LAB University of Applied Sciences

Providing the marketing mix through the analysis of historical data. Case: Divvy

LAB University of Applied Sciences
Bachelor of Business Administration, International Business
2021
Hai Hoang Nguyen

Abstract

Author(s)	Publication type	Completion year
Nguyen Hoang Hai	Thesis, UAS	2021
	Number of pages	
	60	

Title of the thesis

Providing the marketing mix through the analysis of historical data: the case company Divvy

Degree

International Business

Name, title and organisation of the client

Abstract

There is no denying the impact that data has on the functioning of society and business today. Data provides insights to suppliers, especially through historical data. Therefore, through this thesis, it will clearly analyze the historical data of Divvy, which is the case company providing the bike-sharing program, thereby giving the best marketing mix strategies for the organization.

The main theoretical framework illustrated some conceptions relating to marketing and database. The analysis used quantitative methods along with secondary data which is made publicly available by the company on its website, thanks to the accurately utilization case company data, hence, presenting the most effective marketing and operational strategies for the company.

The principal findings of the thesis will refer to a more powerful strategy through the customer journey history records of using Divvy's bike-sharing program service.

Keywords

Marketing mix, strategy, data analysis, historical data, data cleaning and visualization

Contents

1 Introduction

1.1 Background and research problem

Data consciousness has spread throughout economic, political, vocational, social, and personal life. Every organization and individual faces the difficulty of making sense of the fantastic potential given by such a wealth of data. The journey begins with a basic grasp of what data is, where it originates from, what insights it may provide, and how to acquire it. (Ahlemeyer-Stubbe and Coleman 2018,1.)

The potential of big data and analytics brings unexpected opportunities for companies. In other words, big data can assist understanding customer needs and retaining the customers. Over and above, it also encourages enterprises to optimize the company's processes, operations, and management. (Ahlemeyer-Stubbe and Coleman 2018,3.)

The massive amount of data available presently has the potential to radically alter businesses and their business strategies. The increasing relevance of data presents a significant challenge to businesses in terms of intelligent processing, interpretation, and application of this rising volume of data. To obtain significant business insights and become genuinely data-driven businesses, institutions require new technologies and capabilities. (BearingPoint 2021.)

On the other hand, undeniably influenced by technology and machines, enormous amounts of data are potentially generated through any source in a short period of time (Moshirpour et al. 2017, 1). It follows that if a company or organization wishes to better understand its market, to grow sales, or to attract more new customers, data can aid by giving the notable insights (Marr 2017).

Divvy, the case study company, operates its bike sharing system in the Chicago city. Divvy services provide for residents and visitors with a quick, enjoyable, and cost-effective transit option for riding around and visiting Chicago. Divvy’s bikes have a solid and simple design, ensuring safety for users. The bikes can be easily unlocked and stored at different stations, providing flexibility for users. People utilize bike sharing to commute to and from work or to move around in Chicago on private affairs. (Divvy 2021a.)

A bike-sharing program brings tourism value to destinations. It provides experiences and convenient mobility for tourists, who can explore the tourist destinations according to their preferences. (Bryce 2021.)

Moreover, the emergence of bike-sharing systems helps to create a green environment, to provide flexibility of movement, to promote health as well as to cut down expenses, especially for low-income households. In addition to the "indiscriminate" use of personal vehicles, the bike-sharing program also has the optimal definition of "need to use". (Nikitas 2016.)

This study provides marketing strategies for bike sharing programs, using the Divvy data assessment as a case study. This is why, in order to find company insights, it is necessary to dive into the historical operated data of Divvy users and distinguish the different needs of each type of customer. Following that, this research aims to create comprehensive marketing strategies for the Divvy program.

1.2 Research background

On the data science community site Kaggle, a data analyst named Ed Garcia founded this case study and using the R programming application for the work, also, the author conducted some research and analysis on public data from the Divvy company. The author has made some suggestions for resolving the business problem, namely, to assist in the development of strategic marketing to convert regular users into annual users. Additionally, the author used the Divvy public data source to examine the data from May 2020 to April 2021. (Garcia 2021.)

This thesis focuses on and reinforces some of the marketing strategies that the above author lacks in order to develop new research. Some processes of presenting data content are learned from previous author. Furthermore, the dataset's time period is refreshed, starting in January 2021 and ending in September 2021, based on the time of writing this research paper, because this thesis was written in October 2021. In addition, this thesis takes a detailed look at a Divvy company's marketing strategy by adopting the 7Ps marketing framework. It then enables Divvy for not only completing a small task, but also in providing a vision for the future development and expansion of the Divvy service.

1.3 Research objectives

The objective of this research paper is to provide historical data that gives valuable insights to support a well-oriented and effective marketing strategy. Moreover, the thesis also delivers the framework of marketing mix strategy based on its basic concepts.

1.4 Research question

The leading research question of this thesis is about “How can historical data help in providing a marketing strategy for Divvy bikes? “.

Based on the main research question, three sub-questions are formulated. Their purpose is to help to create a thorough view on the thesis topic:

- How can author ensure data quality?
- How can useful information be acquired by taking advantages of historical data?
- What are the results obtained for the marketing strategy?

1.5 Delimitations

First, this research paper focuses mainly on building marketing strategies based on 7Ps of marketing and SWOT analysis, by applying data analysis tools to find insights for the organization based on historical data. However, this thesis does not provide all aspects relevant to marketing. Following that, the theoretical framework is explained in Chapter 3,4 and 5. Furthermore, the empirical research of this study is based on the case company Divvy and the data is limited to the period from January to September 2021. The data collection is described in Chapter 2.

1.6 Structure of the study

The thesis is divided into eight sections:

Firstly, Chapter 1 Chapter 1 introduces the background and research problem of the thesis, the delimitations, the research questions, and methodology. This chapter provides a comprehensive overview of the thesis.

Secondly, in Chapter 2, it describes the analytical method of the study as well as introduces some tools for the analysis.

Thirdly, Chapter 3, 4 and 5 discusses the theoretical definition of marketing and its variety, database conceptions and the kinds of analytics, the implementation of R programming and Tableau as well as some information about Divvy's bike-sharing program to better understand the case company.

In addition, Chapter 6 introduces in detail what a bike-sharing program is and presents the Divvy company and its operations in Chicago.

Besides, Chapter 7 describes all the work progress relating to data manipulation and the properly use of software.

Finally, Chapters 8,9 and 10 summarize the main findings and gives recommendations as well as delivers the discussion and conclusion about the study, whether it was successful or not.

2 Research methodology

2.1 Research method

There are two common research methods that are employed by researchers, including quantitative and qualitative. Because the data utilized to answer the research question is fully in the form of complicated datasets and to demonstrate the impact of historical data on bringing business strategies marketing, **quantitative approach** is used in this study. Additionally, current study only points out on some specific fields of data analytics such as data cleaning, data analysis and data visualization. Therefore, there is no data science in terms of artificial intelligence or machine learning.

Quantitative data is information mostly related to numbers and quantifiable, while for qualitative data it will not be quantifiable, such as interviews. Analyses and statistical interferences are used to process the figures. (Saunders et al. 2016, 569.) The term quantitative research method is frequently used interchangeably with any data gathering tool (such as a questionnaires) or data analysis approach (such as charts and graphs) that employs numerical data. (Saunders et al. 2016,165.)

2.2 Data collection method

Secondary data is collected by other users for purposes other than the original investigation. In the same way, some typical secondary data sources such as government documents, published data or statistical releases. (Sekaran & Bougie 2016, 37.)

In this analysis, the data collected is mainly secondary data from the raw company data. Many of the secondary data sets that are now accessible were once primary data sets that were recombined with other data sets to form bigger data sets (Saunders et al. 2016, 318).

For the purpose of the research paper, and at the same time bringing marketing strategies to case companies like Divvy, because of that, Divvy's use of publicly available data is completely practicable and most effective for the results of this study. Divvy big data is managed by Motivate International Inc. Besides, the data is made available publicly by the owner, legally licensed and easily accessible.

Data License Agreement

Motivate International Inc. (“Motivate”) operates the City of Chicago’s (“City”) Divvy bicycle sharing service. Motivate and the City are committed to supporting bicycling as an alternative transportation option. As part of that commitment, the City permits Motivate to make certain Divvy system data owned by the City (“Data”) available to the public, subject to the terms and conditions of this License Agreement (“Agreement”). By accessing or using any of the Data, you agree to all of the terms and conditions of this Agreement.

1. **License.** Motivate hereby grants to you a non-exclusive, royalty-free, limited, perpetual license to access, reproduce, analyze, copy, modify, distribute in your product or service and use the Data for any lawful purpose (“License”).

Figure 1. Data License Agreement's Divvy (Divvy 2021b)

According to the data licence agreement (Figure 1), Divvy shares data that can be used for any lawful purposes such as modification, access and product analysis needs.

In addition, the secondary data is evaluated through four criteria, which are data period, data correctness, data relevancy, and data pricing (Sekaran & Bougie 2016, 38).

- Timeliness of the data: This secondary data is the most updated version, having been collected between January 2021 and the end of September 2021. Because this thesis is done in October 2021, the data is the most recent period that this research was written.
- Accuracy of the data: This data is obtained from the Divvy company's official website; for that reason, the accuracy and guarantee of this data is entirely perfect for analysis.
- Relevance of the data: Because the purpose of the thesis is to analyse data so as to uplift marketing strategy, using company data is the most relevant to the thesis's outcome.
- Costs of the data: This data is completely free and can be found on the company's website. There is no need to spend any budget on data collection.

2.3 Data analysis tools

With the intention of turning raw data into actionable insights for the case company, this thesis focuses on using two main tools, which are R programming and Tableau.

2.3.1 R programming

Definition

The R Core Team and the R Foundation for Statistical Computing support R as a programming language and open-source framework for numerical computing and its graphics. In general, statisticians and data miners frequently utilize it.

Basic statistical tests, classification and analysis of parts of time are computational and statistical procedures that R can do. (R 2021a.)

Figure 2. R Programming (i2tutorials 2020)

Major functions of R programming

R gives users an absolutely amazing experience with simple operations but can produce high quality images and graphics, in addition, it can be combined with a number of mathematical indicators. (R 2021b.)

R is an open source, so it is extremely economical for users. At the same time, R is very suitable for simulating data through different types of tables and graphs. Thanks to a community of developers and R currently has over 15000 packages in every field of study. R also has a plethora of mathematical approaches, such as fundamental statistics, statistical distribution, and graph practice. It is clear that R is unique in that it is not a general-purpose programming language. It performs some excellent functions, primarily statistical analysis and data visualization. (Miller 2021.)

The R environment

R also has its own text format similar to LaTeX, which provides comprehensive documentation and can be changed in different formats. Although R software is used primarily by statisticians and other users who require a statistical computing and software development environment, it can also be used as a general matrix calculation tool with measurement results.

R also has its own text format, akin to LaTeX, that provides extensive documentation and may be altered in several formats. Nevertheless, statisticians and other users that require a statistical computing and software development environment generally utilize R software, it may also be used as a general matrix computation tool with measurement results. (R 2021b.)

2.3.2 Tableau

Definition

Tableau is an advanced analytic system that is changing the way people use techniques to address issues by empowering businesses and individuals to leverage the most use of the data.

Figure 3. Tableau (Tableau 2021a)

Tableau assists individuals and businesses in becoming more data driven. Tableau is an enterprise data management solution; customers do not have to know how to handle data in graphical form to acquire the data users need from the data source. Once there is a change in the original data, the data analysis process is quick, continuous, and updated on a regular basis. Additionally, Tableau caters to all types of users, allowing them to easily create a great data-driven analytics environment that is attractively displayed and intuitive. Tableau's aim is to help people see and understand numbers, which is why its products are created with the user in mind—whether they are an analyst, data scientist, student, teacher, CEO, or business user. (Tableau 2021a.)

The features of Tableau

There are some special characteristics that many analysts prefer to choose Tableau for their work.

