

Investigation of Social Behaviour Patterns using Location-Based Data – A Melbourne Case Study

Ravinder Singh^{1*}, Yanchun Zhang¹, Hua Wang¹, Yuan Miao¹, Khandakar Ahmed¹

¹Institute for Sustainable Industries and Liveable Cities, Victoria University, Melbourne, Australia.

Abstract

Location-based social networks such as Swarm provide a rich source of information on human behaviour and urban functions. Our analysis of data created by users who voluntarily used check-ins with a mobile application can give insight into a user's mobility and behaviour patterns. In this study, we used location-sharing data from Swarm to explore spatio-temporal, geo-temporal and behaviour patterns within the city of Melbourne. Moreover, we used several tools for different datasets. We used the MeaningCloud tool for sentiment analysis and the LIWC15 tool for psychometric analysis. Also, we employed SPSS software for the descriptive statistical analysis on check-in data to reveal meaningful trends and attain a deeper understanding of human behaviour patterns in the city. The results show that most people do not express strong negative or positive emotions in relation to the places they visit. Behaviour patterns vary based on gender. Furthermore, mobility patterns are different on different days of the week as well as at different times of a day but are not necessarily influenced by the weather.

Keywords: Behaviour Patterns, Social Media, Spatio-Temporal, Mobility Patterns, Swarmapp, Twitter, Sentiment Analysis.

Received on 04 July 2020, accepted on 20 August 2020, published on 26 October 2020

Copyright © 2020 Ravinder Singh *et al.*, licensed to EAI. This is an open access article distributed under the terms of the [Creative Commons Attribution license](#), which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi: 10.4108/eai.26-10-2020.166767

*Corresponding author. Email: Ravinder.singh@vu.edu.au

1. Introduction

Intra-urban mobility has long been a topic of interest across research communities, including urban planners, computer scientists, physicists and geographers [1]. Urban planners are interested in improving transport efficiency by investigating the spatial and temporal differences of travel time and travel flow; geographers are usually interested in the spatial distribution of intra-urban mobility; and computer scientists & physicists are interested in modelling the distribution of travel distance in mathematical ways [2-4]. Similarly, human behaviour related to such intra-urban mobility has been an area of interest for social researchers. These intra-urban mobility patterns can give insight into the behavioural traits of a group of people in a city or at a particular area in a city [5, 6]. The type of places people visit and the time of the day they visit these places tells a lot about their behavioural traits and mobility patterns [7]. The aim of this research is to find new ways to explore human behaviour in today's world of big data and machine learning using Melbourne as a case study. This pilot study proves the effectiveness

of our proposed methodology for conducting large-scale mobility and behaviour studies by utilizing location-based social media data and machine learning techniques for analyses.

Currently, data is being generated at an unprecedented rate, particularly with the advent of social media. In 2015, 7.9 zettabytes of data were generated and this volume is expected to increase to 35.2 zettabytes by 2020 [8-11]. Every activity performed on social media creates data. It has become a part of our daily lives and people use it for various reasons. Facebook, for example, is used to access news and to stay in touch with family and friends. Instagram and Flickr enable users to share photos. Yelp and Foursquare are used to share reviews on services and venues. People use Quora and Reddit as knowledge sharing tools, Pinterest to keep track of things we like and Swarm (a Foursquare subsidiary) to share locations. There is no limit to the number of use cases for social media sites and different people use them for different reasons [12, 13]. The use of social media services is on the rise and these are being integrated into other applications. For example, TV channels give their viewers the ability to discuss a particular show or a sports

event live through their own apps and the apps of others [14, 15]. People indulge in different social media sites for pleasure, education and in some cases under peer pressure. All these activities generate data at an unprecedented rate that can be explored for social behaviour research purposes [10, 11, 16].

1.1. Motivation

Location-based data provides easy access to various human behaviour patterns that may otherwise not be captured by other data collection methods. Some domain areas, such as computer science, have utilized such data sources for some time, however, there is still plenty to learn from this sort of data. An increasing number of users use these platforms to express their emotions, and this online behaviour makes it a useful source for behaviour studies. The data can easily be captured and mined to generate valuable knowledge and aid in decision making [17]. Traditional ways to obtain data on human behaviour and mobility pattern research have become inadequate to meet contemporary policy demands [18]. Data obtained via social media platforms and analysed by machine learning algorithms can make the process efficient and help us discover patterns that are otherwise improbable [19-22]. This research will use check-in data from Swarm, which is a location-based social network (LBSN). Location-based social networks have become important sources of volunteered geographic information (VGI). These networks can be divided into two broad categories, purpose-driven and social driven. A purpose-built network is an application through which people explicitly request another person's location. Some examples of such services are AT&T Family Map and Verizon Family Locator. A social-driven network is an application through which people broadcast their location to family, friends, and followers on the network. Some examples of social-driven networks are Facebook Places, Gowalla and Swarm [23]. Swarm (previously known as Foursquare) is the most popular social-driven LBSN app with the largest user base and daily check-ins, which is why it has been chosen for this research project. The service was initially launched in 2009 as Foursquare and became very popular in a short period of time. The service has more than 60 million registered users and more than 50 million monthly active users. As of October 2017, the platform surpassed 12 billion check-ins and reached an average of 8 million daily check-ins [24]. The service has an app for IOS and Android devices and utilizes the devices' built-in GPS to track the exact location of the user check-in. When it first started, Foursquare was a mobile application that provided local search and discovery services along with check-in features. It provided recommendations for places to visit near the user's location. The app allowed users to leave tips for the different places they had visited as recommendations for others using the app. In May 2014, Foursquare split into two different apps, Foursquare and Swarm. Swarm is mainly used for check-ins and to keep

track of places that users have visited and Foursquare is for tips and recommendations for places of interest. Other reasons people may use the Swarm app are to discover new places, to obtain discounts and special offers and to make new friends, etc. [12]. The platform and the data generated on it can be utilized to conduct large-scaled human behaviour studies.

1.2. Our Approach

In this paper, we propose a methodology to conduct large-scale mobility and human behaviour studies by utilizing location-based social media data. Different people use swarmapp for different reasons, however, the underlying intention for its use is the social aspect [25]. In this study, we analyze user check-ins to discover various mobility and behaviour patterns in Melbourne. In particular, we are interested in finding these patterns based on gender and weather. We also want to explore the various categories of places people visit the most, the busiest check-in times each day, the busiest days of a week and we also undertake sentiment analysis of the messages that users post along with their check-in details. The information obtained from the results of this study can be utilized by local government authorities and businesses to plan travel and business activities in Melbourne. The results can also be utilized to develop behaviour and mobility user profiles, which can further lead to the development of a targeted recommendation system.

1.3. Contribution

Key contributions of the study are: 1) Construction of a medium scale online check-in dataset. 2) Sentiment and Psychometric analysis of check-in data from the Swarmapp. 3) Statistical analysis of check-in data. 4) Topic extraction and knowledge discovery related to behaviour traits using social media data.

The paper is organized into five sections. This section one provides an introduction to the overall research project. Section two details the background and explains how social media data in general, and location-based social media data in particular, have been used to solve several real-world problems. The data has been used in a few research domains but deserves a lot more consideration in others. The section advocates the use of social media data as a main rather than an alternative source of data for academic research. Section three explains the methodology and covers the methods used for data collection, data processing, data transformation, and data analysis. Three types of analyses are carried out in this research project and these are explained in detail in this section. Section four discusses the results of the analyses and section five summarizes the findings, draws a conclusion and also proposes future work that can be carried out based on the findings of this research project.

2. Background

2.1. Social Media Network Data

Online social media has transformed information consumers into information producers. This phenomenon has enticed researchers from various disciplines to study online social media as an important source of data to explore human behaviour in the physical world [26-29]. Social media generates an unprecedented amount of data and has led to new ways to discover urban functions and human behaviour [11]. By aggregating millions of check-ins from platforms such as Swarm and Facebook Places, researchers can uncover distinct visit patterns and busy times for various locations within a city [30]. Based on this information, a recommendation system can be developed to provide real-time information on interesting events and their statistical deviation from past historical trends [31]. Aggregated check-ins can also reveal positive or negative sentiments about places people visit. This can be achieved by conducting sentiment analysis in real-time on the data and the information can then be fed back to a recommender system to make recommendations based on the information [32, 33] which people can use to further plan their outings and visits. Discovering similar trends from data collected by a traditional method, which is also called a ‘top-down approach’ for data collection, will not provide such real-time functionality. Data obtained for urban planning by top-down approaches such as remote sensing, nationwide surveys, and geographic information systems have limitations because this data is intricate and fails to reveal the complex dynamics of a city. It is also difficult to extract the emotions, perceptions, and experiences of people using this approach [34, 35]. Even though such data has been used extensively in the past to study urban functions, urban planning, and geographical information systems, it has limitations in regard to the time and effort required to collect, process and analyze it. On the other hand, data generated by a ‘bottom-up approach’ offers a better alternative and includes user-created data in the form of online blogs, check-ins, and social media posts [36-38]. Researchers are not required to create a questionnaire or a survey to collect data in this form, which is usually a very rich source of information [11]. It can easily be collected using techniques such as APIs, web crawling, software tools such as Capture, and web scraping. By analyzing comments, posts, images, etc. posted by users of social media, researchers can infer important information about them. This may include the type of art people buy, the theatres they visit and the fashion they prefer [39].

