DOCTORAL DISSERTATION

Human Mobility of Social Network Users in an Urban Area: Case Study in Makassar City, Indonesia

August 2018

Yuyun

Graduate School of Environmental Engineering

Dewancker Bart Lab

The University of Kitakyushu, Japan

ACKNOWLEDGEMENT

Alhamdulillah, praise be to Allah the Almighty and the most Merciful that because of His blessings I could finally finish my thesis with the title: Human Mobility of Social Network User in an Urban area: Case Study in Makassar City, Indonesia.

I would like to express my highest gratitude and appreciation to my supervisor, Prof. Bart Julien Dewancher, the person who never gives up to give me advice, ideas, guidance, assistance in completing the thesis. You are so kind, thank you for all you have given to me.

I deeply thanks to jury members: Prof. Hiroatsu Fukuda, Prof. Koji Takasu, Prof. Isamu Akiba for all of the valuable inputs and suggestions.

Thank you for the entire faculty and staff of Kitakyushu University who have provided assistance, guidance, and knowledge to the author.

Thank you for the big family of PPI Kitakyushu for a kinship and friendship relationship.

Thank you for the entire Bart LAB friends (of friendship and togetherness I will not forget) love you guys.

Special thanks to DIKTI for Scholarship support to pursue a doctoral degree at the University of Kitakyushu.

Thanks to my Parents, my Family: Tri and Shinji Kun, Sister, Brother for prayers, support and guidance.

Finally, thank you for all parties who could not be mentioned one by one by the author that has involved a lot of help so that this thesis could be completed.

ABSTRACT

In recent years, social media has developed rapidly. This progress is underpinned by the providers' sites offering the latest innovations in order to attract users. In addition to being a communication channel, social media can also provide a user's location coordinates through the check-in feature. With this feature, individuals can indicate their surroundings by sharing particulars about their location. In this case, the location sharing displays a map that reveals the time and place at which an individual's status was posted; thus, individuals' activities in the virtual world reflect that what is happening in the real world. Therefore, Location check-in reveals the existence of individual hidden activities that show human movement. Furthermore, the check-in feature generates unprecedentedly rich information about urban space. As an effect, location-based social media provides new knowledge that reinforces previous theories about human movement.

In this dissertation, the author analyses 211,922 check-ins on Twitter in Makassar City. Specifically, this study discusses whether human movement sourced from location-based social media can be used as a data source for urban planning. In previous research, the dataset was utilized to analyse people's movement by comparing the population on Twitter with the real urban population. The analysis uses three data sources: Twitter data, the population census, and questionnaire data. Secondly, a mapping approach was used to study the dynamic urban land-use pattern by combining check-in features and individual text-posting activities. Thirdly, using a grid based on an aggregation method to analyse the city center's location and delineate the boundaries of the city. Forth, quantified the mobility of urban inhabitants by examining individuals' movement patterns and calculated how far people travel in the city. Lastly, analysed the activity of social media users in the public spaces and public facilities by identifying the type of places that become a priority for people visit.

This study concludes that there is a correlation between the urban population and the Twitter population in identifying the existence of people in a region. The author also observes that the check-in distribution in an area offers an excellent opportunity to define the land-use function. Due to social media data being dynamic, in certain conditions space will change its function. This situation depends on the time, the day and the users' text-posted status. In this regard, the findings provide input for stakeholders in creating an up-to-date urban land-use pattern. Finally, the author concludes that, as a data source, location-based social media has great potential for helping understand the shape and structure of a city.

Table of Contents

List of Figures

List of Table

Chapter 1
Introduction

1.1 Background

The increase in human mobility in urban areas is a result of the rising global population. Researchers have found that knowledge of human mobility patterns is closely related to planning and management. In recent decades, the study of human mobility has been applied to many fields such as monitoring the spread of infectious diseases [26], detecting urban black holes [27], traffic forecasting [28], transportation planning [29], and the impact of socio-economic factors [30].

Understanding human mobility in urban areas is vital for urban planning [12, 13]. This data produces rich information about locations and also implies the existence of an individual. Knowledge of user mobility in an urban area provides information concerning how people use urban spaces, how they commute on a daily basis, what they are doing, and where they are at that time; this means that people's activities in the places visited (e.g., offices, malls, restaurants, hotels, etc.) on weekdays and weekends results in rich information for urban development. Previous studies have been conducted to collect data on human mobility; information obtained from direct observation and surveys to ascertain how townspeople interacted with their urban neighbourhoods considered the pattern of citizens' travel [14, 15] and analyse individuals' walking patterns from their points of origin and destinations [17]. Other than the above-mentioned points, to capture individuals' movements experts have also used the latest technological devices such as mobile phone call tracking [4], the movement of banknotes [5], subway smart card data [6], GPS information from taxis [7], and wireless signal connections [16].

At the same time, other benefits generated from human mobility data can be used to analyse urban spaces: for instance, the discovery of an urban function zone can capture the socio-economic activities of citizens in different locations [18]; road and transport design can help create new design methods and vehicle lines [19]; pedestrian walking patterns may assist in improving the design and planning of roads [20]; road traffic analysis may help to estimate the traffic stream of a lane during a specific time interval [21, 22]; and characterizing urban dynamics may help in analysing the state and sustainability of a city [24] and in discovering that a region may have different function [25]. Additionally, another advantage of the data is the ability to analyse urban lifestyle patterns from location choices [23]. The many benefits stemming from this dataset have become the motivation for my thesis. By using patterns of movement from Twitter, the author thus utilizes social media data as additional information to understand a city.

Nowadays, the number of social media users has increased significantly. Not only youth but almost all age groups use this app. Statista reports that, at present, the number of social media users in the world has reached 2.62 billion [1] and it predicts that this number will continue to increase to 3.02 billion in 2021 (see Figure 1.1). This means that almost half of the world's population is using social media.

In Indonesia, the statistics regarding internet use as released by the Ministry of Communication and Informatics republic of Indonesia report that currently 63 million people of a population of 262 million use the internet. From that number, 95% have used the internet to access social media. The most accessed social media networking sites are Facebook and Twitter [2]. From a total of 514 cities and districts spread throughout 34 provinces, there are six cities with a large number of internet users, five of which are concentrated on Java Island and one on Sulawesi Island (see Figure 1.2.) [3]. In this regard, social media has become essential for connecting and interacting with friends and also with the environment.

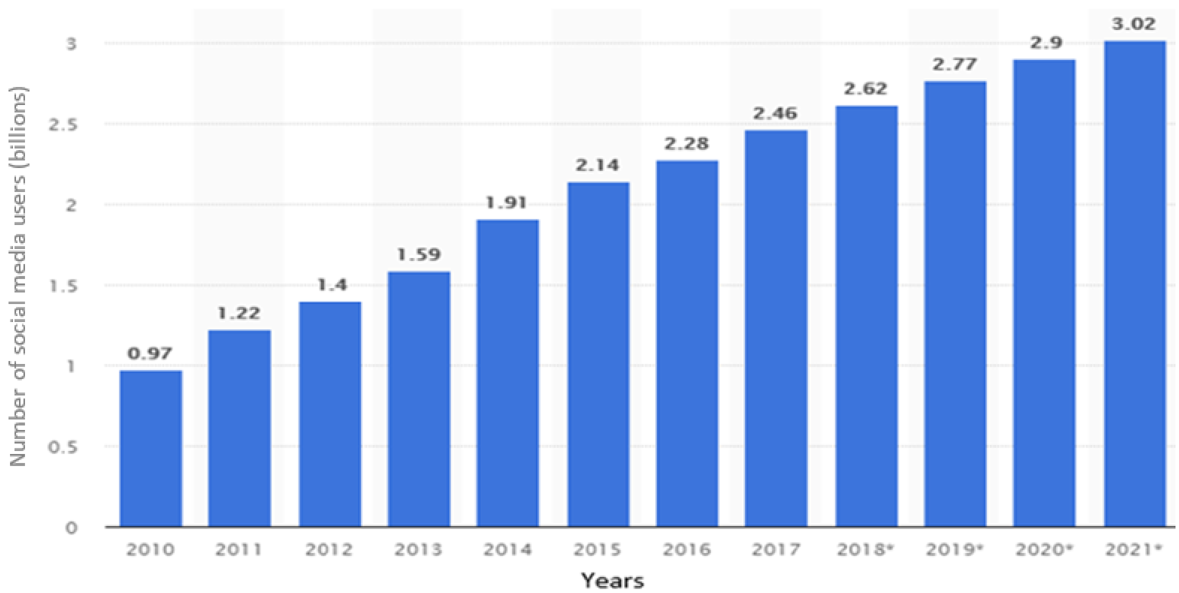


Figure 1.1 Number of social media users worldwide from 2010 to 2021 (in billions)

Source*: www.statista.com*

Social media has become one of the most important communication tools in our daily lives. Beyond being an individual's interaction platform, a user can also share their happiness, pleasure, and opinions about what they see in the places they visit on the status posted. The rapid development of smartphone technology has helped to increase the number of users. Nowadays, almost all smartphones are equipped with a GPS device that allows for a user's position to be tracked. Social media utilizes this device to display the user's specific location. Many new features have been added to increase comfort and convenience for social media users; one of them is a location-based feature. The location feature can show the coordinate points in the form of longitude and latitude by utilizing a GPS sensor embedded in the smartphone to show the user's geographical position in the real world. In addition, this data informs about the individual's activities in specific locations, e.g., shops, parks, hotels, malls, or other places visited. In this context, the author utilized the location feature in the form of check-in activities as one of the main instruments to understand human mobility.

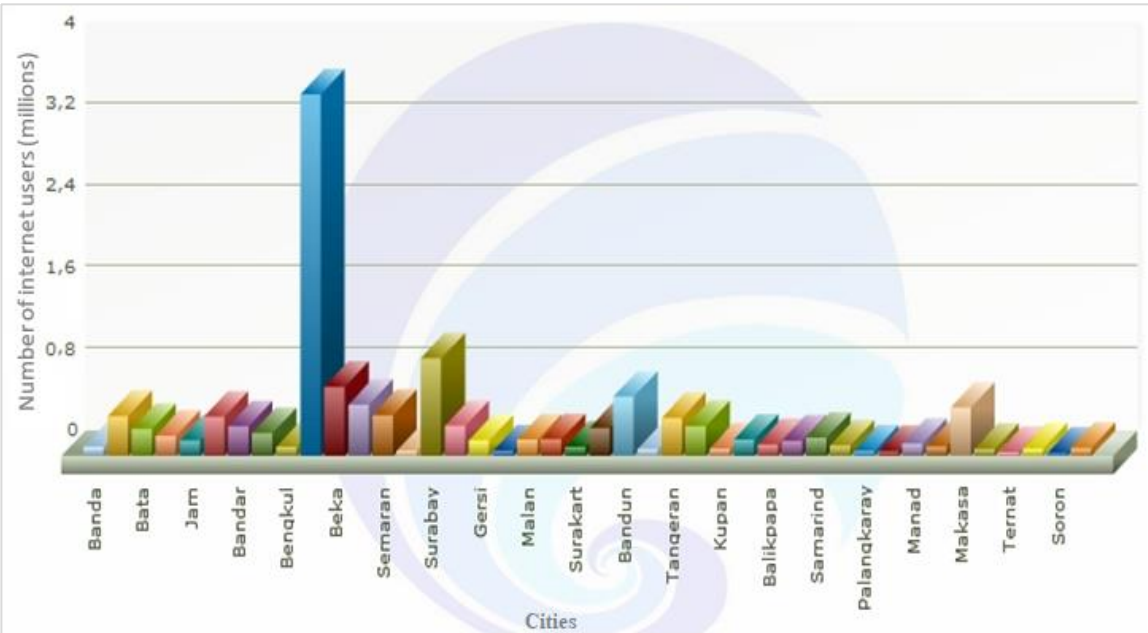


Figure 1.2 Number of internet users based on cities in Indonesia (2013)

Source*: www. statistik.kominfo.go.id*

Ministry of Communication and Information Technology of the Republic of Indonesia

With millions of user-generated check-ins, individual digital footprints regarding amounts of user's embroiled reach were produced. Besides offering new opportunities in the field of mobile applications, this data has the potential to provide a new opportunity to validate the prior theories and approaches used by scholars to gather information related to human movement. But what are the benefits of exploiting user mobility activity in an urban area? In the author's previous studies, this data was used to analyse the urban population, land-use function, urban area mapping, and for quantifying the mobility of urban inhabitants. The author concluded that location-based social network datasets offered a new opportunity for urban planning. Thus, this dataset can be used to compare and validate the existing theories and approaches to understanding individuals' mobility activity in the urban environment.

1.2 Problems Statement

An analysis of individuals' mobility patterns is one of several ways to measure activity in an urban environment. Nowadays, some approaches used by scholars to gather information concerning human movement are direct observation, population surveys, socio-economic characteristics and traveller trip patterns on specific days. These are qualified methods, but there are some issues which arise such as they are usually costly to implement and are weak with respect to covering a large number of individuals and data validation. Meanwhile, alternative approaches used by academics to record individual mobility in urban areas include, e.g., using mobile phone call tracking that indicates where the call occurred, banknote tracking, subway smart card data, and analysis of taxis' GPS information. However, individual behaviour patterns within cities remain hard to understand due to the lack of quantitative validation of the results which is hard to obtain because of privacy concerns [8, 9]. Thus, the main question of this research is whether data from Twitter can be used as a new data source for urban planning applications.

1.3 Research Purpose

An understanding human mobility patterns, specifically in Makassar City, will provide knowledge about the function of a space. In this dissertation, space can be defined as one of the entities in understanding an individual's pattern. In this case, space can express an individual's location, such as when and where an individual is at a given time and what they are doing. In this research, user mobility patterns from the social media application Twitter are used to analyse space functions. The location check-in feature of the system allows people's activities to be recorded. In this case, check-ins can be an instrument to understand human mobility. Thus, this information provides excellent benefits for stakeholders in understanding urban environments. This study aims to investigate the possibilities of using Twitter's social media data as a source of information for urban planning. Specifically, the author will discuss the following points

1. Correlating data from Twitter with census data. In this context, Twitter geolocation data is an essential feature for gaining knowledge about human mobility and the dynamics of urban residents. This paper aims to explore the relation between geolocation data with the existence of people in the urban area. Firstly, the study will analyse the spread of people in a particular area within the city. Secondly, the author will then match and categorize the actual place based on the same individuals visiting it. Then the author will combine the Twitter data from the tracking results and the questionnaire data to obtain the Twitter user's profile. To address this, the writer used a distribution frequency analysis to learn the percentage of visitors and compare it with the local population

Statistics data and land-use mapping released by the city planning department of Makassar's local government to validate the hypothesis.

2. Dynamic land-use map. The identification of land use is intended to reveal accurate sites for future urban development planning. The purpose of this research is to investigate the use of social networking check-in data as a source of information to characterize dynamic urban land use. The data from this study was obtained from the social media application Twitter. The three kinds of data prioritized in this research are check-ins (specific location), timestamps, and a user's status text or post activities. In this study, the writer proposes a grid-based aggregation method to divide the urban area. Two different approaches are compared—rank and clustering methods to group the place's activities. Then the author utilizes time distribution frequency to attain the land-use function. In this case, Makassar City, Indonesia, has been selected as the case study. An analysis shows that the check-in activity and the method that has been proposed can be used to group the actual land-use types.

3. Quantifying the mobility of urban inhabitants. In this regard, check-in locations on social media provide information about an individual's specific position in the real world. In this research, the author used a geolocation service and users' texts posted on Twitter social to quantify human mobility. The research will answer the following two questions: What are the movement patterns of a given citizen and how far do people travel in the city? The trajectory of 201,118 check-ins and 22,318 users over a period of one month in Makassar City, Indonesia is explored. To accommodate individual mobility, the author only analyses users with check-in activity greater than 30 times. To do this, the writer employed a sampling method with a systematic approach to assign individuals. The study found that individual movement shows a high degree of regularity and intensity in certain places. The other finding was that the average distance an urban inhabitant can travel per day is as far as 9.6 km.

4. City center area. Social media generates enormous amounts of data, represents a place where people meet, and makes it possible to measure public sentiment by mining real-time data. Makassar City is becoming more complex and is likely to become polycentric. A better city requires understanding an explicit topology that reveals how city centers are spatially distributed and how to interact with them. A city center must be precisely identified or delineated; it is the core of a city in which, from a spatial perspective, there is a clustering of socio-economic activities. This research collected data sourced from Twitter to examine the center of Makassar City. Two kinds of data used in this study are check-ins and user's posted status text. In general, check-in describes the geographic location of users when they tweet. The individual posted text offers the specific location of the individual. These features become key to catching individual acts in the urban area. As a result, the type of movement data and the model proposed can describe the city center with a proper boundary.

5. Identifying public spaces and public facilities. The aim of this chapter is to categorize the type of place that be a priority for users to visit. The author recognizes six dominant categories: education, banks, administrative, worship, recreation, and health. It was found that beaches are the most visited places with check-in activity at about 34%. Meanwhile, cultural heritage as places have the lowest tweet activity with 5% check-ins. regarding public facilities, universities and schools are the most visited places with 73% check-ins, and banks have the lowest tweet activity with 4%.

1.4 Research Structure

To facilitate the organization of this dissertation, the author has divided the research into 8 chapters. Most of the chapters have been adapted from a paper that has been published in a scientific journal and proceedings. The details are described below:

Chapter 1 This chapter is the introduction. The chapter explains the background, problem statement, research purpose, and research structure.

Chapter 2 This chapter offers information from a literature review. In general, this chapter describes the behaviour and concept of human mobility.

Chapter 3 This chapter contains information about Makassar City and specifically explains its demography and land use. The author then explains the Twitter data collection method.

Chapter 4 This chapter explains the case study and the comparison of the results of the questionnaire survey, and the population and Twitter data. The author aims to understand the profile of a Twitter user.

Chapter 5 This chapter describes a case study regarding how Twitter data can be used to understand the dynamic land-use map. In this chapter land use is divided into four parts: housing, education, business and commercial, and mixed areas.

Chapter 6 This chapter discusses a case study in which the author explains the mobility of urban inhabitants. The chapter answers questions about the movement patterns of citizens and how far people travel in the city. In contrast to previous chapters, the author only analyses users with check-in activities in excess of 30 times.

Chapter 7 Chapter 7 explains a case study in which the author describes how the city center can be formed and detected by using location check-in from social media. To identify the city center, the author divided the research area into groups. From these groups, the author then recognized the kinds of places visited by users.

Chapter 8 This chapter describes a case study in which the author identifies public spaces and public facilities by using individuals' movement data from Twitter. The study will answer the question of how open space and public facilities are used in Makassar City from the perspective of Twitter users.

Chapter 9 In the last chapter, the author explains the findings of each chapter and future research.

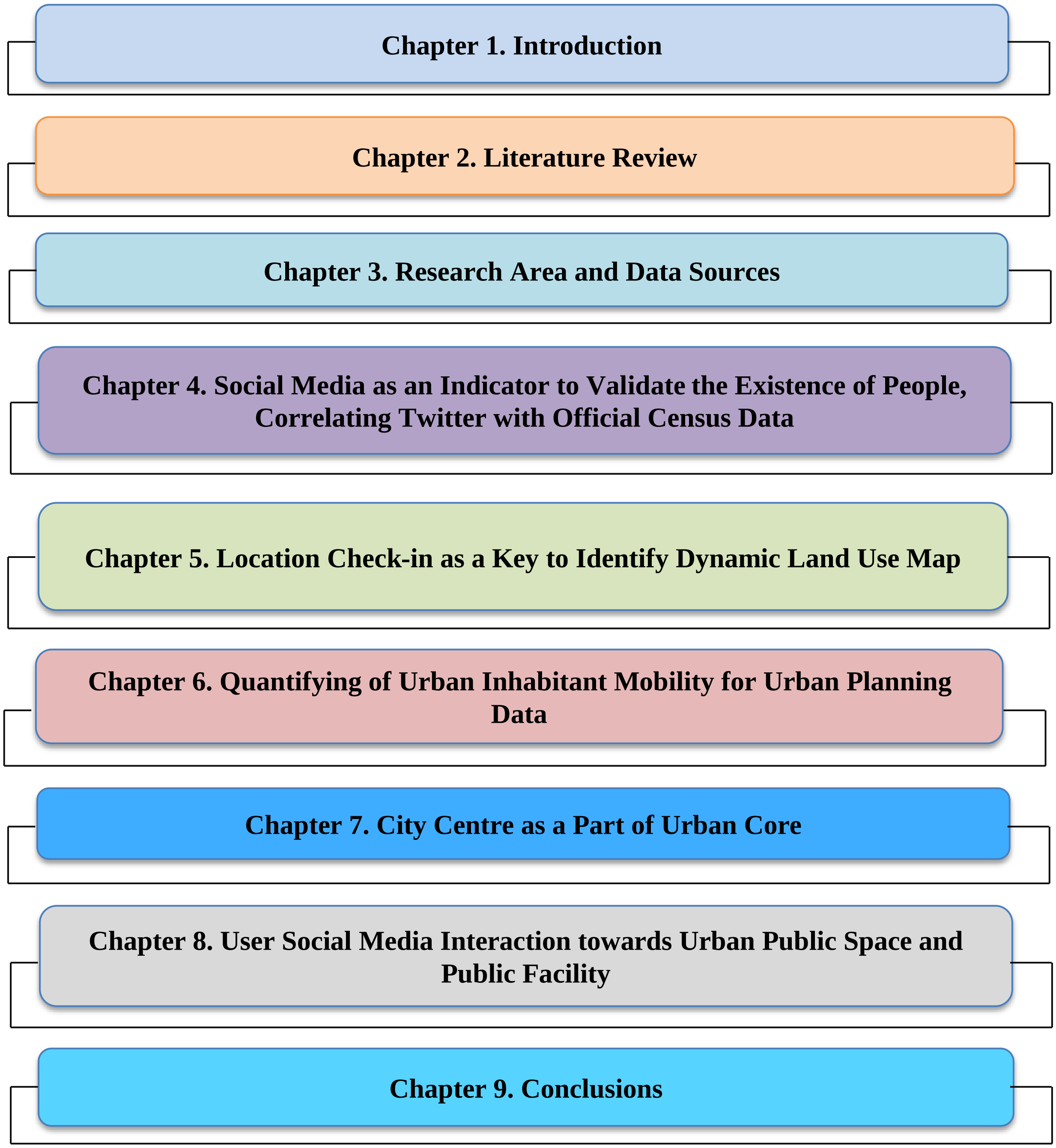


Figure 1. 3 Research Chapters

References

1. Social media statistics data. Available online: https://www.statista.com // (accessed on 27 March **2018**).
2. Internet users in Indonesia. Available online: https:// kominfo.go.id// (accessed on 13 March **2018**).
3. Statistics of a large number of internet user's in Indonesia. Available online: https:// kominfo.go.id// (accessed on 13 March **2018**).
4. Furletti, B.; Gabrielli, L.; Renso, C.; Rinzivillo, S. Analysis of GSM calls data for understanding user mobility behavior. IEEE International Conference. **2013**, pp.550-555
5. Brockmann, D.; Hufnagel, L.; Geisel, T. The scaling laws of human travel. *Nature*, vol. 439, no. 7075. **2016**; 462–465.
6. Hasan, S.; Schneider, C.M.; Ukkusuri, S.V.; González, M.C. Spatiotemporal Patterns of Urban Human Mobility. *J. Stat. Phys.* **2012**, 151, 304–318.
7. Peng, C.; Jin, X.; Wong K-C. Collective Human Mobility Pattern from Taxi Trips in Urban Area. *J. PLoS One*. **2012**, 7, 1-8.
8. Zheng, Y.; Liu, Y.; Yuan, J.; Xie, X. Urban Computing with Taxicabs. In proceeding of the 13th International Conference on Ubiquitous Computing, Beijing, China, 17-21 September 2011, ACM: New York, NY, USA, **2011**, pp; 89-98.
9. Yuan, J.; Zheng, Y.; Xie, X. Discovering Regions of Different Functions in a City Using Human Mobility and POIs Categories and Subject Descriptors. In Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Beijing China, 12-16 August 2012, ACM: New York, NY, USA, **2012**, pp; 186-194.
10. Hyman, G.M. The calibration of trip distribution models. *Journal of Environment and Planning*. **1969**, 1, 105–112.
11. Niemeier, D.; Silvis, J.; D'Souza, R. Social Networks and Travel Behavior: Report from an Integrated travel Diary. In proceeding of the 11th International Conference on Travel Behaviour Research, Kyoto, 16-20 August **2006**.
12. Henderson, T.; Kotz, D.; and Abyzov, I. The changing usage of a mature campus-wide wireless network. *Journal of Computer Network*. *Elsevier* **2008**, 52, 2690–2712.
13. Papadimitriou, E.; Lassarre, S.; Yannis, G. Human Factors of Pedestrian Walking and Crossing Behavior. Transportation Research, vol. 25, Procedia **2017**, pp. 2002– 2015.
14. Yuan, J.; Zheng, Y.; Xie, X. Discovering Regions of Different Functions in a City Using Human Mobility and POIs. In Proceeding of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. Beijing, China, 12-16 August 2012, ACM: New York, NY, USA, **2012**, pp; 186-194.
15. Fuller, R.; Santos, J.A. Human Factors for Highway Engineers. Pergamon Press, Amsterdam, N.L. **2012**, pp; 1-10.
16. Papadimitriou, E.; Lassarre, S.; Yannis, G. Human factors of pedestrian walking and crossing behavior. *Transportation Research. Procedia* **2016**, 25, 2002-2015.
17. Gundlegard, D.; Karlsson, J.M. Road Traffic Estimation using Cellular Network Signaling. In Intelligent Transportation Systems. ISBN: 978-1-60741-588-6, **2009**.
18. Martinez, V.F; Moummi, Y.; Martinez, E.F. Estimation of Traffic Flow Using Passive Cell-Phone Data. In Proceeding of the International Workshop on Data Science for Macro-Modeling, Snow, UT, USA, 22-27 June 2014, ACM: New York, NY, USA, **2014**, pp; 1–2.
19. Hasan, S.; Ukkusuri, S.V. Location Contexts of User Check-ins to Model Urban Geo Life-Style Patterns. *PLoS One*. **2015**, 10, 1–19.
20. Qian, X.; Zhan, X.; Ukkusuri, S.V. Characterizing Urban Dynamics Using Large Scale Taxicab Data. *Engineering and Applied Sciences Optimization, Computational Methods in Applied Sciences*. *Springer* **2015**, 38, 17–33.

21. Mahmud, J.; Nichols. J.; Drews, C. Home Location Identification of Twitter Users. *Journal of Intelligent Systems and Technology*, 5, **2014**.
22. Brockmann, D.; David, V.; Gallardo, A.M. Human Mobility and Spatial Disease Dynamics. **2009**, pp; 55–81.
23. Hong, L.; Zheng, Y.; Yung, D.; Shang, J.; Zou, L. Detecting Urban Black Holes Based on Human Mobility Data. In Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems. Seattle, Washington, 3-6 November 2015. ACM: New York, NY, USA, **2015**.
24. Pan, B.; Zheng, Y.; Wilkie, D.; Shahabi, C. Crowd Sensing of Traffic Anomalies based on Human Mobility and Social Media. In Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. Orlando, Florida 05-08 November 2013, ACM: New York, NY, USA, **2013**, pp; 344-353.
25. Song, X.; Kanasugi, H.; Shibasaki, R. DeepTransport: Prediction and Simulation of Human Mobility and Transportation Mode at a Citywide Level. **2013**, pp; 2618–2624.
26. Martinez, V.F.; Virseda, J.; Rubio, A.; Martinez, E.F. Towards Large Scale Technology Impact Analyses : Automatic Residential Localization from Mobile Phone-Call Data. **2009.**

Chapter 2
Literature Review

2.1 History of Social Media

The development of social media networks began when Johannes Gutenberg invented the letter printing press in 1450 [1]. The invention of this machine became the seed of mass media, known as print media. Due to mass media not being able to spread news directly to the public (as it must wait to be printed) the internet, with websites as a means of deploying information, appeared.

2.1.1 Evolution of the Social Media Site

The social media era began in 1978 from the invention of the bulletin board system (BBS) which connects people using electronic mail. From this discovery in 1979, the first social media, called Usenet, was launched. Social media evolved as follows:

1. Usenet is an online discussion group developed by Duke University graduate students Tom Truscott and Jim Ellis. The critical feature of this site is that users can read and send messages to groups.
2. Listserv was built in 1984. Listserv is an electronic mail software application that allows individuals to exchange information with multiple users. Unlike the current email format, Listserv needs a server to help forward messages to the destination.
3. Internet Relay Chat (IRC) is a multi-user chatting system created by Jarkko Oikarinen in 1998. This device is designed for communication channels between individuals and groups. To enter the group, users are required to join a chat room, commonly called a channel. Through this channel, IRC users can meet.
4. In 1995, GeoCities was born to serve web hosting—the rental data storage service of a website such that the website pages can be accessed from anywhere. GeoCities sparked the establishment of other websites. In the same year, the first social media site, classmates.com, appeared.
5. In 1997, the second social networking site that appeared was sixdegree.com. This social media site is considered to be more of a social networking site in comparison with classmates.com.
6. Then in 1999, a site to create a personal blog, Blogger, appeared. This site offers users the chance to create their own websites; users of Blogger can load anything.
7. In 2002, Friendster was founded. This site is more familiar with the phenomenon of social media.
8. In 2003, LinkedIn was established. This site is not only useful for socializing, but LinkedIn is also helpful for finding a job. As such, it serves a dual function.

9. In 2003, Myspace was established. Myspace offers great ease of use; it is a user-friendly social networking site.
10. In 2003, Flickr was also founded. Flickr is a website for photo sharing and video management.
11. 2004 saw the birth of Facebook, the famous social networking site, which is one of the social networking sites that has the most members.
12. 2006 saw the birth of Twitter. This social networking sites is different from others because updates on the status of users are limited to 140 characters.
13. In 2009, Foursquare was established. This social media focuses more on location check-in as its main feature.
14. In 2010, Path and Instagram were established. Instagram is best known for sharing images, videos and messages.
15. In 2011, the birth of Google+, Google's social networking site.

2.1.2 Characteristics of Social Media

Since social media was created in 1979, other types of social media have emerged. In this research, the characteristics of social media divided into two categories: past and current perspectives [2]. The details of these categories can be seen in Table 2 below.

Table 2.1 Past and present social media functionalities

Past perspective	**Characteristics**
Usernets	User can send and post an article to a common group.
Bulletin board systems	Through a personal computer and modem, the user can interact.
Commercial online services	The first chat application, specifically for commercial service, was launched.
Instant messaging	Users can post updates to their global network, and emotion icons began to emerge.
Social media sharing	Users can exchange media files.
Current perspective	**Characteristics**
Real-time updates	Previously, social media technology relied more on static content; now the trend focuses on real-time updates. Location-based technologies have also increased in popularity, allowing a user to check-in at various places.
Virtual worlds	The user has a virtual community. They can actively participate with others in a community, but they don't know each other, e.g., a game community.

Internet calling	In addition to functioning as text communication channels, social media can be used for voice and video calls.
Blogs/Microblogs	Instead of just reading static content, users can comment on posts, interact with other users, and perhaps engage in dialogue.
Media sharing	The user can view and share media files. Users can also engage with one another via posting and commenting features.
Social news	With the sharing feature, social media can function as an online newspaper.
Livestreaming sites	This is a feature for live events.

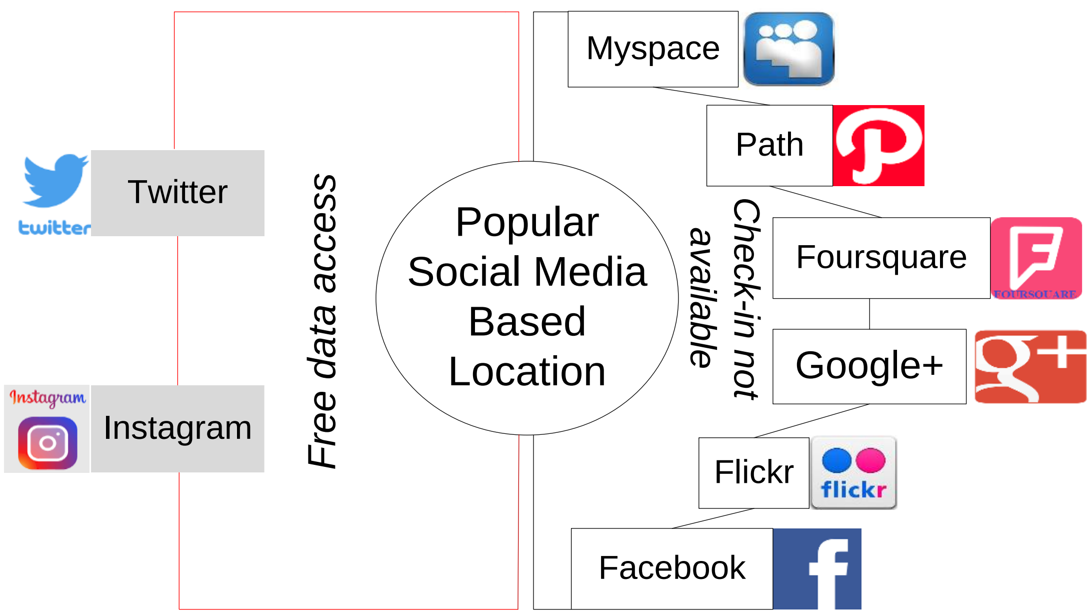


Figure 2.1 Famous location-based social media

Figure 2.1 shows the current social media trends. As described in chapter 3, to access social media user data, users use the application program interface (API) provided by the developer. The author of this research divides social media that gives location features into the following two categories:

1. Free data access is a type of social media that provides access to a download data feature called crawling. Crawling is a process of retrieving or downloading data from a social media server with the help of application programming integration (API). In this research, the writer uses the Twitter API to crawl. The Twitter API is a program that has been provided to facilitate other developers in accessing information on the website. Registration as a Twitter API developer can be completed via https://dev.twitter.com. After registration, developers get a consumer key,

consumer access, an access token and an access token secret that is used as an authentication requirement. The purpose of authentication is to grant a developer permission to download the Twitter data.

2. Closed data access is a type of social media in which the check-ins of users have been set as a private feature so that they cannot be accessed or are unavailable. In general, almost all social media data can be accessed. Some of these datasets are accessible using public APIs provided by the provider site, but usually, some valuable information or metrics are missing from these APIs. For example, the check-in features on Twitter and Instagram are openly accessible. In contrast, check-in features on Facebook, Path, Foursquare, and others are private so that the provider site does not allow the public access; this is one of the reasons why the author uses Twitter data as a single data in this study.

2.2 The Significance of Location Based Social Network

The reason for choosing location-based social media data as significant for human mobility is explained in detail below.

2.2.1 Social Media Location as a Sensor for Human Mobility

Location services on social media are services provided by a device to record the location of users. Location based on social media can function to identify an individual's location or a specific place such as a nearby restaurant, mall, hotel, and other sites. The location service of social media answers the question 'where'; for example, where is the mall? Where am I? Or where is the object's location? This indicates that position is crucial for individuals to arrange their environment and these services continue to show innovation. In addition to a personal digital assistant (PDA) device, innovations have also spread to mobile devices with, for instance, location services on a smartphone, Bluetooth, and wireless location services with the aim of conveniently offering the user the ability to display their geographic location information. Specifically, the author aims to detect mobility activities on the devices individuals use.

2.2.2 The Excess of Location-Based Social Media Service

Location information generated from social media implies a hidden activity that discloses individual travel. In this study, the author tried to identify what distinguishes the current mobility datasets with movement generated from location-based social media.

1. **Check-ins**. Social networking sites offer users the ability to share locations through check-in activities. This feature is attached to the text, image or video in the post to announce the presence

of a user at a place. For example, by conducting a status update in a restaurant, the GPS sensor on the smartphone used will indicate the venue by showing the coordinates of the individual's geographic position in the real world. In addition to displaying a specific location, check-in activity also displays the time and date (days) of the post. Thus, continuous check-in activity will shape a path that reveals an individual's mobility pattern.

2. **Publicly Available Dataset.** One of the common problems in data science is gathering data from various sources. The writer's source of data for this research is Twitter's streaming API. In addition to its being practical, the ease of obtaining significant data has prompted the author to use this. To collect the data, one must have sufficient expertise in computer programming; due to its public availability, anyone can do the research. This will give rise to innovations related to the model and to its analysis. This is in contrast to conventional methods that require costly and considerable resources to obtain a significant amount of data.

2.3 Review of Human Mobility Study Based on Social Media

From a review of location-based social media data, the author found that researchers have used three social media applications to catch individual behaviours on Twitter, Instagram, and Foursquare. The reason that this site is freely available and is open access to download is because of the REST API application provided by the developers. For this part, the author collected a range of methods and applications used to evaluate individuals' activities related to location, as based on social media—specifically those related to human mobility as drawn from academic databases. A total 32 references were reviewed from international journals and articles from 2010 to 2016. The latest findings in each category are then explained and identified in an effort to answer the following three questions: (1) what is the problem being discussed?; (2) How is the research problem resolved (techniques applied)?; (3) What is the conclusion of each article?

Our initial goal is to investigate the extent of the application of location-based social media. To start the data collection, at first, a criterion of geolocation social media is typed into the search engine to get the subject targets. In addition, articles also were sought from journal sites such as PLoS ONE, Science Direct, IEEE Xplore, and the Journal of Information Science in order to supplement the kinds of literature. Then the articles were grouped into several clusters to obtain papers relevant to the research topic. The next process was to analyse the problems discussed, the techniques applied, and the results of each paper. To achieve a high standard of research, the author set some parameters in order to determine the types of papers that would be utilized, such as publication in an international journal and international conference articles. The methodology for the research is illustrated in Figure 2.2.

2.3.1 Keyword of Location-Based Social Media

The location services of social media such as Facebook, Twitter, Foursquare, Instagram, and Path have provided new insights into understanding the shape and structure of a real city. This data has the potential to impact many other areas including travel-demand modelling, ubiquitous computing, epidemiology, urban planning, security, and health monitoring [3]. From the literature list, the author identified three essential elements which are key—namely, movement, location, and distance. Almost all the models that are used improve these elements. For this reason, all the papers are classified into two categories according to the issues that have been discussed in previous research. These categories are defined as follows:

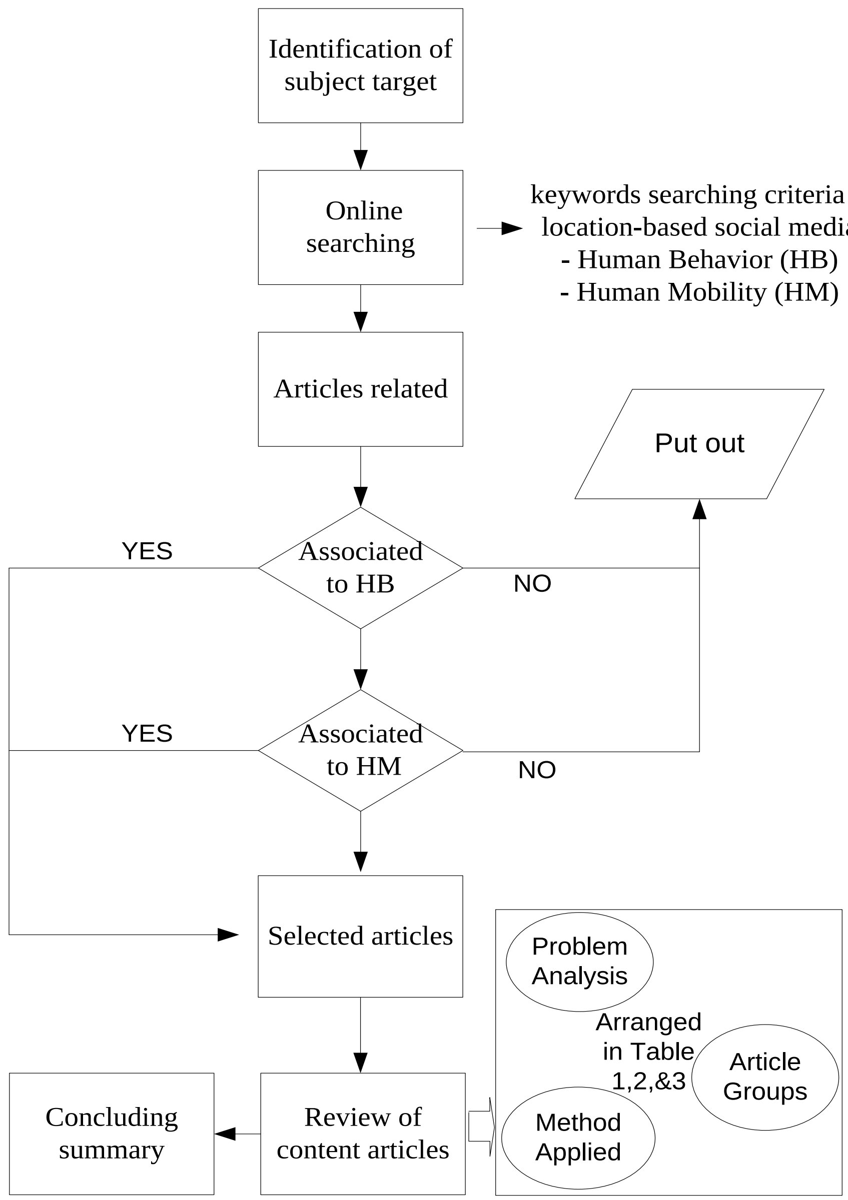


Figure 2.2 Research methodology flow

2.3.2 User Behaviour

Various works have been studied to understand user behaviour related to location-based social media. In this section, the author explains how location-based social media can be modelled to solve existing problems. The author provides a list some literature related to urban applications as a reference for readers (see Table 1). Below, the author provides a summary of the selected articles. Many applications and methods have been proposed to understand the distribution of use of geolocation social media data. Such as [4] presented a new algorithm, namely heuristic classifiers, to predict the home location of Twitter user in different places such as geographic region, city, and state. Their model used the time zone as the criterion to improve prediction accuracy. They analysed movement differences of Twitter users to predict whether a user was traveling in a certain period. They found that this approach works well for predicting Twitter users at their home location.

Then [5, 6] proposed a probabilistic model for estimating Twitter users in the city-level location. They introduced methods for analysing: (i) the tweet content, (ii) the component classification for automatically identifying words in the tweet, and (iii) an estimate of the user's location. They reported that only 100 tweets are required to establish the region. [7] Described location estimation for predicting the home location of Twitter users. They proposed unsupervised methods based on non-localness and geometric-localness and found that about 5112 Twitter users can be predicted within 100 miles. [9] Used the location-based data from Foursquare to analyse the relationship of the places visited to the structures of the social network. They found that a place where people meet has very strong influence upon it being a region. [11] Used k-means clustering to count the data distribution with a distance approach to calculate a place between the closest centers for each user. [16] Developed a topic model to extract multi-day patterns and individual activity patterns. Their model was combined with data from traditional surveys to count the weekly activities pattern and user-specific activity patterns.

[18, 21] proposed the use of geolocation tweets for urban planning and for identifying the urban landscape. The system created used a spectral clustering method. Three cities are presented in the experiments (Manhattan, London, and Madrid). The technique proposed to identify land segmentation and land use. They found geolocation information to be beneficial for urban planners for land-use modelling. [19] Used the same method as mentioned above to model user activity patterns in New York and London. With the geolocation data provided by Foursquare, place information can be detected such as homes, food, parks, and others. [20] Proposed models of aggregation to characterize urban areas by extracting Twitter geolocation data. Then [41] explored geo-tagged social media (Sina-Weibo) to point out the commercial area in Beijing. They used grid-based land-use segmentation and aggregated temporal trends to divide the urban area.

[22] Developed a model to analyse a Twitter text based on the user's position. They found that the best accuracy for text-based location was for men above 40 years old. [23] Demonstrated a distance model for a content-based geolocation tweet. They tested their approach with 17 million multilingual tweets—85% Arabic and 15% English—from 2.6 million twitter users. They reported that the method provides an effective way of grouping trending news topics, geopolitical entities, and hashtags. [24] Used a probabilistic framework to estimate the city-level user location based on the content of the tweets. This research wanted to know whether there was a relation between the location information of the user and their environments. [26] In this study, they used the geographic data from Foursquare's check-in data. The primary objective was to model urban neighbourhood. They devised an optics algorithm for exploring and extracting neighbourhood boundaries in cities. The main points which resulted from developing the model were: (1) activity hotspot detection, (2) measuring area homogeneity, and (3) neighbourhood detection. [35] proposed a probabilistic generative model to infer a city-user`s geolocation on Twitter using their textual content. As a result, 60% of users were identified within 10 km of their locations in the Korean city. [36] Proposed a social graph to model the Twitter geolocation to determine user interests from the locations visited.

Table 2.2 References on the topic of urban applications

Authors	Year	Application Areas	Techniques Used	Countries
Cheng et al.	2010	Location estimate	Probabilistic framework	USA
Chandra et al.	2011	City area	Probabilistic framework	USA
Noulas et al.	2011	Region	Spectral clustering	England
Wakamiya et al.	2011	Urban area	Aggregation model	Japan
Chang et al.	2012	Home location estimate	Unsupervised methods	USA
Chang and Sun.	2012	Individual location	K-means clustering	USA
Eisentein et al.	2012	Region and city	Probabilistic model	USA
Martinez et al.	2012	Urban land scape	ANN	Spain
Brown et al.	2013	Physical spaces	Unknown	England
Khanwalkar et al.	2013	Region	Distance measurement	USA
Zhang et al.	2013	Urban neighborhoods	Unknown	USA and England
Hasan et al.	2014	Urban activity pattern	Topic model	USA
Mahmud et al.	2014	Home identification	Heuristic classifiers	USA
Martines et al.	2014	Urban planning	Spectral clustering	USA
Ryoo & Moon.	2014	City area	Probabilistic generative	South-Korea
Kotzias et al.	2015	City area	Social graph	USA
Pavalanathan et al.	2015	City population	Latent variable model	USA
Wang et al.	2016	Land use	Aggregation Grid	China

The lists in Table 1 are the references which explain the application of geolocation from social media for urban fields. Columns 1 and 2 represent the number and authors of the article; the other columns offer the areas and techniques applied.

2.3.3 Human Mobility

This section aims to explore and analyse an article pertaining to human mobility using location-based social media. From this article by [11], two issues are applied in this research: (1) predicting an individual position, and (2) predicting place. In the article a language model of location using coordinates extracted from geolocation social media was proposed. A geographical model to predict the individual tweet location was created. They found Twitter users at the country, state and city level and achieved a three- to ten-fold increase in accuracy at the zip-code level. [10] Proposed a social-historical model to study the relationship between social ties and the user's check-in behaviour. This model showed how historical effect could help in location prediction.

[12] Utilized Foursquare check-in data to investigate gender in New York City. Ellipse-based and spatial distribution models were used to categorize the activities of individuals. The research results showed that there was a gender difference in the travel and activity pattern of users. This data can be exploited to produce gender-specific daily travel data. The same model and data were used to understand lifestyle choices related to social relationships and individual patterns. They found that the same check-ins of two users will have the same lifestyle in the social media structure. [25] Proposed a system for spatial planners to analyse human distribution using geolocation social media data. Three contributions from the system are evident, namely: (1) use of a graph-based approach for constructing the system, (2) visual analytics to identify the movement between locations, and (3) functional and scalability.

[27] Compared three different methods to detect a city center: (1) cluster detection using LGOG, (2) cluster detection using DBSCAND to counts points, and (3) cluster detection using GN. Three cities (Berlin, Munich, and Cologne) were used to model the system. From their results, they found that: (1) models can identify the city center with a precise boundary, and (2) that they are more flexible for a polycentric city. Hasan & Ukkusuri [28] explored the lifestyle patterns from Foursquare geolocation data and used a probabilistic topic model to solve two issues: (1) patterns of user interests in the different places types, and (2) user visit patterns in different neighborhoods. They found that the exact probability can provide useful information for user choices and interests.

[29] Used data from Foursquare to analyse the user distribution pattern of three different cities (London, Paris, and New York). They found that the distribution of social activity in the three cities was non-linear, especially relate to food and nightlife which have the most active social movement.

Cheng et al.'s [30] study presented a paper which details the use of check-in data to model user mobility patterns. They proved that geolocation on Twitter could potentially be used as data to understand human mobility. They also found that metropolitan areas have a larger number of Twitter user than rural areas. [31] Introduced a rank-based model to study human mobility and urban space in cities. The model calculates the number of places between the point of origin and the destination. The research concluded that this data has the potential to be applied to urban planning and location-based advertisements.

[33] Proposed a novel method to characterize human mobility which focused on the impact of demography. Their methodology utilized the concept of a space-time trajectory to estimate each Twitter user in the urban Chicago area. The data focused on the user at their home location along with demographic information such as ethnicity, gender, and age. They found that user distributions of at-home locations still follow a power-law distribution. Among these three demographic factors, ethnicity has the most significant influence on human mobility. [34] Used the geolocation data from Gowalla and Brightside to analyse the social relations of human mobility. The research concluded that social relationships could explain 10–30% of all human movements, while periodic behaviour explains 50–70%. [40] Aimed to explore the relation between geolocation on Twitter and the presence of people in the urban area. They combined Twitter and questionnaire data to capture the Twitter profile and validated it with the population statistics data. [37] Utilized large datasets from the tweet geolocation to analyse people's movements in real life.

Table 2. 3 References on the topic of mobility application

Authors	Year	Application Areas	Techniques Used	Country
Noulas et al.	2010	Human mobility	Unknown	England
Cheng.	2011	Human mobility	Unknown	USA
Cho et al.	2011	Human mobility	Unknown	USA
Kinsella et al.	2011	Individual location	Language model	Ireland and Spain
Olivier et al.	2011	Human mobility	Unknown	England
Noulas et al.	2011	Human movement	Rank-based	England and Belgium
Gao et al.	2013	User's social behavior	Social-historical model	USA
Chua et al.	2015	Human mobility	Flow sampler	Belgium and Italy
Hasan et al.	2015	Human mobility	Probabilistic topic	Australia and USA
Sun & Li.	2015	Mobility patterns	Spatial distribution	Germany
Comito et al.	2016	Human movement	Sequential pattern mining	Italy
Hasan et al.	2016	Lifestyle choice	Topic model	USA
Luo et al.	2016	Human mobility	Space-time trajectory	USA
Sun et al.	2016	Mobility / city center	Unknown	England
Yuyun et al.	2016	Human mobility	Frequency distribution	Japan

They proposed a novel methodology for detecting relevant semantic locations from geotagged posts. The model was used to calculate the duration of the stay, the time of day and the popularity of places that people visited. The primary objective of the study was to provide the top locations or most popular tourist destinations such as shopping malls, streets, restaurants, and cinemas. Table 2 lists the references in this category.

Figure 2.3 presents the distribution of articles based on the year of publication and Figure 2.4 shows the number of articles published based on country. The articles concern the behaviour and human mobility activity in urban areas using location-based social media data.