At the start, Tableau allows audiences to create simple queries. Analysis speed is always assured when using an in-memory technology platform, regardless of how massive the enterprise's data is. The main thing is to make interactive dashboards on charts, parameterize and interpret data, and this enables users to create thorough reports, to alter and to add extra information without requiring complex understanding of developing analytics.

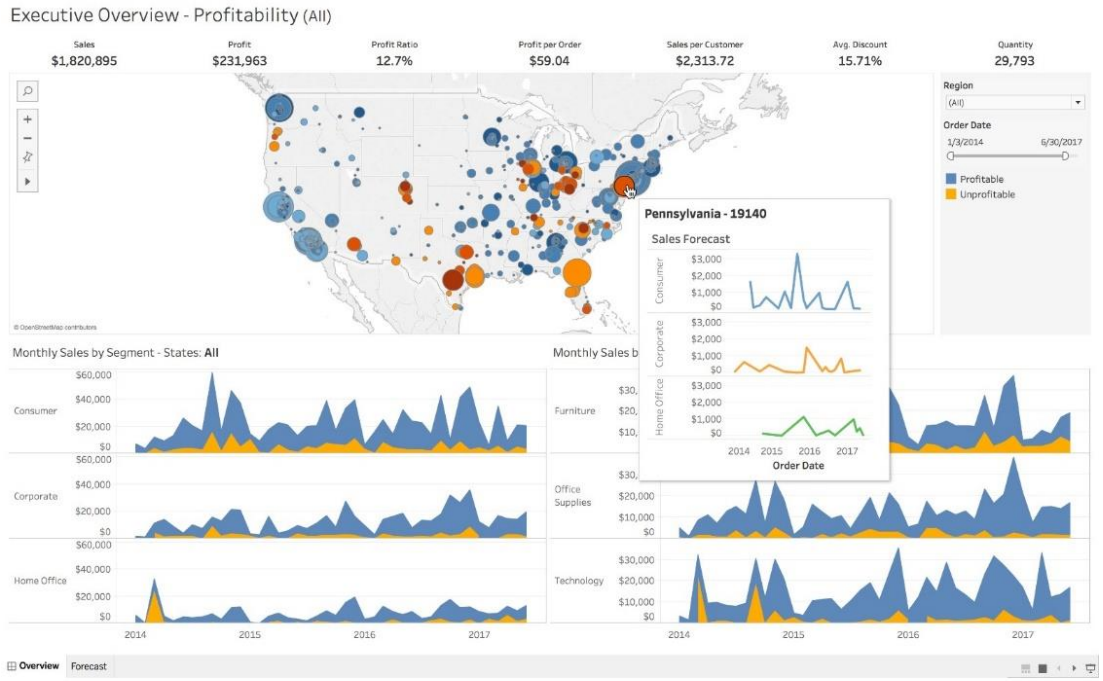

Figure 4. Example of overall data analytics (Tableau 2021b)

Secondly, Tableau delivers nearly every customer need, with advanced analytics capabilities that help users get insights faster and access to machine learning, statistics, data prep.

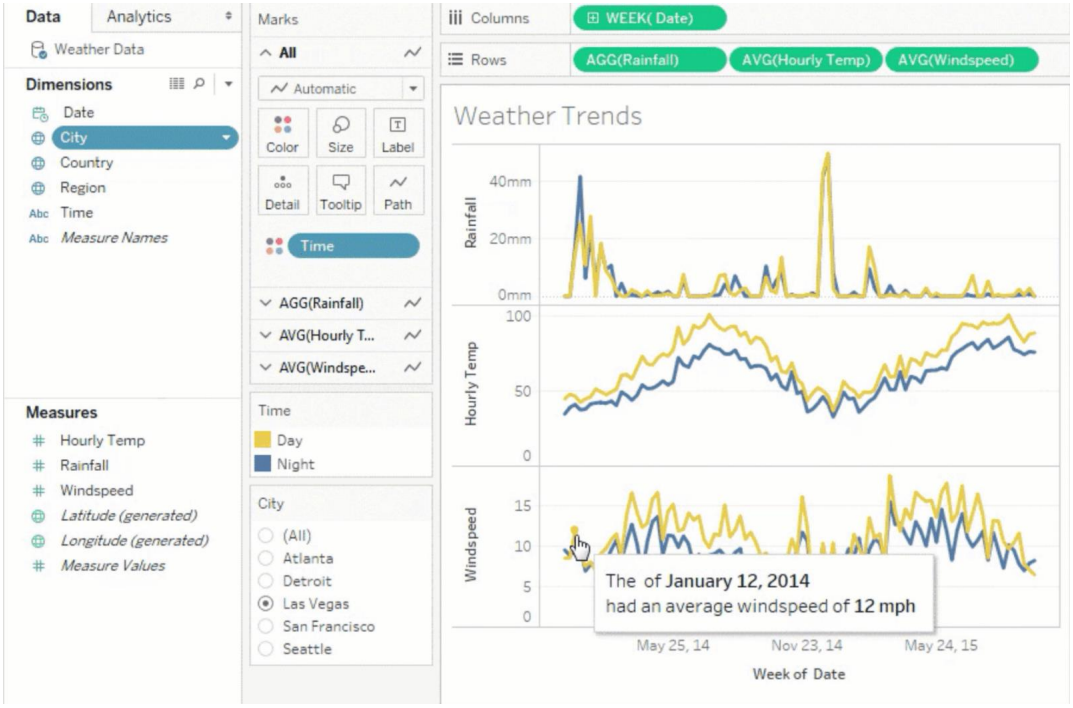

Figure 5. Operational interface of Tableau (Tableau 2021b)

Finally, thanks to Tableau's good self-assessment and analysis, it makes it easy for users to learn and analyse data in the most efficient and fastest way. (Tableau 2021b.)

2.4 Case study

In case studies, the case is the person, group, organization, event, or circumstance that the researcher is interested in. A case study is based on the notion that aiming to get a clear image of an issue, one must analyse the real-life scenario from many aspects and view-points utilizing multiple data collecting methods. Along these lines, a case study may be defined as a research technique that entails an empirical analysis of a specific current phe-nomena within its non-fictional setting utilizing different data collecting methods. (Yin 2009, according to Sekaran & Bougie 2016, 98.)

In this thesis, the case study is going to perform some analysis about Divvy, which is Chi-cago's bike share program. Today, with the development of technology as well as urban modernization, it has supported the upgrading of public transport or transportation systems, making things simpler and more convenient.

However, the population growth combined with the individual needs of each person, makes the use of personal cars more and more, which has a devastating effect on the environment. Public transportation contributes to the sustainability of a city by improving air quality, elim-inating the need for many trips by personal vehicle in dense metropolitan areas, and limiting the use of separate vehicles to help prevent the risk of environmental pollution. Further-more, it also supports the citizen health improvement and stress avoid, which is caused by some problems such as traffic jam, bad weather or noise pollution. (Kathy 2020.)

Correspondingly, the emergence of bike-sharing systems helps to bring a green environ-ment, flexibility of movement, promote health as well as cut down expenses, especially for low-income households. In addition to the "indiscriminate" use of personal vehicles, the bike-sharing program also has the optimal definition of "need to use". It makes people feel bet-ter. (Nikitas 2016.)

3 Marketing

3.1 Definition

As claimed by American Marketing Association (2017), providing value to customers through different communication methods and maintaining a good relationship for both implementers and stakeholders is known as marketing.

In addition, The World Marketing Association has mentioned creating immeasurable value for stakeholders through understanding and discovering the process of meeting organizational and individual needs (Russell 2017, 12).

As stated in The Chartered Institute of Marketing (2015), marketing means the procedure for management in charge of finding, predicting, and financially satisfying requirements of the customers.

3.2 7P’s of marketing

A formula that every business must constantly test and review to ensure that their business is on schedule and achieving the best results for its market share, that is the 7Ps in marketing mix. (Hanlon 2021.)

3.2.1 Product

The product is anything that can be sold or used by another person. It covers the product's quality, the components used, the colour, size, or scent, as well as any physical aspects of the product. Product 'P' refers to everything involved in the creation of a tangible or immaterial object. (Russell 2017, 14.) When an organization or group of people provides a physical condition or service to fulfil what is missing from the user, it is called a product (Twin 2021).

For the service marketing mix, some products and services, such as in the education and tourism industries, are considered intangible products, which can be used by individuals and of high diversity. (Bhasin 2021.)

3.2.2 Price

The price element is the cost that customers must pay to use the supplier's product or service. Because the remaining components will build the value of the product, the pricing part of the 7P’s of the marketing mix model is distinct from the others. Price will deliver profit to the firm. Furthermore, the pricing is affected by external factors such as the product's market

share and the customer's perception of the brand. (Twin 2021.) Price not only covers the expenditures of goods and profit, but it does also accomplish a lot. The desire for high or low quality is also considered through the price of a product, which serves as a metric to classify high and low spending customers. Finally, it tells how the product should be consumed. (Russell 2017, 14.)

In the service marketing mix, pricing services is a little more complex than pricing products. Regarding to restaurant perspective, business usually can charge people for its foods. As a result, these factors must be considered while calculating costs. (Bhasin 2021.)

3.2.3 Place

The distribution system is also another important aspect of the marketing mix. The thing is to consider building a reasonable product and service supply system, which can make it convenient for customers to consume and use. For example, the supply of goods in city centers or supermarkets. (Russell 2017, 15.) Most corporations' ultimate focus is to always endeavour to deliver products to clients at the right place and right time. In any physical store, as well as virtual businesses on the Internet, can serve as a distribution channel as long as it benefits consumers reach the product. Delivering products at the location and time required by the consumer is one of the most critical parts of any marketing strategy that firms strive towards. (Twin 2021.)

In the context of marketing services, specify where the service product will be positioned. On the highway or in the town, for example, is the finest site to open a gas station. (Bhasin 2021.)

3.2.4 Promotion

In marketing mix, promotion refers to any method of promoting the product. It has many different forms of marketing like public relations or any form of direct selling. Profit is another 'P' that is frequently discussed. All commercial enterprises exist to generate a profit. Profits allow firms to maintain the operations and establish new enterprises. (Russell 2017, 15.) Sales assistance and product promotion are operations that assure clients may easily acquire, receive items, and form a favourable opinion of the product. Advertising, cataloguing, and public relations are examples of these activities. (Twin 2021.)

From the service perspective, promotions have evolved into an indispensable component of the service marketing mix. Because services are easily imitated, it is usually the brand that distinguishes one from another. (Bhasin 2021.)

3.2.5 People

People conception refers to the employees and salesmen who have responsible for the company. When business give exceptional customer service, then generating a favourable experience for consumers and, hence, the product will be offered. As a result, current clients could spread the experiences about the exceptional service, and the other potential customers may receive it as referrals. (Queensland Government 2021.) In addition to the function of the product brings to purchasers, external factors such as employees providing products and their interactions with customers are also the internal factors that contribute to a good experience for shoppers (The Chartered Institute of Marketing 2015).

People include everybody who comes into contact with the client, even if just indirectly, so make sure that recruiting top talent at all levels—not just in sales and customer service (Mailchimp 2021).

Concerning to marketing mix in services, it is clear that “people” is one of main ingredients of the service marketing mix. For example, in banking sector, the staff in the certain branch and their interactions with customers will identify the image of company. For that reason, people here can make or destroy a company in the matter of service marketing. (Bhasin 2021.)

3.2.6 Process

Processes that overlap with the client experience should be prioritized. The more exact and seamless the procedures are, the more efficiently the personnel will be able to carry them out. When the staff is not preoccupied with following protocols, they have more time to devote to consumers, which translates straight into personalized and great service. (Mailchimp 2021.) This stage entails providing the company's goods or services to the customer. Processes from production to distribution of products need to be managed and quality maintained to improve revenue potential for the company. (Mass Live Media 2021.)

On the basis of service marketing mix, the service process is the method by which a service is provided to the end user. Moving on to the typical model at two excellent companies: McDonald's and FedEx. Both organizations thrive on their speedy service, which is made possible by their trust in their systems. Furthermore, the demand for these services is such that they must provide optimally without sacrificing quality. As a result, the mechanism through which a service provider delivers its product is critical. (Bhasin 2021.)