User activity modelling has been an area of interest for researchers in the field of pervasive computing. To develop an activity model, researchers usually rely on data from different streams, such as location sensors and smartphones with inbuilt GPS sensors [40, 41]. The data gathered is then analyzed to develop a ‘user behaviour model’ to discover content or

location-dependent knowledge. This knowledge is then fed into a recommender system for targeted advertising and marketing. In particular, such information can be analyzed to determine user activity duration [42]. The activity duration can then lead to other types of predictive analytics. Based on the past time spent at a location, a system can suggest further places of interest to a user. A location-based social network, such as Swarm, offers a new and different avenue for data collection to develop a user activity model [43]. This data can be sufficient on its own, without the need to collect data from other sources to model human activities. Not all check-ins can be used for such a system, however, with millions of check-ins every day, they can be filtered so that those that meet the criteria can be selected and the number can still be substantial. A lot of people check-in regularly when they travel from one place to another. For example, they check-in once they leave office and check-in again when they reach a restaurant or another place of interest. They may check-in again when they reach home after a visit to a place of interest. To develop such a system, only consecutive check-ins from the same users can be analysed. The methodology is to select all tweets within an area of interest for which a user activity model has to be developed. Once collected, these check-ins can be analyzed to estimate the time spent at different locations. However, the time required to travel between two locations needs to be taken into consideration. A user can walk, drive or take public transport to reach the next destination. So, the time spent between destinations has to be estimated and deducted from the total time spent at venues. One easy approach is to take the average of the time it takes by foot, train, and car. Another approach is to find the distance between two locations, and if the distance is shorter than a pre-defined distance, it is likely that the user walked to the next location (for example it does not make sense for a user to catch a taxi or to drive to a location which is only 200-300 meters away). A software program using Google’s location API can be written to estimate the time between locations as well. Once the travel time is estimated, the time spent at a venue can then be calculated. Based on the time spent at a particular location, the user’s interest and partiality for that place can be deduced to develop an efficient activity model [44].

2.2. Data from Social Media and Traditional Sources

Large-scaled location-based data, when compared to other sources of data from questionnaires and surveys, have a lot of advantages as an indicator of human activity categories, especially when it comes to analyzing trends in entertainment, shopping, travel and dining. It provides a fine-grained resolution and is readily available. However, it comes with several limitations in terms of how it represents human mobility [45]. For example, it can contain a group, a venue category or an age bias that

may lead to mobility patterns with certain mechanisms. The data may not be suitable for discovering some kinds of mobility patterns, but it is still a very valuable source for most research studies, especially to study spatial interactions between different categories of places [46-48]. Since spatial interactions are measured via human mobility patterns, check-in data from sources such as Swarm have great potential to be used for discovering spatial interactions, the research area of interest for geographers. Geographers can study consecutive check-ins in a time span shorter than eight hours to ascertain the strength of the relationship between the two or more categories of places. A two-way trip between two places strengthens this pattern and shows a stronger relationship. Similarly, more trips between two places or categories represent that these two are spatially interacted [45, 49, 50]. To realize such patterns, using data collected from surveys and questionnaires is a quite cumbersome, long and expensive process. Location-based social network data is more suitable for this and many other related research problems.

Researchers from some domains have utilized social media data to some extent, however, it can be further utilized for social science studies as well. The work in [51] studied 13,619 Facebook users to show a correlation between their profile information and the online purchases they make on eBay. The research used various machine-learning models with statistical significance to predict the product category from which a user would buy, using information from a user's Facebook profile. The researchers proposed this model to build a cold start recommender system. A cold start recommender system is the only practical solution when an e-commerce company does not have enough information about a user and his purchase history in order to make a recommendation for new products. Most recommender systems are built on two widely used techniques, collaborative filtering methods and content-based methods. Collaborative filtering methods work on the assumption that users with a similar profile (gender, age group, demographics, ethnicity) have similar characteristics and are most likely to buy similar sorts of products. The system works well if the seller has enough information about the user. A content-based method, on the other hand, uses information from the web (blogs, review site, online opinions, tweets, and posts) to rank products and to recommend these to buyers. This approach may work to some extent but not always. A typical recommender system requires a lot of user information, collected through the user's web and purchase history. This information, collected in traditional ways, is not available in this scenario to make a cold start recommender system. The data from the social media fills in the gap very well for such problems.

2.3. Social Media Data for Social Science and Behavioural Science Research

Personal traits and the behaviour patterns of an individual can be discovered from his/her web search history, web browsing history and bank statements [52, 53], however, due to strict privacy restrictions, this information is not available to all. For example, apart from the bank in which a person has an account, only some government organizations have access to a person's account statements. A bank statement can tell the type of places a person has visited and the type of products he/she has bought from physical and online stores. A lot of behaviour patterns can be learned from such data and can be used to offer new products, a personalized search engine, and targeted marketing services. However, due to privacy regulations, even the banks that own this data cannot use it for the aforementioned purposes. A person's web search history and browsing history is only available to some (search engine company and to some developers) and hence cannot be obtained freely by everyone for social research. However, publicly available data set such as check-ins (Swarm, Facebook Places) is easily available to all and can be used for the same. Not every user of Foursquare and Facebook Places shares his/her data but a vast majority of them do. Some users, who do not like to share their activities with others for research purposes, keep their settings private. This prevents their data and activities to be seen by others and to be used for research by the platform they are on, and by others. Only a small number of users keep their profile private because the reason why these users are on the platform is to share information about the things they do and the places they visit which provides a vast source of data for behavioural and social research.

People like to be associated with organizations, places, and activities that they like and this is what platforms such as Swarm and Facebook Places enables them to do [54-56]. By checking into a particular nightclub, a user broadcasts his choice of that place. Similarly, by checking into a particular fashion brand store, a user does the same for that brand. Other traits such as ethnicity, sexual orientation, religious beliefs, and type of food preferences can be discovered from similar check-ins. A study by Kosinski et al. [57] explored how the private traits and attributes of a person are predictable from his/her digital footprints. In their study, the researchers used 58000 volunteers to discover that Facebook likes can be used to accurately and automatically predict a wider range of highly sensitive personal attributes including religion, sexual orientation, political views, ethnicity, happiness, use of an addictive substance, intelligence, gender, age, and parental separation. The model that researchers developed correctly differentiates between heterosexual and homosexual men in 88% of the cases, Republican and Democrats in 85% of the cases and Caucasian American and African Americans in 95% of the cases. The researchers concluded that given the wide variety of personal attributes that the model predicted, a lot more behaviour traits can be discovered if the model is trained with appropriate social media data. The conclusion indicates the need for more research in the area to

discover other aspects of human behaviour using such a dataset that is publicly available to all.

2.4. Real-world application of Location-Based Data

Zhou et al. [58] showed how check-in data from Foursquare and Twitter can be mined to discover urban functions such as transport needs, business distribution and development trends. The authors showed how using check-in data over traditional sources of data such as photography, observations, cognitive maps and remotely sensed imagery makes more sense. Similarly, Hasan et al. [59] used large-scale location-based data from Foursquare to discover urban human activity and mobility patterns in three different cities, New York, Chicago, and Los Angeles. Noulas et al. [60] used check-in data to present the spatio-temporal patterns of users' activities and place transition analysis. The authors collected the Foursquare check-in data from two different sources to avoid any platform bias and found the same patterns from both sets of data. The authors found similar trends from both datasets, validating check-in data as an important source of research data. Check-in data can be a very good avenue by which to discover collective user activities. A higher number of check-ins at a particular bar in a city can imply that the bar is a popular venue. Similarly, a large number of check-ins at a train station can indicate that it is a busy station compared to others in the area. The authors of [60] used a cumulative distribution function to discover the top ten places for check-ins, both on weekdays and weekends.

The patterns discovered were from a large data set and related well to our daily activities. The second part of the study analyzed a different sort of relationship between the places people visited. About 10 percent of users logged two check-ins within a short period of time (less than 60 minutes). These two temporally adjacent check-ins by the same user can signal an important correlation between the two venue types that the user has visited and can indicate a temporal relationship. Similarly, two check-ins made by a single user at two different venues that are not too far away from each other (less than 1 km) can lead to some sort of spatial relationship between the two places. It is possible that both venues fall into the same category such as a bar or restaurant. These can also be from different but related categories of venues such as movies and restaurants. Finding one of these venues can lead to predicting a check-in at the next related venue [61, 62]. A recommender system as a typical business application has been discussed in this paper, however, it is not the only business application of such data. Since the data has predictive potential, it can and has been used in financial markets as an alternative source of data [63-67]. Organizations such as hedge funds and investment banks use check-in data to discover various trends in a geographical region to assist them with investment decisions.

The study by Cheng et al. [68] looked at mobility patterns and human behaviour from three different aspects: repeated check-ins at a same place, whether a user's social status defines the places he visits (check-ins) and the sentiment of those users, (extracted from messages associated with check-ins) when interacting with the places they checked into. It was discovered that most people check-in at places with higher frequency initially and the frequency reduces as time passes. One explanation for this could be that initially the user may be excited to check-in at the same place every day but with time, may find this monotonous. In regard to social status, the researchers studied areas with high net income residents and compared the results with areas with low net income residents. The researchers found that people in a high-income bracket take more long-distance trips when it comes to traveling. The explanation for this could be that affluent people have more resources with which to travel. Another noteworthy observation was that people living in dense areas such as New York City took more trips, but the length of these trips was small compared to trips made by people who lived in the countryside and less busy cities. The author believes the reason for this is that people living in the countryside do not usually have access to a lot of places to go for entertainment and hence they do not go out much. However, when they do, they have to travel a long distance which is the opposite of what people living in dense neighbourhoods, such as New York, LA, and San Francisco do. The researchers found patterns of check-ins over a day and over a week. 9 am, 12 pm and 6 pm are the busiest times when it comes to people checking in on weekdays however the pattern is a little different on weekends when the evening is the busiest time. The number of check-ins increases as the week progresses from Monday to Friday with Friday evening being the busiest. The study also compared check-in patterns of three different cities – New York, Los Angeles, and Amsterdam and showed that people in Amsterdam are early starters when it comes to check-ins. Cranshaw et al. [69] explored mobility patterns in Pittsburgh and also validated the results using a qualitative approach in which they interviewed 27 Pittsburgh residents to see how their perceptions of the city projected onto the research findings there. The results for both the qualitative and quantitative study were very similar, again validating the significance of using LBSN data for research.

All of the aforementioned research projects use some sort of social media data to analyze and explore various human-behaviour-related activities and traits. A lot of these studies have been conducted to explore mobility patterns in different cities around the world [58, 59, 70]. However, none of these studies considers weather as a moderating variable that can impact behaviour patterns. Weather can influence how people interact with and express emotions about the places they visit [71]. To the best of our knowledge, this research project is the first to explore such mobility and behaviour patterns in

Melbourne using location-based social network data. In this study, we have two hypotheses:

- a) People feel and write a lot more positive things when going out and visiting places in summer compared to winter.
- b) A lot of people in Melbourne use languages other than English to express their feelings when checking into different public places.