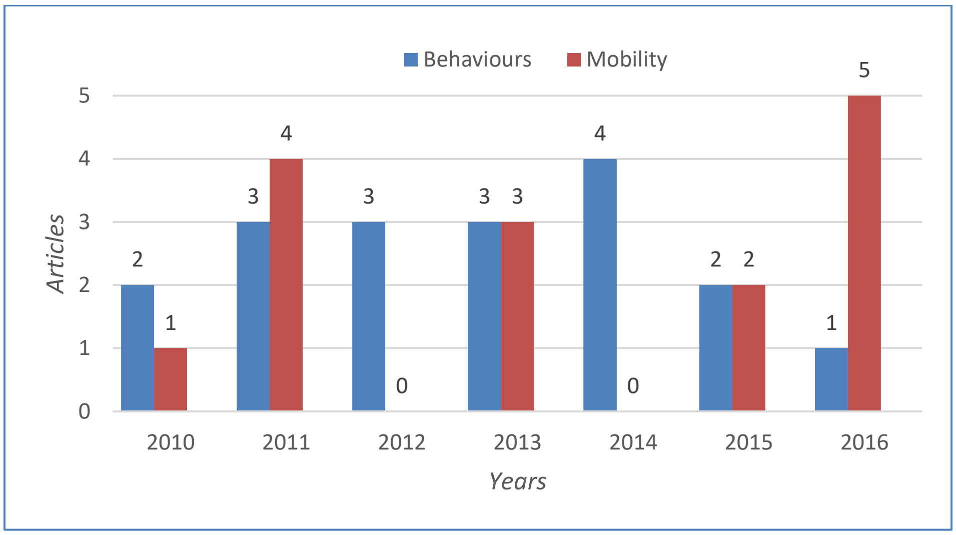


Figure 2.3 Article distribution based on year of publication

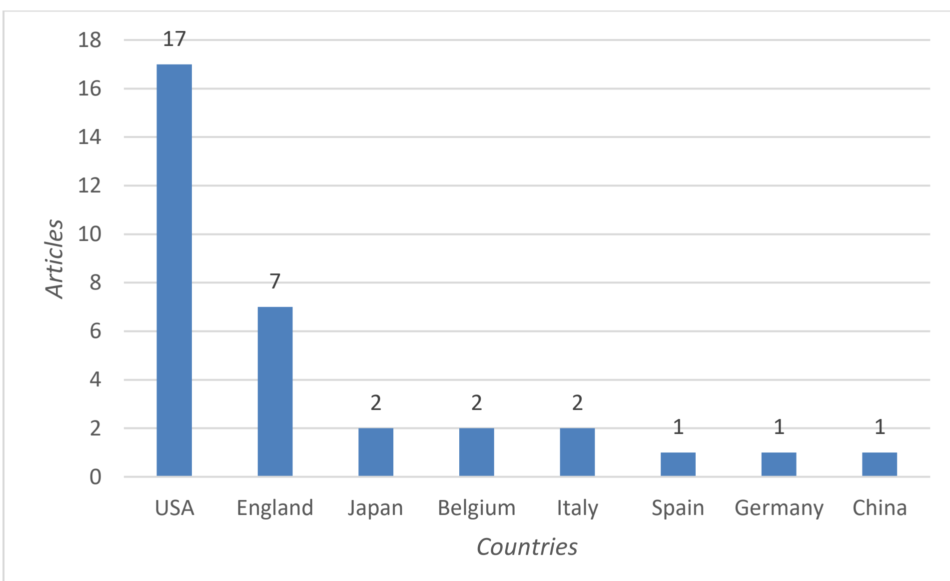


Figure 2.4 Number of articles published per country

References

1. Franklin, B. The History of Printing. in *Jones International and Jones Digital Century, **1994-1999***
2. Korenich, L., Lascu. D.; Manrai, L.; Manrai, A. Social Media: Past, Present and Future. **2013**, 234-249.
3. Hasan, S.; Zhan, X.; Ukkusuri, S.V. Understanding urban human activity and mobility patterns using large-scale location-based data from online social media. In Proceedings of the Second ACM SIGKDD International Workshop on Urban Computing, Chicago, Illinois, 11-13 August 2013; ACM: New York, NY, USA, **2013**, pp. 1-8
4. Mahmud, J.; Nichols. J.; Drews, C. Home Location Identification of Twitter Users. *Journal of Intelligent Systems and Technology*, 5, **2014**.
5. Eisenstein, J.; Connor, B.O.; Smith, N.A.; Xing, E.P. A Latent Variable Model for Geographic Lexical Variation. In Proceeding of EMNLP. **2010**, pp. 1277-1287.
6. Cheng, Z.; Caverlee, J.; Lee, K. You Are Where You Tweet: A Content-based Approach to Geo-locating Twitter Users. In: Proceedings of CIKM, pp.759–768.
7. Chang, H.; Lee, D.; Eltaher, M.; Lee, J. @ Phillies Tweeting from Philly ? Predicting Twitter User Locations with Spatial Word Usage. In Proceeding of the ASONAM, **2012**, pp. 111-118.
8. Kinsella, S.; Murdock, V.; Hare, N.O. I`m Eating a Sandwich in Glasgow: Modeling Locations with Tweets. In Proceeding of the SMUC, **2011**, pp. 61-68.
9. Brown, C.; Noulas, A.; Mascolo, C.; Blondel, V. A place-focused model for social networks in cities. In: Proceeding of Social Com, **2013**, pp.75-80.
10. Gao, H.; Tang, J.; Liu, H. Exploring Social-Historical Ties on Location-Based Social Networks. In: Proceeding of the Sixth international conference on Weblog and Social Media, **2012**, pp. 114–121.
11. Chang, J.; Sun, E.; Ave, S.C. Location 3 : How Users Share and Respond to Location-Based Data on Social Networking Sites. In: Proceeding of the Fifth International AAAI Conference on Weblog and Social Media, **2011**, pp.74–80.
12. Sun, Y.; Li, M. Investigation of Travel and Activity Patterns Using Location-based Social Network Data : A Case Study of Active Mobile Social Media Users. *ISPRS International Journal of Geo-Information*. **2015**, 4, 1512-1529.
13. Dredge, M.; Paul, M.J.; Bergsma, S.; Tran, H. Carmen : A Twitter Geolocation System with Applications to Public Health. In AAAI Workshop on Expanding the Boundaries of Healthy Information Using AI, **2012**, pp. 20–24.
14. Gu, Y.; Sean, Z.; Chen, F. From Twitter to Detector : Real-time Traffic Incident Detection using Social Media Data. *Science Direct Journal of Transportation Research*. **2016**, 67, 321–342.
15. Eduardo, L.; Vasconcelos, G.D.; Ferreira, N. J.; Vasconcelos, L. G. D. Using Tweets for Rainfall Monitoring, **2016**, pp. 1157-1167.
16. Hasan, S.; Ukkusuri, S.V. Urban activity pattern classification using topic models from online geo-location data. *Transportation Research*. **2014**, 44, 363–381.
17. Thomas, B. H. Understanding Social Influence in Activity Location Choice and Lifestyle Patterns Using Geolocation Data. *Journal Frontiers in ICT of mobile and ubiquitous computing*. 2016,1-9.
18. Frias, M.V.; Frias, M.E. Spectral clustering for sensing urban land use using Twitter activity. *Journal of Engineering Applications of Artificial Intelligence*. **2014**, 237–245.
19. Noulas, A.; Scellato, S.; Mascolo, C.; Pontil, M. Exploiting Semantic Annotations for Clustering Geographic Areas and Users in Location-based Social Networks. Association for the Advancement of Artificial Intelligence. **2011**, pp. 32–35.
20. Wakamiya, S.; Lee, R.; Sumiya, K. Urban area characterization based on semantics of crowd

activities in Twitter. *Computer Science*. Springer **2011**, 6631, 108–123.
21. Martinez, V.F.; Soto, V.; Hohwald, H.; Martinez, E.F. Characterizing Urban Landscapes using Geolocated Tweets. In Proceeding IEEE of Social Com. **2012**, pp. 239-248
22. Pavalanathan, U.; Eisenstein, J. Confounds and Consequences in Geotagged Twitter Data. **2015**.
23. Khanwalkar, S.; Seldin, M.; Srivastava, A.; Kumar, A.; Colbath, S. Content-Based Geo-Location Detection for Placing Tweets Pertaining To Trending News on Map. In 4th International workshop on mining ubiquitous and social environments. **2013**.
24. Chandra, S.; Khan, L. Estimating Twitter User Location Using Social Interactions - A Content Based Approach. In: IEEE SocialCom. **2011**, pp. 838-843.
25. Chua, A.; Marcheggiani, E.; Servillo, L. Flow Sampler : Visual Analysis of Urban Flows in Geolocated Social Media Data. In SocInfo Workshop, **2015**, pp. 5-17.
26. Zhang, A.X.; Noulas, A.; Scellato, S.; Mascolo, C. Hood square: Modelling and recommending Neighbourhoods in location-based social networks. In: Proceedings SocialCom, **2013**, pp. 69-74.
27. Fan, H. Identifying the City Centre Using Human Travel Flows Generated from Location-based social networking data. Environment and planning: planning and design, **2015**, pp. 1-19.
28. Hasan, S.; Ukkusuri, S.V. Location contexts of user check-ins to model urban geo life-style patterns. PLoS ONE. **2015**.
29. Phithakkitnukoon, S.; Olivier, P. Sensing Urban Social Geography Using Online Social Networking Data. In Proc of the fifth international conference on web blogs and social media. **2011**, pp. 36-39.
30. Cheng, Z.; Caverlee, J.; Lee, K.; Sui, D. Z. Exploring Millions of Footprints in Location Sharing Services. In Proceeding of the fifth international conference on weblogs and social media. **2011**, pp. 81-88.
31. Noulas, A.; Scellato, S.; Lambiotte, R.; Mascolo, C.; Pontil, M. A tale of many cities : universal patterns in human urban mobility. *PNAS*. **2011**, 1-11.
32. Noulas, A.; Scellato, S.; Mascolo, C.; Pontil, M. An Empirical Study of Geographic User Activity Patterns in Foursquare. In Fifth International AAAI Conference on Weblogs and Social Media. **2010**, pp. 570–573.
33. Luo, F.; Cao, G.; Mulligan, K.; Li, X. Explore spatiotemporal and demographic characteristics of human mobility via Twitter : A case study of Chicago. *Applied Geography*. **2016**, 70, pp. 11-25.
34. Cho, E.; Myers, S.A.; Leskovec, J. Friendship and mobility. In 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, **2011**.
35. Ryoo, K.; Moon, S. Inferring Twitter User Locations with 10km Accuracy. In the International World Wide Web Conference Committee. **2014**, pp. 643-648.
36. Kotzias, D.; Lappas, T.; Gunopulos, D. Home is where your friends' are : Utilizing the social graph to locate twitter users in a city. *Scientists Journal of Information Systems*. **2015**, 1–11.
37. Comito, C.; Falcone, D.; Talia, D. Mining human mobility patterns from social geo-tagged data. *ScientDirect Journal of Pervasive and Mobile Computing*. **2016**, 33, pp. 91-107.
38. Twitter. https://dev.twitter.com/rest/public. (Accessed on 15 December **2016**)
39. Foursquare. https://developer.foursquare.com/docs/. (Accessed on 15 December **2016**)
40. Yuyun, Dewancker. B.J. Analysis of Urban Population Using Twitter Distribution Data: Case Study of Makassar City, Indonesia. *In World academic of Science, Engineering and Technology*. **2016**, 10, 1627-1631.
41. Wang, Y.; Wang, T.; Tsou, M.H.; Li, H.; Jiang,W.; Guo, F. Mapping Dynamic Urban Land Use Patterns with Crowdsourced Geo-tagged Social Media (sina-weibo) and Commercial Points of Interest Collections in Beijing, China. *Sustainability*. **2016**, 8, 1202.

Chapter 3

Research Area and Data Sources

3.1 Location Study

3.1.1 Deciding the City Study

In the national development plan, Indonesia is divided into two regions, Eastern Indonesia (KIT) and Western Indonesia (KIB) [12]. The city of Jakarta is the centre of KIB and Makassar city is central to the development of KIT. Thus, this is the first reason for selecting Makassar City as the object of this the study. With regard to demographics, Makassar has the largest population in Eastern Indonesia. The ease of data access and the ability to identify objects around the city lend themselves to Makassar being the city of choice for this study.

Figure 3.1 Makassar City in the south of the Sulawesi region

Source: *www.google.co.jp/maps/place/Makassar,+Makassar+City,+South+Sulawesi,+Indonesia*

3.1.2 An Overview of Makassar City

Geographically Makassar City lies on west coast of South Sulawesi at the coordinates 119°18'27,97" 119°32'31,03" east longitude and 5°00'30,18"-5°14'6,49" south latitude. It has an area of 175.7 km^2 with the following boundaries:

1. North side: Pangkajene Islands District
2. South: Gowa District
3. East: Maros District
4. West: Makassar Strait

In general, Makassar City is a coastal area located 0.5-10 meters above sea level.

In addition to having a mainland area, Makassar City also has an archipelago along the coastline. The island is part of two sub-districts, namely Ujung Pandang and Ujung Tanah. These islands, as many as 12, are a group of coral islands. They consist of Lanjukang, Langkai, Lumu-Lumu, Bone Tambung, Kodingareng, Barrang Lompo, and Barrang Island. Administratively this city consists of 14 sub-districts and 143 urban villages. The northern parts of the city are the Biringkanaya, Tamalanrea, Tallo and Ujung Tanah sub-districts. The southern part consists of the Tanalatte sub-district and Rappocini. The eastern parts of the city are Manggala and Panakkukang. Finally, the western part consists of the Wajo, Bontoala, Ujung Pandang, Makassar, Mamajang and Mariso sub-districts. The administrative map of the 14 sub-districts of Makassar City can be seen in the Figure 3.2, and the details related to the area of each sub-district are as follows:

Table 3.1 Area percentage by sub-district, 2016

Area code	Sub-District	Area (km^2)	Percentage of Sub-District Area (%)
10	Mariso	1.82	1.03
20	Mamajang	2.25	1.27
30	Tamalatte	20.21	11.42
31	Rappocini	9.23	5.21
40	Makassar	2.52	1.42
50	Ujung Pandang	2.63	1.49
60	Wajo	1.99	1.12
70	Bontoala	2.10	1.19
80	Ujung Tanah	5.94	3.36
90	Tallo	5.53	3.12
100	Panakkukang	17.05	9.63
101	Manggala	24.14	13.64
110	Biringkanaya	48.22	27.24
111	Tamalanrea	31.84	17.99
111	Sangkarang Island	1.54	0.87
7371	Makassar City	175.77	100.00

Source: RT-RW Makassar City, 2015-2034

3.1.3 Demography

The main source of population data is the census. Nationally, since Indonesia became independent in 1945, a population census which is conducted every ten years, has been held six times

(1961, 1971, 1980, 1990, 2000, and 2010). In the population census, data collection is conducted throughout the territory of Indonesia as well as for foreign citizens. The method used is an interview between an officer of the census and a permanent resident—an individual who is usually resident. Officers go directly to where the respondents live and conduct the interviews there. For non-permanent residents, where the officers found them is noted. In addition to the population census, every year the population is recorded by the local government. For Makassar municipality in 2016, 1,469,601 people—742,287 males and 727,314 females—were recorded. The ratio of the male population to female is 9:8.

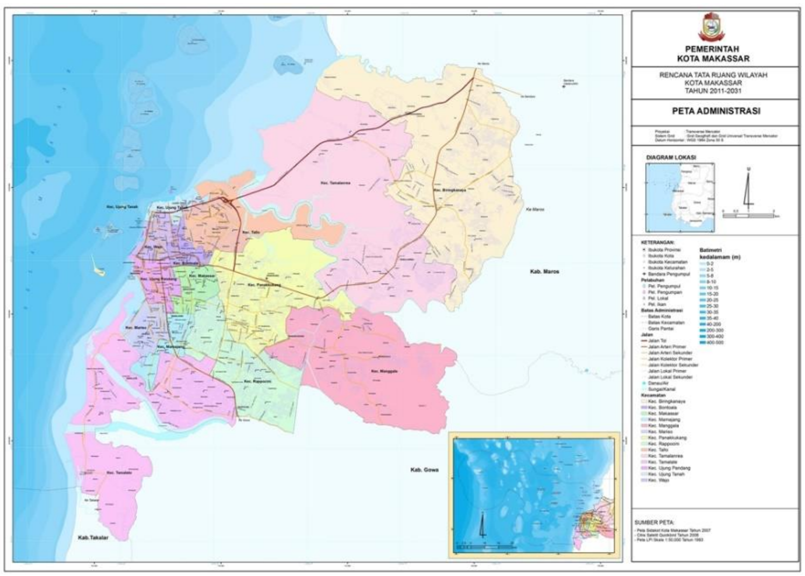


Figure 3.2 Makassar City administrative map
Source: RT-RW Makassar City, 2011–2031

The population spread in Makassar City according to the sub-districts shows that residents are concentrated in the Biringkanaya sub-district which has 202.520 inhabitants, or 13, 78% of the total population. This is followed by the Tamalate sub-district with 194.493 people (13, 23%). The Rappocini sub-district is about 164.563 people (11, 20%) and the lowest population number is in the Ujung Pandang sub-district which has around 28.497 people (1, 94%). In terms of population density, the Makassar sub-district is most densely populated at 33.634 inhabitants per km^2, then the Mariso sub-district (32.578 inhabitants per km^2, followed by the Mamajang sub-district (27.114 per km^2). The district of Tamalanrea is the district with the lowest population density of 3.522 people per km^2. The sub-district of Biringkanaya has 4.199 inhabitants per km^2, Manggala has 5.743 inhabitants per km^2, the Ujung Tanah sub-district has 8.286 per km^2), and the Panakkukang sub-district has 8.667 people per km^2 (see Table 3.2).

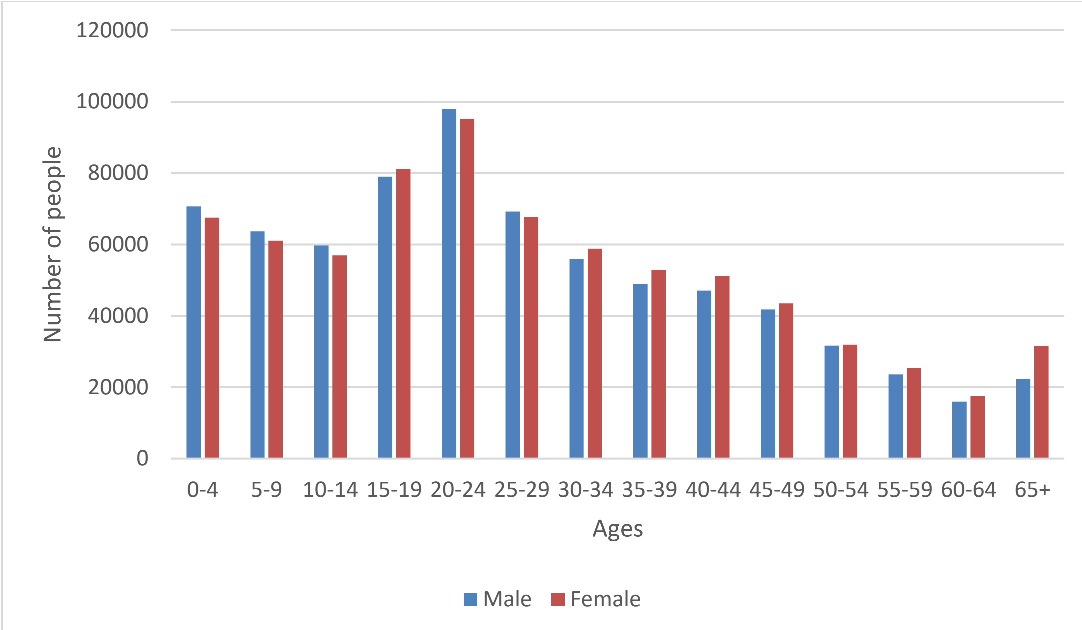


Figure 3. 3 Population pyramid by Age Group and Sex
Source*: BPS Makassar City, 2016*

Table 3. 2 Population number and density by sub-district, 2016

No	Sub-district	Population	Percentage	Area (km^2)	Density
110	Biringkanaya	202520	13.78	48.22	4200
30	Tamalate	194493	13.23	20.21	9624
40	Rappocini	164563	11.20	9.23	17829
100	Panakkukang	147783	10.06	17.05	8668
90	Tallo	139167	9.47	5.53	25166
101	Manggala	138659	9.44	24.14	5744
111	Tamalanrea	112170	7.63	31.84	3523
50	Makassar	84758	5.77	2.52	33634
20	Mamajang	61007	4.15	2.25	27114
10	Mariso	59292	4.03	1.82	32578
80	Bontoala	56536	3.85	2.1	26922
80	Ujung Tanah	49223	3.35	5.94	8287
70	Wajo	30933	2.10	1.99	15544
60	Ujung Pandang	28497	1.94	2.63	10835
	Total	1.469.601	100%	175.47	229.668

Source: BPS/Indonesia Population Projection 2011–2035, BPS

Table 3.3 Population by Age Group and Gender in Makassar Municipality

Age group	*Male*	*Female*	*Male+Female*
0-4	70626	67503	138129
5-9	63647	61087	124734
10-14	59704	56957	116661
15-19	79016	81117	160133
20-24	97985	95241	193226
25-29	69180	67707	136887
30-34	55959	58771	114730
35-39	48957	52927	101884
40-44	47053	51121	98174
45-49	41816	43511	85327
50-54	31661	31932	63593
55-59	23543	25364	48907
60-64	15956	17597	33553
65+	22210	31452	53662
Makassar City	727313	742287	1469600

Sourced: BPS Makassar, 2016

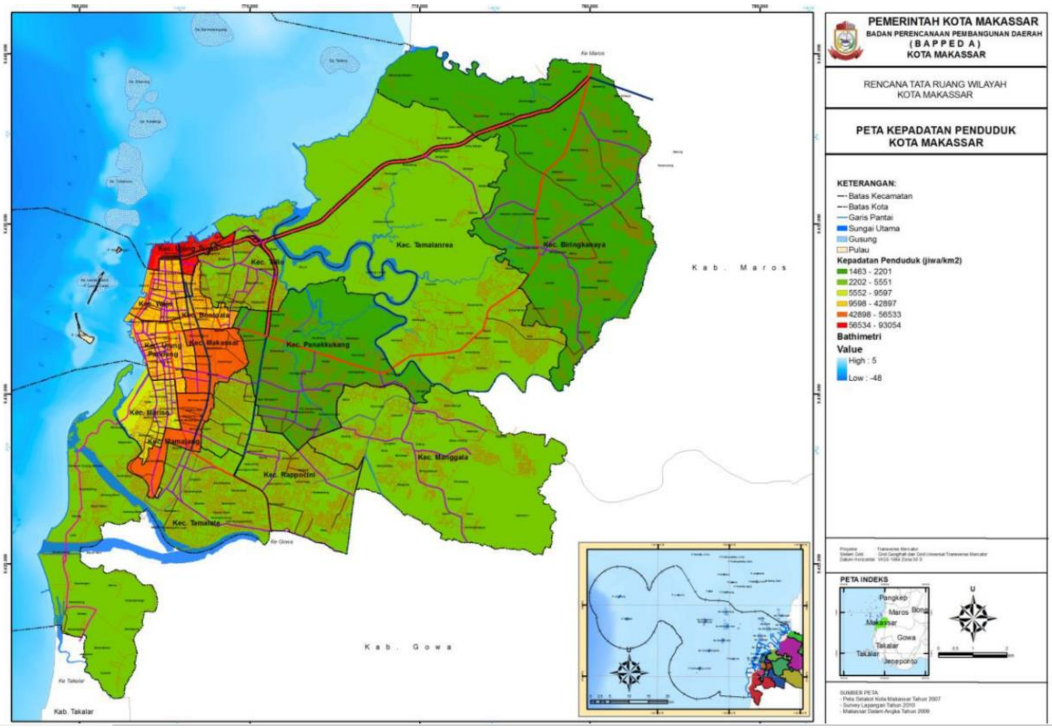


Figure 3. 4 Map of population density

Source: RT-RT Makassar City

3.1.4 Land Use

In general, the spatial pattern of Makassar City consists of protected and cultivated areas.

A. Protected area

Substantially, the determination of protected area includes the following areas:

1. Green open Spaces

Green areas in the city of Makassar are divided into two parts, namely the protected green area and the green zone. The protected green area is a part of green areas that have natural characteristics which need to be conserved for local habitat protection as well as for the protection of the larger area. The green zone is part of a green area outside of the protected green area which is use for greening through planting, development and maintenance.

Table 3. 4 The condition of green open space

No	Sub-distirict	Hec-tare	The Existing of Green Open Space							Green Open Space (Hectare)	Percen-tage
			City Forest	Green Line	Field	Park	Cemetery	Mangg-roves	Sempadan		
1.	Biringkanaya	4822	62,9306	8,6414	69,1640	52,3041	14,3822	10,0990		217,52	1,24
2.	Bontoala	210		0,4521		4,6963	1,1620			6,31	0,03
3.	Makassar	252		2,6342	0,2935	3,8700	1,8758			8,67	0,05
4.	Mamajang	225		0,1474	0,2597	1,7398	4,4404			6,59	0,04
5.	Manggala	2414			11,7922	2,0649	37,0512		4,7029	57,35	0,33
6.	Mariso	182	0,5438	1,9251	5,0202	2,0393				9,53	0,05
7.	Panakkukang	1705	17,9466	8,9194	13,7499	9,54	13,3391			63,50	0,36
8.	Rappocini	923		9,3156	3,8255	3,0930	1,2459			17,48	0,10
9.	Tallo	583		4,3992	3,9216	7,1144	13,1018	364,0627		391,85	2,23
10.	Tamalanrea	3184	44,5131	16,1707	9,8345	7,3920	5,3108	20,9905	74,5290	105,10	0,60
11.	Tamalate	2021	0,7581	6,4276	11,2939	2,3399	6,4056		161,8264	187,71	1,07
12.	Ujng Pandang	263		2,9813	8,4631	4,4419				15,89	0,09
13.	Ujung Tanah	594		4,2440	3,1506	1,55	0,3145			9,26	0,05
14.	Wajo	199		1,1607	0,0157	0,7288	0,0320			1,94	0,01
	Total	**17.577**	**44,52**	**18,45**	**119,18**	**53,78**	**65,00**			**1.098,7**	**6,25**

Source: RTRW Makassar city, 2014

Green open spaces in Makassar City are very limited; hence, the government decided that 30% of the total land should be green areas. Table 3.4 is a grouping of green land use in Makassar City. The concept of extensive development of green open space in Makassar City is divided into 3 areas: the city area that has been built, the city area that has not yet been built, and a reclamation area. The existing use of green open space can be seen in Figure 3.4.

2. Cultural Heritage Areas

The cultural heritage areas in Makassar City have spread in some parts of the city. In these areas, there are buildings or ancient cultural heritage sites that require preservation. Cultural heritage areas in Makassar City include non-building environments and buildings which consist of:

a Raya Mosque Building (Figure 3.5) located in the sub-district of Bontoala

b Somba Opu Fortress (Figure 3.6) located in the Tamalate sub-district

c Kings of Tallo Cemetery (Figure 3.7) in the Tallo sub-district

d Fort of Rotterdam (Figure 3.8) in the Ujung Pandang sub-district

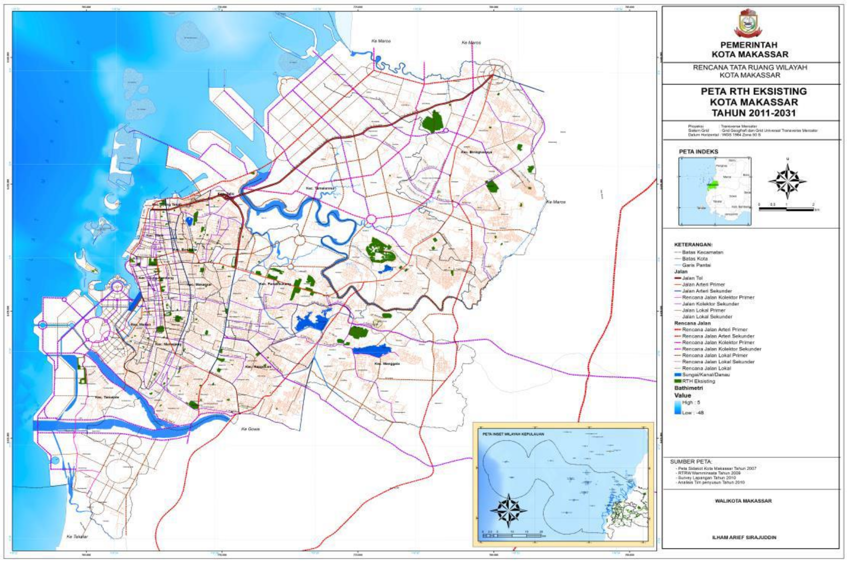


Figure 3. 5 Map of green open space
Source: RT-RW Makassar City

Figure 3. 6 Somba Opu Fortress

Source: www.google.co.jp/maps/place/Benteng+Somba+Opu

Figure 3. 7 Kings of Tallo Cemetery
Source: www.google.co.jp/maps/place/Kings+of+Tallo+Cemetery

Figure 3. 8 Rotterdam Fort
Source*: www.google.co.jp/maps/place/Fort+Rotterdam*

B. Cultivation area

A cultivation area is an area designated for the preservation and development of natural resources, human resources, and artificial resources. The cultivation areas include:

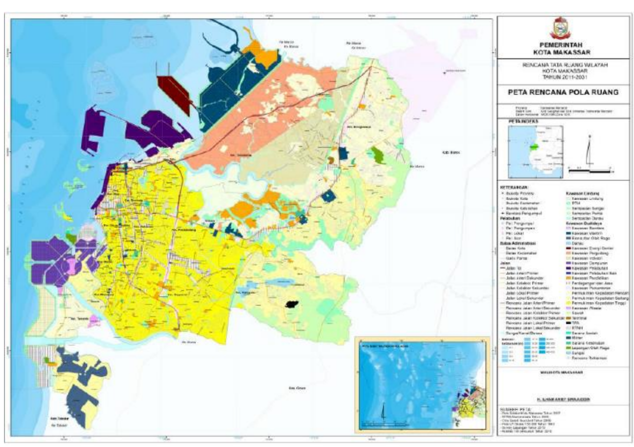


Figure 3.9 Map of space pattern palling
Source: RT-RW Makassar City

1. Housing Areas

Based on a spatial plan for the development of the residential area of Makassar city, the housing areas are divided into three categories, namely high, medium and low-density areas.

a. High-density housing areas include: Bontoala sub-district, Makassar sub-district, Mamajang sub-district, part of Mariso sub-district, part of Panakkukang sub-district, part of Rappocini sub-district, Tallo sub-district, Tamalate sub-district, Ujung Pandang sub-district, Ujung Tanah sub-district and Wajo sub-district.

b. Medium-density housing areas include: part of the Biringkanaya sub-district, part of the Manggala sub-district, part of the Mariso sub-district, part of the Panakkukang sub-district, part of the Rappocini sub-district, part of the Tallo sub-district, part of Tamalanrea, and part of the Ujung Tanah sub-district.

c. Low-density housing areas include: part of the Biringkanaya sub-district, part of the Manggala sub-district, part of the Panakkukang sub-district, part of the Tallo sub-district, part of the Tamalanrea sub-district, part of the Tamalate sub-district, part of the Ujung Pandang sub-district, and part of the Ujung Tanah sub-district.

d. Trading Areas

The trading areas in Makassar city consist of traditional markets, shopping centers and modern stores. Areas within this category are located in the Mariso sub-district.

e. Office Areas

The office areas of Makassar city include government offices at the province level and sub-district level, and private offices.

f. Industrial Areas

The industrial areas consist of large industrial estates located in the Biringkanaya and Tamalatte sub-district, medium industrial areas, and small industrial areas located in the Ujung Pandang sub-district.

g. Warehousing Areas

The warehousing areas are located in the Biringkanaya and Tamalandrea sub-districts. The development plan for warehousing areas consists of warehousing at port areas, warehousing at airports, and warehousing at maritime areas.

h. Tourism Areas

The tourism areas of Makassar city include areas of cultural tourism (Benteng Fort Rotterdam, Benteng Somba Opu, Makam Raja-Raja Tallo, Makam Pangeran Diponegoro, Emmy Saelan monument, City Museum, Masjid Raya, Katedral church, Klenteng Ibu Agung Bahari, and China Town), natural tourism areas which consist of Losari beach, Akkarena beach, Kayangan

Island, Samalona Island, Kodingareng Keke Island, and La'jukang Island, and artificial tourism areas. Makassar City's spatial pattern can be seen in its entirety in the following figure:

3.2 Data Collection Method

Social media has experienced dramatic growth over the past few years. It can provide space to expand our social relationships. Many new features are added to enhance the convenience for social media users; one of these is the location feature. The location feature helps social media users (Facebook, Twitter, and Foursquare) to disclose user location by posting it on their status update.

3.2.1 Twitter Dataset

To gather tweet data, the author used the Twitter streaming application program interface (TS-API). It is an application window that allows developers to access the program. The Twitter API provides various information related to a tweet's attributes than can be accessed freely, such as tweet a new tweet, read the author profile, follower data, time zone and location information. The twitter API serves as a link between the systems built with Twitter. The Twitter API requires a consumer key, consumer access, an access token and secret access tokens obtained by registering for the Twitter API app at http://dev.twitter.com. Next, searching the data from Twitter is done via two means: the user and the keyword criteria. The system built using PHP programming language with searching data is based on both the user and keywords. In the last process, the data was downloaded and saved into databases for tweet and user.

Our final Twitter data set consists of 42 days (six weeks) starting from August 24th until October 5th, 2016. It is constructed using Indonesian language tweets which were filtered to reveal those tweets that contain the geolocation. In this study, 253,968 data records from Twitter were analysed. In the next step, only data that contained the user's location was processed.

Table 3.5 Dataset Detail

Original dataset	Total
Number of users	25.346
Number of check-ins	211.922

To facilitate the collection of data, it was necessary to build an application for collecting and storing the tweets' information. The application architecture can be seen in Figure 3.10. Briefly, the information from the tweets is collected using the following steps:

1. When the application is executed, the system will automatically download the data from Twitter.

2. The system created will identify the Twitter data using the web service API with location coordinate parameters as provided by the developer (see Figure 3.13).
3. When the parameter is identified by the Twitter server, the data will be displayed in the application.
4. To facilitate the exchange of data between the Twitter server and client (application build), Twitter provides data in JSON format. In this case, JSON acts as a carrier of data and information required by the client.
5. When the data is connected to the system, then the application will update automatically and save the information with the latest changes.

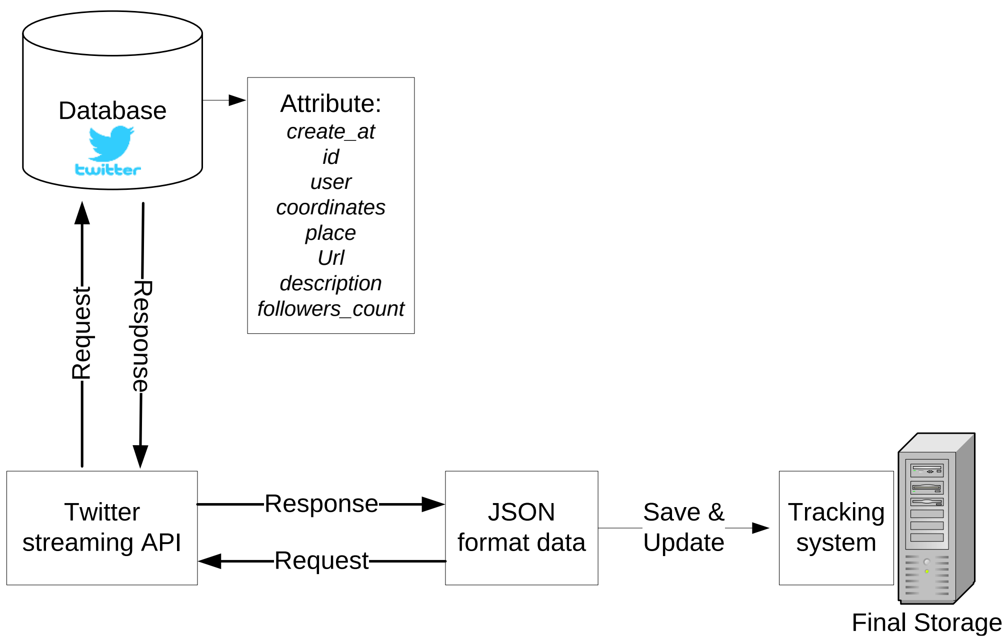


Figure 3.10 Application architecture

Figure 3.11 shows the Twitter user profile page. When the user conducts a status update, some attributes are automatically attached to the uploaded status such as such as time, date, day, and location. Figure 3.12 shows the characteristics of the data displayed on the system which consists of:

1. User profile (user name) is the name on someone's Twitter profile that can be created with spaces and various punctuation.
2. User name (screen name) is the name used when a user is mentioned by other Twitter users such as city name (@makassar), building name (@hotel…), and person's name (@yuyun). The name is created without any spaces and various punctuation. Usually the user name distinguishes between individual users and other users.
3. Text is the user's status updates posted on Twitter.

4. Date and time are features on Twitter to display when a post occurs. Other than this, components also indicate the day on which a post is made.
5. Latitude and longitude features are geographic coordinate systems used by Twitter to determine the location at which a user is tweeting.

Figure 3.11 shows the Twitter user profile page. This page serves to display an individual's activity on Twitter. As such, when users give a status update, there are some attributes, such as time, date, day, and location, attached to the post. With millions of data produced, it is impossible to collect a user's information by manually copying and pasting it to other storage.

Figure 3.11 Characteristics of Twitter data

To facilitate data collection, it is necessary to build a system to track and catch the Twitter activities of users. The system which has been created works based on location coordinate parameters. The measurement point is cantered at kilometre 0 of Makassar City with a coordinate point at -5.131813, 119.405258 (Figure 3.12). The system performs a 360-degree search for the Twitter user within a radius of 25 km (Figure 3.13).

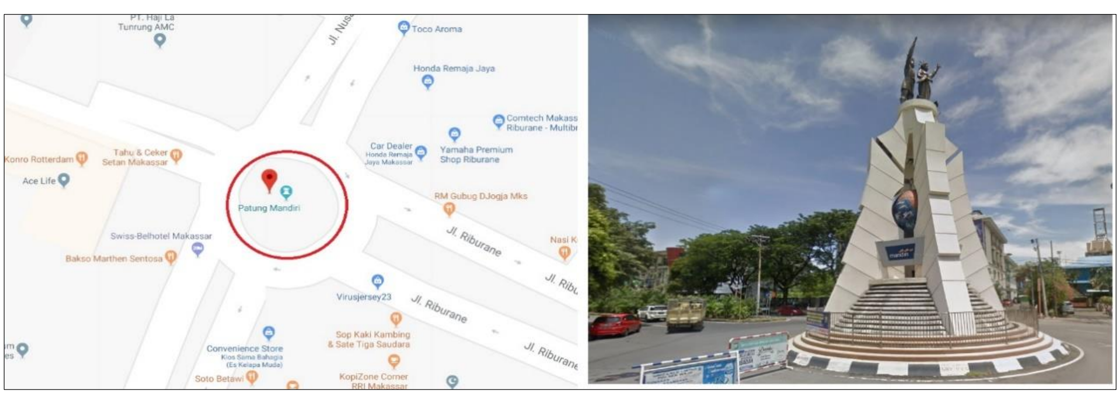


Figure 3.12 The Patung Mandiri (0 km of Makassar City)
Source: www.google.co.jp/maps/places/patungtuguh

Due to Makassar City not being globular, geographically speaking, tweet activity is sourced from outside the region's boundaries entered in the system's reach. To solve this problem, it is necessary to sort the data by analyzing the latitude and longitude coordinates. If the coordinate points come near to the center of measurement (-5.131813, 119.405258), this means that users are in the area of Makassar City. If they are away from the central point, then users are outside the city boundaries (see Figure 3.14).

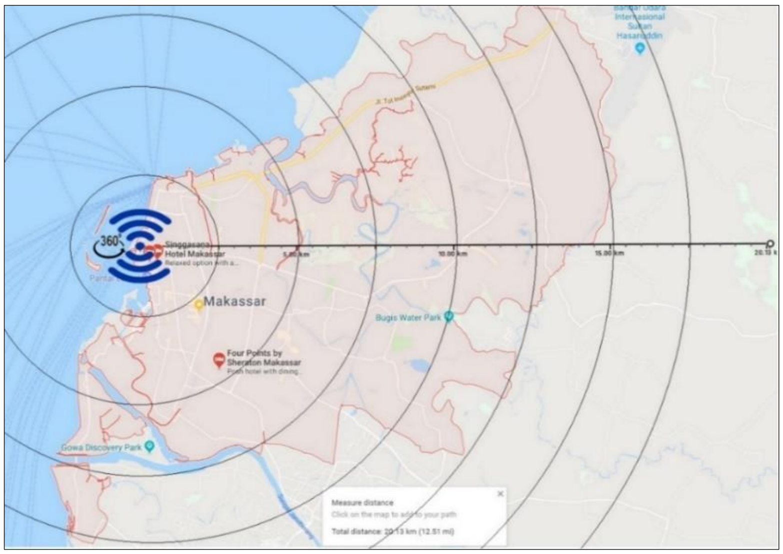


Figure 3.13 System range of the Twitter user search

user_name	Day and Date	Time	Latitude	Longitude
Nur Faida	Monday, August 29, 2016	5:53:52 AM	-4.97389747	119.5743435
Pramuka Smaga Maros	Thursday, September 1, 2016	12:00:10 AM	-4.97389747	119.5743435
IDRUS MADRIDISTA	Monday, September 5, 2016	11:43:45 PM	-4.98779419	119.5749523
Autumn	Thursday, August 25, 2016	2:06:22 PM	-4.99186	119.5721
GooGuns Lulz	Friday, August 26, 2016	8:36:00 PM	-4.99712207	119.9803754
Bang Zafran	Friday, August 26, 2016	2:06:36 PM	-4.99748	119.57355
NurulSelviani	Monday, September 19, 2016	11:22:30 AM	-4.99748078	119.5731557
THitye	Friday, September 9, 2016	2:45:37 PM	-4.99954	119.57238
Ahmad AQil Al Banna	Friday, August 26, 2016	9:41:33 PM	-5.00515	119.57147
Rijal M Ramli	Tuesday, August 30, 2016	9:42:13 PM	-5.00542	119.5[illegible]292
A R S T Y	Friday, August 26, 2016	4:18:03 PM	-5.00582	119.57371
ewithadwiaramadhania	Thursday, August 25, 2016	7:51:27 PM	-5.00586	119.61509
Cakra Usman Djiwang	Tuesday, August 30, 2016	11:53:57 AM	-5.00594415	119.6404669
Puang Koro	Tuesday, September 6, 2016	10:24:32 AM	-5.00633	119.57365
#BaraniaCoffee	Saturday, September 10, 2016	12:07:32 AM	-5.00674527	119.5769249
Nur Magfirah Ashri	Sunday, September 4, 2016	8:37:39 AM	-5.00683206	119.7709129
Mariana TH	Tuesday, September 20, 2016	10:13:15 AM	-5.00691405	119.5720568
latifah umrah	Friday, August 26, 2016	2:21:00 PM	-5.00694	119.572
Rani Soraya ?	Saturday, August 27, 2016	2:19:15 AM	-5.00694	119.572
betsy yosia	Tuesday, August 30, 2016	1:56:40 AM	-5.00694	119.572
Evelyn?	Monday, August 29, 2016	7:43:57 AM	-5.13448	119.52292
SUARDI PA'BA	Wednesday, September 7, 2016	9:21:22 PM	-5.1344828	119.4425689
riskaaisyh	Wednesday, August 24, 2016	3:01:25 PM	-5.13449	119.44498
firda nur annisa	Wednesday, August 24, 2016	1:10:02 PM	-5.13449	119.49428
Muli Pakila	Thursday, September 8, 2016	8:57:53 AM	-5.13449	119.49397
cita utami	Wednesday, September 14, 2016	12:57:46 PM	-5.13449	119.41503
Indra Rukmana Rusli	Wednesday, September 21, 2016	11:00:21 AM	-5.13451	119.4473
Hot 'n Cold	Thursday, September 8, 2016	12:03:07 AM	-5.13451251	119.41386
Sri Sahrianii	Thursday, August 25, 2016	4:43:18 PM	-5.13453	119.49862
a khaira o r	Monday, August 29, 2016	11:25:24 AM	-5.13453	119.41919
halija	Tuesday, September 6, 2016	1:15:38 PM	-5.13453	119.41888
Ade Isran Hidayat	Friday, September 2, 2016	11:39:51 AM	-5.13454	119.40653
St. Nurinra Mega M	Friday, August 26, 2016	4:33:34 PM	-5.13456	119.50245

Outside of city bondaries

Inside of city boundaries

Figure 3.14 Tweet activities inside and outside the city boundaries

The next step is filtering. This process aims to identify whether the users' check-in activities are in the city or outside Makassar City. Figure 3.15 shows maps of user distribution in different places before the filtering process. Then Figure 3.16 shows the spread of user check-in activities after the filtering step.

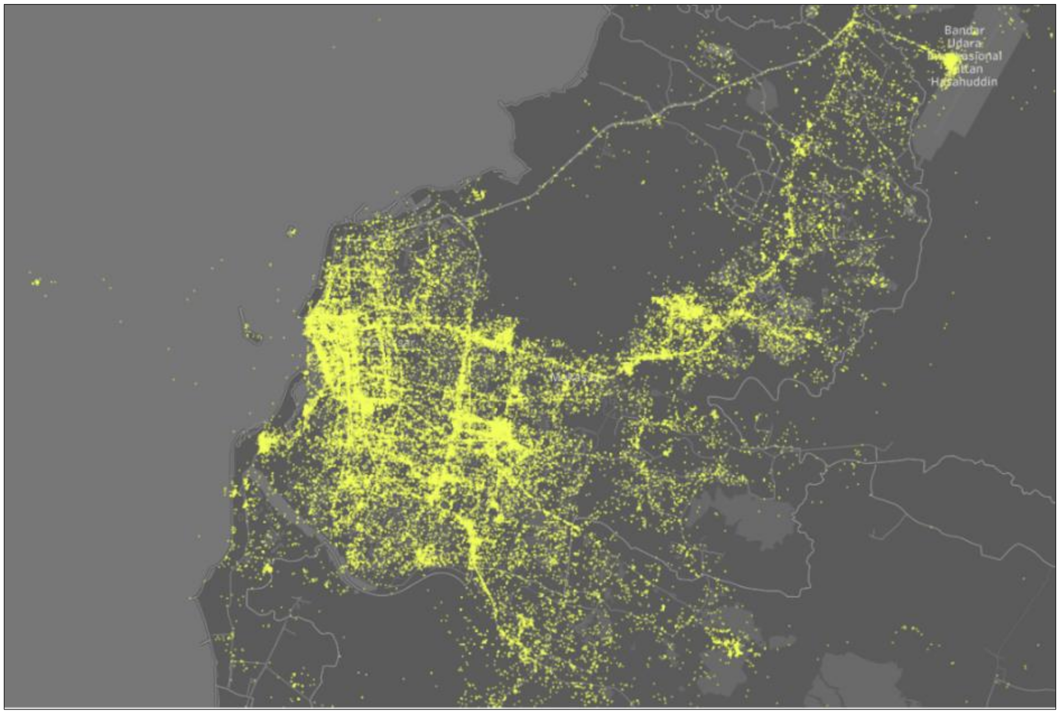


Figure 3.15 Check-in distribution data before the filtering process

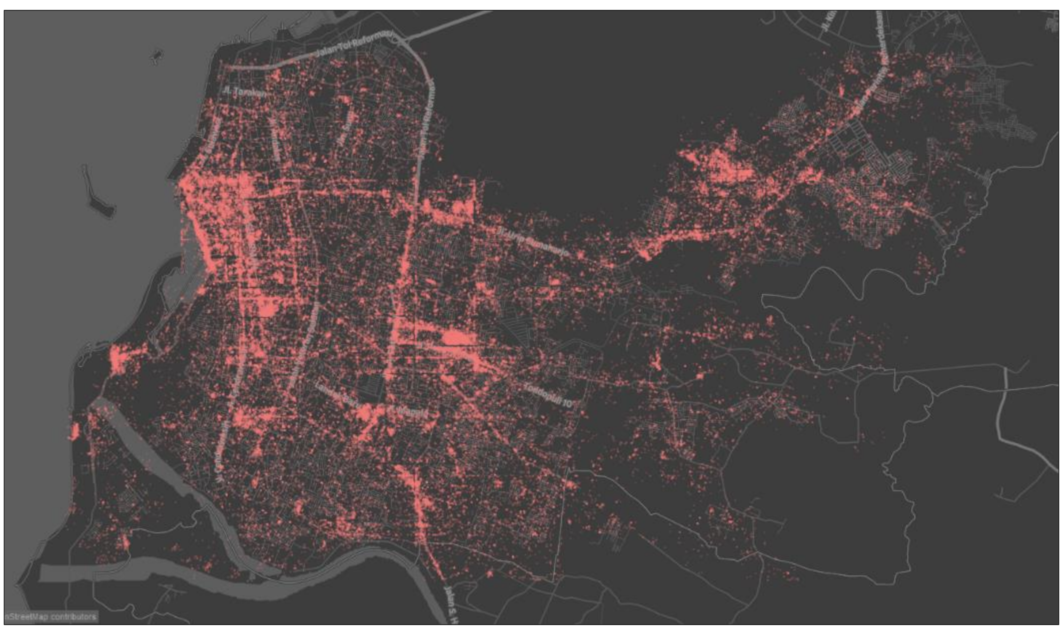


Figure 3.16 Check-in distribution data after the filtering process

References

1. BPS/Indonesia Population Projection 2011–2035, 2014.
2. RT-RW Makassar City, 2011-2031, 2014.
3. Central Bureau Statistic. Makassar in Number, 2016.
4. Makassar city map. Available online: www.google.co.jp/maps/place/Makassar // (accessed on 13 March 2018).
5. Rotterdam Fort. Available online:www.google.co.jp/maps/place/Fort+Rotterdam (accessed on 13 March 2018).
6. Kings of Tallo Cemetery. Available online: www.google.co.jp/maps/place/Kings+of+Tallo+Cemetery (accessed on 13 March 2018).
7. Somba Opu Fortress. Available online: www.google.co.jp/maps/place/Benteng+Somba+Opu (accessed on 13 March 2018).

Chapter 4

Social Media as an Indicator to Validate the Existence of People, Correlating Twitter with Official Census Data

4.1 Deciding on Twitter as a Source of Data Research

Today almost all social media has added a geolocation feature to attract users. Twitter was chosen as the single social media application to study due to the company being able to provide access data free for individuals and developers through their API (application program interface) facility. Other social media (e.g., Facebook, Instagram, Path, Foursquare, and others) due to privacy factors are not able to share publicly. Despite the Twitter data covering a large population, there are some open issues on a Twitter user's profile (such as age, gender, and occupation). Because the Twitter data cannot cover the whole range of user profiles, the author combined the Twitter data from the tracking results and the questionnaire data to obtain the Twitter user profile.

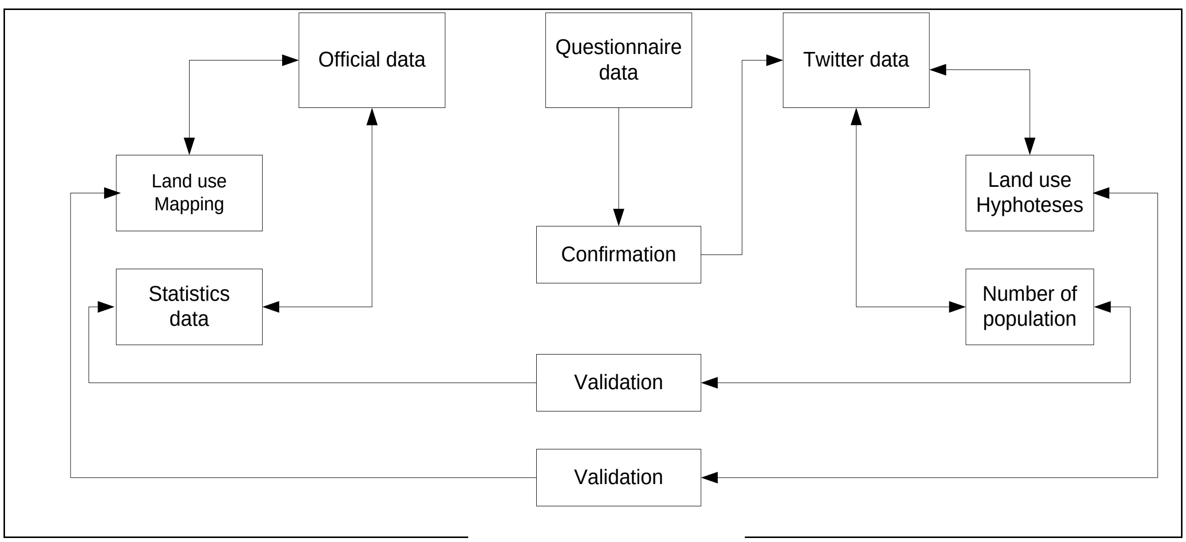


Figure 4.1 System Flow

This chapter aims to explore the relationship between Twitter geolocation and the existence of people in the city of Makassar with the two main focuses. Firstly, the study will investigate the spread of people in a particular area, specifically highlighting the spread of Twitter users in every sub-district. As validation, the author used local data released by the statistics department of the Makassar City local government to compare with the Twitter data as shown in Figure 4.1. The second objective is to define the land function based on the pattern of individuals' visits. To validate these findings, the land-use mapping released by the city planning department of the Makassar local government was used in this research, as well as a distribution frequency analysis to learn the percentage of visitors.

4.2 Questionnaire Data Collection

In addition to the Twitter data collection questionnaires were also used. The target of the survey was to identify and validate the profile of Twitter users (e.g., gender, age, occupation, etc.) The questionnaire was prepared in two parts. Part 1 consisted of 6 questions to reflect the respondent's profile, including gender, age, occupation, domicile status, residence status, sub-district domicile, and use of social media. Part 2 consisted of 4 questions covering the attributes concerning the types of social media, such as kinds of social media used, duration of use, and place for conducting social media activity.

4.2.1 Part 1 Profile

Table 4. 1 Distribution of demographic respondents

Attributes	N=205	Number	Percentage	Total
Gender	Male	132	67%	100%
	Female	73	33%	
Age	<15	3	1%	100%
	15-19	22	11%	
	20-29	121	59%	
	30-39	48	24%	
	40-49	9	4%	
	>50	1	0.5%	
Occupation	Student	70	34%	100%
	Teacher	36	18%	
	Employee	61	30%	
	Entrepreneur	22	11%	
	Others	15	15%	
Sub-district living area	Tamalandrea	36	18%	100%
	Biringkanaya	28	14%	
	Panakkukang	26	13%	
	Rappocini	21	10%	
	Manggal	20	10%	
	Tamalate	15	7%	
	Tallo	14	7%	
	Makassar	13	6%	
	Mamajang	7	3%	
	Wajo	7	3%	
	Bontoala	6	3%	
	Mariso	5	2%	
	Ujung Pandang	5	2%	
	Ujung Tanah	2	1%	

Profile information was used to understand the background of a respondent as it relates to the profile of a user of social media. The questionnaire was made using the google form application. Distribution was through an online survey spread by social media discussion groups. It was also sent to relatives and forwarded to others. A total 205 respondents answered the questionnaire as shown in Table 4.1.

The type of question related to a respondent's profile is shown in the figure below:

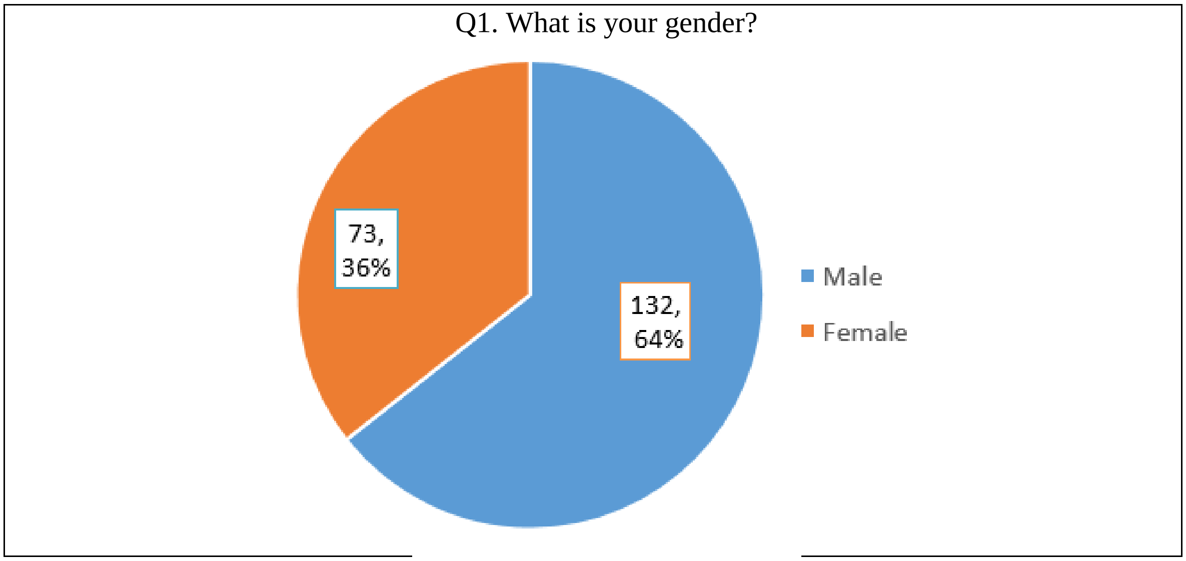


Figure 4.2 Distribution of gender

Figure 4.2 shows the distribution of gender (total number of respondents, N = 205). Male respondents have a high level of contribution (67%), and female respondents have a lower one (33%).

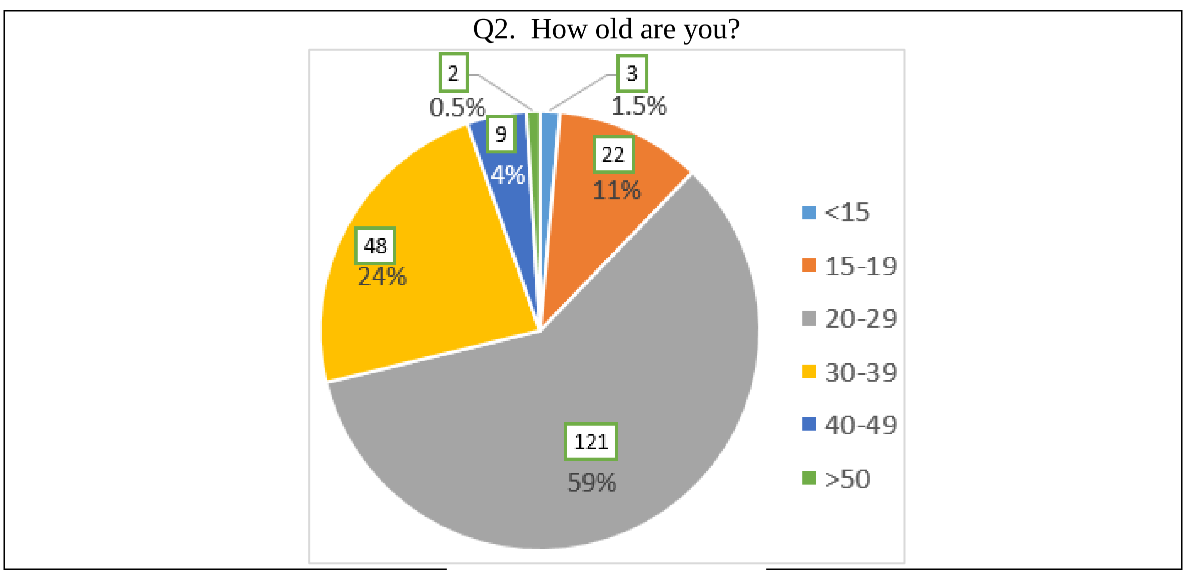


Figure 4.3 Age distribution

Figure 4.3 shows the distribution of age. Respondents were grouped into six age categories—under 15, 15–19, 20–29, 30–39, 40–49 and more than 50 years old. The chart shows that the highest number of respondents were in the 20–29 age group (59%), 23.5% of respondents were 30–39, about 10.8% of respondents were 15–19, 4.4% of respondents were 40–49 about, under 15 comprised 1.5% of respondents, and those older than 50 comprised about 1% of respondents.