3.2.7 Physical evidence

Physical evidence is a collection of practical experiences in a Service Marketing environment. This is a collection of physical factors created by humans and nature that have the ability to positively or negatively affect the performance of a service business. Along with it, businesses that provide products and services seek to manage the physical proof of their service since it affects the total consumer experience. (Red Bike Marketing 2021.) Good examples of proof of purchase include new haircuts in the salon and braces in the dental office. Addedly, for transaction receipts, it is like a real proof of the service or product they used or owned. It forms a sense of the value of the purchase. (Mass Live Media 2021.)

In respect of services, the final component of the service marketing mix is crucial. As previously stated, services are intangible in nature. However, in an effort to provide a better client experience, tangible things are included with the service. Consider a restaurant that merely has chairs and tables and provides good cuisine, as opposed to one that has ambient lighting, great music, and a good seating arrangement and also serves good food. (Bhasin 2021.)

3.3 SWOT analysis

SWOT is an acronym that stands for the first letters of the term's strengths, weaknesses, opportunities, and threats. Based on the four criteria listed above, this is a well-known model in business analysis. Businesses may conveniently perceive their goals as well as issues inside and outside the organization by harnessing SWOT theory.

Figure 6. SWOT analysis (Kenton 2021)

Strengths: Strength is the advantage of the business, project or product. These must be the outstanding, unique characteristics that the business is holding when compared to the competition, particularly high technologies for instance.

Weaknesses: Weaknesses include things that the business has not done well. In cases where a company is having difficulty identifying flaws, looking at factors such as bad debt or underperforming human resources are some examples of imperfections.

Opportunities: Opportunities are understood as external influences that bring advantages to the company in competition. In business, opportunity often derives from a number of changes such as a rapidly growing market, a new technology opportunity, or a competitor in a disadvantaged position.

Threats: Besides these latent opportunities, risks and challenges also appear to hinder the development of the company. Therefore, entrepreneurs must properly identify the problem and guard against the negative change of external factors. (Kenton 2021.)

3.4 Service marketing and its features

3.4.1 Service marketing

In accordance with Kotler and Keller (2007), services are usually activities for another party with intangible assets and often use physical or immaterial to fulfil the desires. Service marketing is activities aimed at satisfying people's needs for services through the purchase and sale of services. In order to satisfy the needs of service consumers, service marketing needs to do the task of uncovering the service-related needs of service consumers through market research and develop services precisely to meet the needs of service consumers. On that account, the mission of convincing the end-users is totally high priority. (Prachi 2019.)

3.4.2 The attributes of service marketing

Different from traditional marketing, service marketing requires a number of distinct properties, and it must be accomplished rigorously to achieve good results. The following are some of the essential qualities that distinguish service marketing with traditional marketing:

Figure 7. Characteristics of service marketing (Prachi 2019)

- Intangibility

A service is an intangible object that cannot be seen but should instead be experienced. When purchasing or experiencing a kind of service, consumers' knowledge about the service helps them to make a decision to use the service. Furthermore, clients put in a lot of time and effort when making decisions, especially with respect to services that require substantial investments and risks.

- Use of tangible products

To give adequate service to consumers, service providers use a variety of tangible or real things. The products used to provide services add value to the standard of living of purchasers. For example, travel services often provide a ticket or a product to pass through the station to the customer. Over and above that, for public transport services, the provision of vehicles is also a tangible material.

- Unnecessary stock

These numerous of service marketing are not tangible products and are supplied on-the-spot; consequently, unlike items, no inventory is required in such circumstances. As mentioned above, just relating with traditional forms of marketing, the quality of products and services determines most of the profits of the business. A photographer, in other words, is renowned by their talents, clicks, and portfolio, and there is no need to maintain it stocked.

- Lack of pricing sensitivity

In some ways, today's clients are typically more concerned with the quality of the service than with the cost. The highest earning potential is found in products and services that are free of flaws. It should be pointed out, hence, that the consumer's price sensitivity to the relevant service is dissimilar to the expected performance. The below chart exhibits the sensitivity of price in the marketing:

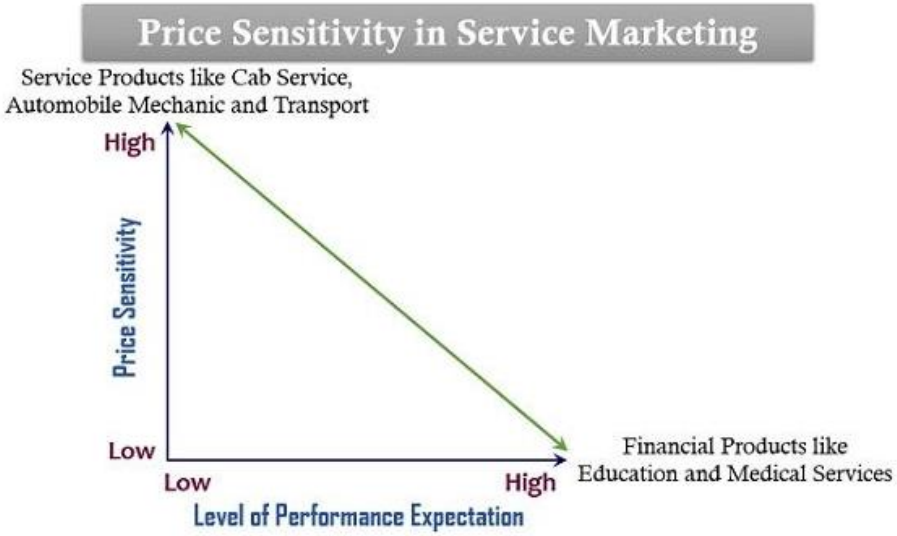

Figure 8. Price sensitivity in service marketing (Prachi 2019)

Some products and services are less price sensitive than, which are in teaching and training, however, some car maintenance services are more price sensitive.

- Process of building up values

Factors related to people, service processes or basic 4P's principles in marketing assist to bring the surplus value to customers, it gives a rise to a combination of marketing and services. Creating services that are flexible, user-friendly, and easy to use are also the ways of making the services becoming the face of the brand. (Prachi 2019)

4 Database

4.1 Data

Data is an unstructured collection of facts, such as numbers, words, images, intended to measure, observe, or simply describe things. Developments in the field of technology, especially in smartphones, have resulted in text, video and audio being included in the data. (Itl Education Solutions Limited 2010, 1.)

Decisions related to a certain action through the processing of collected data. (Migrant & Seasonal Head Start Technical Assistance Center 2006).

4.2 Big data

Big data refers to enormous, varied collections of information that are growing at an exponential rate. The fundamental three Vs of big data include: volume, variety and velocity. Volume represents the amount of information of the data, variety shows the variety that the data brings, and velocity unveils the processing speed of the data.

Big data can be structured (often numerical, readily prepared and saved) or unstructured (more free-form, less quantifiable). Frequently, apart from small data types, big data is contained in large systems and requires some separate software to work with it. (Segal 2021b.)

4.3 Database

A database is an organized collection of data that is usually accessed from a computer system or exists as a file in a database management system. Users can get the desired information based on different tools. (Itl Education Solutions Limited 2010, 2.)

The database can then be easily accessed, modified, updated, and organized. Most databases use structured query languages to write and query data. Database plays a key role in business activities such as ensuring the security of data as well as easy storage in a systematic way. (Lutkevich 2021.)

4.4 Functions of database

Data is kept in databases is used by businesses to achieve its business objectives. The following are some examples of how businesses use databases:

- Enhance business processes: Companies gather information regarding company processes such as sales, order processing, and customer support. The data is studied in order to improve the procedures, expand their business, and increase revenue.
- Keep track of users: Databases frequently store information about individuals, such as users and customers. Social media sites, for example, utilize databases to store user information such as names, email addresses, and usage patterns. The data is used to recommend content to users and meet consumer needs.
- Protect personal health information: In terms of healthcare industry, databases are used by healthcare practitioners to securely store personal health data with a view to inform and ensure better treatment.
- Save personal data: Personal information can also be stored in databases. Individual users, for example, can utilize personal cloud services to archive media such as images. (Lutkevich 2021.)

5 Analytics

Analysis is a method of dividing the whole into parts in order to deepen the perception of those parts. For some areas where large amounts of data are collected, there are many advantages to understanding the techniques for analyzing the underlying data and turning it into useful information for the organization's operations.

Today, with the speed of technology, organizations can collect hundreds of thousands of data within minutes, thereupon, it is very necessary for all industries. Some examples include finding relationships between customer needs and the prices of goods, or the influence of people's lifestyles on disease. (Techopedia 2021.)

5.1 Data analytics

Data analysis is a defined process of collecting, cleaning, modifying as well as dividing data into different arrays to uncover useful information for organizational or business decisions. In short, data analytics are capable of facilitating a corporation improve its performance. (Frankenfield 2021b.)

5.2 Types of data analytics

There are four primary forms of data analytics:

- **Descriptive analytics** is the evaluation of historical data to have a superior understanding about company trends. Descriptive analytics refers to the utilization of historical data to do some reviews. A specific example of descriptive analysis is economic reports, such as annual financial statements of a company's performance, annual profit and sales figures. (Frankenfield 2020a.)
- **Diagnostic analytics** refers to the approaches are using to ask its data about the happening. This is digging deep into the data to get relevant insights. The first step in most firms' data analysis is descriptive analytics, which is a simpler procedure that records the facts of what has already occurred. Diagnostic analytics goes a step further by clarifying the logic behind specific results. (Sisense 2021.)
- **Predictive analytics** is the use of data along with statistical and analytical tools to obtain predictive outcomes of a business in the future. In addition, predictive analytics can use current data to evaluate and find the intrinsic possibilities of events that may occur in the future. Today's predictive analytics, thanks to modern analytical

tools, does not require in-depth knowledge of data, but can still perform the analysis on its own to help build a good strategy for the corporation. (Halton 2021.)

- **Prescriptive analytics** is a step beyond descriptive and predictive analytics pertaining to recommend appropriate actions and to predict possible outcomes instead of dealing with initially available information. This type of modern data analysis requires not only historical data but also external information due to the nature of statistical algorithms. The tools used in this type of analysis are usually algorithms or machine learning. (Segal 2021a.)

6 Bike-sharing program

6.1 Definition

Bicycle sharing system (BSS) is a program by a group or organization to provide public transportation by bicycle or e-bike from various stations located in an area. Each public bike ride is always given a certain amount of time, and the pick-ups or drop-offs can be made flexibly. (European Cyclists' Federation 2021.)

Figure 9. Bike-sharing system (Capo Velo 2017)

6.2 Its operations

The original

Although the origins of public bicycles have been around for a long time in parts of Europe in the 1960s, it was in the early 2000s that bike-sharing campaigns were really popular. The service using bike sharing is combined with modern tools such as mobile phones and e-wallets, so the process of using the service is quite simple and easy to follow. With each different system, there are different payment methods.

Bikes and its stations

The category of bicycles and its stations are the most visible and tangible elements of any bike-sharing program. Public bicycles today mostly have a certain size but with many different shapes and colours. (Martucci 2020.)

6.3 The case company: Divvy

6.3.1 Introduction

Divvy is a shared and operated public bike-sharing system in Chicago, United States. Divvy services provide residents and tourists with a cost-effective way to explore the city of Chicago. Divvy ensures the quality of their bikes by providing constant upgrades, durability, and a plethora of docked stations throughout Chicago. Divvy bicycles also provide a comfortable experience for a variety of customers with varying bike usage abilities and heights. (Divvy 2021a.)