3. Methodology

3.1. Data

The data source for this research project is check-in data from the Swarm app (formerly known as Foursquare). The mobile app allows users to check-in when they are at a particular venue. Once checked-in, the details of the check-in and location are shared with the user's contact. Check-in data is classified as large-scale location-based data and mainly contains text along with the time and date. Typical check-in data contains a user's name, a message along with the check-in time, day, date, venue category, latitude and longitude of the location and gender category. Most users of Swarm check-in at a venue using a mobile app, but some also share their check-in data via Twitter using an in-built feature in the Swarmapp. The check-in then is broadcast to the user's followers on the Twitter platform. Tweets containing Swarm check-ins are the data source for this project. Data for this study were collected manually unlike most other similar studies that use Twitter API or some sort of software tool. Every tweet containing a Swarmapp check-in was copied directly from Twitter and was pasted into an Excel spreadsheet for further processing. The limitation of using a Twitter API, web scrapping or software tools to collect tweets is that one can only go back a week in time to collect data. Twitter does not allow users on a free subscription model to access data that is more than one week old. A paid version of the service, called Firehose, allows access to any amount of historic data for a price. Almost all software tools that are available on the market to collect Twitter data have a free subscription API working in the background and hence are dictated by the same Twitter API protocol. This study needs data from the summer and winter period of 2017 and to collect data freely and without any cost, manual data collection was the only option. Another limitation of using web scraping, Twitter API or a software tool to collect data from Twitter is that Twitter only allows a random 1% of its data to be collected with these methods, without paying any subscription and access fee. Manually, we were able to collect all the check-in tweets that were created in the Melbourne metropolitan area and shared on the Twitter platform during the specified time periods. Two 30-day time periods were selected, one from summer 2016-2017 (summer in Melbourne is from December to February) and the other from winter 2017 (winter in Melbourne is

from June to August) The winter data was collected from 1st June 2017 to 30th June 2017 and the summer data was collected from 15th Jan 2017 to 13th Feb 2017.

To collect the Swarm check-in data on Twitter, Twitter's advanced search tool was used to select the check-ins from the Melbourne area for the selected dates. A total of 3332 check-ins were collected from both the winter (1499) and the summer (1833) months. Swarm assigns every check-in to a venue category. In Swarm, there are ten top-level categories, and these are further divided into second-level categories. Second-level categories are further divided into third-level categories and so on. Considering the nature and scope of our study, we developed our own venue categories. We increased the size of the top-level categories to twenty to cover the breadth and reduced the depth from 4 to 2. This was done for simplicity and to streamline the venue categories. All the check-ins that we collected belonged to a second level and first-level category. First-level categories are airports, banks, bars, educational institutes, entertainment facilities, food places, grocery outlets, gyms, yoga studios, aerobic facilities, spas, saunas, home, hotels, landmarks, medical centres, neighbourhoods, outdoors, public transport, religious places, salons, shopping centres, sports venues, workplaces. Each of the venue categories is assigned a number. For example, 1 represents a check-in at an airport, 2 represents a check-in at a bank, and 3 represents a check-in at a bar and so on. Similarly, each day of the week is also assigned a number ranging from 1 to 7. For gender classification, males are assigned a value of 1, females 2 and couples 3. Days from both thirty-day periods are assigned values from 1 to 30 respectively. Transforming time data was a little difficult because it was in a 12-hour format (am, pm) when first collected. So, the first step was to convert it to a 24-hour format (e.g. 13:30:15). Once this was done, we divide the 24-hour day into four parts - morning (5 am – 12 pm), afternoon (12 pm – 5 pm), evening (5 pm – 9 pm), and night (9 pm - 5 am). Once this was done, each time slot was allocated a number from 1 to 4. 1 represented a check-in logged in the morning, 2 represented afternoons and so on.

3.2. Data Analysis

In this study, we perform three types of data analyses:

3.2.1 Text Analyses

A lot of people also like to express their feelings and thoughts when checking-in at different venues. They usually write a small note expressing their sentiments, emotions or any other thoughts they might have at the time of the check-in. An example of a short message is:

“ Laura @ViataDulce This mall has changed so much. Starbucks and Grilld for dinner. #Eastland (@ Eastland Shopping Centre)”

These messages can offer meaningful insight into a user's behaviour and emotions. There are different techniques that can be used to analyze these text messages. In this study, we use MeaningCloud, a machine learning-based software system that enables text analytics and the semantic processing of text data [72]. The tool uses different pre-trained machine learning models for different types of analysis. It is a commercial grade software system used both in industry and academia. We utilized the free subscription available to researchers and academics.

3.2.1.1 Language Identification with MeaningCloud

This feature identifies the language in which a document is written. However, in our case, we had different tweets related to different check-ins. We fed the excel column containing all the tweet data into the software. Analyses were done on both the summer and winter data by combining them together in one data set. The software traversed through each cell and identified the language in which the text was written.

3.2.1.2 Topic Extraction

MeaningCloud's text extraction feature identifies the main theme/topic of a text. The software goes through every word in a text and analyses the semantics, related syntactic structures, phrases, clauses, and paragraphs. The software tries to establish a relationship between different words in a text and the message they convey. The next step in this analysis is to establish the meaning of every sentence relative to the whole text and identify a theme. In our case, since our text data is in different spreadsheet cells, the software traversed through every cell to identify the main theme of that cell.

3.2.1.3 Text Clustering

Text clustering forms a cluster of the most common words in a document. The MeaningCloud software identifies all the words and counts the number of times they appear in a document. These words are then listed in ascending order. The most common word used in a document has the highest count and the least used has the lowest. We performed text clustering on our data from both the summer and winter periods separately to see which words are the most commonly used in both seasons and to determine what people are talking about, depending on the weather at the time.

3.2.1.4 Sentiment Analysis

Humans are emotional creatures and like to express their emotions when visiting different places. Sentiment analysis was undertaken to discover the emotions people express for different places in Melbourne. This experiment was also conducted to gauge the percentage of people who express some sort of emotion when checking in. We analyzed the summer and winter data separately to observe the difference in the way people express emotions, depending on the weather. The software analyzed each tweet and assigned to it one of the

following sentiments: Strong Negative, Negative, Neutral, Positive and Strong Positive. These were then aggregated to analyze all the emotions expressed.

3.2.2 Psychometric Analysis

We performed a psychometric analysis of our check-in data to discover the emotions that are not covered by the five sentiments discussed in the previous section. We used the Linguistic Inquiry and Word Count (LIWC2015) tool to perform these analyses. This tool enables computerized text analyses to reveal thoughts, feelings, personalities, and motivations [63]. The everyday words that people use can provide rich information about their fears, anxiety, anger, and beliefs [73]. This tool can be used to study various cognitive, emotional and structural components present in an individual's written samples and verbal speech. The software consists of a dictionary with more than 6400 words, word stems, and emoticons. These words are associated with one or more word categories and sub-dictionaries. An example of the dictionary structure is that the word 'cry' is associated with five different categories: negative emotions, sadness, overall affect, past focus and verbs. All sadness words belong to the broader 'negative emotions' and 'overall effect words' categories and so on. For each text, the software outputs 90 different variables. This includes word count, language summary, general description, linguistic dimensions, psychological construct, word categories, punctuation categories, and language markers, etc. [63]. We analyzed our check-in data for three psychometric properties: anxiety, anger, and sadness. Considering the scope of this study, these three psychometric analyses are the most relevant.

3.2.3 Statistical Analysis

Descriptive statistical analyses were performed on the data set to explore the mobility and behaviour patterns. Analyses were undertaken to explore how people move around in Melbourne based on the time of day and the day of the week. Different venue categories were explored to see which were the busiest in different time frames. Behaviour and mobility patterns based on gender were also analysed. The analyses considered different times of the day and different days of a typical week to gauge the difference in mobility patterns between male and female users. The results are discussed in the next section

4. Result and Discussions

4.1. Text Analysis

4.1.1 Language Analysis

Language analysis was undertaken to determine whether the users wrote messages associated with check-ins in a language other than English. Our hypothesis is that even though Melbourne is mainly an English-speaking city, there are a lot of people who write comments in different

languages. The results show that our null hypothesis is true and the use of languages other than English is quite prominent in Melbourne. Arabic, Chinese, Japanese and Korean are some of the other languages used extensively in the comments. It is not clear whether the people posting messages in these languages are natives of Melbourne or travelling from other places. Melbourne is a multicultural city with people from many different countries who have settled here. Therefore, it is highly possible that some of the non-English tweets were posted by people who live here permanently.

4.1.2 Topic Extraction

After conducting topic extraction on the comments posted, we identified some of the most common topics related to check-ins and the category to which they belong. We found Melbourne was the word most used in all the check-in tweets followed by Tullamarine (Airport), shopping, train and so on. A lot of check-ins did not contain extra comments and were generic check-ins such as I'm at _ _ _ _ .

Table 1. Topic Extraction

Trending	Category
Melbourne	Neighborhood
Tullamarine	Airport
Shopping	Shopping
Train	Public Transport
Drink	Bar
Dinner	Food
Game	Sports
Lunch	Food
MCG	Sports
Breakfast	Food

Since the data was collected in the Melbourne metropolitan area, it is not surprising to find the city name as the topic of discussion in a lot of check-ins. The trending words presented in Table-1 are in line with some of the most popular venue categories in which people check-in in Melbourne. People in Melbourne love their sports, which is why it is known as the sports capital of Australia. This explains why the words *game* and *MCG* (*Melbourne Cricket Ground*) are in the top ten topics of discussion. Both summer and winter are a busy time for sports in Melbourne. Cricket and tennis tournaments cover most of summer and footy (AFL) covers most of autumn and winter (March-Sep). Apart from sports, Melbourne is also one of the best places in the world for

different kinds of food and drink. There are many places people can visit to have a meal or a drink and enjoy the varied cuisines that Melbourne has to offer. The café and bar culture in Melbourne is world-renowned. This may be one of the reasons that the topics *'drink'* and *'lunch'* are among the top ten topics discussed.

