,3>	127	18	8
400	<2,3>	138	24	7

Table 2.2: General Statistics of Guimaraes Sequences

A pattern M is referred to as the <L, W> pattern if the subsequent sub-patterns contains at least W and L operations as specified by the Teiresias program or algorithm of the novel approach method. From the table above, there are 5 recurring patterns form each of the Photo-to-POI threshold which is conspired to be sub-patterns.

Guimaraes Sequential Patterns using L=2 and W=3

Photo-to-POI Threshold	# of Input Sequences	Sequence Patterns
200	5	Guimaraes Historical Center → Nossa Senhora da Oliveira
	3	Guimaraes Castle → Church of Sao Miguel de Castelo
	3	**Nossa Senhora da Oliveira → Church of Sao Miguel do Castelo**
	3	Church of Sao Miguel do Castelo → Nossa Senhora da Oliveira
	2	**Church of Sao Miguel do Castelo → Nossa Senhora da Oliveira**
400	4	Guimaraes Historical Center → Nossa Senhora da Oliveira
	4	**Guimaraes Castle → Nossa Senhora da Oliveira**
	3	**Nossa Senhora da Oliveira → * → Nossa Senhora da Oliveira**
	2	Church of Sao Miguel do Castelo → Nossa Senhora da Oliveira
	3	Guimaraes Castle → Church of Sao Miguel do Castelo

Table 2.3: Guimaraes Sequential Patterns

Similarly, the novel approach was also tested and implemented in Berlin, Germany at the tourists' attraction sites. The process was meant to identify the mobility sequential patterns of the tourists visiting Berlin precisely the listed attraction sites in the city. Germany is also regarded as one of the potential destination for most of the international tourists. In recent years, it has experienced growth in the industry in terms of the surging numbers of inbound tourists. Some of the tourist attraction sites in Berlin include: Museum Island, Reichstag Building, Brandenburg Gate, Berlin Wall Graffiti Remains, and the Holocaust Memorial. According to the tourism board, each of the destination records a high number of tourists on an annual basis hence the need to carry out mining of the sequential patterns to understand the travel and activity behaviors of the tourists.

In 2008, approximately 17 million people visited Berlin for tourism activities of which about 7 million were categorized as foreign tourists according to the European cities site [76]. By application of the novel approach, 71,821 photos were retrieved from Flickr between the 2005 and 2009. Using the same distance threshold of 200 and 400 meters, the photos were assigned to the respective points of interest (POI). The Teiresias algorithm also has the capacity of identifying the most frequent patterns. After completion of the approach, the statistics of the sequence patterns were presented in table 5 below. Out of 8954, 2488 were found to be valid sequences for the 200 m distance threshold. On the other hand, 2845 sequences were valid out of the possible 8968 as summarized in table 2.4:

Berlin Travel Sequence Statistics

Photo-to-POI Threshold	<L, W>	# of People in Sequences	# of Valid Sequences	# of Sequence Patterns
200	<2,3>	8952	2844	2047
	<3,4>			186
	<4,5>			9
400	<2,3>	8968	2845	2086
	<3,4>			195
	<4,5>			11

Table 2.4: Berlin Travel Sequence Statistics

Owing to the sub-pattern specifications, the algorithm revealed that there were 2047 sequential patterns of length 2, 186 of length 3, and 9 sequence patterns of length 4 respectively. For the 400 meters distance threshold as shown in table 5, 2086 sequence patterns were of length 2, 195 of length 3 and 11 patterns of length 4 respectively [76]. Further analysis of the sequence patterns, the first three sequence patterns from both distance thresholds showed that the tourists began photographing from Brandenburg Gate before moving to other places. Also, the first sequence pattern showed that people started at Brandenburg Gate while the majority of them ending their day at the Reichstag Building. The application of the density-based clustering technique, more frequent patterns were revealed as shown in table 2.5:

Sequential Patterns, Berlin with <L, W> Specifications

Photo-to-POI Threshold	# of Input Sequences	Sequential Patterns
200	74	Brandenburg Gate → Reichstag
	53	Brandenburg Gate →Memorial to the Murdered Jews of Europe
	46	Brandenburg Gate → * → Reichstag
	41	Reichstag → Brandenburg Gate
	36	Pariser Platz → Brandenburg Gate
400	71	Brandenburg Gate → Reichstag
	51	Brandenburg Gate →Memorial to the Murdered Jews of Europe
	47	Brandenburg Gate → * → Reichstag
	43	Reichstag → Brandenburg Gate

	34	Reichstag → * → Reichstag

Table 2.5: Sequential Patterns, Berlin with <L, W> Specifications

The adoption of the novel approach in the two cases of Guimaraes and Berlin showed the complexity of mining geographical data. Although there are many spatial problems associated with geographic data mining, data quality ought to be maintained to ensure smooth processes and reliable outcomes of the sequence patterns. However, data quality depends on how the data was collected and processed before the application of the algorithms. Poor quality of the data often results in complicated analysis and the eventual unrealistic outcomes in terms of the sequence patterns. The quality of the data obtained from Portugal and Germany case was high hence the easy implementation of the novel approach which uses both Teiresias and Density-based clustering algorithm and technique respectively. Although frequent patterns are always used as the point of reference, it was not the case for the Berlin case study. It, therefore, means that frequent sequential patterns may not always translate into being interesting patterns of the research study. In most cases, the frequent patterns are considered to be obvious hence most of the researchers are going to the less interesting patterns which are explored by few people. In regards, they try to find out the rationale of why people were captured in those patterns.

The result of the Portugal and Germany sequence patterns analysis revealed that altering the threshold distance between the photos and the POI may result in a slight effect on the number of the sequence patterns obtained. From the case studies, human involvement is also paramount in the mining of sequence patterns by selecting the appropriate parameters, controlling the analysis and inspecting and cleaning the data. In as much as the automation process is required for the creation of the sequence patterns, human contribution is also vital to yield credible and reliable outcomes [75] [76] [77]. Also, other data mining techniques such as clustering may also be incorporated in the process to unearth hidden patterns which are often considered to be less interesting. The novel approach for mining traveling sequences can yield patterns of any light hence regarded to be more flexible. The cases of Berlin and Guimaraes revealed that the novel approach can be used in spatial places with both few and large tourists.

In any given dataset, the time factor is an important ingredient in the understanding of the activity patterns of individuals. With time, one is able to create a sequence of activities for instance activities done in the morning hours are given priority and appear at the start of the sequence. While events happening late in the night often appear at the end of the sequence. However, when creating a sequence of travel patterns or locations, time factor may not be necessary since one may decide to travel from one destination to another at their time of convenience. In this case, one is required to record the destination from one place to another which can ultimately be useful in the creation of sequence patterns. The Global Positioning System (GPS) is one of the technologies used to track and record time at which the activities are carried out. In a research study conducted in Catalonia, Spain to understand the activity patterns of people who visit the theme park, the technology was used to track and record time-space trajectories at the Port Aventura theme park. The findings showed that the visitation to the theme park assumed both diurnal and intradiurnal behavioral patterns. In regards, most of the activities in the theme park happened during the day and in accordance to a periodic interval [9]. As compared to temporal activity patterns, spatial activity patterns are more straightforward and less complex hence easier for ordinary readers to understand human mobility in the theme parks. In other environments such as the beach, wild, and zoo, the temporal activity patterns are most repetitive as compare to the spatial activity patterns as revealed by the research in Catalonia.

In tourism research, there has been growing concern in the understanding of the tourists' behaviors in terms of both activities and traveling. In regard, managers are able to utilize the information in strategic planning to enhance the creation of a sustainable tourism sector. Recently, Hong Kong has risen to be one of the target destinations for inbound tourists thanks to the investment by the movement in the tourism industry [24]. The analysis of their mobility patterns including tourism activates is paramount in helping to improve the service delivery to international tourists. However, there has been the challenge of the inadequate technologies used to capture and collect the data required for mining of the sequence patterns. Hong Kong is one of the South-Asian country rich in numerous tourists' attraction sites such as the Temple Street, Ocean Park, Lantana Island, TSIM Sha TSUI, Hong Kong Disneyland, and the Victoria Peak. The availability of the internet and social media platforms has made it possible to gather photos which contain the relevant information to be used in the process of mining the sequence patterns.

To understand the travel and activity patterns of the tourists in Hong Kong, a sample of 2100 tourists and 29,443 photos were used in the Sequential Pattern Mining (SPM) process. For the purpose of destination management, tourism managers have been looking for insights into tourism behaviors. With the knowledge of the activity patterns and travel behaviors, the tourism managers in Hong Kong are able to address the challenge of the transportation flows including placing adequate measures to improve service delivery in the country. Flickr has become a common resource for the travel behavioral analysis. However, there is a need to be careful and learn the tricks involved to be able to identify fake photo posted. The travel information from the photos can be used to infer the activities being taken in those moments of the photo taking. After the collection of the retagged photos, density clustering and market chain data mining techniques are useful in creating a suitable representation for the travel analysis task.

The data was found from Flickr by searching only the Hong Kong Inbound tourist information. Also, the data obtained was categorized into Western tourists and Asian tourists. The latter are those inbound tourists from adjacent countries such as Singapore and South Korea. The metadata extracted from Flickr contained the following parameters: Photo ID, Owner ID, GPS location with latitudes and longitudes, Location of the origin of the owner, and the time is taken. The time the photo was taken indicates the travel footprint of the tourists. The main purpose of the research study is to establish post popular destinations visited by the inbound tourist in the country. Also, it includes the mobility analysis and the activity pattern analysis in which the information was deduced from the retagged photos. Based on the analysis of the data obtained, it was found that the majority of the inbound tourists to Hong Kong come from Asia-Pacific, North America, and European countries. The place of origin of the users helps in the classification of the tourists to either Asian or Western. The categorization helps in identifying the similarities and differences between the Western and the Asian tourists in association to their respective traveling behaviors and activity patterns. The breakdown of the data collection statistics is shown in table 2.6 below:

Data Collection Statistics of the Hong Kong Inbound Tourists

Group	Number of tourists	Number of photos
Western tourist	1036 Tourists	15,990 Photos
Asian tourist	1064 Tourists	13,453 Photos
Total	2100 Tourists	29,443 Photos

Table 2.6: Data Collection Statistics of the Hong Kong Inbound Tourists

As evident from table 2.6, there were slightly more Asian tourist as compared to the Western tourists with figures 1064 and 1036 respectively. The almost similar number helps to increase a fair edge in the analysis whereby no group has a better advantage in terms of a number of tourists.

Due to the noise present in the geographic information, the DBSCAN algorithm was applied to remove it and avoid misleading information entering into the data analysis phase. The density clustering approach was used in the analysis to provide the most popular sites as well as the subsequent mobility and activity patterns of the tourists. The benefit of the density clustering method algorithm is that the classification of the destination sites is created automatically by the program. From the analysis, it was established that there were seven clusters representing the seven areas of interest by the inbound tourists in Hong Kong. Although the study did not address the sequential pattern analysis which includes the ordering of the activities. The algorithm only specifies the destination sites and the activity mostly involved by the tourists visiting the country. Based on the application of the density cluster algorithm, the seven areas identified include Hong Kong Central, TSIM Sha TSUI, Times Square Towers, Peak Tower, Hong Kong Metropolitan Area, Hong Kong International Airport, and the Tian Tan Buddha.

The TSIM Sha TSUI and the Hong Kong Central were found to be the most frequented destinations by both the Western and the Asian tourists. The Hong Kong Central is considered to be the ‘heart’ of the country and the hub of the tourism sector. In contrast, the western tourists showed more interest in the Peak Tower, and the Center Mong Kok while the Asian counterparts preferred the Hong Kong International Airport and the Times Square Tower. The summary of the popularity of interest of the identified areas is shown in table 2.7.

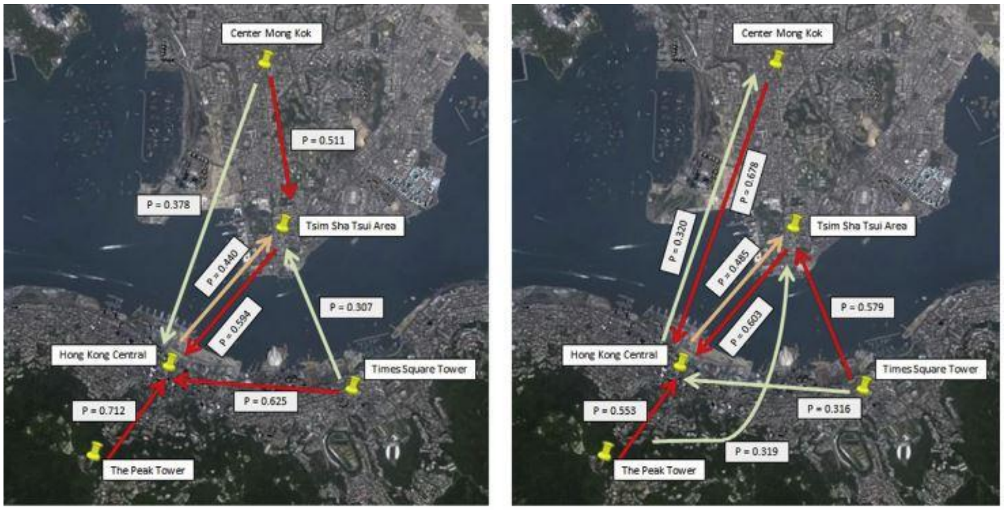
Areas of Interest by Popularity in Hong Kong

Group	Area of interest	Percentage (%)	Popularity rank
Asian tourist	Hong Kong Central	40.35	1
	Tsim Sha Tsui Area	38.80	2
	Times Square Towers	20.08	3
	Hong Kong International Airport	18.34	4
	The Peak Tower	14.19	5
	Center Mong Kok	10.52	6
Western tourist	Hong Kong Central	47.92	1
	Tsim Sha Tsui Area	44.64	2
	The Peak Tower	19.74	3
	Center Mong Kok	15.14	4
	Times Square Towers	12.22	5
	Hong Kong International Airport	10.99	6
	Tian Tan Buddha Statue	10.43	7

Table 2.7: Areas of Interest by Popularity in Hong Kong

Movement pattern analysis helps in the identification of the travel flows including the commonly used routes hence making substantive strategic decisions. The development of the travel management plans in Hong Kong heavily relies on the outcomes of the movement pattern analysis of the routes in terms of the paths of preferences to their various destination sites. The time factor extracted from the geo-tagged photos is a representation of the movement trajectory between the identified areas of interest from the density clustering analysis. Further analysis found out that majority of the tourists from both groups preferred to travel to destinations that are closer to each other perhaps to have more time spent on the attraction sites rather than traveling. Using the market chain technique and the areas of interest (AOI), the transition probability matrix from one point to another was created in a bid to analyze the movement of both groups from location 1 to location 2 and vice versa. Based on the results, most of the Asian tourists tend to flow to the Hong Kong Central from the surrounding areas with a higher probability value of 0.6. On the other hand, the likelihood of tourists moving from the Center MONG KOK to the TSIM Sha TSUI was above 50% (0.5). Having a probability value of 0.678, Asian tourists tend to move from Center MONG KOK to Hong Kong Central which is deemed to be the hub of the tourism sector in the country.

Movement Pattern Analysis of Asian and Western Tourists



(A) Asian Tourist (B) Western Tourist

Figure 2E: Movement Pattern Analysis of Asian and Western Tourists

Based on the market chain technique and the transition probability matrix, there are different likelihoods in terms of the respective probabilities of moving from one AOI to another. For example, the movement from The Peak Tower to the Hong Kong Central was 0.712 which is one of the highest likelihood of tourists moving in that direction regardless of the group or place of origin.

Apart from the movement pattern nasals, the managers are also interested in the daily activities of the tourists at the respective destinations. In order to address the problem of the overloads where individuals visit a certain destination over a short period of time, it is important to understand their traveling patterns to be able to make scheduling planning appropriately. The plotting of the probabilities against time helps in the identification of the specific times the groups of tourists visit a certain area [24]. The plots revealed that there were places where both groups are likely to visit at the same time hence a possible cause of congestion in the area. Further results established that both groups were likely to visit Hong Kong central between 11:00 hrs. and 16:00 hrs. Consequently, the tourism management needs to schedule and distribute the visiting hours where the assignment may be allowed to visit earlier or for certain days of the week.

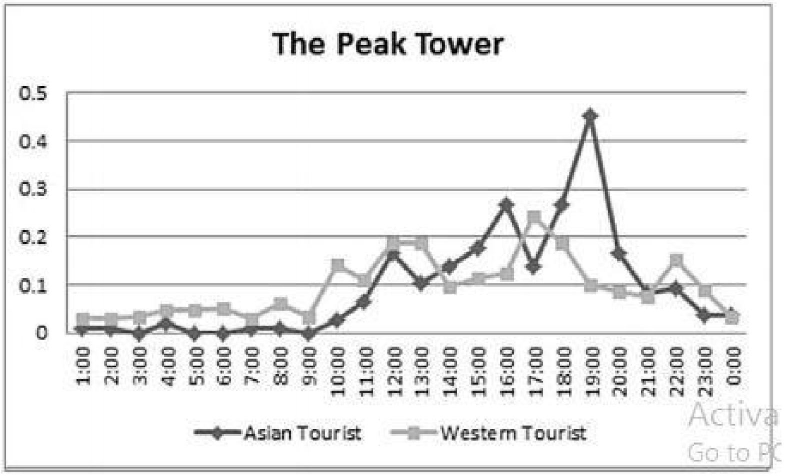


Figure 2F: The Peak Tower Visitation Likelihood

The majority of Western tourists are likely to visit the Peak Tower at 17:00 hrs. while most of the Asian tourist is likely to visit at 19:00 hrs. There is no possibility of overloads in the areas as compared to other destinations.

Figure 2G: The Hong Kong Central Visitation Likelihood

The highest Likelihood for Asian tourist to visit Hong Kong Central is between 12:00 and 13:00 hrs. The situation is the same for the Western tourists. In the early morning hours of the day, there are fewer activities noticeable at the Hong Kong Central destination sites as shown in figure 9.

In tourism research, most of the sequential pattern mining methods use the Geographical Information System (GIS) techniques which is crucial in the collection and organization of the data. However, this method shares common interactive dimensions such as location, time, and activity which are an important factor in the construction of the sequence activity patterns. The adoption of the GIS technique helps in avoiding the complexity of the multivariate pattern generalization which often affects the resulting outcomes of the tourist activity patterns. The analysis of the activity-travel behavior can involve a combination of methods or techniques to yield the result. Nonetheless, it is vital to ensure that the methods used are compatible with each other based on the parameters and characteristics associated with them. The United States is also one of the destination's targets for the majority of international tourists eager to explore more and learn about the culture and the living standards of the Americans.

Visualization is also another technique that is used in exploring the big data and deducing some of the problems that may be associated with each. Ideally, visualization is applicable at the data preprocessing stage of any given dataset. The technique is based on the analogy that people do learn better by visuals rather than in textual context.

A travel and activity survey conducted in Portland, metropolitan area in the United States resulted in viable data that can be used in the analysis of the travel-activity patterns. A sample of 7,090 households was used in the survey in which it was split into two phases. The first phase involved the recording of all travel-activities within a duration interval of 30 minutes. The second phase included a diary of the daily in-home activities' recorder within a similar time interval. From the sample, only 4,451 households having 10,084 individuals returned completed surveys. Moreover, the dataset included coordinates of the activity locations other than the workplace or at home. Besides, the study also included the geographic information (latitude and longitude) of the Metropolitan area. The inclusion of the coordinates in the association of the travel-activity data with the geographic environment of the metropolitan area during the visualization of the data.