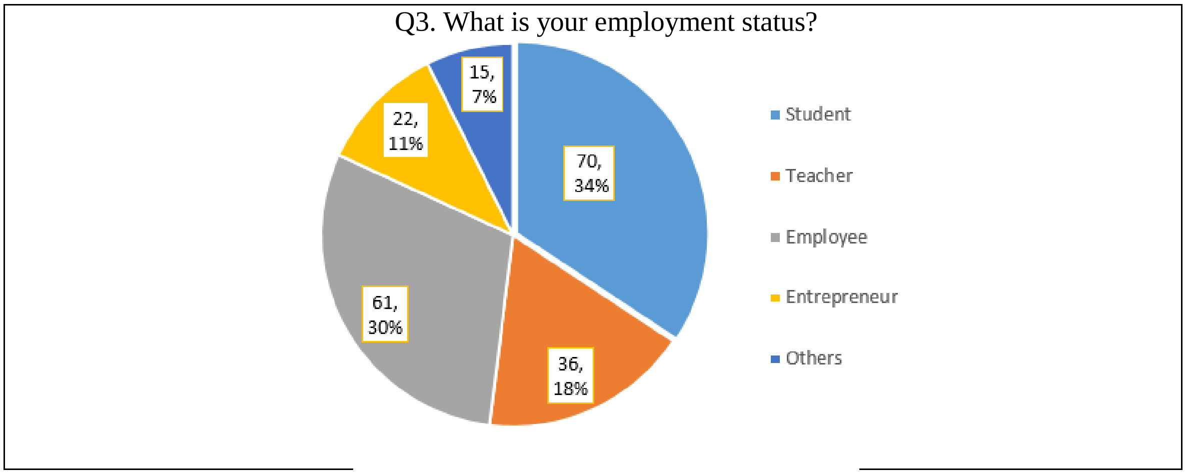


Figure 4.4 Distribution of employment status

Figure 4.4 shows the percentage chart of employment status. The author found that the respondents were students, teachers, employees, entrepreneurs, housewives, and others. In this regard, 34% of respondents were students, 30% were employees, 18% were teachers, 11% were entrepreneurs, 2% were housewives and 5% were others (e.g., unemployed, freelance, and no answer).

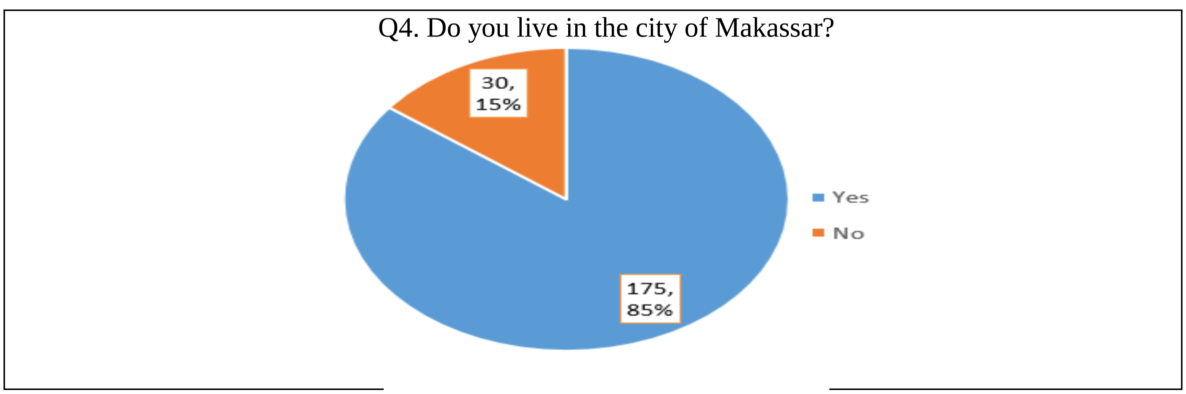


Figure 4. 5 Distribution of living status

Figure 4.5 shows the living status of respondents, where 85% of respondents lived in the city and 15% outside the city. Fifteen percent of respondents were someone who resided in another city and who came for specific purposes, such as commuters for work.

Q5. What is your resident status in Makassar city?

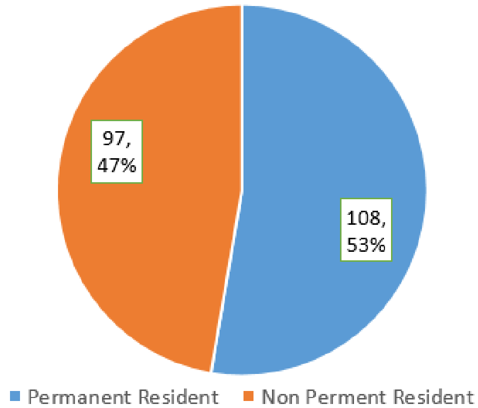


Figure 4.6 Distribution of resident status

Fifty-three percent of respondents were permanent residents and 47% were non-permanent residents. Non-permanent residents were those who were living there for a certain period of time, such as students and workers.

Q6. In which sub-district do you live?

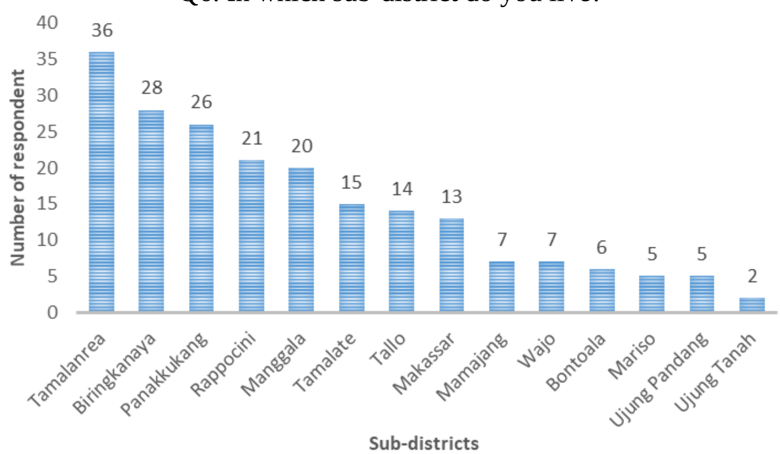


Figure 4.7 Graph distribution of sub-districts

Figure 4.7 shows the distribution of respondents based on sub-districts. The majority of respondents (18%) lived in the Tamalandrea sub-district, 14% in Bringkanaya, 13% in Panakukkang, 13% in Rappocini, 10% in Manggala, 7% in Tamalate, 7% in Tallo, 6% in Makassar, 3% in Mamajang, 3% in Wajo, 2% in Mariso, 2% in Ujung Pandang, and 1% and Ujung Tanah. From the questionnaire results it can be concluded that 18% of respondents from the Tamalanrea sub-district were students.

4.2.2 Part 2-Social Media

In the second part of the questionnaire user activities on social media were discussed. This part was designed to understand the characteristics of the social media used, and the frequency and place where social media activity occurs. The kinds of question and answers can be seen in the following figures:

Q1. What kind of social media are you currently using? You can choose more than one answer.

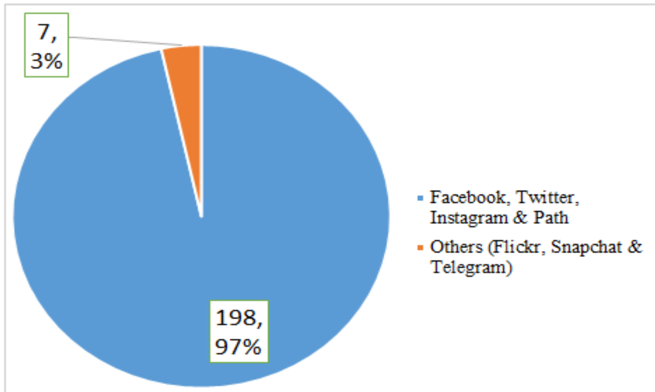


Figure 4.8 Kinds of social media used

The first question offered several options as responses. Each respondent was able to choose more than one answer. Based on questionnaire data, it was found that, on average, each respondent had more than one type of social media. These social media can be seen in Figure 4.8 where 97% respondents answered that they have Facebook, Twitter, Instagram, and Path. The percentage of respondents who said they have social media other than those mentioned above was only 3%.

Q2. How long have you been using social media?

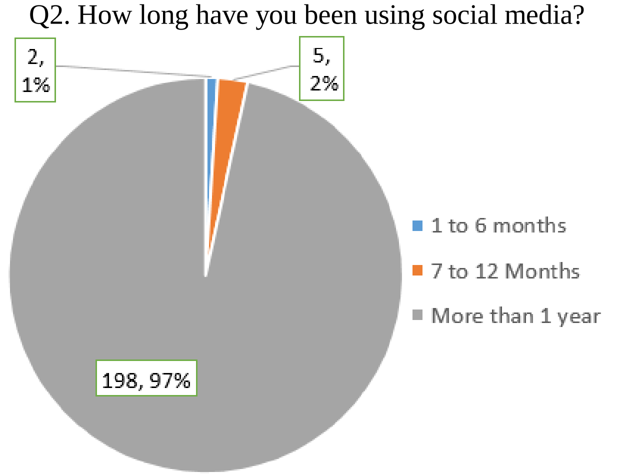


Figure 4.9 Pie chart distribution of time using social media

The second question asked how long the respondent had been using social media. The purpose of this question was to know the respondent's traces against the use of social media, and to ascertain whether they were new or old users. This will allow the author to find out how important social media is to them.

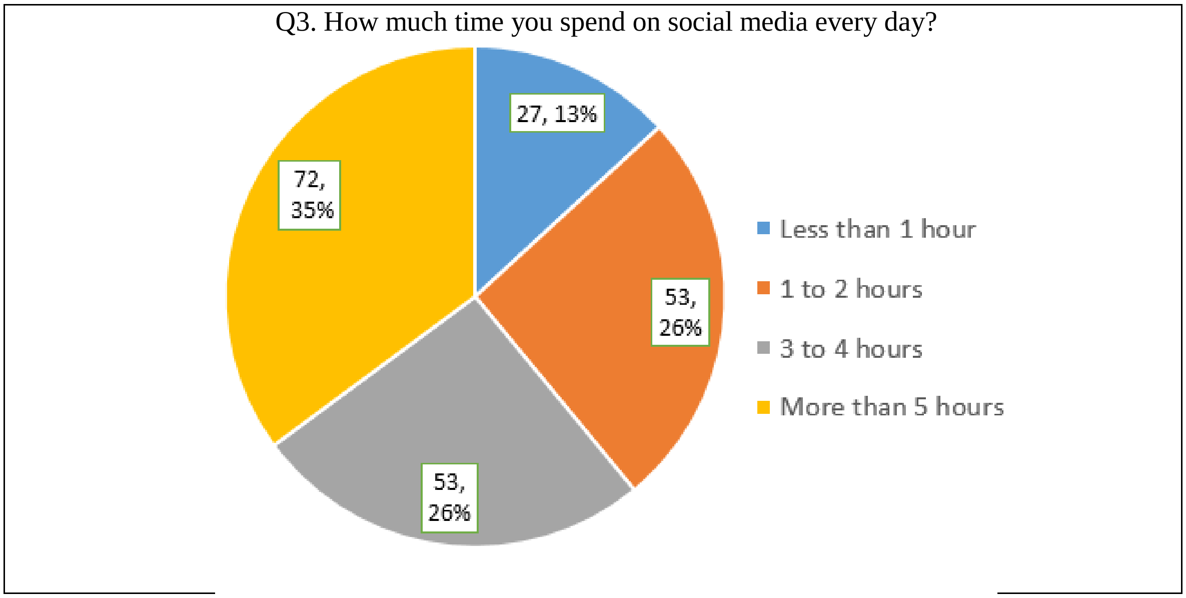


Figure 4.10 Graph distribution of the frequency of social media usage

Graph 4.10 shows the frequency of social media usage in respondents' daily lives. Answers are divided into four times groupings. Each respondent can select from the several answers provided. Thirty-five percent of respondents were engaged in social media activity more than 5 hours per day, 26% were engaged in it for 3 to 4 and 1 to 2 hours, and only 13% selected less than 1 hour. From that data, the author argues that social media has become an import part of their lives.

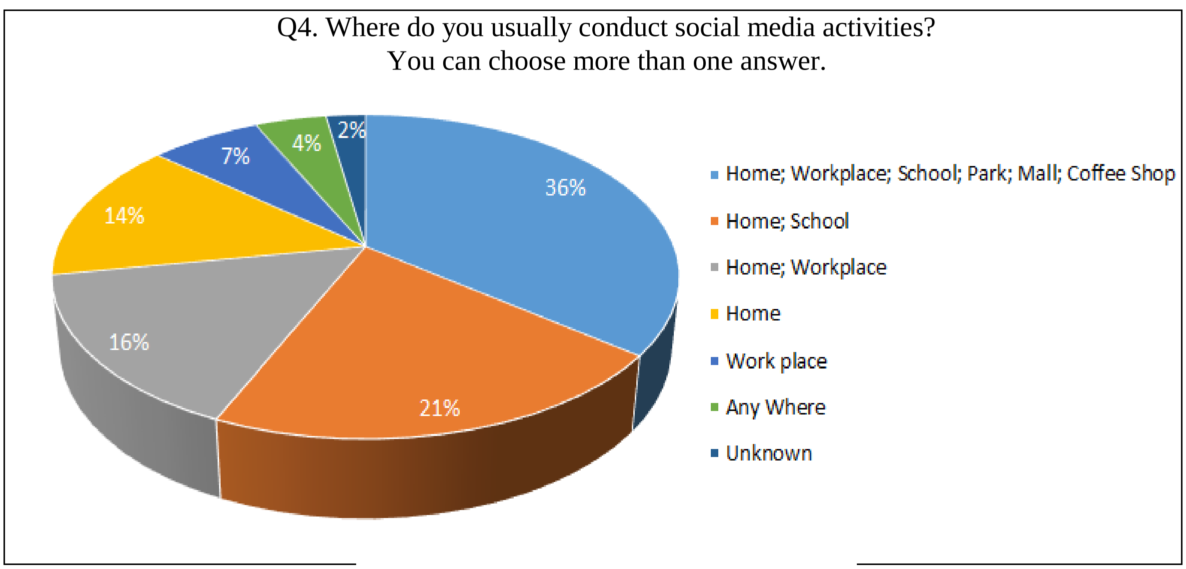


Figure 4.11 Graphic of distribution of places

Figure 4.11 shows the distribution of places. In this question, author asked where respondents usually conducted social media activities. Thirty-six percent of respondents answered home, workplace, school, park, mall, or coffee shop; 21% said at home and school; 16% at home and workplace; 14% at home; 17% at workplace; 4% said anywhere; and 2% said unknown.

4.3 Comparison of Results of Questionnaire Survey, Population, and Twitter Data

To improve accuracy in measurement, the author also used population data and questionnaire results as additional data (see Table 4.2). Both sources of data were used to compare the number of Twitter users with a population in each sub-district. It is, furthermore, important to know the size of the population inhabiting a region; a combination of these 3 data sources can be used to analyze land-use functions.

Table 4. 2 Comparison of each data source between Makassar city population [1], Questionnaire, and Twitter

Sub-Districts	Area Km^2	Population	Average	Questionnaire	Average	Twitter	Average
Biringkanaya	48.22	190802	14%	27	14%	10266	8%
Bontoala	2.1	55910	4%	6	3%	4200	2%
Makassar	2.52	84014	6%	13	7%	9000	3%
Mamajang	2.25	60236	5%	7	4%	8400	3%
Manggala	24.14	127915	10%	20	10%	18000	7%
Mariso	1.82	56547	4%	5	3%	24322	9%
Panakukkang	17.05	56610	4%	25	13%	52580	20%
Rappocini	9.23	160499	12%	20	10%	19400	7%
Tallo	8.75	135216	10%	13	7%	2712	1%
Tamalandrea	31.86	109471	8%	35	18%	28720	11%
Tamalate	20.21	186921	14%	15	8%	35768	14%
U Pandang	2.63	27141	2%	5	3%	33000	12%
Ujung Tanah	5.94	48531	4%	2	1%	2200	1%
Wajo	1.99	31947	2%	7	4%	5400	2%
Total	175.47	1.331.760	100%	205	100%	253.968	100%

From the questionnaire results, an analysis was conducted to determine the Twitter users' profiles. Our initial assumptions show that the age of users is in the range of 15 to 50 years old, and the average is dominated by the 20 to 39-year-old age group. The gender category is 67% male and 33% female. For respondents' occupations, this was mostly dominated by students and employees, at 34% and 30% respectively.

Figure 4.13 illustrates a visualization of users in different locations. Each coordinate area is divided into 14 groups. Each group represents a district. The dots represent the number of check-ins, and the colour indicates the different categories.

From a group's results, the author can then offer a hypothesis about the number of Twitter users and the type of land function. To verify the author's assumptions, evaluation results are compared to the official population data released by the department of statistics of the Makassar City local government (see Table 4.2). Then, to ascertain the land function, the author used the official land-use data published by the Department of Spatial Planning of the Makassar City local government (see Fig. 4.14). The author observed that 8% of Twitter users could be found in the Biringkanaya sub-district which has the largest population and the largest area. And when compared to the questionnaire data, 14% of respondents reside in this sub-district. In fact, most of the Twitter user activity took place around the airport, not in the residential area.

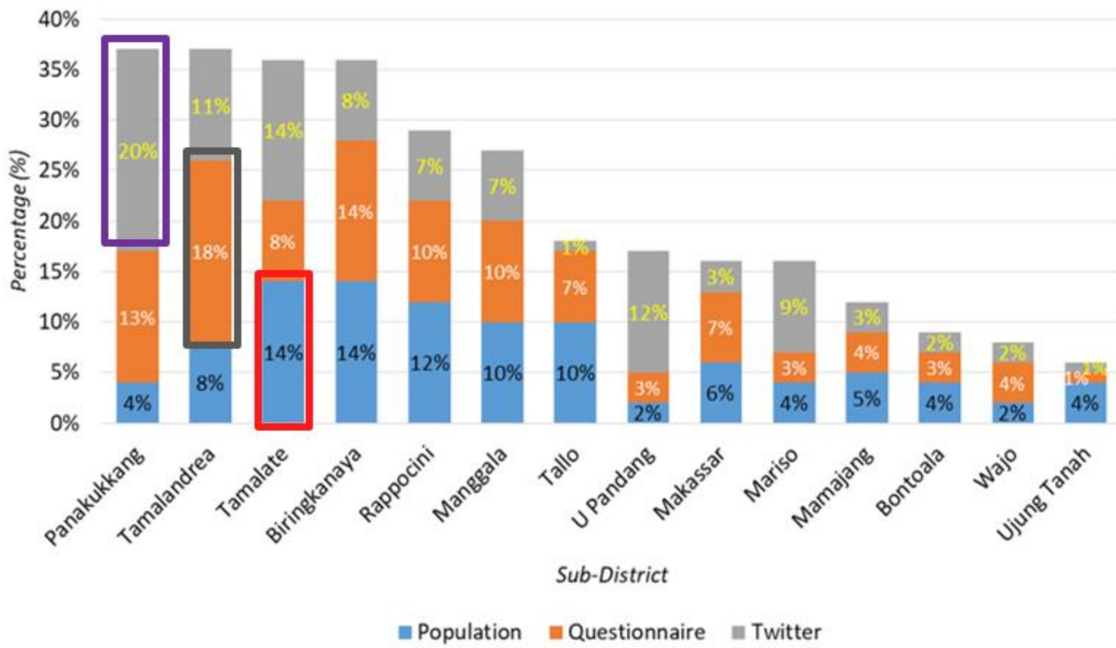


Figure 4.12 Graph of comparison data between the population, questionnaire, and Twitter

Percentages of Twitter users generated from the two sub-districts of Bontoala and Tallo, were 2% and 1%, respectively. According to the existing land function, the area is designated for warehousing. Our assumption is that this area is not attractive to visit due to the lack of facilities and infrastructure in the surrounding environment to attract people. Twitter activity is only done around the home, which also affected the number of respondents, which was just 7% and 3%.

Another finding is related to Tamalandrea sub-district which has the highest number of respondents (18%) and Twitter users (11%) of which the majority are students. Based on the mapping activity (see Figure 4.13), the brown color for people's activities is dominant in the university area. Then, our hypothesis is that the land function is for the education area. That is because the majority of students carry out their daily routines at home and in the university. This can be proved by comparing the occupation types of respondents on the questionnaire and the land-function map released by the local government.

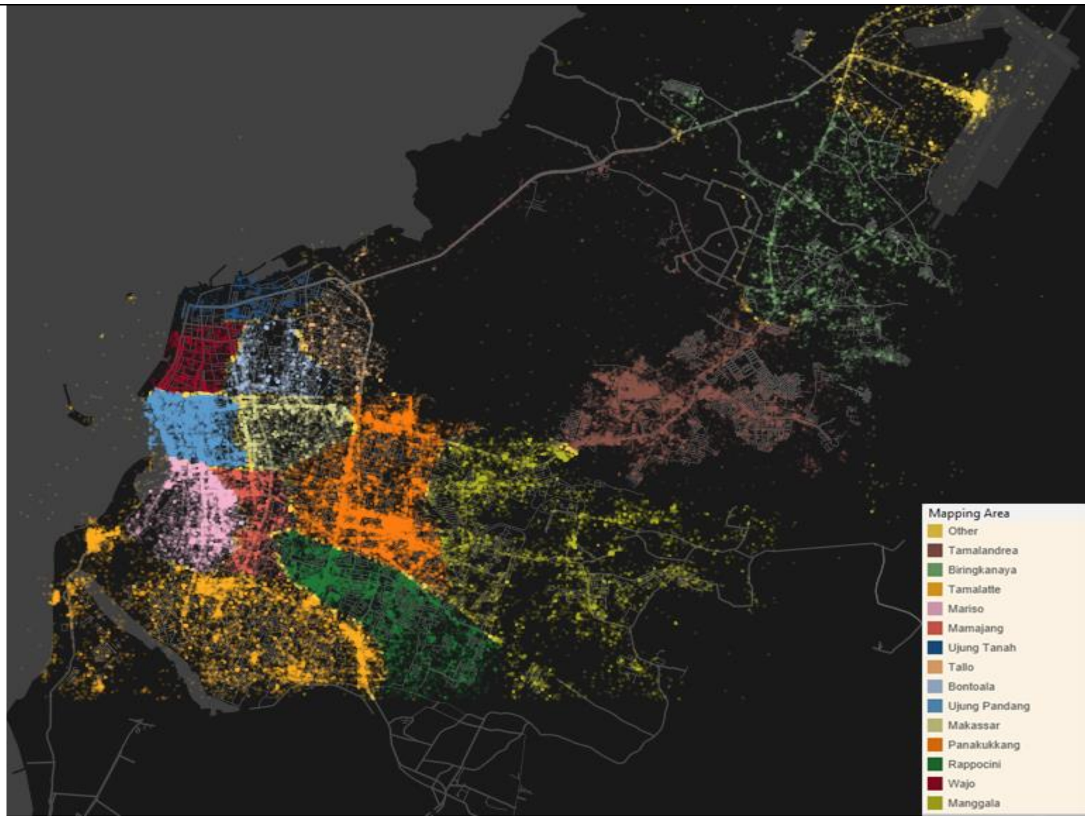


Figure 4.13 Comparison of the distribution of user check-ins for each district

Meanwhile, 14% of Twitter activity was sourced from the Tamalate sub-district. When compared to the other regions, it was the second largest for Twitter distribution. Based on the land-use map, this area is designated as a tourism and culture area. There are two sub-districts which represent the business area: Mariso 9% and Ujung Pandang 12%. Regarding the residential area, three districts have been identified; Panakukkang 20%, Rappocini 7%, and Manggala 7%.

4.4. Section Conclusion

Based on the analysis, it is shown that firstly, the three largest sub-districts for Twitter users in Makassar city are Panakukkang (20%), Tamalatte (14%) and the Ujungpandang district (12%). Secondly, based on our hypothesis, the largest distribution of individual mobility is contained in the residential area. Thirdly, concerning the movement of people, the writer has concluded that land function could also characterize the behavior of people. It can be proved that most of the respondents are students and connected to the Tamalandrea district which is an education region. It means that there is a correlation between land function and the people who inhabit a particular area as associated with their occupation. Fourthly, questionnaire data can be used as a data source to determine the user profile of the social media application Twitter. For urban planning, this data has the potential to allow estimates of the city's population in the daytime and at night.

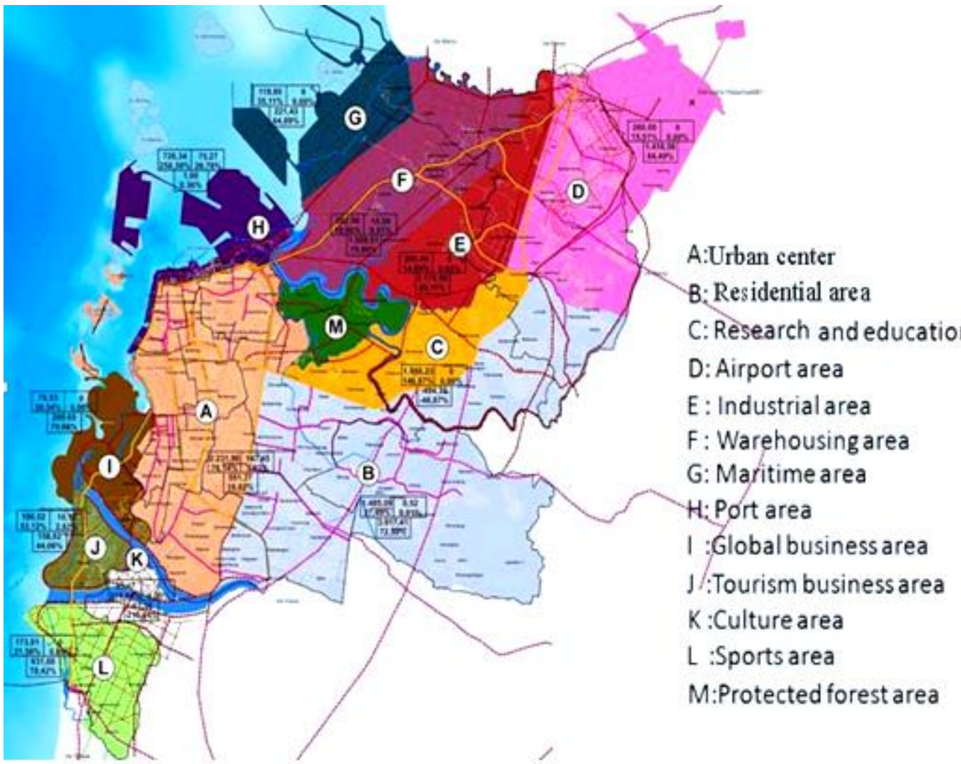


Figure 4.14 Official land-use map of the Makassar City government [2]
Source: *http://darimakassar.com/rtrw-kota-makassar-2010-2030-2*

References

1. Central Bureau Statistic. Makassar in Number (2014).
2. Spatial planning map of Makassar city 2010-2030. (8 August 2016), *http://darimakassar.com/rtrw-kota-makassar-2010-2030-2*

Chapter 5

Location Check-in as a Key to Identify Dynamic Land Use Map

5.1 Current Situation of land Use Data Collection

Urban planning is a technical process in the formation, arrangement, and development of a city. One kind of study on urban planning is land-use mapping, related to accurate land determination for urban zoning. The problem on urban land-use mapping is deciding upon the particular region for certain land use. Previous studies have been conducted to detect land use over time, such as the use of aerial photographs for mapping and quantifying the change in forest land-use patterns [1], remote sensing [2], geographic information systems techniques [3], and Landsat images via satellite, which provide an efficient means for land-use detection [4,5]. However, these approaches have some weaknesses, such as the inability of numerous sensors to obtain data and information in cloudy areas. Clouds make the resolution of the satellite imagery too coarse for detailed mapping and for distinguishing small contrasting areas, yet high-resolution satellite imagery is very costly and time-consuming [6].

With the development of an embedded system planted on the smartphone, a user's movement could be tracked [7]. Researchers use the mobile phone's footprint to predict the user's behavior [8], Bluetooth traces [9], Global Position System (GPS) hint [10], and smart card data [11]. In the literature, author find that some researchers use these devices for land-use identification—for instance, the demonstration of GPS data for discovering a region and sensing human activity [12], urban Wi-Fi characterization [13], land-use and landscape identification using cell-phone data [14–16]. However, these models concentrate on a particular region in a specific area, the lack of information from this data [17] and difficult to identify the user's footprint.

To overcome these research challenges, some scientists use location-based-on-social-network (LBSN) data to capture people's travel behavior as an alternative approach. These data contain information on their interests, hobbies, and place activities. Recently, the data source of social media's geolocation has provided new information in terms of understanding an individual's activity pattern. In the literature, author found that some researchers discuss social media—that is, foursquare check-in data—to catch people's social events distribution, such as by investigating human travel activity patterns [18], inferring individual lifestyle patterns [19], and predicting the next venue [20]. Additionally, many researchers have used Twitter's check-in data to capture the individual's activity in the urban area, such as in home-location identification [21], and to estimate the user's location [22, 23].

On the basis of the above description, the information on the people visiting a particular place will be pertinent to form a new area. In the perspective of urban planning, geolocation becomes an indicator to identify a specific urban area. In this paper, author analyze social media data from Twitter for detecting the dynamics of urban land use. The data includes the period (time-stamped), the user's status text or post information (tweet), and the geolocation or specific location that is the point of interest where and when people undergo check-in activity. To analyze the data, author propose a grid-based aggregation method and text mining to split the Twitter land map. The proposed method uses a grid to divide the urban area and text-mining activity to count popular keywords among different categories. Author compare two distinct methods: a rank method and *k*-means clustering to classify different areas. To validate the analysis, author combine the individual's travel time spread on weekdays and weekends as the parameters to define the land-use. In this part, our focus is not only the check-in data but also involving the Twitter text record, where author uses a specific filter on the location name search

5.2 Land Use Detection

Makassar is a city with the largest population in eastern Indonesia. The 2010 census of population registered 1.34 million residents in an area of 175.7 km2 [24]. From the data collection, author identified that Twitter users have an average age of 15–40 years, where 34% are males and 66% are females [25]. This research is essential, as the land-use map of Makassar City is not up-to-date, while the current design for projection is 20 years ahead [26]

5.2.1 Text Mining for Place-Name Identification

The main purpose of text mining is to support the process of knowledge discovery on large document collection. In principle, text mining is a science field that involves information retrieval, text analysis, natural language processing, and a logic-based learning machine [27]. In this regard, text mining specifies the places at which the individuals make the tweets. Through this service, the check-in locations are grouped using the clustering method, and the place-names are individually identified from the user's status post on Twitter marked with the symbols # and @ to define the place-name (e.g., "eating at #thexxxrestaurant" and "playing soccer at @theyyystadium"). Because of this, the Twitter application does not insert the location name on the APIs' search engines but includes the geographic location in the form of latitude and longitude coordinates. Author use a Voyan tool, an open-source web-based application used to discover most frequently used words, to analyse and count the documented texts and to ease text separation.

Figure 5.1 shows the data flow and methods proposed for urban land-use identification, where two data-grouping methods are compared. To conclude the land hypothesis, we used the daily time distribution on weekdays and weekends activities.

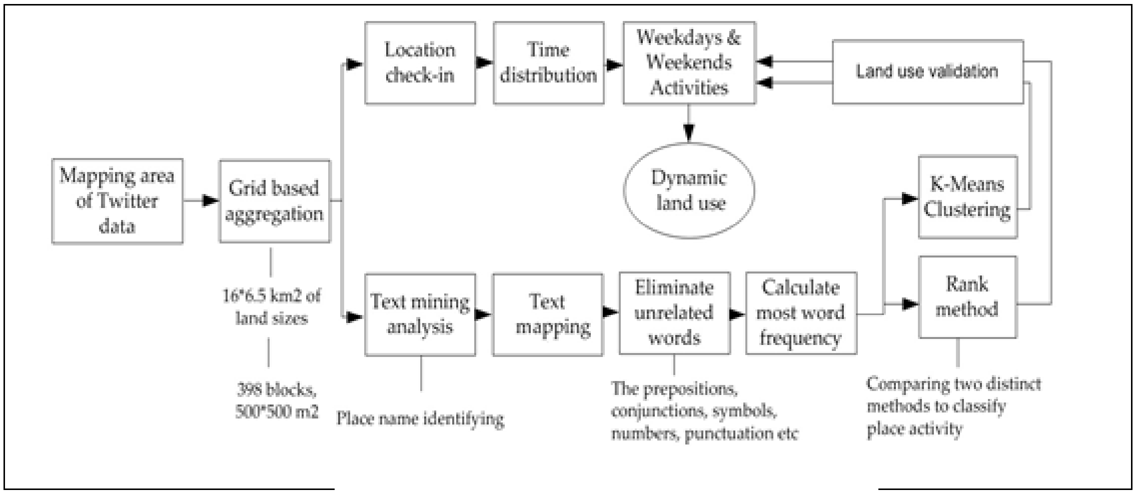


Figure 5. 1 Data flow diagram of method used

5.2.2 Aggregation Grid for Dividing Land Area

In a Twitter dataset, check-ins are separate; thus the issue arises of how to unite their spreads in one or several information units. Author propose a grid-based aggregation method to identify each area for detecting urban land use, a technique to combine distinct objects into different groups. Figure 5.2 shows the 16 × 6.5 km^2 tweet distribution map of Makassar City.

To facilitate the analysis, we divided the grid into 500 × 500 m^2 areas and produced 558 blocks. Author then removed the blocks without check-in activity. A total of 160 blocks were removed, and 398 blocks with tweet activity were used. The figure below illustrates the spread of twitter check-ins. The dots represent the locations, and the block gradations indicate the frequency of each block.

To recognize the place type on each block, author used the user's text-posting activity on Twitter. A total of 85 venues were found from the whole blocks. Author then divided the area into six categories (Table 4.1). From this result, author could see the description about the information of the land.

Table 5. 1 Location categories visited by user.

Category	Place
Art and Entertainment	Cinema, street park, bar, karaoke, hall, meeting building, monument, wedding hall, fort, photography studio, radio station, television station.
Business and Service	Hotel, guesthouse, housing, bank, diagnostic center, pharmacy, skincare clinic, hospital, telecommunication service headquarter

Community and Office	Electricity company, church, office, mosque, university, school, library
Food and Drink	Café, coffee shop, restaurant (culinary, meatball, seafood, noodle, chicken porridge, ice cream, fried chicken, pizza, donut, steak, snack, lunch, dinner, fried rice, sushi, and udon), tea house, canteen, cake shop, kiosk/corner stand (roasted corn, and fried banana)
Shopping	Mall, shop, store, fresh market, bike shop, bookstore
Sports and Recreation	Indoor soccer field, basketball court, beach, stadium, sports area, jogging track, garden, swimming pool, field, gym

To calculate the number of check-ins on each block, author grouped each block into 32 classes with an interval of 100 check-ins. The grouping provided a description of the frequency of data diversity. Figure 5.3 shows the graph of block allocation based on each class (e.g., the class C100 contains 112 blocks).

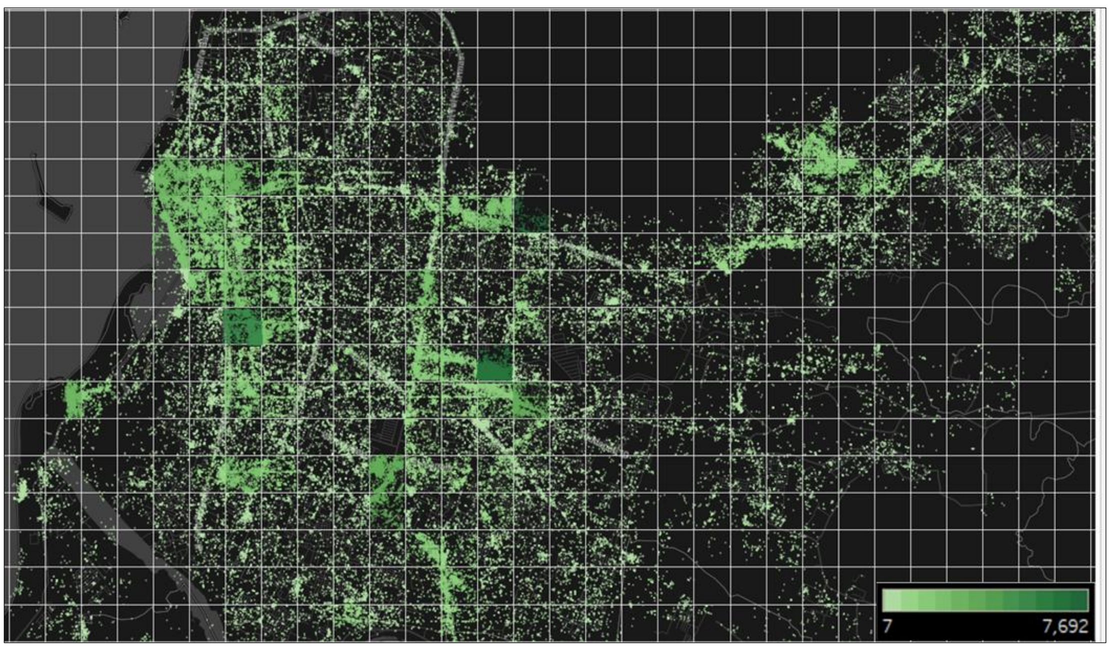


Figure 5. 2 Grid distribution of check-ins with 500 × 500 m2 blocks. The dots represent the user location tags, and the color describes the Twitter activity frequency

5.2.3 K-means Clustering for Land Use Characterizing

Clustering is a method to group objects into classes with identical characteristics [28]. The *k*-means clustering is one algorithm of unsupervised learning that uses a nearest mean approach. This reliable algorithm can quickly process huge amounts of data [29]. The *k*-means clustering attempts to group objects into two or more clusters so that the objects within one cluster share similarities. To measure the similarity among objects, *k*-means clustering utilizes the distance function as the parameter to determine the group members. The *k*-means algorithm uses the following steps:

- Decide the number of clusters (in this research, five clusters are specified).
- Determine the centroid value (center of measurement) randomly.
- Calculate the distance between the centroid points and the point of each object. To measure, author use the Euclidean distance:

$$D_e = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2} \quad (1)$$

  Where D_e is the Euclidean distance, i is the number of the object, (x, y) are the object coordinates, and (s, t) are the centroid coordinates.
- Assign object to closest cluster.
- Go back to step 2 and recalculate the centroid value until the cluster members do not move to other clusters.

From the place activity (see Appendix A), author then grouped the data and produced five clusters. Table 4.2 shows the different places visited by people.

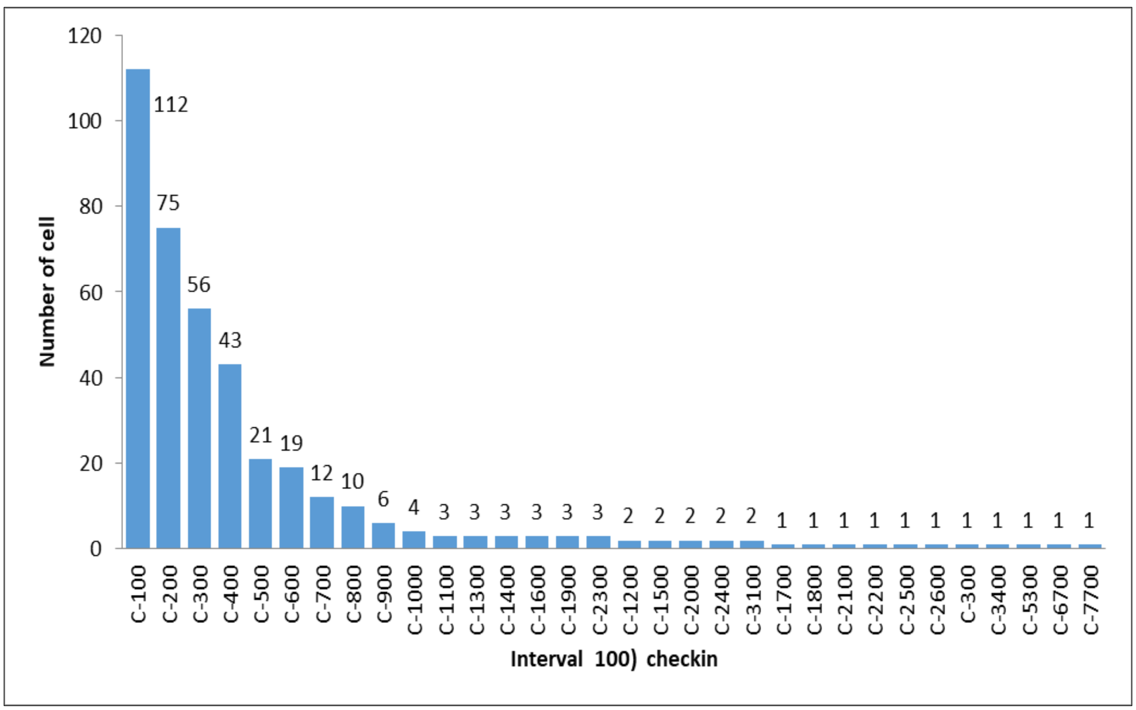


Figure 5. 3 Frequency distribution classes with each group of 100 check-ins

Table 5. 2 K-means clustering result for land use type.

Place	Check-In	%	Place	Check-In	%	Place	Check-In	%
Cluster 1			Soccer	597	0.032	University	11187	0.805
Housing	2974	19.101	KFC	450	0.024	**Cluster 4**		
Mall	1691	10.861	Worship	620	0.033	Office	2330	0.150
Cinema	1841	11.824	Shop	278	0.015	School	1742	0.112
School	1869	12.004	Park	247	0.013	University	1969	0.127
University	1780	11.432	Seafood	278	0.015	Beach	427	0.027
Coffee	1777	11.413	Karaoke	421	0.023	Coffee	3692	0.238
Hotel	958	6.153	University	436	0.023	Hotel	568	0.037
McDonald's	958	6.153	Cinema21	269	0.014	Housing	438	0.028
Street	877	5.633	Hall	658	0.035	KFC	1082	0.070
Stadium	845	5.427	Fitness	182	0.010	Mall	544	0.035
Cluster 2			Housing	186	0.010	McDonald's	444	0.029
School	1017	0.054	Bookstore	133	0.007	Restaurant	2309	0.149
Hotel	1003	0.054	Hall	680	0.031	**Cluster 5**		
Culinary	1372	0.073	Tea	590	0.027	University	3517	0.160
Bank	616	0.033	Hotel	438	0.020	Restaurant	2938	0.133
Restaurant	2811	0.151	Meatball	430	0.020	Coffee	2500	0.113
Office	1164	0.062	Mall	429	0.019	Hospital	1889	0.086
Coffee	963	0.052	Beach	375	0.017	Culinary	1570	0.071
Street	889	0.048	Fort	296	0.013	Cinema21	1329	0.060
Mall	676	0.036	Supermarket	250	0.011	Office	1235	0.056
Noodles	1322	0.071	Stadium	191	0.009	KFC	1148	0.052
Café	376	0.020	**Cluster 3**			Street	820	0.037
Hospital	506	0.027	Housing	1350	0.097	McDonald's	704	0.032
Pizza	534	0.029	Hotel	1354	0.097	Pizza	698	0.032

Figure 5.4 shows the time distribution pattern on weekdays and weekends from *k*-means clustering. To analyze the land use type, the method will be compared with the group result from the ranking method to determine the potential land use.

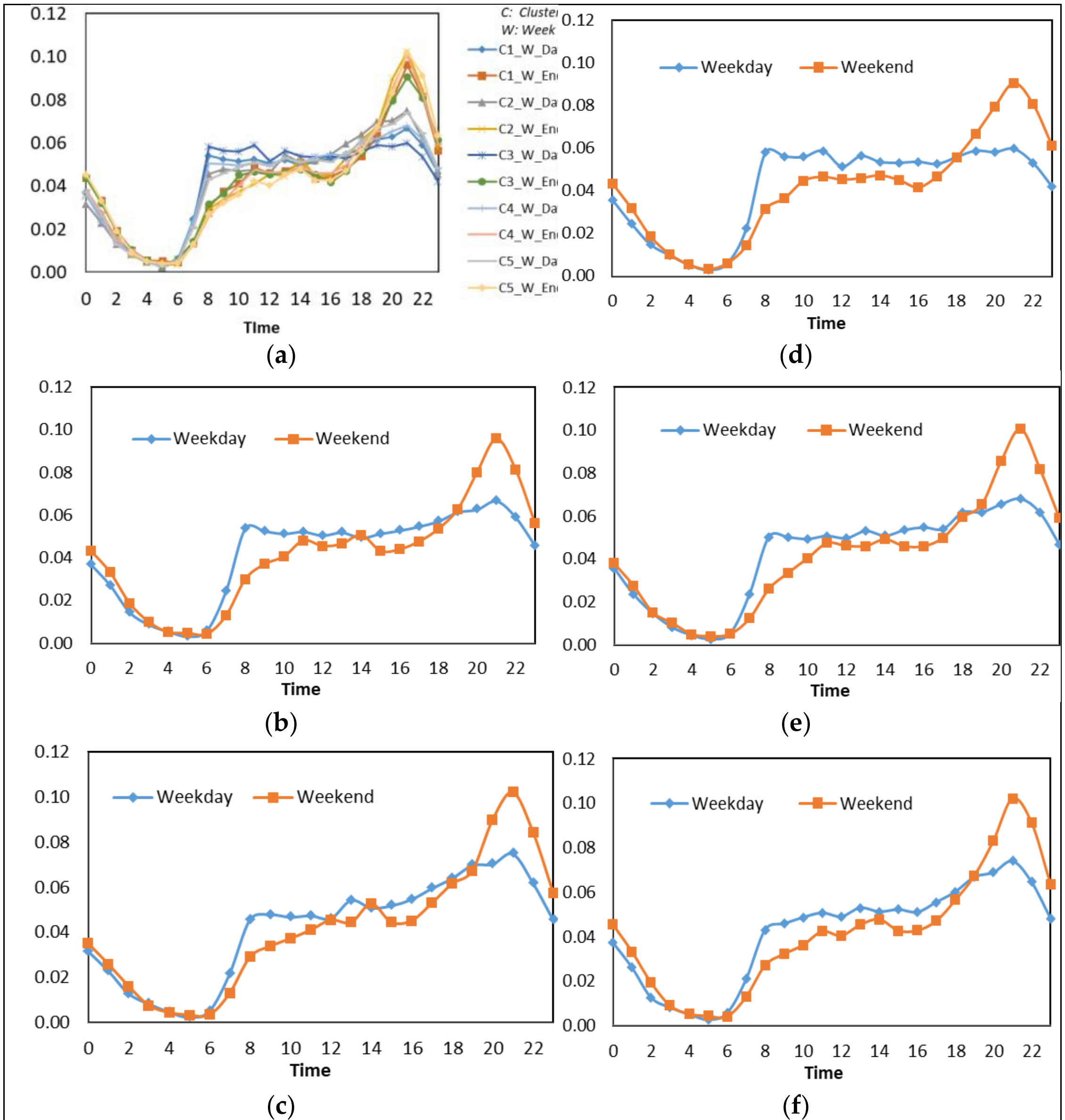


Figure 5. 4 the graph results of k-means clustering in different time frequencies on weekdays and weekends. (a) Clusters' comparison on weekdays and weekends; (b) cluster 1; (c) cluster 2; (d) cluster 3; (e) cluster 4; (f) cluster 5.

5.3 Land-Use Segmentation

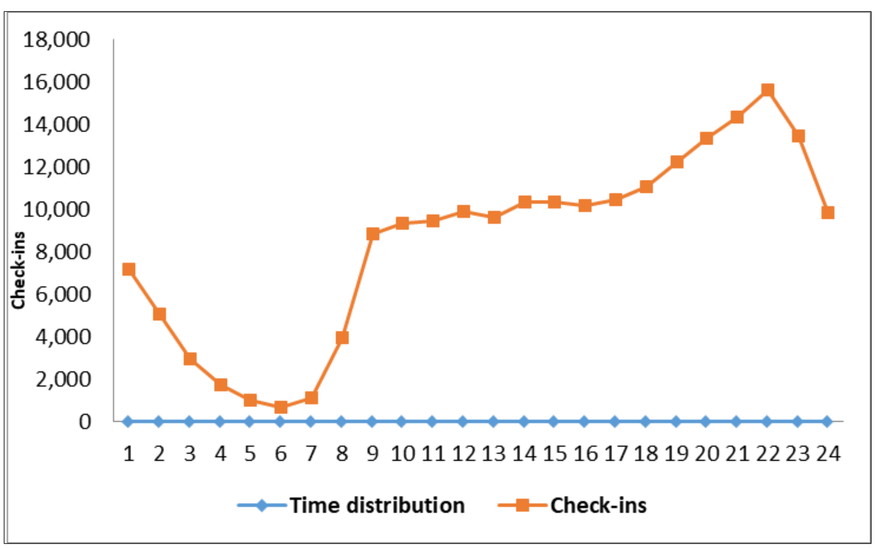
Author tried a method of grid-based aggregation to divide the urban area. After class grouping (see table 5.3), author then characterized each region to understand the type of land use. To identify the land area, author grouped the check-in activity blocks on the basis of the following:

- Author determined the frequency of places visited by comparing the percentage data of each block. Author then combined the blocks into several class and grouped the class into several clusters. In this case, each cluster was decided by the place with the highest frequency as a decision-making indicator. For example, on the basis of tweets, author found that class C100, C200, C300, and C400 were dominated by the individual's activities in residential areas (see Table A1). Thus, the combination of these class was called cluster 1.
- To identify the land-use type, author ranked every place on each cluster to determine the most visited venue (see Table 5.4).
- Author then analyzed the time distribution frequency on each class to determine the trends of each region by comparing weekday and weekend check-in patterns. In doing so, the identification of land use could be detected.

On the basis of the above criteria, author classified the check-ins (see Table 5.3) into four clusters. The clusters illustration can be seen in the following table:

Table 5. 3 Grouping of check-in class.

Cluster	Classes (C)
Cluster 1	C100, C200, C300, C400
Cluster 2	C500, C1200, C1300, C1400, C1500, C2200, C2300
Cluster 3	C1000, C1900, C2000, C2100, C2200, C2400, C2500, C2900, C5200
Cluster 4	C600, C700, C800, C900, C1100, C1600, C1700, C3300, C3600

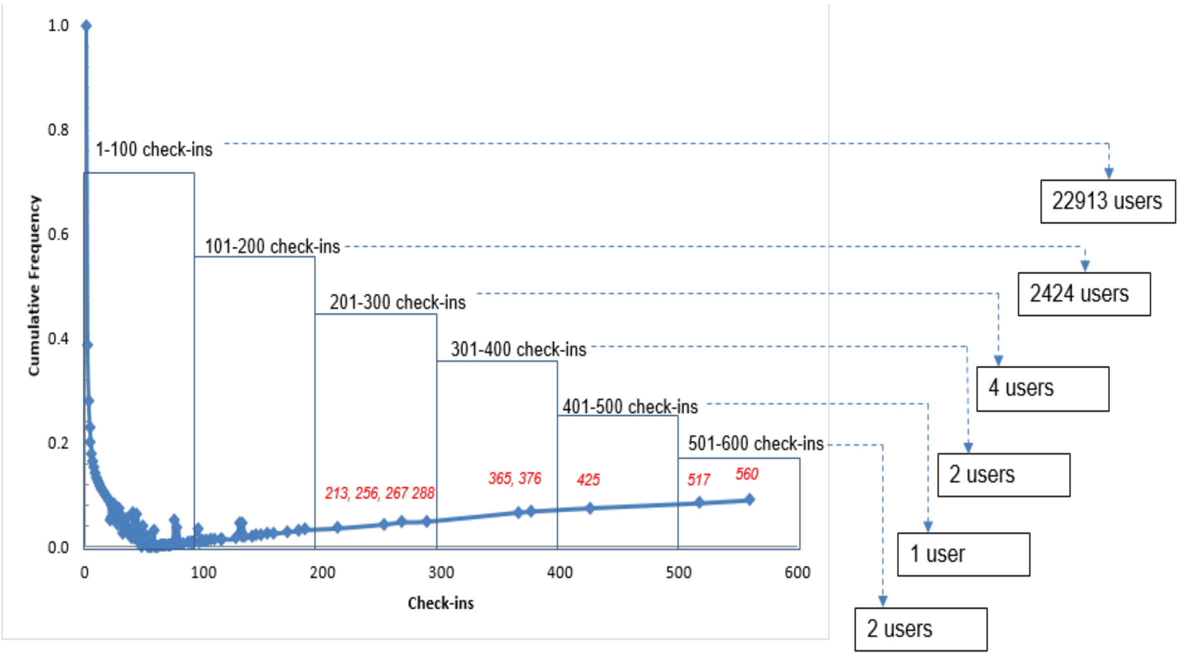


(a)

Table 5. 4 Place ranking for land-use-type clustering.

Cluster 1		Cluster 2		Cluster 3		Cluster 4	
Place	**Check-in (%)**	**Place**	**Check-In (%)**	**Place**	**Check-In (%)**	**Place**	**Check-In (%)**
Housing	4324 26.149	University	8033 46.019	Culinary	2787 13.654	Coffee	2943 16.121
University	2089 12.633	Street	2206 12.637	Coffee	2630 12.885	Hotel	2640 14.461
Coffee	1634 9.881	School	996 5.706	Restaurant	2170 10.632	Office	1820 9.969
Street	1596 9.652	Restaurant	814 4.663	McDonald's	2008 9.838	Culinary	1767 9.679
Culinary	1291 7.807	Pizza	728 4.170	Hotel	1440 7.055	Stadium	1205 6.601
School	1156 6.991	Office	684 3.918	KFC	1333 6.531	Mall	1073 5.878
Office	1112 6.725	Monument	606 3.472	Office	1269 6.217	Restaurant	1003 5.494
Restaurant	608 3.677	McDonald's	457 2.618	Mall	1184 5.801	School	873 4.782
Hospital	341 2.062	Mall	440 2.521	School	1145 5.610	KFC	867 4.749
Beach	304 1.838	Library	438 2.509	University	928 4.547	Hall	780 4.273
Park	286 1.730	KFC	371 2.125	Stadium	523 2.562	Cinema XXI	672 3.681
Office	277 1.675	Housing	307 1.759	Hospital	440 2.156	Street	555 3.040
KFC	268 1.621	Hotel	233 1.335	Beach	438 2.146	Hospital	510 2.794
Hotel	259 1.566	Hospital	228 1.306	Street	370 1.813	McDonald's	390 2.136
Soccer	244 1.476	Hall	222 1.272	Cinema	364 1.783	University	314 1.720
Seafood	225 1.361	Culinary	164 0.940	Pizza	329 1.612	Pizza	283 1.550
Mosque	180 1.089	Coffee	161 0.922	Bar	318 1.558	Corner	174 0.953
Building	154 0.931	Cinema21	152 0.871	Karaoke	256 1.254	Fresh market	142 0.778
Shop	96 0.581	Building	109 0.624	Monument	243 1.191	Bank	123 0.674
Hall	92 0.556	Beach	107 0.613	Supermarket	236 1.156	Building	122 0.668

Figure 5.5a illustrates the user's daily frequency times. Author observe that the peak of individual activity occurs at 10 p.m. and the lowest check-in activity at 6 a.m. On figure 5b, author see that majority of user frequency is between 20 up to 100 check-ins. From the 85 places (see Table 5.1), author then identified the venue type and found 31 places with significant check-ins. Table 5.4 depicts the spatial distribution cluster showing the check-in numbers and percentages in each place. This cluster would provide an overview of potential land use.