4.1.3 Text Clustering

Table 2 shows the top ten words used in the check-in tweets in Melbourne during the summer 2016-17 and the winter 2017 periods. This analysis was conducted on the summer and winter data separately to see if there are any differences in the words used to imply the places visited. Vic, Melbourne, and Victoria are the top three words used in all tweets.

Table 2. Text Clustering

Text Clustering	
<u>Summer</u>	<u>Winter</u>
Vic	Vic
Melbourne	Melbourne
Victoria	Victoria
Tullamarine	Tullamarine
South	Station
Station	Shopping Centre
Shopping Centre	South
Outdoor	East
Australia	Epping
Coffee	Southbank

The explanation for this could be that every check-in has a locality address at the end. Since all the tweets were collected from Melbourne, they all contain at least two of these three words. It can be seen that the words are quite similar and almost in the same order in both the winter and summer periods, implying that there is no significant difference in the type of places in which people check-in, and talk about, during summer and winter periods. As shown in Table 2, the word *outdoor* was used a lot more in summer compared to winter. Warm weather brings a lot more people out to be involved in outdoor activities, so it was no surprise to see this word is used often in summer and not as much during winter. The results also show the word *coffee* is used more numerous in the summer tweets, indicating that more people go out for coffee in warmer weather than in cooler weather. *Shopping centre* was among the top ten words for both seasons. This may imply that shopping is one of the main activities in which most people engage for pleasure and fun. Melbourne is

home to many shopping centres including the world-famous Chadstone Shopping Centre.

4.1.4 Sentiment Analysis

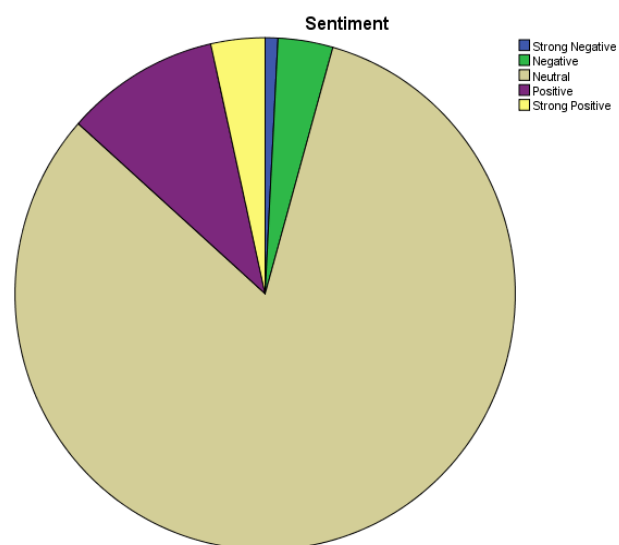


Figure 1. Sentiment Analyses for summer.

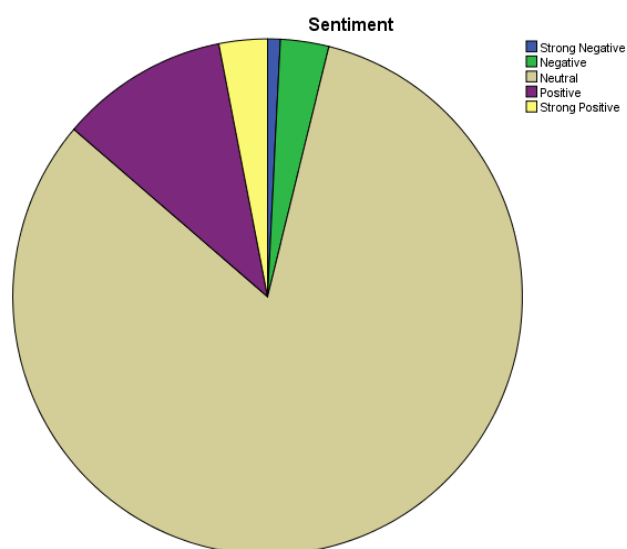


Figure 2. Sentiment Analyses for winter.

Sentiment analyses were performed on the summer and winter data set separately to see how people express their emotions when visiting places in these two seasons, and whether these emotions are impacted by weather. We found that the sentiments expressed during both kinds of weather were quite similar as opposed to our hypothesis that more people write positive things about the places they visit during summer compared to winter. Some of the literature in behavioural science suggests that people are

happier in summer than in winter [60, 74], but this was not obvious in the results from the summer data sets. Figure 1 and Figure 2 show the results of the sentiment analyses for summer 2016-17 and winter 2017 respectively. Each colour represents a different sentiment expressed during the two seasons. The results of our analysis show that a lot of check-in tweets are neutral in nature and this means most users in Melbourne have not expressed strong emotions in relation to the places they have visited. As can be seen from the pie charts in Figures 1 and 2, the percentage of neutral tweets in summer was 82.2%, which is almost identical to the percentage (82.4%) in winter. In summer, 0.8% of the tweets were strong negative, which is exactly the same as the percentage in winter. Similarly, the percentages for negative, positive and strong positive tweets in summer are 3.5%, 9.9% and 3.5% and in winter are 3.1%, 10.7% and 3.1%, respectively. There is no significant difference in the results for summer and winter. People express similar sorts of sentiments about the places they visit in both kinds of weather. Roughly 18% of check-in tweets express either a negative or positive sentiment related to the places they have visited. Almost all the negative tweets express a complaint about the places and almost all the positive tweets express happiness and excitement after checking-in to venues. The results also show that in both seasons, there are more positive tweets than negative tweets. This may imply that most people feel happy and excited when they check-in at venues. A small number of negative sentiments may be the result of bad service or a negative experience that the users encountered during their visit to a particular venue. Since the number of negative tweets is smaller than the number of positive tweets, this shows that most people are neutral or happy rather than unhappy while visiting places or going out. The weather does not seem to have much of an effect on the user's sentiments towards the places they visit.

4.2. Psychometric Analysis

Psychometric analysis was conducted on the check-in tweets to see whether people were anxious, angry or sad when visiting places. Table 3 shows some examples of messages from our dataset and the psychometric properties they exhibit. A cell numbered 1 next to a message represents the presence of the related psychometric property and 0 indicates its absence. We analyzed data for summer and winter separately to determine whether there are any behaviour differences due to weather, with the results shown in Table 4.

Table 3. Psychometric Analyses 1

Message with Check-in	Anxious	Anger	Sad
Panic!At The Disco (@The Disco in West Melbourne, VIC)	1	0	0
Sam ??Where are you? (@ Tallboy and Moose in Preston, VIC)	1	0	0
Ahhh I'd forgotten how rude the staff are here. Welcome back Robstar!	0	1	0
Where every part of the order is wrong. Why? This place generallily sucks (@ The Coffee Club in Airport West, VIC)	0	1	0
olives here suck. maybe they do better gibson martinis? ??(@ Cookie in Melbourne, VIC)	0	1	0
Missed the train by just a few seconds :) WHY IS THERE EVEN AN EXTRA PLATFORM HERE ??(@ Surrey Hills Station)	0	0	1
Last time I used public transport was in Greece. Gosh I miss it	0	0	1
im gonna miss this place ??(@ Grill'd in Cheltenham, VIC)	0	0	1

We found that the results are quite similar regardless of the weather. Again, this is contrary to the common belief that people are happier and feel more positive in summer than in winter. As discussed in the previous section, a lot of users do not write/share any comments when checking-in. Such check-ins are classified as neutral. As can be seen from the results, most messages in both the summer and winter periods are neutral. A few messages express anxious, angry or sad emotions; however, the number is very small at 1%for both summer and winter. There are at least two explanations for such a low number of check-ins associated with these words. Firstly, users are generally in a good mood when visiting places and are looking forward to having fun, which is why they are not anxious, angry or sad. Another explanation is that even if some users are anxious, angry or sad when they are at a particular venue, they probably like to keep their emotions to themselves and do not want to broadcast these feelings.

Table 4. Psychometric Analyses 2

	(Summer)		(Winter)	
Anxiety				
	Frequency	Percent	Frequency	Percent
Neutral	1829	99.8	1497	99.9
Anxious	3	0.2	2	0.1
Total	1832	100.0	1499	100.0
Anger				
	Frequency	Percent	Frequency	Percent
Neutral	1813	99.0	1492	99.5
Anger	19	1.0	7	0.5
Total	1832	100.0	1499	100.0
Sad				
	Frequency	Percent	Frequency	Percent
Neutral	1816	99.1	1485	99.1
Sad	16	0.9	14	0.9
Total	1832	100.0	1499	100.0

4.3. Statistical Analysis

4.3.1 Spatio-Temporal User Activity Patterns

this subsection, we present the results of our statistical analysis for spatio-temporal activity patterns. Our results suggest that activities in Melbourne differ over the course of the day and also of a week. We identified several meaningful patterns that are closely related to human activity from a spatial and temporal point of view [7]. Figure 3 shows how the activities in the city increase as the day passes. More people are active and moving around in the afternoon than in the morning and this activity increases in the evening before it starts slowing down after 9 pm. The evening period between 4 pm -9 pm is the busiest time in the city with 36.3% of check-ins occurring at this time, followed by 28% in the afternoon, 24.3% in the morning and only 11.4% at night. An explanation for the evening being the busiest time of the day is that most people finish work during the 4 pm -9 pm time frame and visit places such as train stations, restaurants, bars, sports centres, movies, and the airport, etc.