In a space-time trajectory of the given dataset, the Z provides a vertical dimension of particular human activity. Using the geographic coordinates X and Y, each of the activities involved is located using the Z values. Each of the activities is expected to begin and end at a certain time. However, the activities recorded were done during the daylight excluding those done in the night. Following the use of the interactive geo-visualizations, the areas surrounding downtown Portland had the highest concentrations in terms of the percentage of the non-employment activities. Also, some of the clusters of the non-employment activities were noticed around the east of Gresham and west of Beaverton. Notably, the employment activities in the suburban as of the metropolitan areas are often one in the morning hours while the non-employment activities done during the lunch break were realized in the densely populated areas of the downtown. Also, the majority of the non-employment activities were found to be completed within a very short time with virtually 94% of the duration being under 5 mins. In the metropolitan areas of the United States, there is a strong association between location and non-employment activities based on the 3D space-time representation [78]. Another technique or geo-visualization method is the use of the 3D active density surfaces while comparing the density activity patterns of various activities within a real geographic sake.

Beijing, China has also emerged recently to be a major global tourist destination hence there is a need to understand the travel behavior and activity patterns to aid in better management. The understanding of tourist mobility helps in the marketing of the attractions as well as the design of newer attractions. A heuristic method uses the trajectories of focal tourists including the movement of past tourists in the country [63]. The modern tourist's experience has been on the mobility paradigm of the tourists. The concept of mobility has not only been focused on the local daily transportations but also the large-scale movement of objects and individuals across the globe. Recently, there has been an increasing urge to understand the movement patterns of tourists with the adoption of various methods such as behavioral analysis and market segmentation. With the advancement in technology, it is possible to keep track and record the paths taken by the individual tourists from one point to another with respect to time. Both pattern analysis and clustering techniques are used in the process. The heuristic approach is executed in three systematic steps namely: data collection, application of grid-based clustering, and the use of a heuristic algorithm to decipher activity attires from the data. Using the GPS technology, movement information of 117 tourists were collected in Beijing, China. Based on the results of the independent sample t-test, the heuristic approach was found to be more appropriate as compared to others.

The era of the internet has enabled the collection of movement data from the city, thus making the process to be completed in less time. The study of the aspect of the movement of tourists from one destination to another begun in the 1960s where it involved the mapping of the paths taken by tourists between destinations. It is assumed that destinations are not independent of each other and have a complementary relationship. From the assessment of the past movements of tourists from one destination to another, there is a possibility of forecasting the next destination in which the tourists are expected to take. In the Asia-pacific region, over a decade, there has been a spatial pattern of travel flows of tourists from one point to another with respect to time [63]. There is some preference in visiting a particular place or region during a certain time of the day. Although there are some factors of motivation and individual preferences, time plays a key factor including the climatic factor while visiting a destination.

In the United States, several models have been proposed in the analysis of the multiple trips taken by inbound international tourists. In Australia, the study of international tourists in terms of the movement patterns has attracted interest from researchers in recent years. In Hong Kong, an inductive approach was used to astonish the factor influencing the movement of tourists from one destination to another. Notably, space and time are the common attributes used in the formulation and the contracting of the travel sequences of tourists from the various destination in a country. In Hong Kong, the inductive approach in the study of the travel behaviors resulted in 78 discrete movement patterns which were subsequently classified into eleven movement styles. The study of the movement patterns of tourists has been hampered by the difficulty to collect information about the locations. However, this problem has been addressed with the incorporation of the emerging and existing technology to gather the information up to a micro level. Due to the diversity of the tourists traveling patterns, it has become a challenge to model inter-attraction patterns such as national parks and theme parks. A similar study in Scotland found that the movement patterns of tourists are highly depended on the road network. In regard, the majority of tourists tend to use road transport to move from one destination to another. For international tourists, flights were preferred to move from one country to another which was deemed to be less time consuming and safe [63] [65]. The analysis of the movement patterns of the tourists has included the qualitative factors affecting the traveling activity which has broadened the scope of the tourists' movement pattern analysis in the past few years.

Owing to the mapping and modeling of the tourists' movement patterns, little attempt has been made on the forecasting of the trends of the tourists in the coming years. Consequently, Markov chains have been used to predict the movement patterns of the tourists in the past two years. The method was proposed and utilized effectively by incorporating the time dimension. The Markov-based method was implemented on the assumption that the probability of the next location is independent of the past movement of the tourists. However, several studies have relied on the past historical records of the movements of the tourists to predict the future movements. The concept of the reliance of the past tourist movements has ensured the accuracy of the forecasts of the travel pattern movements. The data collected by administering questionnaires is prone to faults and errors hence the need to stick to the appropriate technology to collect the required data for the research study. Ideally, the movement of people has been deemed unpredictable since they can change abruptly although there has been a desired schedule executed before. The undertaking of the movement patterns has enhanced the urban planning and improvement of the transportations systems within the country of focus such as Hong Kong and Canada.

The incorporation of data mining techniques and Markov chain methods have resulted in the accurate prediction of human mobility in recent years. Notably, the accuracy is dependent on the validity of the historical data gathered in the past. The growing concern in the tourist mobility prediction hence the need to resolve the tradeoff between efficiency and accuracy. The Markov model has developed the phenomenon of phasing out the past data. Subsequently, the debate among the researchers has continued on the viability of using past historical records on the movement of tourists to predict the future trends. The Markov-based methods forecast the future trends of the movement of tourists by ignoring the contribution of the historical data. The disregarding of the historical data may have an effect on the prediction performance of the movement patterns. Instead of the reliance of the Markov-based methods with the uncrating in the use of the historical data in the prediction of the next location, the adoption of the heuristic approach helped in solving the issue. The implementation of the heuristic prediction algorithm has been found to strike a balance between efficiency and accuracy of the travel movement patterns of the tourists.

The major difference between the Markov chain technique and the heuristic prediction algorithm is that the latter values and uses the historical data of the movement of the tourists in order to make predictions of the future trend. For the Markov model, the emphasis is put on the prediction of the next location based on the current location. In forecasting, the activity information or data is paramount in order to effectively predict the future trend, hence making heuristic prediction algorithm to be appropriate as compared to the Markov modeling technique. The algorithm relies on the data collected by the GPS technology. In the data preparation phase, the grid-based clustering method is applied to break down the raw data collected and organize them meaningfully including extracting the possible patterns through the visualization technique. The movement mining forms the basis of the analysis which also includes the incorporation of the pattern-matching algorithm. After the clustering has been made, the mining of the movement patterns is the next phase which also includes the matching of the similar patterns. If a pattern matching fails, the pruning method is used by improving the parameters of the algorithm. With the help of the heuristic prediction algorithm, a probability matrix is developed to be able to predict the next location based on previous patterns obtained in the second phase. The movement prediction is largely dependent on the probabilistic analyses carried out during the execution of the process.

According to the current statistics, the summer palace in Beijing China is one of the most popular tourist attractions, especially for inbound tourists. It is considered to be an ideal destination during the summer seasons of the year. Quite a number of both local and international tourists tend to visit this attraction site on an annual basis. Based on the evaluation by UNESCO the summer palace was recognized as one the of the world heritage sites making into the list of the international tourist market in the millennium year [63]. Owing to its large area, the palace has over 100 tourist spots and is suitable for hiking because of the stretch of the longevity hill and the Kunming Lake. Residential and scenic areas are also present in the summer palace. A survey conducted in December 2010 took approximately two days with the survey team able to gather demographic characteristics and the spatial-temporal behaviors of the tourists. Also, Questionnaires were administered to the tourists to supplement the survey done using the handheld GPS tracking devices. To avoid bias, random sampling technique was used to arrive at a sample of 117 tourists. The tracking devices were able to capture the data on the traveling and activities of the tourists and recorded them accordingly. The process was done by giving the volunteers the GPS tracking devices at the entrance and then collect them at the exit point. Notably, the use of the tracking devices was a challenge since most of the tourists were reluctant into accepting the tracking devices. The assurances of personal safety and confidentiality of the information collected were issued to the participants to uphold the ethical consideration of the study. Upon returning the GPS tracking devices, the respondents were requested to complete a short questionnaire about their touring experience at the Summer Palace. With minimal error, the devices were able to gather location information including the latitudes and the longitudes of the location visited. Also, time, velocity, and the information about the directions taken were also recorded and channeled to a central server for storage and future use. The study adopted both manual and automatic methods in collecting the data required for the analysis of the movement patterns of the tourists. The questionnaire was split into five sections namely: personal information, route information, activity choices, consumption information, and the sites visited by the participant. Furthermore, the respondents were asked to draw the route taken from the entrance point to the time of the exit based on the guide map of the Summer Palace. While administering the questionnaire, pictures of the various sites of the Summer Palace were issued to the respondents to help them refresh the memories hence making the process to be fast. The validity rate of the survey was 94%. In regards, 111 respondents participated fully and completed the survey. Out of the 111 responded, the male participants were 53% and 47% were female respondents. Interestingly, 30.2% of the respondents were categorized as first-time tourists meaning that majority of the tourists visiting the Summer Palace were repeat visitors who would have at some point in the past visited the area.

Although the error is inevitable, the GPS technology has been found to be most effective in the collection of the data of the movement patterns of tourism in a given attraction site. To date, most of the scholar and researchers are relying on the technology to obtain data for their research studies. For example, an error may occur when the tracking device reports indirectly a location point within the attraction site. Such errors are often carried into the analysis as it is difficult to point out the error easily form the recorded information. To be able to avoid such errors, there is a need to improve the performance and efficiency of the devices, for instance by upgrading the memory and the processor to be able to process the information promptly. The errors may be attributed to the poor performances of the devices while in use [75]. From the 111 respondents of the tourists engaged, a total of 5,847 location points were captured by the GPS tracking devices. Afterwards, the clustering method was implemented appropriately to figure and mine the possible movement patterns and the relationship between the accuracy and the difficulty. The GPS density function assisted in the establishment of the association between the characters and the GPS points gathered from the Summer Palace. There are minimum point values which determined the points that ought to be phased out in the process.

Based on the characteristic trajectories, as the minimum points value increased, the more grid cells are removed from that given trajectory. The characteristic trajectories were classified into 0, 10, 20, and 30 minimum points' values respectively. Ideally, the greater the minimum point values, the less the grid cells within the space chosen. In regard, the grid cells for the four characteristic trajectories based on the minimum point values were 141, 118, 85, and 61 respectively [75]. The purpose of the analysis involving the characteristic trajectories and the grid cells analysis is to underline the importance of reading the complexity of the data processing while ensuring the most important information is retained. The concept of minimum point's value had the possibility of affecting the performance of the prediction of the movement trend of the tourists at the Summer Palace in Beijing. The prediction accuracy is carried out by evaluating the performance of the heuristic approach. The evaluation is one by using heuristic probability inspection method to split the data into two while making another set as the training set. Based on the prediction accuracy and the experimental results, it was found that the heuristics approach is more accurate as compared to the Markov-based method as shown in figure 2H below.

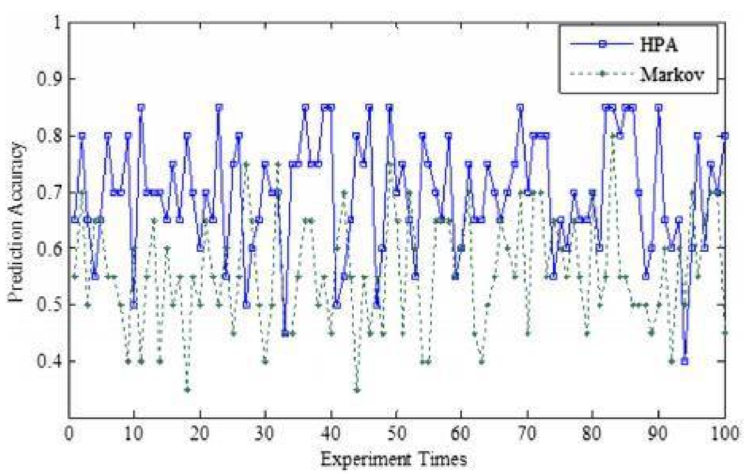


Figure 2H: Prediction Accuracy of Heuristic Prediction Algorithm and Markov Chain Method

Based on this figure, the heuristic prediction algorithm has a higher prediction accuracy of the next movement location of the tourists as compared to the Markov chain technique. Therefore, the heuristic approach performs better in predicting the movement patterns of the tourists in a given attraction site, in this case, the Summer Palace in Beijing China. The estimation and the prediction of the tourist's location has an impact on the laying out of the service points such as toilets, and in crowd management. Using the heuristic approach, all the 111 GPS trajectories were used to forecast the spatial distribution of tourists in the Summer Palace. The time intervals used include: 11:00, 11:30, 12:00, and 12:30 representing a 30 min time interval. The result of the prediction can be used to mine valuable information which helped in the planning of the management of the crowds especially during the high peak season of the international tourists. One of the applications of predicting individual tourists' paths using the heuristic approach is the development of personalized tour guides programs or apps which need users to provide demographic information upon registration which then reveals the direction including the next location to take automatically. From the prediction results obtained from the 100 spots in the Summer Palace, 45% of the tourists moved to spot 22 immediately after completing visiting spot 13. On the other end, the tourists who successively visited spots 4, 5, and 13 were found that only 20% of them visited spot 22 as their next destination. The algorithm has some viable applications for the big date in making a prediction of the movement patterns. Having chosen the "spot 13, Spot 22" combination, out of the 45 tourists who visited the Summer Palace, 37 of them chose to exit the attraction site only after visiting two spots [63]. In regular time intervals, using GPS density, congestion areas can be noticed from the map hence better management of the crowd in certain times of the day.

In tourism research, the tourists' travel and activity behavior has been affected by the seasonal variation of the climatic conditions in the attraction's sites. The seasonality aspects have had an impact on the tourists' experience including the production of different time-space consumption patterns. South Korea is one of the consistent destinations for most of the international tourists. Ideally, the majority of the inbound tourists in the country tend to visit the Seoraksan National Park. Therefore, there is a need to investigate the activity patterns in the attraction sites to be able to understand the activities in which tourists style edge in. The information regarding the spatial activity patterns of the tourists helps the tourism industry in Korea to better manage the sector especially in cases where they expect to receive a higher number of tourists into the country. The analysis of the activity patterns includes both local and international tourists. However, there may be a significant difference in terms of the spatial activity patterns of the local and the inbound tourists because of the varied taste and preferences. Local tourists may have different priorities because they are used to the Summer Palace as compared to the first time tourists in the attraction site. Data was collected in 2015 from January to December which resulted in the capturing of the activities of 1,206 mobile application users. Virtually every method or approach in the sequential pattern mining adopts a mobile technology to collect the data of the geographic location of the attraction sites and those of the tourists. Out of all the spatial and analytical techniques, the study was dependent on the GIS technology to provide the analysis of the spatial patterns of the tourists. The results and findings revealed variations in terms of activity distribution in the respective points. Therefore, the results were helpful to the park managers to be in a position to manage the negative impacts of the natural resources as well as enhancing the tourists exorcise in the Seoraksan National Park in South Korea [75]. By doing so, there is the possibility of first-time tourists returning to the attraction sites in the future. In every destination, there is always the presence of first-time tourists hence it is important to offer better tourism experience to boost the chances of the first-time individuals returning back to the attraction site or destination.

The national park is often the preferred destination by tourists because they offer more tourism opportunities and diversity in recreation as compared to other attraction sites. As a result of the annual rising in the number of tourists registered at the national parks, it has been a challenge of the park management agencies to strike a balance between protecting the natural resource and creating an exemplary tourism experience for the tourists. Due to the climatic variation at the Seoraksan National Park in South Korea, the tourists and recreation activities are deemed to be a seasonal experience. In regard, there is a need to understand the activities which are mostly available at certain seasons of the years to enhance the tourists' experience, especially for the first-time tourists. In Australia, seasonal variation was noticed in all the 23 protected national parks in the country. Across all the six climatic zones namely: the Grassland, Desert, Tropical, Unitarian, Alpine and Subalpine, there was seasonal variation in terms of the spatial activity patterns of the tourists. Both alpine and subalpine area are influenced by institutional factors such as holidays.

The GPS tracking techniques haves been utilized in the collection of data about the activities in numerous national parks including other tourism destinations. In essence, the tracking techniques have the capacity to evaluate the spatio-temporal patterns of the tourists including identification of the hotspots for mountain bike riders, runners and the hikers in the park. It is the responsibility of the park manager to create the boundary for the animals and the recreation activities to avoid human-wildlife conflict which may result in jeopardizing the safety of the tourists. In prior studies, researchers have used GPS tracking methods in the assessment of the activity patterns and tourists' movement flows for instance in the Dwingelderveld National Park in the Netherlands [75]. Generally, the GPS tracking methods have been proving to be more effective as compared to the traditional survey techniques in the assessment of the activity patterns in the national parks. The limited battery life of most of the past GPS tracking devices resulted in the restriction of the sample sizes used in the analysis of the activity patterns. In an era of widespread smart phones, the development of mobile user applications which are more user-friendly has been on the rise. Individuals are being accustomed to the technology and use the tracking applications which are able to record time, as well as the location of the activities. The mobile exercise applications were used at the Seoraksan National Park to track the activities of the tourists with respect to time. The findings of the tracking of the activities of the spatial pattern analysis helped the park managers in the understanding of the tourist flow across all the seasons and days in South Korea.

Among the parks in South Korea, the Seoraksan National Park (SNP) was chosen because it was the most popular and is one of the designated destinations according to the UNESCO biosphere reserve. In 2015, the SNP attracted approximately four million tourists wanting to experience the variety of cultural and natural attractions at the site. The influx in the numbers of tourists not only resulted in a management crisis but also posing a threat to environmental conservation. Therefore, the understanding of the spatial activity patterns of the tourists across the various seasons in the country helps in better management of the tourist and enhancing the exercise as well as protecting the natural resources from degradations. The ability of the digital tracking technology being able to offer successive and accurate information aided in the choice of the use of the GPS-based mobile application in the collection of the data in the study area [70] [75]. In South Korea, Triangle, one of the popular mobile exercise applications, was used in the collection of the data at the SNP. The application concealed the identity of the users while capturing the data of time, activity type, and the coordinates of the location in which the activity happened. The dataset was extracted from the Triangle database having the coordinates do the location and the starting and ending points of the daily activities.

The geographical information about the boundaries of the park was useful in the spatial pattern activity analysis. From the analysis of the dataset, most of the participants use the exercise application Triangle for hiking at 94.3%. Subsequently, the application was used for walking, jogging, cycling, and other activities at 4%, 0.7%, 0.7%, and 0.3% respectively. The GIS-based spatial analysis and mapping of the tourist's activities is still an evolving tool for mining activity patterns of the tourists. As a result, the method is poised to improve in the near future. In the analysis of the dataset, the SNP was aggregated into 69 cells thus being able to evaluate the activities mostly done in the larger part of the park. The results of the spatial activity pattern analysis reveals a seasonal distribution of the activity points. During the winter season of between December and January, there were few activity points noticed while on the other end, the largest number of activity points happened during the fall season of between September and November. Less activity points accounting for about 13.5% during the spring which normally occurs between March and May in South Korea. Primarily, the summer has the most of the activity points at 31.6% and in fall at 44%. Majority of the tourists tend to visit the SNP and engage in hiking activities during the summer season unlike during the winter and spring seasons [75].

The distribution of the activity points was done in two formats namely: weekdays (Monday-Friday) and weekends (Saturday-Sunday). From the findings, it was revealed that most of the tourists' activities were concentrated on the weekends more than during the weekdays. Across all the four seasons, the weekends accrued most of the activity points based on the spatial analysis using the GIS-based systems. Precisely, the activity points for the weekend were more pronounced during spring and summer as opposed to winter and fall. In all the seasons, the fall season accrued most of the activity points indicating that the tourists went to the SNP in Korea in the same season while majority engage in hiking [75]. Therefore the sequence can be represented as: Hiking > Walking > Jogging > Cycling > Other Activities.