(**b**)

Figure 5. 5 Daily time distribution activity (a) and trip flow distribution for each user (b)

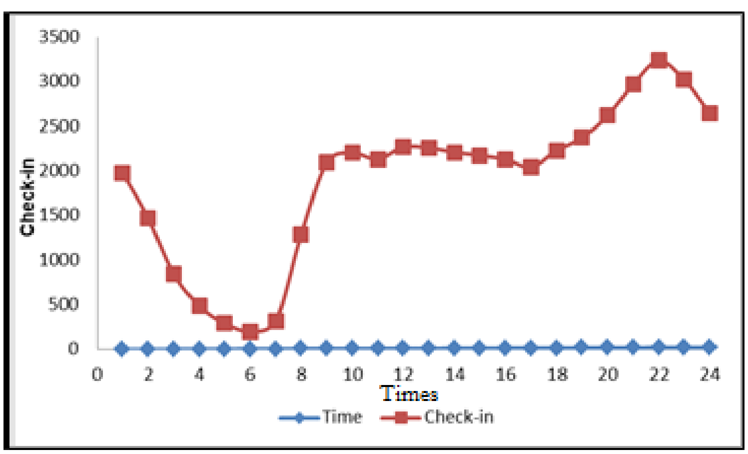
5.3.1 Housing Area (Cluster 1)

To understand the land use of this region, author compared the classes by considering the most frequently visited places. Author observed that in general, the tweet activity in cluster 1 was closely related to the activities of people who were around the residential area (see Figure 5a). We found about 26% of the tweet activity covered by this group (see Figure 5.6b). Author then analyzed the daily tweet pattern and found that the peak of tweet activity occurs at 10:00 p.m. (Figure 5.6a), related to the individual's activity before bed. Meanwhile, other activities, such as being in or going to a university, a café, and others, were done during the day and peaked from 11 a.m. to noon. Author observed about 70% of this area was covered by this cluster. Thus, this cluster can be associated with the housing area. If author compare this to the *k*-means clustering, then this group is identical with clusters 1 and 3 (see Table 5.2 and Figure 5.4). Thus we associate this area to housing.

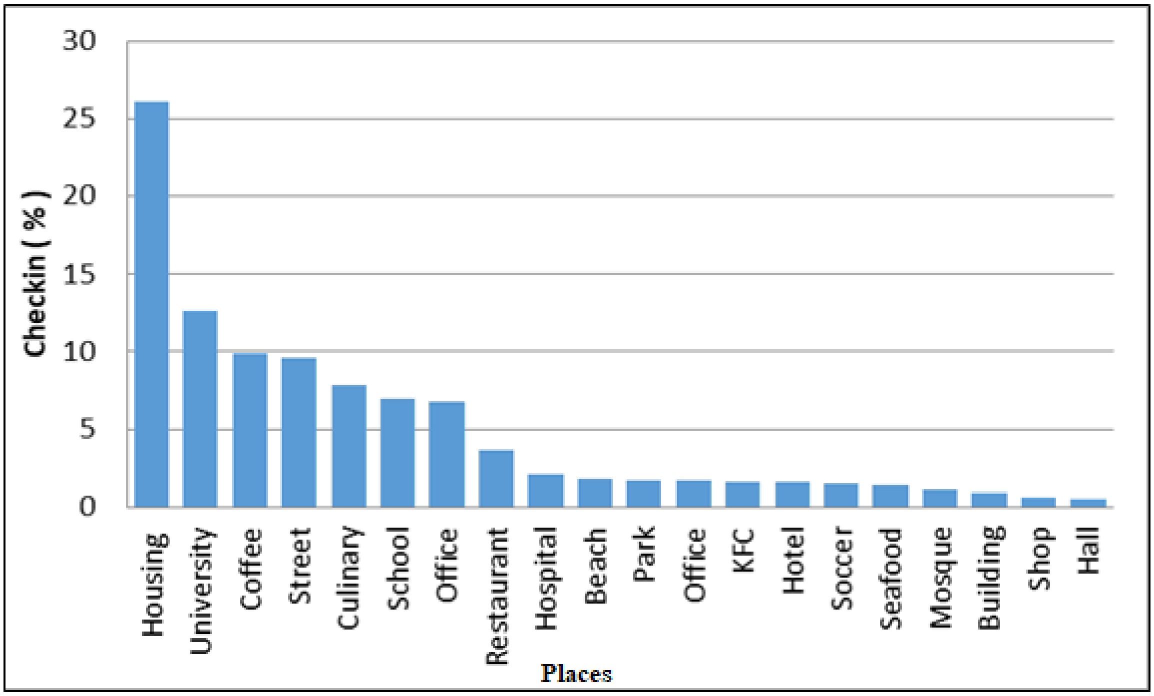
(a)

(b)

Figure 5. 6 The words frequency (a) and housing distribution map (b)



(**a**)



(b)

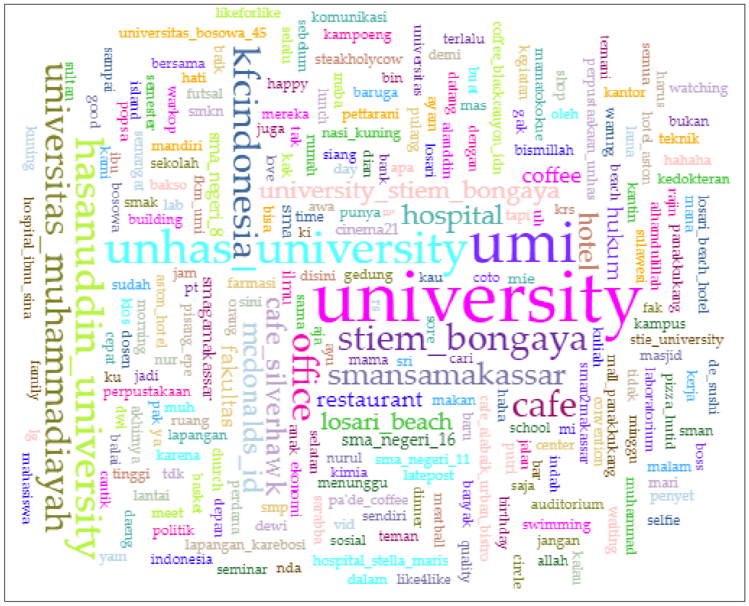
Figure 5. 7 The daily time spread (a) and percentage of check-ins in different places in cluster 1 (b)

.

5.3.2 Education Area (Cluster 2)

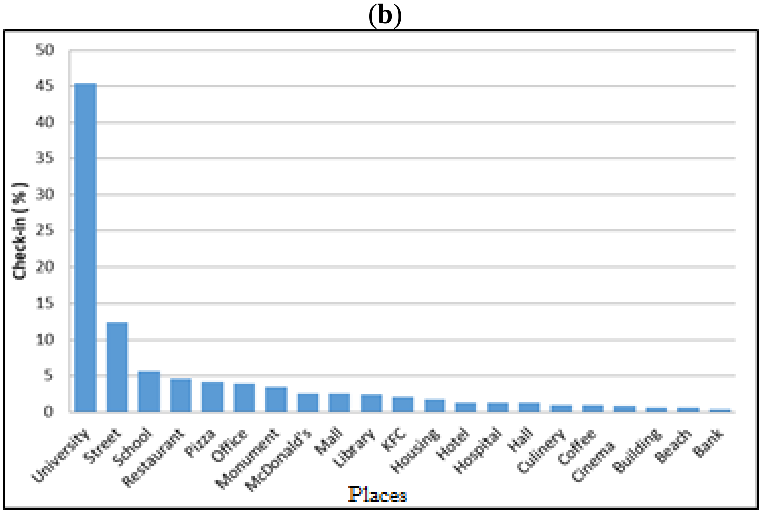
As shown in Figure 5.8d, author compared the pattern of weekday and weekend activities. During weekday, tweet activity increased at 8 a.m. Author observed a changing trend between 10:00 a.m. and 2:00 p.m. Then on the weekends, the peak activity was at 8:00 a.m. and 11:00 p.m. Author compared the pattern of weekdays and weekends and found a very significant difference in that, on the weekends, the tweet activities decreased. This was because on weekdays, the frequency of university visits increases, while on weekends, only a handful of individuals come to the university.

In general, this cluster was more populated in places such as universities and schools. The existence of other venues such as restaurants—Pizza Hut and McDonald's—malls, and others was because of the university and was not influenced by other regions. On the basis of this analysis, author then concluded that this cluster is related to education. This can be seen in the word frequency and graph percentage of each place (Figure 5.8a, c). This group is similar to clusters 2 and 5 (see Table 5.2 and Figure 5.4) from the *k*-means result. If author observe the difference between Figures 5.8c, f, author find that there are contrasting activities during weekdays and weekends, except for during night.

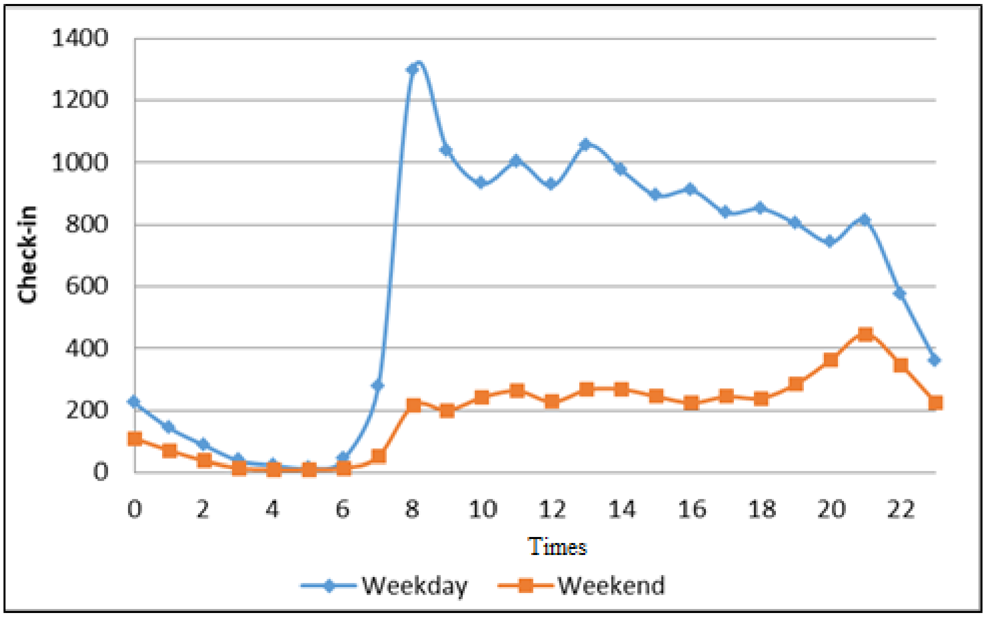

university
umi
unhas
university
hasanuddin_university
universitas_muhammadiyah
kfcindonesia
stiem_bongaya
smansamakassar
office
cafe
hospital
university_stiem_bongaya
restaurant
losari_beach
mcdonalds_id
cafe_silverhawk
hotel
hukum
coffee
fakultas
sma_negeri_16

(a)

(b)


Check-in (%)
50
45
40
35
30
25
20
15
10
5
0
University
Street
School
Restaurant
Pizza
Office
Monument
McDonald's
Mall
Library
KFC
Housing
Hotel
Hospital
Hall
Culinary
Coffee
Cinema
Building
Beach
Bank
Places

(c)



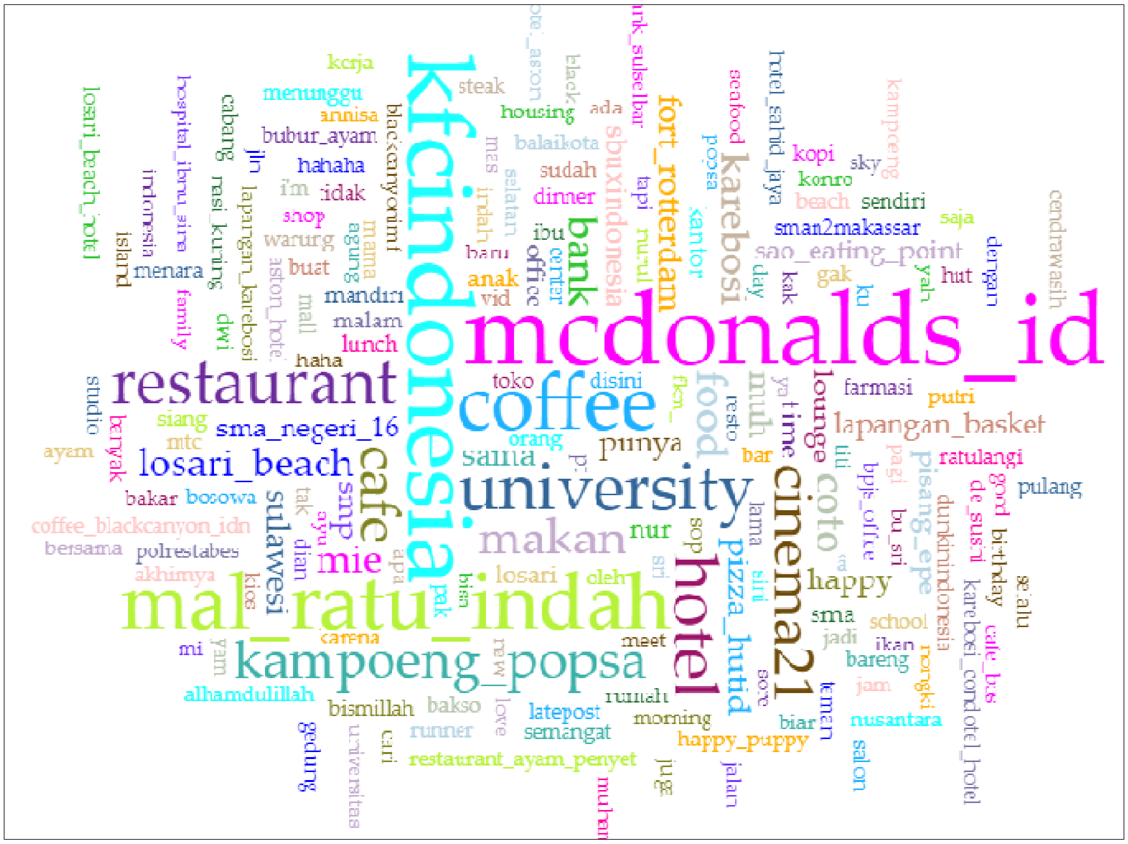
(d)

Figure 5. 8 The analysis of user text posted (a), the map of the education area (b), a graph of different visited places in the education cluster (c), and difference activity on weekdays and weekend (d).

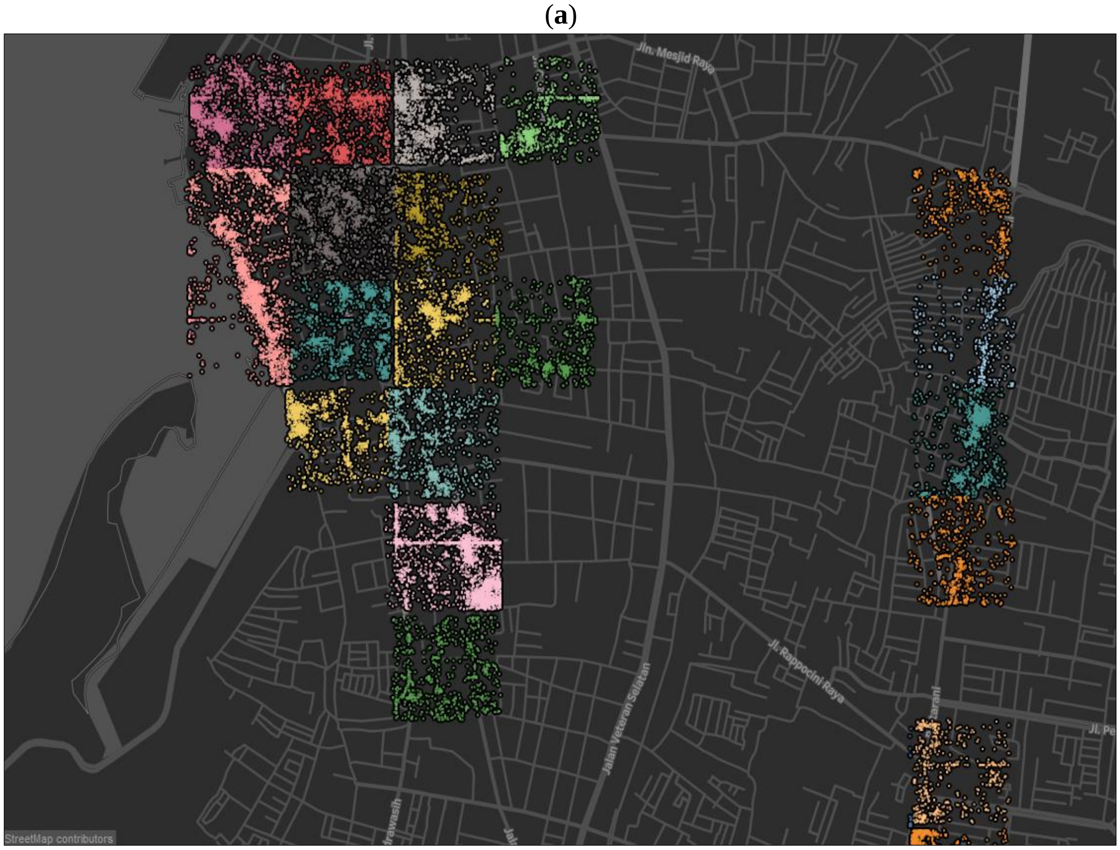
5.3.3 Commercial, Business, and Work Area (Cluster 3)

In cluster 3, author divided the time spread into two parts (evening and morning). In the evening, the peak of tweet activity occurred at 9 p.m. Author observed that this cluster was dominated by individual activity at places such as culinary venues, coffee, and restaurants (see Figure 5.9c). It is therefore most likely that people go out for dinner. Author would argue that this cluster represents the commercial area for eating or other culinary activities, which can be proven by the decrease of check-in activity one hour later (see Figure 5.9d).

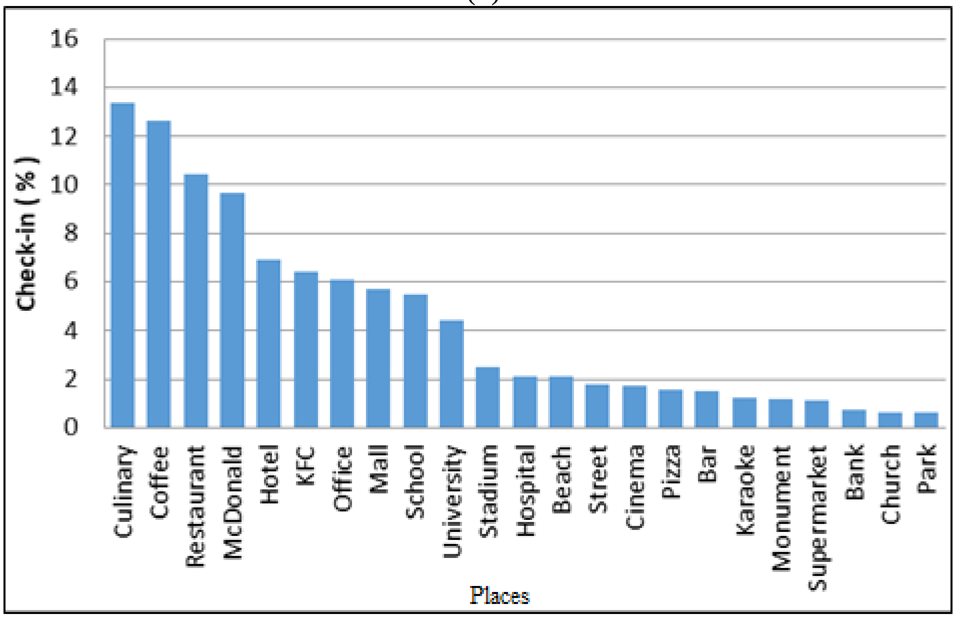
Then in the morning, the peak occurred at around 8–9 a.m., and then the trend fluctuated until noon or 2 p.m. (see Figure 5.9d). Author observed that this cluster was populated in places such as hotels, offices, and malls. Author argue that in addition to visitors, this check-in was also made by employees and office staff. Author therefore concluded that this was a working or business area. There was a large difference when we compared the tweet pattern on weekdays and weekends; weekends showed a decrease in tweet activity when compared to weekdays. Thus, we concluded that check-in at work places started from the morning and continued until noon. Then in the afternoon (returning home from work), people would look for other activities, such as shopping or going to dinner. Comparing this with the *k*-means result, author find that cluster 5.4 (see Table 5.2) has a similarity with the group pattern of the rank method. Author concluded that this is a work area.


mcdonalds_id
kfcindonesia
mal_ratu_indah
university
restaurant
coffee
kampoeng_popsa
hotel
cinema21
cafe
makan
losari_beach
sma_negeri_16
karebosi
fort_rotterdam
bank
food
pizza_hut
coto
lapangan_basket
sao_eating_point
sulawesi
mie
happy
lounge
time
sbuxindonesia
menunggu
bubur_ayam
hahaha
kerja
steak
housing
balaikota
sudah
dinner
kopi
sky
beach
sendiri
sman2makassar
saja
farmasi
putri
ratulangi
pulang
toko
disini
orang
sama
punya
bar
mandiri
malam
lunch
haha
siang
mtc
ayam
bakar
bosowa
coffee_blackcanyon_idn
bersama
polrestabes
akhirnya
losari
nur
meet
sma
school
jadi
ikan
bareng
jam
nusantara
biar
salon
alhamdulillah
bismillah
bakso
runner
latepost
semangat
rumah
morning
happy_puppy
restaurant_ayam_penyet
menara
warung
shop
anak
hut
dengan
cendrawasih
losari_beach_hotel
hotel_sahid_jaya
karebosi_condotel_hotel
dunkinindonesia
pisang_epe
universitas
gedung

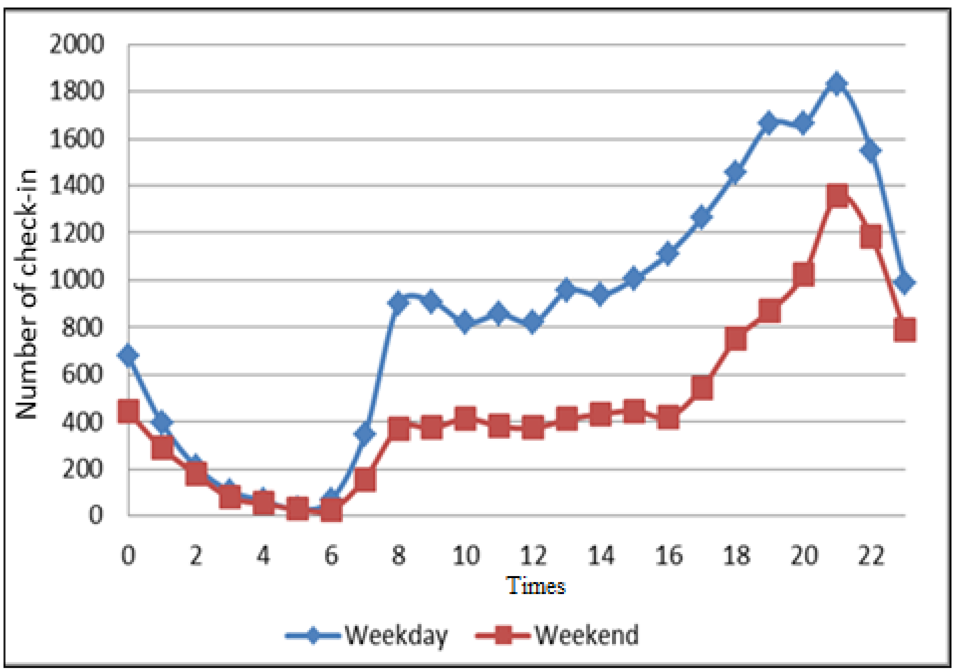
(**a**)


Jln. Mesjid Raya
Jl. Rappocini Raya
Jalan Veteran Selatan
StreetMap contributors

(b)



(c)

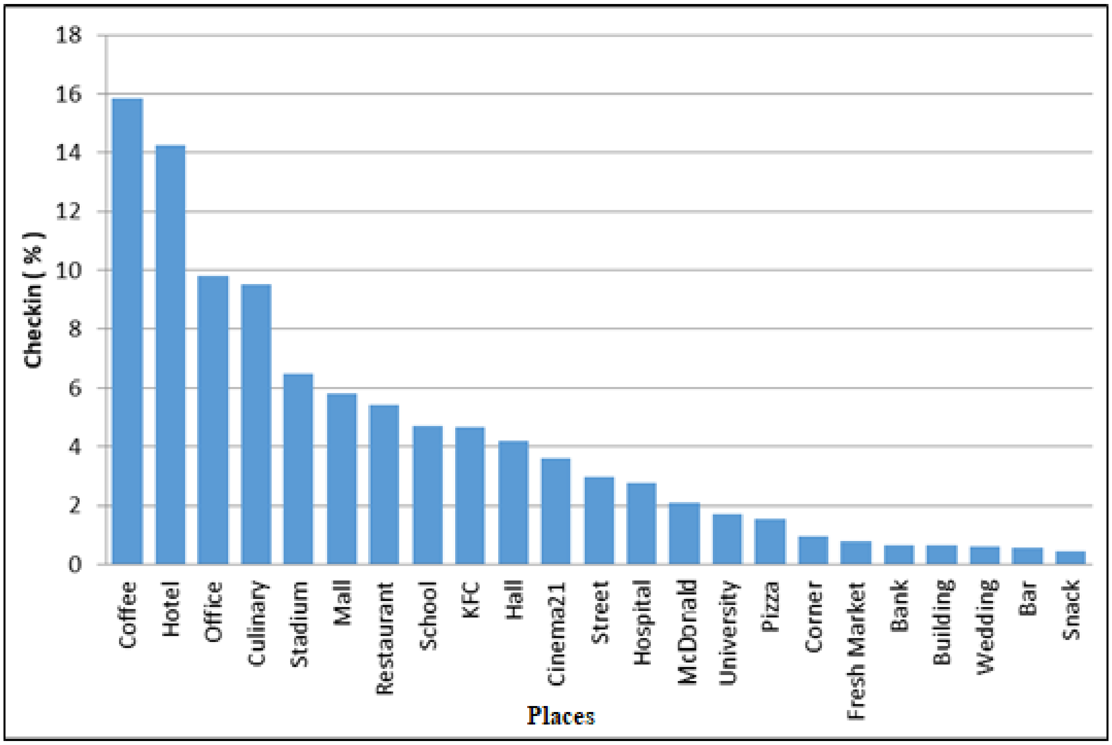


(d)

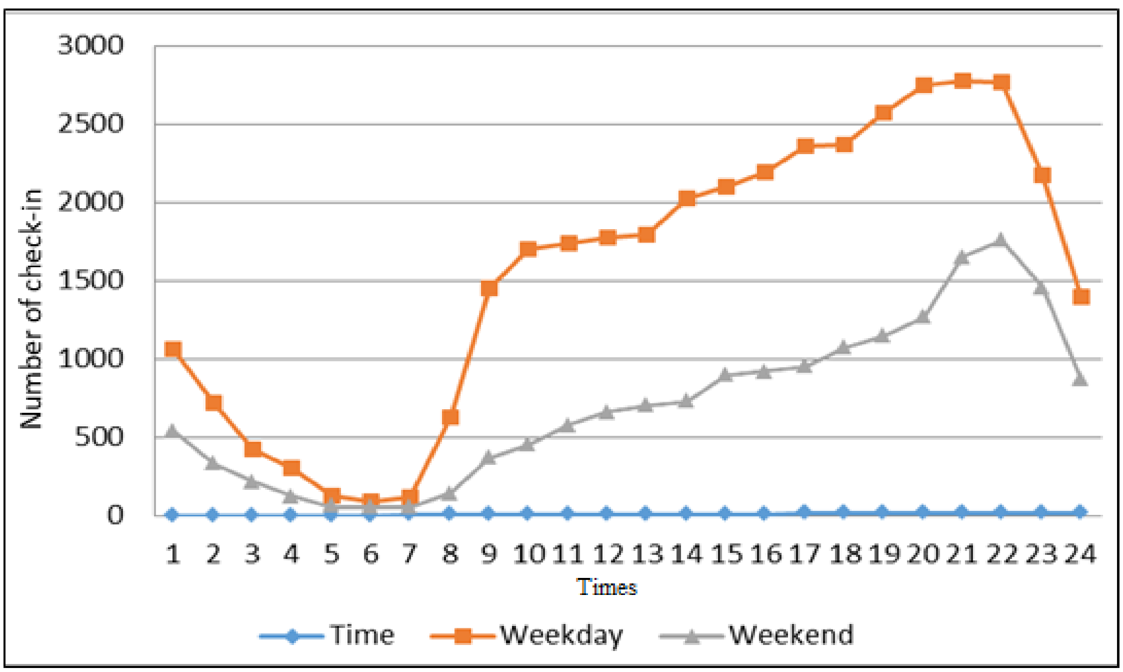
Figure 5. 9 The word frequency analysis (a), the user distribution map in cluster 3 (b), check-in activity in different places (c), and the time difference of user distribution on weekdays and weekends (d).

5.3.4 Mixed Area (Cluster 4)

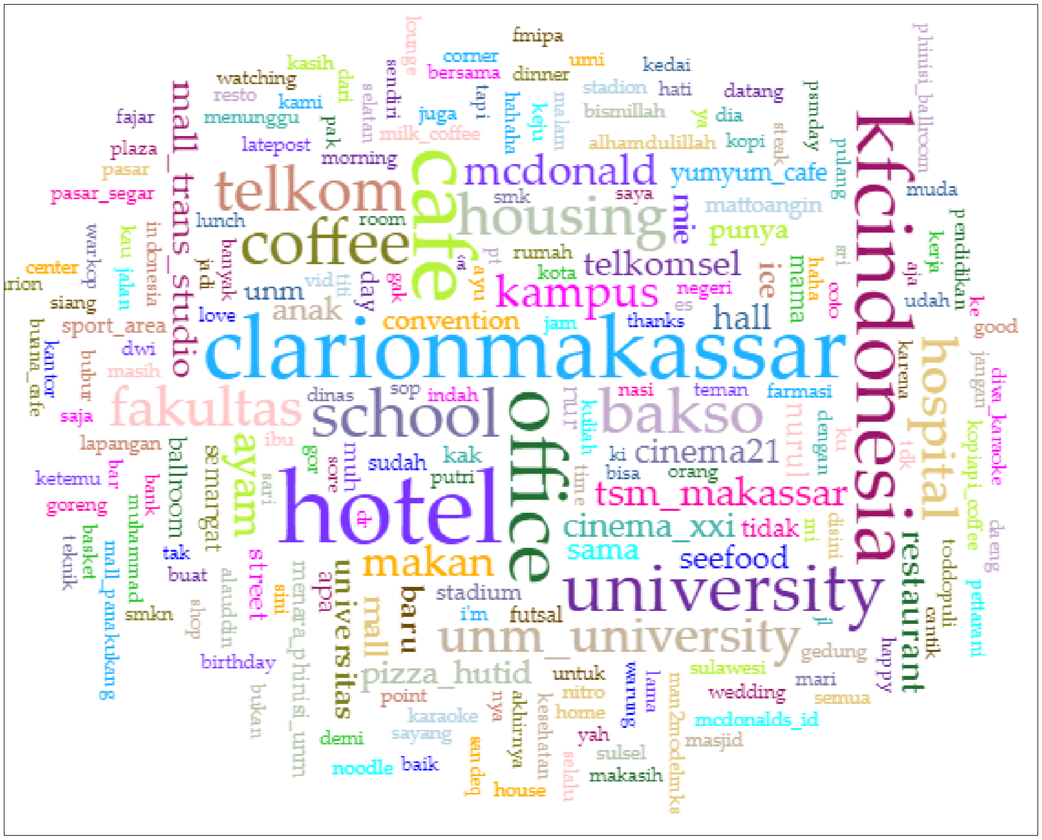
Author could not explain specifically the land use of this region. Author called this the mixed cluster, because in this region, there were various activities in venues such as hotels, shopping centers, office centers, and sports centers (see Figure 5.10a). In the morning, check-in activity for this cluster began at 7 a.m. and increased until the afternoon. The spread of time on weekdays and weekends had similar patterns. Author concluded that this area was the most active area as the tendency of check-in activity did not decrease until 10:00 p.m. (see Figure 5.10b).



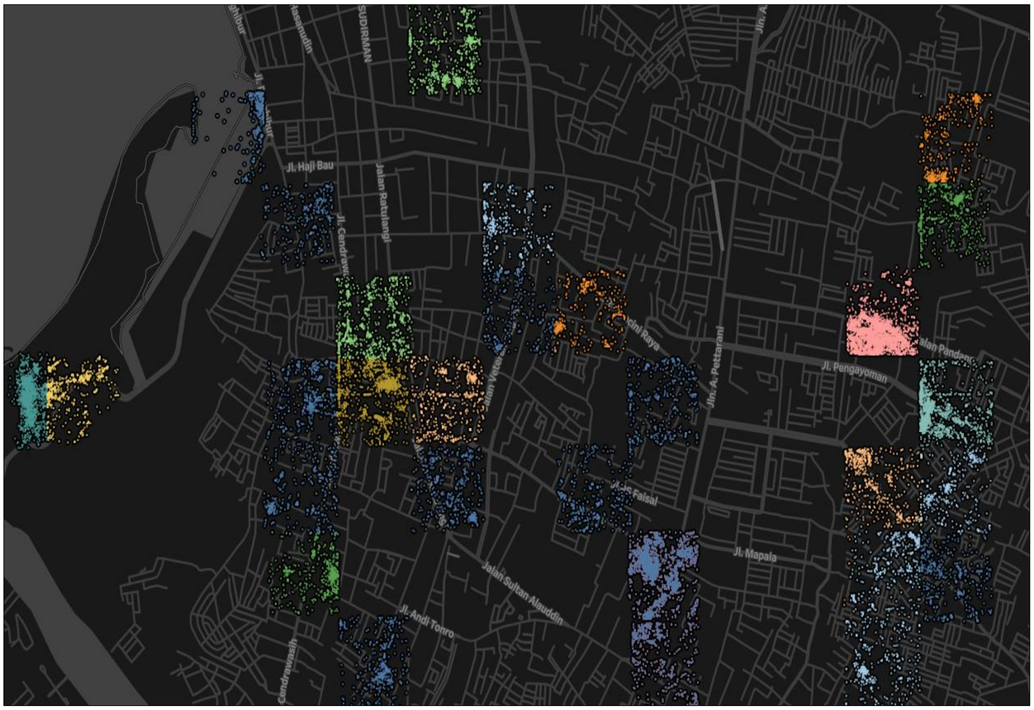
(a)



(b)


clarionmakassar
kfcindonesia
hotel
office
university
cafe
coffee
telkom
housing
mcdonald
kampus
school
bakso
fakultas
hospital
restaurant
unm_university
tsm_makassar
cinema21
cinema_xxi
telkomsel
mall_trans_studio
makan
ayam
pizza_hutid
convention
hall
yumyum_cafe
mattoangin
punya
seefood
sama
stadium
futsal
ballroom
street
universitas
mall
baru
anak
unm
mie
ice
mama
hari
dinner
watching
kasih
fmipa
umi
kedai
stadion
bismillah
alhamdulillah
kopi
milk_coffee
morning
latepost
menunggu
resto
fajar
plaza
pasar
pasar_segar
center
siang
sport_area
saja
lapangan
ketemu
goreng
teknik
basket
mall_panakukang
muhammad
smkn
birthday
menara_phinisi_unm
noodle
baik
house
makasih
masjid
mcdonalds_id
wedding
sulawesi
gedung
pettarani
phinisi_ballroom
pendidikan
good
diva_karaoke
kopiapi_coffee
toddopuli
karaoke
sayang
point
demi
home
nitro
untuk
seafood
sulsel

(c)


SUDIRMAN
Jl. Haji Bau
Jalan Ratulangi
Jl. A. Pettarani
Jl. Pengayoman
Jl. Mapala
Jalan Sultan Alauddin
Jl. Andi Tonro

(d)

Figure 5. 10 The graph of check-ins at different places (a), user time deployment activity over 24 h (b), word frequency for analysis and place identification (c), and the physical layout of tweeting activity in cluster 4 (d).

5.4 Section Conclusion

In this chapter, author used Twitter as a source of data to analyze urban land use. To investigate the regional profile, author collected information from Twitter in the form of users' text posts, time zones, and coordinates. In this paper, author proposed a grid-based aggregation method to explore urban areas. The proposed approach divided the region in the form of a grid, where on each grid, there was a 500×500 m^2 block, thus yielding 398 blocks. Author divided the area into 32 classes, where each class had 100 check-in intervals, and then classified the existing classes into some clusters. Land identification was determined on the basis of, firstly, the highest number of check-ins and, secondly, the result of a comparison of check-in patterns on weekdays and weekends.

The method proposed could characterize the urban area, particularly for land-use identification. The model used produces a polycentric area—not centered on one particular region—which means that in the city, there will be more than one similar land-use type (see Figure 5.11). For example, the education and commercial areas are not only centered on one area but also spread over several regions. Author concluded that Twitter check-in data can be used to understand the actual urban land use. Our new method can contribute additional data or input for city planners and stakeholders to solve these problems, specifically the analysis of urban land use. As such, the method author propose is cheap to implement and easy to use. In this regard, this research could become a part of the city's sustainability, specifically for the development of urban land use. To obtain maximum measurement results, this method depends and relies on the size of the used grid. For this, larger grid sizes will provide at least twice as many land-use functions in a region. In this regard, grid-size standardization is necessary for the partition of land types. This challenge needs to be considered for future research.

If author compare the ranking and *k*-means clustering methods, author found that the rank method measures on the basis of the order of data; the highest-ranking order became a standard to determine the state of the region. Meanwhile, the *k*-means clustering method used a similarity-and-distance approach to group the data. Other than being reliable, both methods can solve huge amounts of data.

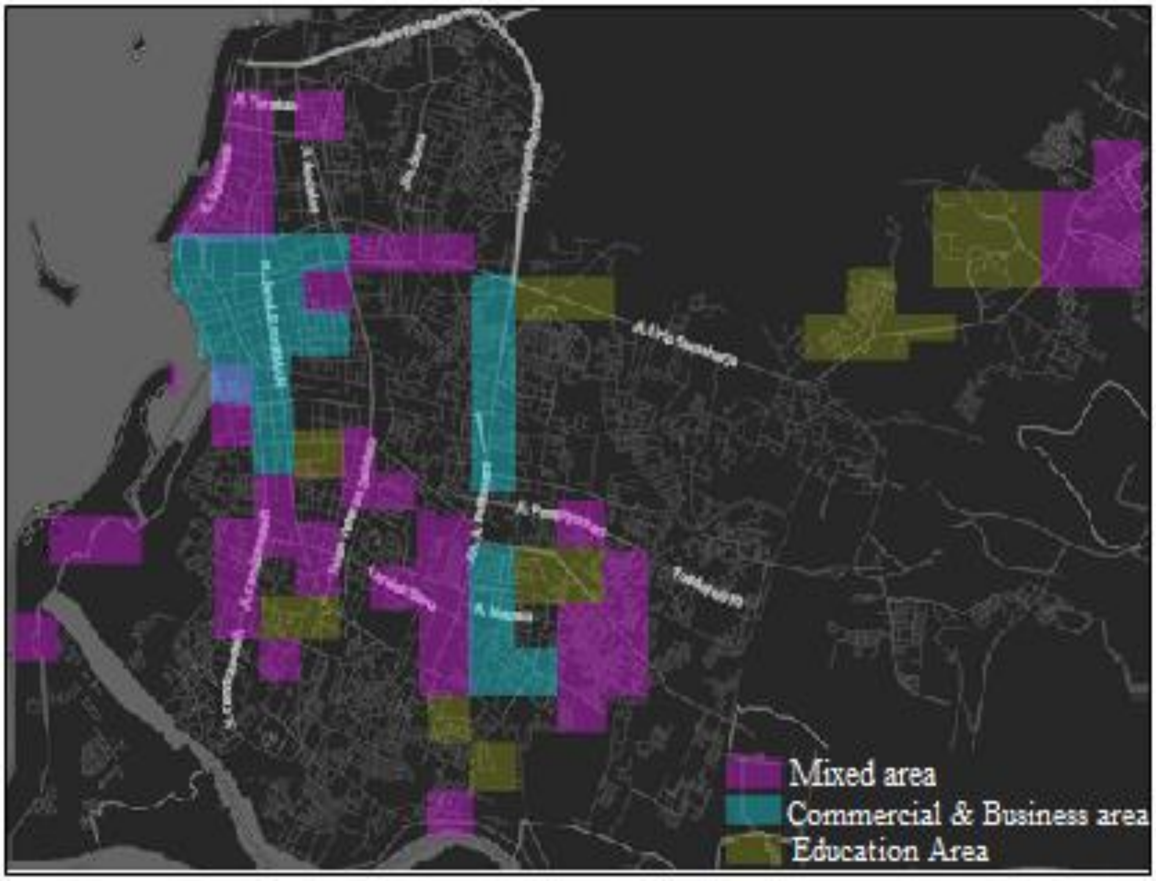


Figure 5. 11 Land use hypothesis (education, commercial and mixed area)

Appendix A

Table A1. The classes group of place activity.

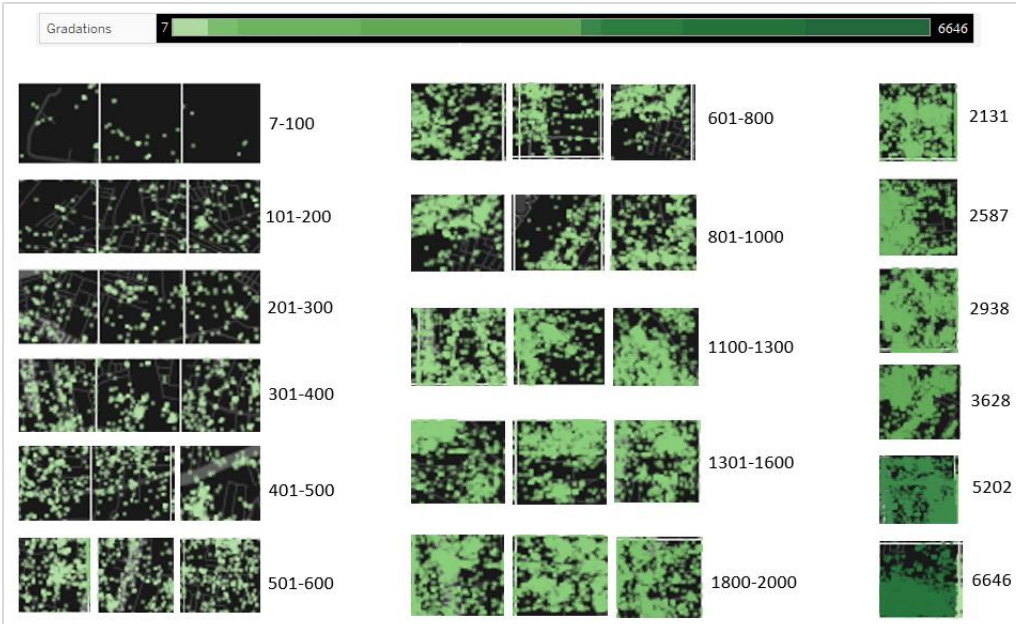
C200	Check-In	C300	Check-In	C400	Check-In	C500	Check-In
Housing	1350	Housing	1135	Housing	1060	School	866
University	469	University	885	Street	877	University	696
Office	400	School	407	University	600	Housing	457
Street	303	Street	305	Coffee	619	Coffee	556
School	299	Coffee	267	Office	427	Hospital	318
Restaurant	269	Café	253	School	349	Office	189
Coffee	454	Office	248	Culinary	268	Hotel	119
Pool	133	Park	194	KFC	268	Bank	109
Seafood	117	Meatball	165	Meatball	255	Ice	107
Beach	116	Culinary	154	Hospital	222	Street	107
Culinary	112	Restaurant	149	Beach	188	University	106
Shop	96	Hospital	119	Eating	181	Meatball	104
Park	92	Noodle	113	Hotel	150	Eating	97
Cinema21	88	Hotel	109	Noodle	114	Culinary	66
Meatball	82	Hall	92	Seafood	108	Chicken	62
Field	81	Mosque	83	Mosque	97		

C900	Check-In	C1100	Check-In	C5200	Check-In	C700	Check-In
KFC	427	McDonald	390	Mall	845	Coffee	763
Coffee	450	Ice	217	KFC	631	Hospital	268
Housing	288	Stadium	191	Cinema21	364	University	366
Hospital	242	Restaurant	319	McDonald	333	Street	212
Eating	329	Office	181	Eating	245	Office	206
Mall	164	Coffee	105	Coffee	501	Meatball	175
Noodle	126	Noodle	101	Pizza	216	Housing	174
Pizza	97	Meatball	91	Street	148	Seafood	164
Office	149	Café	68	Hotel	133	Eating	124
Soccer	73	Hotel	64	Karaoke	121	School	118
Hotel	64	Karaoke	51	Restaurant	117	Restaurant	101
Street	60	Shop	45	Culinary	117	Skincare	98
Porridge	59	Mall	44	Supermarket	100	Cheese	88
Noodle	55	Housing	39	Office	165	Eating	85
Cinema21	53	Church	37	Shop	91		
C1200	**Check-In**	**C3300**	**Check-In**	**C600**	**Check-In**	**C1000**	**Check-In**
School	592	Coffee	606	Hotel	1354	Coffee	488
Church	91	KFC	219	Hall	341	University	344
Coffee	86	Cinema21	194	University	314	School	322
Culinary	147	Market	142	Café	168	Culinary	140
Coffee	73	Mall	122	School	136	Restaurant	119
Restaurant	138	Hotel	52	Corner	87	Housing	108
Office	51	Street	87	Office	66	Cinema21	107
Hotel	42	Bar	80	Street	65	Noodle	73
Culinary	38	Tea	77	School	65	Shop	67
Mall	38	Eating	72	Building	57	Mall	56
Clinic	35	Karaoke	65	Wedding	56	Office	48
Store	29	Culinary	182	Swimming	55	Eating	45
Mall	28	Pizza	63	Garden	47	Bank	38
Donuts	26	Snack	58				
C1800	**Check-In**	**C2500**	**Check-In**	**C6600**	**Check-In**	**C1300**	**Check-In**
Pizza	291	Culinary	544	Mall	1145	University	928
Coffee	361	Hotel	263	Cinema	1025	McDonald	371
University	224	Office	114	Tea	347	KFC	199
Culinary	190	Bar	92	Supermarket	250	Hospital	186
School	305	Mall	159	Pizza	191	Office	153
Beach	187	Culinary	102	Mall	188	Street	152
Restaurant	482	Tower	82	Coffee	209	Coffee	247
Bar	131	Park	60	Eating	102	Restaurant	77
Meatball	119	Bank	52	Restaurant	313	Monument	161
Office	106	Hospital	73	Bank	61	Pizza	60
Hall	84	Eating	44	Bookstore	70	Noodle	56
Bank	70	Coffee	32				
C7600	**Check-In**	**C3600**	**Check-In**	**C2900**	**Check-In**	**C1900**	**Check-In**
Mall	1720	Hotel	1106	Restaurant	444	McDonald	914
Cinema	935	Office	571	Fort	296	Coffee	344
KFC	159	University	325	Office	239	Office	172
Tea	243	Café	165	Coffee	200	Eating	163

Eating	183	School	187	Park	67	Culinary	271
Coffee	294	Ballroom	100	Food	44	Hotel	77
Pizza	141	Happy	98	Bar	87	Steak	77
Restaurant	180	Corner	87	Culinary	76	Ice	74
Snack	95	Wedding	56	Hotel	36	University	64
Bookstore	70	Street	131	Eating	137		
C1700	**Check-in**	**C2100**	**Check-in**	**C2400**	**Check-in**	**C800**	**Check-in**
Mall	687	Field	332	School	236	University	2138
Cinema21	318	KFC	125	KFC	137	Office	564
Restaurant	101	School	246	Culinary	179	School	392
Coffee	119	Field	86	Hotel	142	Culinary	280
Tea	39	Office	152	Field	67	KFC	221
Dinner	33	Mall	70	Bank	60	Seafood	139
Lunch	27	Street	70	Coffee	137	Pizza	123
Bank	25	Pizza	53	Hospital	41	Coffee	116
Snack	25	Coffee	96	Restaurant	64	Soccer	108
Fitness	18	Bank	42				
C2000	**Check-in**	**C2200**	**Check-in**	**C1500**	**Check-in**	**C100**	**Check-in**
Restaurant	568	University	1897	University	1715	Housing	779
Hotel	175	Beach	438	Café	992	University	135
Café	169	Restaurant	258	Cinema21	228	Street	111
Bar	139	KFC	241	Mall	241	School	101
Guesthouse	84	Culinary	200	Building	130	Café	41
Office	42	Coffee	325	Library	233	Coffee	37
Hospital	34	Hotel	508	Meatball	61	Restaurant	36
Eating	52	Hall	84	Hotel	59	Office	29
Culinary	62	Hospital	106	School	57	Culinary	28
C1600	**Check-in**	**C3000**	**Check-in**	**C1400**	**Check-in**	**C2300**	**Check-in**
Stadium	849	Mall	1600	University	2117	University	574
Office	101	Restaurant	342	Office	603	School	958
Photography	62	Cinema	225	Hospital	191	Futsal	39
Soccer	57	Coffee	150	Building	92	Hospital	13
School	45	Bar	77	School	126	Mosque	18
University	87	Snacks	133	Hall	164		
Culinary	22	Eating	31	Canteen	52		
Television	17	Fitness	30				

Appendix B

An examples blocks with check-in activity



References

1. Al-Tahir, R.; Rajack, F.; Oatham, M. Aerial photographs for detecting land use changes in Valencia Wildlife Sanctuary and Forest Reserve, Trinidad. *Caribb. J. Earth Sci*. **2005**, *38*, 35–42.
2. Modara, M.; Belaid, M.A. Mapping and assessing land use/land cover change in Muharraq Island based on GIS and remote sensing integration. *Remote Sens. Spat. Inf. Sci*. **2013**, *XL-4/W1*, 57–63.
3. Reis, S. Analyzing land use/land cover changes using remote sensing and GIS in Rize, North-east Turkey. *Multidiscip. Digit. Publ. Inst*. **2008**, *8*, 6188–6202.
4. Fonji, S.F.; Taff, G.N. Using satellite data to monitor land-use land-cover change in North-eastern Latvia. *Springerplus* **2014**, *3*, 61, 1-15
5. Hu, T.; Yang, J.; Li, X.; Gong, P. Mapping urban land use by using landsat images and open social data. *Remote Sens. Spat. Inf. Sci*. **2016**, *8*, 1–18.
6. Kawakubo, F.S.; Morato, R.G.; Nader, R.S.; Luchiari, A. Mapping changes in coastline geomorphic features using landsat TM and ETM imagery: examples in South Eastern Brazil. *Int. J. Remote Sens.* **2011**, *32*, 2547-2562.
7. Haeusler, M.H. Enabling low cost human presence tracking. In Proceeding of the International Conference of the Association for Computer-Aided Architectural Design Research in Asia CAADRIA; Melbourne, ON, Australia, 30 Marc–2 April **2016**; pp. 45–54.
8. Song, J.; Tang, E.Y.; Liu, L. User behavior pattern analysis and prediction based on mobile phone sensors. In Proceedings of the 2010 IFIP International Conference on Network and Parallel Computing, Zhengzhou, China, 13–15 September 2010; Springer: Berlin/Heidelberg, Germany, **2010**; pp. 177–189.
9. Zhang, A. Detecting Human Behavior Patterns from Mobile Phone. *J. Comput. Inf. Syst*. **2012**, *8*, 2671–2679.

10. Jankowska, M.M.; Schipperijn, J.; Kerr, J. A framework for using GPS data in physical activity and sedentary behavior studies. *Exerc. Sport Sci. Rev*. **2015**, *43*, 48–56.
11. Munizaga, M.; Devillaine, F.; Navarrete, C.; Silva, D. Validating travel behavior estimated from smartcard data. *Transp. Res. Part C* **2014**, *44*, 70–79.
12. Van der Spek, S.; van Schaick, J.; de Bois, P.; de Haan, R. Sensing human activity: GPS tracking. *Sensors* **2009**, *9*, 3033–3055.
13. Farshad, A.; Marina, M.K.; Garcia, F. Urban wifi characterization via mobile crowdsensing. In Proceeding of the IEEE NOMS, Krakow, Poland, 5–9 May **2014**.
14. Soto, V.; Martinez, E.F. Automated land use identification using cell-phone records. In Proceeding of the 3rd ACM International Workshop on MobiArch, Bethesda, MA, USA, 28 June 2011; ACM: New York, NY, USA, **2011**; pp. 17–22.
15. Toole, J.L.; Ulm, M.; González, M.C. Inferring land use from mobile phone activity. In Proceeding of ACM SIGKDD International Workshop on Urban Computing, Beijing, China, 12 August 2012; ACM: New York, NY, USA, **2012**; pp. 1–8.
16. Ratti, C.; Pulselli, R.M.; Williams, S.; Frenchman, D. Mobile Landscapes : using location data from cell-phones for urban analysis. *Environ. Plan.* **2006**, *33*, 727–748.
17. Hasan, S.; Zhan, X.; Ukkusuri, S. V. 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,* Chicago, IL, USA, 11–14 August 2013; ACM: New York, NY, USA, **2013**; p. 6.
18. Sun, Y.; Li, M. Investigation of travel and activity patterns using location-based social network data : A case study of active mobile social media users. *ISPRS Int. J. Geo. Inf*. **2015**, *4*, 1512–1529.
19. Hasan, S.; Ukkusuri, S.V. Location contexts of user check-ins to model urban geo life-style patterns. *PLoS ONE* **2015**, *10*, e0124819.
20. Noulas, A.; Scellato, S.; Lathia, N.; Mascolo, C. Mining user mobility features for next place prediction in location-based services. *Proc. IEEE Int. Conf. Data Min. ICDM* **2012**, doi:10.1109/ICDM.2012.113.
21. Mahmud, J.; Nichols, J.; Drews, C. Home Location Identification of Twitter Users. *vACM Trans. Intell. Syst. Technol*. **2013**, *5*, 1–47.
22. Williams, E.; Gray, J.; Dixon, B. Improving geolocation of social media posts. *J. Pervasive Mob. Comput.* **2017**, *36*, 68–79.
23. Kong, L.; Liu, Z.; Huang, Y. SPOT : Locating social media users based on social network context. *VLDB* **2014**, *7*, 1681–1684.
24. Central Bureau of Statistic. Available online: http://sp2010.bps.go.id/ (accessed on 18 October 2017).
25. Wabula, Y.; Dewancker, B.J. Analysis of urban population using twitter distribution data : Case study of makassar city, indonesia. *Int. Comput. Electr. Autom. Control Inf. Eng. Waset* **2016**, *10*, 1627–1631.
26. Land Use Map of Makassar City. Available online: http://darimakassar.com/rtrw-kota-makassar-2010-2030-2/ (accessed on 18 October **2017**).
27. Irfan, R.; King, C.K.; Es, G.; Ewen, S.; Khan, S.U.; Madani, S.A.; Kolodziez, J.O.; Wang, L.; Chen, D.; Rayes, A.; et al. A survey on text mining in social networks. *Knowl. Eng. Rev.* **2015**, *30*, 157–170.
28. Varghese, B.M.; Jose, J.T.; Unnikrishnan, A.; Poulose, K.J. Clustering Student Data to Characterize Performance Patterns. *Int. J. Adv. Comput. Sci. Appl. Spec. Issue Artif. Intell.* **2011**, *2*, 138–140.
29. Mihai, D.; Mocanu, M. Statistical considerations on the *k*-means algorithm. *Ann. Univ. Craiova Math. Comput. Sci. Ser.* **2015**, *42*, 365–373.