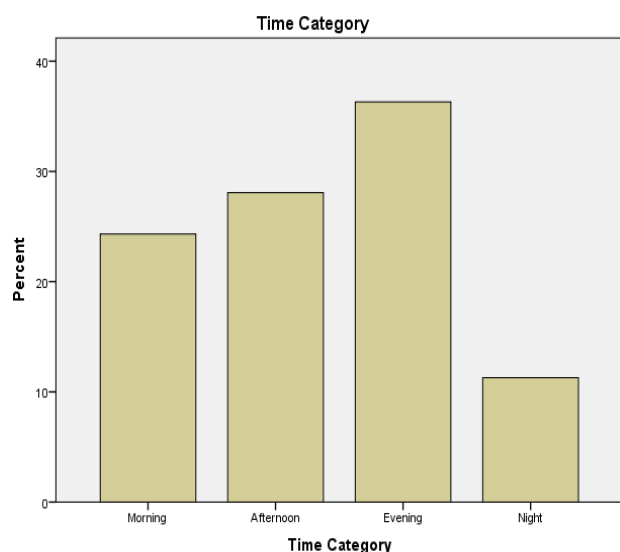


Figure 3. Spatio-temporal User Activity Patterns during a Day

Most people sleep at night so there is not much activity in the city. This might not have been the case if we had collected data from a city such as New York, where people are active at almost any time of the day. Table 5 shows the activity during the day based on different categories of venues in the city. We can see the food category has the highest number of check-ins in the evening which is when a lot of people go out for a meal. In addition to restaurants, a lot of people also visit bars and pubs for after-hour drinks. Table 5 shows some activity during the night and most of this is at Tullamarine Airport, late nightclubs and restaurants. 12.5 % of these check-ins at night are at home or at a hotel. These people are not really moving around in the city but are at home. The check-ins at hotels and homes are highest at night compared to other time frames, indicating people are returning home after finishing their day. Most check-ins in the morning are at café's, the airport and work, with cafés being the busiest. Twenty-four of the total two hundred check-ins at work are at night, indicating that around 12% of the total workforce undertake some sort of night work/shift. This sort of work is most likely to be in places such as hospitals, nursing homes, hotels, nightclubs, etc. Gym check-ins are highest in the evening and morning, showing that people prefer to exercise in the evening and morning even though most gyms in Melbourne are open 24 hours a day.

The Table-5 also shows that around 8.3 % of gym-goers prefer to exercise late at night. Most people visit banks or ATMs in the afternoon. This resonates with people making quick visits to banks and ATMs during their lunch break while at work. Visits to the doctor, hospitals and grocery stores are steady during the day. Public transport is busy in the morning when most people go to work and in the evening when they return home.

Table 5. Spatio-Temporal User Activity Patterns 1

VENUE TYPE * Time Category						
		Time Category				Total
		Morning	Afternoon	Evening	Night	
VENUE TYPE	Airport	122	38	74	31	265
	Bank	2	7	4	1	14
	Bar	6	34	116	65	221
	Education	70	52	40	7	169
	Entertainment	12	45	88	23	168
	Food	219	354	426	93	1092
	Grocery	27	35	37	8	107
	Gym	22	5	28	5	60
	Home	5	14	35	23	77
	Hotel	16	6	12	24	58
	Landmark	10	25	23	11	69
	Medical	18	17	16	1	52
	Neighbourhood	13	10	13	16	52
	Outdoor	10	45	30	5	90
	Public	64	46	56	17	183
	Religious	3	5	4	1	13
	Salon	1	7	10	0	18
	Shopping	47	100	73	8	228
	Sports	45	41	95	13	194
	Work	98	49	29	24	200
Total		810	935	1209	376	3330

The airport is busiest in the morning with 46% of check-ins taking place during this time. The activity indicates that most flights are scheduled early in the morning and in the evening. Similar to activity variations during the day, activities also vary during the week. Figure 4 shows how mobility and activity increase as the week progresses. The number of check-ins is highest on Sundays followed by Saturdays and Fridays. Most people do not work on weekends which enables them to spend time doing what they like the most, such as watching/playing sports, going out for meals, going to the movies, shopping, etc. which is why weekends are busier than weekdays as people have more time to go out.

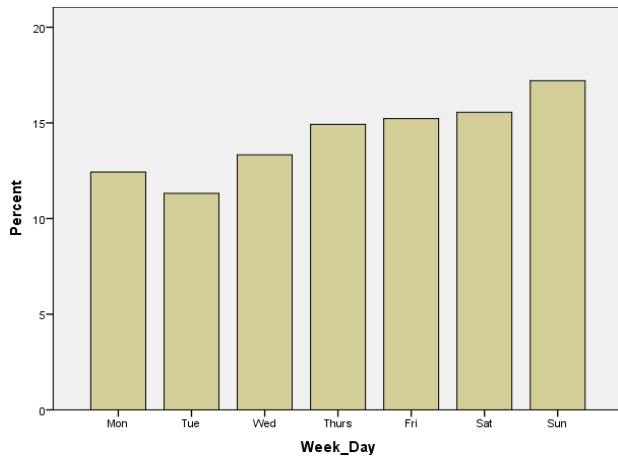


Figure 4. Spatio-Temporal User Activity Patterns during a Week

Tuesdays are the quietest days, followed by Mondays, and as the week progresses, people start moving around. One explanation for this could be that at the start of the week, most people focus more on their work and as the week progresses, they feel more relaxed and start going out for different activities. Table 6 shows a breakdown of the temporal patterns based on different venue categories during a week. The findings also show that as the week progresses, people indulge more in leisure activities. It can be seen from the table that check-ins are highest at venues such as bars, places of entertainment, food places, outdoors, shopping centres, and sports centres on weekends. In contrast, activities at work are slowest on the weekends. Airport activities are similar throughout the week with Wednesdays and Sundays being a little busier than other days. In the education category, which includes colleges, universities, schools, and libraries, there is much less activity on weekends compared to weekdays. These findings are in line with the previous literature which shows most students like to work hard on weekdays and party hard on weekends. Bars & pubs become busier from Fridays onwards with Friday being the busiest day. A lot of people like to end their stressful week with a drink or catching up with family and friends. Sunday is the busiest day for grocery shopping with all the other six days showing similar sorts of activity. One would imagine that Saturdays would be as busy as Sundays, however, this does not seem to be the case.

As indicated by the results in Table 6, the reason for this is that activities associated with leisure such as sports, going out for meals, entertainment, and bars have higher check-ins on Saturdays than Sundays. This shows that most people like to enjoy, relax and have fun on Saturdays and leave grocery shopping, which is a more mundane task than a leisure activity, for Sundays. Apart from grocery shopping, Sunday is also a preferable day to address other shopping needs as this day has the highest check-ins in shopping centres. Similar to shopping,

Sundays also seem to be a popular day for visiting religious venues such as churches.

Table 6. Spatio-Temporal User Activity Patterns 2

	VENUE TYPE * Week_Day							Total
	Week_Day							
	Mon	Tue	Wed	Thurs	Fri	Sat	Sun	
Bank	3	2	2	0	3	3	1	14
Bar	17	16	16	28	53	46	45	221
Education	29	27	37	31	26	9	10	169
Entertainment	18	17	13	21	27	36	36	168
Food	132	133	149	153	160	193	172	1092
Grocery	13	7	14	14	11	13	35	107
Gym	6	12	13	8	10	4	7	60
Home	13	12	8	11	12	9	12	77
Hotel	8	6	10	8	11	4	11	58
Landmark	9	5	15	11	9	8	12	69
Medical	8	8	10	6	11	5	4	52
Neighbourhood	1	5	4	8	14	7	13	52
Outdoor	17	3	5	8	7	21	29	90
Public Transport	29	15	28	37	22	17	35	183
Religious Venue	1	2	0	0	1	0	9	13
Salon	0	1	3	6	3	3	2	18
Shopping Centre	28	32	20	37	33	35	43	228
Sports	18	23	17	36	18	47	35	194
Work	29	26	31	37	38	25	14	200
Total	414	377	444	497	507	518	573	3330

4.3.2 Gender-based Activity Analysis

Based on the results shown in Figure 5, males represent 58% of all check-ins and females represent only 40.7% of all check-ins. The results show that males are almost 1.5 times more likely to go out and check-in at one of the 20 venue categories. Figure 6 shows the gender-based activity for the 20 venue categories. In 16 of the 20 categories, more males than females visit a place and log a check-in. The figure also shows an interesting finding in that 4 of 5 check-ins made in bars are by males. On the contrary, 4 in 5 check-ins made in shopping centres are by females [7]. One would imagine that men need to go to shopping centres to buy clothes, shoes, etc. so why is their check-in percentage so low at shopping centres compared to women’s? Similarly, one can argue that the percentage of women going to a bar or a pub for a drink has to be

more than 20%. Why do we have such differing results from our dataset?

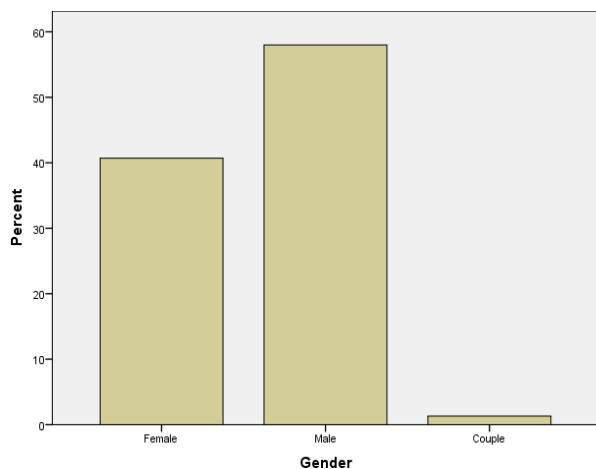


Figure 5. Gender-based Activity Analysis

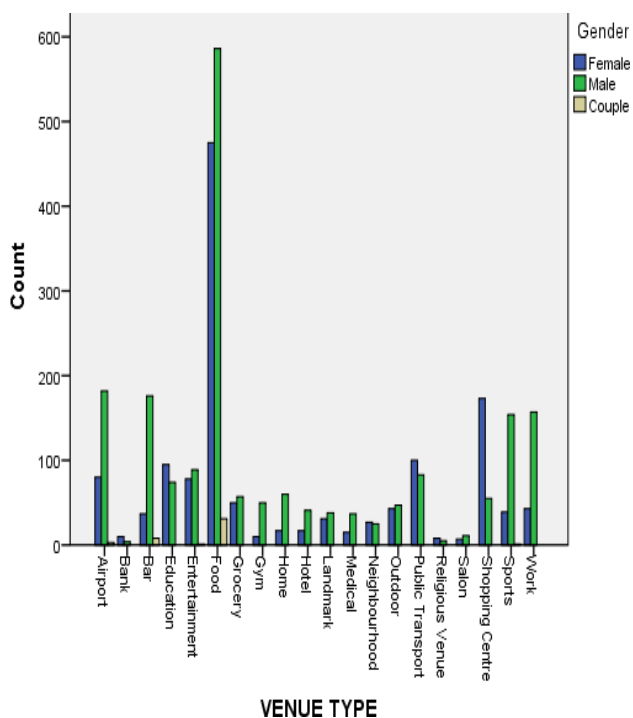


Figure 6. Gender and Venue-based Activity Analyses

Due to our societal norms, a female rather than a male is associated with household chores which include grocery and other types of shopping. The low percentage of male check-ins in shopping centres may be because males do not want to be associated with these chores and may not want to broadcast this when they are involved in such activities. They may feel it is a better image to be