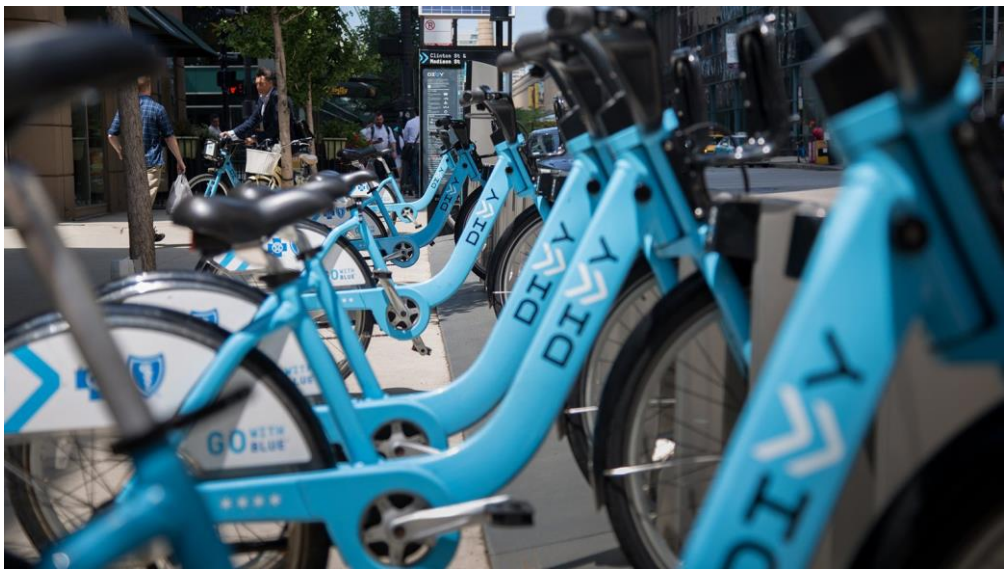

Figure 10. Divvy bikes in a station (Ruppenthal 2019)

6.3.2 Divvy operations

The process to rent a Divvy bike is quite ordinary because it has been integrated through the application on the mobile phone, making it easy for users to connect and to use. It consists of four steps as shown below:

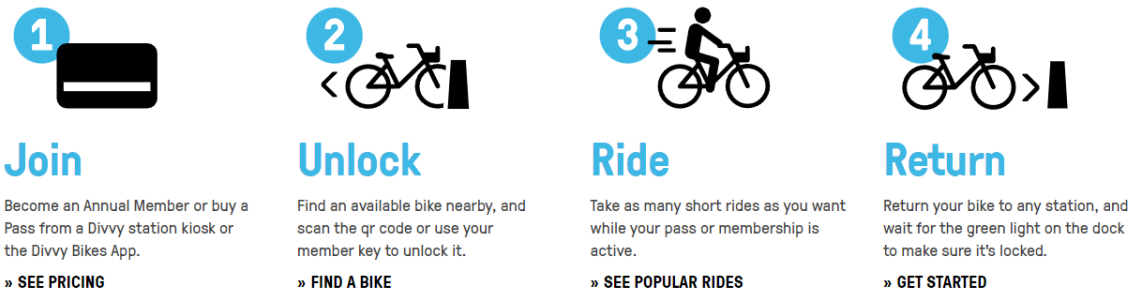

Figure 11. Four steps to use Divvy services (Divvy 2021c)

Firstly, users can control the Divvy public bicycle service by downloading the Divvy app to their phone, which allows them to use bicycle packages for each type of customer and connect to online payment methods available on the Divvy app. Followingly, subscribers might possibly turn on the Divvy app to search for available bike stations, and then find available bikes in the area by scanning the QR code or entering the membership code to unlock the bike. Thirdly, customers are able to participate the Divvy bicycle service as much as individuals want as long as individuals return their bicycles at different stations and wait until the green light turns on to complete the return of the bicycle. However, the number of minutes that the customer is able to cycle the bike, will be determined by the type of package that the customer owns. For each minute that the customer exceeds the time limitation, the customer will be charged according to the price listed in the customer classification below. (Divvy 2021c.)

6.3.3 Category of customers

With a view to provide usage packages for each type of customer, Divvy currently has two main packages for annual members and casual users:

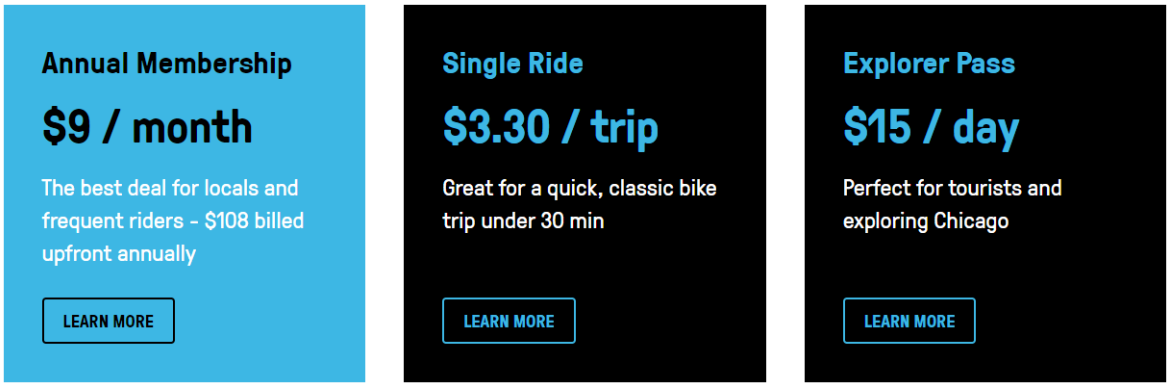

Figure 12. Class of users in Divvy (2021c)

- Annual members: the official annual member have a package that costs 108 USD per year (equivalent to 9 USD per month), and users are capable of running unlimited times within the first 45 minutes (Divvy 2021d).
- Casual members: casual clients have two smaller bundles, which are Single Ride and Explorer Pass.

In terms of Single Ride, starting cost for each trip within 30 minutes is 3.30 dollars (Divvy 2021e). On the subject of Day Pass package, riders have to pay 15 dollars for a day trip with 24-hour free ride, but each trip is limited to three hours, and it need to be returned

before 3-hour trip if client does not expect to pay an extra fee. This ticket is appropriate for tourists visiting nearby landmarks or cycling along Chicago's lakefront. (Divvy 2021f.) If riders wish to travel for a greater duration of time than the time allotted for each type of membership listed above, they must pay an additional fee of 0.15 dollars per minute (Divvy 2021d).

6.3.4 Group of bikes

Divvy offers two types of vehicles: traditional bikes and electric bikes.

Classic bicycles are designed by Divvy to be flexible and durable for users, streamlining for riders to enjoy the travel experience while in Chicago. (Divvy 2021h.)

Figure 13. Classic bicycle (Divvy 2021g)

Divvy's e-bikes are powered by a pedal motor to aid boost the user's travel pace, allowing customers to enjoy Chicago sightseeing with minimal effort. With top speeds of up to 20 miles per hour, these one-of-a-kind pedal-assist e-bikes allow subscribers to conquer bridges, cutting the commute time, and exploring more neighbourhoods in less time. Everything is to simply start pedalling, and the power begins to flow. (Divvy 2021i.)

Figure 14. Electric bike (Divvy 2021g)

Similarly, ensuring that Divvy vehicles are managed and operated within an authorized area, Divvy parking at its stations is free; however, users must pay an additional 2 dollars if parking at public stations. In addition to the violation of the parking location, the Divvy-er will be fined 25 dollars.

In addition, the cost of using a bike for casual users will be 3.30 dollars plus 0.20 dollars per minute, while annual members pay 0.15 dollars per minute for unlocking an electric bicycle. Before per-minute pricing is implemented, the e-bike trips begin with included minutes at no additional charge in specific zones of the service region. (Divvy 2021i.)

7 Research process

7.1 Retrieving data from the company website

This research process is set up to make it easier for readers to understand the analysis process.

Firstly, downloading the data from the official website of Divvy company:

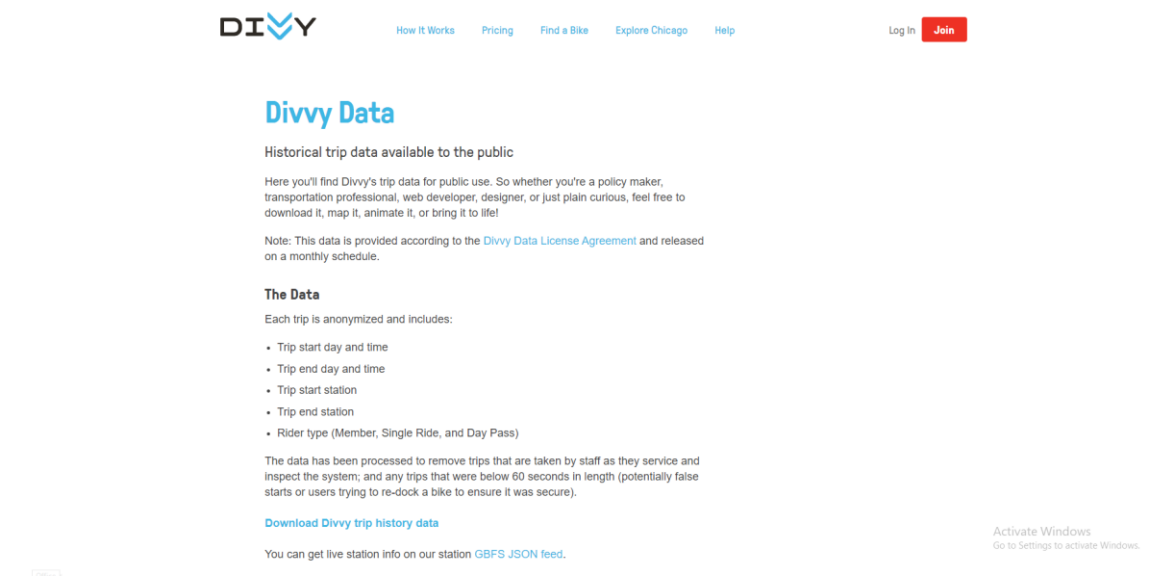

Figure 15. Divvy website

Clicking on “Download Divvy trip history data “and it turns to the main data storage page:

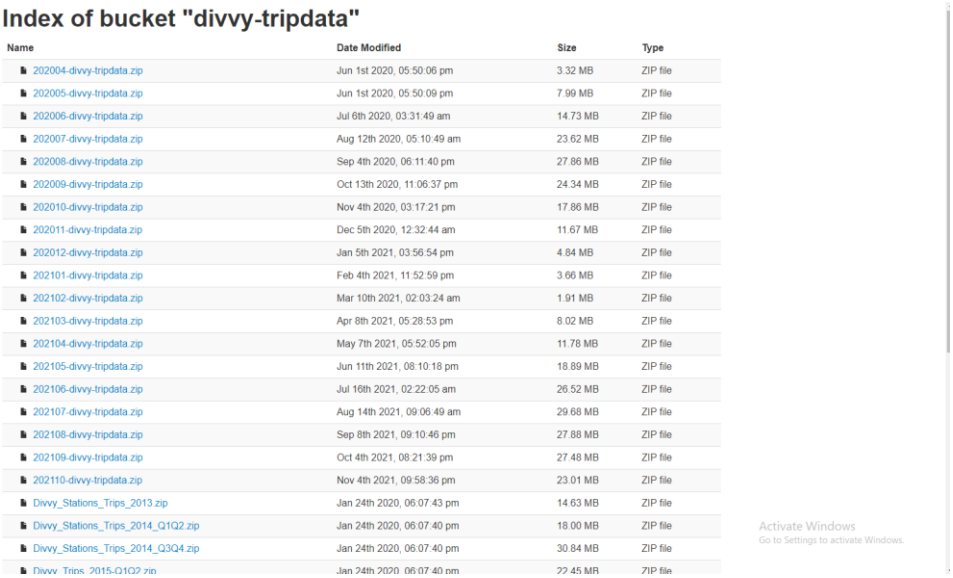

Figure 16. Divvy datasets

Downloading the data from the beginning of January 2021 to the end of September 2021 because this thesis evaluate the dataset in this period.