Chapter 6

Quantifying of Urban Inhabitant Mobility for Urban Planning Data

6.1 Research Purpose

Knowledge of human mobility patterns within cities is central to better urban planning. Researchers have proven that human mobility plays vital roles in planning urban infrastructure [1], urban development and human migration [2], and the development of transportation facilities [3]. In the previous approach, there were many ways to measure the flow of citizens. Such information was usually gathered through a traditional survey method or by using questionnaires that attempt to capture how citizens interact with their environment [4, 5, and 6], and the urban demographics data related to where people live and work [7]. The presence of technological devices produces individual traces and human spatial behaviours that have not been discovered before. Data on mobile phones users [8], personal digital assistants [9], and GPS devices have provided individuals' mobility information [10]. Through GPS devices, individual travel activities to places visited can be recorded such as information about times, days, and even the types of transportation used. In addition, the smartphone can offer the information about the location at which the call occurred. This data has become important since most citizens have a smartphone. Thus, this device can become a sensor to explain people's movements.

In addition to the above mentioned, the author uses social media as additional data to understand human mobility in Makassar City, Indonesia. First each social media user was identified by analysing the people with certain check-ins (see Figure 6.1). This is necessary because the involvement of active users makes it easy to investigate human movement. Then the distance travelled by each user (km) was calculated and the type of places that people visit was identified. To recognize the name of a venue, the text that the user posted on Twitter was used as a key to determine the name of the place.

The objective of the research is to understand the pattern of mobility of urban inhabitants using a geolocation service and user texts posted on Twitter. More specifically, the research will answer the following questions:

1. What are the movement patterns of a citizen?
2. How far do people travel in the city?

Figure 6.1 Check-in distribution of all locations with 546 users during a 30-day period

6.2 Material and Methods

6.2.1 Dataset Structure

In recent years, the number of social networking users in the world has grown by leaps and bounds. Millions of data generated from these sites provide information about human movement. Many features are provided by social media developers to make it easier for a user to communicate. Besides the status update feature, users also can attach the location embedded in the posted message. The location information shared indicates a place where a person conducts a social media activity

The study took place in Makassar City, Indonesia. The dataset consists of 30 days (four weeks), starting from September 1st to the 30th, 2016 with 201,118 check-ins and 22,318 users.

Table 6. 1 Dataset Detail

Original dataset	Check-in
Number of Check-ins	38185
Number of Users	546
Study Sample	
Number of Check-ins	2570
Number of Users	54

6.2.2 Data Analysis

To accommodate individual mobility, the author only analyzed users with more than 30 check-ins. The next step was the filtering process to obtain the 38,185 check-ins with 546 users (see Table 6.1) used in this study. To determine the number of samples in the research, the author used the formula

S = 1/10*P, (S is sample, and P is population), thus producing 54 users.

To spread the population evenly, the author used a sampling method with a systematic sampling approach. The technique takes a sample based on alphabetical sequences of the Twitter username. For example, every user who has the first letter M will be taken twice. If the first letter of each sequence of the alphabet is processed twice, then (26*2 = 54). This amount is equal to the number of research samples (see Table 6.2).

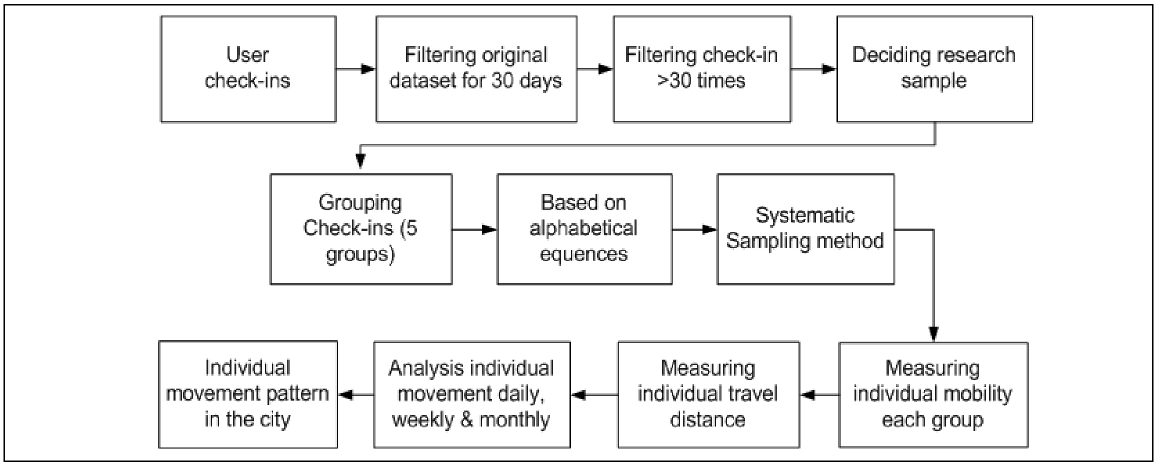


Figure 6.2 Data flow diagram of method used

6.3 Analysis of Individual Travel Distance

To analyze the individual movement of the study sample, the check-in activity was split into five groups. Each group contains a place activity and an individual's mobility distance during the study period—in this case, how far they travel when they take a trip. Then, in each group, the types of places visited were also identified. Figure 6.3 shows the users' deployment and check-ins based on groups.

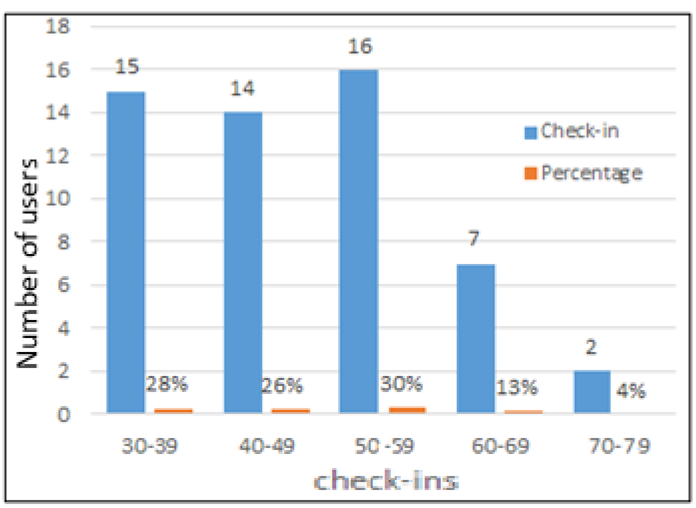


Figure 6.3 Check-ins and percentage of users for each group

6.3.1 Individual Movement Pattern of Group 1

Referring to Table 2 of group 1, users in this group have a frequency of check-in activity of around 30 to 39 times with 15 users. In the first week, an individual's average travel reached 48.03 km. The author observed that their movement distance varied from 1 km to 48 km in the 1st week, 1 km to 159.7 km in the 2nd week, 1 km to 54 km in the 3rd week, 4 km to 132 km in and the 4th week. Then the user's average distance per week was analyzed, which was 48.0 km in week 1, 85.0 km in week 2, 52.9 km in week 3, and 84.8 km in week 4. From the results, it can be concluded that the daily average mobility of people in group 1 was about 9.2 km.

In this group, the type of location that people visit was identified. Almost all the users show check-in activity at places such as the university (25%), school (19%), hotel (17%), home (13%), dormitory (9%), café (9%), and McDonalds (8%) (see Figure 6.4).

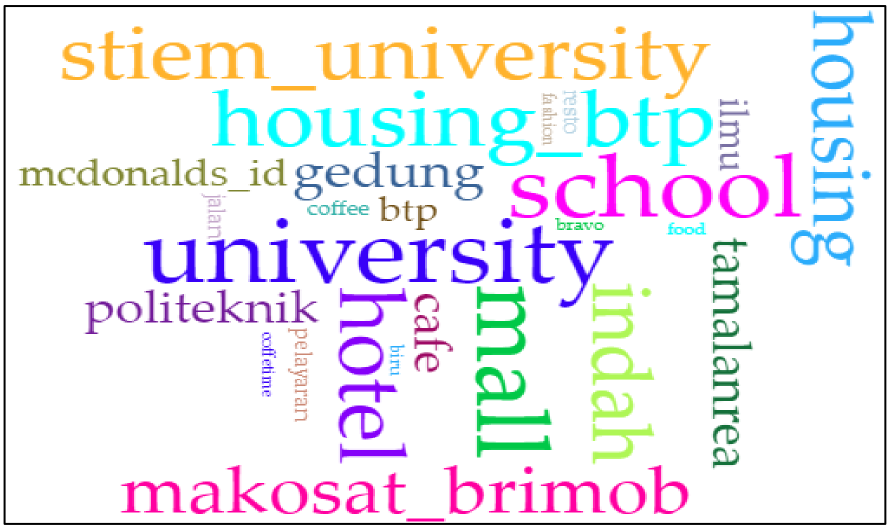


Figure 6.4 The word frequency for places visited by group 1

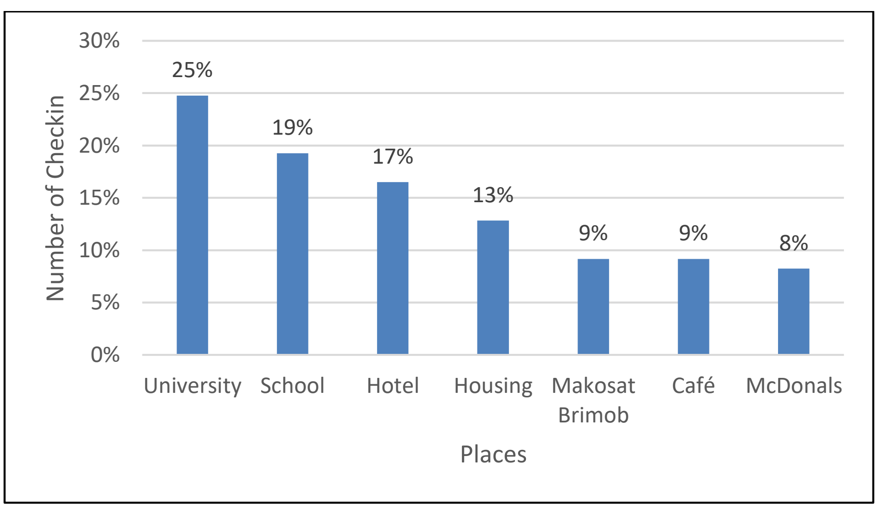


Figure 6.5 The percentage graph of different places visited by group 1

6.3.2 Individual Movement Pattern of Group 2

In this group, the average check-in activity frequency is 40 to 49 times. An individual's mobility per week was observed first. From the results of the analysis (see Table 2 of group 2), the maximum movement of a user reached 81.2 km with the following characteristics: in week 1, the average total journey length was about 45.12 km, with the shortest distance being 1.6 km and the furthest distance being 48.3 km. In week 2, the distance of the total average journey was 68.25 km, with the shortest and the longest journey lengths being 9.7 km and 77 km, respectively.

Meanwhile for week 3, the average total movement reached 78.97 km, with the shortest distance being 2.5 km and the longest being 81.2 km. While in week 4, the total average length of journeys for group 2 reached 82.03 km, with the shortest and longest distances being 21.3 km and 64.4 km, respectively. It was concluded that the average length of individual trips per day for four weeks in group 2 was approximately 9.15 km.

At the same time, the author identified the kinds of places that people visited. Figure 6.6 shows the word frequency percentage of check-in venues. Analysis of the results shows that, generally, user check-ins were at places such as tour and travel shops (20%), cinemas (12%), restaurants (13%), coffee shops (11%), a dormitory (11%), a diner (9%), a mall (9%), a stadium (8%), and a gym (7%) (see Figure 6.7).

Figure 6.6 The word frequency for places visited by group 2

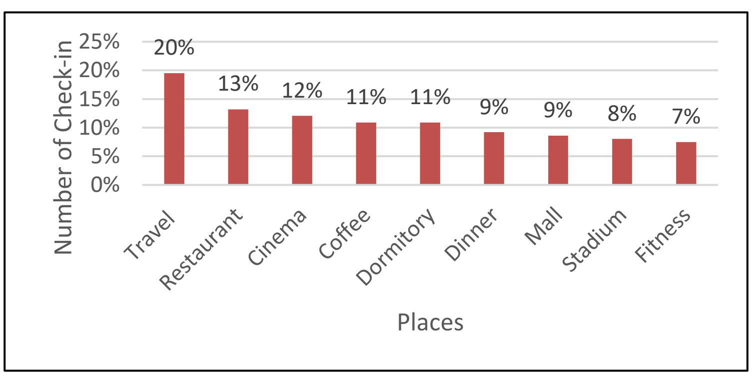


Figure 6.7 The percentage graph of different visited places by group 2

6.3.3 Individual Movement pattern of Group 3

As shown in Figure 6.8, the places related to group 3 were analyzed. The majority of tweet activity in this group is covered by location, such as for instance, the mall (22%), office (20%), faculty (11%), hotel (8%), nursing college (8%), KFC (7%), university (7%), hajj dormitory (6%), and the cinema (8%). This group is dominated by place activity, e.g., the mall, office, and university faculty.

In this group, 40–49 check-ins by 16 users were analyzed. Referring to Table 2 of group 3, that the highest mobility was found to be 150 km. The average mobility of subjects was 84.34 km, with the shortest and longest distances being 3.3 km and 84.4 km, respectively. Then, in the second week, mobility increased with the shortest and longest distances being 13 km and 94.3 km. While in week 3 and week 4, the shortest and longest distances reached 6.7 km and 150 km, and 9.2 km and 94.9 km, respectively. It can be concluded that the average distance traveled per day during the four-week period was 14 km.

Figure 6. 8 The word frequency for places visited by group 3

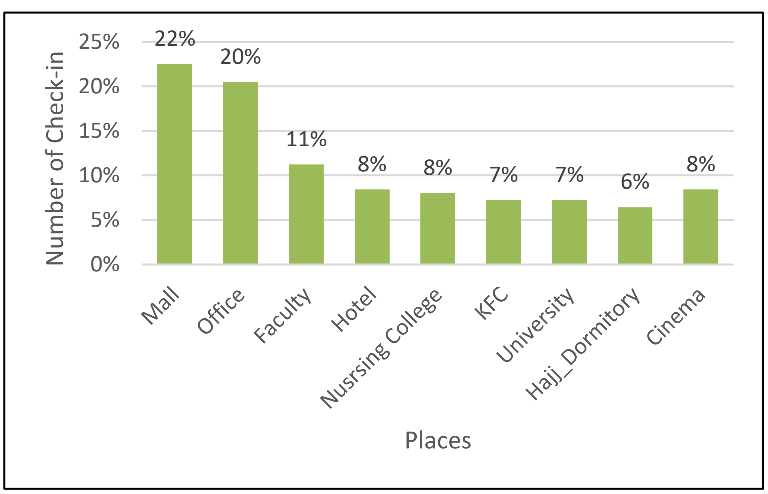


Figure 6. 9 The percentage graph of different visited places by group 3

6.3.4 Individual Movement pattern of Group 4

In general, the user activity in this group was tightly related to the individual's activity in the university (see Figure 6.10). Due to their activities around the university, it was concluded that their status was that of a student. It was found that about 53% of user check-ins were sourced from the polytechnic, 14% from school, 13%, housing, 13% Poltekkes (health polytechnic), 12% dormitory, 12% law faculty, 10% basket, 10% school, 9%, KFC, and 7% coffee. The existence of venue activities such as basket, coffee, and dormitory, were individual actions conducted around the college.

This group displays the spatial distribution of seven users with a check-in frequency between 60-69 times. From Table 6.2 in group 4, their highest spatial movement was 123.2 km with the following characteristics: in the first week, the lowest mobility was 23.7 km and the highest was 84 km. Then, during the second week, the minimum distance of their trips was 15.8 km and the maximum distance of their travel was 78.9 km. Meanwhile, the shortest and the greatest distances for the third week and the fourth week were 19 km and 123 km, and 7.9 km to 103.7 km, respectively. It was concluded that their daily travel average was 6.56 km.

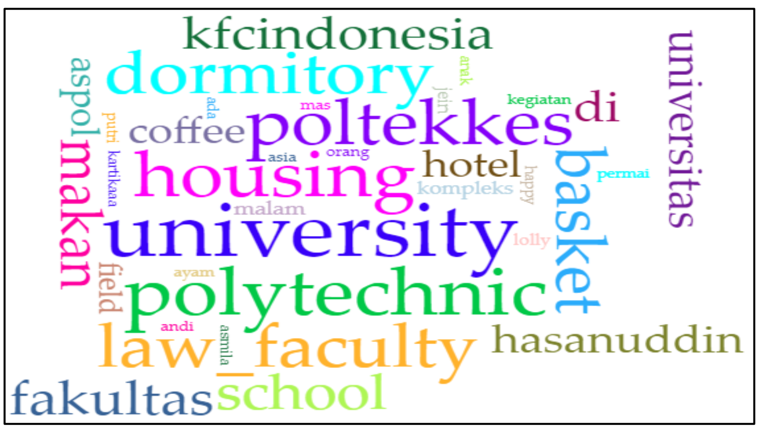


Figure 6.10 The word frequency for places visited by group 4

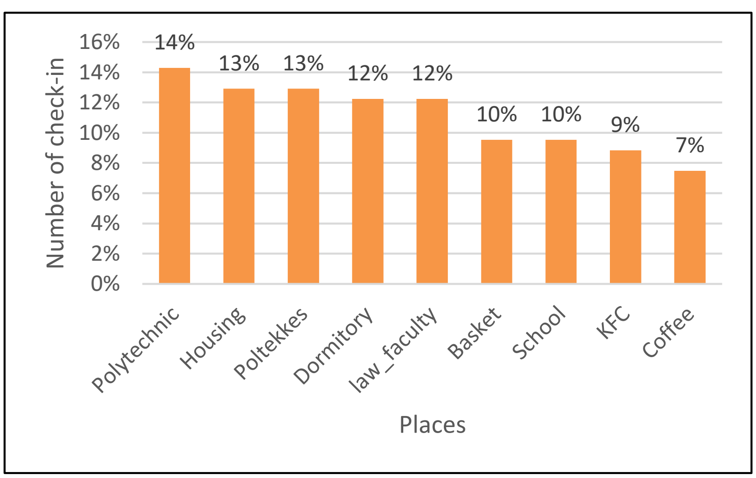


Figure 6.11 The percentage graph of different places visited by group 4

6.3.5 Individual Movement pattern of Group 5

As shown in Figure 6.12, the majority of tweet activity was conducted at the university and beach. The places percentage was dominated by activities at the university (30%), beach (22%), high school (18%), hotel (10%), McDonalds (7%), culinary shop (meatball) (7%), and photo studio (7%).

The check-in activity for this group was between 70-79 check-ins for two users. From Table 2 of group 5, it was observed that the individual with the highest mobility distance was 166 km. For week 1 and week 2, the average user journey length was 30.59 km and 78.7 km, respectively. For week 3 and week 4, the average user journey length was 14.66 km and 13.05 km, respectively. It was concluded that the total of the individual average distance travelled daily of group 5 was 9.63 km.

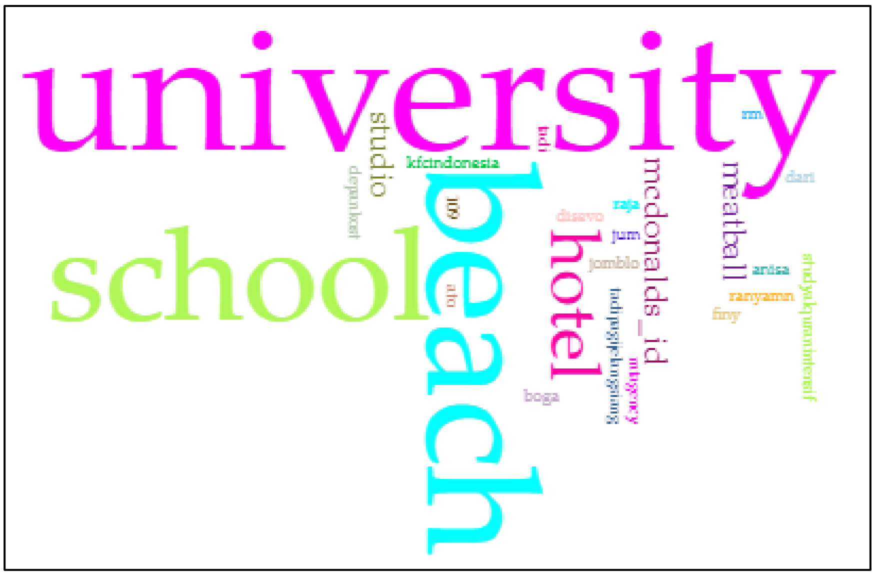


Figure 6. 12 The word frequency for places visited by group 5

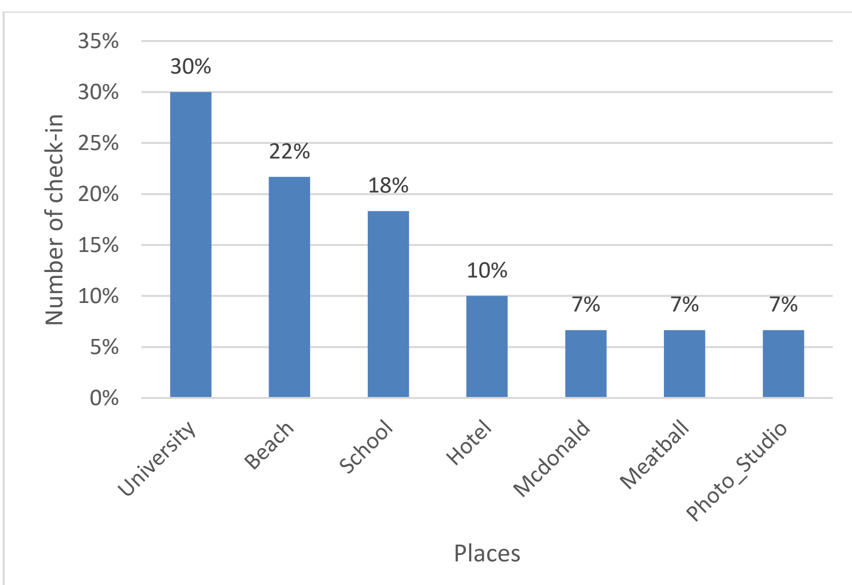


Figure 6. 13 The percentage graph of different places visited by group 5

6.4 Section Conclusion

The mobility dataset used in this study was collected via the Twitter streaming application program interface (API). Three feature data were used: check-in (specific location), time stamp, and user's status text or post activities. From this, the study measured the displacement distance of each user (daily and weekly periods) from one point to another point based on the check-in parameters. In this analysis, a systematic sampling approach was used to decide the number of research samples from the Twitter user population. This chapter presented a method for analyzing human mobility in Makassar City.

Analysis of the results illustrates that individual movement shows a high level of regularity and intensity in a specific location and at a certain time. Individuals tend to check-in at locations where their daily activities take place. For example, almost every day, the participants in this study use social media at the university, which can be seen in the user text activities posted in each group (see Figures 4 (a) and (c)–(e)). Secondly, the tendencies of the subjects were almost the same; aside from the university, the next most visited destination was a shopping mall. It is worth noting that for this activity, the author cannot be sure if the purpose of the visit was to shop or be engaged in another activity (e.g., meeting friends at a coffee shop). In general, the movement pattern of the subjects in the study is: university-mall-home, university-dormitory, office-mall-home, and office-home and other. Then their average weekly mobility were 1st week: 30.45km, 2nd: 43.72, 3rd: 40.27, 4th: 44.88.

Table 6. 2 Distribution of individual weekly travel distance

		Trips per week				
User Code	**Username**	**1st week (km)**	**2nd weeks (km)**	**3rd weeks (km)**	**4th weeks (km)**	**Number of Check-ins**
			Group 1			
A54	4rifxxxx	48.1	n/a	n/a	76.4	31
A39	Fuxxxx	26.8	159.7	n/a	46.9	31
A25	musxxxx	12.9	43.6	22.7	18.2	31
A21	ONxxxxx	22.0	18.5	30.5	39.7	31
A3	yuzxxxxx	1.3.0	9.30	15.8	7.30	31
A44	dwixxx	4.40	69.7	37.7	38.0	32
A34	iyxxxxx	15.0	25.3	23.3	24.7	32
A52	ayxxxxxxxx	48.0	37.2	54.7	21.3	34
A42	Ernhxxxxxxx	42.4	n/a	n/a	132.6	34
A2	zulkxxxxx	9.60	51.5	49	19.8	34
A26	luxxxxxxxxxxx	29.2	66.2	30.4	4.40	35
A33	Iykxxxxxxxxxx	14.7	37.6	30.9	27.6	36
A7	vivxxxxxxx	46.6	8.4	12.7	64.6	36
A53	Aaxxx	6.90	44.0	37.8	2.40	37
A20	Oyxxxxxxxxx	8.20	24.1	24.7	69.4	38

Group 2						
A18	Pyaxxxxxxxxxxx	1.60	38.1	59.0	39.3	41
A16	Qhyxxxxx	33.3	10.2	35.1	59.1	41
A45	Dwiixxxxx	18.4	36.5	50.9	21.8	42
A31	judxxxxxx	25.4	9.7	26.2	25.9	42
A22	nysaxxxxxx	25.9	20	43.9	47.3	42
A19	puttxxxxxx	36.9	54.2	13.9	56.7	42
A50	ayuxxxxx	48.3	17.2	50.4	21.3	43
A30	junxxxxxx	6.00	50.7	41.4	48.2	43
A11	Tyxxxxxx	22.5	33.0	81.2	32.5	43
A24	Muxxxxx	36.5	39.7	28.8	23.2	44
A1	Zultxxxxxx	n/a	0.20	2.50	42.0	44
A43	dyxxxx	25.7	45.4	53.1	47.5	45
A41	GeRxxxx	33.7	45.5	13.0	42.9	45
A35	husexxxx	1.00	77.6	53.4	66.4	46
Group 3						
A28	Kusxxx	48.1	41.7	55.6	37.3	50
A23	nuzxxxxxxxxx	21.2	89.8	54.5	37.8	50
A48	budixxxxx	14.3	13.0	27.2	94.9	51
A38	Ghifxxxx	24.1	48.8	33.7	87.3	51
A46	creaxxxx	38.5	43.6	34.5	18.5	52
A37	Gitaxxxxx	40.5	43.4	89.4	83.4	52
A29	Kirxxxx	3.3	70.3	15.9	77.0	53
A12	Tryxxxxx	41.0	46.2	74.7	53.1	53
A27	lukixxxxx	9.10	51.0	6.7.0	n/a	54
A17	qadxxxx	96.0	23.2	31.2	62.8	54
A32	Joxxxx	84.4	15.0	22.7	34.4	56
A8	Vixxxx	18.1	33.3	79.1	29.7	56
A49	blaoxxxxxxx	70.1	52.1	18.2	38.7	57
A4	Yuyxxxx	21.1	34.5	150.2	9.20	57
A47	cracxxxx	22.1	61.5	71.2	49.6	59
A10	Uwxxxx	38.4	94.3	91.1	17.1	59
Group 4						
A13	syaxxxxxx	23.7	47.3	52.8	103.7	64
A13	Syaxxxx	n/a	n/a	17.1	84.8	64
A14	Ryaxxx	43.2	78.9	22.4	47.5	65
A36	Hndrxxx_	54.4	42.4	27.5	7.90	66
A51	Ayxxx	82.2	15.8	123.2	28.3	67
A15	Rumxxx	84.0	76.8	19.6	78.8	69
A9	uyxxx	50.0	50.7	30.6	82.4	69
Group 5						
A5	wulaxxxx	n/a	166	70.6	38.3	73
A6	wirxxx	63.8	48.2	32.1	53	74

AVERAGE	30.45	43.72	40.27	44.88
Individual daily mobility: 9.56 Km				

References

1. Gonzalez, M.C.; Hidalgo, C.A.; Barabasi, A.L. Understanding individual human mobility patterns. *Nature*, **2018**, 453, 779-782.
2. Noulas, A.; Salvatore, S.; Lambiotte, R.; Pontil, Mascolo, C. A tale of many cities: universal patterns in human urban mobility. *PloS one*, **2012**, 7.
3. Song, C.; Koren, T.; Wang, P.; Barabasi, A. L. Modelling the scaling properties of human mobility. *Nature Physics*, **2010**, 6, 818-823.
4. Ewing, R.; Cervero, R.; Travel and the built environment: A synthesis", *Transportation research record*, 01-3515, 87–113.
5. Vilhelmson, B. Daily mobility and the use of time for different activities", *Geo-Journal*, **1999**, 48, 177–185.
6. Hanson, S. Perspectives on the geographic stability and mobility of people in cities. Proceedings of National Academy of Science of the USA. **2015**, 102, pp. 15301-15306.
7. Thomas, E.; Serwicka, I.; Swinney, P. Urban Demographics Where People Live and Work, *DAC beachroft*. **2015**.
8. Boldrini, C.; Passarella, A. HCMM: Modelling spatial and temporal properties of human mobility driven by users' social relationships. *Computer Communications*, **2010**, 33, 1056–1074.
9. Song, C.; Qu, Z.; Blumm, N.; Barabasi, A.L. Limits of predictability in human Mobility. *Science*, **2010**, 327: 1018–1021.
10. Cao, X.; Cong, G.; Jensen, C.S. Mining Significant Semantic Locations from GPS Data. In the *Proceedings of the VLDB Endowment*, **2010**, 3, pp. 1009-1020.
11. Twitter. Open twitter streaming api. (2015, August 26). *https://dev.twitter.com/docs/streaming-api*.

Chapter 7
City Center as a Part of Urban Core

7.1 Research Purpose

A city center is a central place for activity among the urban population. In this zone there is a central area with high accessibility to economic activities. The city center serves a range of important economic roles: it is a place of trading and the civic heart of the community [1] and plays a central part in business transactions conducted. Development of the city center can reflect the broader dynamics of the digital era and shifting demographic trends [2]. Concerning the city center, it is necessary to create boundaries in anticipation of the possible overlap in determining the direction of sustainable urban development.

In this chapter, the author tries to use other methods and datasets to analyse the location of the city center. Our data, sourced from Twitter social media data, attempts to delineate the boundaries of the city core. To construct the research, the author proposes a grid-based aggregation method and text mining to split the Twitter land map. The proposed method uses a grid to divide the urban area and text-mining activity to count popular keywords among different categories. Two kinds of data that are presented in this research are check-ins and a user's status text or post activities.

7.2 Data Analysis

7.2.1 Grid Based on Agregation Method

In this part, the aggregation grid is defined as the relationship between one object (check-in) with another object, where each object and other objects are separate but united in some group. To formulate the method, the author split the urban area in the form of a grid or block. Every block contains the check-in activity. To facilitate the analysis, each block area is divided into 500 x 500 m^2; this generated 558 blocks. Figure 7.1 illustrates the spread of Twitter check-ins in the form of grids. The dots represent the locations, and the cell gradations indicate the check-in frequency. The steps in the method used to analyse the city center can be seen on Figure 7.3.

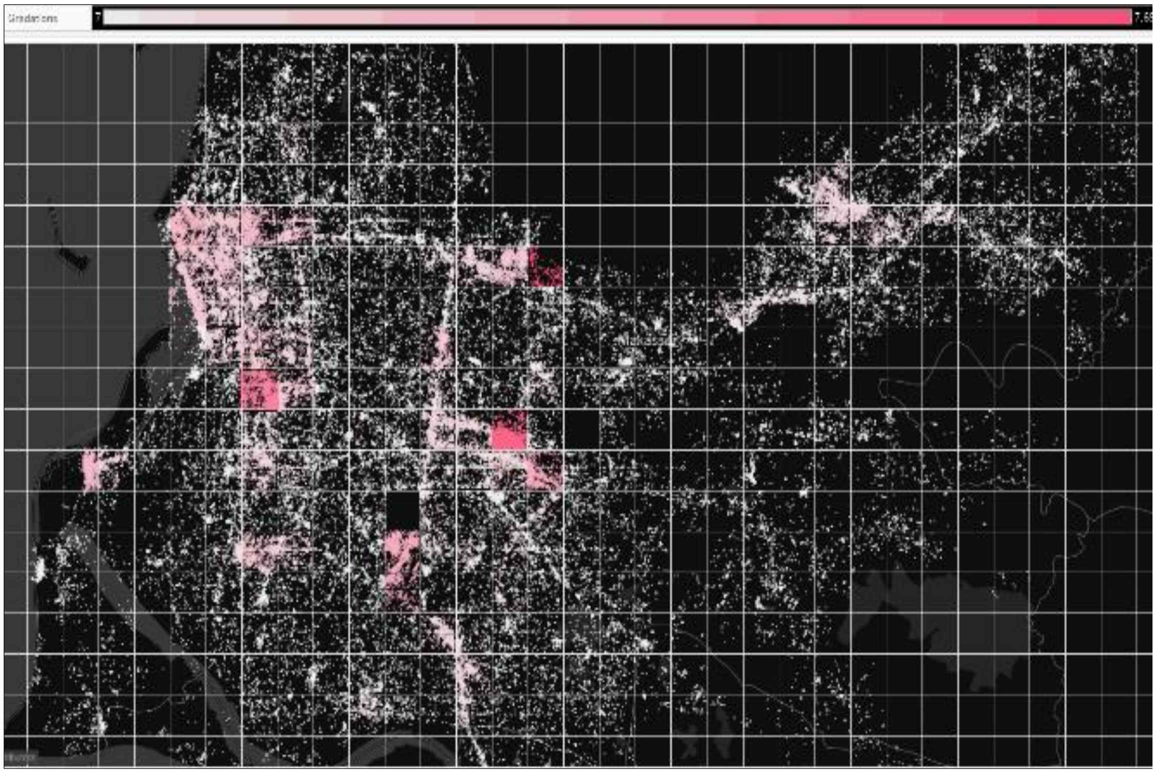


Figure 7.1 Grid distribution of check-ins

7.2.2 Text Mining for Place Name Identification

Due to Twitter only showing the location's position in coordinate form, the specific location's name is not identified. To identify the place name the author utilized the text post activity feature. It includes a symbol (@ and #) as criteria to determine the location name (e.g., "education at school," "shopping @the mall," and "eating #the restaurant"). From these texts, then author identified the most frequently used words and counted the percentage of document text produced. The goal is to know the type of place people are interested in on Twitter. The result of this modeling was used to identify the most commonly visited areas.

7.2.3 Land Mark Identification

To facilitate the analysis, the author designed specific criteria to define the city center. These criteria aim to identify the essential factors in coming to a decision. Below are the steps to identify the city center:

a. Grouping the area based on the busiest part (see Figure 7.2)
b. Landmark identifications of each group (e.g., plaza, beach, hotel, mall, restaurant, and university) as a centre of measurement. From the grouping results, seven groups were produced as candidates for the city centre.

c. Grouping the candidate landmarks, then calculating the frequency and percentage of each place. Counting the value of the place category by the formula:

$$\mathbf{p} = \frac{f}{n} * 100$$

Where p = percentage, f = frequency, and n = total.

d. Specifying the city center. In this step, the groups are matched with nearby candidate landmarks. Determination of the city center is arrived at based on the number of check-in activities in places such as commerce, entertainment, shopping, and business areas. In this regard, the decision is made if the criteria are suitable. Otherwise, if the requirements do not suit, then is not suggested as the candidate.

Figure 7.2 The area grouping with the busiest part

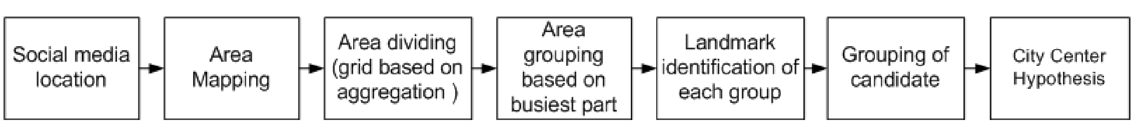


Figure 7.3 The steps to identify the city center area

7.3 Results

To characterize the type of individual places on each block, the writer used the user's text posted on Twitter. A total of 57 places were found from all the blocks. The author then divided the area into seven categories. The description of the groups and venue can be seen in Table 7.1

Table 7. 1 Location categories activities

Category	Place
Entertainment & Business	Cinema, Street, Café, Bar
Office	Electricity Company, TV and Radio, Government Building, Private Building, Telecommunication company, Bank
Food and Drink	Coffee, Pizza Restaurant, Culinary (Meatball, Fried Chicken, Noodle, Seafood, Chicken Porridge, Fried Banana, Tea House, Fried Rice, Donut, Steak, Snack, Sushi, Udon, Ice Cream, Donut, MacDonald, KFC
Shopping	Mall, Store, Fresh Market, Bookstore
Sports & Recreation	Soccer Field, Basketball Court, Beach, Stadium, Gym
Healthcare	Hospital, Clinic, Medical Center, Nursing Home
Education	University, Nursing School, Midwives School, Industrial High School, High School, Junior High School

7.3.1 Classifying places

After classifying the area, the author identified the landmark in each category. In this research, the landmark is used as an indicator to find out where people are concentrated and the reason why individuals gathered there. Then, the author identified the check-in activity to check an individual's movement and identify where they are. To facilitate the analysis, the author grouped the area into seven categories. Each category contains the user activity in the place visited. The categories description is as follows:

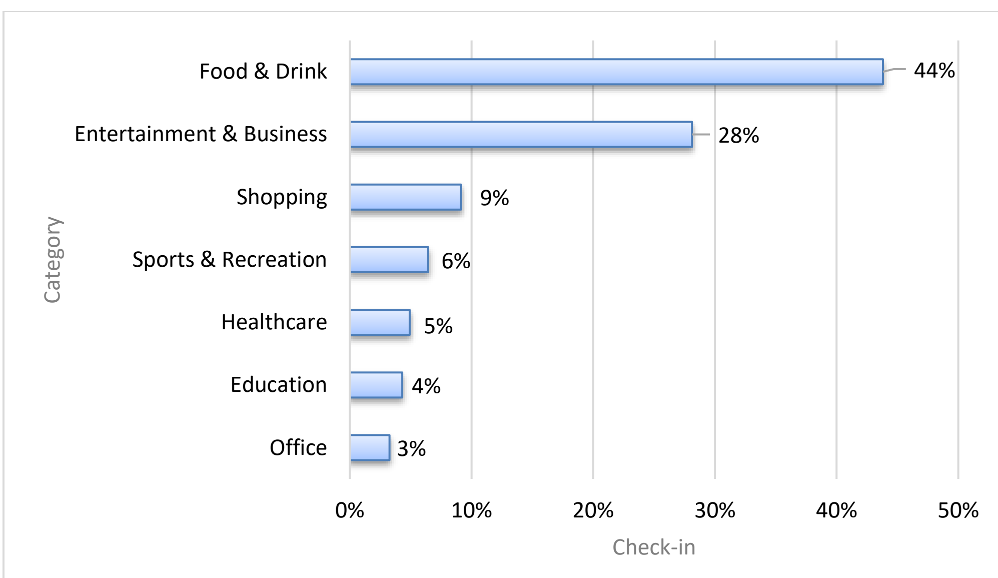


Figure 7.4 Category distribution off type 1

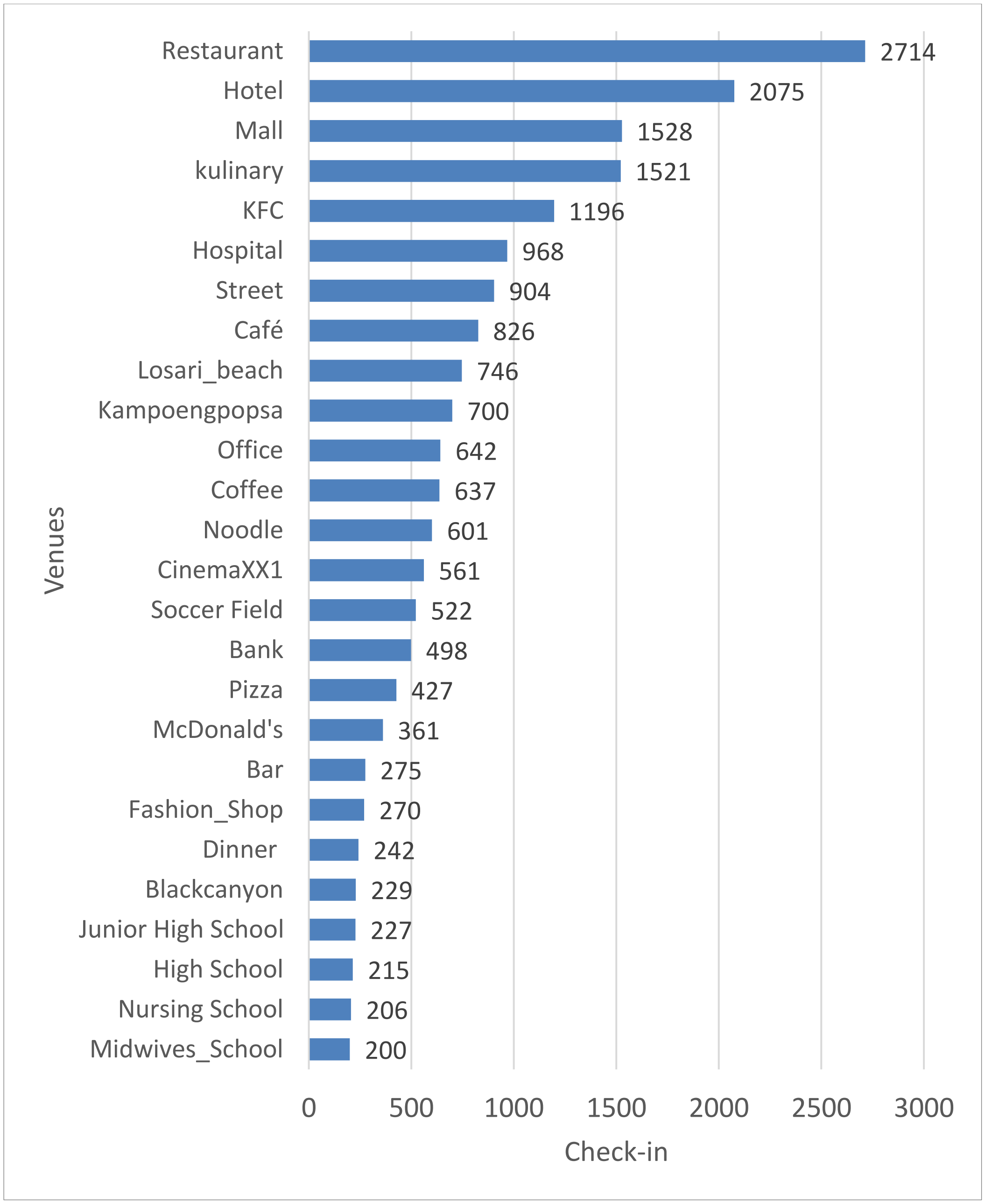


Figure 7.5 Places of activity for type 1

Type 1. Almost all the categories are in this type. The author observed that the areas covered by this category's activity were food and drink, entertainment and business, shopping, sport and recreation, healthcare, education, and office categories (see Figure 7.4). To recognize the places, the author used text that users posted on Twitter as a key to identifying the name of the place (see Figure 7.6). In this regard, the majority of this area was covered by food and drink categories (e.g., restaurant, KFC, coffee, noodle, and culinary places. Then, for the entertainment and business categories, users tended to focus on places such as hotels, banks, streets, cinemas, and cafés. Furthermore, the shopping category is dominated by mall and shop activities (see Figure 7.4).

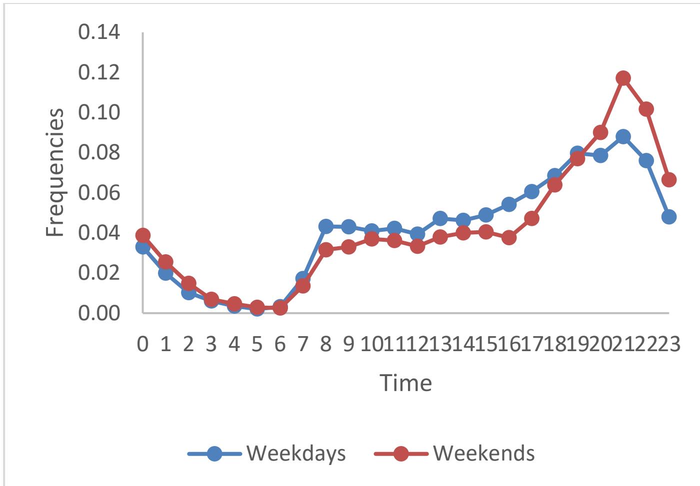


Figure 7.4 Time distribution for type 1

Figure 7.5 The word frequency for type 1

Type 2. As shown in Figure 7.8, the author analysed that this group was dominated by the shopping category, where almost 100% of people's activity was in the mall. The existence of venues such as cinema, coffee, Starbucks, Zafferano, and others, are individual activities conducted in the mall area (see Figure 7.9)

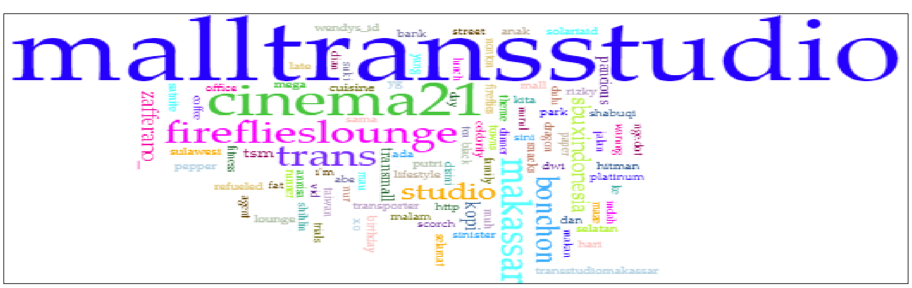


Figure 7.6 The word frequency analysis of type 2

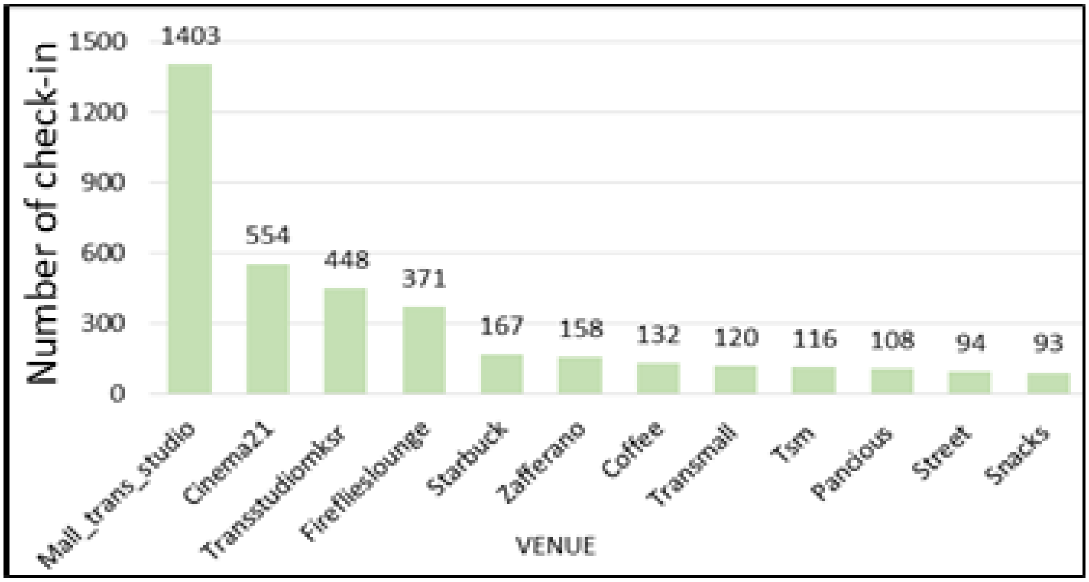


Figure 7. 7 Places of activity for type 2

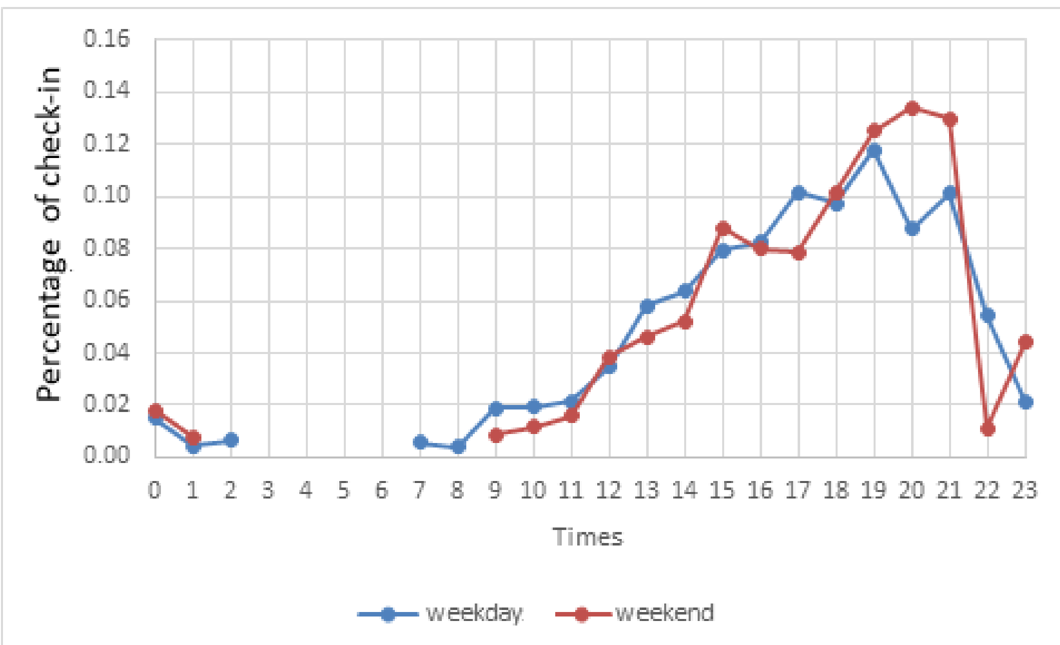
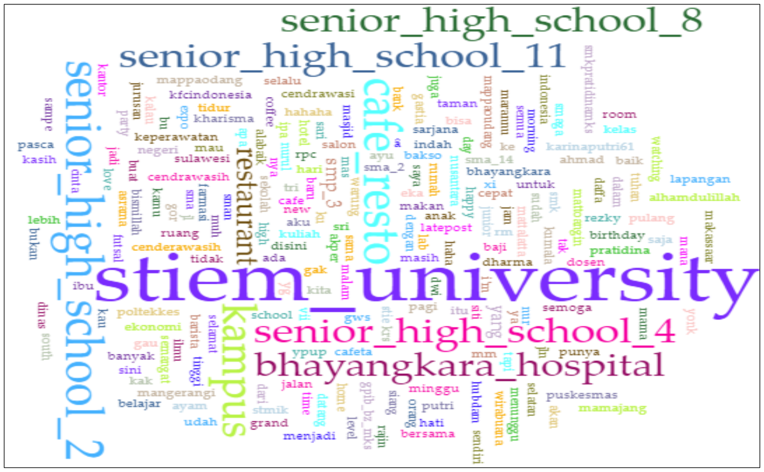
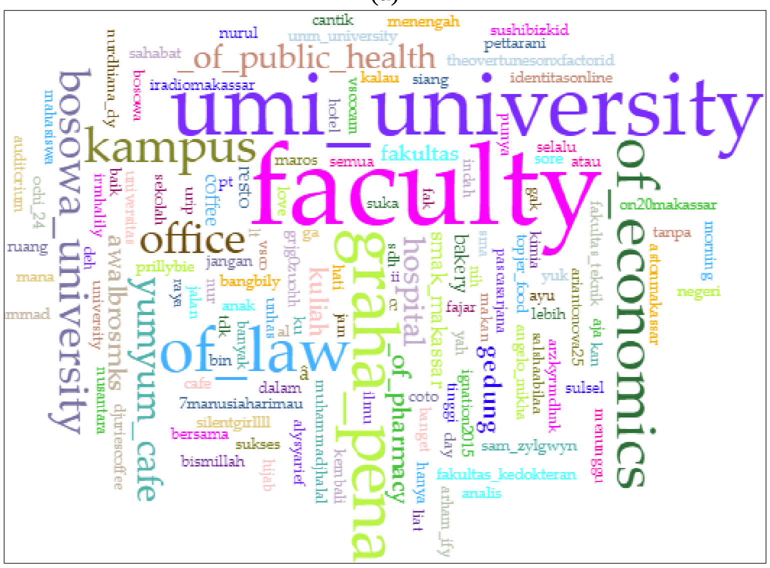


Figure 7.8 Time distribution for type 2

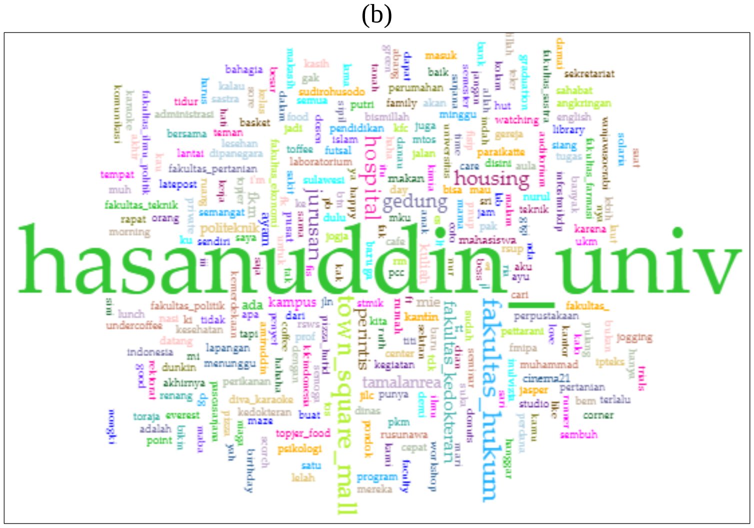
Type 3, Type 6, and Type 7. These types have similar activities. The author found that these types are dominated by activity in places such as the university and school (see Figure 7.9). In general, these types display the spatial distribution of users in the university area. There are mixture places such as hospitals and offices, but they are not significant. The majority of people tweeting were observed in the university faculties (see Figure 7.11).