associated with a drink hence they like to broadcast their check-ins every time they are in a bar having a drink with a friend. Because it is more socially acceptable for a man to have a drink, they are more likely to check-in to a bar compared to a shopping centre. For the same societal norms, most women do not like to be associated with alcohol. This may also be due to family pressure and norms. So even though the percentage of women going out for a drink is most likely higher than 20%, the reason for the low number of check-ins could be because most women do not like to let others know when they are having a drink. Similarly, the percentage of males going shopping is most likely higher than 20%, however, the reason for the low number of check-ins could be because most men prefer not to broadcast the fact that they are engaged in such an activity. Figure 7 shows the gender-based activity on different days of a week. It can be seen that male activity is in line with the overall activity in Melbourne i.e. it starts slowly at the beginning of the week and increases as the week progresses. Saturdays and Sundays are the two busiest days of the week. However, for females, the activities do not increase as much as the week progresses and remains very similar throughout the week but increases slightly on Sundays. The difference in these patterns can be explained by the report published by the Australian Government’s Workplace Gender Equality Agency in 2016 [75] which states that only 36.7% of all full-time employees are women compared to 63.3% of full-time employees being male, which is almost double the number of female employees. Since most people in Australia work from Monday – Friday, most men are at work on these days and when the working week finishes on Friday, their mobility activities increase. As for women, since a large number of them either do not work or work only part-time or casually, they are more flexible on weekdays to go out and visit places.

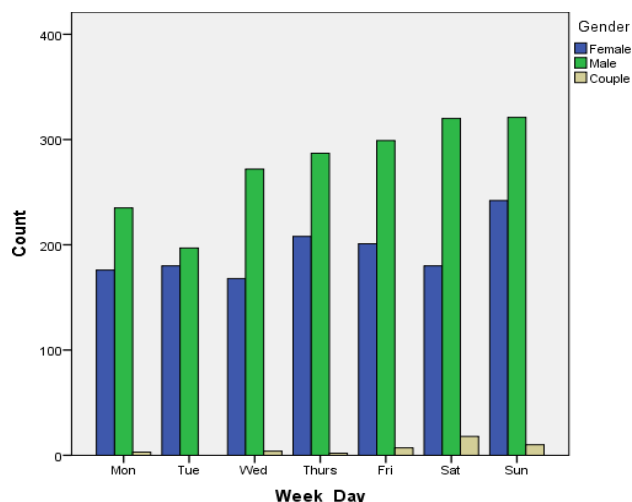


Figure 7. Gender and Weekday-based Activity Analysis

Figure-7 also shows that female activity is high on Thursdays and Fridays. Since 4 in 5 shopping centre check-ins are by females and shopping centres are open late on these two days in Melbourne, this may imply that more women check-in at shopping centres on Thursdays and Fridays. Sunday has the highest activity when it comes to grocery shopping and it also has the highest number of check-ins by females. This may indicate that most grocery check-ins on Sundays are logged by females, implying that females do most of the grocery shopping in their respective households. Figure 8 shows the distribution of activities by male and female users based on different time slots in a day.

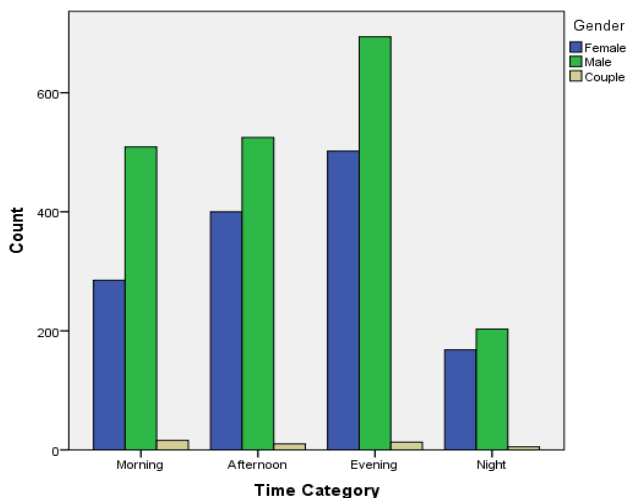


Figure 8. Gender and Time of the Day-based Activity Analysis

These findings reinforce the previous discussion on male and female activity patterns. The results show that male activities are low in number and very similar in the mornings and afternoons because this is when most people, especially males (with 63.3% of all full-time jobs being held by males) are at work. Male activity increases when they finish work and visit places and activity eventually decreases at night. On the other hand, female activity increases during the day and eventually decreases at night in line with the overall activities of both genders. As previously discussed, fewer females participate in full-time employment which gives them more flexibility to visit places, even during the afternoon, when most men are at work. So, the findings show how male and female activities vary not only during the day but also during the week.

4.3.3 Geo-Temporal Patterns

In this section, we discuss the results of our geo-temporal analysis. Figure 10 shows the popular places that people visit and the check-ins at different times of the day. It can be seen from the number of check-ins that food- and drink-related places are the most popular, followed by

shopping centres, the airport, work, sports centres, and public transport. Venues such as religious places, banks, and salons are not very popular places for check-ins. This may imply that when visiting these places, some people choose not to broadcast their presence there as it may not be considered to be as fashionable or exciting compared to places like bars, shopping centres or restaurants. There may be many reasons and explanations for such behaviour. In regard to banks, safety and security may be the main cause. A salon is more of a private place and people may not like others to know that they are having some sort of treatment to make them look better. According to Foursquare [24], most of its users are middle-aged individuals who may not feel that it is appropriate to be associated with religion, which may explain why a lot of people do not check-in when they visit these types of venues [7]. Activity in each venue category varies depending on the time of day. The green and light brown bars represent afternoon and evening activities, showing that these two times are quite busy for almost all venue categories except work, public transport, and the airport, which are busy in the morning. Most people go to work in the morning and check-in when they get there and most people in Melbourne use public transport to commute to work. This explains why the mornings are busy for these two types of venue categories.

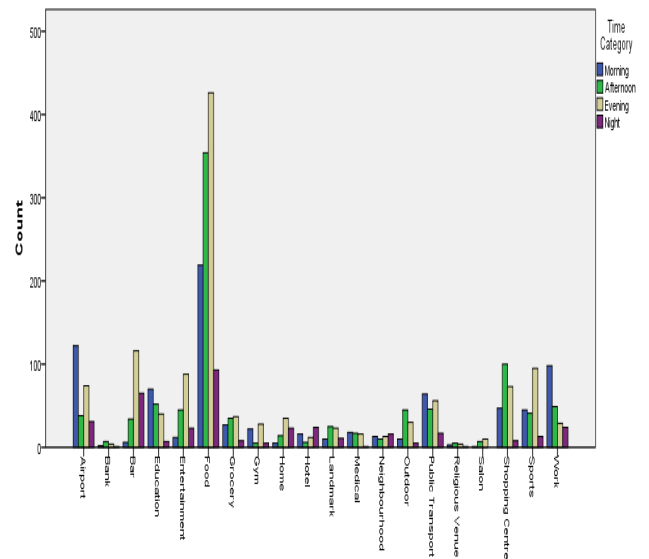


Figure 9. Geo-Temporal Patterns Analysis

The high number of check-ins at the airport indicates that most airlines organize their flights early in the morning. Regardless to whether these are incoming or outgoing flights, the airport is busy in the morning. The second busiest time at the airport is evenings. This may be due to domestic flights because most people who fly into Melbourne for work in the mornings fly home in the evenings. Apart from work, public transport and the

airport, other categories that are also somewhat busy in the morning are education, food, gyms, and hotels. Most educational institutes are busy in the mornings and afternoons. Busy food places mean a lot of people are buying coffees and breakfasts. As previously discussed, early morning and evening are the two most popular times for gym-goers, as indicated in the figure.

4.3.4 Check-in Dynamics

This section presents the results of the check-in (mobility activity) dynamics analysis. We studied all the check-ins as a whole representing human mobility patterns within Melbourne at different times of the day. Figure-10 shows some of the findings. The results indicate that most of the check-ins that represent mobility patterns take place during the evenings. Evenings are busier than any other time period every day except on Sundays. On Sundays, afternoon activities are highest. An explanation for this could be that since Monday is the first working day after the weekend, most people tend to take it easy and stay home on Sunday evenings and in some cases, prepare for work on Monday. Most people also prefer to stay home and catch up on their favourite movies on Sunday evenings. However, afternoons on Sundays are the busiest leading to relatively quieter evenings. From the graph, we can also see that night-time activities are the highest on Fridays followed by Saturdays.

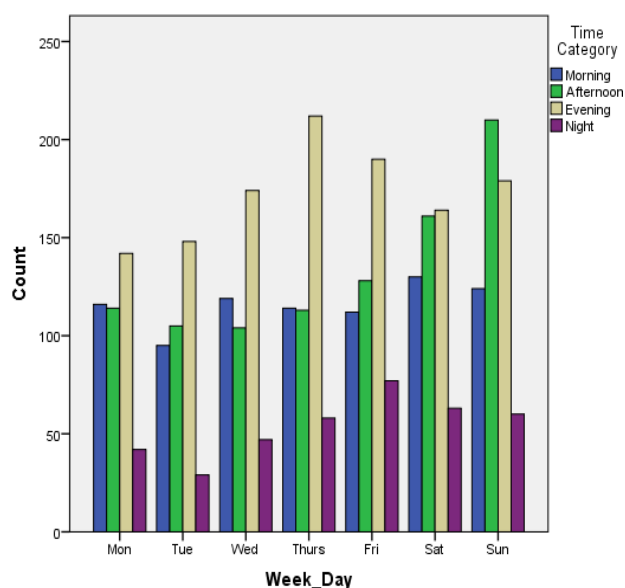


Figure 10. Temporal Patterns Analysis

Most people are in a party mood on the weekend and they start their activities on Friday nights and continue this on Saturday nights. Activity reduces on Sunday nights. The results show that activity at night is low at the start of a week and increases as the week progresses with the exception of Monday nights. This may be explained

by the fact that a lot of young people who work on weekends go out on Monday nights. Morning activity is stable throughout the week. So even though most people do not go to work on weekend mornings, they are busy checking-in at places such as café's, gyms, sports centres, the airport, shopping centres, and outdoors. The results also show a positive correlation between the number of afternoon activities and the day of the week. The number of afternoon activities increases as the week progresses. Overall, afternoon activity is higher on weekends compared to weekdays as most people are at work on weekday afternoons [7].