These datasets name as below with the type of data is Microsoft Excel Comma Separated Values:

- 202109-divvy-tripdata
- 202108-divvy-tripdata
- 202107-divvy-tripdata
- 202106-divvy-tripdata
- 202105-divvy-tripdata
- 202104-divvy-tripdata
- 202103-divvy-tripdata
- 202102-divvy-tripdata
- 202101-divvy-tripdata

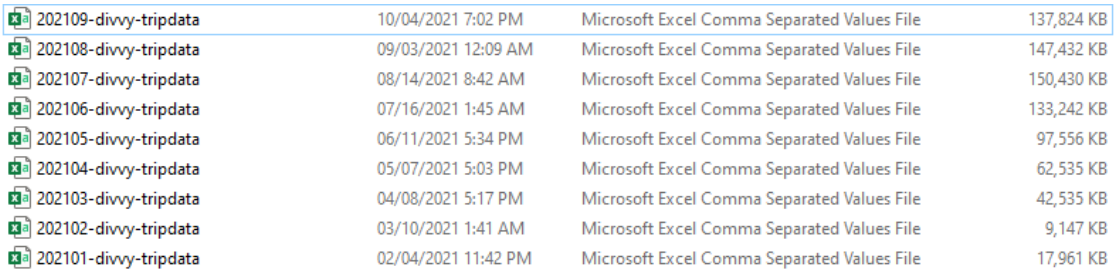

Figure 17. Targeted files for analysis from January 2021 to September 2021

In the description of original data, it has 15 columns with many data types that are explained below:

Column Name	Data Type	Description
ride_id	STRING	Unique Ride ID Number
rideable_type	STRING	Bike Type (Classic Bike, Docked Bike, Electric Bike)
started_at	TIMESTAMP	Ride Start Date (YYYY-MM-DD HH:MM:SS UTC)
ended_at	TIMESTAMP	Ride End Date (YYYY-MM-DD HH:MM:SS UTC)
start_station_name	STRING	Ride's Starting Station Name
start_station_id	STRING (Data from January - September 2021, INTEGER)	Starting Station ID Number
end_station_name	STRING	Ride's Ending Station Name
end_station_id	STRING (Data from January - September 2021, INTEGER)	Ending Station ID Number
start_lat	FLOAT	Starting Station Coordinates (Latitude)

Figure 18. Data types 1

Column Name	Data Type	Description
start_lng	FLOAT	Starting Station Coordinates (Longitude)
end_lat	FLOAT	Ending Station Coordinates (Latitude)
end_lng	FLOAT	Ending Station Coordinates (Longitude)
member_casual	STRING	User Type (Annual Member or Casual User)
ride_length	TIME	The numbers of times that user rided
day_of_week	INTEGER	The day that the customer used services , sorted by number from 1 to 7

P/S: Day_of_week explanation:

Day_of_week	Number
Monday	2
Tuesday	3
Wednesday	4
Thursday	5
Friday	6
Saturday	7
Sunday	1

Figure 19. Data types 2

7.2 Data wrangling, data cleaning with R programming

In order to process the data with R programming, it is necessary to download RStudio Desktop from the main website and it is no charge:

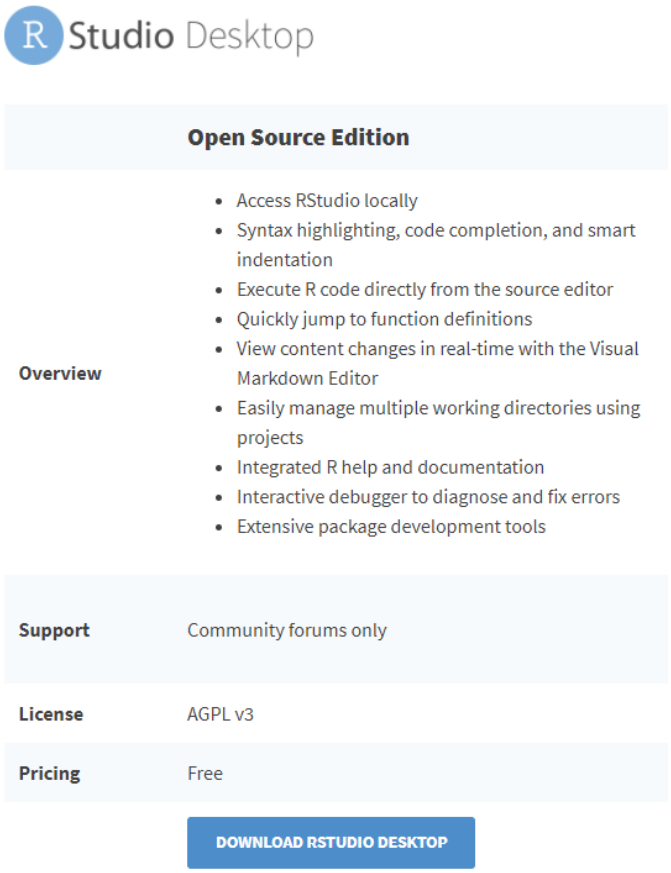

Figure 20. RStudio Desktop download

Once Rstudio has been downloaded, users can run it and prepare to perform data transformation and analysis. For R software, first to be able to implement different applications, users must understand that it needs to have packages that are suitable for the user's needs.

In the case of data selection and analysis as above, the author uses two main packages, namely "tidyverse" package and "lubridate" package. The "tidyverse" package helps users to easily interact with data in adding, editing, changing, visualizing or removing unwanted data for their needs. (Vidhya 2019.)

The " lubridate " package is used to help users work with date and month data types. In addition, " lubridate " package is now also integrated in "tidyverse", so almost just need to download " tidyverse" and users can manipulate almost anything with data. (Spinu et al. 2021.)

Moving on to the main steps, firstly, installing and loading the packages with the specific codes:

- install.packages(“tidyverse”)

- install.packages(“lubridate”)

Nextly, using library() code for activating all the libraries are included in packages that user has:

- library(tidyverse)
- library(ggplot2)
- library(readr)

Because data and directory are not linked yet. Due to the fact, it is necessary to have an easy operation to link the Rstudio operation to where the specified data files are located. So use getwd() to set working directory:

- getwd()

```
getwd()
```

```
[1] "D:/R thesis file/thesisDivvy"
```

Figure 21. Setting up the main directory

From now, the user can read the 9-file datasets that there are already downloaded before by using read_csv(“name of the file. file type”) code and using “ <- “ for naming the data loaded into Rstudio environment:

- t1_2021 <- read_csv(‘202101-divvy-tripdata.csv’)
- t2_2021 <- read_csv(‘202102-divvy-tripdata.csv’)
- t3_2021 <- read_csv(‘202103-divvy-tripdata.csv’)
- t4_2021 <- read_csv(‘202104-divvy-tripdata.csv’)
- t5_2021 <- read_csv(‘202105-divvy-tripdata.csv’)
- t6_2021 <- read_csv(‘202106-divvy-tripdata.csv’)
- t7_2021 <- read_csv(‘202107-divvy-tripdata.csv’)
- t8_2021 <- read_csv(‘202108-divvy-tripdata.csv’)
- t9_2021 <- read_csv(‘202109-divvy-tripdata.csv’)

After that, when the data is all loaded in the R environment, user can see it look like this one on the right corner of the Rstudio environment:

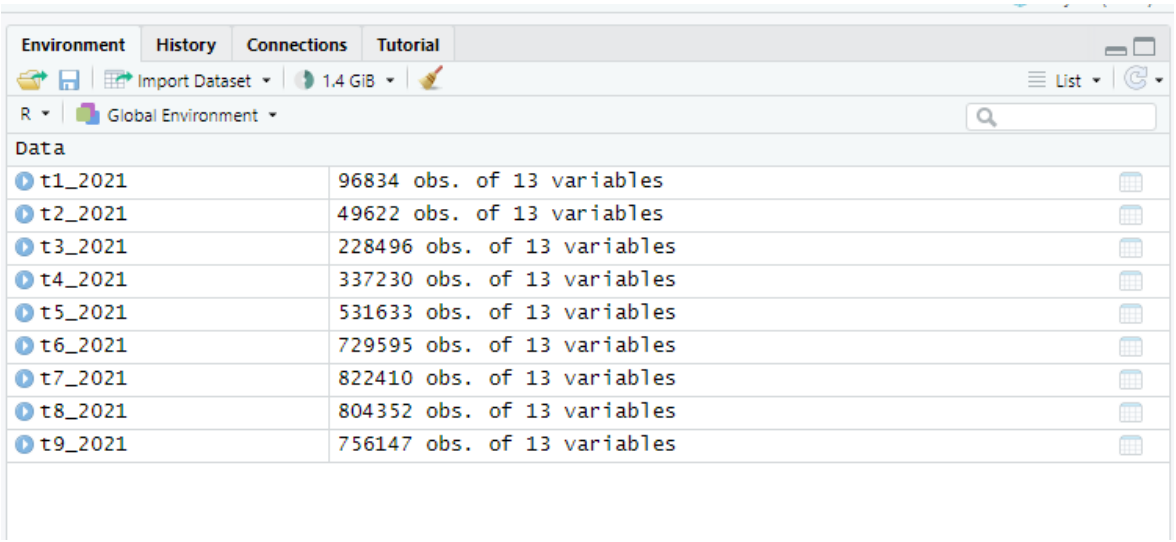

Figure 22. Loaded data in Rstudio

Now that everything is installed and the data is uploaded to Rstudio, the analysis begins.

First, when looking at the first few columns like rideable_type and ride_id both are in STRING format, so to make it easier to count and sort the values inside, we use the mutate() code. This code has the function to help transform and create a new value through an existing value:

- t1_2021 <- mutate(t1_2021,ride_id = as.character(ride_id),rideable_type = as.character(rideable_type))
- t2_2021 <- mutate(t2_2021,ride_id = as.character(ride_id),rideable_type = as.character(rideable_type))
- t3_2021 <- mutate(t3_2021,ride_id = as.character(ride_id),rideable_type = as.character(rideable_type))
- t4_2021 <- mutate(t4_2021,ride_id = as.character(ride_id),rideable_type = as.character(rideable_type))
- t5_2021 <- mutate(t5_2021,ride_id = as.character(ride_id),rideable_type = as.character(rideable_type))
- t6_2021 <- mutate(t6_2021,ride_id = as.character(ride_id),rideable_type = as.character(rideable_type))

- t7_2021 <- mutate(t7_2021,ride_id = as.character(ride_id),rideable_type = as.character(rideable_type))
- t8_2021 <- mutate(t8_2021,ride_id = as.character(ride_id),rideable_type = as.character(rideable_type))
- t9_2021 <- mutate(t9_2021,ride_id = as.character(ride_id),rideable_type = as.character(rideable_type))

After using mutate(), user can check its types again by using summary() code and applying the first dataset names “t1_2021” for example:

- summary(t1_2021)

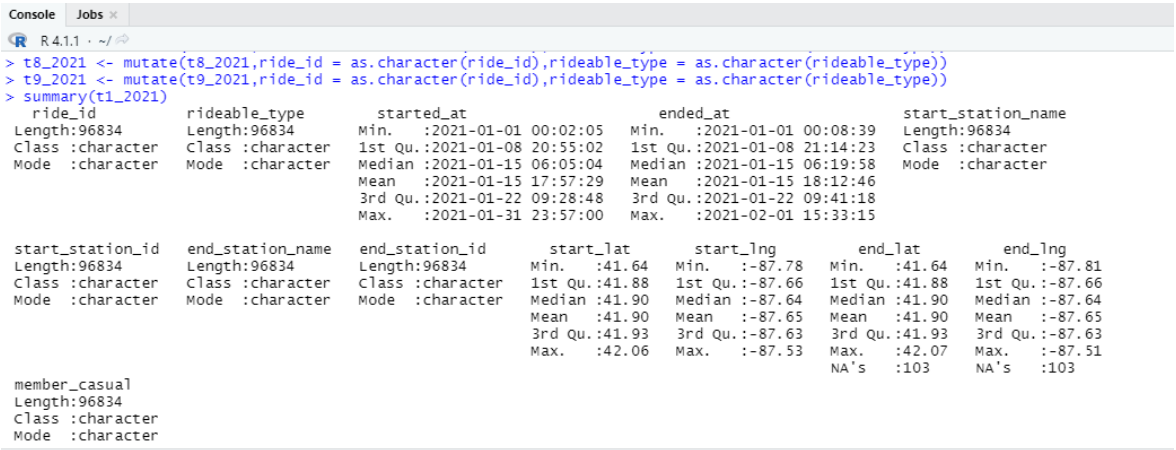

Figure 23. The result of summary(t1_2021)