The activity point's spatial analysis also revealed the activity hotspots for all the seasons. Apart from the winter and fall, the spring and summer have the same number of activity hotspots of seven. The winter season had 6 hotspots while the fall had 8 hotspots. Out of the 69 cells, only seven of them were found to be statistically significant as shown in table 2.8.

Spring Activity Hotspots

Cell 8	ONKYO waterfall, SAIGOL valley
Cell 16	OSAEK spring, GEUMGANG MOM, JUJEONGOL valley, YONGSO waterfall
Cell 23	SEUNG waterfall, HAUNTER TEIGN valley
Cell 24	SEORAK waterfall, DOUGAL valley
Cell 42	YANGPOK waterfall
Cell 52	SHINHEUNGSA valley, GWONGEUMSEONG peak, GWONGEUMSEONG rock, BISEONDAE rock
Cell 61	HEUNDEUL BAWI rock, ULSANBAWI rock

Table 2.8: Spring Activity Hotspots

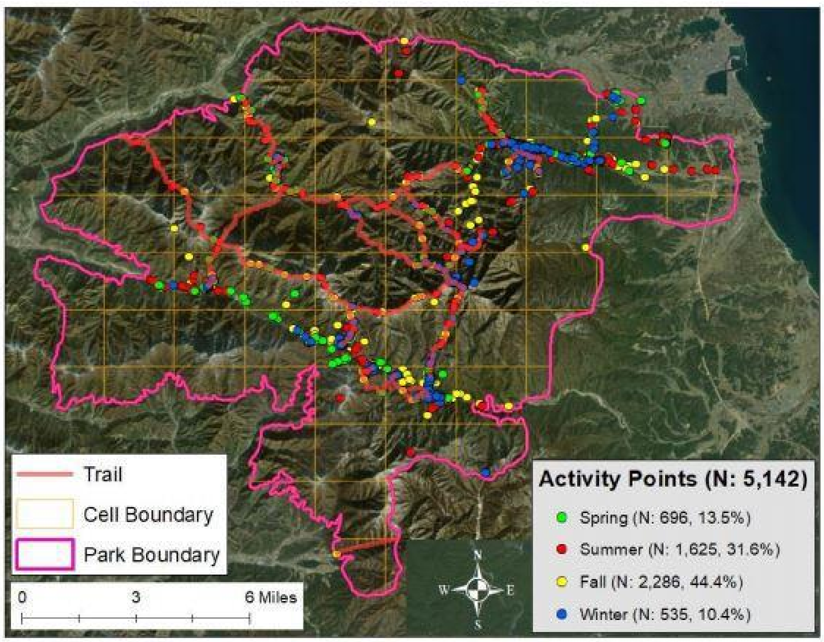


Figure 2I: The Seasonal Distribution of the Activity Points at the SNP

The fall season has the highest composition of the activity points followed by summer, spring, and winter respectively. Ideally, all seasons recorded some tourist activities.

Ideally, there have been several attempts to evaluate and assess the traveling behaviors of the tourists across the world by most of the scholars and researchers. Some of the researchers have focused on analyzing the activity and movement patterns of the tourists using different methods. To date, several methods have been proposed and implemented successfully. In a nutshell, the use of technology in the collection of the data for the analyses of the tourist behavior has been consistent for most of the researches. Notably, attempts have been given on understanding the movement of tourists from one location to another while also focusing on the activities carried out at those locations. Also, some of the research studies have provided the most interesting patterns in terms of the frequency by the tourists. However, there is a gap in trying to decipher the sequences of the activity of these patterns with respect to time. The time factor has been sparingly utilized in most of the research. To bridge that gap, this study will focus on establishing the sequential activity patterns of tourists in Singapore. Although there has been past research on the southern axis, the existence of this gap in tourism research prompted the need to conduct the research study in Singapore.

CHAPTER 3: METHODOLOGY

3.1 AN OVERVIEW

There are many algorithms and data mining techniques that can be applied onto this study, however, the aim is to choose the most efficient and reliable method according to the dataset. The Prefixspan algorithm has been proven to outperform the APRIORI algorithm including other emerging algorithms for sequential tourist activity pattern mining [79]. In essence, the Prefixspan algorithm runs considerably much faster as compared to both APRIORI and Free Span algorithms. In regard, the Pattern-Growth method which uses Prefix-Scan algorithm will be used in the study.

Notably, the Prefixspan algorithm in sequential pattern mining uses the prefix subsequence from the database of a certain attribute to predict the postfix subsequences in the database. For a given projected database, the frequent patterns are obtained by exploring the local frequent patterns. It can be done by either level by level projection or bi-level projection. Based on the performance evaluation, the bi-level projection is better for large datasets. Without the loss of generality, the activities or items in the patterns can be arranged in any given order. Primarily, the patterns of the tourists' activities may be organized in any given manner.

The pattern-growth methodology is used in mining structured and multidimensional sequential patterns from a large sequential database. The prefix span is a more efficient method as compared to the free span because of the reduced projection rate of the databases and also it adopts an ordered growth phenomenon [80]. The prefix span algorithms uses tourists database to solve the problems of mining sequential patterns using the minimum support threshold provided. Using both real and synthetic datasets, an experimental study was conducted to establish the performances of the different algorithms namely, free span, prefix span, SPADE, and GSP. The algorithms are both APRIORI-based and pattern-growth based. All the algorithms were run in Microsoft Windows server using the C++ programming language. With the lower minimum support threshold, the prefix span performed much faster than the GSP and SPADE. Also, in terms of runtime and memory usage, the prefix span consumed less memory and has faster execution time to yield complete sets of sequential patterns.

There are various reasons that have been given as to why the prefix-span emerges to be the appropriate method for mining sequential patterns in the modern world. The further implementation of the pseudo-projection techniques helps in improving the performance of the algorithm which is under the pattern-growth method. Unlike the commonly used APRIORI-based approach which relies on the candidate generation to obtain the sequential patterns, the prefix span only relies on the counts of the frequent local items obtained from the dataset. Consequently, the item sets are appended together to form the sequential patterns [79] [80]. The rate of growth of the number of complete sets of sequential pattern rises whenever the minimum support threshold drops drastically. Interestingly, the prefix span methods work by growing long sequential patterns from the short ones by focusing on the subspace while dividing the search space. The use of the candidate generation and test approach consumes much time as exhibited by the GSP algorithm in the experimental study. As the prefix span focuses on growing longer sequence using the shorter ones, the APRIORI approach often scans all the frequent sequences including many irrelevant sequences which are eventually added to the overhead. In regard, it tends to increase the number of execution times required to generate complete sets 'n' of sequential patterns from the dataset [80] [81] [83].

Ideally, the prefix span approach uses the divide and conquer methodology and it does not involve the generation of candidates hence consumes less and stable memory size of the computer. The SPADE and GSP method relying on candidate generation need a substantial amount of the memory space to hold the generate candidate sets [79] [80].

3.2 PRIVACY CONCERNS

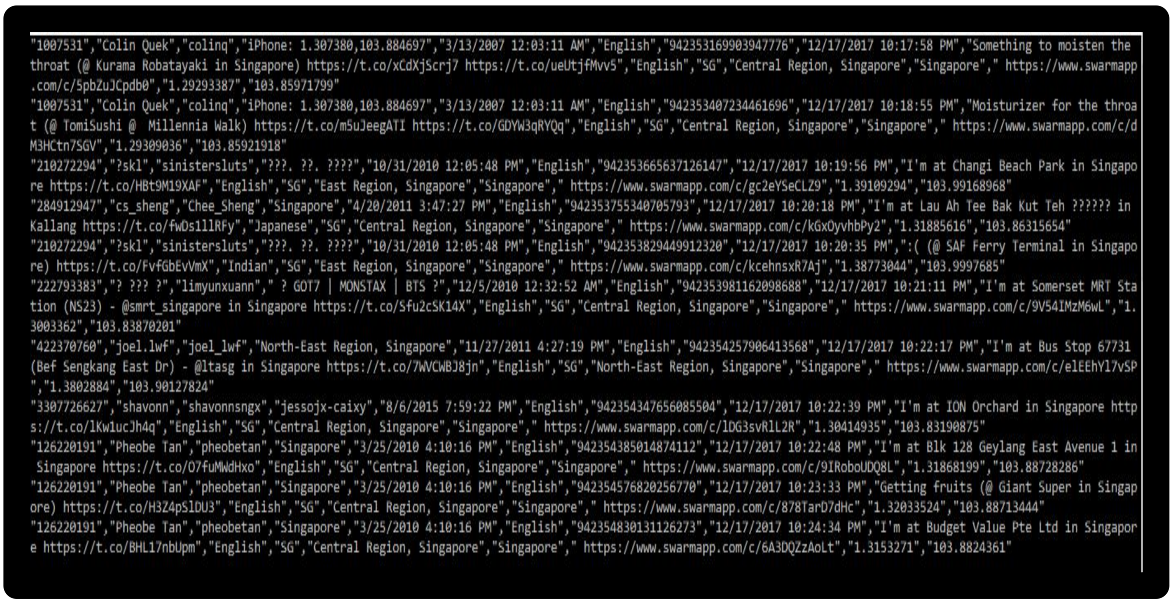
The research is focused in collecting sensitive data about tourist travel location, date, times and activities. The respect for privacy enriches social and personal interaction by providing contexts for the development of various kinds of relationships and multiple dimensions of personality. Privacy is needed to enable a person to deliberate upon and establish his/her views and opinions. Privacy has been widely recognised as a value important to both individuals and society. This recognition, however, has been unsuccessful in converting the value into a clearly defined, protectable legal standard. Keeping the privacy concerns in mind, the research will be implying and using the general privacy policy of the Foursquare platform. The users that create an account on Foursquare to use its consumer services have already entered an agreement with it, thereby giving Foursquare the right to collect and use their data for the purpose of improvement, customization and further enhancement. The data collected is thus anonymised, aggregated and pseudonymous to make sure the research does not pose any threat to users' data privacy laws. A detailed list and account of Foursquare data privacy laws can be viewed at https://foursquare.com/legal/privacy.

The issue of privacy was addressed in this research through the variety of tactics utilised to ensure that the personal tourist information used from which data was collected was encrypted and therefore anonymously categorised. As the information that was accessed is regarded as highly personal; ranging from exact locations of check ins, times and dates, travel information and activities, there were a variety of protocols in place to ensure the protection of the consumers privacy that were in anyway implicated through this data collection and research. These protocols include inspecting Foursquare data privacy laws and regulations that state that any data collected through users, consumer services and third parties is safe to use for research and development purposes. Users that are using Foursquare to make such information publicly available have therefore agreed to Foursquare terms and regulations that include operating, improving and maintaining the consumer service, communicating with users, protection of rights, safety and property of Foursquare, the consumer services and users and employees as well as any third party.

Foursquare further state on their platform privacy policies that they provide their data with third party partners and service providers to provide or perform services on their behalf. These providers include the following categories; cloud storage providers, IT service providers, advertising networks and analytics and search engine providers. Foursquare further state that they disclose anonymous data to other third parties including developers who use API or SDK or other enterprise customers for analytical or other purposes. With these laws, regulations and policies in mind, the privacy of all consumers data implicated has been protected in this research through a series of encryptions, anonymous categorisations and aggregations.

3.3 DATA COLLECTION

During the early data collection process, tweets with tags that contained the keywords such as "Holidays in Singapore", "Travelling to Singapore", "Vacation in Singapore", "Singapore holidays" and "Singapore Tourism" were filtered via the twitter live streaming. However, the final dataset contained only the tweets which were associated with Foursquare Check-ins. In order to retrieve the data, the Twitter application programming interface (API) has been used. Like all others APIs, this program allows the programmers to access the Twitter's server to gain the specific data needed for the research. The Twitter-API full record can be found here: https://developer.twitter.com. The time duration of the research was one year, starting in October 2016 and ending in October 2017. The data has been collected by using Twitter Developer API, which extracted the tweets of various tourists, who had visited Singapore during the one-year duration.



"1007531","Colin Quek","colinq","iPhone: 1.307380,103.884697","3/13/2007 12:03:11 AM","English","942353169903947776","12/17/2017 10:17:58 PM","Something to moisten the throat (@ Kurama Robatayaki in Singapore) https://t.co/xCdXjScrj7 https://t.co/ueUtjfMvv5","English","SG","Central Region, Singapore","Singapore"," https://www.swarmapp.com/c/5pbZuJCpdb0","1.29293387","103.85971799"
"1007531","Colin Quek","colinq","iPhone: 1.307380,103.884697","3/13/2007 12:03:11 AM","English","942353407234461696","12/17/2017 10:18:55 PM","Moisturizer for the throat (@ TomiSushi @ Millennia Walk) https://t.co/m5uJeegATI https://t.co/GDYW3qRYQq","English","SG","Central Region, Singapore","Singapore"," https://www.swarmapp.com/c/dM3HCtn7SGV","1.29309036","103.85921918"
"210272294","?skl","sinistersluts","???. ??. ????","10/31/2010 12:05:48 PM","English","942353665637126147","12/17/2017 10:19:56 PM","I'm at Changi Beach Park in Singapore https://t.co/HBt9M19XAF","English","SG","East Region, Singapore","Singapore"," https://www.swarmapp.com/c/gc2eYSeCLZ9","1.39109294","103.99168968"
"284912947","cs_sheng","Chee_Sheng","Singapore","4/20/2011 3:47:27 PM","English","942353755340705793","12/17/2017 10:20:18 PM","I'm at Lau Ah Tee Bak Kut Teh ?????? in Kallang https://t.co/fwDs1llRFy","Japanese","SG","Central Region, Singapore","Singapore"," https://www.swarmapp.com/c/kGxOyvhbPy2","1.31885616","103.86315654"
"210272294","?skl","sinistersluts","???. ??. ????","10/31/2010 12:05:48 PM","English","942353829449912320","12/17/2017 10:20:35 PM",":((@ SAF Ferry Terminal in Singapore) https://t.co/FvfGbEvVmX","Indian","SG","East Region, Singapore","Singapore"," https://www.swarmapp.com/c/kcehnsxR7Aj","1.38773044","103.9997685"
"222793383","? ??? ?","limyunxuann"," ? GOT7 | MONSTAX | BTS ?","12/5/2010 12:32:52 AM","English","942353981162098688","12/17/2017 10:21:11 PM","I'm at Somerset MRT Station (NS23) - @smrt_singapore in Singapore https://t.co/Sfu2cSK14X","English","SG","Central Region, Singapore","Singapore"," https://www.swarmapp.com/c/9V54IMzM6wL","1.3003362","103.83870201"
"422370760","joel.lwf","joel_lwf","North-East Region, Singapore","11/27/2011 4:27:19 PM","English","942354257906413568","12/17/2017 10:22:17 PM","I'm at Bus Stop 67731 (Bef Sengkang East Dr) - @ltasg in Singapore https://t.co/7WVCWBJ8jn","English","SG","North-East Region, Singapore","Singapore"," https://www.swarmapp.com/c/elEEhYl7vSP","1.3802884","103.90127824"
"3307726627","shavonn","shavonnsngx","jessojx-caixy","8/6/2015 7:59:22 PM","English","942354347656085504","12/17/2017 10:22:39 PM","I'm at ION Orchard in Singapore https://t.co/lKw1ucJh4q","English","SG","Central Region, Singapore","Singapore"," https://www.swarmapp.com/c/lDG3svRlL2R","1.30414935","103.83190875"
"126220191","Pheobe Tan","pheobetan","Singapore","3/25/2010 4:10:16 PM","English","942354385014874112","12/17/2017 10:22:48 PM","I'm at Blk 128 Geylang East Avenue 1 in Singapore https://t.co/O7fuMWdHxo","English","SG","Central Region, Singapore","Singapore"," https://www.swarmapp.com/c/9IRoboUDQ8L","1.31868199","103.88728286"
"126220191","Pheobe Tan","pheobetan","Singapore","3/25/2010 4:10:16 PM","English","942354576820256770","12/17/2017 10:23:33 PM","Getting fruits (@ Giant Super in Singapore) https://t.co/H3Z4pSlDU3","English","SG","Central Region, Singapore","Singapore"," https://www.swarmapp.com/c/878TarD7dHc","1.32033524","103.88713444"
"126220191","Pheobe Tan","pheobetan","Singapore","3/25/2010 4:10:16 PM","English","942354830131126273","12/17/2017 10:24:34 PM","I'm at Budget Value Pte Ltd in Singapore https://t.co/BHL17nbUpm","English","SG","Central Region, Singapore","Singapore"," https://www.swarmapp.com/c/6A3DQZzAoLt","1.3153271","103.8824361"

Figure 3A: Data Collection from the Twitter API Interface

3.4 FOURSQUARE

Foursquare is a local search-and-discovery service mobile app which provides search results for its users. The app provides personalized recommendations of places to go to near a user's current location based on users' previous browsing history, purchases, or check-in history. The tourists can book their trips outside their home country with Foursquare which can inspire proposals on attractions to visit, give restaurants and general city guides — all dependent on Foursquare information. The tourists can check-in, to various places which gives area-based information to the absolute greatest retail names in the country. First is the Foursquare City Guide application that enables users to discover new places, with proposals from different users. The second application is an expansion of the 'registration' diversion that underscored its underlying ubiquity.

Foursquare gives users a chance to score real-world perks each day by registering with the different place as well. As Foursquare has developed and accumulated more information, it has turned into an extraordinary wellspring of suggestions which helps its users / tourists in multiple ways.

Foursquare check-ins were used in this research owing to the ease, accuracy and speed of data collection as well as the growing popularity of Foursquare around the globe. The check-in data was collected via twitter streaming using the twitter developer API to extract all tweets of tourists visiting various destinations in Singapore. Using a specialized coded program, the raw dataset was filtered to obtain only the tweets with Foursquare check-ins of the tourists or users. The data was collected for a period of between seven to nine months which include the category of venues to help figure out which activities the tourists engaged in at a particular time. Each check-in had the following attributes: check-in ID, user ID, time and geographic coordinate (latitude and longitude), category and subcategory of the check-in's location, i.e., the type of place where it occurred. Notably, the ideal dataset was collected for the sample period of 6-8 months.

3.5 DATA PRE-PROCESSING

As soon as the raw dataset had been collected from the Twitter API, it was filtered by using a customized coded program. The tweets that were collected for this research were connected with the Foursquare check-ins. The filtered dataset consisted of tweets of 700 tourists, which were collected for a period of 6 to 8 months. The Foursquare category information was extracted using a coded program to determine the tourist behavior in a given time period.