(a)

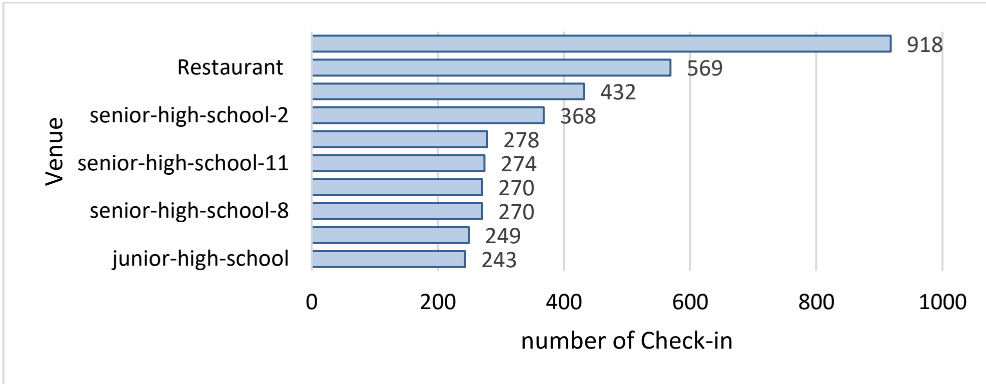


(b)

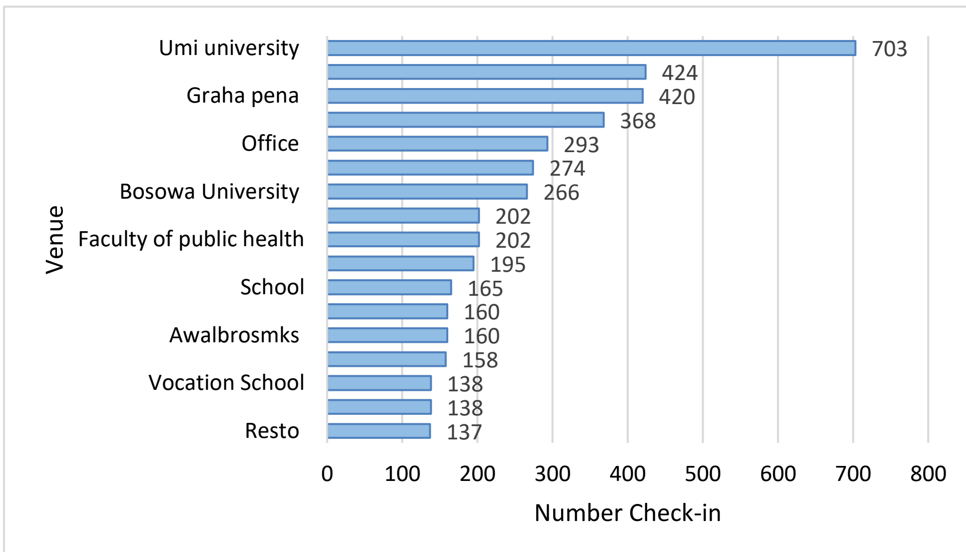


(c)

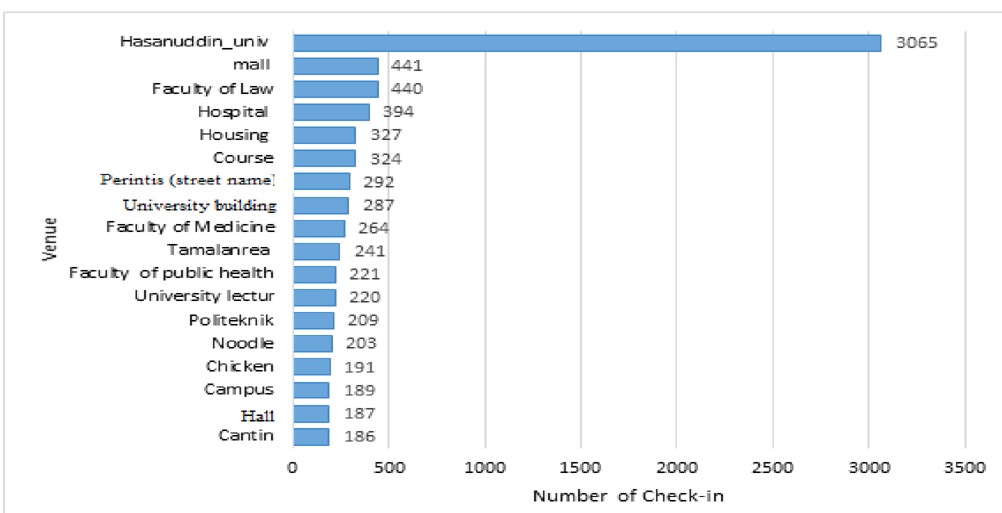
Figure 7.9 The word frequency for (a) type 3; (b) type 6; (c) type 7



(a)

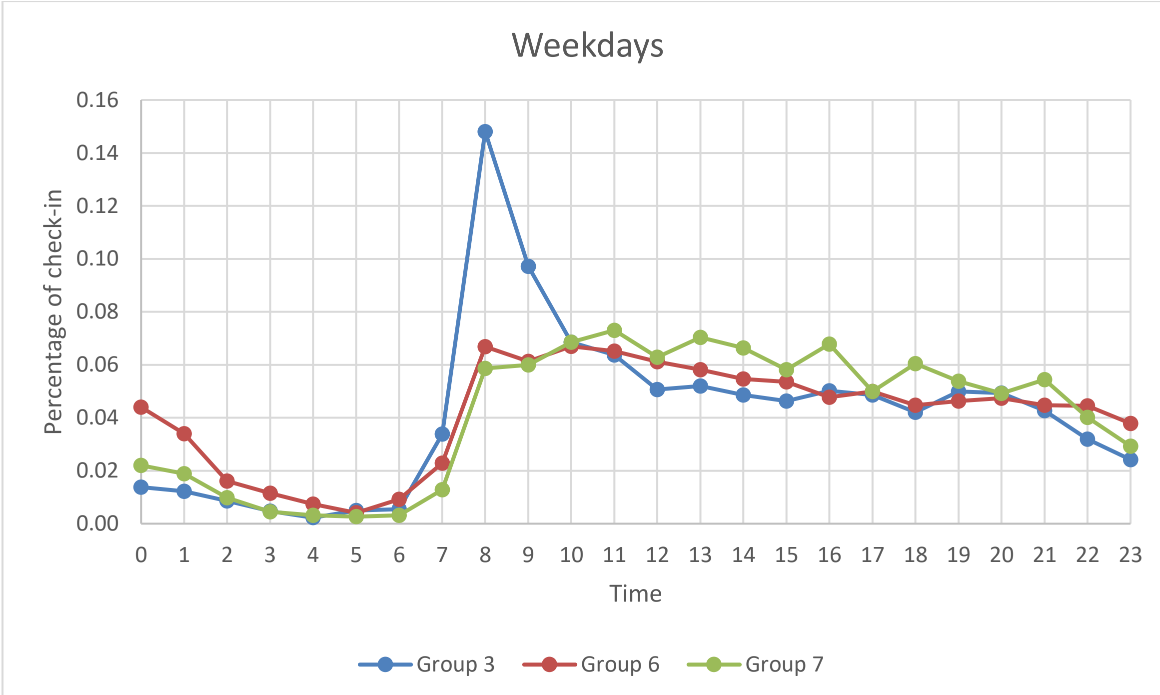


(b)

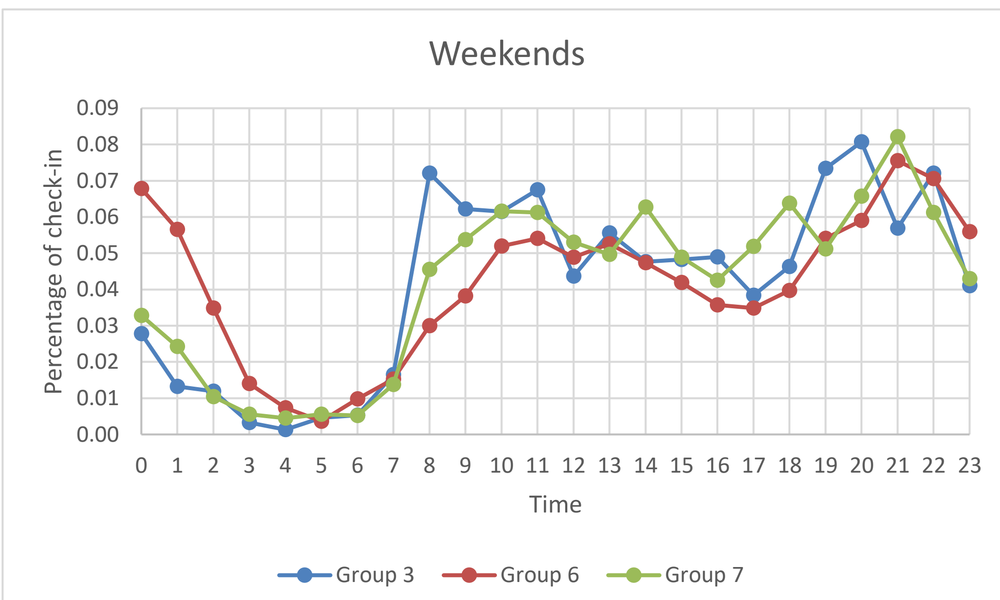


(c)

Figure 7.10 Places activities of (a) type 3; (b) type 6; (c) type 7



(a)



(b)

Figure 7.11 Time distribution for group 3, group 6 and group 7—(a) weekdays and (b) weekends

Type 4. Figure 7.14 shows the percentage for each category. The majority of tweet activity in this category is covered by education, entertainment and business, food and drink, and office areas. The author observed that the education category was dominated by places, e.g., the university and school. Then for the entertainment and business categories, the majority of people were present at hotels and streets.

The categories of food and drink as well as the office category are dominated by places such as McDonalds, KFC, meatball, coffee, and telecommunication offices. Figure 7.15 shows the word count generated from the area corresponding to group 4. Information about places visited can be seen in detail in Figure 7.16.

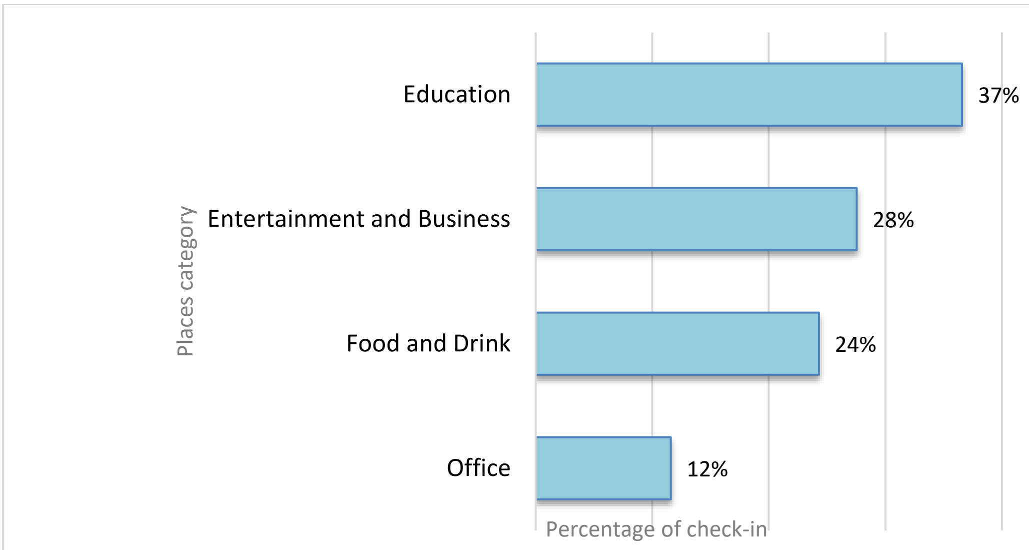


Figure 7.12 Place category comparison of type 4

Figure 7.13 The word frequency for type 4

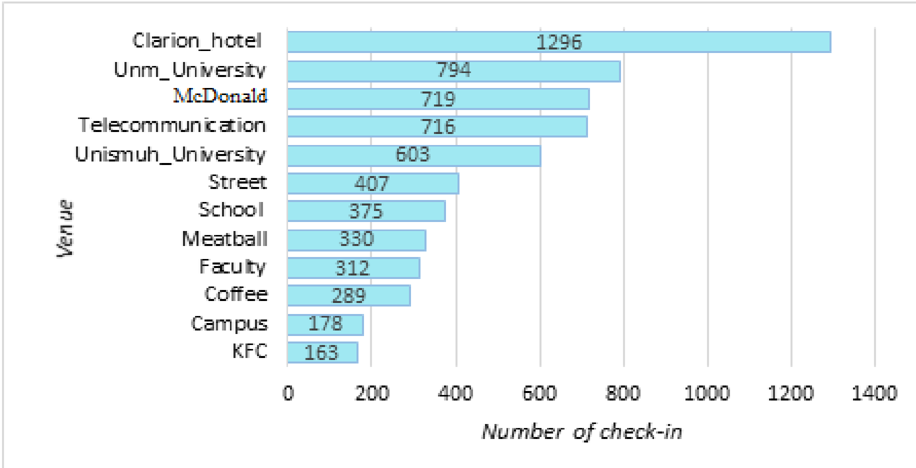


Figure 7.14 Places of activity for type 4

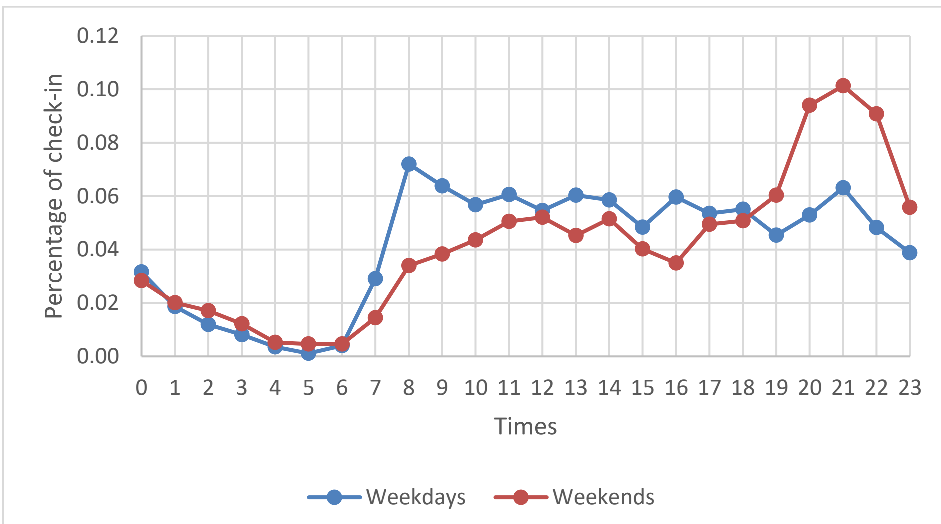


Figure 7.15 Time distribution for type 4

Type 5. In general the user activity in this type was tightly related to the individual's activity around the categories of food and drink, entertainment and business, shopping, and healthcare (Figure 7.18). The author found about 53% of user activity was sourced from the food and drink category. The kinds of places visited in this category were coffee, tea, McDonalds, KFC, pizza, donut, and other.

An analysis of the entertainment and business category showed that 26% of the area was covered by activity at the cinema, hotel, bar, and café. Meanwhile, the other categories such as shopping and healthcare covered about 19% and 2% of places like the mall and hospital (see Figure 7.19). To prove the existence of the areas, an estimation of user text posted on Twitter was displayed (Figure 7.20). The text message on this figure talked about the activity in places in this type.

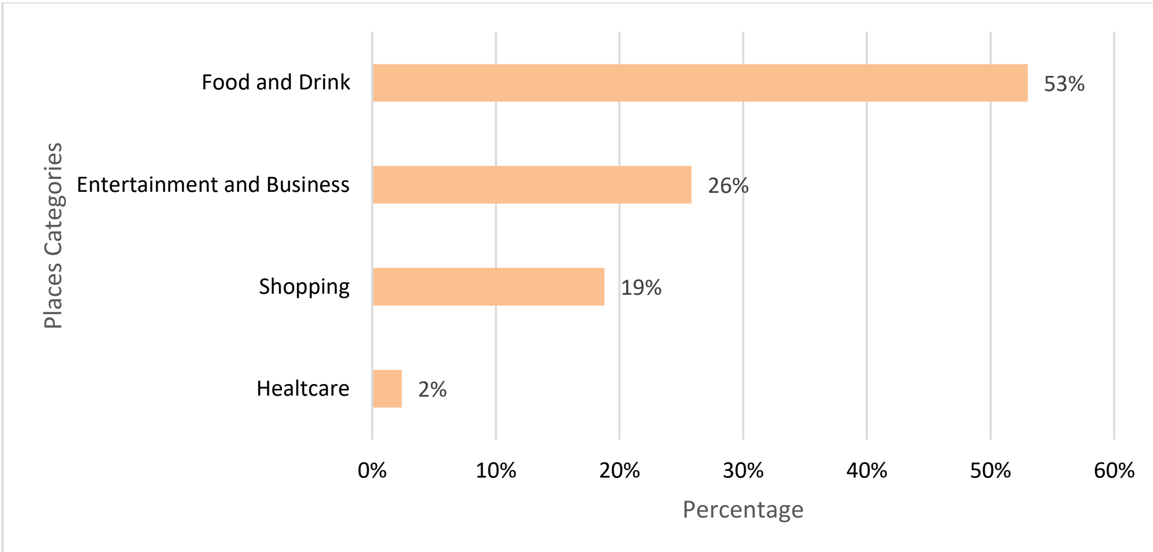


Figure 7.16 Place category comparison of type 5

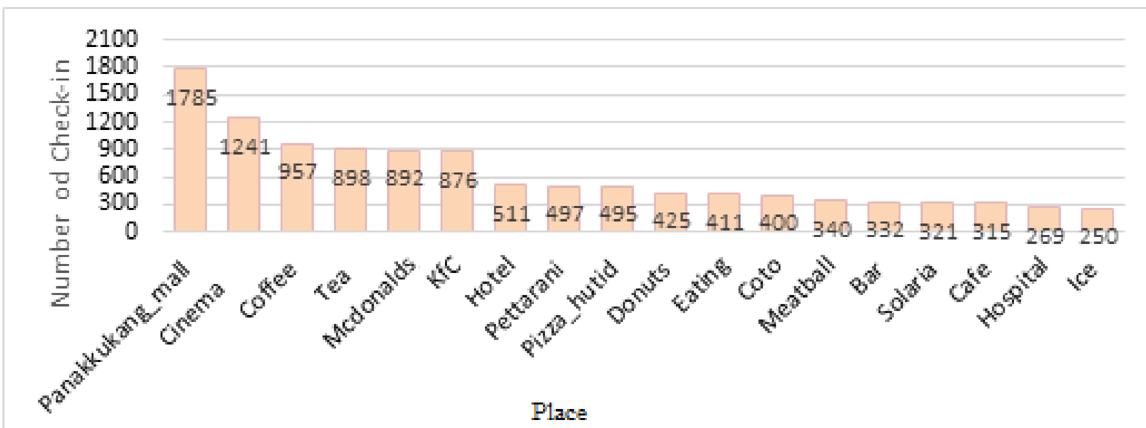


Figure 7.17 Places of activity for type 5

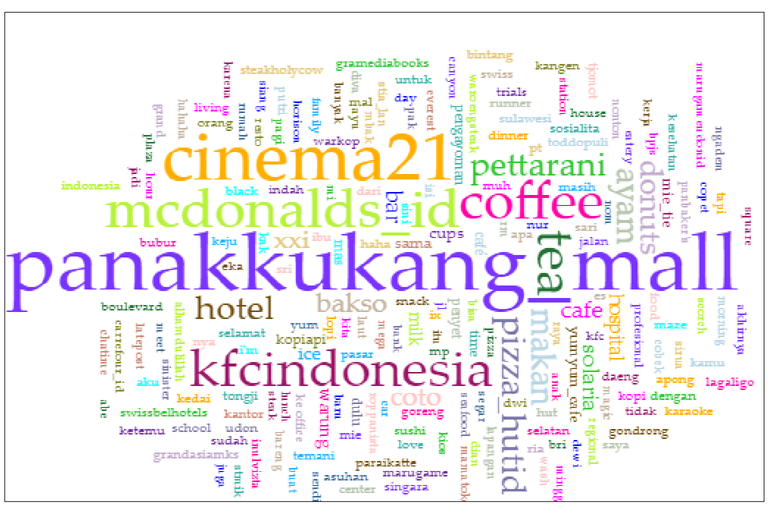


Figure 7.18 The word frequency for type 5

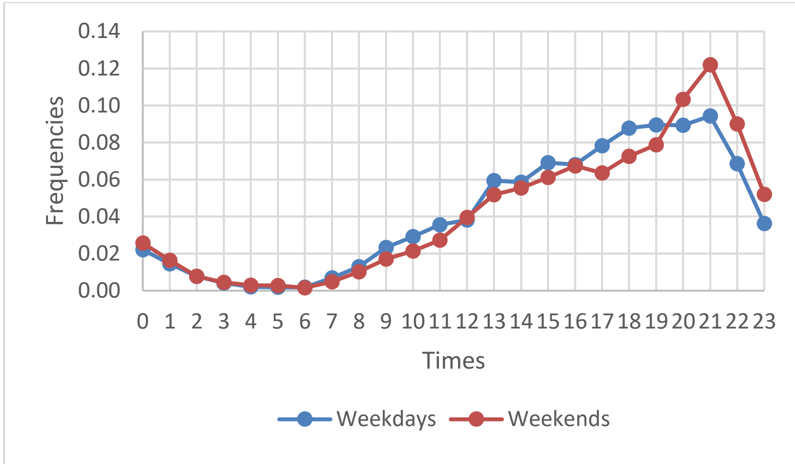


Figure 7.19 Time distribution for type 5

7.4 Section Conclusion

In this chapter, the author used Twitter social media data as a source of information to analyse the city center. The data feature which was examined from Twitter was in the form of users' status text update and check-in activity. To build this research, the writer proposed a grid based on an aggregation method to divide the urban area into a grid. To facilitate the detection of the downtown, the author grouped the urban area based on the busiest part. From the classifying results, the author categorized places and identified the landmarks in each category. In this study, landmarks were used to detect the kinds of places where people gather.

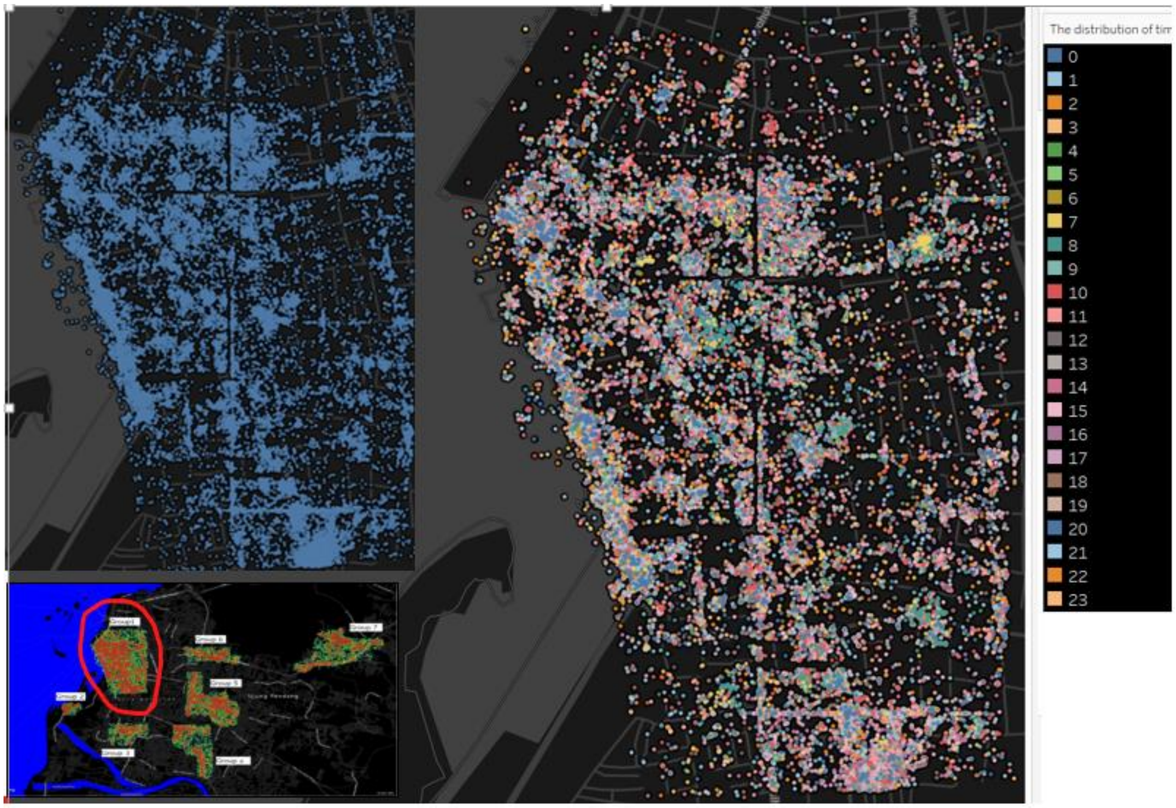
From the results of the analysis, the author concluded that type 3, type 6 and type 7 do not reflect the downtown area due to the majority of check-ins from this area stemming from the education category in which the users' activity occurs in places like the university and school. Then. The author analysed type 2 and found that almost all the users' activity in this group took place at the mall. This indicates that there is only one category in this place. On account of this, the area is less likely to be the city core.

Then, the author analysed type 1, type 4, and type 5 (figure 7.22) concluded that these areas have the possibility of being the city center. In these types there are categories which support it being the downtown, such as the categories of entertainment and business, office, food and drink, shopping, sport and recreation, healthcare, and education. The author concluded that human travel flow generated from Twitter data in the city of Makassar is likely to be polycentric in that an individual's distribution in the town of Makassar not only focuses on one region but is spread to several areas.

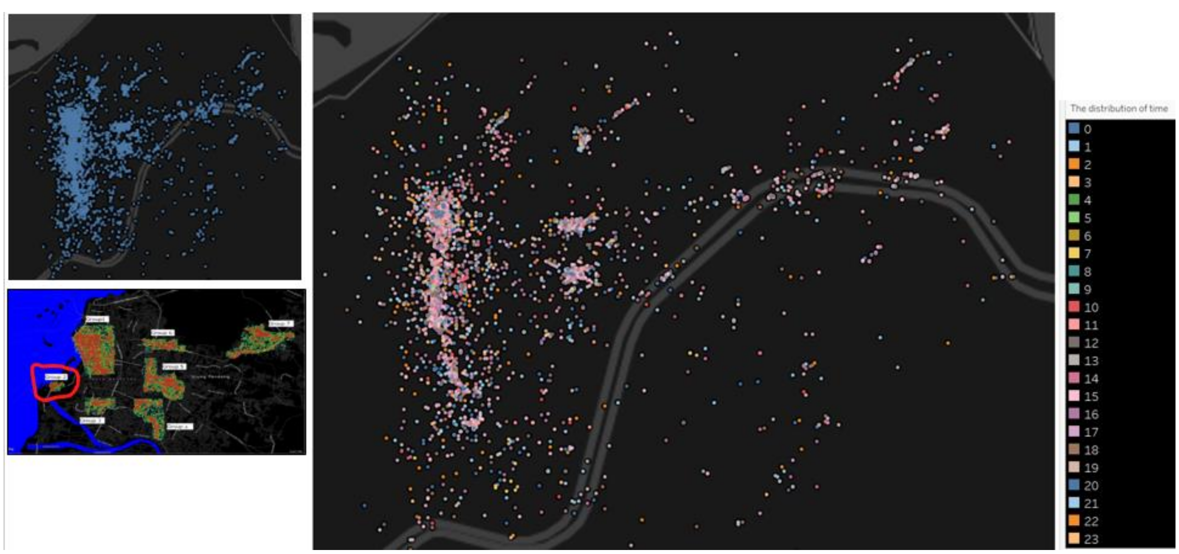
Figure 7. 20 City center hypotheses

Appendix

Check-in and times ditribution map of type 1

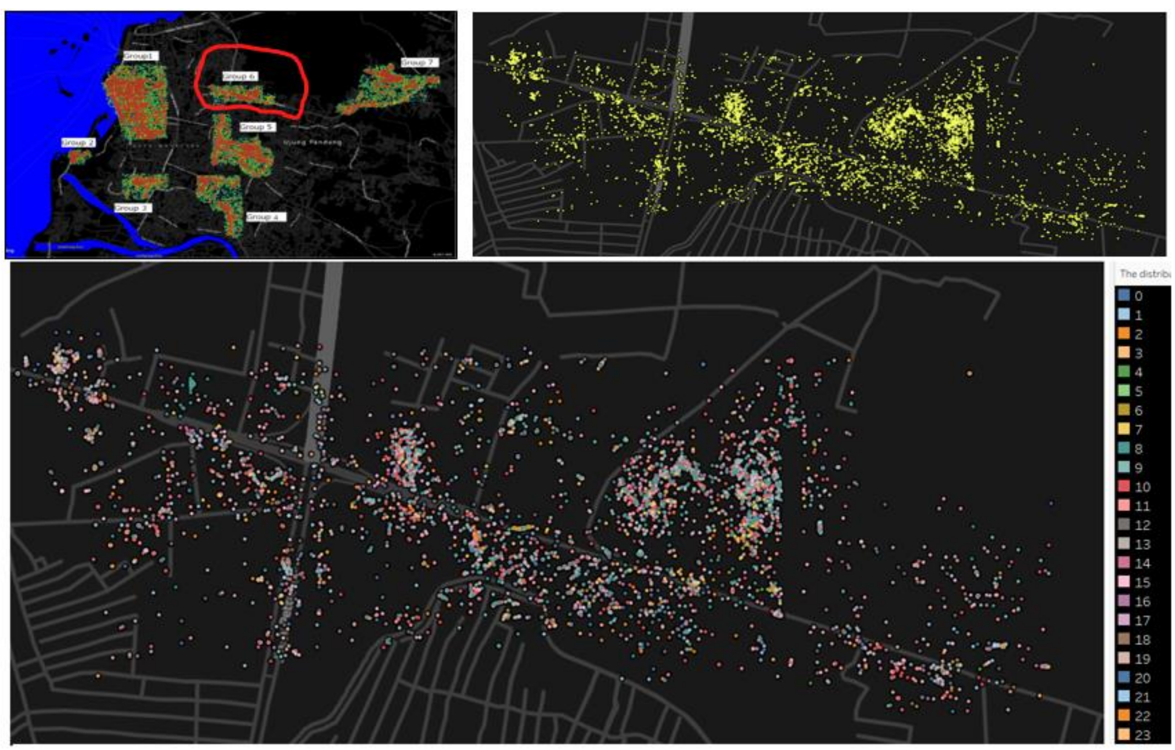


Check-in and times ditribution map of type 2

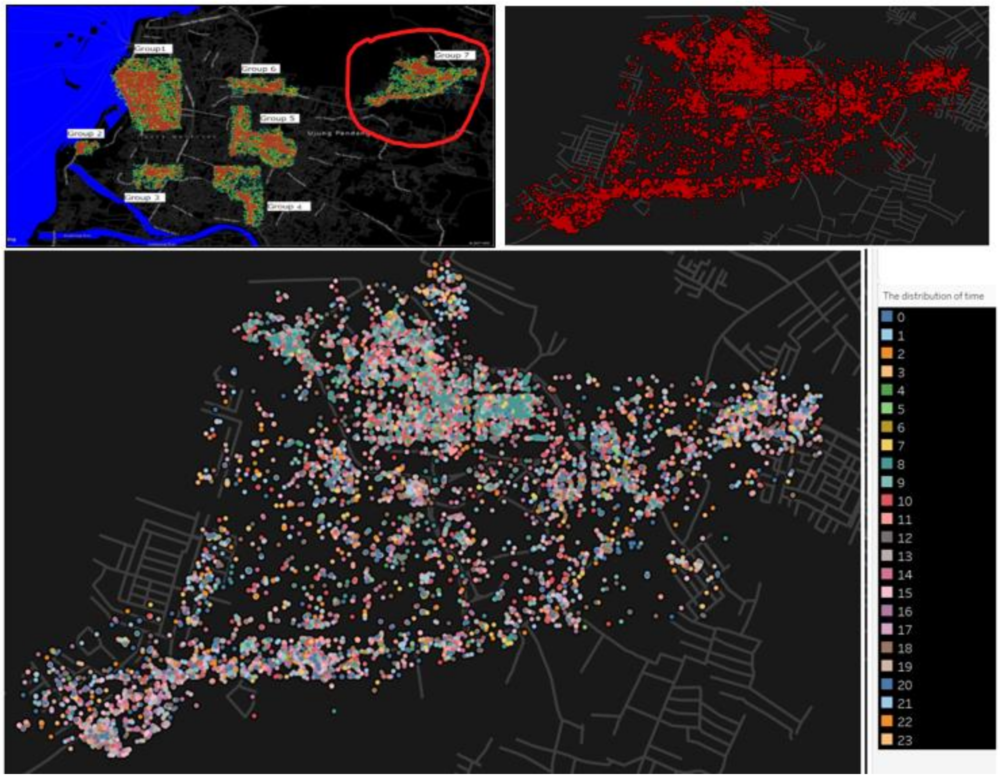


Check-in and times ditribution map of type 3

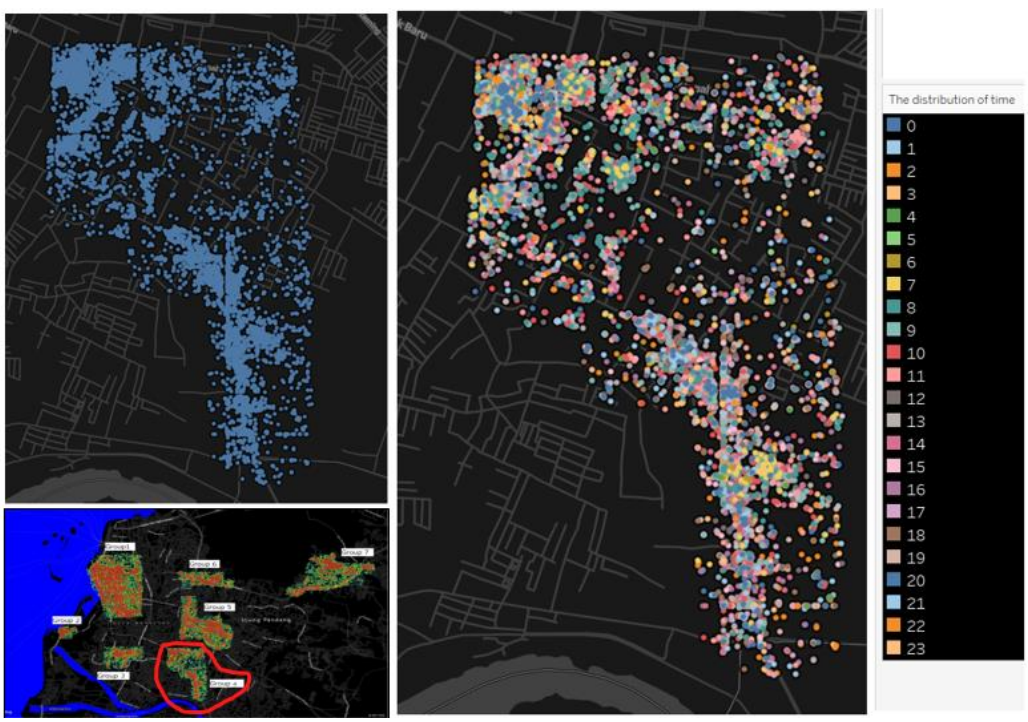
Check-in and times ditribution map of type 6



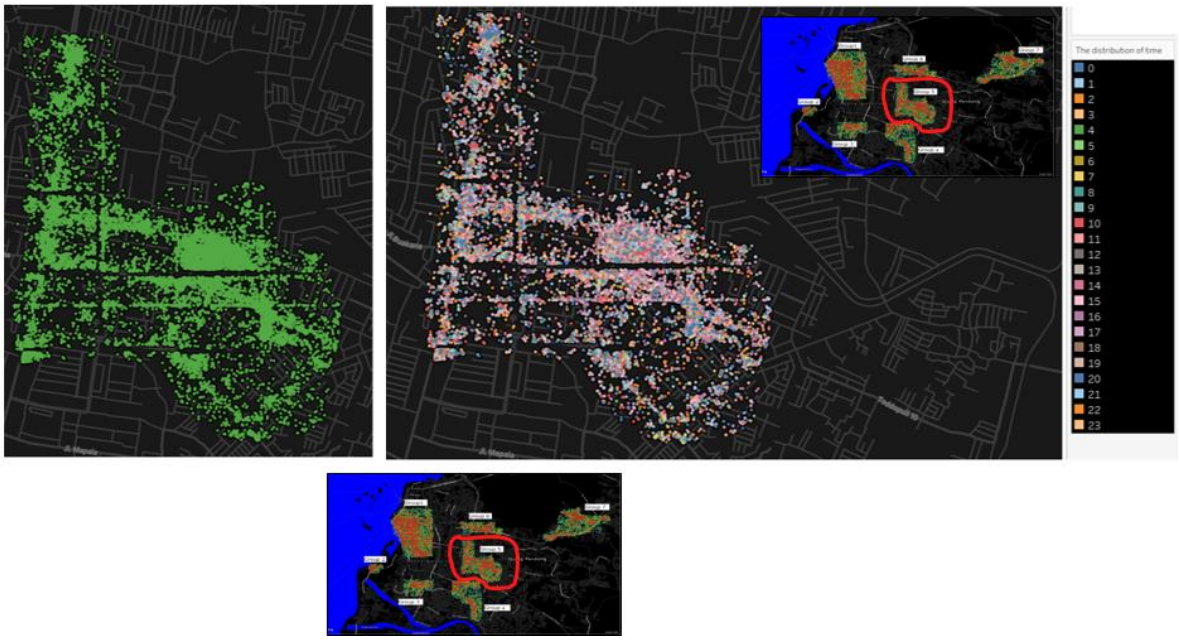
Check-in and times ditribution map of type 7



Check-in and times ditribution map of type 4



Check-in and times ditribution map of type 5



References

1. Cortright, J. Surging City Center Job Growth. **2015**.
2. Kotkin, B. J. The Future of the Center : The Core City. 264.

Chapter 8

User Social Media Interaction towards Urban Public Space and Public Facility

8. 1 Visualization of Urban Public Space and Publict Facility

8.1.1 Purpose of the Chapter

Public space is a space in which people gather to engage in activities for a specific purpose, to relax, or to perform certain activities [1]. In urban planning, public space is defined as "open space", meaning the streets, parks and recreation areas, plazas and other publicly owned and managed outdoor spaces, as opposed to the private domains of housing and work [2]. Nowadays, the existence of public space has become an important part of the daily life of urban society, for example, as a place to play, a place where city communities get together and interact, a place for sports activities, or a place for recreation. Given that public space is a part of the constituent elements, then public space can provide its own character to a city. Thus, the existence of public space becomes a necessity that must be met in the formation or development of a city.

The purpose of this chapter is to discover the activity of social media users in the public spaces and public facilities in Makassar City by identifying the type of place that become a priority for people to visit. In this study, the research question is how are the open spaces and public facilities in the city of Makassar used from the perspective of Twitter users?

8.2. Method of Data Analysis

In Twitter social media data, there are four important features used to explore individual activities in a public space. The data consists of locations marked with the geographical location in the form of latitude and longitude coordinates. Individual posted text is used to identify the type of place visited. Time is a feature used to explain the time at which the post is created and the date or day is to show when the post is made.

The writer identified six categories of public spaces and public facilities that were prioritized for people conducting check-in activities. In each category, the author identified the type of place that corresponded to each category. Then, in each location, the author identified the positioning of users based on daily, weekdays and weekends activities. It is essential to know the number of users performing check-in activities in the places visited. Figure 8.1 shows the procedure for identifying public spaces and public facilities.

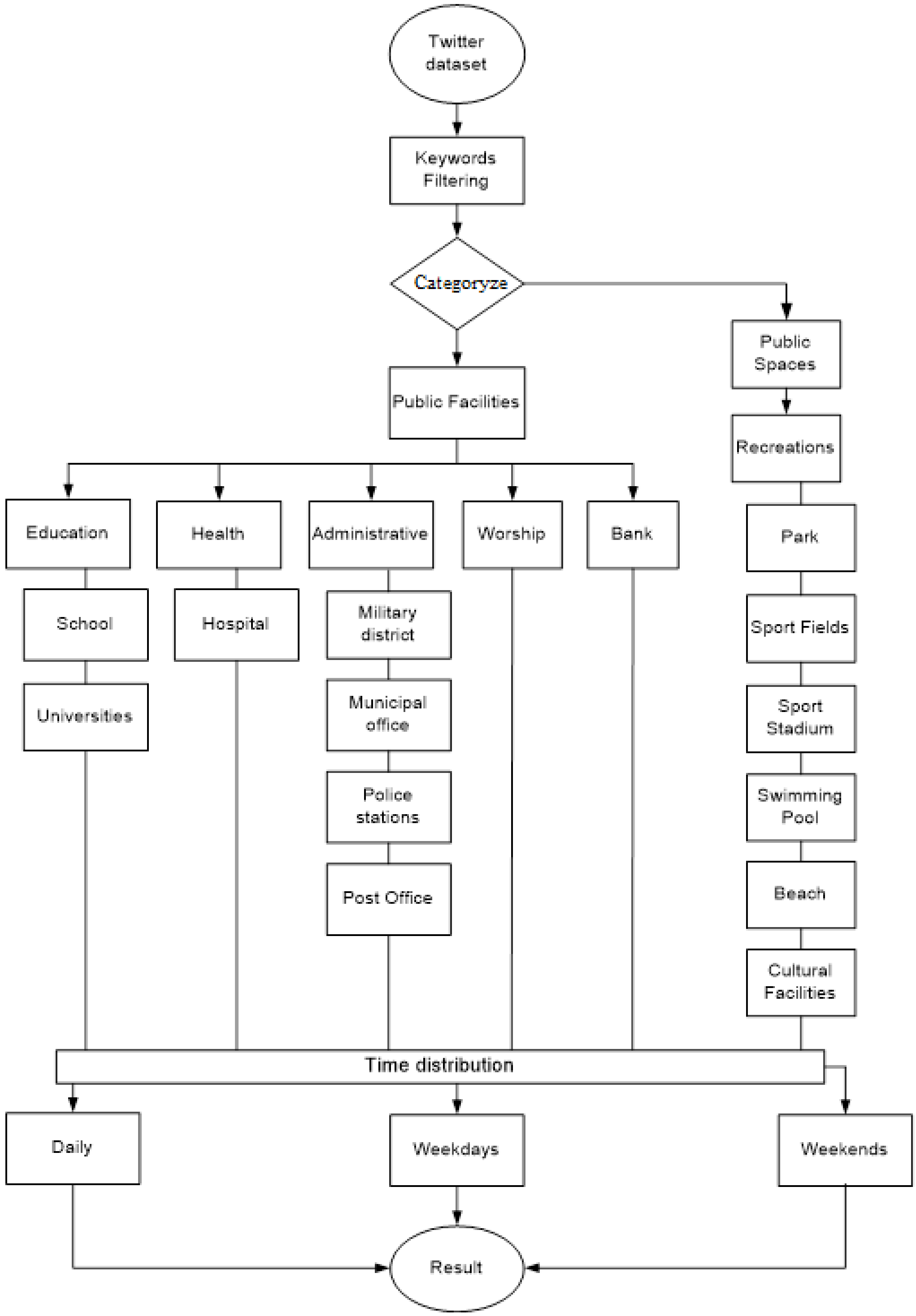


Figure 8.1 Procedure to identify public spaces and public facilities

With hundreds of thousands of data in the author's database, it is necessary to identify the type of criteria for classifying the place according to category. In defining the rules, the author introduced the type of keywords used by social media users when conducting post activities. The list of keywords used by the author to identify the kinds of places is shown in Table 8.1.

Table 8.1 The list of language keywords to recognize the kinds of places

Categories	Places	Indonesian	English
Education	School	SMA	Senior high school
		SMP	Junior high school
		SMK	Vocational high school
		Sekolah	School
	University	Universitas	University
		Fakultas	Faculty
		Jurusan	Department
		Sekolah tinggi	High school
		Akademi	Academy
		Laboratorium	Laboratory
		Perpustakaan	Library
		kantin	Kantin
Bank	Types of banks	Bank Mandiri	Mandiri bank
		Bank BNI	BNI bank
		Bank BCA	BCA bank
		Bank BRI	BRI bank
Administrative	Military district	Kodim	District military command
		Kodam	Regional military command
	Police station	SPN	Police academy school
		Polsek	Sectoral police
		Polres	Regency police
		Polda	Regional Police
	Office	Kantor / Dinas	Office
Worship	Type of worship	Mesjid	Mosque
		Gereja	Church
		Pura	Temple
		Etcetera	
Recreation	Type of recreation	Taman	Park
		Lapangan	Fields
		Gym	Gym
		Stadion	Stadium
		Pantai	Beach
Health	Type of health	Rumah sakit	Hospitals
		Klinik	Clinics
			Etcetera

Below are the five steps used to select and determine individuals' activities in public spaces and and public facilities:

a. Filter the user's text posted according to keywords (see Table 8.1)

b. Group the places based on the activity category (see Table 8.2)

c. Count the number of check-ins in each place to get information about how often the individual visits

d. Analyze user deployment by investigating the trends of each check-in daily and for weekday and weekend periods

e. Summarize the check-in activities in each category

Table 8.2 Place categories of public spaces and public facilities

Category's activity	Places
Administrative	Military district, municipal office, police station, and post office
Bank	Types of banks
Education	School, University
Health	Hospital
Recreation	Park, sport fields, sports stadium, swimming pool, beach, cultural facilities
Worship	Types of worship

8.3 Dataset Description

The dataset used in this chapter was sourced from Twitter's social media. This chapter discusses the individual check-in activities in the city space more specifically—in this case, in public spaces and public facilities. After completing a series of analyses, the author identified 18,612 check-ins and 7156 users scattered throughout the urban area (see Table 8.3) and produced six groups as shown in Table 8.2.

Table 8.3 Makassar City Dataset detail

Original dataset	Number of check-in
Users number	25.346
Check-ins Number	211.922
Research sample	
Users number	7.156
Check-ins Number	18.612

8.4 Individual Movement around the City

The writer now presents the approach to identify how often people visit a place in an urban space. In this section, the author aims to understand the extent of the utilization of places designed for public purposes by, for example, counting check-in activities in open spaces. Then the author analyzes the percentage for each user to ascertain the number of check-ins at each place. Figure 8.2 shows a distribution map of people into six categories. It is seen that check-in activity is dominated by individual activities in the education and recreation categories. To facilitate the analysis, in the next step, the discussion about the place will be divided into several sections.

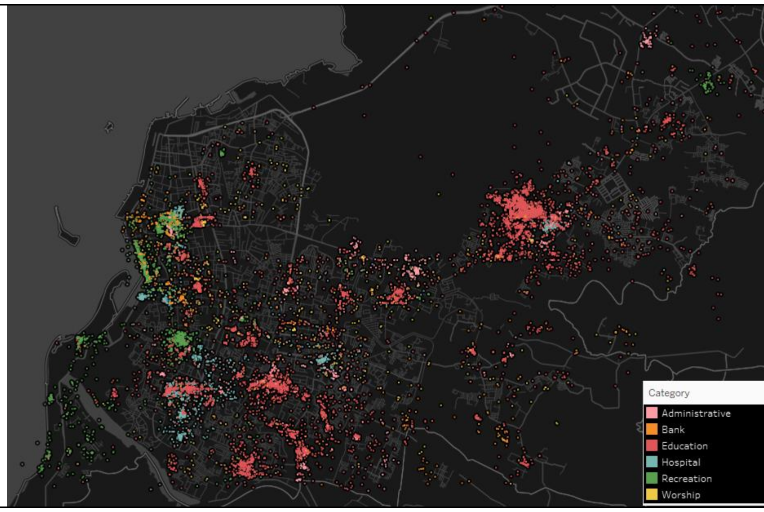


Figure 8.2 Spatial distribution of Twitter venues of public spaces and public facilities

8.4.1 Public Spaces

In this chapter, public space is defined as an area or place outside a building or an open space. This space allows for humans to interact because this space can be accessed by the public as an open space. The author classifies public space as a recreation category. As mentioned in Figure 8.1, six areas such as parks, sports fields, sports stadiums, swimming pools, beaches and cultural heritage were generated. The distribution map of the recreation area can be seen in Figure 8.3.

Figure 8.3 Physical layout of check-in places in public spaces

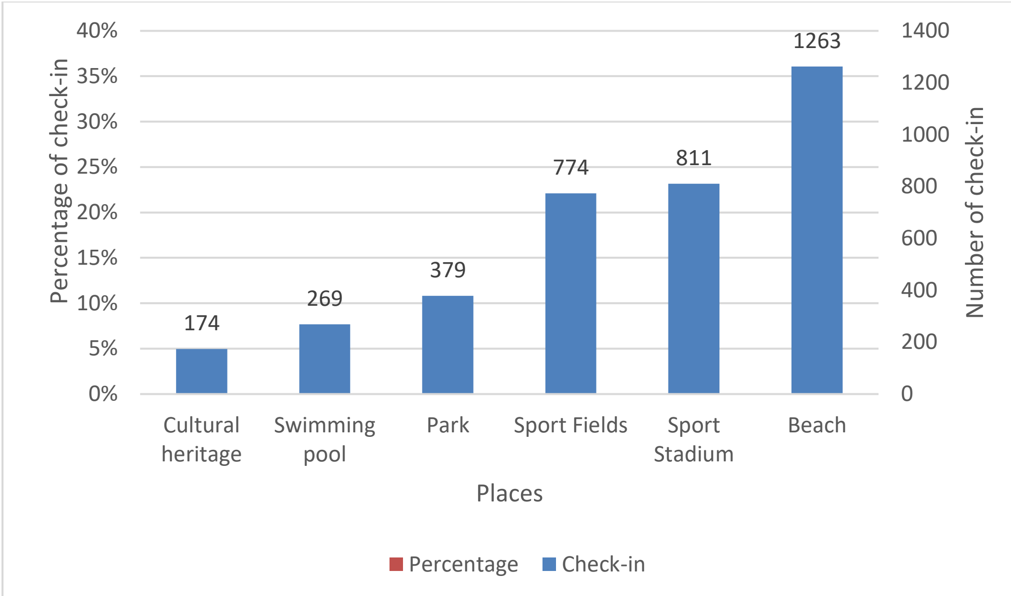


Figure 8.4 Comparison of check-in activities in public spaces

Figure 8.4 shows the check-in activities in public spaces specifically for the recreation category. From the analysis, it was found that the beach was the most visited place. This place covers about 34% of individual activities in the beach. Other exciting places are sports stadiums (22%), sports fields (21%), parks (10%), swimming pools (7%), and cultural heritage (about 5%). The spread of check-in activities for the recreation category is described below.

a. Parks

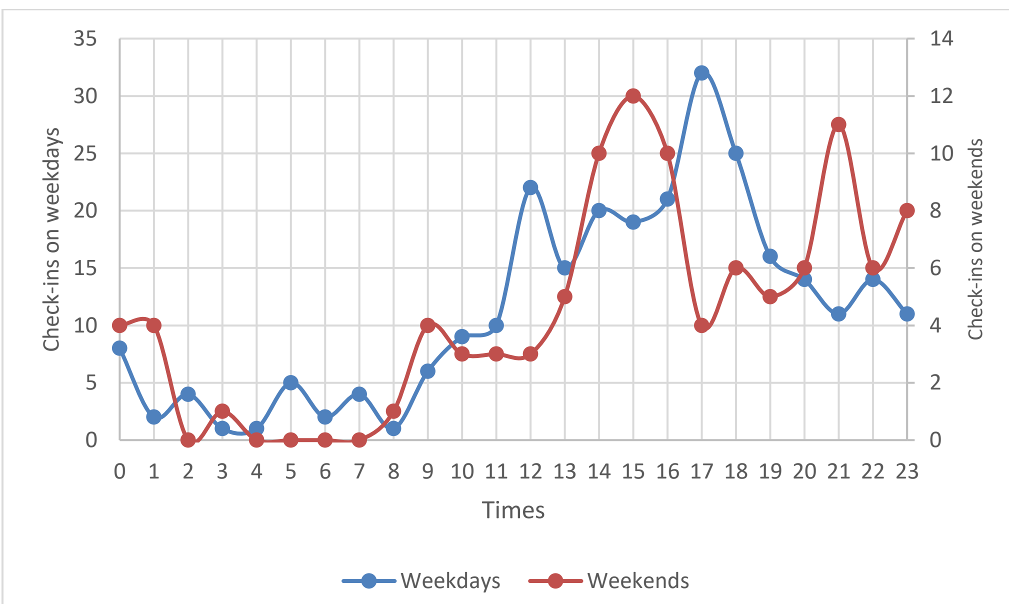


Figure 8.5 Differences in individual check-in activities on weekdays and weekends in the parks

In this place, there are 379 check-in activities. From the total visitors to public spaces spread throughout parks in the city of Makassar, 10% was covered by this area. Figure 8.5 shows the difference in check-in activities based on the distribution over weekdays and weekends. On weekdays, the author divided the time into two parts: a daily period during which the peak of tweet activity occurred at 12 p.m. and 5 p.m.; and the weekdays during which the peak of individual check-ins was at 5 p.m., and on weekends at 3 p.m. It indicates that there was a significant difference between weekdays and weekends in people's routines in the parks.

b. Sports fields

The sports field is defined as an open space for sports activities. In this section, the author found that Karebosi field was the most visited place. This place is interesting because it is one of the landmarks and icons of Makassar City. Please refer to Figure 8.6 to see the place word frequencies. Then the author analyzed the daily tweet patterns and found that the highest number of check-ins occurred at 9 p.m. (see Figure 8.7). Furthermore, the author compared the tweet activity on weekdays and weekends and found that on weekdays, the peak of individual acts occurred at 6–9 p.m., and on weekends at 9 p.m.

Figure 8.6 The word frequency for sports fields

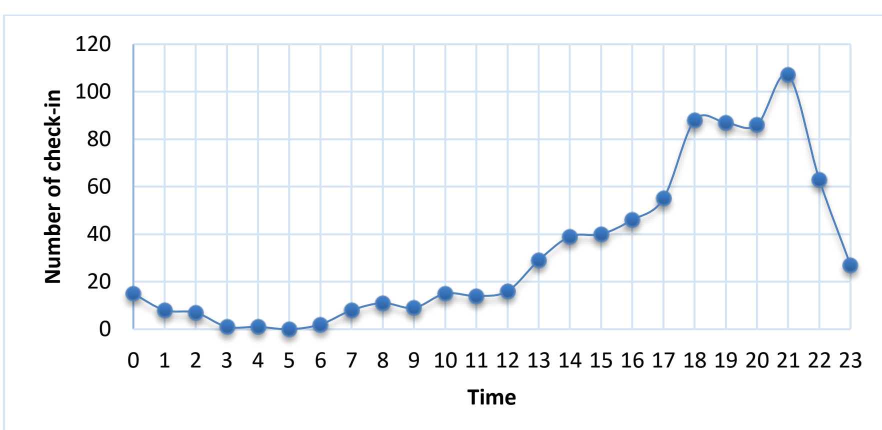


Figure 8.7 Time frequency for sports fields: daily

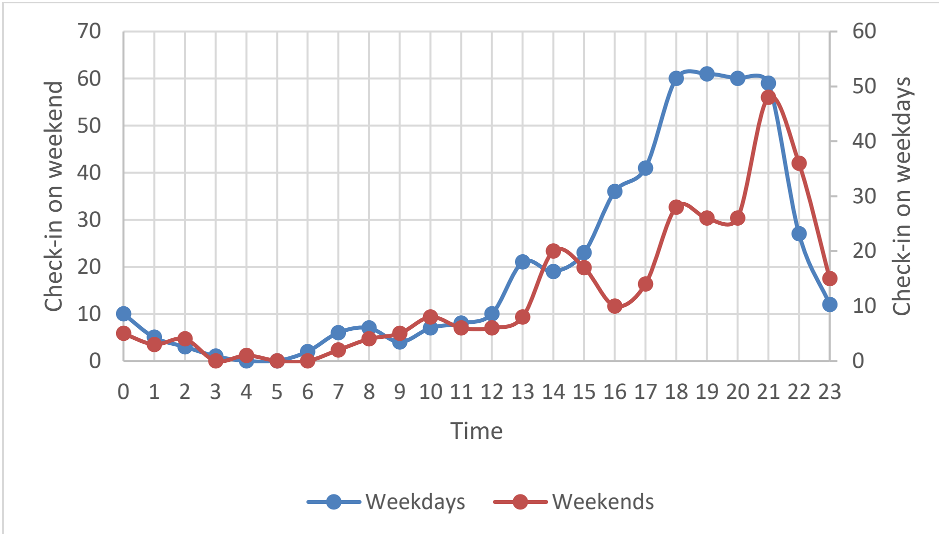


Figure 8.8 Differences in individual check-in activities on weekdays and weekends at the sports fields

c. Sports stadiums

The sports stadium is defined as a building used for sporting events. In this place, there were 811 check-in activities in which 71% of users mention specifically the name of the stadium in their status update post, 27% mention swimming pool, and 2% mention basket court (see Figure 8.9). The daily distribution check-in peak occurred at 5 p.m. and 7 p.m. (Figure 8.10). Figure 8.11 shows the spread of Twitter users on weekdays and weekends. During work days, most check-in activities occurred at 7 p.m., whereas on weekends it was at 4 p.m.