And then the result proved that user finished the previous step. Next, all 9 data files owned have the same column types, so it is possible to combine these 9 data files into a single data file for easy further analysis:

- all_2021 <- bind_rows(t1_2021,t2_2021,t3_2021,t4_2021,t5_2021,t6_2021,t7_2021,t8_2021,t9_2021)

After the combination, ”all_2021” dataset is established. In addition, it is important to check it again all about the column in order to make sure for the consistency, the user apply colnames() code for this situation:

- colnames(all_2021)

```
 [1] "ride_id"            "rideable_type"      "started_at"         "ended_at"
 [5] "start_station_name" "start_station_id"   "end_station_name"   "end_station_id"
 [9] "start_lat"          "start_lng"          "end_lat"            "end_lng"
[13] "member_casual"
```

Figure 24: The number of columns in all_2021 dataset from January 2021 to September 2021

Next, for the purpose of calculating the amount of time people have spent on the Divvy service, it is crucial that creating a new column, which is to calculate the ride length of users:

- all_2021$ride_length <- difftime(all_2021$ended_at,all_2021$started_at, units = "mins")

Then checking it out again by using summary(all_2021):

```
   ride_id          rideable_type        started_at                    ended_at                    start_station_name
 Length:4356319     Length:4356319     Min.   :2021-01-01 00:02:05   Min.   :2021-01-01 00:08:39   Length:4356319
 Class :character   Class :character   1st Qu.:2021-05-23 16:11:48   1st Qu.:2021-05-23 16:42:05   Class :character
 Mode  :character   Mode  :character   Median :2021-07-08 17:25:10   Median :2021-07-08 17:43:02   Mode  :character
                                       Mean   :2021-07-01 07:12:39   Mean   :2021-07-01 07:36:01
                                       3rd Qu.:2021-08-18 21:33:27   3rd Qu.:2021-08-18 21:55:50
                                       Max.   :2021-09-30 23:59:48   Max.   :2021-10-01 22:55:35

 start_station_id   end_station_name   end_station_id       start_lat       start_lng         end_lat         end_lng
 Length:4356319     Length:4356319     Length:4356319     Min.   :41.64   Min.   :-87.84   Min.   :41.51   Min.   :-88.07
 Class :character   Class :character   Class :character   1st Qu.:41.88   1st Qu.:-87.66   1st Qu.:41.88   1st Qu.:-87.66
 Mode  :character   Mode  :character   Mode  :character   Median :41.90   Median :-87.64   Median :41.90   Median :-87.64
                                                          Mean   :41.90   Mean   :-87.65   Mean   :41.90   Mean   :-87.65
                                                          3rd Qu.:41.93   3rd Qu.:-87.63   3rd Qu.:41.93   3rd Qu.:-87.63
                                                          Max.   :42.07   Max.   :-87.52   Max.   :42.17   Max.   :-87.49
                                                                                           NA's   :3952    NA's   :3952
 member_casual      ride_length
 Length:4356319     Length:4356319
 Class :character   Class :difftime
 Mode  :character   Mode  :numeric
```

Figure 25. The summary of all_2021

Now, ride_length column is created.

As can be seen in the overview of total Divvy statistics from the beginning of January 2021 to the end of September 2021, there are some key data points to pay attention to, particularly member casual and ride length. As a result, all the information about the two data columns mentioned above. However, in the ride length column, a conversion is required for the calculation. The next task is to convert " ride_length" column to numerical type (although the datatype is already numeric mode , the author is still doing this step for data processing for other cases).

- is.factor(all_2021$ride_length)

```
> is.factor(all_2021$ride_length)
[1] FALSE
```

Figure 26. Coding outcome

Now applying the as.numeric() code for the conversion:

- all_2021$ride_length <- as.numeric(as.character(all_2021$ride_length))

Then checking it again about whether it is numerical data or not:

- is.numeric(all_2021$ride_length)

```
> is.numeric(all_2021$ride_length)
[1] TRUE
```

Figure 27. Coding outcome 2

The data of ride_length column has been converted to metric, next continuing to review the summary() data again for a preliminary check on this metric:

```
> summary(all_2021$ride_length)
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
   -7.05     7.25    12.85    23.36    23.22 55944.15
```

Figure 28. Coding outcome 3

The above result demonstrates the inconsistency of the ride length data, because all Divvy bike rides must be greater than or equal to 0, so all numbers less than one (< 1) must be removed to keep the data consistent. Additionally, as Divvy data mentioned, all data with a start to end time of less than 1 minute is proven to be due to technical processing on the part of Divvy staff and it should be deleted for research purposes. (Divvy 2021k.)

- all_2021a <- all_2021[!(all_2021$ride_length<1),]

Now running summary(all_2021a$ride_length) for re-checking it:

- summary(all_2021a$ride_length)

```
> summary(all_2021a$ride_length)
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
    1.00     7.47    13.07    23.71    23.45 55944.15
```

Figure 29. Coding outcome 4

The data of ride_length has been cleansed, as seen above.

Next, considering the customer's rental time, we can extract the bike running time through two main columns, which are started_at and ended_at columns. From here, the author creates some more columns to categorize days, months and years specifically for more in-depth analysis.

There are some codes relating to building date-type columns:

- all_2021a$date <- as.Date(all_2021a$started_at)
- all_2021a$month <- format(as.Date(all_2021a$date), "%m")

- all_2021a$day <- format(as.Date(all_2021a$date), "%d")
- all_2021a$year <- format(as.Date(all_2021a$date), "%Y")
- all_2021a$day_of_week <- format(as.Date(all_2021a$date), "%A")

Now it shows:

Data	
all_2021	4356319 obs. of 14 variables
all_2021a	4290839 obs. of 19 variables
t1_2021	96834 obs. of 13 variables
t2_2021	49622 obs. of 13 variables
t3_2021	228496 obs. of 13 variables
t4_2021	337230 obs. of 13 variables
t5_2021	531633 obs. of 13 variables
t6_2021	729595 obs. of 13 variables
t7_2021	822410 obs. of 13 variables
t8_2021	804352 obs. of 13 variables
t9_2021	756147 obs. of 13 variables

Figure 30. Datasets

The all_2021a dataset has been added 5 more columns, and now it is necessary that investigating some formats of these variables:

- table(all_2021a$month)

```
> table(all_2021a$month)

    01     02     03     04     05     06     07     08     09
 95415  48649 225523 332422 523243 717699 809760 792943 745185
```

Figure 31. The data of month column in all_2021a dataset

Since the display of the month column's value is numeric, there is a necessary correction to the literal value and rearranging the order of values (from January to September in the order)

- all_2021a <- all_2021a %>% mutate(month = recode(month,"01"="Jan","02"="Feb","03"="Mar","04"="Apr","05"="May","06"="Jun","07"="Jul","08"="Aug","09"="Sep"))
- all_2021a$month <- ordered(all_2021a$month, levels=c("Jan", "Feb", "Mar", "Apr", "May", "Jun", "Jul", "Aug", "Sep"))

Then checking it again for the consistency:

- table(all_2021a$month)

```
> table(all_2021a$month)

   Jan    Feb    Mar    Apr    May    Jun    Jul    Aug    Sep
 95415  48649 225523 332422 523243 717699 809760 792943 745185
```

Figure 32. The ordered data of month column

All data contained in the all_2021a dataset in RStudio environment has almost been fully converted, filtered and edited, from here the author prepares to perform analysis based on the above data.

7.3 Data analysis with R programming

First of all, it is fundamental to check the type of riders by coding table(all_2021a$rideable_type):

- table(all_2021a$rideable_type)

```
> table(all_2021a$rideable_type)

 classic_bike   docked_bike electric_bike
      2645236        275190       1370413
```

Figure 33. The type of customers in Divvy

According to the picture above, there are three core classes of biker: classic, docked and electric ones. In detail, users preferred to choose classic and electric bikes rather than docked bikes.

Then now to determine the top ten longest Divvy service trips, first create a condition (condition) for the all_2021a dataset so that it can be filtered out:

In order to create a condition for the question, a value can be produced by order() function:

- dv_ridelength <- order(all_2021a$ride_length, decreasing = T)

Values	
dv_ridelength	Large integer (4290839 elements, 17.2 MB)

Figure 34. The result of creating a condition in the top right corner in RStudio environment

Nextly, using head() function for filtering out the top 10 rows with decreasing value of ride length in Divvy:

- head(all_2021a[dv_ridelength,],10)

A tibble: 10 x 19

ride_id <chr>	rideable_type <chr>	started_at <S3: POSIXct>	ended_at <S3: POSIXct>	start_station_name <chr>
F043F0F6A1AA4F85	docked_bike	2021-06-05 02:27:26	2021-07-13 22:51:35	Michigan Ave & Lake St
7F0578ABF030FC83	docked_bike	2021-06-04 22:03:33	2021-07-13 14:15:14	Streeter Dr & Grand Ave
BDA1217EC8532C7B	docked_bike	2021-05-02 02:56:07	2021-06-08 13:37:43	State St & Van Buren St
E6E1E5BFFD72ECEC	docked_bike	2021-06-05 23:33:51	2021-07-12 13:55:14	Clark St & Lake St
99415B003B7E7EEE	docked_bike	2021-07-08 19:29:49	2021-08-11 21:56:58	Kedzie Ave & Milwaukee Ave
E84DF812305C9C9F	docked_bike	2021-04-02 17:50:00	2021-05-05 22:06:42	Wabash Ave & Wacker Pl
8016DC50082D9E93	docked_bike	2021-06-05 21:47:40	2021-07-08 13:18:31	Lake Shore Dr & Ohio St
89F33397EABB55E7	docked_bike	2021-07-08 15:13:08	2021-08-06 13:18:39	Streeter Dr & Grand Ave
E25E58F5E94EE351	docked_bike	2021-08-01 18:53:10	2021-08-30 16:42:20	Michigan Ave & Jackson Blvd
D4E35EE211346690	docked_bike	2021-07-10 15:59:21	2021-08-07 22:43:57	Dusable Harbor

1-10 of 10 rows | 1-5 of 19 columns

Figure 35. Top 10 decreasing value of ride_length column in all_2021a dataset

Followingly, the author analyzes departure and stop-over stations to find out how active Divvy stations in Chicago are by using sort() function:

- top_5_start_station_names <-sort(table(all_2021a$start_station_name), decreasing=TRUE)[1:5]
- knitr::kable(top_5_start_station_names, col.names = c("Starting Station Name", "Number of Rides"), caption = "Top 5 Starting Stations (Jan 2021 - Sep 2021)")

Top 5 Starting Stations (Jan 2021 - Sep 2021)

Starting Station Name	Number of Rides
Streeter Dr & Grand Ave	72052
Michigan Ave & Oak St	38709
Wells St & Concord Ln	34975
Millennium Park	34851
Theater on the Lake	32700

Figure 36. Top 5 starting stations

- top_5_ending_station_names <- sort(table(all_2021a$end_station_name), decreasing=TRUE)[1:5]
- knitr::kable(top_5_ending_station_names, col.names = c("Ending Station Name", "Number of Rides"), caption = "Top 5 Ending Stations (Jan 2021 - Sep 2021)")

Top 5 Ending Stations (Jan 2021 - Sep 2021)

Ending Station Name	Number of Rides
Streeter Dr & Grand Ave	72276
Michigan Ave & Oak St	39032
Millennium Park	35312
Wells St & Concord Ln	35219
Theater on the Lake	32937

Figure 37. Top 5 ending stations

In terms of kind of riders, it is important that establishing two conditions which are annual members and casual riders in order to achieve some top used stations in these type of users.