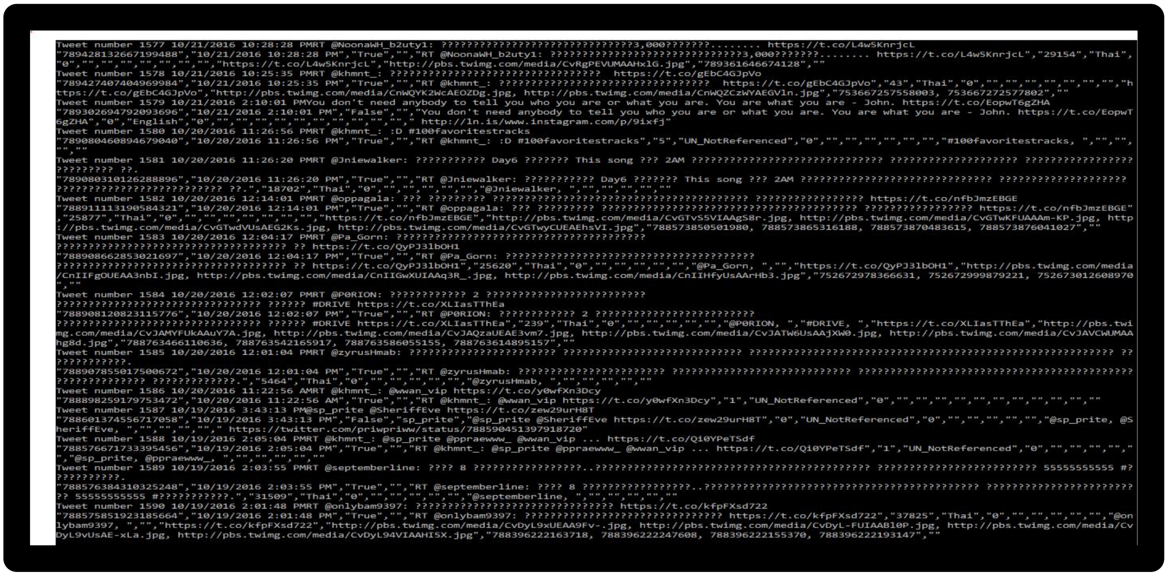


Figure 3B: Raw Data Filtering

The raw data obtained from the Twitter's API contains numerous anomalies such as outliers, missing values, and inconsistencies of entries. The pre-processing stage helps in removing the inconsistencies hence preparing the dataset for analysis and subsequent application of the sequential pattern mining algorithms. Using the SPSS software, the data was run through several validation tests to ensure that the data is normal and fit for analysis. With the software, the missing values and outliers were handled accordingly. Due to the nature and the purpose of the research study, some of the attributes such as check-in ID and Subcategory were removed to narrow done on the importance attributed in line with objectives of the study. Having completed all the procedures, the dataset was exported to an excel file to be ready for analysis.

Below is an example of one tourist file; a female tourist from Thailand with userID 13evl_masa and has a total of 21 Check Ins within 3 days time period (from 13^{th} Oct 2017 - 15^{th} Oct 2017).

1	UserID	Gender	UserLocatic	CheckInID	CheckInLink	YMD-Time	VenueID	Venue Nam	Latitude	Longitude	VenueCate	VenueCategoryID
2	13evl_masa	female	Thailand	59e304b581	https://www.swarmap	2017 10 15 14 48 2	4b0bd124f9	Singapore (	1.3566	103.9884	Airport	4bf58dd8d48988d1ed931735
3	13evl_masa	female	Thailand	59e2e248de	https://www.swarmap	2017 10 15 12 21 2	4b058815f9	Mustafa Ce	1.3099	103.8557	Departmen	4bf58dd8d48988d1f6941735
4	13evl_masa	female	Thailand	59e2c383c8	https://www.swarmap	2017 10 15 10 10 1	4b9df38cf9	ArtScience	1.2863	103.8595	Art Museur	4bf58dd8d48988d18f941735
5	13evl_masa	female	Thailand	59e28a826e	https://www.swarmap	2017 10 15 6 6 58	4b2c8c2ef9	The Merlio	1.2869	103.8544	Outdoor Sc	52e81612bcbc57f1066b79ed
6	13evl_masa	female	Thailand	59e2152064	https://www.swarmap	2017 10 14 21 46	4b05880ef9	Clarke Qua	1.2902	103.8457	Canal	56aa371be4b08b9a8d573562
7	13evl_masa	female	Thailand	59e1d531a7	https://www.swarmap	2017 10 14 17 13 2	4b949b12f9	Gardens by	1.2823	103.8642	Garden	4bf58dd8d48988d15a941735
8	13evl_masa	female	Thailand	59e1bf62e6	https://www.swarmap	2017 10 14 15 40 1	4c7157b834	Marina Barr	1.2808	103.8709	Scenic Look	4bf58dd8d48988d165941735
9	13evl_masa	female	Thailand	59e1aa1216	https://www.swarmap	2017 10 14 14 9 2	4b058815f9	Plaza Singa	1.3001	103.8452	Shopping M	4bf58dd8d48988d1fd941735
10	13evl_masa	female	Thailand	59e18a0711	https://www.swarmap	2017 10 14 11 52 3	4bca492c51	Telok Blang	1.2791	103.8112	Park	4bf58dd8d48988d163941735
11	13evl_masa	female	Thailand	59e1720b08	https://www.swarmap	2017 10 14 10 10 1	4b57b23df9	Henderson	1.2762	103.8152	Scenic Look	4bf58dd8d48988d165941735
12	13evl_masa	female	Thailand	59e1646075	https://www.swarmap	2017 10 14 9 12 0	4b480ec2f9	Marang Tra	1.268	103.8212	Trail	4bf58dd8d48988d159941735
13	13evl_masa	female	Thailand	59e0929b9c	https://www.swarmap	2017 10 13 18 16 5	4b949891f9	Sands SkyP	1.2847	103.8609	Roof Deck	4bf58dd8d48988d133951735
14	13evl_masa	female	Thailand	59e07eb31	https://www.swarmap	2017 10 13 16 52	57b51ef749	Century Se	1.281	103.8436	Hostel	4bf58dd8d48988d1ee931735
15	13evl_masa	female	Thailand	59e063c5fe	https://www.swarmap	2017 10 13 14 57	4b8f48cff96	Food Repul	1.2948	103.859	Food Court	4bf58dd8d48988d120951735
16	13evl_masa	female	Thailand	59e04cfe87	https://www.swarmap	2017 10 13 13 19 5	4b05880ff9	Masjid Sult	1.3021	103.8592	Mosque	4bf58dd8d48988d138941735
17	13evl_masa	female	Thailand	59e04a13b5	https://www.swarmap	2017 10 13 13 7 3	4b058818f9	Haji Lane	1.3009	103.859	Road	4bf58dd8d48988d1f9931735
18	13evl_masa	female	Thailand	59e03ed481	https://www.swarmap	2017 10 13 12 19 3	4b05880ff9	Singapore (	1.2795	103.8452	Art Gallery	4bf58dd8d48988d1e2931735
19	13evl_masa	female	Thailand	59e0163810	https://www.swarmap	2017 10 13 9 26 1	4a73e804f9	Buddha Toc	1.2814	103.8443	Buddhist Te	52e81612bcbc57f1066b7a3e
20	13evl_masa	female	Thailand	59dff11910	https://www.swarmap	2017 10 13 6 47 5	4b166e4ef9	Chinatown	1.2845	103.8443	Metro Stati	4bf58dd8d48988d1fd931735
21	13evl_masa	female	Thailand	59dfa092a5	https://www.swarmap	2017 10 13 1 4 18	4b0bd124f9	Singapore (	1.3566	103.9884	Airport	4bf58dd8d48988d1ed931735

Figure 3C: A Sample Tourist Check-in File

The following table summarizes the statistics of the tourist dataset that will be used to find out the sequential activity patterns.

Number of Tourists	1057					
Age Group	20-25yrs (41%)	25-30yrs (37%)	30-35yrs (12.5%)	35-40yrs (9.5%)		
Nationality	Bangkok	Thailand	China	Japan	Kuwait	Malaysia
Type of Stay	Long	Short				
Categories Visited	Art Entertainment	Food Hospitality	Museum	Music Culture	Shopping	College University
Purpose of Visit	Tourism Recreation	Family Visit	Business Visit	Visiting Friends		

Table 3.1: A Snapshot of Tourist Statistics

3.6 ALGORITHM DEVELOPMENT

With such a convention, the expression of a sequence is unique. Then, we examine whether one can fix the order of item projection in the generation of a projected database. Intuitively, if one follows the order of the prefix of a sequence and projects only the suffix of a sequence, one can examine in an orderly manner all the possible subsequences and their associated projected database (Pei et al, 2004). Thus, we first introduce the concept of prefix and suffix

Definition 1 (Prefix). Suppose all the items within an element are listed alphabetically [80] [81]. Given a sequence a=<e1,e2,e3,…en >(where each ea. corresponds to a frequent element in S), a sequence B=<e1’,e2’…en’> is called a prefix of A if and only if 1) ei’ = ei for (i<=m-1); 2) em’ subset of em ; and 3) all the frequent items in (em-em’) are alphabetically after those in em’.

Definition 2 (Suffix). Given a sequence A=<e1,e2,..en> (where each eye corresponds to a frequent element in S). Let B=<e1e2..em-1em’> (m<=n) be the prefix of A. Sequence S=<em”Em+1…En> is called the suffix of A with regards to prefix B, denoted as C=A/B, where e”=(em-em’)^2We also denote A=B.C Note, if B is not a subsequence of B, the suffix of _ with regards to A is empty.

Definition 3 (Projected database). Let A be a sequential pattern in a sequence database S. The _-projected database, denoted as SJ_, is the collection of suffixes of sequences in S with regards to prefix _. To collect counts in projected databases, we have the following definition:

Definition 4 (Support count in the projected database) [81]. Let A be a sequential pattern in sequence database S, and B be a sequence with prefix A. The support count of B in A-projected database SJ, denoted as support SJ (B), is the number of sequences in SJ such that B subset of A.C

3.6.1 ***<u>The Pattern-Growth Method: Prefix span Algorithm</u>***

The Pattern-Growth Method of mining sequential patterns [81-83] involves the use of the Prefix span algorithm which can be executed in the following steps;

1. Finding the length of the sequential patterns: The given database is scanned for all the frequent items of a given length. The sequences patterns ought to be of the same length.
2. Partitioning of the sequential patterns into subsets: The partitioning is done based on the attached prefix.
3. Finding subsets of the sequence patterns: From of the subsets of the sequential patterns, the projected database is constructed recursively.

The parameters involve include; S is the sequence database, α is a sequential pattern, l is the length of α, SP is the alpha projected database. The major cost of the prefix-scan is the generation of the projected database [81] [82] [83]. The process of the pattern-growth method involves a series of steps which requires input, method, and then output. The output, in this case, is the complete set of sequential patterns. Minimum support thresholds and confidence levels are required for the implementation of the algorithm.

Method

- Scan the sequence database S to yield al the frequent items b where b can be assembled to form sequential patterns or <b> may be appended together to α to form a set of sequence patterns.
- Each of the frequent items forms the projected database b is appended to form α which is the output.
- In each of the given sequential pattern, α the sequential pattern is constructed through the call of the prefix-scan (α, l+1, SP)

The subsets can be formed by the resulting projected databases as shown in the table below.

Projected Database and Sequential Patterns

Prefix	Projected (postfix) database	Sequential Patterns
a	((abc) (ac) d (cf)), ((_d) c(bc) (ae)), ((_b) (df) cb), ((_f) cbc)	(a), (aa), (ab), (a(bc)), (a(bc)a), (aba), (abc), ((ab)), ((ab)c), ((ab)d), ((ab)f), ((ab)dc), (ac), (aca), (acb), (acc), (ad), (adc), (af)
b	((_c)d(cf)), ((_c) (ae), ((df)cb), (c)	(b), (ba), (bc), ((bc)), ((bc)a), (bd), (bdc), (bf)
c	((ac)d(cf)), ((bc)(ae)), (b), (bc)	(c), (ca), (cb), (cc)
d	(cf) (c(bc)(ae)), ((_f)cb)	(d), (db), (dc), (dcb)
e	(_f)(ab)(df)cb), ((af) cbc)	(e), (ea), (eab), (eac), (eacb), (eb), (ebc), (ec), (ecb), (ef), (efb), (efc), (efcb)
f	((ab) (df) cb), (cbc)	(f), (fb), (fbc), (fc), (fcb)

Table 3.2: Projected Database and Sequential Patterns

In a given sequence dataset, in this case, the Singapore tourist database, the Pattern-Growth method using the Prefix-Span algorithm helps in identifying all the frequent sequence patterns in the dataset. There are two parameters which are executed at the start of the algorithm, namely; minimum support and maximum prefix [81] [82] [83]. The latter helps by providing the length of the sequence which is quite crucial while analyzing large databases. On the other end, the minimum support parameter is obtained by dividing the pattern with the number of sequences in the dataset. Using the projected database, the Prefixspan algorithm first finds out the lengths of the sequential patterns then divide the space and mine the projected database to yield the patterns.

A Sequence Database

ID	Sequences
S1	(1), (2), (1 2) , (3), (1 3), (4 5), (6)
S2	(3 4), (3), (2 3), (1 4)
S3	(4 5), (2), (2 3 4), (3), (1)
S4	(4), (5), (1 6), (3), (2), (7), (1)

Table 3.3: An Example of the Sequence Database

Table 3.3 represents an example of the sequence dataset which contains four sequences of patterns having items sets. For instance, the first sequence contains seven item sets of frequent patterns [83]. Similarly, the analysis of the venues category and check-in time shall yield the subsequent patterns from the selected sequence of the Singapore database. The application of the Prefix Span entails the scanning of the sequence database to obtain the frequent items [83] [84]. Thereafter, the frequent items are appended together to form a sequential pattern which is then used in the construction of the projected database. The prefix span algorithm shall be implemented in java pseudo code although it can also be run using python and R programming languages.

3.6.2 ***Prefix Span Pseudocode***

PrefixSpan is a frequent sequence-mining algorithm. The summarized pseudocode for the PrefixSpan algorithm is as follows:

PREFIXSPAN ($\mathbf{D}_r$, $\mathbf{r}$, *minsup*, Ƒ): $\mathbf{D}_r \leftarrow \mathbf{D}$, $\mathbf{r} \leftarrow \emptyset$, .Ƒ← Ø

foreach $s \in \Sigma$ *such that* $sup(s, \mathbf{D}_r) > minsup$ **do**

$\mathbf{r}_s = \mathbf{r} + s$ // extend **r** by symbol s

Ƒ←Ƒ∪ $\{(\mathbf{r}_s, sup(s, \mathbf{D}_r))\}$

$D_s \leftarrow \emptyset$ // create projected data for symbol s

Foreach $\mathbf{s}_i \in \mathbf{D}_r$ **do**

$\mathbf{s'}_i \leftarrow$ projection of $\mathbf{s}_i$ w.r.t symbol s

Remove any infrequent symbols from $\mathbf{s'}_i$

Add $\mathbf{s'}_i$ to $\mathbf{D}_s$

if $\mathbf{D}_s \neq \emptyset$ **then** PREFIXSPAN ($\mathbf{D}_s$, $\mathbf{r}_s$, *minsup*, Ƒ)

Input: A sequence database S, and the minimum support threshold minsup.

Output: The complete set of sequential patterns.

Method: Call Prefix Span (‹›, 0, S)

Subroutine Prefix Span (α, l, S|α)

Parameters: α: a sequential pattern; l the length of α; S|α: α-projected database, α ≠ ‹›; otherwise, the sequence database S.

Method: 1. Scan S|α once, find the set of frequent items b such that (a) b can be assembled to the last element of α to form a sequential pattern; or (b) ‹b› can be appended to α to form a sequential pattern. 2. For each frequent item b, append it to α to form a sequential pattern α', and output α'; 3. For each α', construct α'-projected database S|α', and call Prefix Span (α', l+1, S|α') [85].

The prefixspan algorithm is suited for small datasets. The execution time is poised to increase with the increase in the size of the input sequence datasets. In the future, due to the continued increase in the size of dataset especially in the tourism industry, the prefixspan approach may no longer be appropriate to mine sequence patterns. However, with the continuous advancement in technology, more sophisticated and efficient methods will be proposed to solve the problem. The minimum support values help while running the algorithm on large datasets [81] [82] [85]. There is one of the new parameters referred to as the maximum prefix lengths that is imposed on the algorithm to effectively handle large databases. Based on projection, the pattern growth method adopts the prefixspan algorithm in mining sequence patterns from a given dataset. Since the projection is made on the prefix that is most frequent, less processing time is required thus improving the efficiency of the algorithm. In order to successfully generate the sequence patterns, there are two parameters that ought to be included at the beginning of the process namely; minimum support and the maximum prefix length. In small datasets, the latter is not necessary since it is only useful while dealing with large datasets to fasten the execution time.

In large datasets, the minimum support and maximum prefix length values are set before executing the algorithm to generate the sequence patterns. In essence, the algorithm works differently with different sizes of the data sets depending on two factors namely; complexity and memory utilization. The maximum prefix length value is vital in the execution of large sequential datasets. Both memory utilization and time complexity are computed using the values of minimum support and maximum profit length. The minim support values are obtained by dividing the pattern by the number of total possible sequences in the database. Basically, the sequence patterns having values greater than the set minimum support are often extracted from the databases. Both minimum support and the maximum prefix lengths often affect memory utilization as well as the time complexity [81] [83]. Different datasets exhibit different sequence patterns depending on the different memory space and time complexity.

The concept of frequency is largely applied in the pattern-growth method. Due to the increase in the sizes of databases, the frequency alone cannot be relied upon in the generation of the sequence patterns from the given database. In regard, there is need to incorporate other constraints such as maximum prefix length, compactness threshold, and monetary constraints. The additional proposed constraints help in minimizing the processing time and reduce memory utilization hence improving the efficiency of the prefixspan approach. For instance, in a dynamic environment such as the supermarkets, the purchasing behaviors of customers tend to change properly and in an unpredictable manner. Consequently, it is important to add compactness and monetary constraints to ease the problem of mining the sequence patterns. Also, in the tourism industry, the unpredictable nature of tourists' activity has made it a challenge to mine the sequence patterns from databases. Therefore, there is the need to incorporate compactness thresholds with the minimum values threshold to help in the generation of the sequence patterns. In a business environment, the monetary constraint is paramount. The benefit of compactness constraints is that it allows the patterns to be mined based on a reasonable time span. Time is an important factor in the minimum of the sequence activity patterns from a given dataset. Both compactness and monetary constraints help in the generation of significant sequence patterns using the prefix-span algorithm [85] [86]. In addition to minimum support, compact threshold and monetary thresholds are added in the input stage of the execution of the prefixspan algorithm. The Compact-Frequency-Monetary (CFM) prefix span is an improvement of the prefixspan algorithm.

The compactness and monetary concepts are derived from the aggregate and duration constraints which helps in the development of more valuable sequence patterns. Ideally, the CFM prefix span is an improvement of the prefix span algorithms. Both algorithms use the divide and conquer strategy in the generation of the sequence patterns. In the dataset on the Singapore tourists, the addition of the aggregate and duration constraints helps in adding substance to the output of the prefix span which are complete sets of sequence activity patterns [86]. Since tourists is also a business venture, the monetary concept is paramount in the execution of the prefix span algorithms.

The pattern-growth method uses both transactional and sequential databases. There are to approaches namely; prefix span and free span. The former approach has been proposed and utilized by the majority of the researchers in the process of mining sequence patterns from databases or datasets. One of the important features of the Prefixspan algorithm is the ability to maintain the original characteristics of the data particularly when mining sequence patterns from large databases. In the study of a medical database to extract the disease trajectory patterns, a frequent subsequence analysis which is part of the prefix scan was used in the process. In the sequence pattern mining field, the prefixspan approach through the pattern-growth method has been utilized successfully in the generation of sequential orders of frequent items from big datasets. In Singapore, there is an increasing number of tourists visiting various destinations in the country hence the prefixspan approach is suitable for mining the sequence tourist’s activity patterns. By doing so, it helps the management of the Singapore Tourism Board in improving the service delivery. Also, the knowledge of the tourists’ activity patterns across the year’s aids in proper preparation prior to their visitation to the country. Although the pattern growth methodology had been widely applied in another field such as healthcare and entrepreneurship, it has been hardly implemented in the tourist's industry especially in southern Asia.