4.3.5 Research Implications

The research has some real-world implications. This can be considered as a pilot study for a possible large-scale research project. In this study, we collected data manually as it was a free option and we only wanted data from one city, however, for large-scale studies, similar data can be collected using Twitter's Firehose and Swarm's paid data collection services that allow organisations to collect any amount of historic and current data for a fee. Since collecting data in this way is far more convenient and faster than the traditional survey and questionnaire type of data collection, techniques that are generally used for such behavioural studies, the ability to conduct large-scale human behaviour studies can give organisations and governments the insight they need to provide efficient services to their clients and citizens. One such organisation, Dataminr [76], has recently started doing something similar. It captures the latest trends and patterns from various public datasets such as Twitter, Facebook, Reddit, etc. and provides knowledge extracts to organisations and governments alike. Some of the commercial organisations that work with Dataminr and use its services are news outlets and investment banks. Governments around the world would benefit from learning about their citizen's behaviour patterns when interacting with public infrastructure. Data can be collected for a longer period and from different cities within a country to gauge the different behaviour and psychological patterns of the nation's citizens and the information can then be utilized to plan for public services and future infrastructure. The government of the United States of America utilized the information and knowledge collected from such public data sets for national security purposes. Palantir [77] is the company that is contracted by the government to perform this task. Since such data sets and techniques are already being used by governments for their national security and counter-terrorism efforts, these can also be utilized to conduct large-scale mobility and behaviour studies, as proposed in this research paper.

5. Conclusion

In this paper, we performed sentiment analysis, psychometric analysis and various descriptive statistical

analyses on location-sharing check-in data from Swarm to explore human mobility and behaviour patterns in Melbourne. Even though location-based data has been used for various studies in the past, its use is relatively new in social & behavioural science studies. Large-scale data handling techniques along with machine learning analysis methods have made these studies more feasible and viable using location-based social media data. In this study, we show the effectiveness of such a methodology for behaviour studies by producing meaningful results. The contributions of this work are: (1) the creation of a data set of location-based check-ins from Swarmapp; (2) a pilot study that proves the effectiveness of our proposed methodology; and (3) knowledge discovery about some of the mobility and behaviour patterns of people living in the Melbourne metropolitan area. The results show that restaurants, café's bars, shopping centres, gyms, and sports centres are some of the most visited places in the city and activity at these places increase as the week progresses. The sentiment analysis of messages associated with check-ins showed that most people do not express strong emotions in relation to the places they visit. Descriptive analysis revealed that mobility patterns change at different times of a day and a week. Night activity is slow in Melbourne at the start of the week and increases as the week progresses and is highest on Fridays. Activities outside of work were low in the mornings and afternoons and increased in the evenings for males but remains nearly the same throughout the day for females.

Even though the study is quite comprehensive, it has some limitations. The dataset is from one city and for a limited time span. We could have undertaken a lot more analyses if the sample size was larger and from different cities, enabling us to eliminate the possibility of bias. The results show some meaningful patterns that can lay the foundation for future work. Deeper analysis can be performed to discover the busy parts of the city at different times of a day or a week which can lead to the development of a predictive analytics framework on user behaviour and mobility activities. Check-in data can also be used to mine other human-psychological traits.

References

- [1] B. T. van Zanten, D. B. Van Berkel, R. K. Meentemeyer, J. W. Smith, K. F. Tieskens, and P. H. Verburg, "Continental-scale quantification of landscape values using social media data," *Proceedings of the National Academy of Sciences*, p. 201614158, 2016.
- [2] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási, "Limits of predictability in human mobility," *Science*, vol. 327, no. 5968, pp. 1018-1021, 2010.
- [3] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi, "Understanding individual human mobility patterns," *nature*, vol. 453, no. 7196, p. 779, 2008.
- [4] C. Song, T. Koren, P. Wang, and A.-L. Barabási, "Modelling the scaling properties of human mobility," *Nature Physics*, vol. 6, no. 10, p. 818, 2010.
- [5] G. B. Colombo, M. J. Chorley, M. J. Williams, S. M. Allen, and R. M. Whitaker, "You are where you eat: Foursquare checkins as indicators of human mobility and behaviour," in *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2012 IEEE International Conference on*, 2012: IEEE, pp. 217-222.
- [6] T. H. Silva, P. O. V. de Melo, J. M. Almeida, M. Musolesi, and A. A. Loureiro, "You are what you eat (and drink): Identifying cultural boundaries by analyzing food and drink habits in foursquare," in *Eighth International AAAI Conference on Weblogs and Social Media*, 2014.
- [7] R. Singh, Y. Zhang, and H. Wang, "Exploring Human Mobility Patterns in Melbourne Using Social Media Data," in *Australasian Database Conference*, 2018: Springer, pp. 328-335.
- [8] E. Bakshy, I. Rosenn, C. Marlow, and L. Adamic, "The role of social networks in information diffusion," in *Proceedings of the 21st international conference on World Wide Web*, 2012: ACM, pp. 519-528.
- [9] V. Mayer-Schönberger and K. Cukier, *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt, 2013.
- [10] S. Yin and O. Kaynak, "Big data for modern industry: challenges and trends [point of view]," *Proceedings of the IEEE*, vol. 103, no. 2, pp. 143-146, 2015.
- [11] H. Wang, Z. Xu, H. Fujita, and S. Liu, "Towards felicitous decision making: An overview on challenges and trends of Big Data," *Information Sciences*, vol. 367, pp. 747-765, 2016.
- [12] J. Lindqvist, J. Cranshaw, J. Wiese, J. Hong, and J. Zimmerman, "I'm the mayor of my house: examining why people use foursquare-a social-driven location sharing application," in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2011: ACM, pp. 2409-2418.
- [13] A. Whiting and D. Williams, "Why people use social media: a uses and gratifications approach," *Qualitative Market Research: An International Journal*, vol. 16, no. 4, pp. 362-369, 2013.
- [14] B. L. Fossen and D. A. Schweidel, "Social TV: How Social Media Activity Interacts With TV Advertising," *GfK Marketing Intelligence Review*, vol. 9, no. 2, pp. 31-36, 2017.
- [15] C. Oh and S. Yergeau, "Social capital, social media, and TV ratings," *International Journal of Business Information Systems*, vol. 24, no. 2, pp. 242-260, 2017.
- [16] D. Ruths and J. Pfeffer, "Social media for large studies of behavior," *Science*, vol. 346, no. 6213, pp. 1063-1064, 2014.
- [17] K. Weller and M. Strohmaier, "Social media in academia: How the social web is changing academic practice and becoming a new source for research data," *IT-Information Technology*, vol. 56, no. 5, pp. 203-206, 2014.
- [18] M. A. Zook and M. Graham, "Mapping DigiPlace: geocoded Internet data and the representation of place," *Environment and Planning B: Planning and Design*, vol. 34, no. 3, pp. 466-482, 2007.
- [19] C. C. Aggarwal, *Data mining : the textbook*. Cham : Springer, 2015., 2015.
- [20] J. Han, J. Pei, and M. Kamber, *Data mining: concepts and techniques*. Elsevier, 2011.