- annual_member <- all_2021a[!(all_2021a$member_casual == "casual"),]
- casual_user <- all_2021a[!(all_2021a$member_casual == "member"),]

Now counting top stations by each kind of customers:

- top_5_member_start <- sort(table(annual_member$start_station_name), decreasing=TRUE)[1:5]
- knitr::kable(top_5_member_start, col.names = c("Starting Station Name", "Number of Rides"), caption = "Annual Members, Top 5 Starting Station Names (Jan 2021 - Sep 2021)")

Annual Members, Top 5 Starting Station Names (Jan 2021 - Sep 2021)

Starting Station Name	Number of Rides
Clark St & Elm St	18535
Wells St & Concord Ln	18206
Kingsbury St & Kinzie St	16775
Wells St & Elm St	16012
Dearborn St & Erie St	14551

Figure 38. Top 5 starting stations of annual members

- top_5_member_end <- sort(table(annual_member$end_station_name), decreasing=TRUE)[1:5]

- knitr::kable(top_5_member_end, col.names = c("Ending Station Name", "Number of Rides"),caption = "Annual Members, Top 5 Ending Station Names (Jan 2021 - Sep 2021)")

Annual Members, Top 5 Ending Station Names (Jan 2021 - Sep 2021)

Ending Station Name	Number of Rides
Wells St & Concord Ln	18746
Clark St & Elm St	18738
Kingsbury St & Kinzie St	16976
Wells St & Elm St	16585
Dearborn St & Erie St	15122

Figure 39. Top 5 ending stations of annual members

- top_5_casual_start <- sort(table(casual_user$start_station_name), decreasing=TRUE)[1:5]
- knitr::kable(top_5_casual_start, col.names = c("Starting Station Name", "Number of Rides"), caption = "Casual Users, Top 5 Starting Station Names (Jan 2021 - Sep 2021)")

Casual Users, Top 5 Starting Station Names (Jan 2021 - Sep 2021)

Starting Station Name	Number of Rides
Streeter Dr & Grand Ave	58472
Millennium Park	28346
Michigan Ave & Oak St	26628
Shedd Aquarium	19519
Lake Shore Dr & Monroe St	19413

Figure 40. Top 5 starting stations of casual users

- top_5_casual_end <- sort(table(casual_user$end_station_name), decreasing=TRUE)[1:5]
- knitr::kable(top_5_casual_end, col.names = c("Ending Station Name", "Number of Rides"), caption = "Casual Users, Top 5 Ending Station Names(Jan 2021 - Sep 2021)")

Casual Users, Top 5 Ending Station Names(Jan 2021 - Sep 2021)

Ending Station Name	Number of Rides
Streeter Dr & Grand Ave	60153
Millennium Park	29009
Michigan Ave & Oak St	27766
Theater on the Lake	20508
Shedd Aquarium	18205

Figure 41. Top 5 ending stations of casual users

It seems that all the data has been filtered and clearly classified according to each type. Finally, the last thing to do before switching to data visualization is to save the all_2021a dataset as a file, so that it can be easily linked to Tableau for future analysis. It is vital to run write.csv() function to create a file based on data received:

- write.csv(all_2021a,"D:\\R thesis file\\thesisDivvy\\all2021data.csv")

Figure 42. all_2021a dataset as a file

7.4 Data visualization with Tableau

Regarding to Tableau visualization, opening Tableau Desktop and starting to link with the above all_2021a file that was downloaded from RStudio:

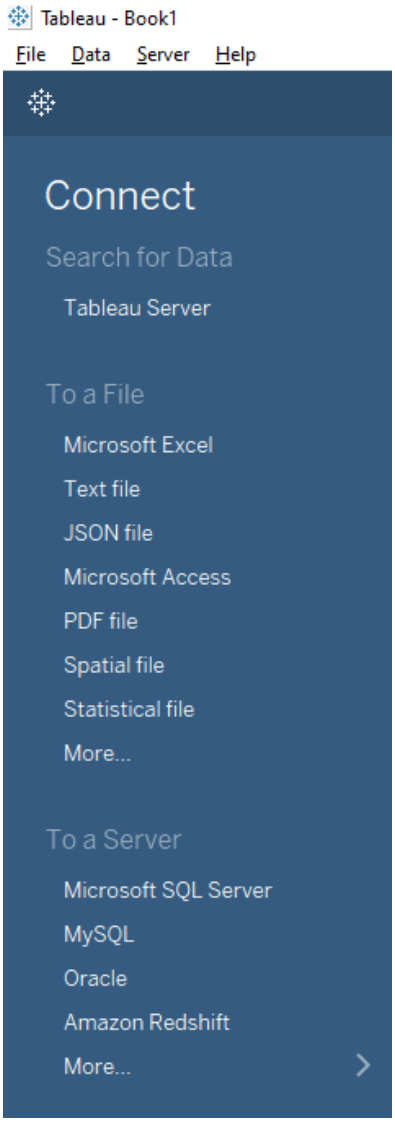

Figure 43. Tableau Desktop Menu

Then finding “To a File” section and choosing “More” option if audiences are not sure what specific file type the data file is, it can support to users automatically classify the file type. And select the file all2021data which has been saved as a .csv file.

For Tableau app, it is simple to do data visualization, the following description is an example for the average ride duration by each day for annual and casual members.

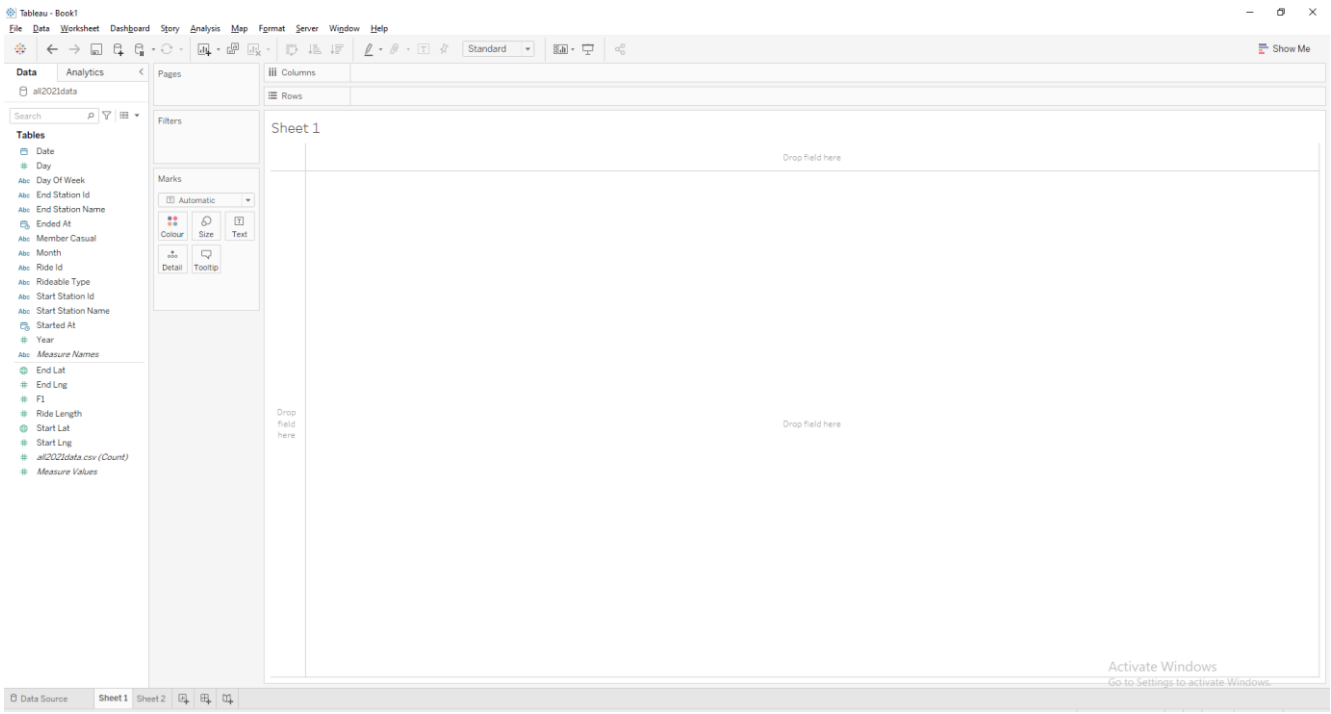

Figure 44. Basic interface of all2021data file in Tableau

It can be seen that there are two columns and rows at the top, where the user can put the data from the left-hand side in, and from there, it outputs the desired data:

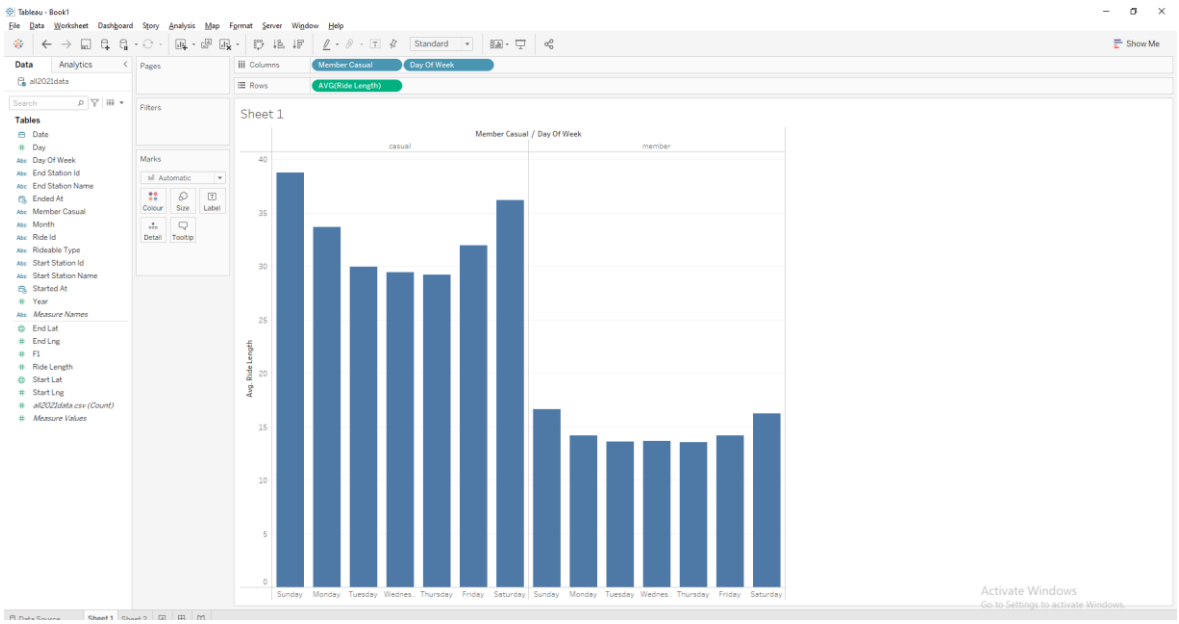

Figure 45. Result in Tableau

This is a picture that three main factors are “Member Casual”,” Day of Week” and “Ride Length” columns had been put on columns and rows functions. However, the data seems to be quite difficult to classify, thereupon, some more operations are needed to make the above graph more reasonable.