With the user-specified minimum support values, the sequence pattern mining method obtains all the frequently occurring subsequences. Ideally, all the sequence patterns having values greater than the minimum values set, are extracted into the projected database. One of the advantages of the pattern growth method is that it has the ability to extract relevant subsequences patterns in either transactional or sequential database. The Singapore dataset is classified as a sequential database hence the implementation of the prefix-span approach will yield significant sequence activity patterns for decision making in the tourism industry. Let *S* be a sequence database, which is composed <*s_id, s*> where *s_id* is known as the sequence identifier and *s* is the sequence patterns of frequent items [87]. The table below gives an illustration of a sequential database:

Sequence Id	Sequence
1	<a(abc)(ac)d(cf)>
2	<(ad)c(bc)(ae)>
3	<(ef)(ab)(df)cb>
4	<eg(af)cbc>

Table 3.4: An Illustration of a Sequential Database

For instance, in the sequence *id* 1, the sequence contains five elements representing diseases which were diagnosed at five different time states [87]. Similarly, also other sequence Ids exhibit a different sequence of varied lengths consisting of disease trajectory patterns. Without the loss of generality, the items in each of the elements can be ordered in an alphabetic manner this the formulation of the sequence patterns. The first step of the prefixspan approach is the scanning of the sequence database to yield all the frequent items organized in sequence patterns of different lengths. The prefix of the elements in the sequence for example in sequence id 1 <an (ABC) (ac) d (cf)>, <a> is taken as the prefix which is used in the projection of the database. With the projected dataset, the process of scanning for frequent items is repeated recursively till all the sequence patterns with the prefix <a> are extracted.

Through the divide and conquer strategy, prefix span mines complete set of sequence patterns while reducing the need for candidate generation. The projection and mining are based on the frequent prefix of the elements in the sequence. The search space partitioning is one of the crucial features of the pattern growth method [77]. The partitioning effect helps in the management of the memory hence better executing time of the algorithm. A data structure is first drafted before the implementation of the prefix span algorithm. The formulation of the data structure is done at the initial stages of the sequence pattern mining process.

From the web access sequence database (WASD), the implementation of the prefix-span algorithm yielded the most interesting sequence patterns [80] [81]. Using the minimum support threshold of 3, the scanning \of the WASD was done to obtain item set sequential patterns. The frequency of the items is given as (a:4, b:4, c:3, d:2, e:2, f:2). Given the rule that frequent items having greater equal value of the minim support are only extracted, the resulting frequency item set includes; {a: 4, b: 4, c: 3}. The completed set of the sequence pattern may be categorized into three parts depending on the prefix of the elements, i.e. a, b, and c. the item sets with similar prefix are appended together to form the resulting sequence patterns. The process is thereafter repeated recursively including dividing the projected dataset into subsets of sequential patterns. Using the WASD dataset as the input and the prefixspan algorithm, the execution is done in the Java programming language in the Microsoft Windows server leads to the resulting output of a complete set of frequent sequential patterns.

Over time, there have been proposed modifications to the prefix spa algorithms in a view of improving its performance. The inclusion of the compactness and monetary constraints led to the formation of CFM-prefixspan algorithm which is an improvement of the prefix span. Also, the addition of the maximum prefix length parameter at the initial stages before the execution phase has improved the efficiency of the algorithm resulting to the generation of more precise and reliable sequence patterns. The CIC-prefix span generated sequence patterns by combining pseudo-projection and prefix span. The new version of the prefix span uses the maximum forward path technique where it excludes the non-human user sessions to find frequent transactions in a given database [34]. The use of the CIC-prefix span has proved to yield accurate sequence patterns with low execution time and high efficiency during the process. The algorithms generate duplicate projections and reduce memory space in a bid to find the most appropriate and frequent patterns. However, it is not possible to find patterns or paths from frequent substructures within a pattern. The continued modifications on the prefix span through the pattern growth method has resulted in high efficiency and faster execution time in the mining of sequence patterns, especially for large datasets.

CHAPTER 4: DATA ANALYSIS AND FINDINGS

4.1 GENERAL DEMOGRAPHICS OF DATA

Tourism receipts and visitor arrivals hit new record highs last year, topping the stellar performance from 2016. Preliminary estimates from the Singapore Tourism Board (STB) showed tourism receipts climbed 3.9 per cent to $26.8 billion, spurred by growth in visitor arrivals across all top 10 markets - China, Indonesia, India, Malaysia, Australia, Japan, the Philippines, South Korea, United States and Vietnam. The increase in tourism receipts was also boosted by more visitor arrivals from high-spending markets such as China, South Korea, the US and the United Kingdom. China emerged top in tourism receipts for the third straight year, although Indonesia logged a 7 per cent decline, India dipped 1 per cent and Japan fell 9 per cent, mostly because of fewer business travel and meetings, incentive travel, conventions and exhibitions visitor arrivals.

Visitor arrivals rose 6.2 per cent to 17.4 million, with 13 of the top 15 markets showing growth. Singapore's second straight year of record tourism performance was the result of the combined efforts of the STB and its industry partners, against the context of better-than-expected global economic recovery, continued growth in Asia-Pacific travel and increased flight and cruise connectivity to Singapore. Visitor arrivals and tourism receipts for 2017 hit record highs for the second time in two years. Arrivals to Singapore increased by 6.2 per cent to 17.4 million, with seven of the top 15 markets - China, India, Vietnam, Philippines, United States, United Kingdom and Germany - also hitting record visitor arrivals.

According to the country's most recent census in 2015, nearly 23% of Singaporean residents (i.e. citizens and permanent residents) were foreign born (which means about 10% of Singapore citizens were foreign-born naturalized citizens) and if non-residents were counted, nearly 43% of the total population were foreign born [25]. As of mid-2018, the estimated population of Singapore was 5,638,700 people, 3,471,900 (61.6%) of whom were citizens, while the remaining 2,166,800 (38.4%) were permanent residents (522,300) or foreign students/foreign workers/dependents (1,644,500).

Singapore Population Census Chart

Estimated Population	Citizens of Singapore	Permanent residents of Singapore	Foreign Students in Singapore
5,638,700	3,471,900	2,166,800	522,300
	Increasing citizen in 2018	**Increasing permanent resident in 2018**	**Increasing students in 2018**
	61.6%	38.4%	8%

Table 4.1: Singapore Population Census Chart

According to the Singapore Tourism Board of Statistics 74.1% of residents were from China, 13.4% from Malaysia ,9.2% from India, and 3.3% of other countries.

The following table define the counts of the visitors and locals in Singapore.

Diversification of Visitors in Singapore

Type of Visitors in Singapore	Number of Visitors
Long-term visitors	2,282
Short-term visitors	835
Ambiguous visitors	782
Locals	8,597 1

Table 4.2: Diversification of Visitors in Singapore

The Exploratory Data Analysis further revealed the demographics of tourist population in Singapore. Ideally the data-set comprised of 3,830 female tourists, 3,577 male tourists while 207 did not wish to declare their gender. The data-set comprised of more than 10,000 Foursquare Check-ins, which were generated by 1,057 number of tourists visiting Singapore during a period of 4-6 months. On an average, each tourist recorded 8-10 live check-ins on Foursquare which were included in this study to examine and extract the most interesting and valuable sequential activity patterns.

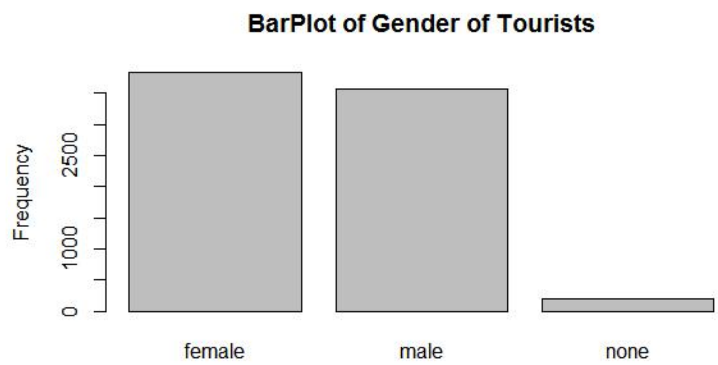


Figure 4A: Bar Plot of Gender of the Tourists

As shown in the figure above, there were more female tourists than male tourists visiting different venues in the country. From the analysis, majority of the female tourists came from Malaysia, NaN, Thailand, Peru, and BBK. On the other end, majority of the male tourists came from Japan, NaN, Kuwait, Selangor, Thailand, and Indiana.

4.2 TOURIST SEQUENTIAL ACTIVITY ANALYSIS

Next we study the differences between long and short trips by a number of measures including check-in distributions over venue categories/subcategories, check-in distributions over time and check-in intensity.

TYPES OF VISITORS IN SINGAPORE		
Check-in count	short-term	long-term
1	388	114
2	141	57
3	85	54
4	44	43
5	33	42
> 5	144	1972

Table 4.3: Tourist Classification in Singapore

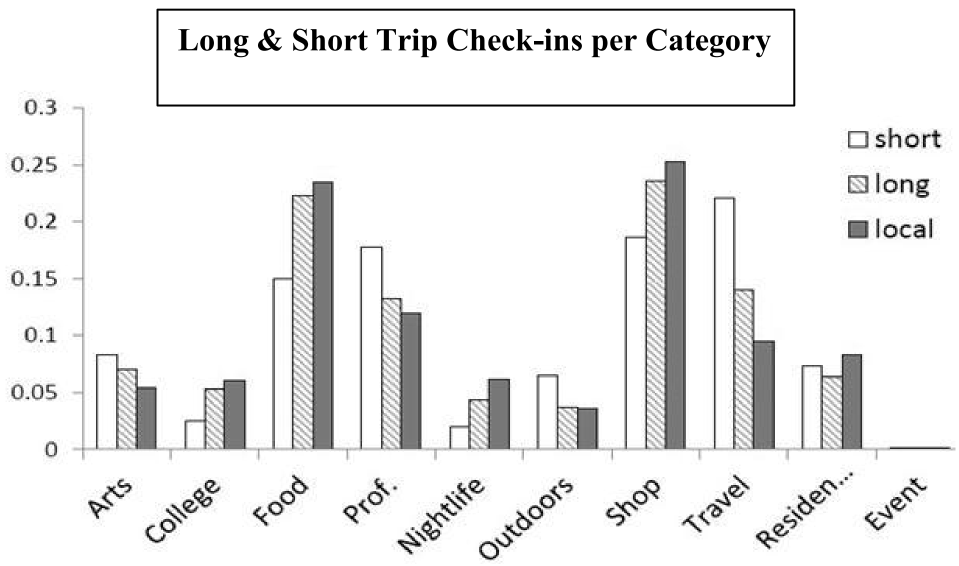


Figure 4B: Long & Short Trip Check-ins of Different Categories

Trips to a city are made for various purposes, affecting both the trip duration and the categories of venues visited. In this section, we examine the distribution over venue categories to understand trip purposes. Our task is facilitated by the fact that Foursquare already categorizes venues into 10 top-level categories which indicates their functions. These are: Arts & Entertainment (Arts), College & University (College), Food, Professional & Other Places (Prof.), Nightlife Spot (Nightlife), Outdoors & Recreation (Outdoors), Shop & Service (Shop), Travel & Transport (Travel), Residence and Events in the country of Singapore.

The figure 4a shows the check-in distribution over venue categories for Singapore respectively. For example, in Singapore, the probability of having a check-in from a long trip at a shopping venue (the 'Shop' category) is 0.24. The same probability is lower at 0.19 when the check-ins come from a short trip. For comparison, we also include the probability if the check-in is from a local. Hence, local check-ins have the highest probability for 'Nightlife', followed by long trips and with short trips last in place. As a side note, since visitors' ethnic composition affect their travel patterns, one can conduct interesting analysis of a city's ethnic composition or to quantify how cosmopolitan or mixed a population is.

Earlier, we have seen that for check-in distributions over main venue categories, long trips are more similar to local check-ins, than short trips. As main categories are coarse and each can comprise many subcategories, we further analyze check-in distributions over subcategories as well. However, it is not informative to display the complete distributions here due to the large number of subcategories (>700) in Singapore. Instead, we examine most probable travel-related subcategories where differences are more discernible. The complete set of Foursquare Venue Categories and Sub-categories can be found at 'https://developer.**foursquare**.com/docs/resources/**categories**'.

The Exploratory Data Analysis of the tourists dataset revealed some interesting demographics of the Singapore tourists which will be discussed herewith. The following is the item frequency plot of the Apriori generated rules. As expected for all the users, Airport, Shopping Mall, Hotel, Theme Park and Garden are the top activities in all the travel patterns.

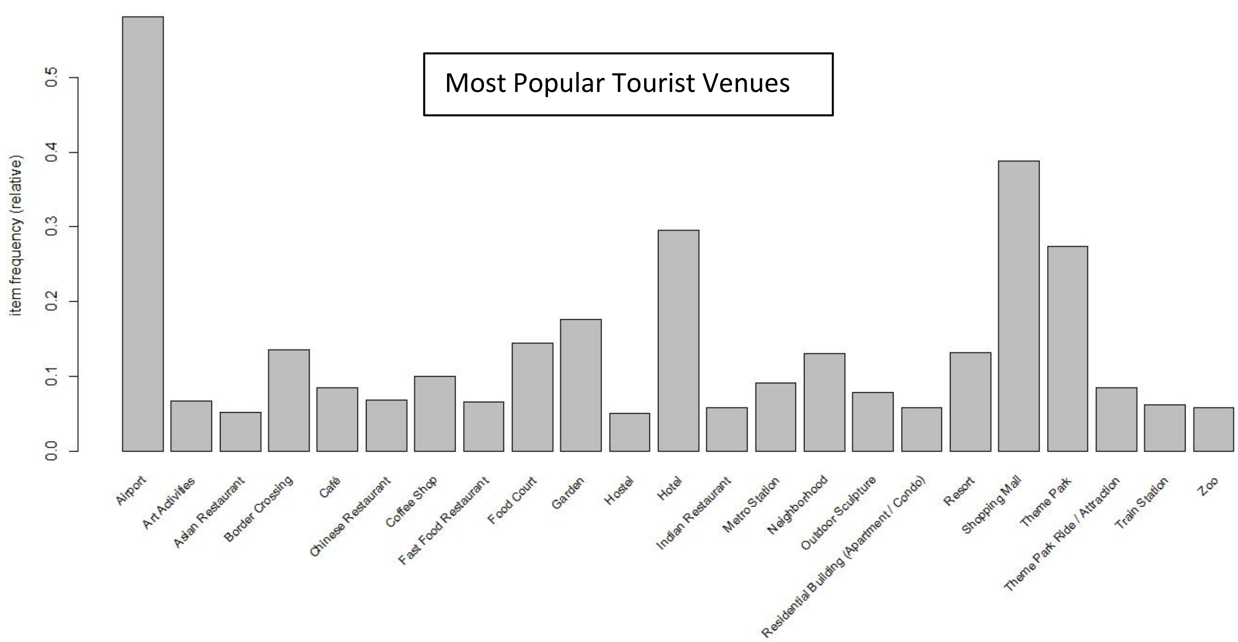


Figure 4C: Bar Plot of Most Popular Tourist Venues

Next, we filter the combinations with maximum lift. The following table shows the antecedents and the consequents for the sequences, since most of the patterns contains Airport within the patterns, so lift has been used as the measure of extracting results, else Airport will show up in the rules. This is the situation, where lift is considered as the best attribute for analysis.

	lhs		rhs	support	confidence	lift	count
[1]	{Airport,Coffee Shop,Food Court,Shopping Mall}	=>	{Asian Restaurant}	0.01041667	0.5238095	10.057143	11
[2]	{Airport,Coffee Shop,Shopping Mall}	=>	{Clothing Store}	0.01041667	0.2244898	9.482449	11
[3]	{Shopping Mall,Theme Park,Train Station}	=>	{Metro Station}	0.01041667	0.8461538	9.307692	11
[4]	{Airport,Coffee Shop,Food Court}	=>	{Asian Restaurant}	0.01136364	0.4615385	8.861538	12
[5]	{Theme Park,Train Station}	=>	{Metro Station}	0.01136364	0.8000000	8.800000	12
[6]	{Coffee Shop,Shopping Mall}	=>	{Clothing Store}	0.01231061	0.1969697	8.320000	13
[7]	{Airport,Food Court,Train Station}	=>	{Metro Station}	0.01136364	0.7500000	8.250000	12
[8]	{Coffee Shop,Food Court,Shopping Mall}	=>	{Asian Restaurant}	0.01041667	0.4230769	8.123077	11
[9]	{Airport,Food Court,Shopping Mall,Train Station}	=>	{Metro Station}	0.01041667	0.7333333	8.066667	11
[10]	{Hotel,Train Station}	=>	{Outdoor Sculpture}	0.01041667	0.6111111	7.869919	11
[11]	{Airport,Food Court,Metro Station}	=>	{Bridge}	0.01041667	0.3548387	7.494194	11
[12]	{Airport,Coffee Shop}	=>	{Clothing Store}	0.01136364	0.1714286	7.241143	12
[13]	{Coffee Shop,Food Court}	=>	{Asian Restaurant}	0.01136364	0.3750000	7.200000	12
[14]	{Bridge,Food Court,Shopping Mall}	=>	{Metro Station}	0.01041667	0.6470588	7.117647	11
[15]	{Airport,Bridge,Food Court}	=>	{Metro Station}	0.01041667	0.6470588	7.117647	11
[16]	{Airport,Shopping Mall,Train Station}	=>	{Metro Station}	0.01704545	0.6428571	7.071429	18
[17]	{Food Court,Metro Station,Shopping Mall}	=>	{Bridge}	0.01041667	0.3333333	7.040000	11
[18]	{Asian Restaurant,Coffee Shop}	=>	{Food Court}	0.01136364	1.0000000	6.947368	12
[19]	{Asian Restaurant,Coffee Shop,Shopping Mall}	=>	{Food Court}	0.01041667	1.0000000	6.947368	11
[20]	{Airport,Asian Restaurant,Coffee Shop}	=>	{Food Court}	0.01136364	1.0000000	6.947368	12

Figure 4D: Results Generated through SPM using the Apache Software

The above table shows top 20 rules sorted on the basis of the lift. Let's examine the first rule, LHS (Left Hand Side) denotes the antecedants and RHS (Right Hand Side) denotes the consequents, which is the next location the user might be interested in. As we see the first results, it indicates that if the user has last check-ins at Airport, Coffee Shop, Food Court, Mall then the next destination would be Asian Restaurant and so on for the other rules. Now, let's see the graph of rules determined, please note that the size of the circle (a location) depicts the support of the particular destination and on the other hand the colour darkness depicts the lift.

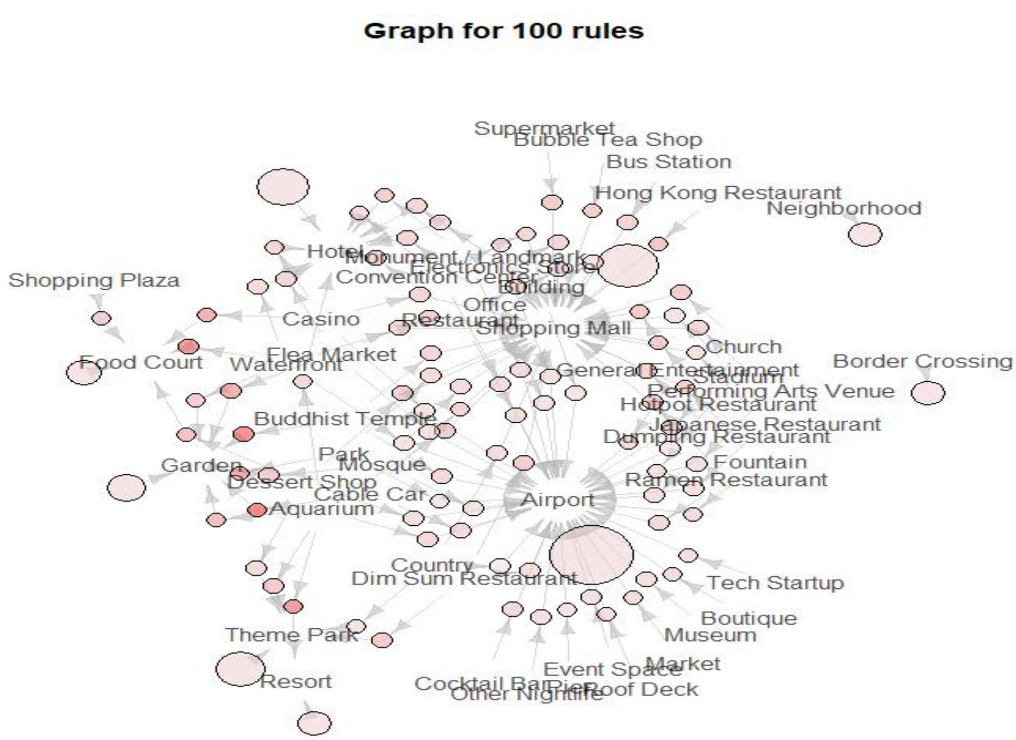


Figure 4E: First 100 SPM Patterns in Graphical Format

In a similar way, the graph below shows the grouped matrix for the most prevalent and interesting patterns. On the columns of the matrix, there are rules and the number of times the rule is found in all rules which have been extracted, in the rows, there is the target destination, which could be the next location, the traveller is to go. The circles size and colour, bigger and darker, increase the chances of his particular destination.