- [21] N. Japkowicz and J. Stefanowski, *Big Data Analysis: New Algorithms for a New Society*. Springer, 2016.
- [22] R. Kitchin, "Big Data, new epistemologies and paradigm shifts," *Big data & society*, vol. 1, no. 1, p. 2053951714528481, 2014.
- [23] J. Y. Tsai, P. Kelley, P. Drielsma, L. F. Cranor, J. Hong, and N. Sadeh, "Who's viewed you?: the impact of feedback in a mobile location-sharing application," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2009: ACM, pp. 2003-2012.
- [24] Foursquare. "Foursquare Labs, Inc, Location based social network." Foursquare. <https://enterprise.foursquare.com/> (accessed 12th June, 2017).
- [25] J. Frith, "Communicating Through Location: The Understood Meaning of the Foursquare Check-In," *Journal of Computer-Mediated Communication*, vol. 19, no. 4, pp. 890-905, 2014.
- [26] M.-A. Abbasi, S.-K. Chai, H. Liu, and K. Sagoo, "Real-world behavior analysis through a social media lens," in *International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction*, 2012: Springer, pp. 18-26.
- [27] R. M. Chang, R. J. Kauffman, and Y. Kwon, "Understanding the paradigm shift to computational social science in the presence of big data," *Decision Support Systems*, vol. 63, pp. 67-80, 2014.
- [28] R. Singh *et al.*, "A Framework for Early Detection of Antisocial Behavior on Twitter Using Natural Language Processing," in *Conference on Complex, Intelligent, and Software Intensive Systems*, 2019: Springer, pp. 484-495.
- [29] M. R. Islam, M. A. Kabir, A. Ahmed, A. R. M. Kamal, H. Wang, and A. Ulhaq, "Depression detection from social network data using machine learning techniques," *Health information science and systems*, vol. 6, no. 1, p. 8, 2018.
- [30] H. Yin, B. Cui, Z. Huang, W. Wang, X. Wu, and X. Zhou, "Joint modeling of users' interests and mobility patterns for point-of-interest recommendation," in *Proceedings of the 23rd ACM international conference on Multimedia*, 2015: ACM, pp. 819-822.
- [31] M. Sklar, B. Shaw, and A. Hogue, "Recommending interesting events in real-time with foursquare check-ins," in *Proceedings of the sixth ACM conference on Recommender systems*, 2012: ACM, pp. 311-312.
- [32] G. Preethi, P. V. Krishna, M. S. Obaidat, V. Saritha, and S. Yenduri, "Application of Deep Learning to Sentiment Analysis for recommender system on cloud," in *Computer, Information and Telecommunication Systems (CITS), 2017 International Conference on*, 2017: IEEE, pp. 93-97.
- [33] P. V. Krishna, S. Misra, D. Joshi, and M. S. Obaidat, "Learning automata based sentiment analysis for recommender system on cloud," in *Computer, Information and Telecommunication Systems (CITS), 2013 International Conference on*, 2013: IEEE, pp. 1-5.
- [34] M. Shepard and A. Greenfield, "Urban computing and its discontents," *New York: The Architectural League of New York, New York USA*, 2007.
- [35] H. Li, Y. Wang, H. Wang, and B. Zhou, "Multi-window based ensemble learning for classification of imbalanced streaming data," *World Wide Web*, vol. 20, no. 6, pp. 1507-1525, 2017.
- [36] A. Crooks *et al.*, "Crowdsourcing urban form and function," *International Journal of Geographical Information Science*, vol. 29, no. 5, pp. 720-741, 2015.
- [37] J. Du, S. Michalska, S. Subramani, H. Wang, and Y. Zhang, "Neural attention with character embeddings for hay fever detection from twitter," *Health information science and systems*, vol. 7, no. 1, p. 21, 2019.
- [38] J. Huang, M. Peng, H. Wang, J. Cao, W. Gao, and X. Zhang, "A probabilistic method for emerging topic tracking in microblog stream," *World Wide Web*, vol. 20, no. 2, pp. 325-350, 2017.
- [39] E. Currid and S. Williams, "The geography of buzz: art, culture and the social milieu in Los Angeles and New York," *Journal of Economic Geography*, vol. 10, no. 3, pp. 423-451, 2010.
- [40] L. Pei *et al.*, "Human behavior cognition using smartphone sensors," *Sensors*, vol. 13, no. 2, pp. 1402-1424, 2013.
- [41] M. Han, J. H. Bang, C. Nugent, S. McClean, and S. Lee, "A lightweight hierarchical activity recognition framework using smartphone sensors," *Sensors*, vol. 14, no. 9, pp. 16181-16195, 2014.
- [42] T. Huang, Y.-J. Gong, S. Kwong, H. Wang, and J. Zhang, "A niching memetic algorithm for multi-solution traveling salesman problem," *IEEE Transactions on Evolutionary Computation*, 2019.
- [43] J. Zhao, T. Wang, X. Xu, and Y. Yang, "Personalized LBSN Recommendation System," in *Proceedings of the 2017 International Conference on Management Engineering, Software Engineering and Service Sciences*, 2017: ACM, pp. 119-123.
- [44] J. Melià-Seguí, R. Zhang, E. Bart, B. Price, and O. Brdiczka, "Activity duration analysis for context-aware services using foursquare check-ins," in *Proceedings of the 2012 international workshop on Self-aware internet of things*, 2012: ACM, pp. 13-18.
- [45] Y. Sun, "Investigating "locality" of intra-urban spatial interactions in New York city using foursquare data," *ISPRS International Journal of Geo-Information*, vol. 5, no. 4, p. 43, 2016.
- [46] H. Jiang, R. Zhou, L. Zhang, H. Wang, and Y. Zhang, "Sentence level topic models for associated topics extraction," *World Wide Web*, vol. 22, no. 6, pp. 2545-2560, 2019.
- [47] J. Ma, L. Sun, H. Wang, Y. Zhang, and U. Aickelin, "Supervised anomaly detection in uncertain pseudoperiodic data streams," *ACM Transactions on Internet Technology (TOIT)*, vol. 16, no. 1, pp. 1-20, 2016.
- [48] Y.-H. Zhang, Y.-J. Gong, Y. Gao, H. Wang, and J. Zhang, "Parameter-Free Voronoi Neighborhood for Evolutionary Multimodal Optimization," *IEEE Transactions on Evolutionary Computation*, vol. 24, no. 2, pp. 335-349, 2019.
- [49] Z.-G. Chen, Z.-H. Zhan, H. Wang, and J. Zhang, "Distributed individuals for multiple peaks: A novel differential evolution for multimodal optimization problems," *IEEE Transactions on Evolutionary Computation*, 2019.
- [50] Z.-J. Wang *et al.*, "Automatic niching differential evolution with contour prediction approach for multimodal optimization problems," *IEEE Transactions on Evolutionary Computation*, vol. 24, no. 1, pp. 114-128, 2019.

- [51] Y. Zhang and M. Pennacchiotti, "Predicting purchase behaviors from social media," in *Proceedings of the 22nd international conference on World Wide Web*, 2013: ACM, pp. 1521-1532.
- [52] M. Kosinski, Y. Bachrach, P. Kohli, D. Stillwell, and T. Graepel, "Manifestations of user personality in website choice and behaviour on online social networks," *Machine learning*, vol. 95, no. 3, pp. 357-380, 2014.
- [53] L. Li, A. Li, B. Hao, Z. Guan, and T. Zhu, "Predicting active users' personality based on micro-blogging behaviors," *PloS one*, vol. 9, no. 1, p. e84997, 2014.
- [54] C. Stavros, M. D. Meng, K. Westberg, and F. Farrelly, "Understanding fan motivation for interacting on social media," *Sport Management Review*, vol. 17, no. 4, pp. 455-469, 2014.
- [55] C. Dijkmans, P. Kerkhof, and C. J. Beukeboom, "A stage to engage: Social media use and corporate reputation," *Tourism Management*, vol. 47, pp. 58-67, 2015.
- [56] L. Dessart, C. Veloutsou, and A. Morgan-Thomas, "Consumer engagement in online brand communities: a social media perspective," *Journal of Product & Brand Management*, vol. 24, no. 1, pp. 28-42, 2015.
- [57] M. Kosinski, D. Stillwell, and T. Graepel, "Private traits and attributes are predictable from digital records of human behavior," *Proceedings of the National Academy of Sciences*, vol. 110, no. 15, pp. 5802-5805, 2013.
- [58] X. Zhou and L. Zhang, "Crowdsourcing functions of the living city from Twitter and Foursquare data," *Cartography and Geographic Information Science*, vol. 43, no. 5, pp. 393-404, 2016.
- [59] S. Hasan, X. Zhan, and S. V. Ukkusuri, "Understanding urban human activity and mobility patterns using large-scale location-based data from online social media," in *Proceedings of the 2nd ACM SIGKDD international workshop on urban computing*, 2013: ACM, p. 6.
- [60] A. Noulas, S. Scellato, C. Mascolo, and M. Pontil, "An Empirical Study of Geographic User Activity Patterns in Foursquare," *ICWSM*, vol. 11, pp. 70-573, 2011.
- [61] Y. Zhuang, S. Fong, M. Yuan, Y. Sung, K. Cho, and R. K. Wong, "Location-based big data analytics for guessing the next Foursquare check-ins," *The Journal of Supercomputing*, pp. 1-16, 2016.
- [62] H. Gao, J. Tang, and H. Liu, "gSCorr: modeling geo-social correlations for new check-ins on location-based social networks," in *Proceedings of the 21st ACM international conference on Information and knowledge management*, 2012: ACM, pp. 1582-1586.
- [63] J. W. Pennebaker, R. L. Boyd, K. Jordan, and K. Blackburn, "The development and psychometric properties of LIWC2015," 2015.
- [64] I. Zheludev, R. Smith, and T. Aste, "When can social media lead financial markets?," *Scientific reports*, vol. 4, p. 4213, 2014.
- [65] S. Y. Yang, S. Y. K. Mo, and A. Liu, "Twitter financial community sentiment and its predictive relationship to stock market movement," *Quantitative Finance*, vol. 15, no. 10, pp. 1637-1656, 2015.
- [66] P. Xie, "Predicting Digital Currency Market With Social Data: Implications Of Network Structure And Incentive Hierarchy," Georgia Institute of Technology, 2017.
- [67] M. Aljohani, A. Nisbet, and K. Blincoe, "A survey of social media users privacy settings & information disclosure," 2016.
- [68] Z. Cheng, J. Caverlee, K. Lee, and D. Z. Sui, "Exploring millions of footprints in location sharing services," in *Fifth International AAAI Conference on Weblogs and Social Media*, 2011.
- [69] J. Cranshaw, R. Schwartz, J. I. Hong, and N. Sadeh, "The livelihoods project: Utilizing social media to understand the dynamics of a city," 2012.
- [70] P. Aliandu, "Sentiment Analysis to Determine Accommodation, Shopping and Culinary Location on Foursquare in Kupang City," *Procedia Computer Science*, vol. 72, pp. 300-305, 2015.
- [71] L. Böcker, M. Dijst, and J. Prillwitz, "Impact of everyday weather on individual daily travel behaviours in perspective: a literature review," *Transport reviews*, vol. 33, no. 1, pp. 71-91, 2013.
- [72] MeaningCloud. "Extract valuable information from any text source." Meaning Cloud. <https://www.meaningcloud.com/developer/apis> (accessed 29th October, 2017).
- [73] L. A. Gottschalk and G. C. Gleser, *The measurement of psychological states through the content analysis of verbal behavior*. Univ of California Press, 1969.
- [74] S. Bimonte and V. Faralla, "Happiness and outdoor vacations appreciative versus consumptive tourists," *Journal of Travel Research*, vol. 54, no. 2, pp. 179-192, 2015.
- [75] A. G. E. Workplace. "Gender workplace statistics at a glance." Australian Government- Workplace Gender Equality Agency. https://www.wgea.gov.au/sites/default/files/Stats_at_a_Glance.pdf (accessed 21st October 2017, 2017).
- [76] DataMinr. "Datamnr." DataMinr. <https://info.datamnr.com/> (accessed 9th December 2019, 2019).
- [77] Palantir. Palantir. www. palantir.com (accessed 19th December, 2019).