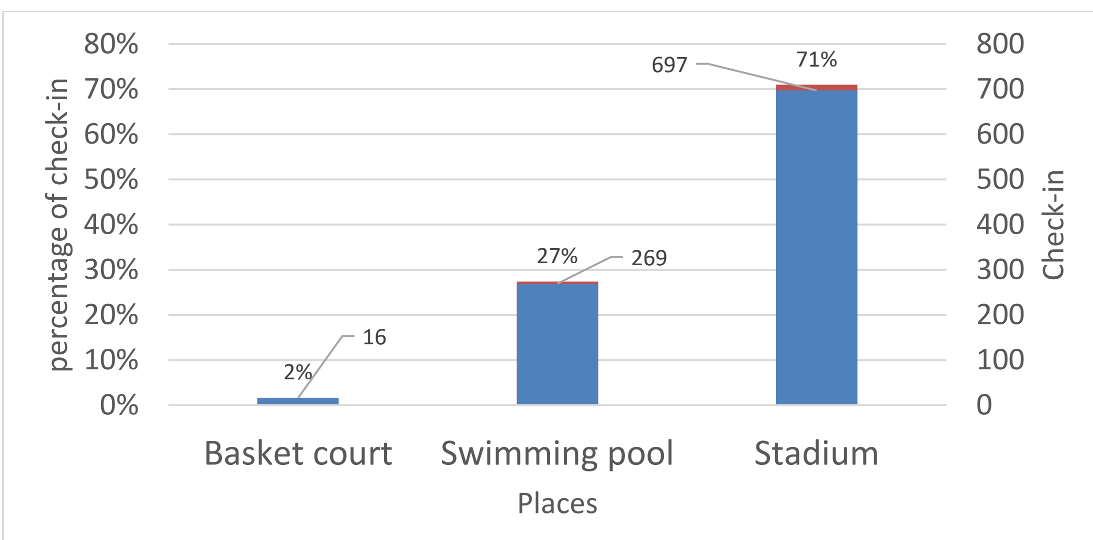


Figure 8.9 Check-in activities at the sports stadium facilities

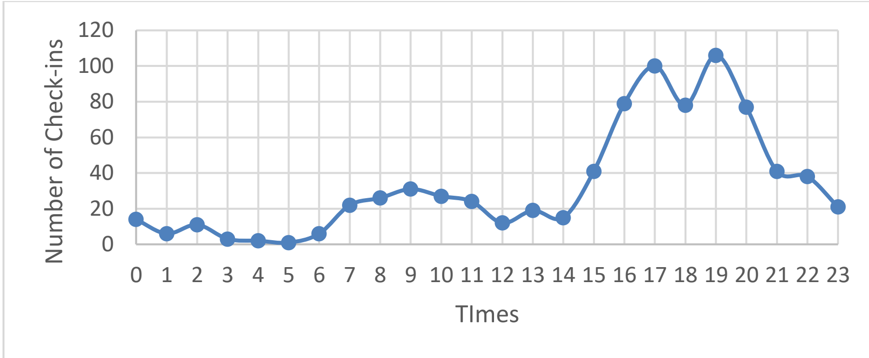


Figure 8.10 Time frequency for stadium fields: daily

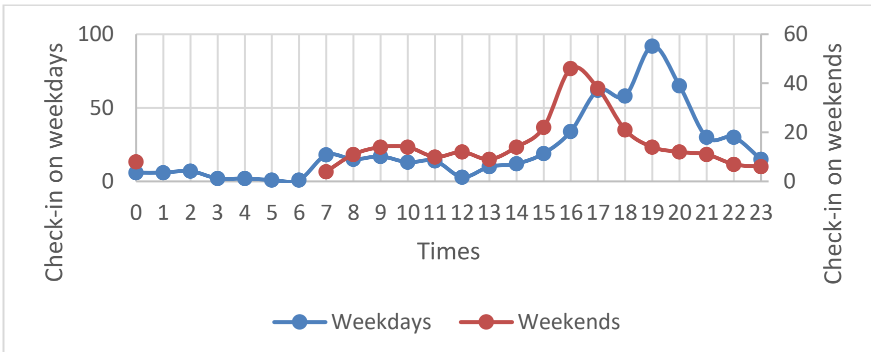


Figure 8.11 Differences in individual check-in activities on weekdays and weekends at the sports stadiums

d. Swimming pools

In this place, there were 269 tweets. On a daily basis, the check-in peak occurred at 5 p.m. (see Figure 8.12). When the author compared the individual actions on weekdays and weekends, there was no significant difference as the peak check-in on working days and weekends was 5 p.m. (Figure 8.13).

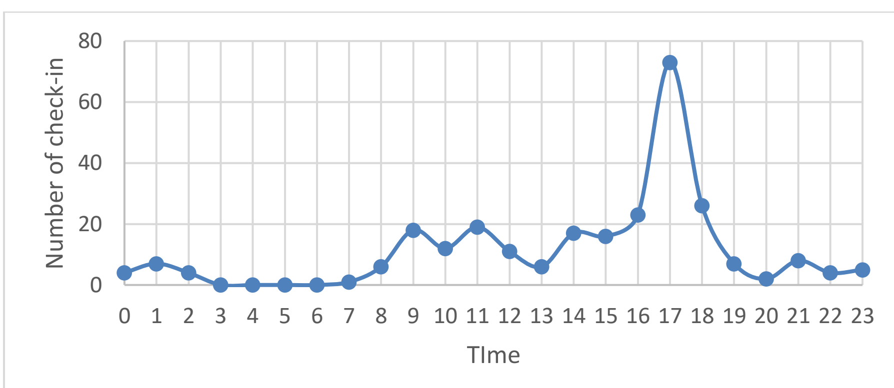


Figure 8.12 Time frequency for swimming pool: daily

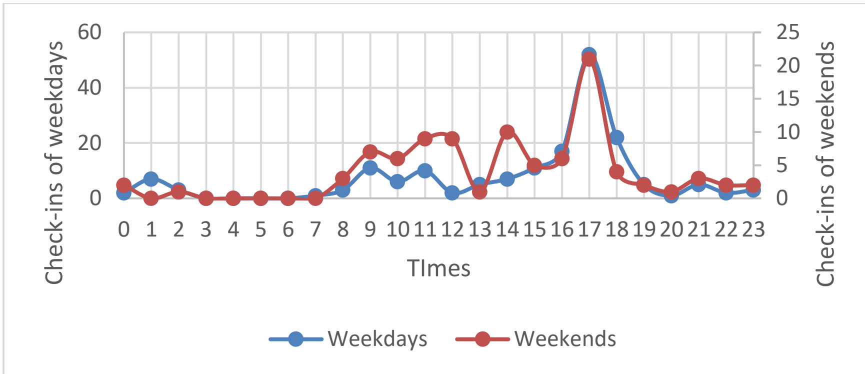


Figure 8.13 Differences in individual check-in activities on weekdays and weekends at the swimming pool

e. Beach

In addition to Karebosi field, Losari beach is also an icon of Makassar City. Here there were 1262 check-in activities: 663 check-ins occurred at Losari beach, 211 at Akarena beach, and 77 at Tanjung Bayang beach (see Figure 8.14). At the same time, the writer observed the daily check-in period and found that peak tweet activity occurred at 7 p.m. (Figure 8.15). The peak time for tweets on weekdays occurred at 7 p.m. and on weekends at 9 p.m. (Figure 8.16).

Figure 8.14 The word frequency for beach

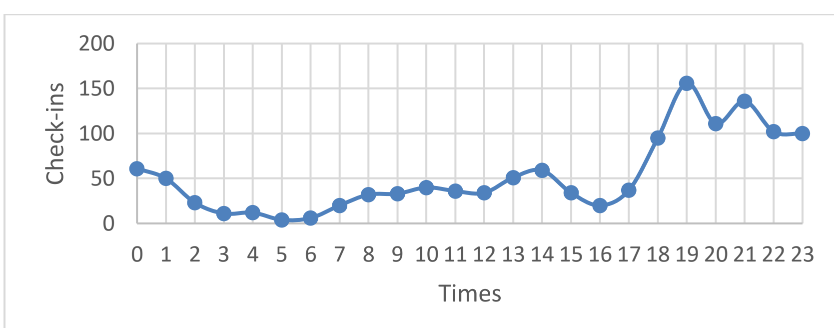


Figure 8.15 Time distribution for beach: daily

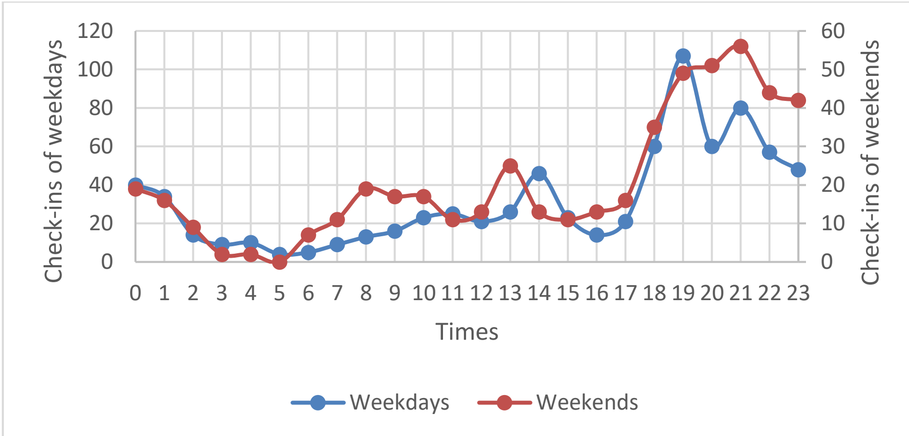


Figure 8.16 Differences in individual check-in activities on weekdays and weekends at the beach

f. Cultural heritage

The place with the lowest check-in activity was cultural heritage with 174 check-in activities. In the daily time distribution, the peak for check-ins took place at 12 a.m. This place is different from other areas where the majority of check-ins occurred in the afternoon.

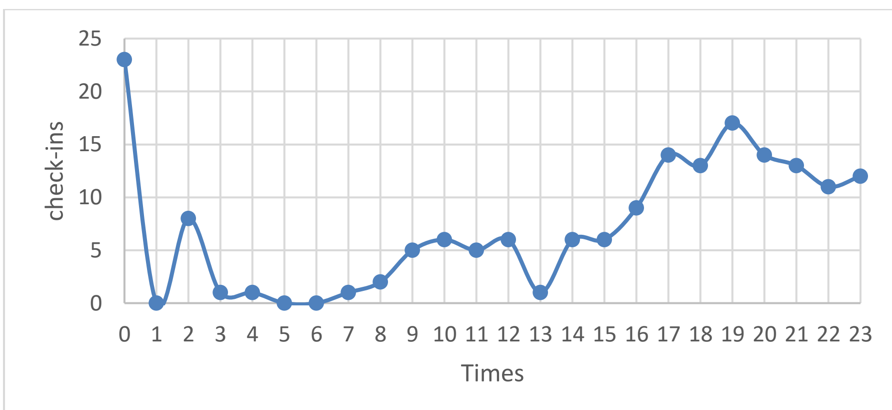


Figure 8.17 Time distribution for cultural heritage: daily

8.4.2 Public Facilities

To make it easier to identify public facilities, it is necessary to define what items are included. In this case, the author defines public facilities as facilities and infrastructure provided by the government to serve the public interest. There was a total of 15,228 check-in activities in this group consisting of five categories: education 11,139 (73%), health 1815 (12%), administrative office 879 (6%), worship 776 (5%), and bank 619 (4%). The percentages for these categories are described in Figure 8.18.

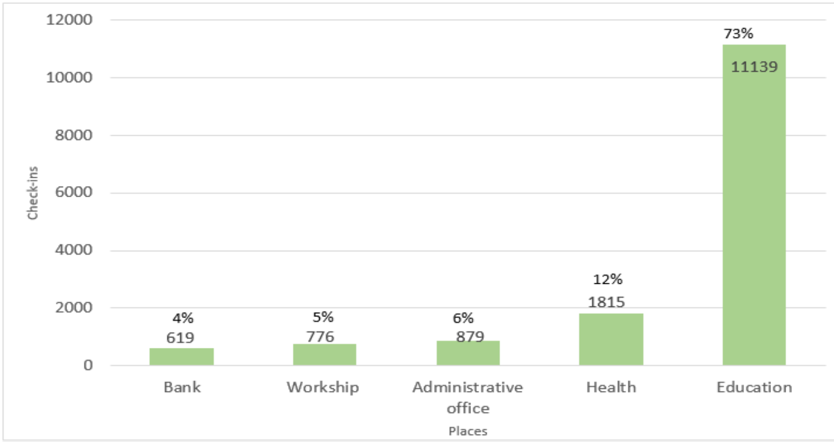


Figure 8.18 Place distribution for public facilities

8.4.2.1 Education facilities

For this part, two places were a priority for Twitter users. Besides school, the university was the most visited place. An explanation of these places is as follows:

a. School

Figure 8.19 shows a comparison of individual activities on weekdays and weekends. From the analysis, it was found that 8 a.m. was the peak time for check-in activity. The initial analysis concluded that individual activity occurred when a user arrived at school. There is a contrast between workdays and weekends as on weekends there is a decrease in check-in activity to below one hundred. The school facilities which people visit are: the laboratory with 29 (1%) check-ins, 26 (0.9%), field 23 (0.8%), basketball court 16 (0.6%), canteen (0.4%) and 96.3% of check-ins do not specifically mention the place in the school. Generally, they mention the school as the place where the Tweet activity is conducted (see Figure 8.20).

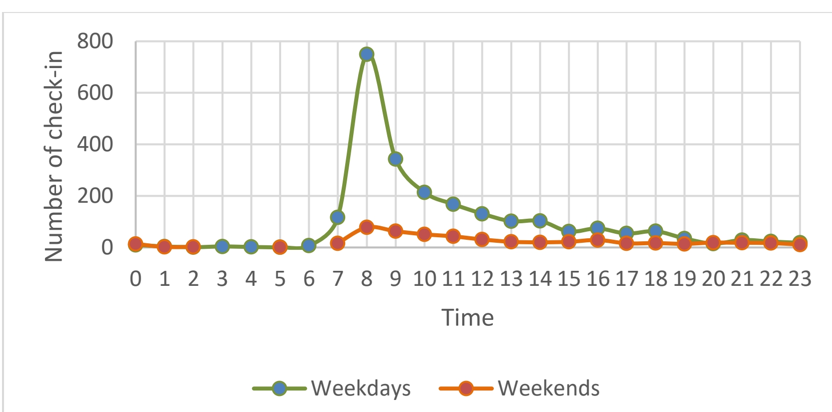


Figure 8.19 Differences in individual check-in activities on weekdays and weekends at the school

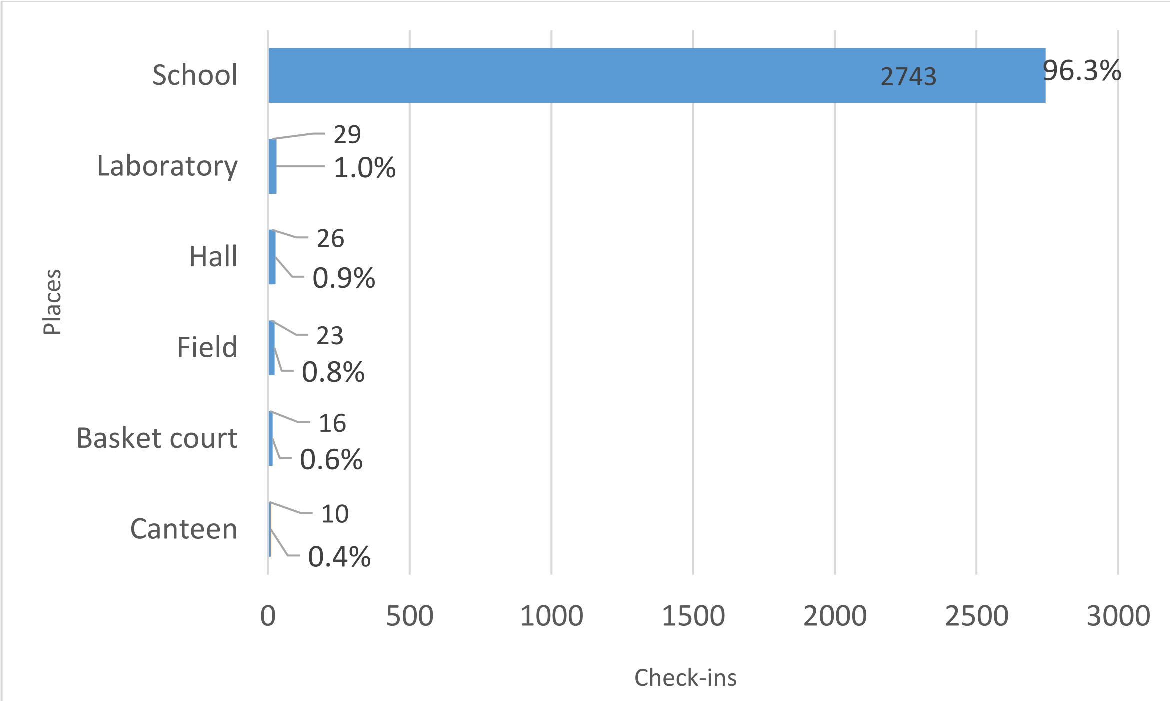


Figure 8.20 Check-in activities at the school facilities

b. University

The next facility that became a priority for people was the university. In the daily time distribution, the writer found that check-in activity started at 7 a.m.; the peak occurred at 11 a.m. (see Figure 8.21). There is a significant difference when comparing weekdays and weekends. On weekdays, the check-in peak took place at 9 a.m., whereas on weekends it occurred at 11 a.m. It is clear that on weekends there is a decrease in activity by up to 200 check-ins (see Figure 8.22).

Figure 8.23 shows the university facilities that people visit. Twenty-five check-ins or 0.3% occurred in bank, 30 in a place of worship (0.4%), 50 at a basketball court (0.6%), 56 at a football court (0.7%), 76 at a hall (0.9%), 154 at a library (1.8%), 161 at a laboratory (1.9%), 195 at a canteen (2.3%), and 7561 check-ins (90.7%) mentioned in the university did not mention the name of the university facilities in detail. Figure 8.24 shows the comparison of check-in activities at university facilities.

Beside campus facilities, users also mentioned at which faculty they performed check-in activities. A total of 2032 check-ins were made at a university faculty. The percentage of each faculty are: faculty of law—414 check-ins or about 20%, economics—334 (16%), medicine—322 (16%), mathematics—221 (11%), public health—173 (9%), engineering—120 (6%), politics—116 (6%), pharmacy—113 (6%), agriculture—73 (4%), forestry—53 (3%), fisheries—47 (2%), farming—25 (1%), and culture—21 (1%). The percentages for each faculty are described in Figure 8.25.

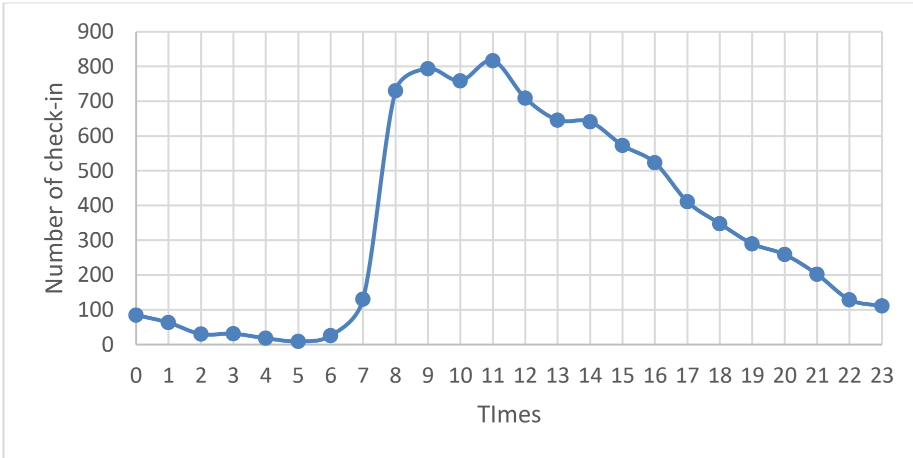


Figure 8.21 Time distribution for university: daily

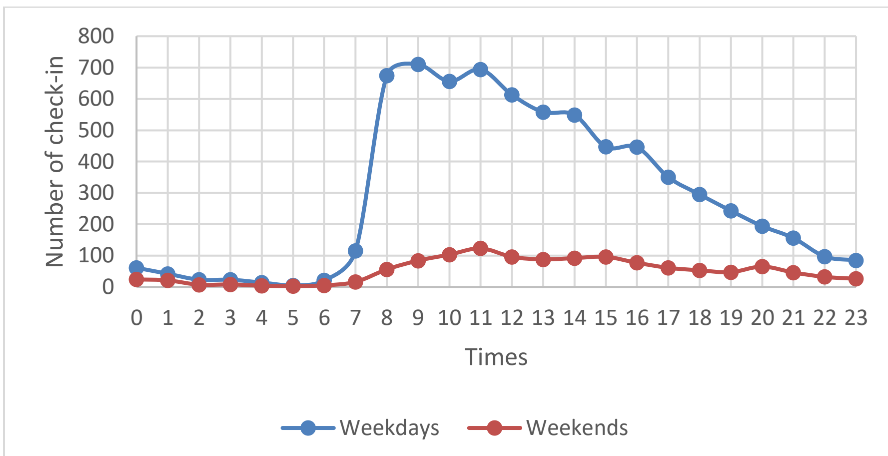


Figure 8.22 Differences in individual check-in activities on weekdays and weekends at the university

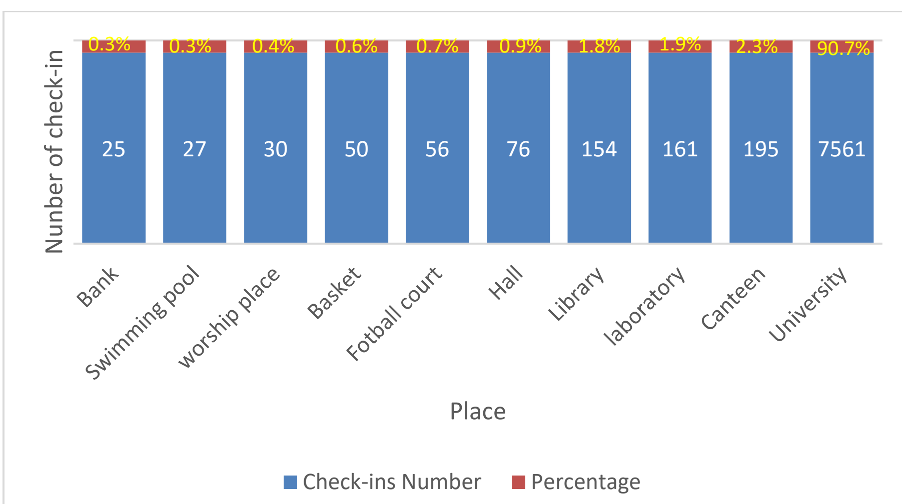


Figure 8.23 Differences in individual check-in activities at the university and university facilities

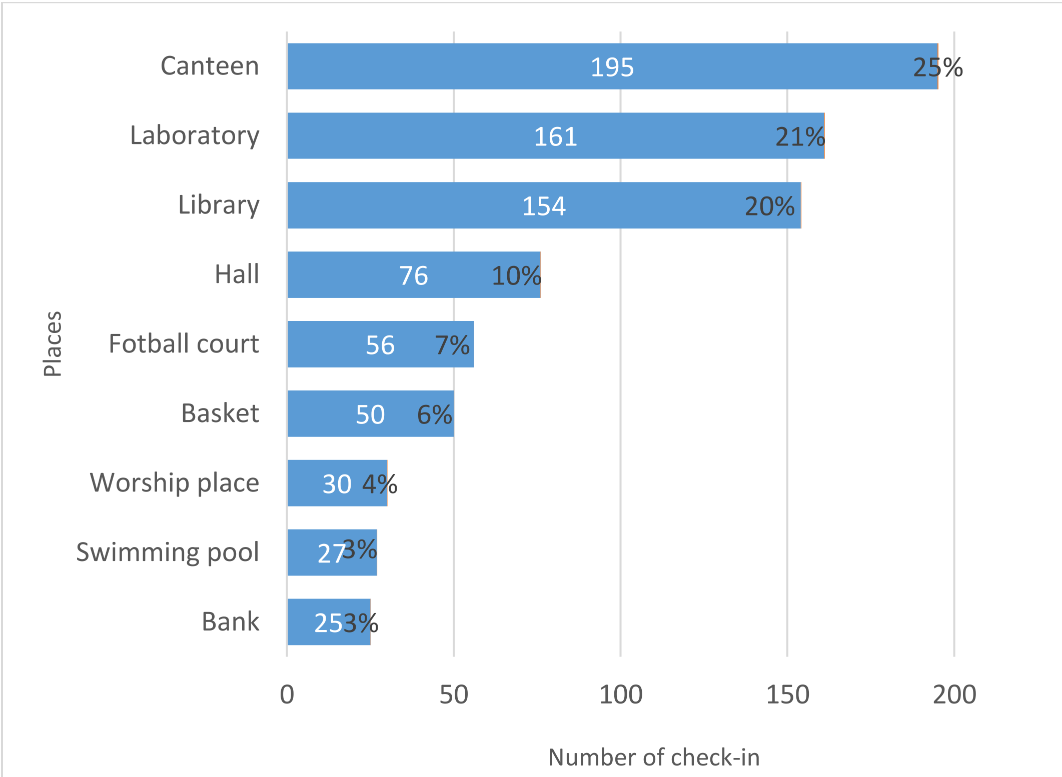


Figure 8.24 Differences in individual check-in activities in the university facilities

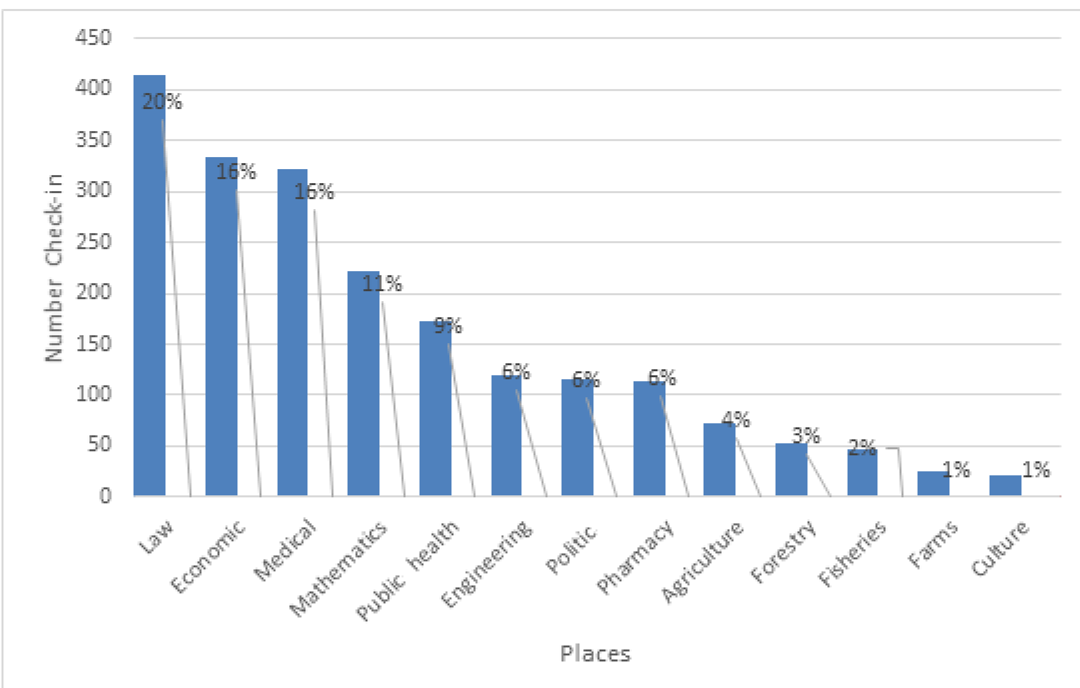


Figure 8.25 Check-in activities in the university faculties

8.4.2.2 Health

In this category, almost all check-in activities are conducted in the hospital. Figure 8.26 shows in detail the daily mobility of people with the peak check-in occurring at 9 p.m. Figure 8.27 shows the different check-in activity on weekdays and weekends. On weekdays, peak check-in took place at 8 a.m. and 5 p.m. whereas on weekends, peak check-ins occurred at 7 p.m. and 9 p.m.

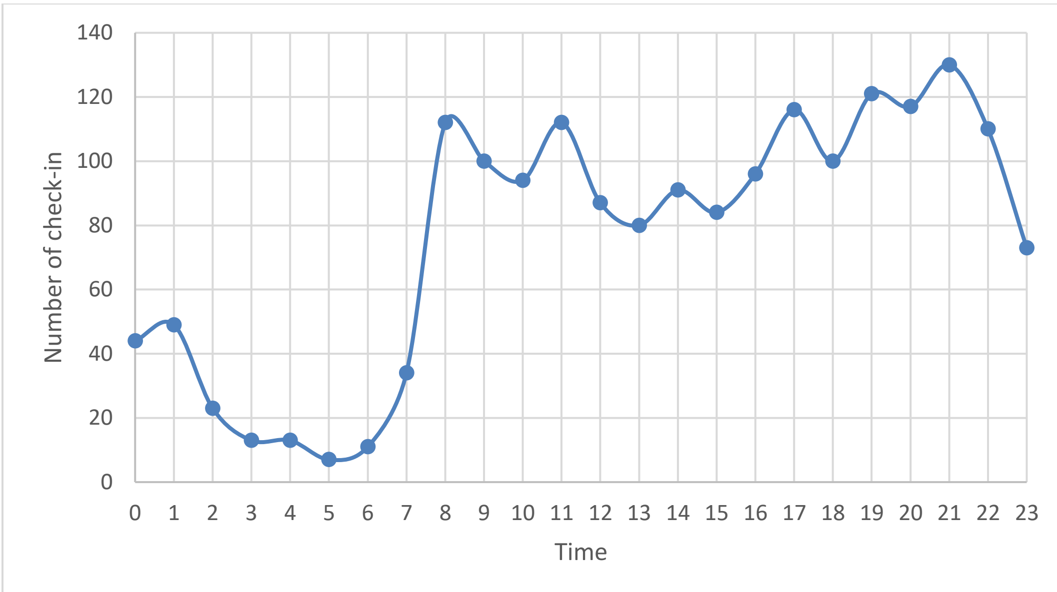


Figure 8.26 Time distribution for hospital: daily

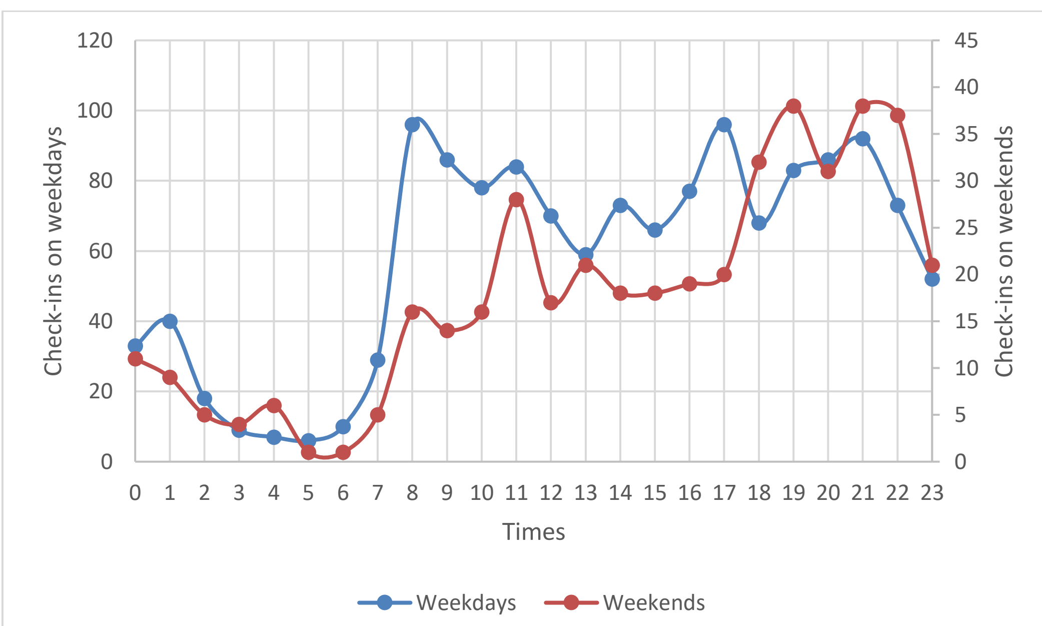


Figure 8.27 Differences in individual check-in activities on weekdays and weekends at the hospital

8.4.2.3 Places of worship

In this category, the author found two places of worship that were a priority for users—the mosque and church. The author observed that in the morning, the check-in peak occurred at 7 a.m., noon, at 1 p.m., and in the evening at 7 p.m. (Figure 8.28). The existence of other activities at certain hours perhaps occurred in the public spaces provided by the facility.

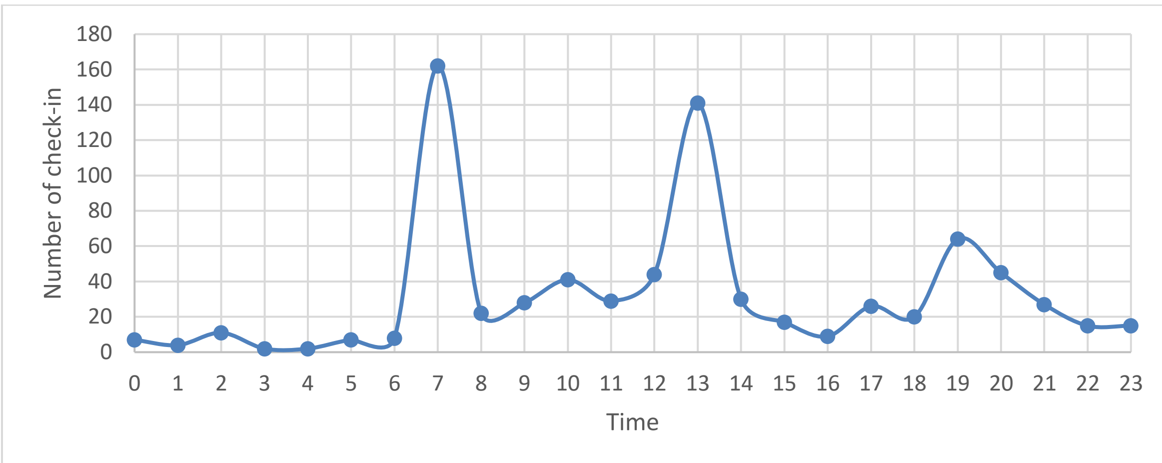


Figure 8.28 Time distribution for places of worship: daily

8.4.2.4 Banks

The next public facility is banks. On a daily basis, the check-in peak occurred at 9 a.m. (Figure 8.29). There was the difference between weekdays and weekends. On weekdays the check-in peak took place at 11 a.m. and on weekends at 9 a.m. On weekends there was the decrease in check-in activity with an average of below ten check-ins each hour.

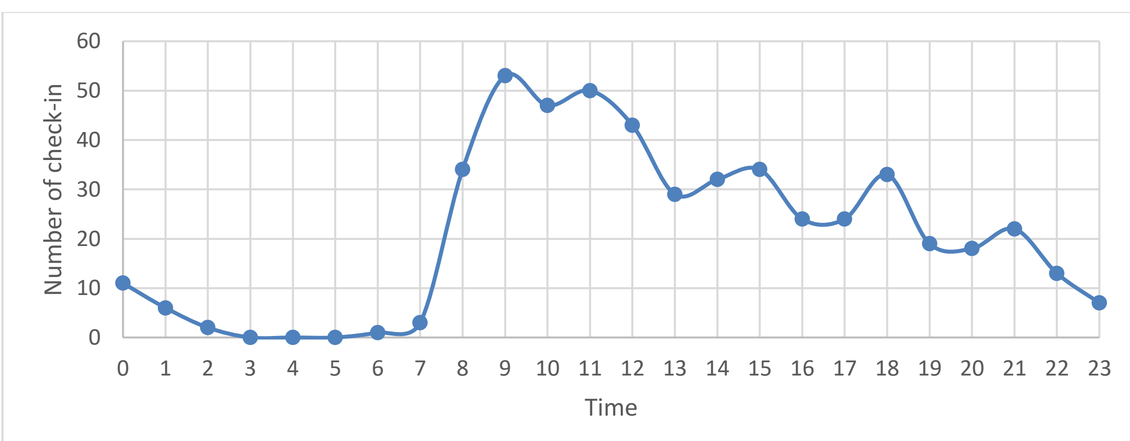


Figure 8.29 Time distribution for places of worship: daily

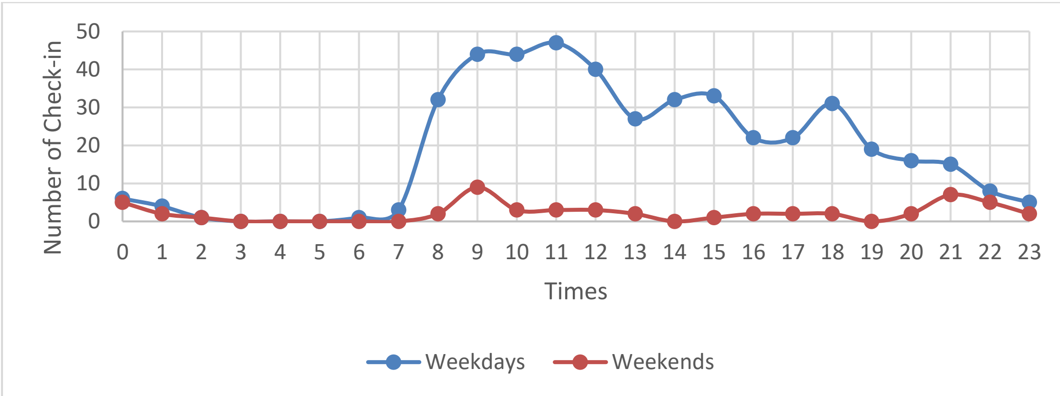


Figure 8.30 Differences in individual check-in activities on weekdays and weekends at the bank

8.4.2.5 Administrative offices

The author defines an administrative office as a place and infrastructure provided by the government for public services. These offices include the administration of residences, public utilities, and services offices, such as those for water, electricity, telephone, and security and safety service posts.

Figure 8.31 displays the public facilities check-ins for administrative offices. There were ten places that were a priority for people to visit: military district with 121 check-ins or about 19%, police station—120 (18%), insurance—76 (12%), electric—75 (11%), education—67 (10%), mixed office—64 (10%), governor—44 (7%), training center—42 (6%), high court—39 (6%), and post office—6 (1%).

The author then analyzed the daily distribution of users and found peak check-in activity to be at 9 a.m. A comparison of the mobility of users on weekdays and weekends revealed that on weekdays the peak for check-ins was at 9 a.m. and on weekends, it was also at 9 a.m. On weekends there was a decrease in tweet activity with an average of under twenty check-ins each hour.

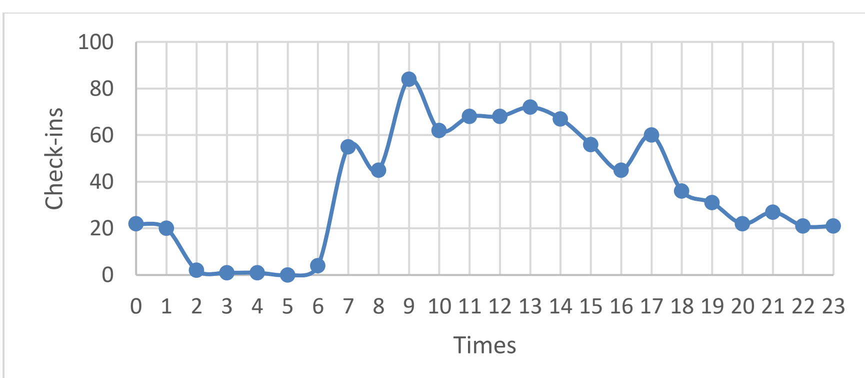


Figure 8.31 Time distribution for administrative offices: daily

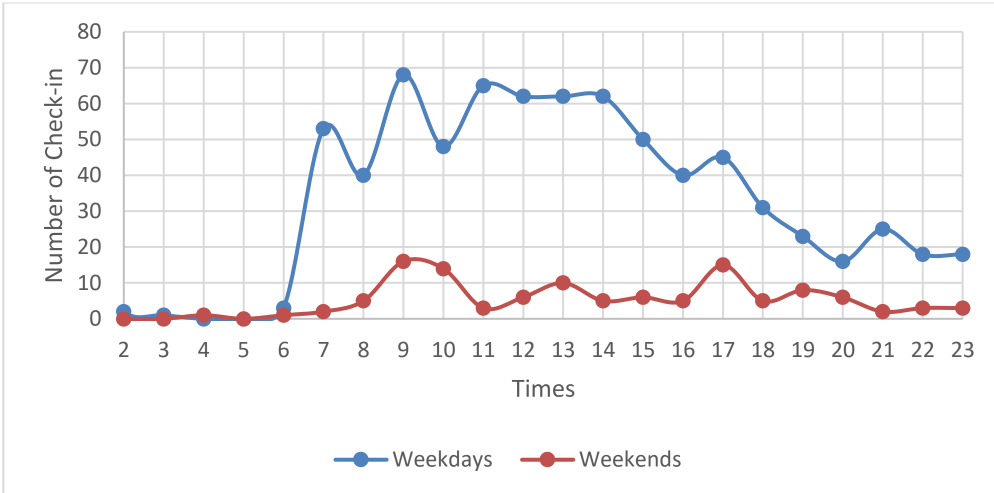


Figure 8.31 Differences in individual check-in activities on weekdays and weekends at the administrative offices

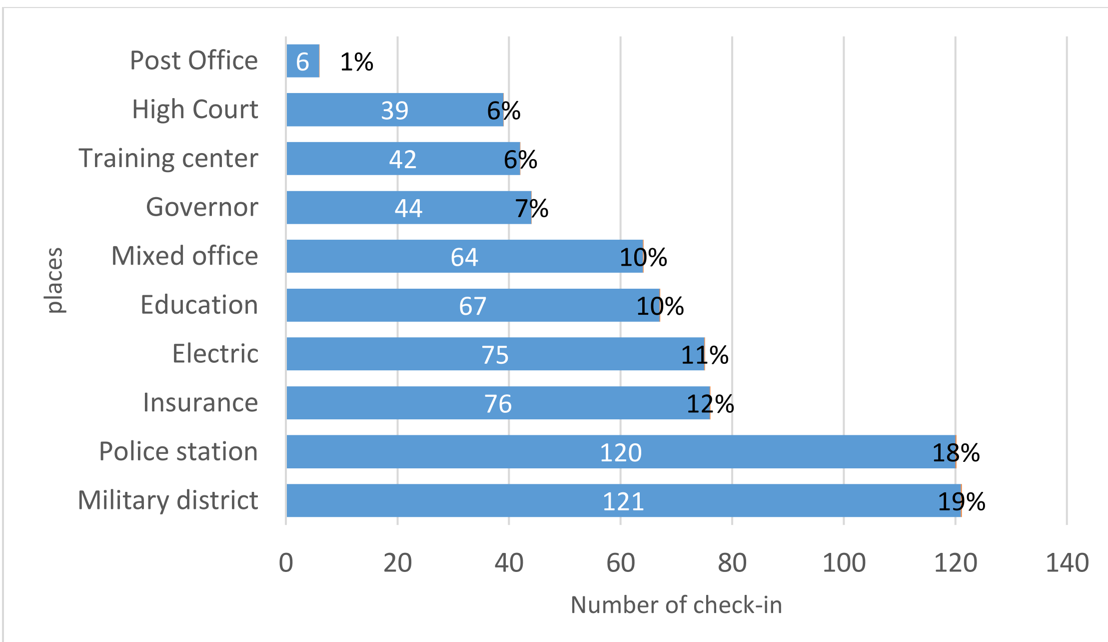


Figure 8.33 Check-in activities at the administrative office facilities

8.5 Section Conclusion

In this chapter, the author focused on individual distribution check-ins in public spaces and public facilities. For the public spaces, it was found that the beach was the most visited place with 1263 check-in activities or about 34%. Cultural heritage places had the lowest tweet activity with 174 check-ins or approximately 5 %. With respect to public facilities, universities and schools were the most visited places with 11,139 check-ins or 73%, and banks were the places with the lowest tweet activity with 619 check-ins or 4% of total visitors.

In the time distribution, there are some places that have check-in peak times. Figure 8.34 and Figure 8.35 shows the check-in peak times in the same places for weekdays and weekends. On weekdays, the peak check-in at the sports stadium and the beach was at 7 p.m. The park, swimming pool, and health was at 5 p.m. and the university was at 9 a.m. On weekends, for sports fields and the beach it was at 9 p.m., for health and cultural heritage at 7 p.m., and for government offices along with banks at 9 a.m. This means that on weekdays and weekends, at the same time and in different places, the same peak check-in activities occur.

Table 8.4 List of check-ins peak on weekdays, weekends, and daily period

Place	Weekdays time	Weekends time	Daily time
Park	17	15	17
Sport Field	18	21	21
Sport Stadium	19	16	19
Swimming pool	17	17	17
Beach	19	21	19
Cultural heritage	24	19	24
School	8	8	8
University	9	11	11
Healthy	17	19	21
Workshop place	7	10	7
Bank	11	9	9
Administrative office	9	9	9

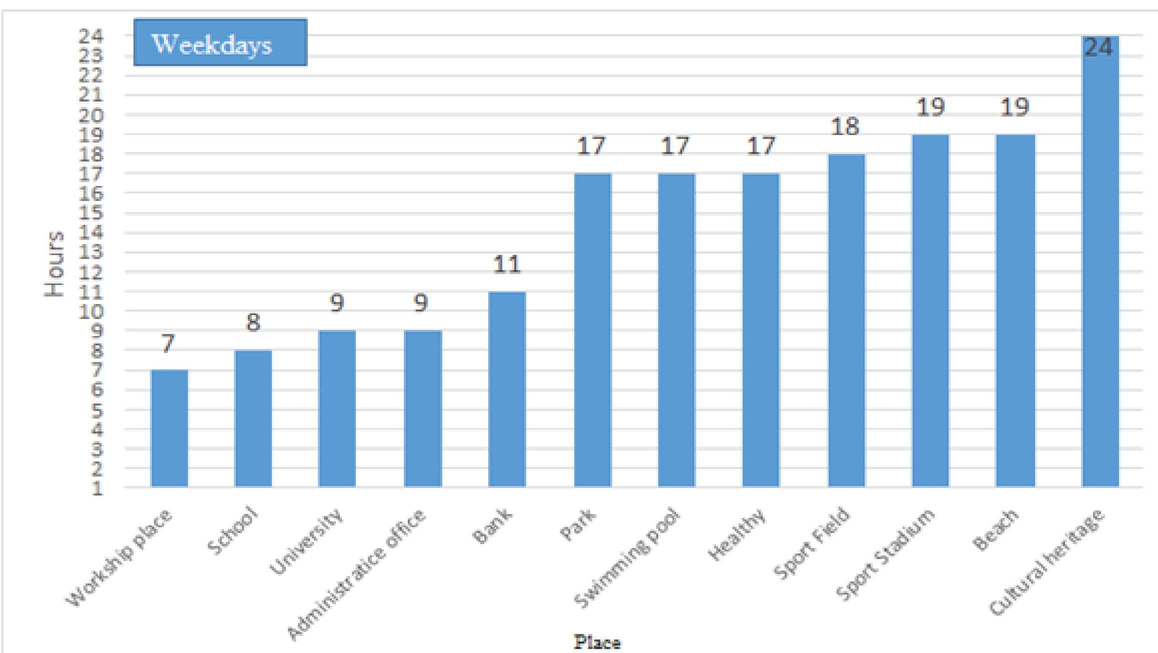


Figure 8.32 places which have check-in peak time in the same places on weekdays

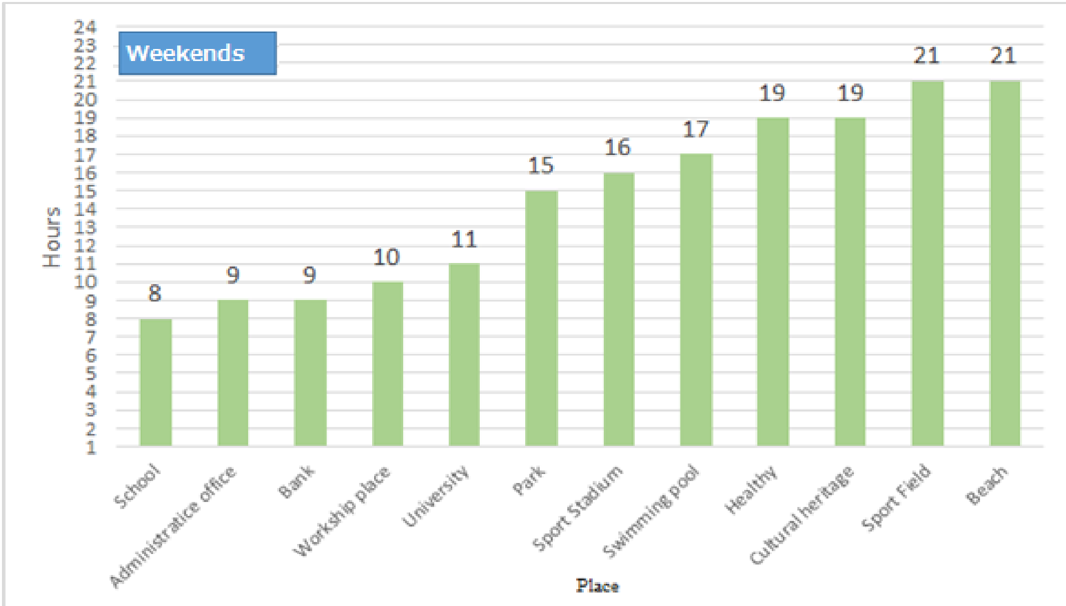


Figure 8.33 places which have check-in peak time in the same places on weekends

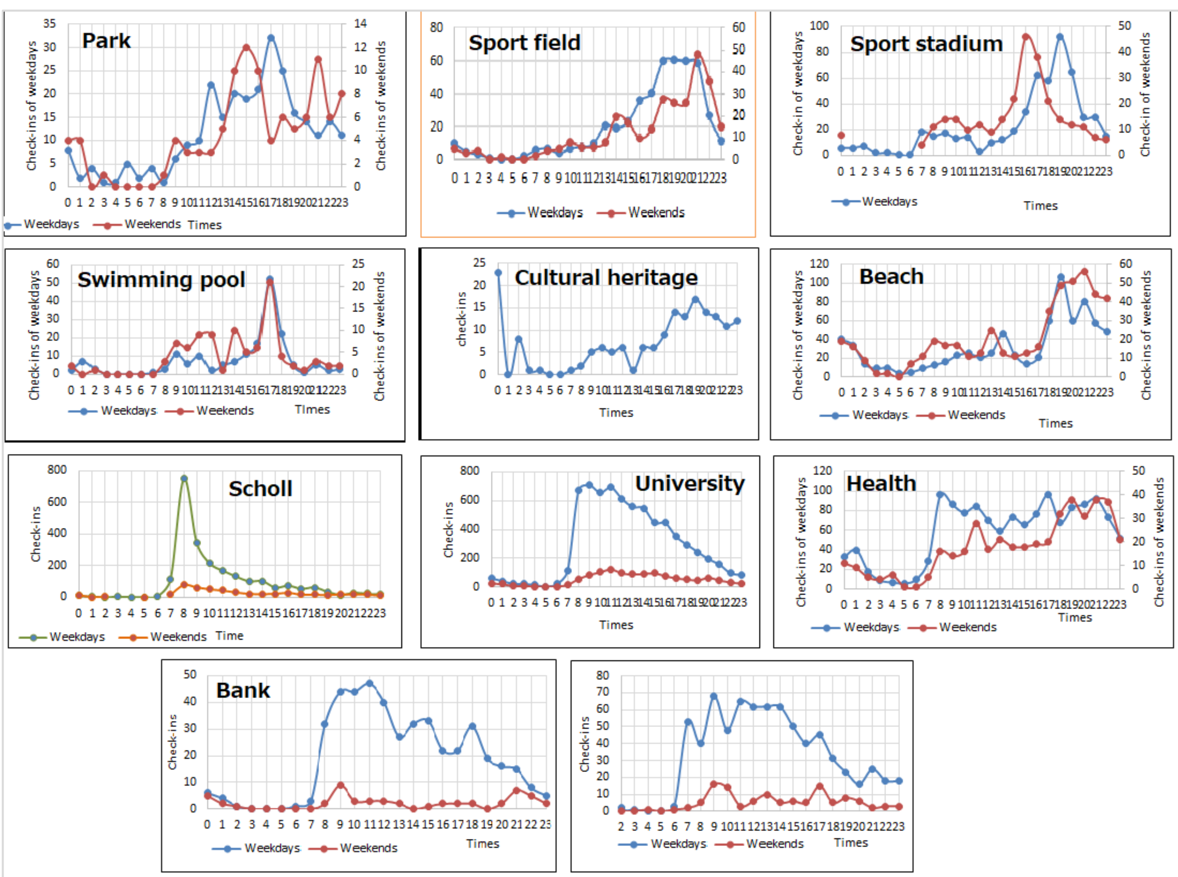
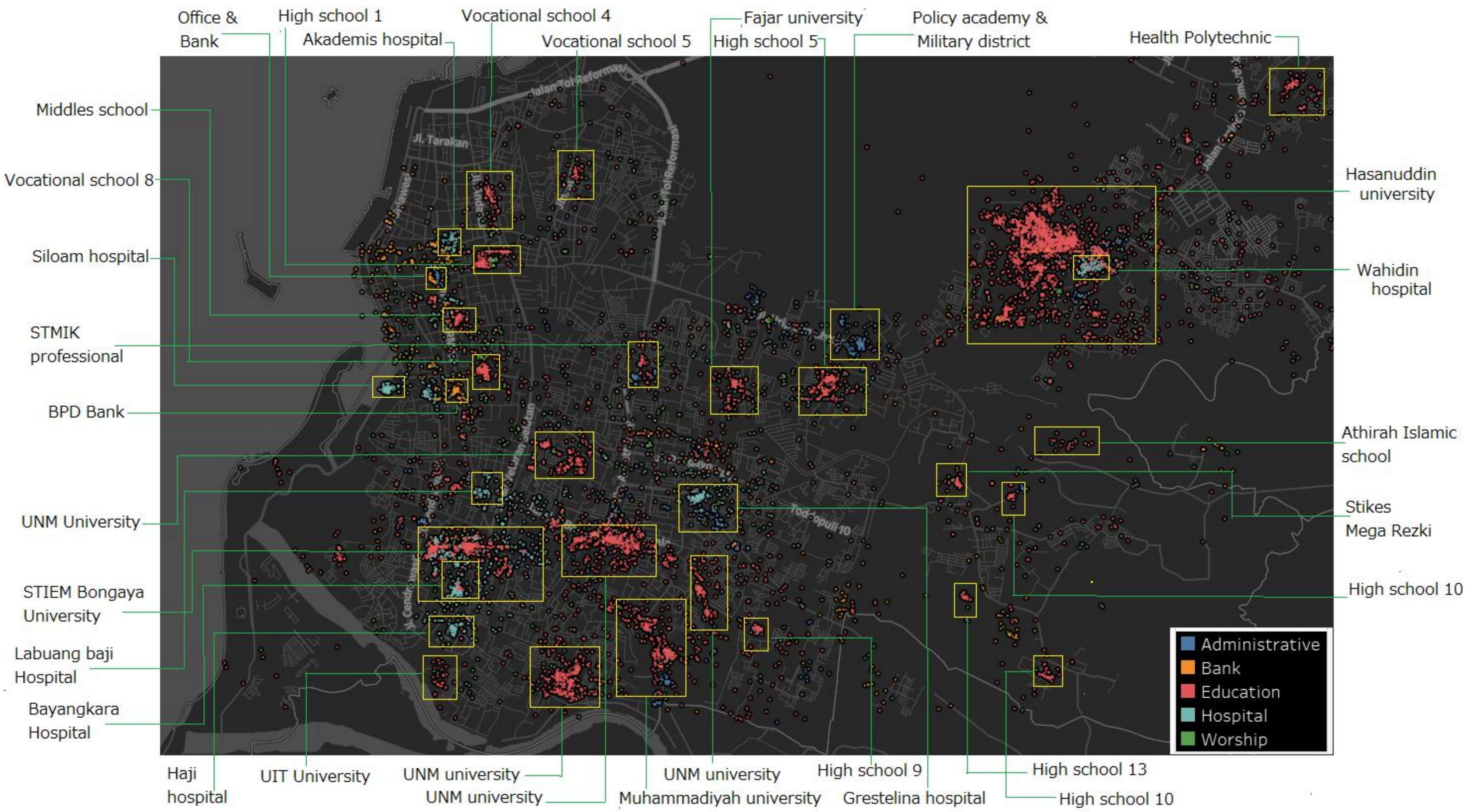