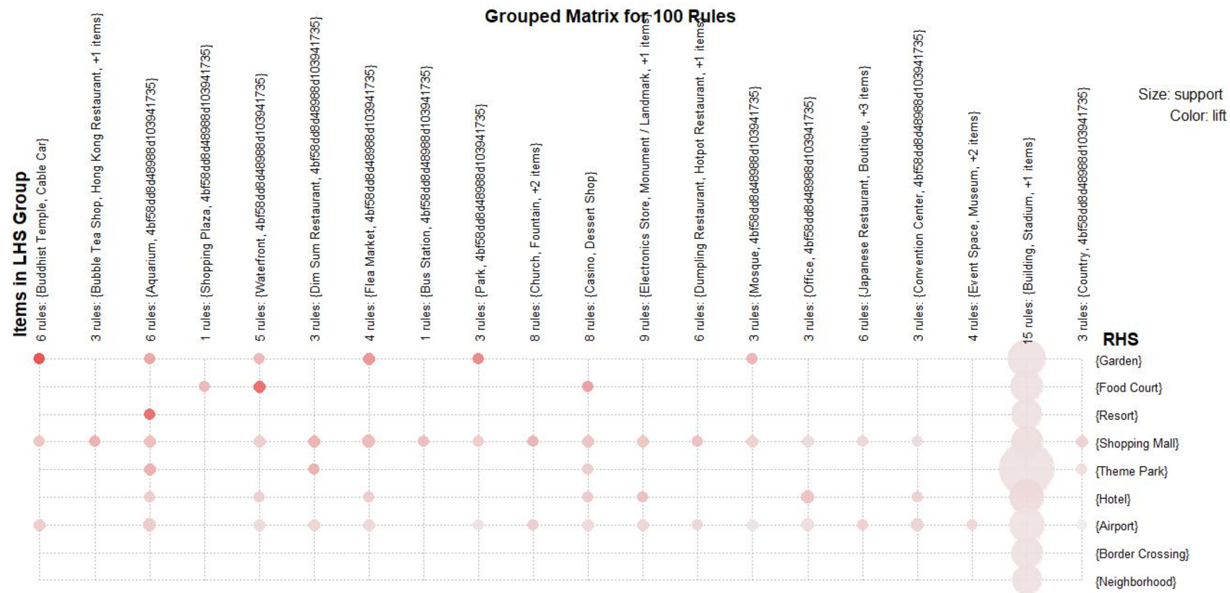


Figure 4F: Group Matrix for the 100 Sequential Patterns

For one more case, we chose to analyse the patterns for all users, the below screenshot of the table is showing the most important 20 rules for the tourists in Singapore. If we see the patterns here, we can conclude that there are more chances that Metro Station, Food Court, Garden are the next stops or check in destinations for tourists if they visit the previous LHS locations.

	lhs		rhs	support	confidence	lift	count
[1]	{Airport,Shopping Mall,Train Station}	=>	{Metro Station}	0.02348337	0.8000000	8.176000	12
[2]	{Food Court,Train Station}	=>	{Metro Station}	0.02152642	0.7333333	7.494667	11
[3]	{Airport,Train Station}	=>	{Metro Station}	0.02739726	0.7000000	7.154000	14
[4]	{Shopping Mall,Train Station}	=>	{Metro Station}	0.02544031	0.6190476	6.326667	13
[5]	{Airport,Asian Restaurant,Shopping Mall}	=>	{Food Court}	0.02152642	0.8461538	5.923077	11
[6]	{Asian Restaurant,Shopping Mall}	=>	{Food Court}	0.02152642	0.7857143	5.500000	11
[7]	{Chinese Restaurant,Shopping Mall}	=>	{Food Court}	0.02348337	0.7500000	5.250000	12
[8]	{Airport,Shopping Mall,Train Station}	=>	{Food Court}	0.02152642	0.7333333	5.133333	11
[9]	{Airport,Café,Shopping Mall}	=>	{Food Court}	0.02152642	0.7333333	5.133333	11
[10]	{Botanical Garden}	=>	{Garden}	0.02544031	0.7647059	5.009804	13
[11]	{Bridge,Shopping Mall}	=>	{Garden}	0.02348337	0.7500000	4.913462	12
[12]	{Beach}	=>	{Food Court}	0.02152642	0.6875000	4.812500	11
[13]	{Bridge,Shopping Mall}	=>	{Food Court}	0.02152642	0.6875000	4.812500	11
[14]	{Café,Shopping Mall}	=>	{Food Court}	0.02348337	0.6315789	4.421053	12
[15]	{Metro Station,Train Station}	=>	{Food Court}	0.02152642	0.6111111	4.277778	11
[16]	{Airport,Metro Station,Shopping Mall}	=>	{Food Court}	0.02739726	0.6086957	4.260870	14
[17]	{Airport,Train Station}	=>	{Food Court}	0.02348337	0.6000000	4.200000	12
[18]	{Airport,Neighborhood,Shopping Mall}	=>	{Food Court}	0.02348337	0.6000000	4.200000	12
[19]	{Food Court,Hotel,Shopping Mall}	=>	{Garden}	0.02152642	0.6111111	4.003561	11
[20]	{Airport,Zoo}	=>	{Garden}	0.02348337	0.6000000	3.930769	12

Figure 4G: Sequential Patterns with Confidence, Lift and Support Values

Now, let’s look at the top rules again in graphical manner.

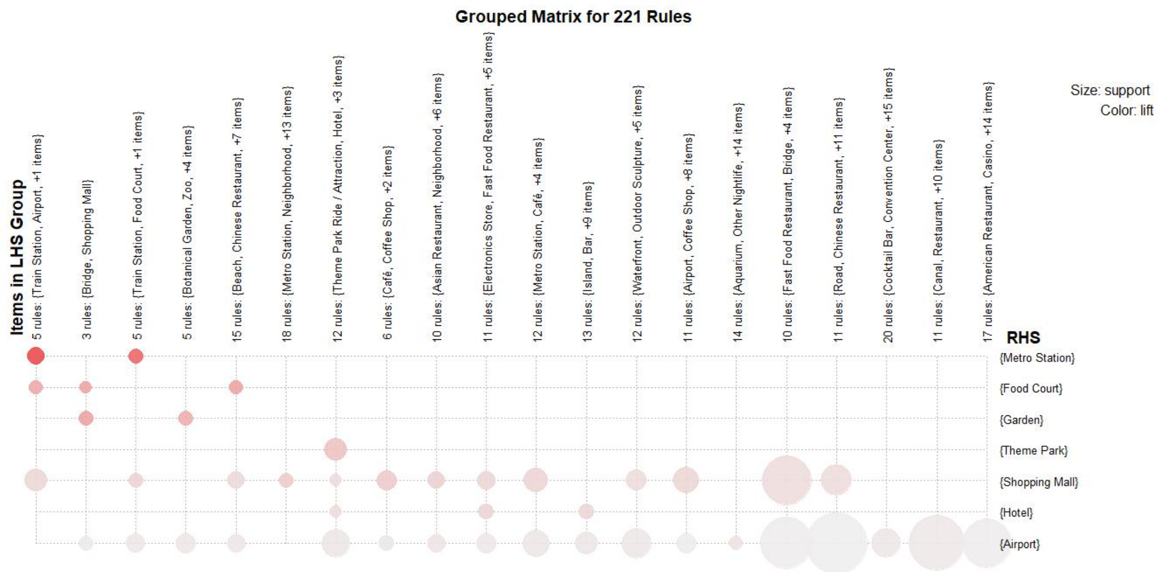


Figure 4H: Group Matrix for the 200 Sequential Patterns

The following conclusions can be made from the above graph, if we see the consequents, the size of circles for Airport destinations, it is bigger in size of high support since Airport is the part of every tourist’s journey, hence we will be ignoring this. Other locations are Metro, Food Court, Theme park, Mall and Hotel, which are important destinations here for the rules which are shown in the antecedents such as there are 5 rules, where Train Station, Airport exists and there are many chances that next destination could be any of Metro, Food Court, Theme park, Mall and Hotel etc. There are 3 rules, where Bridge and Shopping Mall exists. We can easily deduce that for all tourists, the most favourable patterns and the consequents comprised of diverse locations as they are more interested in dining in restaurants of different cuisines (Chinese, Japanese, Dim Sum, Dumplings) as well as the art attractions and places like church, and mosque.

Now we look at the tourists preferences for food in Singapore as evident from our EDA.

Top Food Categories in Singapore
• Bak Kut Teh (lit. Meat Bone Tea/ Pork Ribs Soup)
• Wanton MEE
• Fried Carrot Cake
• Dim Sum
• Kaya Toast & Soft-Boiled Eggs. The one and only traditional Singaporean breakfast – Kaya Toast with Soft-Boiled Eggs
• Crabs (Chili Or Pepper)
• Curry Fish Head
• Laksa.

Table 4.4: Food Preferences of Singapore Tourists

The table above shows the categories of the food which the tourist prefer while visiting Singapore. Singapore has been described as a playground for the rich, and it's true that the small city-state does have a certain sheen of wealth. But Singapore offers more than just high-end shopping malls, luxury hotels, and fine dining. There is also a vibrant history and diverse ethnic quarters to discover, along with many family friendly attractions and lovely public spaces that make visiting this slightly futuristic city worthwhile. Singapore has an excellent public transportation system that makes getting around convenient and easy. Once the tourists get a sense of the metro map, they have no problem zipping from one part of town to the next.

The ways to deal with customary travel journal development depend on tourist cooperation and manual account; subsequently, they are tedious as well as restricted in the scale and the number of tests. Online informal community stages have been utilized as elective information hotspots for catching the developments and travel examples of visitors at an expansive scale. They neglect to give definite logical data on visitor activities for further examination. In this research, we present the data analysis to construct tourists activity trajectory based on the venue check-in data available in mobile social media with rich information on locations, time, and activities. Our contextual investigation centers around the inbound travel industry in Singapore utilizing an informational index made out of 17,355 registration produced by approximately 1000 visitors. We exhibit how the proposed travel journal can give valuable viable ramifications to applications in the area the executives, transportation executives, affect the board, and tourist encounter advancement among others.

The greatest generator of the travel industry are the free market activity provinces. The abundance of an extensive number of various characteristics, authentic and social travel industry potential is an essential course of a progression of totally extraordinary visitor activities. Officially present types of the travel industry offer, for the most part, nearby character, outing, entertainment, business and travel industry, we should include the wellbeing, religious (journey), sports and rustic travel industry, as new types of particular travel industry destination.

Primarily, the Pattern-Growth Method using Prefixspan Algorithm and utilizing the sequential pattern mining method, the dataset was analyzed based on the respective tourists to determine the sequential patterns in terms of venues visited with respect to the check-in times. Using the Pattern-Growth method and the Prefixspan algorithm many sequential patterns were obtained. The Apache software which runs on java was used to implement the algorithm following the systematic steps to yield the tourists sequential activity patterns. It is evident that there are some interesting patterns which can easily be drawn. They will be discussed in the following pages.

4.3 SNAPSHOT OF TOURIST SEQUENTIAL PATTERNS

4.3.1 *SEQUENCE 1: Most Popular Tourist Morning Activities*

The following table captures the most interesting activity patterns of tourists during the morning hours (7:00 - 14:00). The frequency and confidence interval, which is the specified probability that the value lies within each activity, is mentioned in each case below.

Activity 1> 2> 3	Frequency	Min-Support	Confidence
Travelling > Religious > Dining	120	75%	0.45
Shopping > Recreation > Entertainment	120	75%	0.78
Hiking > Outdoor > Refreshments	85	59%	0.90
Entertainment > Religious > Shopping	89	45%	0.61
Shopping > Hiking > Dining	65	26%	0.53
Nature > Refreshments > Dining	65	`55%	0.42
Archives > Shopping > Nature walk	23	34%	0.74
Dining > Travelling > Nature walk	113	67%	0.56
Hiking > Nature > Shopping	113	58%	0.55
Dining > Nature > Nature	87	24%	0.33
Sport > Dining > Nature	65	17%	0.24
Nature > Archives > Nature	50	41%	0.19
Religious > Nature > Shopping	25	67%	0.41
Entertainment > Nature > Refreshments	114	71%	0.24
Dining > Hiking > Shopping	111	70%	0.30

Table 4.5: Tourist Morning Activity Patterns

Although primarily tourists had different times that they were involved in various activities, there were notable activities such as visiting the park, scenic lookout, and outdoor sculpture. Also visiting the garden and border crossing were other activities that were commonly done in the morning hours. Hiking, nature, and religion dominated the morning hours of the tourist schedules based on the patterns. As compared to other time periods in the afternoon and evening, people prefer to visit different restaurants to eat in the morning hours. Based on the figure above several tourists share most of the sequential activity patterns.

4.3.2 ***SEQUENCE 2: Most Popular Tourist Afternoon Activities***

Activity Sequence 1 > 2 > 3	Frequency	Min-Support	Confidence
Dining > Walking > Shopping	115	56%	0.67
Field > Exhibition > Shopping	115	89%	0.35
Field > Dining > Shopping	185	40%	0.52
Shopping > Gaming > Walking	170	41%	0.36
Dining > Walking > Shopping	165	39%	0.28
Nature > Nature > Walking	144	67%	0.71
Travelling > Dining > Walking	123	74%	0.12
Shopping > Refreshments > Refreshments	113	87%	0.14
Nature > Dining > Walking	113	23%	0.26
Shopping > Field Walk > Nature	187	76%	0.83
Dining > Field > Hiking	165	45%	0.74
Archives Visit > Travelling > Resting	140	21%	0.67
Scenery > Dining > Entertainment	125	76%	0.48
Dining > Archives > Walking	114	54%	0.27
Archives > Dining > Field	111	90%	0.78

Table 4.6: Tourist Afternoon Activity Patterns

The above table shows the frequency and confidence for afternoon activities' patterns in Singapore. The afternoon hours range from midday to late night (14:00 - 22:00). An interesting fact to note here is that most of the afternoon activities are outdoors, which attributes to the year round tropical weather of Singapore. With long 13-14 hours day during the peak tourist season, it gives an ample time of saturation at tourist venues and hence tourists can make the most of their times and raise their satisfaction at individual levels. Increased tourist satisfaction amounts to repeat visitation and higher tourist demands in the growing season.

4.3.3 ***<u>SEQUENCE 3: Popular Tourism Months To Visit Singapore</u>***

Month of the Year	**Frequency**	**Min-Support**	**Confidence**
January	**233**	**67%**	**0.22**
February	226	28%	0.44
March	385	36%	0.43
April	379	45%	0.65
May	365	67%	0.45
June	400	29%	0.95
July	313	31%	0.34
August	565	87%	0.21
September	820	51%	0.12
October	940	55%	0.87
November	710	79%	0.45
December	710	39%	0.90

Table 4.7: Popular Tourism Months

From the sequences above, it is evident that most tourists prefer to visit various destination and get involved in different activities in the last quarter of the year. Ideally, the period between August and December recorded more tourists' activities from the sequential activity patterns drawn above. Subsequently, October is the common month that the tourists in Singapore preferred to visit.

4.3.4 ***SEQUENCE 4: Most Frequently Visited Tourist Destinations***

Activity	Frequency	Min-Support	Confidence
Shopping Malls	200	61%	0.66
Nature Attractions	300	45%	0.28
Dining	785	68%	0.73
Outdoor	870	41%	0.65
Religious	344	68%	0.67
Exhibition	523	78%	0.59
Gaming	613	89%	0.56
Entertainment	723	23%	0.67
Scenery	965	35%	0.34

Table 4.8: Popular Tourist Destinations

Both male and female tourists were largely involved in shopping, nature attractions, and dining. The tourists prefer to visit the variety of hotels and restaurants to explore different types of food owing to the rich Asian world. The activity pattern Shopping > Field > Nature walk had the highest frequency indicating that most of the tourists engaged in the activities in that pattern based on the analysis of the dataset.

Utilizing the Pattern-Growth method to identify the activities of tourists and the algorithm of Prefix-Span for the accompanying arrangement designs were acquired. The software of Apache which keeps running on java was utilized to actualize the calculation following the systematic steps to yield the tourist sequential activity patterns.

The greater part of arrivals of international tourists (53% or 632 million) were motivated by holidays, entertainment and relaxation related travel. Around the same time, business and expert travel represented 14% of all worldwide arrivals of tourists and another 27% of the tourists went for different reasons, for example, visiting the friends and relatives, for shopping as well. The motivation behind movement for the remaining 6% of entries was not determined. As indicated the patterns will remain to a great extent stable into 2030, when it is anticipated that relaxation, diversion and holidays will speak to 54%, business and expert travel 15% for different purposes 31% of every single arrival for tourists.

4.3.5 ***<u>SEQUENCE 5: Most Popular Tourist Venues In Singapore- In Sequential Order</u>***

The table below shows the sequential patterns of the venues visited by the tourists.

VENUES SEQUENCE 1 > 2 > 3	
Venue Sub-Categories	**Frequency**
Outram park > Yum cha > Tanah Ferry	1200
Shoppes > Esplanade > Woodlands	1085
Golden Mine > Gusti Bed & Breakfast > Masjid Sultan	1070
Song Fa > Ventura View > The Tuckshop	1165
VivoCity > Singapore Zoo > Orchard Gateway	1244
Woodlands > Singapore Center > Woodlands Center	1223
Marina bay > Universal studios > ZAM	1213
Regent Singapore > Universal Studios > Lucky Plaza	1123
Arab Street > Geylang > **Sentosa** Island	1087
Cinema Complex > Bugis Junction > SEA Aquarium	1165
Singapore Changi > Intercontinental > Singapore Sports Hub	965
Singapore Changi > Zouk > Swee Choon	1012
Cinema Complex > Maxwell Food Center > Carl’s Jr.	865
Festive Hotel > Jurassic Park > Sephora	760

Table 4.9: Popular Tourist Venues

The sequential pattern which received the highest number of tourists include VivoCity > Singapore Zoo > Orchard Gateway while Festive hotel > Jurassic park > Sephora was the least involved activity pattern.

4.3.6 *SEQUENCE 6: Female Tourist Sequential Activity Patterns*

The table below shows the sequential activity patterns of the female tourists.

ACTIVITY SEQUENCE 1 > 2 > 3	
SEQUENTIAL ACTIVITY PATTERN	**FREQUENCY**
Dining > Travel > Religious	480
Dining > Refreshment > Archives	650
Hiking > Nature Walk > Dining	400
Refreshments > Dining > Arcade	365
Dining > Shopping > Nature	278
Entertainment > Archives > Dining	415
Hiking > Nature Walk > Entertainment	498
Field > Nature > Outdoor	370
Travel > Dining > Nature Walk	589
Nature > Shopping > Resting	220
Archives > Outdoor > Sporting	120

Table 4.10: Sequential Activity Patterns of Female Tourists

The above activity patterns show that most of the female tourists preferred to engage in shopping followed by nature walk, whereas fewer were involved in outdoor and sporting activities. This could also be attributed to the varying age groups of the female tourists that were included in this research.

4.3.7 ***SEQUENCE 7: Male Tourist Sequential Activity Patterns***

The table below shows the sequential activity patterns of male tourists in Singapore which are quite different to those of females.

ACTIVITY SEQUENCE 1 > 2 > 3	
SEQUENTIAL PATTERNS	**FREQUENCY**
Hiking > Travel > Recreation	200
Travel > Archives > Entertainment	500
Hiking > Entertainment > Gaming	550
Dining > Religious > Nature Walk	400
Archives > Refreshments > Gaming	265
Dining > Recreation > Exercise	178
Gaming > Dining > Market	315
Dining > Travel > Field	498
Exercise > Outdoor > Religious	370
Field > Travel > Gaming	189
Archives > Sporting > Entertainment	120
Sporting > Outdoor > Dining	120

Table 4.11: Sequential Activity Patterns of Male Tourists

It can be deduced from the above chart that male tourists in Singapore are more prone to Sporting, Gaming and Entertainment activities. Singapore being a hub of multiculturalism and tourism, offers a variety of activities and ventures that satisfies its consumers.

4.3.8 ***<u>SEQUENCE 8: Places Mostly Visited by Male Tourists</u>***

The table below shows the sequential pattern of places mostly visited by male tourists.

VENUE SEQUENCE 1 > 2 > 3	
SEQUENTIAL PATTERNS	**FREQUENCY**
Chinatown > Pinnacle Skybridge > Thian Temple	390
Clarke Quay > **Covenant** > **Copthorne**	213
Disco > Din Tai > Duala Lo	212
East Coast > Esplanada > Singapore City Gallery	167
Gardens > Entertainment Complex > Gym	436
Harbour F > Hard **Rock** > Highland	410
Ion Orch > Jem > Jew Kit	568
Johor-Si > **Johor**-S > Jumbo Se	56
Marina B > Makan Sut > Masjid Sultan	745
National > Night Sa > Newton F	540
Orchard > Outram P > One Farr	440
Woodland > Vivo City > Yum Cha	810

Table 4.12: Places Visited by Male Tourists in Sequential Order

Chinatown > Pinnacle Skybridge > Carl's Jr is ranked as the most popular series of places visited by the tourists in Singapore. This may be attributed to the fact that the venues are all in close proximity and marketed as the top tourist destinations in Singapore.

4.3.9 ***SEQUENCE 9: Tourist Sequential Activity Patterns on Daily Basis***

The most common activity found is that tourists and visitors visit tanning salon, use the bus station and go to the mall. In terms of entertainment, the activity sequences demonstrate that tourists travel to the mall, movie theater and nightclub for their entertainment. The most common mode of transportation is through the bus, followed by the train.

ACTIVITY SEQUENCE – A>B	
SEQUENTIAL PATTERNS	**FREQUENCY**
Residential---->Bus	405
Tanning Salon---->Home	381
Building---->Mall	373
Bus Station---->Residential	367
Mall---->Building	346
Movie Theater---->Cineplex	335
Metro---->Mall	331
Mall---->Cineplex	330
Residential---->Mall	312
Education---->Mall	305
Nightclub---->Mall	299
Mall---->Education	298
Mall---->Residential	298
Cineplex---->Mall	298
Mall---->Food Court	297
Home---->Housing Development	291
Food Court---->Mall	291
Train Station---->Mall	289
Cineplex---->Movie Theater	288
Housing Development---->Home	287
Mall---->Nightclub	285

Table 4.13: Tourist Daily Activity Patterns

4.3.10 *SEQUENCE 10: Tourism Receipts for Visitor's Arrival*

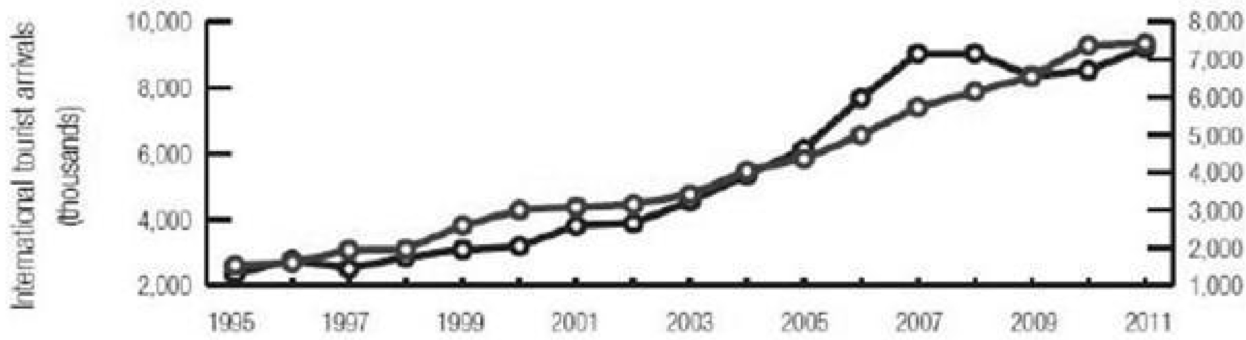


Figure 4I: International Tourist Arrival in Singapore

MARKET	MAJOR COMPONENTS TOURISM RECEIPTS PER CAPITA EXPENDITURE (S$)				
	SHOPPING	ACCOMMODATION	FOOD & BEVERAGE	OTHER TR COMPONENTS[1]	TOTAL[2]
TOTAL	**354**	**345**	**152**	**365**	**1,216**
AMERICAS	**215**	**444**	**175**	**435**	**1,268**
Canada	310	271	134	238	953
USA	201	489	187	503	1,380
SOUTHEAST ASIA	**306**	**221**	**117**	**302**	**945**
Indonesia	300	172	94	332	898
Malaysia	252	166	89	209	717
Philippines	254	304	171	252	981
Thailand	291	325	146	279	1,041
Vietnam	430	292	144	316	1,181

Table 4.14: Global Tourism Receipts

The table show the tourist receipt in many countries like America, Canada, USA and so on.

For 2018, the tourism receipts to be in the range of S$27.1 to S$27.6 billion (+1 to +3%) and international visitor arrivals to be in the range of 17.6 to 18.1 million (+1 to +4%).

With the global economic outlook looking favorable and the tourism poised to expand, which is generally optimistic about tourism prospects for the year ahead for the country of Singapore. There are challenges that remain, however, especially geopolitical tensions that might affect consumer travel sentiments and intensifying regional competition.

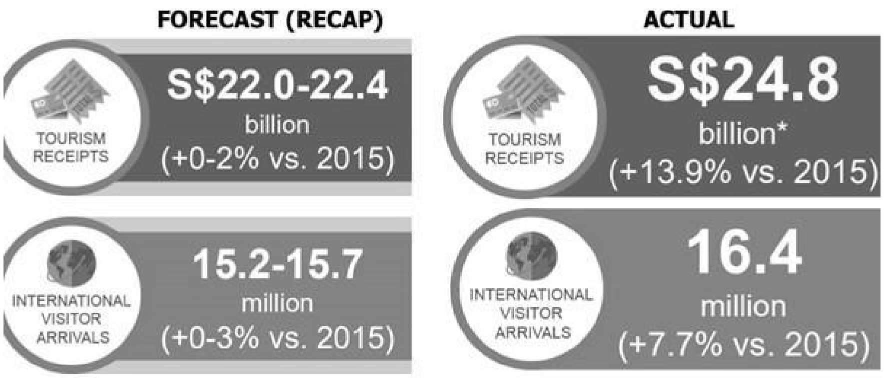


Figure 4J: Tourist Arrival in Singapore - Forecast & Actual

The past research has been useful in understanding travelers' conduct, in any case, little data is increased about tourist arrivals in Singapore as shown in above figure at a specific destination which is essential for the travel industry the board and can aid numerous ways. This examination will fill the hole in the travel industry writing by considering traveler exercises in the successive request, in giving rich data about visitor decisions, inclinations and choices while visiting a specific destination.

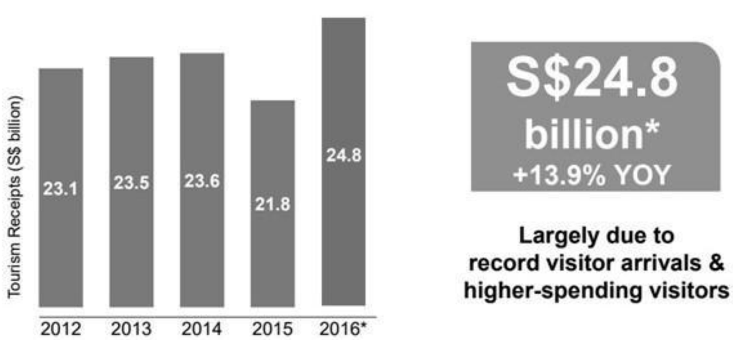


Figure 4K: Annual Tourism Receipts in Singapore

The tourist really did at that specific area, which exercises they were associated with, and how much time they spent at that specific movement in different countries.

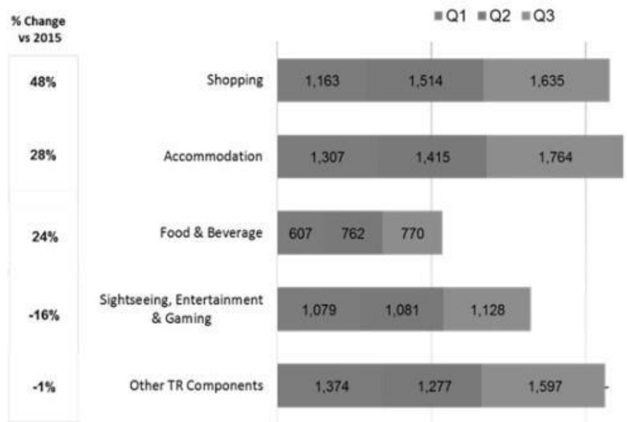


Figure 4L: Cumulative Tourist Interests in Singapore

It gives a sign that tourists have diverse tastes and inclinations while visiting a specific area in the world. The outcomes give sufficient data to the tourist administration experts in Singapore to enable them to devise better methodologies regarding tourist visits.

4.4 SUMMARY

To summarize, different tourists have unique sequences in terms of check-in times for certain destinations or activity. Using the Apache software, the execution of the prefix span yielded the sequence tourists activity patterns. Nonetheless, there are some common sequences that can be picked from the analysis such as the sequences 14 and 15 as well as sequences 2 and 3 from table 3. The typical day according to the respective check-in times of the tourists was split into three sections namely; morning, afternoon, and Evening. Most of the tourists spent their mornings traveling by either train or air. In the afternoon the spend shopping or at the food courts while in the evening they visit various places such as art gallery museums, gardens, islands, and theme parks. However, it was not the case for all the tourists according to the individual sequential patterns. The Pattern-Growth Prefix Span framework process was effective in the study since it provided adequate patterns of the tourist activities in a sequential manner. The tourism management of Singapore will be able to use the insights of the most interesting patterns to improve delivery of services whenever necessary.

CHAPTER 5: DISCUSSION, RECOMMENDATIONS AND CONCLUSION

5.1 DISCUSSION

Tourists are motivated by different factors such as adventure and social influence to visit different destinations including attraction sites across the globe. However, the majority of them have individual taste and preferences which are independent of the motivating factors. The understanding of the tourist activity patterns helps in offering guidance to the management team in the tourism sector to enhance its service delivery. Time which is often referred to as the 'check-in time' is one of the essential factors in the determination of the activity patterns of tourists. In recent years, there has been a consistent annual increase in the number of international tourists across the world. Based on the statistics from the United Nations World Tourism Organization (UNWTO), the number of international tourists arrival is projected to cross the 60 million mark reached in 2015. With the evaluation of the literature review, it was established that most of the researchers have focused on the investigation of the tourists mobility patterns only. This research study realized a gap with the incorporation of an order with respect to the particular activity patterns in the tourist destinations along with their mobility patterns. In regards, the tourists' activities were arranged from morning to evening such that activities that occurred in the morning hours were given priority and appeared at the top of the sequence.

The main objective of this research study was to find out the most appropriate design method that will be used to generate the tourist sequential activity patterns in Singapore. The data of the study was obtained through the Foursquare program using the Twitter API Application. Also, GPS technology played a key role in identifying the location of an individual at a particular time of the day. Ideally, tourists tend to move from one destination to another while taking pictures and posting them through social media platforms. Due to the advancement in technology, the devices are able to record the time as well as the locations of the users including the date of the visit to a particular destination. Over the past years, Singapore has received high numbers of international tourists from across Europe and the neighboring Asian countries. Having evaluated different methods of sequential pattern mining (SPM), the pattern-growth method which uses the prefix-span algorithms, was chosen, to develop and generate the sequence activity patterns of tourists. The SPM algorithms runs on different programming languages of the machine language such as java and python. Some of the algorithms such as the prefix span are only compatible with the Windows operating system. Different scenarios or cases were used while generating the sequence activity patterns of the tourists.

The analysis was conducted in several different scenarios using the pattern-growth method by executing the prefix-span algorithms. The results revealed that tourists have common tastes and preferences while visiting different destinations and attraction sites in Singapore. There were more female tourists than males, based on the Singapore dataset user in the study. In most cases, there is no significant gap between the male and female tourists since the majority of them visit as a couple or partners. Subsequently, most of the male tourists were from Japan and Kuwait while the majority of the female tourists were from Thailand and Malaysia including Peru. Ideally, a larger percentage of the tourists captured were from the South-Asian countries with a few of them being from Europe and across the world. With the application of the prefix-span, minimum support and confidence values play a key role in the generation of the sequence activity patterns of the tourists. The obtained sequence activity patterns help the management of the tourism board in Asia to make an informed decision regarding the delivery of service and the improvement of the program standards at the destinations in the country.

The ten case scenarios documented in this writing involved generating the sequential activity patterns of the tourists during the morning hours (7am – 2pm). There were different frequencies from the respective sequential activity patterns obtained, indicating that tourists have their own preferred schedules of visiting places and choosing activities in Singapore. In the morning hours, most of the tourists preferred Travelling > Religious Activities > Dining and Shopping > Recreation > Entertainment. The dining activity came at last of the sequence which marked the end of the morning period. Nature > Refreshments > Dining was the least frequented activity pattern by the tourists. Religious activities, nature walks, and dining dominated most of the tourists' schedules in the morning hours. On the other end, during the afternoon hours, the majority of the tourists engaged in field walks, visiting historical places and engaging in a variety of outdoor activities . In essence, the afternoon patterns exhibited Field > Exhibition > Nature and Dining > Walking > Health Activities (Gym, Spa/Sauna, Cycling). In the afternoon, Dining

> Archiving > Walking was the least frequented sequential activity pattern. This may be attributed to the changing temperatures which are not conducive for field walk and hiking.

The most predominant sequential activity pattern among both male and female tourists includes: Shopping > Field > Nature Walk and Hiking > Dining > Shopping throughout the day. Interestingly there is a disparity in terms of the sequential activity patterns for males and female tourists in Singapore. A case for analyzing the sequential activity patterns for female tourists revealed that Shopping > Travel > Nature Walk was the widely followed sequential pattern among the female tourists. On the other end of the scenario, the majority of the male tourists preferred to follow Hiking > Entertainment > Gaming. However, this may keep changing from time to time. Essentially male tourists prefer entertainment more than females although this has been changing significantly in recent years with more female tourists preferring entertainment. The time frame taken into account in this study was between 07:00 hrs. and 22:00 hrs. The prefixspan algorithm has a way of automatically generating the sequential activity patterns depending on the pseudo code and the input parameters placed.

5.2 RECOMMENDATIONS

Sequential Pattern Mining (SPM) is one of the evolving fields with new approaches and design methods being proposed and utilized by researchers worldwide. The advancement in technology has resulted in the development of new robust methods and algorithms used to mine the sequential activity patterns of tourists in the tourism industry. From the findings and the case scenarios analyzed in the study, it is evident that tourists exhibit different sequential activity patterns while visiting a destination. Singapore is considered to be a modern tourist destination with numerous attractions sites. As a result, international tourists have a variety of activities to choose from thereby assuming different schedules during their visitation. Since more efficient methods and algorithms are being innovated in the coming years, there is a need to keep on generating new sequential activity patterns periodically and not to rely on the past activity patterns to make informed decisions at the management level. Tourists tend to change their minds and evolve frequently thus the evolution of the generations make it more possible that the young tourists will assume different sequential activity patterns as compared to the elderly. In regards, it is important to constantly generate the sequential activity patterns.

Data inconsistency is also an area of concern for sequential pattern mining. There is a need to ensure that the collected data is credible and reliable and solely used for the purpose of mining sequential activity patterns regardless of the method used. The more credible the data, the more reliable outcomes can be expected in terms of the sequential activity patterns. In recent years, data privacy and integrity has been several compromised by the security intruders which are interested in obtaining the tourist data thereby jeopardizing their social lives. Consequently, it is important for tourist operators to uphold data privacy be installing monitoring systems and constantly checking on potential intruders. The risk of insecurity attributed to the travel-related activities has prompted tourists to share less through the social media platforms such as Twitter which makes it more difficult to obtain the required data for carrying out the sequential pattern mining. Tourist agencies need to partner with the national governing bodies to ensure the security of the country such as Singapore to ensure that the tourists visiting are safe and free to interact through the online platform.

In order to have meaningful and comprehensive sequential patterns, it is important to include factors such as the season culture and social status which are potential influences in the pattern generation. For example, tourists from the middle class would have different taste and preferences as compared to the tourists form a higher class. Tourists from other Asian countries will become easily accustomed to the culture and beliefs of the people of Singapore. In regards, it is paramount to consider these factors among other such as entrance fees, time of the year, and the language barrier. Also, security of the country plays a key role in either encouraging or discouraging visitation & number of international tourists. In as much as the tourism managers may rely on the tourist sequential activity patterns generated, it is also necessary to consider these factor in a bid to enhance customer service delivery and improve the standards of the respective destinations.

5.3 CONCLUSION

The history of the sequential pattern mining is dated in the computer systems and manufacturing systems where professionals were using the technique to understand the various processes in the industry. There have been various methods and approaches that have been proposed and utilized in the mining of the sequential patterns from a given dataset. Databases or datasets can either be classified as transactional or sequential. Most of the methods which use algorithms work well with the sequential databases as composed to the transactional databases. The APRIORI-based method is one of the oldest approaches implemented towards the generation of sequential patterns from datasets. With the advancement in technology, there has been continued improvement of the existing methods as well as the emergence of new methods such as the pattern-growth method which uses the prefix-span or free-span algorithms.

Furthermore, numerous challenges have come with the implementation of the sequential pattern mining methods on datasets. Some of the difficulties have been attributed to incomplete datasets making it hard for the algorithms to be executed successfully. In the research study, the Singapore dataset was considered to be a sequential database because of the presence of repeated entries. As a result, it made it easier to generate the sequential activity patterns of the tourists. The Pattern-Growth method which used the prefixspan algorithm was used to mine the tourist sequential activity patterns for the research study. The data was obtained from the Foursquare program and Twitter API application. Before running the algorithm and implementing the Pattern-Growth method, data was first cleaned to remove incomplete and duplicate data. There were five different scenarios that were identified in the mining of the sequential activity patterns. First, the analysis was based on the activities of the tourists during the morning hours and in the afternoon hours. It was established that most tourists preferred hiking and nature walks in the morning hours and doing shopping in the afternoon hours. In other cases, male tourists preferred entertainment and gaming as opposed to female tourists who prefer to go shopping in various malls. The obtained sequential activity patterns help the management of the tourism sector in Singapore to enhance operation and service delivery. Notably, there is a possibility of the mined patterns to change drastically over a certain period of time owing to the evolution of the generations which lead to a shifting in the taste and preferences of tourist activities.