Figure 8 34 Comparison individual check-in activities on weekdays and weekends in the public spaces and facilities and

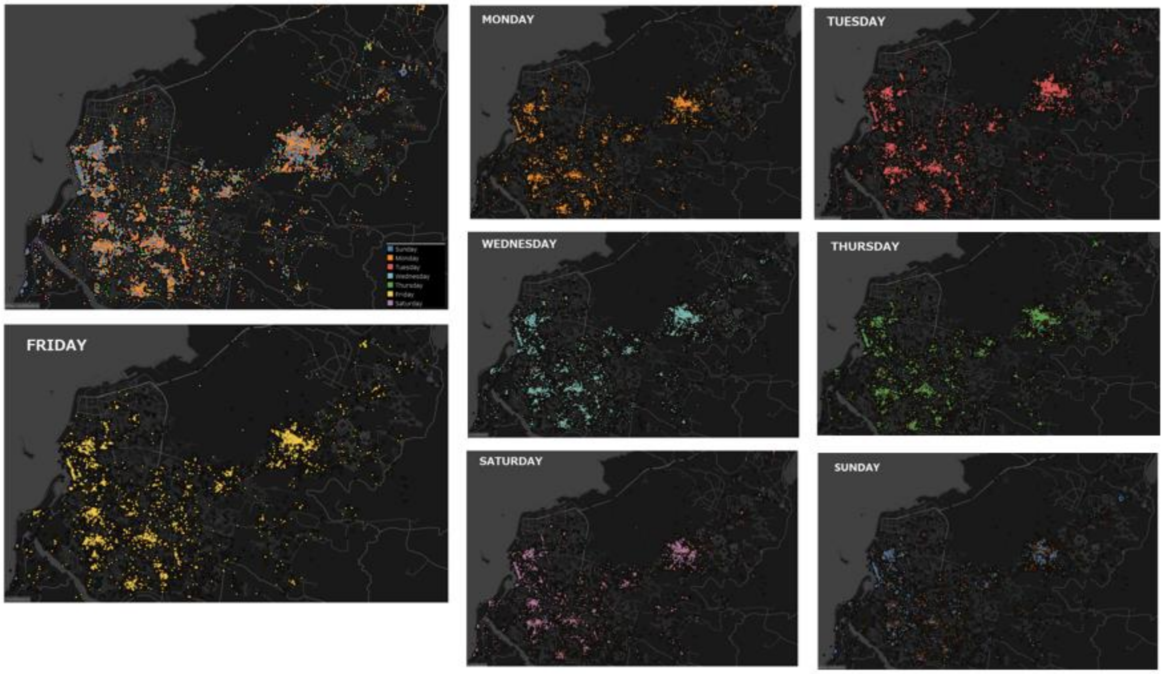
Appendix

1. Spatial distribution map of public Facilities

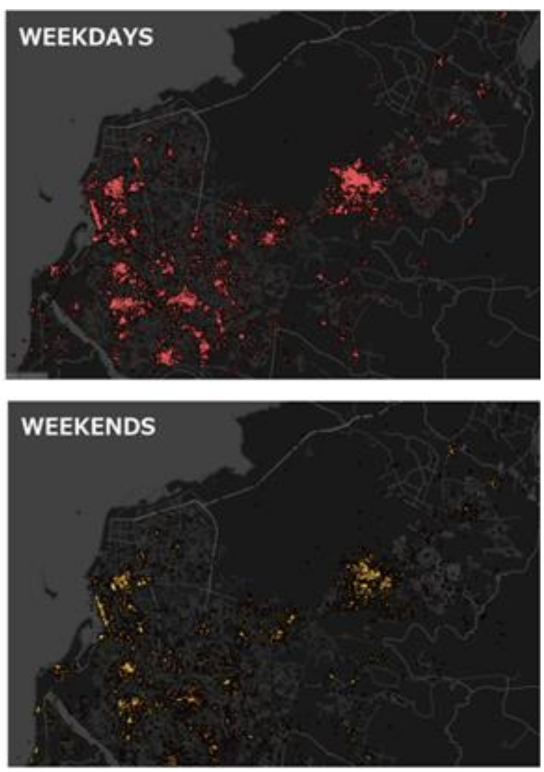


2. Spatial distribution map of public Spaces

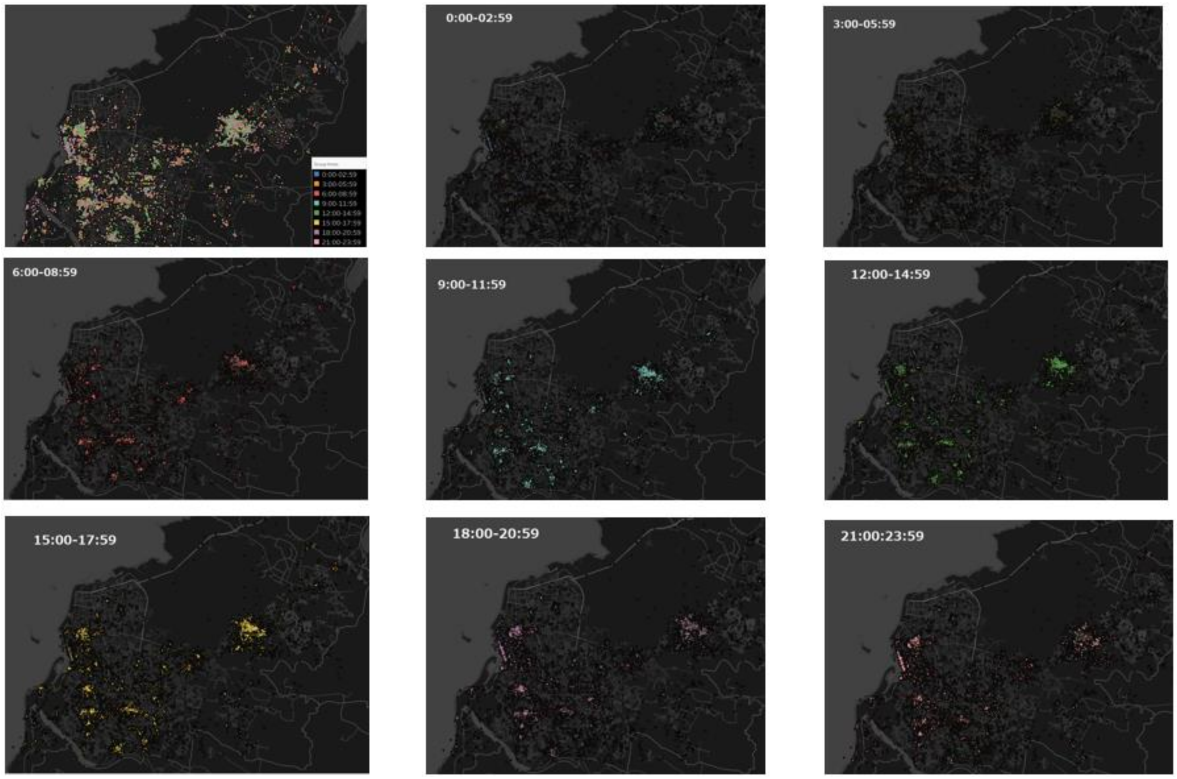
3. Spatial distribution maps of Twitter venues of public spaces and public facilities on Monday to Sunday



4. Spatial distribution maps of Twitter venues of public spaces and public facilities on Weekdays and Weekends



5. User Social Media Interaction towards Urban Public Space and Public Facility



References

1. Stephen, C.; Mark, F.; Leanne G.; Rivlin; Andrew, M.S. Public space. New York: Cambridge University Press, **1992**.
2. Tonnelat. S. the Sociology of urban public spaces. Atlantis press, **2010**.

Chapter 9
Conclusion

9.1 Result

In this disertation, the writer has analyzed 211,922 check-ins on Twitter in Makassar City. The research answers the question of whether the geolocation of social media data can be useful in the creation of novel and dynamic indicators for urban planning. From the research results in the previous chapters (4, 5, 6 and 7) it is clear that the dataset has significant benefits in this regard. The results indicate that:

1. The combination of Twitter and questionnaire data offers a great opportunity to assess the region's population. By using the geographic date and time features of Twitter, this dataset can determine the number of people at a specific time. Table 4.2 shows the Twitter dataset by sub-district and displays the comparison between the questionnaire results and the real population data. As some features are private, with the result that the user's profile on social media is inaccessible, the author thus used the questionnaire data to identify the gender, age, and work of the respondent. If analyzed more deeply, the distribution of this data can have the potential to estimate urban populations. In particular, it can be used to calculate the daily population—in this case, the daytime population and the nighttime population.
2. With the same data, the author attempted to utilize Twitter data to investigate the use of check-in data as a source of information to characterize dynamic urban land use. Three kinds of data prioritized in this research were check-ins (specific location), timestamps, and a user's status text or post activities. The analysis shows that the check-in activity and the method that was proposed can be used to group the actual land-use types. Figure 5.11 shows a land-use hypothesis that identifies an area according to function, as, for example, a commercial, business, education or mixed area. Thus, the advantages of this dataset are that it can display daily and dynamic land use.
3. In another study, the author analyzed the distribution of Twitter users to understand urban inhabitants' mobility patterns. In this study, the author answers the following questions: What is the movement pattern of a citizen? How far do people travel in the city? From the results of the analysis, it can be concluded that the average daily mobility of individuals in the city of Makassar is 9.6 km. As shown in Table 6.2, the distribution of individual weekly travel distances based on the accumulation of users' daily travel is displayed. The writer found four models for individual mobility patterns of places visited (1) university - mall - home, university - dormitory, office - mall – home, and office - home and others. From the distribution of respondents (Table

4.1), individual activities are dominated by students. In other words, their relationship to the shopping center (mall) is very high as can be seen from their mobility pattern. Thus, the results of this analysis can be additional data for city planners to address transportation problems and traffic congestion in particular.

4. In the chapter 7, the writer analyzed and identified the city center area. The experimental results indicate that individual movement resulting from user distribution of Twitter can delineate the boundaries of the city core such as displayed in Figure 7.22 concerning the city center area hypothesis.
5. The last chapter, the author identified the individuals check-in activity in the publict spaces and publict facilities. The research concluded that beach was the place with ther most places and the cultural heritahe had the lowest tweet activity in the publict spaces. Meanwhile in the publict facilities, school and university were the most visited places and banks were the places with the lowest tweet activity.

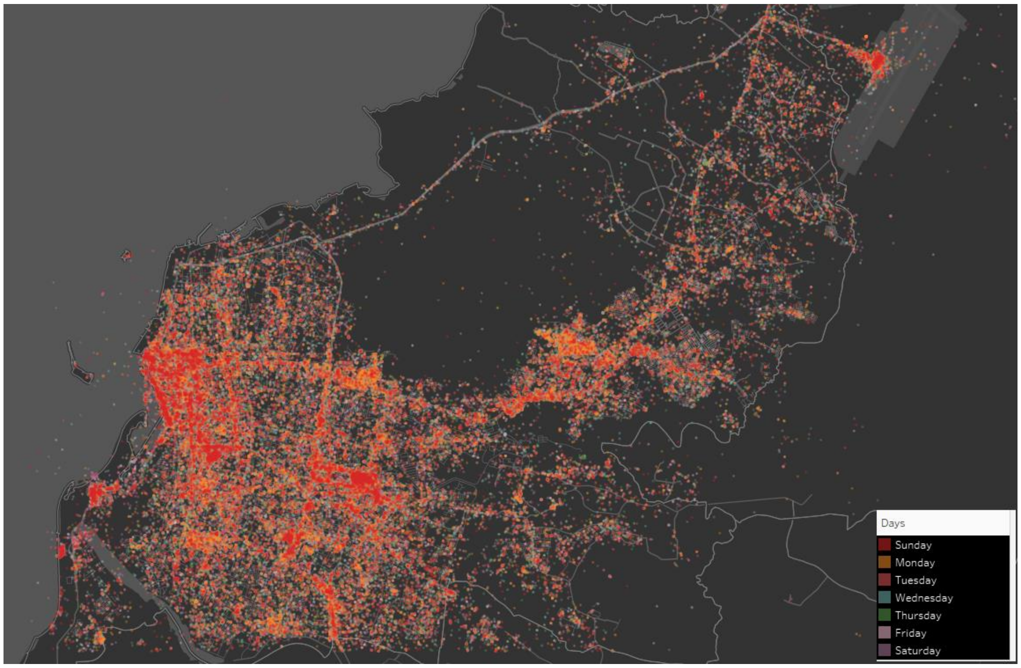
9.2 Future Work

Due to this research being limited to one city and it only using geolocation data as the parameter, future research should compare the data with other cities and the use of social media content, e.g., tags and comments, to know what kinds of things individuals are talking about and whether what they discuss is related to urban issues. Particularly in chapter 6, to improve the accuracy of the research results, it is necessary to analyze the users who have check-in activities in excess of 50 times. Investigating users with many check-in activities will affect the outcome of an experiment.

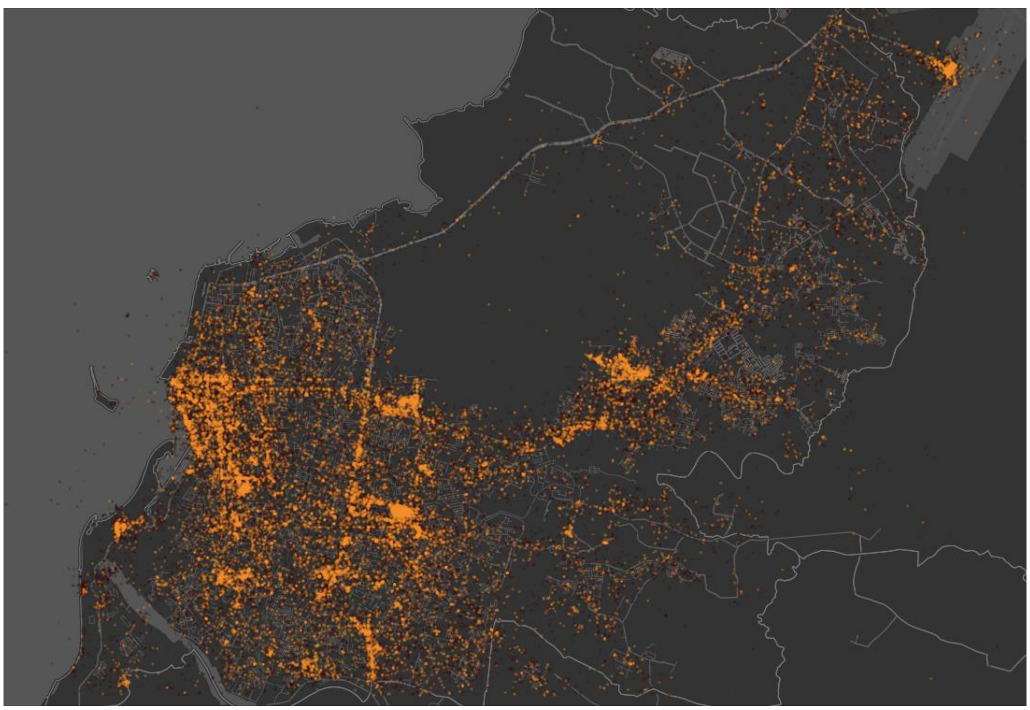
APPENDIX 1

MAPS DISTRIBUTION OF INDIVIDUAL CHECK-INS ON MONDAY TO SUNDAY

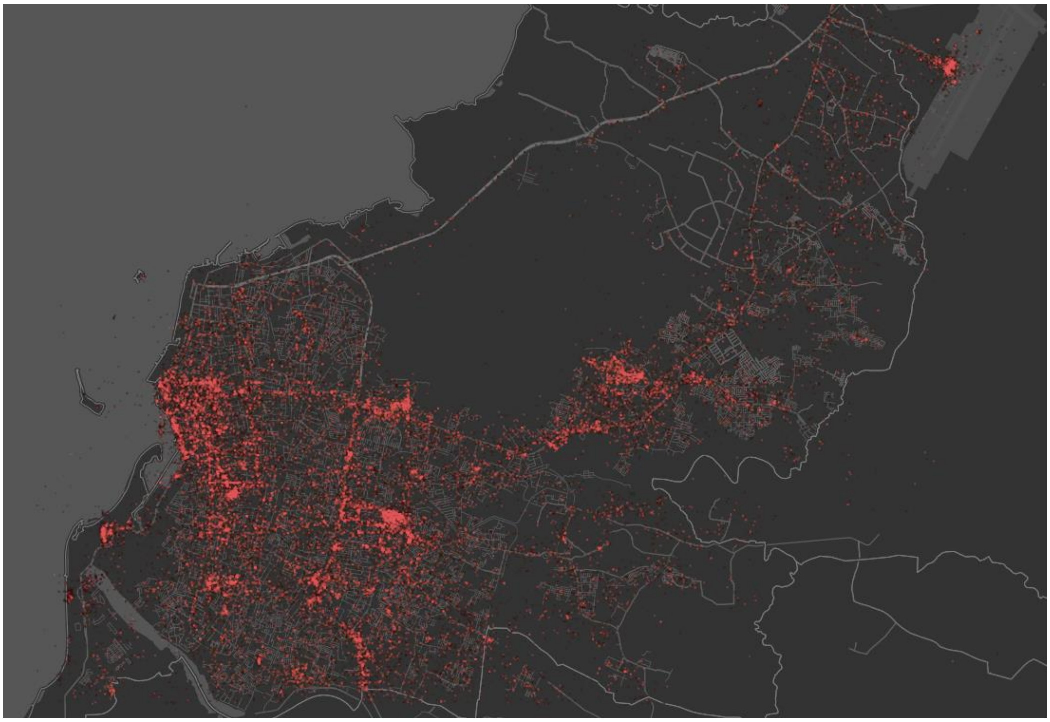
1. Map of individual check-ins distribution for all days



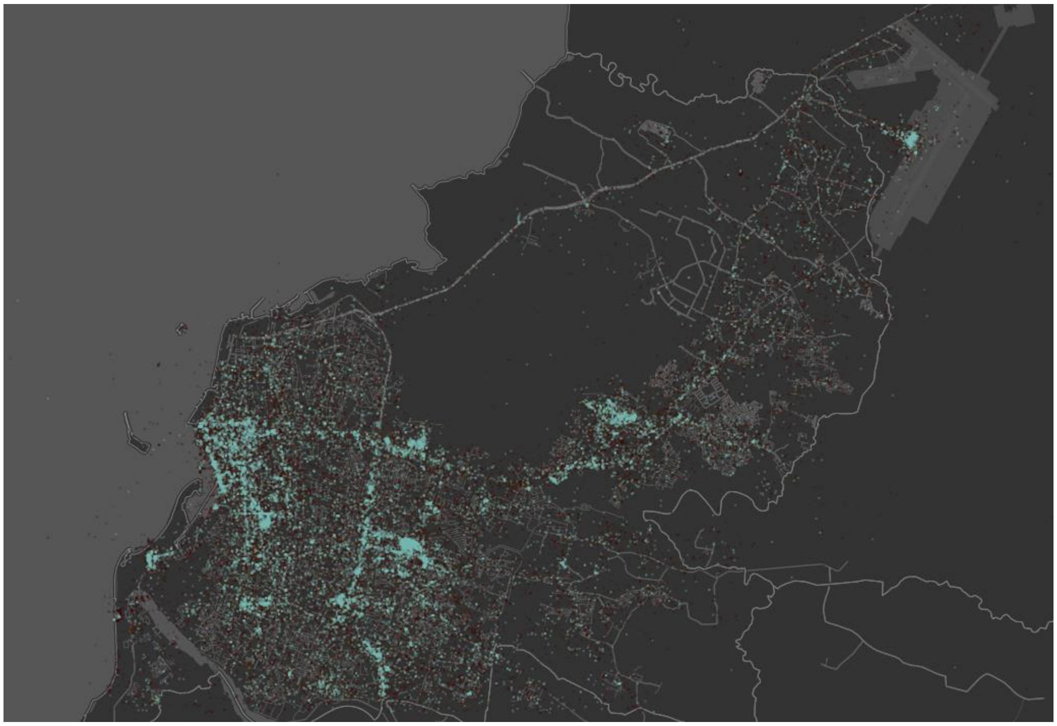
2. Map of individual check-ins on Monday



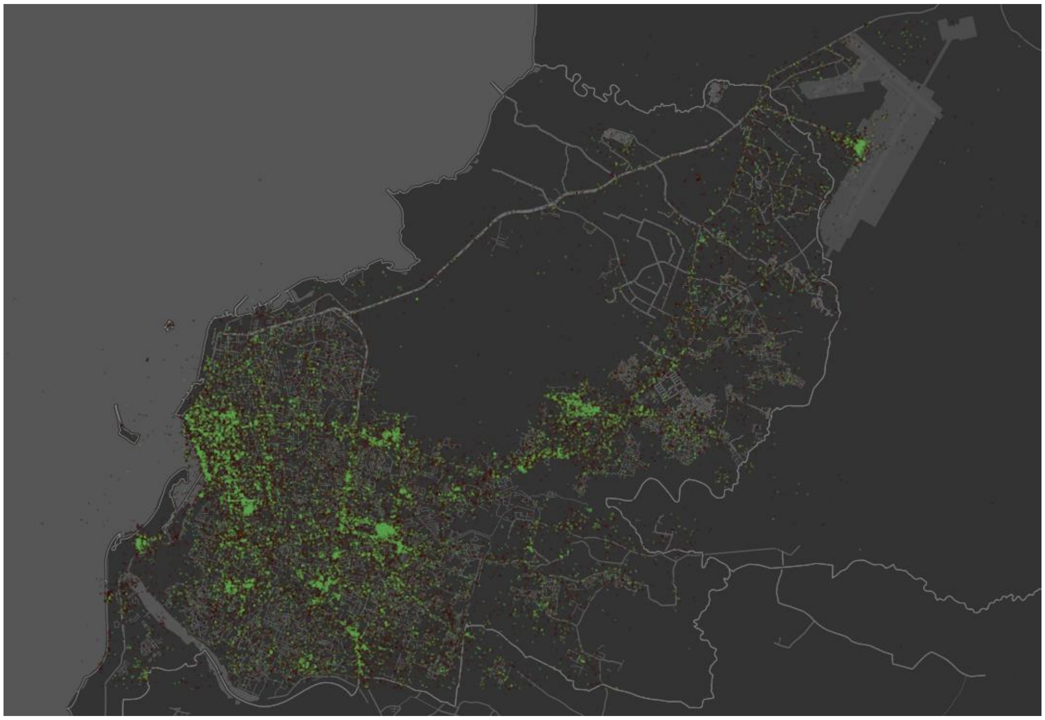
3. Map of individual check-ins on Tuesday



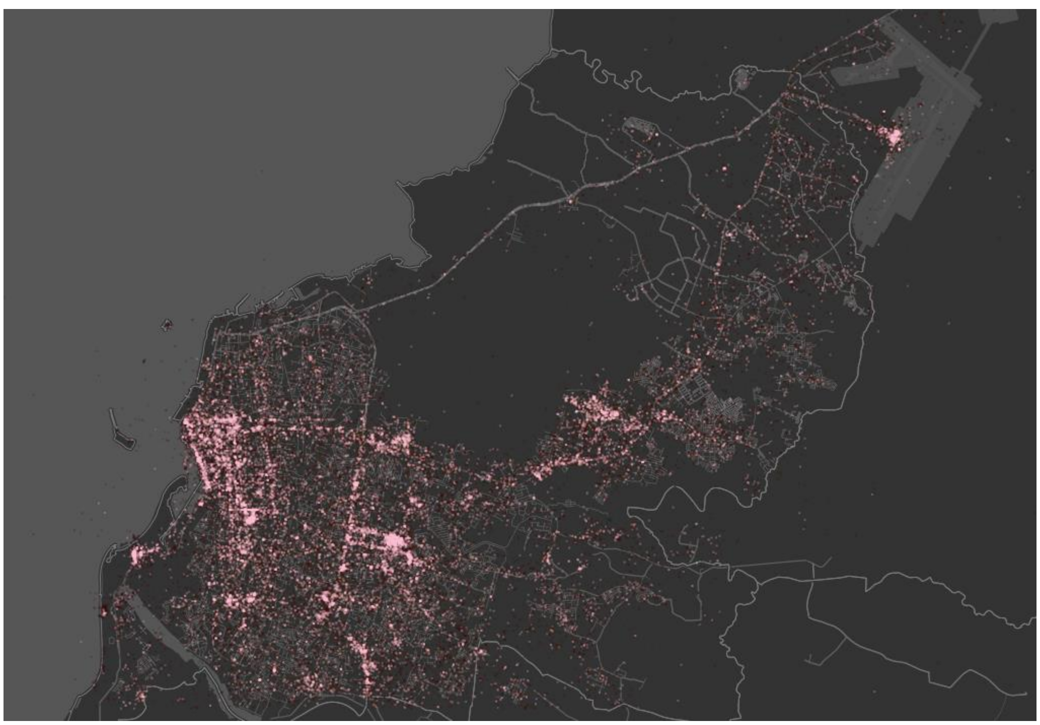
4. Map of individual check-ins on Wednesday



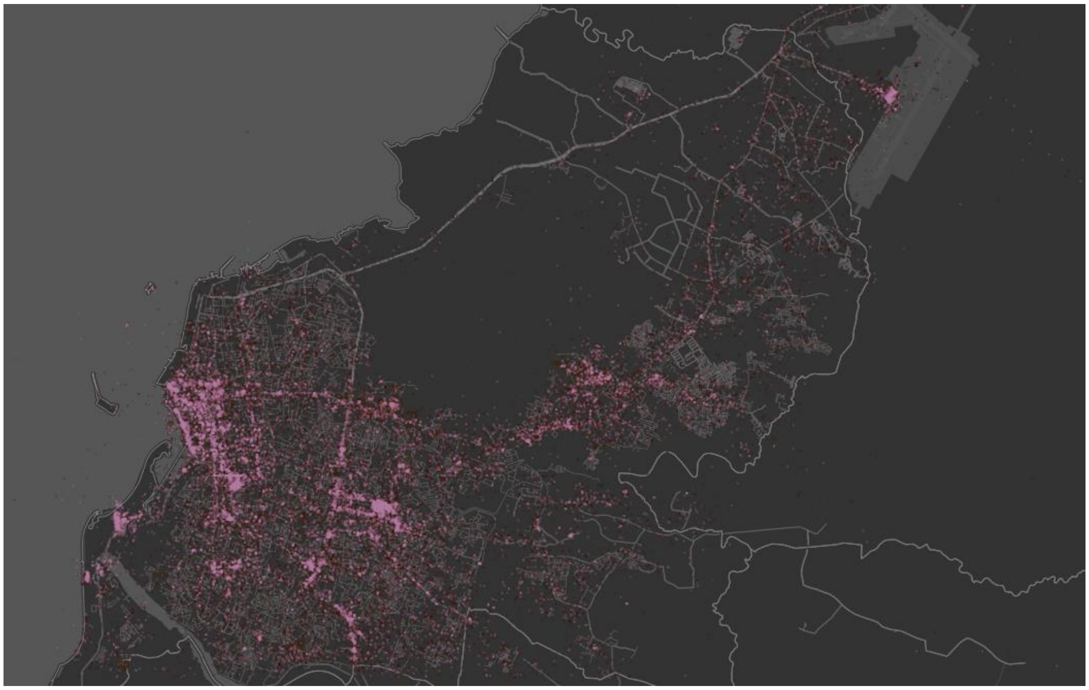
5. Map of individual check-ins on Thursday



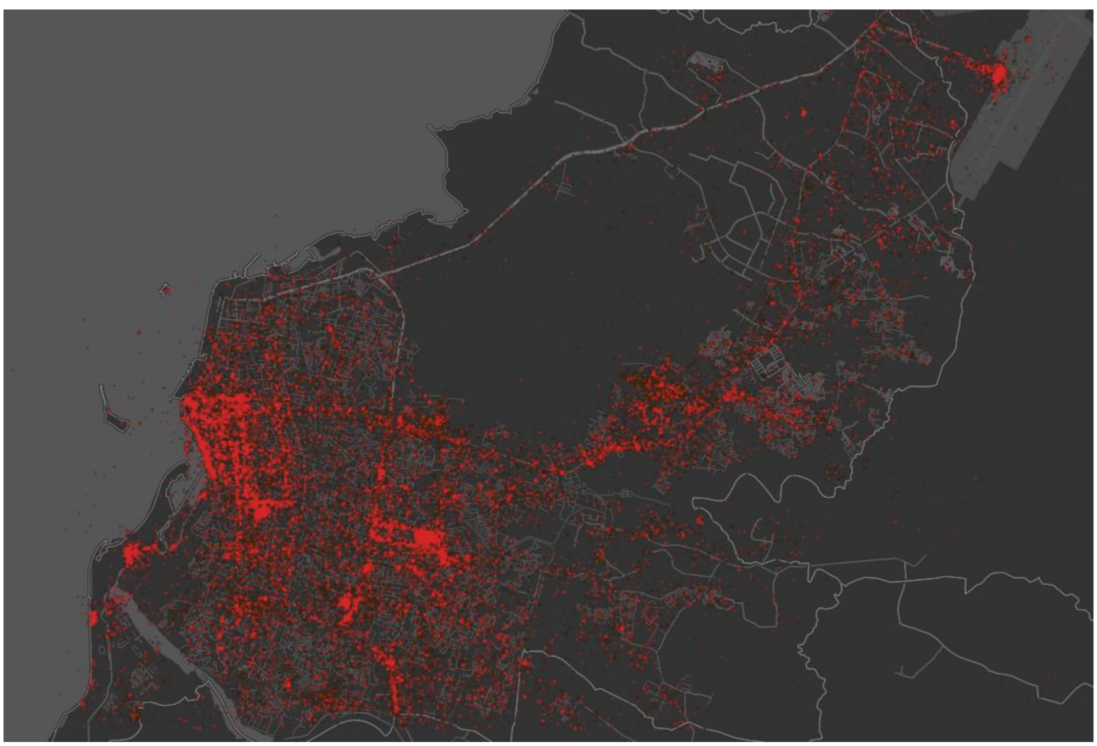
6 Map of individual check-ins on Friday



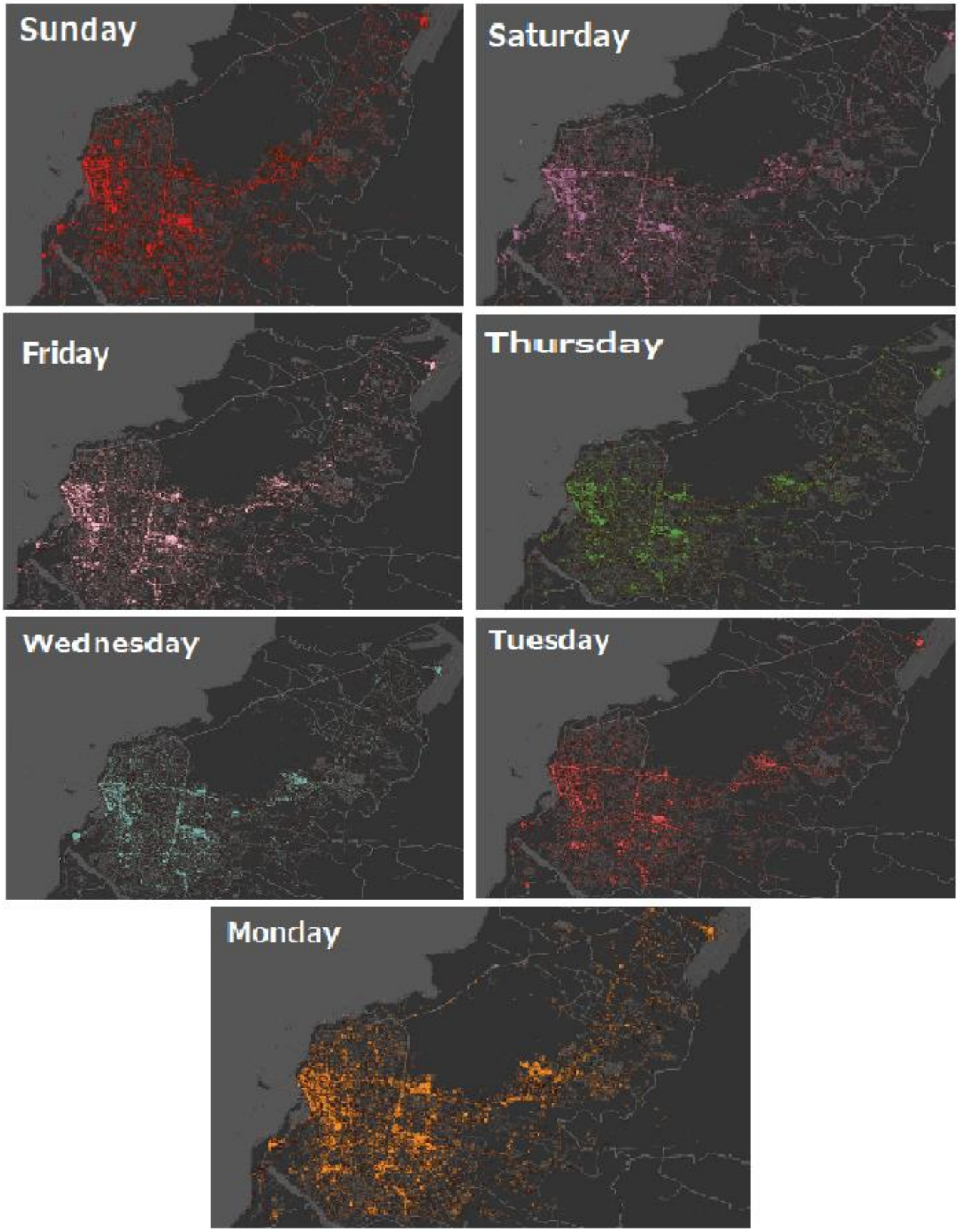
7. Map of individual check-ins on Saturday



8 Map of individual check-ins on Sunday



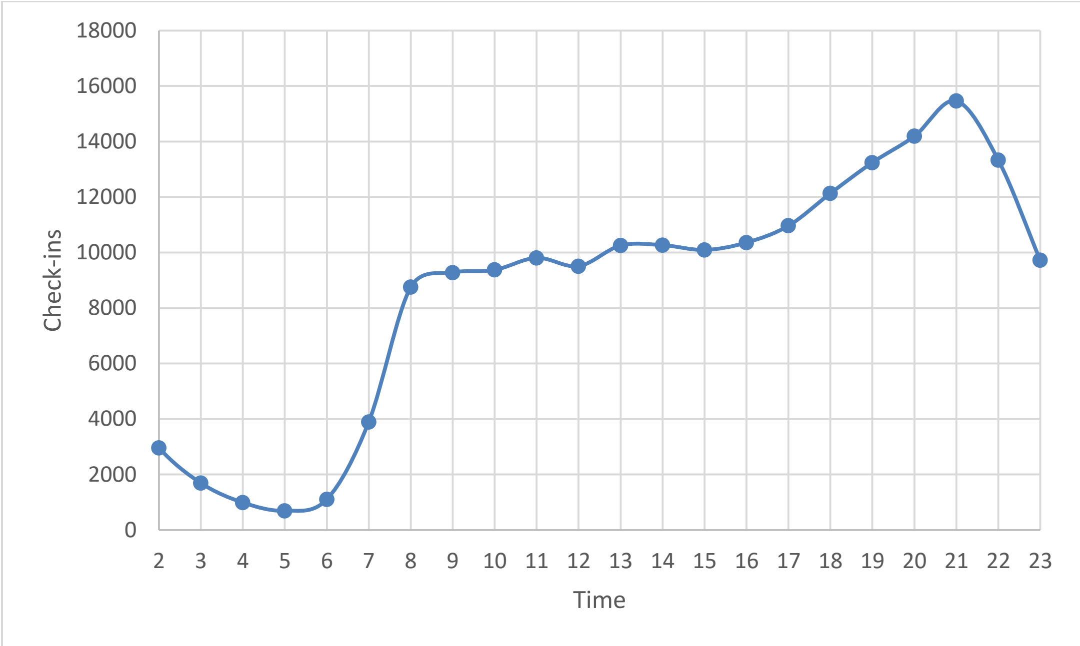
9. Maps of individual check-in distribution on Monday to Sunday



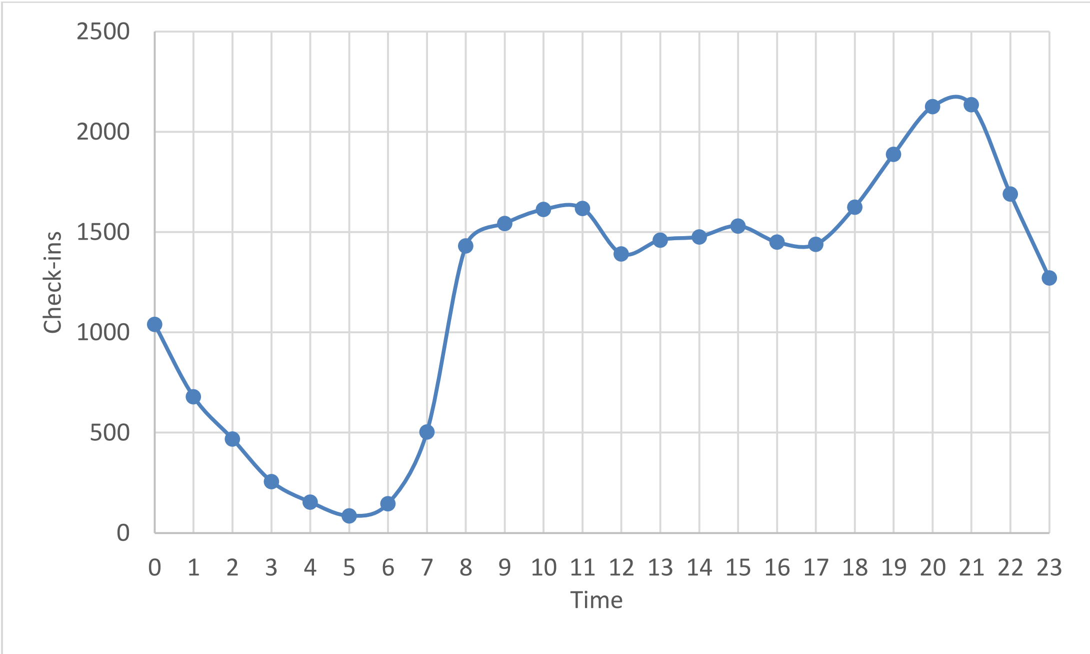
APPENDIX 2.

DAILY TIMES DISTRIBUTION OF INDIVIDUAL CHECK-INS ON MONDAY TO SUNDAY

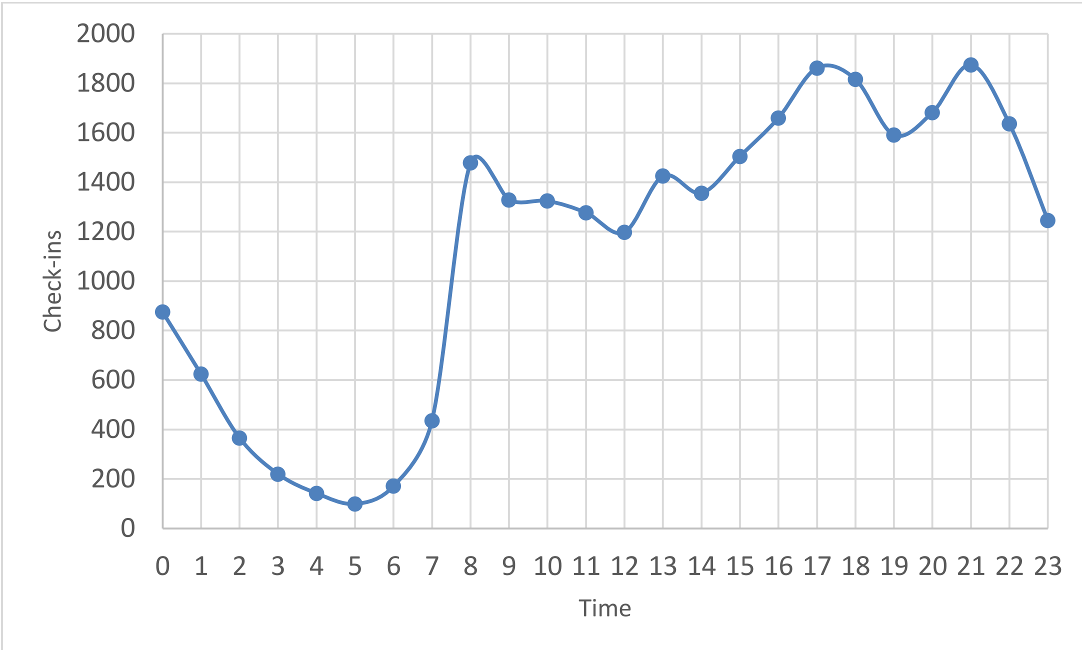
1. Individuals check-in times for all days



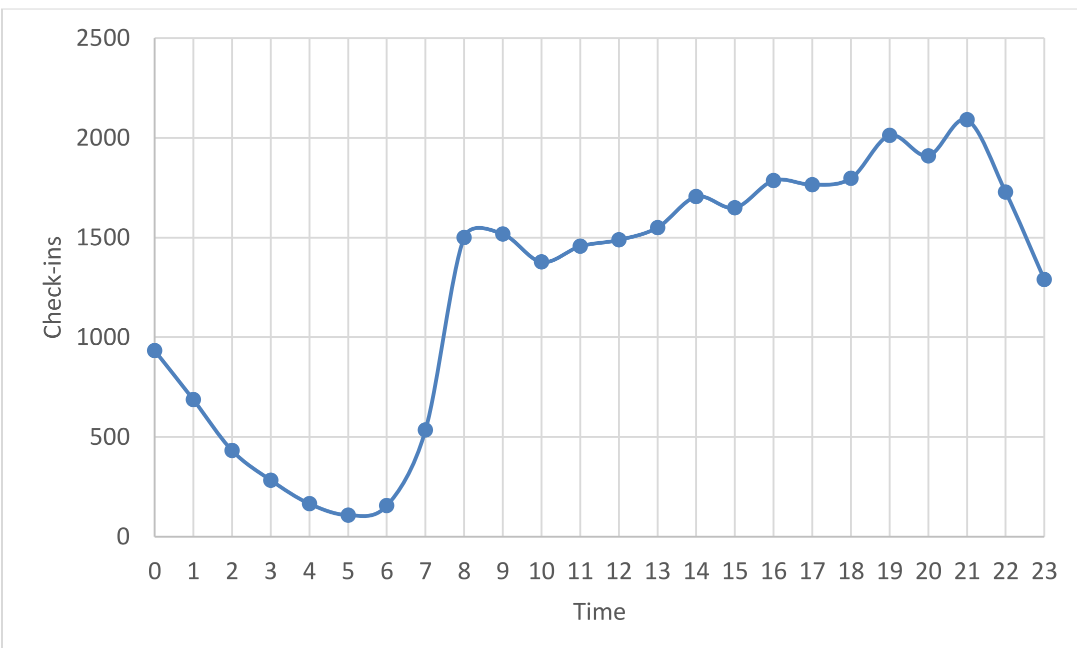
2. Individuals check-in times on Monday



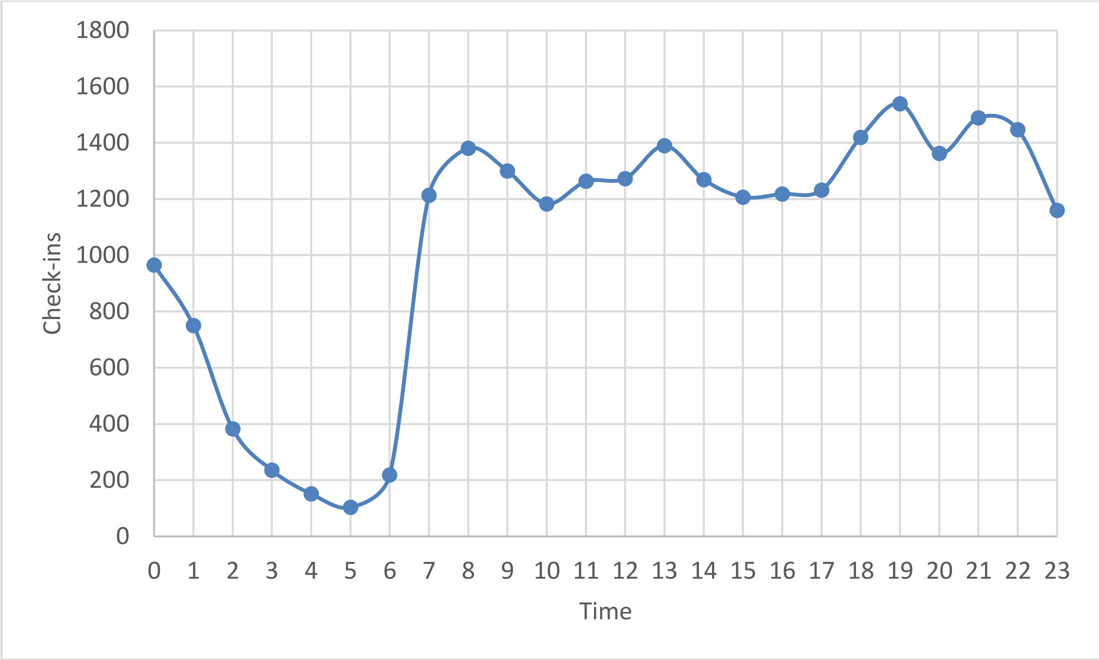
3. Individuals check-in times on Tuesday



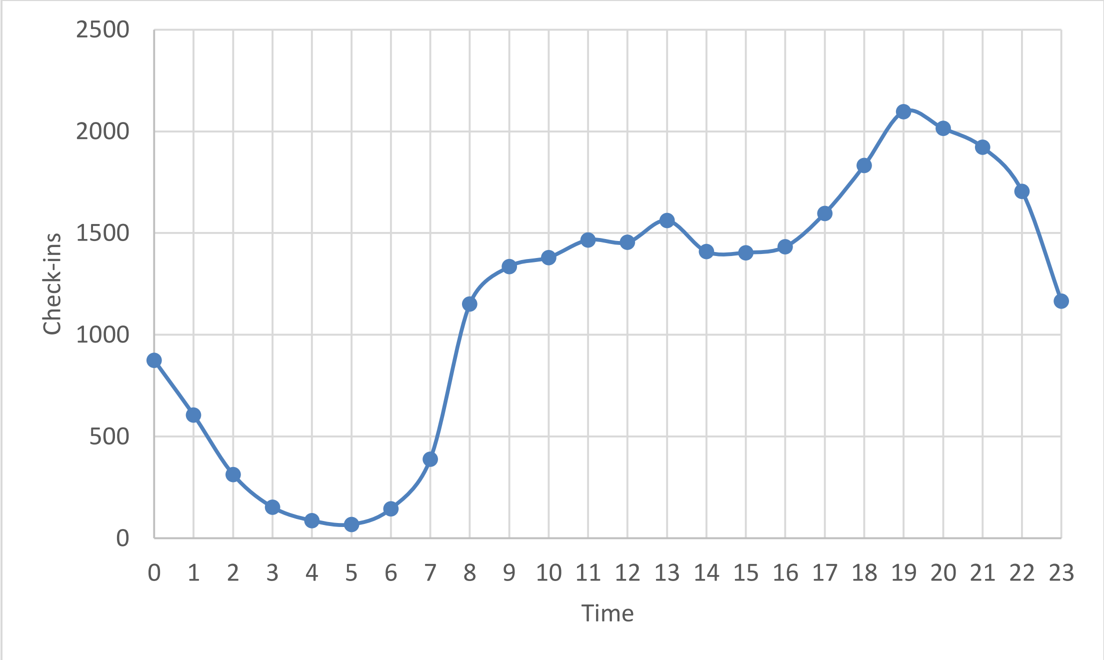
4. Individuals check-in times on Wednesday



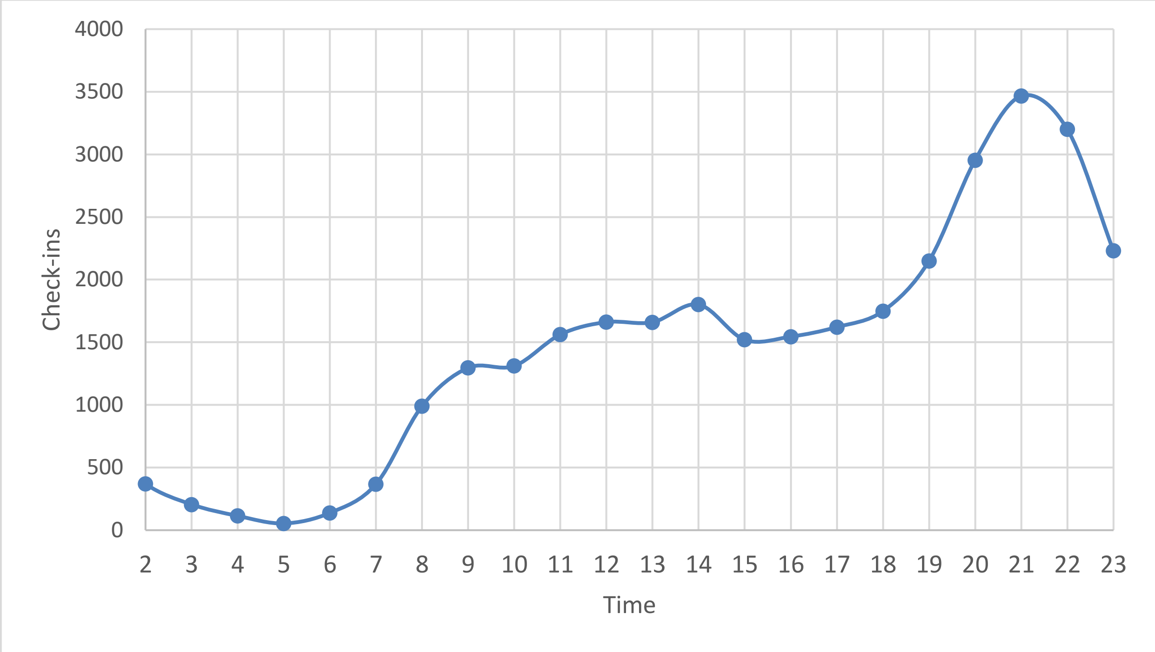
5. Individuals check-in times on Thursday



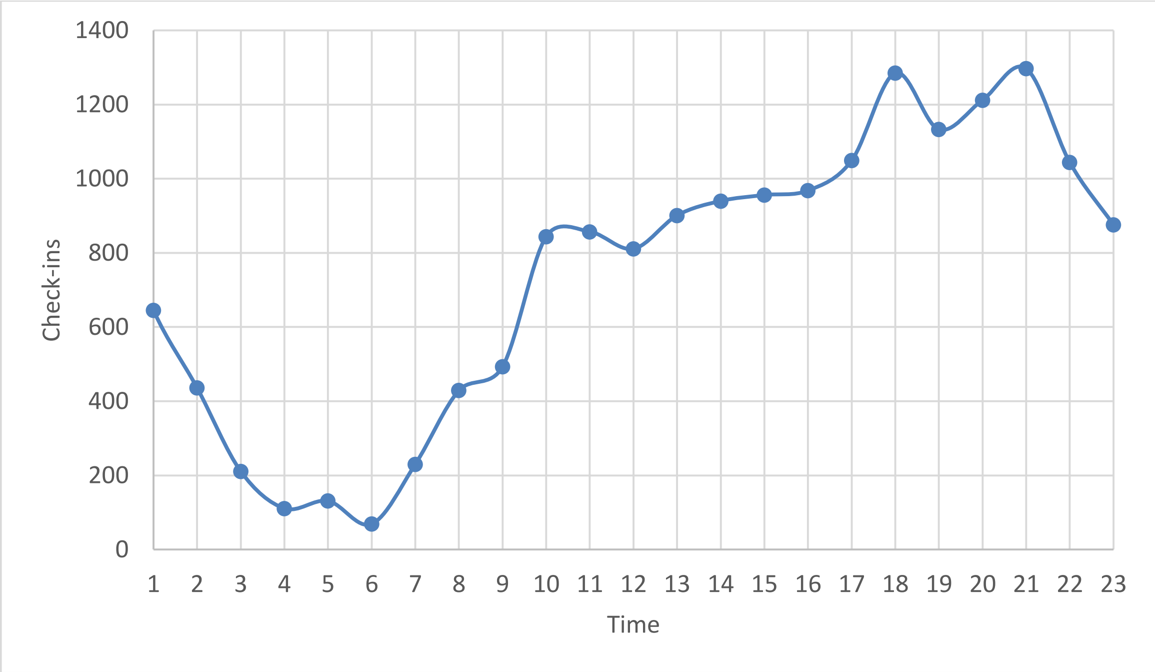
6. Individuals check-in times on Friday



7. Individuals check-in times on Saturday



8. Individuals check-in times on Sunday



9. Times of individual check-in comparison on Monday to Sunday