Using the comparison analysis, the Prefixspan algorithm was proven to be more effective as compared to other algorithms such as SPM and APRIORI. Consequently, the prefixspan algorithm was used in the mining of the sequential activity patterns using the Singapore datasets. The transformation in the needs and wants from one generation to another call for the proposed use of the SPM technique with the incorporation of the most effective methods and algorithms to ensure that comprehensive and reliable outcomes are achieved. In the near future, improvement in the existing methods and emergence of new approaches will be able to combat the current challenges thus making them inevitable with the evolution of technology.

5.4 ACKNOWLEDGEMENTS

This research was successfully accomplished and documented through the help and support from everyone, including: parents, teachers, family and friends. Foremost, I would like to express my sincere gratitude to my mentors Hua Wang, Huy Quan and Yuan Miao for their continuous support during the research, for their patience, motivation, enthusiasm and immense knowledge. Their guidance, encouragement and insightful comments helped me during the time of research and writing of this thesis. I could not have imagined having a better group of people as mentors for my studies. My sincere thanks also go to my family for their unconditional support throughout my journey.

Due to the nature of this research and considering the low risk and minimal budget, the research was not funded in anyway by a grant.

A Humans Resource Ethics Committee (HREC) approval was not necessary for this research as this information that the users had made freely available through Foursquare to the public domain. The data collected was exclusively on publicly available information that had been made public by the individual users in regard to Foursquare privacy laws and regulations. However, the humans whose data was implicated in this thesis were protected as the individual users are not identifiable through the research. As this project is deemed of negligible risk, it does not require human ethics review, yet the privacy of the tourist's data is being protected as the data is anonymous and the user ID's are encrypted and do not trace back to the individual user accounts thereby protecting their privacy. Bearing in mind the privacy issues, the work must suggest and use the Foursquare platform's general privacy policy. Users who build an account on Foursquare to use their customer services have already entered into an agreement with it, granting Foursquare the right to collect and use their data for enhancement, customization and further development purposes which therefore dismisses the necessity of a HREC approval for this research.

REFERENCES

[1] Glaesser, D, Kester, J, Paulose, H, Alizadeh, A & Valentin, B 2017, 'Global Travel Patterns: An Overview', Journal of Travel Medicine, vol. 1, pp. 1-27.

[2] Lau, G & McKercher, B 2006, 'Understanding tourist movement patterns in a destination: A GIS approach', Tourism and Hospitality Research, pp. 39-49.

[3] Versichele, M, Groote, L, Bouuaert, MC, Neutens, T, Moerman, I & Weghe, NV 2014, 'Pattern mining in tourist attraction visits through association rule learning on Bluetooth tracking data: A case study of Ghent, Belgium', Tourism Management, vol. 44, pp. 67-81.

[4] Zhu, YX & Qun, W 2014, 'Exploratory space-time analysis of inbound tourism flows to china cities', International Journal of Tourism Research, vol. 16, pp. 303-312.

[5] Hallo, JC, Beeco, JA, Goetcheus, C, McGee J, McGehee, NG & Norman, WC 2016, 'GPS as a method for assessing spatial and temporal use distributions of nature-based tourists', Journal of Travel Research, vol. 51, no. 5, pp. 591-606.

[6] Bujosa, A, Riera, A & Pons, PJ 2015, 'Sun and beach tourism and the importance of intra-destination movements in mature destinations', Tourism Geographies, vol. 17, no. 5, pp. 780-794.

[7] Kadar, B 2014, 'Measuring tourist activities in cities using geotagged photography', Tourism Geographies, vol. 16, no. 1, pp. 88-104.

[8] Yang, Y & Wong, KKF 2012, 'Spatial distribution of tourist flows to China's cities', Tourism Geographies, vol. 15, no. 2, pp. 338-363.

[9] Birenboim, A, Clave, SA, Russo, AP & Shoval, N & 2013, 'Temporal activity patterns of theme park visitors', An International Journal of Tourism Space, Place and Environment, vol. 15, no. 4, pp. 601-619.

[10] Chen, JS, Wang, W & Prebensen, NK 2015, 'Travel companions and activity preferences of nature-based tourists', Tourism Review, vol. 71, no. 1, pp. 45-56.

[11] Cantis, S, Ferrante, M, Kahani, A & Shoval, N 2016, 'Cruise passengers' behavior at the destination: Investigation using GPS technology', Tourism Management, vol. 52, pp. 133-150.

[12] Orellana, D, Bregt, AK, Ligtenberg, A & Wachowicz, M 2012, 'Exploring visitor movement patterns in natural recreational areas', Tourism Management, vol. 33, pp. 672-682.

[13] Ting, HX & Hu, WB 2012, 'Intra-attraction tourist spatial-temporal behavior patterns', Tourism Geographies, vol. 14, no. 4, pp. 625-645.

[14] Yeung, MW, Kim, S & Schuckert, M 2016, 'Japanese tourists to Hong Kong: their preferences, behavior, and image perception', Journal of Travel & Tourism Marketing, vol. 33, pp. 730-741.

[15] Dejbakhsh, S, Arrowsmith, C & Jackson, M 2011, 'Cultural influence on spatial behavior', Tourism Geographies, vol. 13, no. 1, pp. 91-111.

[16] Klahn, DT, Roosen, J, Gerike, R & Hall, CM 2015, 'Factors affecting tourists' public transport use and areas visited at destinations', An International Journal of Tourism Space, Place and Environment, vol. 17, no. 5, pp. 738-757.

[17] Beeco, JA, Huang, WJ, Hallo, JC, Norman, WC, McGehee, NG, McGee, J & Goetcheus, C 2013, 'GPS tracking of travel routes of wanderers and planners', Tourism Geographies, vol. 15, no. 3, pp. 551-573.

[18] Wang, H & Sharaf, MA 2014, 'Databases Theory and Applications: 25th Australasian Database Conference', Brisbane, QLD, Australia, July 14-16, 2014, Proceedings - New York: Springer.

[19] Li, Z, Deng, S & Moutinho, L 2014, 'The impact of experience activities on tourist impulse buying: an empirical study in China', Asia Pacific Journal of Tourism Research, vol. 20, no. 2, pp. 191-209.

[20] Li, Y, Xiao, L, Ye, Y, Xu, W & Law, A 2016, 'Understanding tourist space at a historic site through space syntax analysis: The case of Gulangyu, China', Tourism Management, vol. 52, pp. 30-43.

[21] Bauder, M & Freytag, T 2015, 'Visitor mobility in the city and the effects of travel preparation', Tourism Geographies, vol. 17, no. 5, pp. 682-700.

[22] Zheng, Y & Zhou, X 2011, 'Location-based social networks: Users. In Computing with Spatial Trajectories', Report Springer, viewed on 20th Dec 2016.

[23] Zheng, Y 2016, Mobile LBS Market in the US 2015-2019 – Reportlinker Review.

[24] Vu, HQ, Li, G, Law, R & Ye, BH 2015, 'Exploring the travel behaviors of inbound tourists to Hong Kong using geotagged photos', Tourism Management, vol. 46, pp. 222-232.

[25] Paresh, K 2010, 'Tourism and Economic Growth: A Panel Data Analysis for Pacific Island Countries', Journal of Economies, vol.1, no.2, pp.5-15.

[26] Buhalis, D & Michopoulou, E 2011, 'Information-enabled tourism destination marketing: Addressing the accessibility market', Current Issues in Tourism, vol. 14, no. 2, pp.145-168.

[27] Koutoulas, D & Zoyganeli, S 2007, 'Analyzing the destination marketing task at the regional level: The case of prefecture promotion committees in Greece', Paper presented at 1st Biannual International Conference in September 2007.

[28] Chaitip, P, Chaboonsri, C, Kovacs, S, Balogh, P 2008, 'A Structural Equation Model: Greece Tourism Demand for Tourist Destination', Applied Studies in Agribusiness and Commerce, vol. 1, pp. 75-83

[29] Buhalis, D 2000, 'Tourism in Greece: Strategic Analysis and Challenges', Current Issues in Tourism, vol. 4, no. 5, pp.440-480

[30] Gan, W, Lin, JC, Fournier-Viger, P & Yu, PS 2018, 'A Survey of Parallel Sequential Pattern Mining', Journal of Latex Class Files, vol.1, pp. 1-21.

[31] Huang, JZ, Cao, L & Srivastava, J 2011, 'Advances in Knowledge Discovery and Data Mining: 15th Pacific-Asia Conference', Shenzhen, China, May 24-27, 2011, Proceedings, Part 1. Berlin: Springer.

[32] Caplinskas, A 2013, 'Databases and Information Systems VII: Selected Papers from the Tenth International Baltic Conference', DBIS 2012, Amsterdam: IOS Press.

[33] Mooney, CH & Roddick, JF 2013, 'Sequential Pattern Mining: Approaches and Algorithms', ACM Computing Surveys, pp. 1-46.

[34] Cios, KJ, Pedrycz, W, Swiniarski, RW & Kurgan, LA 2007, 'Data Mining: A Knowledge Discovery Approach', New York: Springer Science & Business Media.

[35] John, C, Arentze, T, Hofman, F & Timmermans, H 2002, 'Activity Pattern Similarity: A Multidimensional Sequence Alignment Method', Transportation Research Part B Methodological, pp. 385–403.

[36] Hyuk-Jin, L & Joh, C 2010, ‘Tourism Behaviour in Seoul: An Analysis of Tourism Activity Sequence using Multidimensional Sequence Alignments’, Tourism Geographies, vol.1, pp. 487-504.

[37] Xu, Z, Kang, JE & Chen, R 2017, ‘A Random Utility-Based Estimation Framework for the Household Activity Pattern Problem’, Transportation Research Procedia, pp. 809-826.

[38] Bathelt, H., & Zeng, G. (2015). Temporary Knowledge Ecologies: The Rise of Trade Fairs in the Asia-Pacific Region. Massachusetts: Edward Elgar Publishing.

[39] Li, X, Meng, F & Uysal, M 2008, ‘Spatial Pattern of Tourist Flows Among the Asia-Pacific Countries: An Examination over a Decade’, Asia Pacific Journal of Tourism Research, pp. 229-243.

[40] Kotler, P & Gertner, D 2002, ‘Country as a Brand, Product, and Beyond: A place Marketing and Brand Management Perspective’, Journal of Brand Management, vol. 4, no. 5, pp. 249-261.

[41] Hall, MC & Page, S 2012, ‘Tourism in South and Southeast Asia’, Abingdon: Routledge.

[42] Yang, L, Wu, L, Liu, Y & Kang, C 2017, ‘Quantifying Tourist Behavior Patterns by Travel’, International Journal of Geo-Information, pp. 2-18.

[43] Zimanyi, E 2017, ‘Business Intelligence: Third European Summer School’, Dagstuhl Castle, Germany, July 7-12, 2013, Tutorial Lectures, New York:Springer.

[44] Aggarwal, C. C. (2013). Managing and Mining Sensor Data. Berlin: Springer Science & Business Media.

[45] Kim, S & Reijnders, S 2017, ‘Film Tourism in Asia: Evolution, Transformation, and Trajectory’, New York: Springer.

[46] Echtner, C, Ritchie, B 2003, ‘The Meaning and Measurement of Destination Image’, The Journal of Tourism Studies vol. 14, no. 1, pp. 37–48.

[47] Mistilis, N & Daniele, R 2005, ‘Challenges for Competitive Strategy in Public and Private Sector Partnerships in Electronic National Tourist Destination Marketing Systems’, Journal of Travel & Tourism Marketing, vol. 17, no. 4, pp. 63-73.

[48] Nezirovic, S 2013, ‘Tourism, Politics and Regional Development – Guidelines of the Development of Northeastern Bosnia, in Contemporary trends in Tourism and Hospitality’, A Collection of Papers, pp. 31-37.

[49] Sheehan, L, Ritchie, JRB & Hudson, S 2007, ‘The Destination Promotion Triad: Understanding Asymmetric stakeholder Interdependencies Among the City, Hotels and DMO’ Journal of Travel Research, vol. 46, pp. 64-74.

[50] Dwyer, L, Livaic, Z & Mellor, R 2003, ‘Competitiveness of Australia as a tourism destination’, Journal of Hospitality and Tourism Management , vol. 10. no. 1, pp. 60-78.

[51] Mazilu M, 2012, ‘Sustainable Tourism of Destination, Imperative Triangle Among: Competitiveness, Effective Management and Proper Financing’, Sustainable Development - Policy and Urban Development - Tourism, Life Science, Management and Environment, INTECH.

[52] Angella, F., Go, F., (2010), Tale of two cities’ collaborative tourism marketing: Towards a theory of destination stakeholder assessment, Tourism Management Vol.30, pp.429 –440

[53] Lopes, S 2011, ‘Destination image: Origins, Developments and Implications, PASOS’, Revista de Turismoy Patrimonio Cultural, vol. 9, no. 2, pp. 305-315.

[54] Yun, HJ & Park, MH 2014, ‘Time-space movement of festival visitors in rural areas using a smartphone application’, Asia Pacific Journal of Tourism Research, vol. 20, no. 11, pp. 1246-1265.

[55] Bauder, M 2015, 'Using GPS supported speed analysis to determine spatial visitor behavior', International Journal of Tourism Research, vol. 17, pp. 337-346.

[56] Long, X, Jin, L & Joshi, J 2013, 'Towards understanding travel behavior in Location-Based Social Networks', Symposium on Selected Areas in Communications, School of Information Science.

[57] Maghrebi, M, Abbasi, A, Rashidi, TH & Waller, ST 2015, 'Complementing travel diary surveys with Twitter Data: Application of text mining techniques on activity location, type and time', IEEE 18th International Conference on Intelligent Transportation Systems, School of Engineering & IT, UNSW.

[58] Grinberger, AY, Shoval, N & McKercher, B 2014, 'Typologies of tourists' time-space consumption: a new approach using GPS data and GIS tools', Tourism Geographies, vol. 16, no. 1, pp. 105-123.

[59] Kitajima, M, Tahira, H, Takahashi, S & Midorikawa, T 2012, 'Understanding tourists' in situ behavior: A cognitive chrono-ethnography study of visitors to a hot spring resort', Journal of Quality Assurance in Hospitality & Tourism, vol. 13, pp. 247-270.

[60] Zoltan, J & McKercher, B 2015, 'Analysing intra-destination movements and activity participation of tourists through destination card consumption', Tourism Geographies, vol. 17, no. 1, pp. 19-35.

[61] Wang, D, Qian, J, Chen, T, Zhao, M & Zhang, Y 2014, 'Influence of the high-speed rail on the spatial pattern of regional tourism – taken Beijing-Shanghai high-speed rail of China as example', Asia Pacific Journal of Tourism Research, vol. 19, no. 8, pp. 890-912.

[62] Yang, Y, Fik, T & Zhang, J 2013, 'Modeling sequential tourist flows: Where is the next destination?', Annals of Tourism Research, vol. 43, pp. 297-320.

[63] Zheng, W, Huang, X & Li, Y 2017, 'Understanding the tourist mobility using GPS: where is the next place?', Tourism Management, vol. 59, pp. 267-280.

[64] Vu, HQ, Leung, R, Rong, J & Miao, U 2016, 'Exploring Park Visitors' Activities in Hong Kong using geotagged photos', International Tourism and Hospitality, pp. 183-196.

[65] Leung, R, Vu, HQ, Rong, J & Miao, Y 2016, 'Tourists visit and photo sharing behavior analysis: A case study of Hong Kong temples', International Tourism and Hospitality, pp. 197-209.

[66] Chua, A, Servillo, L, Marcheggiani, E & Moere, AV 2016, 'Mapping Cilento: using geotagged social media data to characterize tourist flows in southern Italy', Tourism Management, vol. 57, pp. 295-310.

[67] Oh, H 2009, Handbook of Hospitality Marketing Management, Abingdon, Routledge.

[68] Triantaphyllou, E & Felici, G 2006, 'Data Mining and Knowledge Discovery Approach Based on Rule Induction Techniques', Berlin: Springer Science & Business Media.

[69] Uysal, M, Sirakaya-Turk, E & Li, X 2008, Chapter 15, Push-pull dynamics in travel decisions, Abingdon: Routledge.

[70] Majid, A, Chen, L, Mirza, HT, Hussain, I & Chen, G 2015, 'A system for mining interesting tourist locations and travel sequences from public geo-tagged photos', Data and Knowledge Engineering, vol. 95, pp. 66-86.

[71] Zhang, J 2017, 'Life-Oriented Behavioral Research for Urban Policy', New York: Springer.

[72] Goulet-Langlois, G 2013, 'Exploring Regularity and Structure in Travel', Massachusetts: Massachusetts Institute of Technology.

[73] Wilson, C 2008, ‘Activity patterns in space and time: Calculating representative Hagerstrand trajectories’, Transportation, pp. 485-499.

[74] Shapoval, V, Wang, MC, Hara, T & Shioya, H 2017, ‘Data Mining in Tourism Data Analysis: Inbound Tourists to Japan’, Journal of Travel Research, pp. 1-42.

[75] Kim, J, Thapa, B, Jang, S & Yang, E 2018, ‘Seasonal Spatial Activity Patterns of Tourists with a Mobile Exercise Application at Seoraksan National Park’, South Korea Sustainability, pp. 1-21.

[76] Kisilevich, S, Keim, D & Rokach, L 2010, ‘A Novel Approach to Mining Travel Sequences using Collections of Geotagged Photos’, Springer, pp. 163-182.

[77] Boghey, R., & Singh, S. (2013). Sequential Pattern Mining: A Survey of Approaches. International Conference on Communication Systems and Network Technologies, 671-674.

[78] Kwan, MP 2000, ‘Interactive visualization of activity-travel patterns using three-dimensional geographic information systems’, Transport Research Part C, pp. 185-203.

[79] Han, J & Aggarwal, C 2013, ‘Frequent Pattern Mining’, Berlin: Springer Media.

[80] Pei, J, Han, J, Wang, J, Pinto, H, Mortazavi-asl, B, Chen, Q & Hsu, MC 2004, ‘Mining Sequential Patterns by Pattern-Growth: The PrefixSpan Approach’, IEEE Transactions on Knowledge and Data Engineering, pp. 1-17.

[81] Pratik, S, Sedamkar, RR & Rathi, S 2015, ‘PrefixSpan Algorithm for Finding Sequential Pattern with Various Constraints’, International Journal of Applied Information Systems, pp. 37-41.

[82] Rincy, TN & Pandley, Y 2013, ‘Performance Evaluation on State of Art Sequential Pattern Mining Algorithms’, International Journal of Computer Applications, pp. 8-15.

[83] Ruan, D, Chen, G, Kerre, EE & Wets, G 2005, ‘Intelligent Data Mining: Techniques and Applications’, Berlin: Springer Science & Business Media.

[84] Han, JW, Pei, J & Yan, XF 2004, ‘From Sequence Pattern mining to Structured Pattern Mining: A Pattern Growth Approach’, Journal of Computer Science & Technology, vol.1, pp. 257-279.

[85] Ozdzynski, P & Zakrzewska, D 2017, ‘Using Frequent Pattern Mining Algorithms in Text Analysis’, Information Systems in Management, pp. 213-222.

[86] Mallick, B, Garg, D & Grover, PS 2012, ‘CFM-PREFIXSPAN: A Pattern Growth Algorithm Incorporating Compactness And Monetary, International Journal of Innovative Computing, Information, and Control, pp. 4509-4520.

[87] Chen, CJ, Pai, T, Lin, S, Yeh, CC & Liu, MH 2016, ‘Application of PrefixSpan Algorithms for Disease Pattern Analysis’, New York: IEEE Press Conference.