First, moving the Sunday column behind the Saturday column to make it easier for the user to see the data in chronological order. Thereafter, navigating to the " Marks " section and inserting a column of data called " Member Casual " into the “Colour” section to facilitate users to distinguish between the two key groups of Divvy customers. After completing the preceding procedure, the following results were obtained:

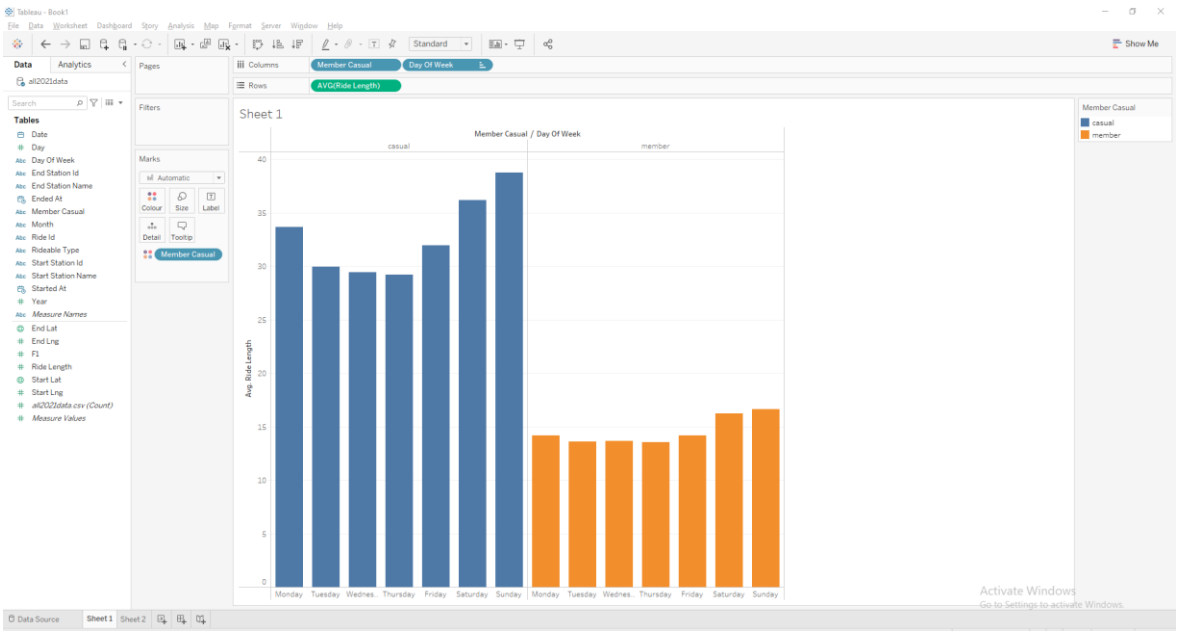

Figure 46. The second result in Tableau

Lastly, users simply need to click on the title place as "Sheet 1" and the names of columns x and y to rename it. Subsequently, audiences are equal to easily obtain the average time spent on using Divvy bikes of the different segments of customers. Audiences can also perform the same as described above, according to the graphs in the Findings section below.

In the case of the geographic graph below, the audience can do the following with the column and row positions:

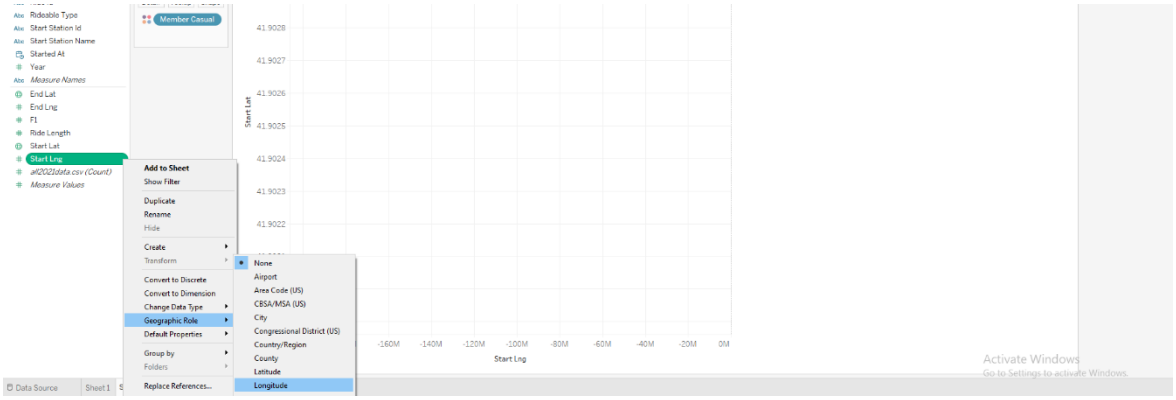

Figure 47. Changing to geographical data type

If the existing data is not in the geographical location format, the user can convert it as shown above.

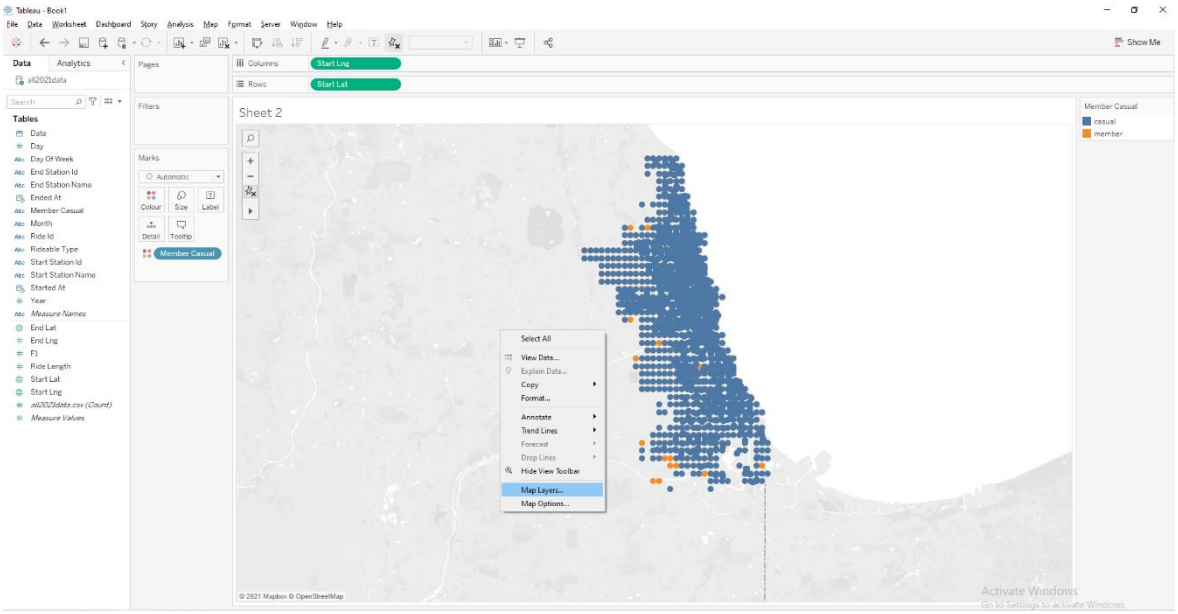

Figure 48. Map Layers in Tableau

In addition, audiences can pick “Map Layers” to add locations, areas, or borders related to geographic location. To get the same results as in Findings, drag " Ride Id " column into the Filter area and right-click and select " Count " measurement to select stations with more than 2000 pick-ups and drop-offs. In “Filter “area, it can be right clicked on “Ride Id “column and then select “Range of values” that audience desires to analyze. In this evaluation, author chose the range starting from 2000.

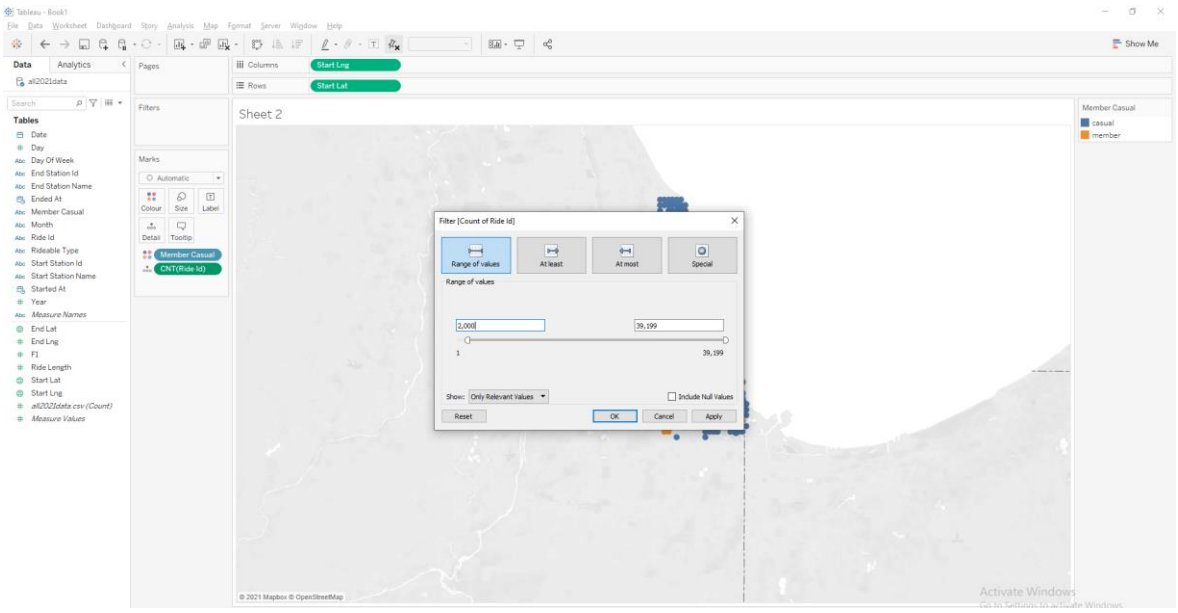

Figure 49. Range of values in Filter area

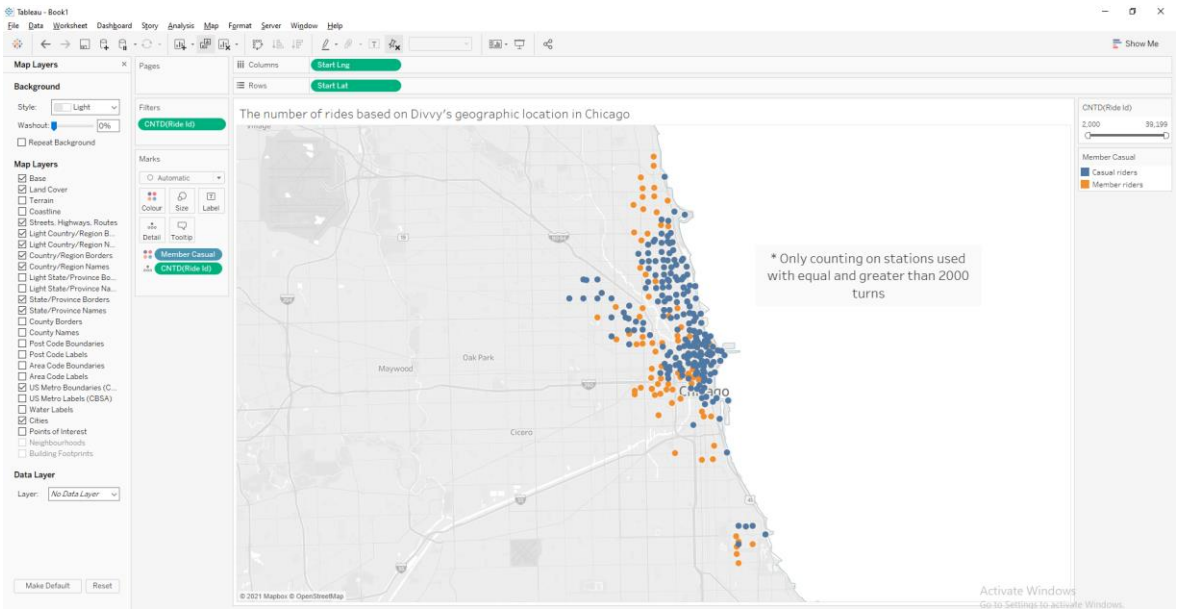

Figure 50. Geographical graph after applying the above filters and map layers function

8 Findings and recommendations

8.1 Findings

There are some visualizations and insights obtained after examining Divvy's historical data by means of R programming and Tableau:

The average ride duration by each day for annual and casual members

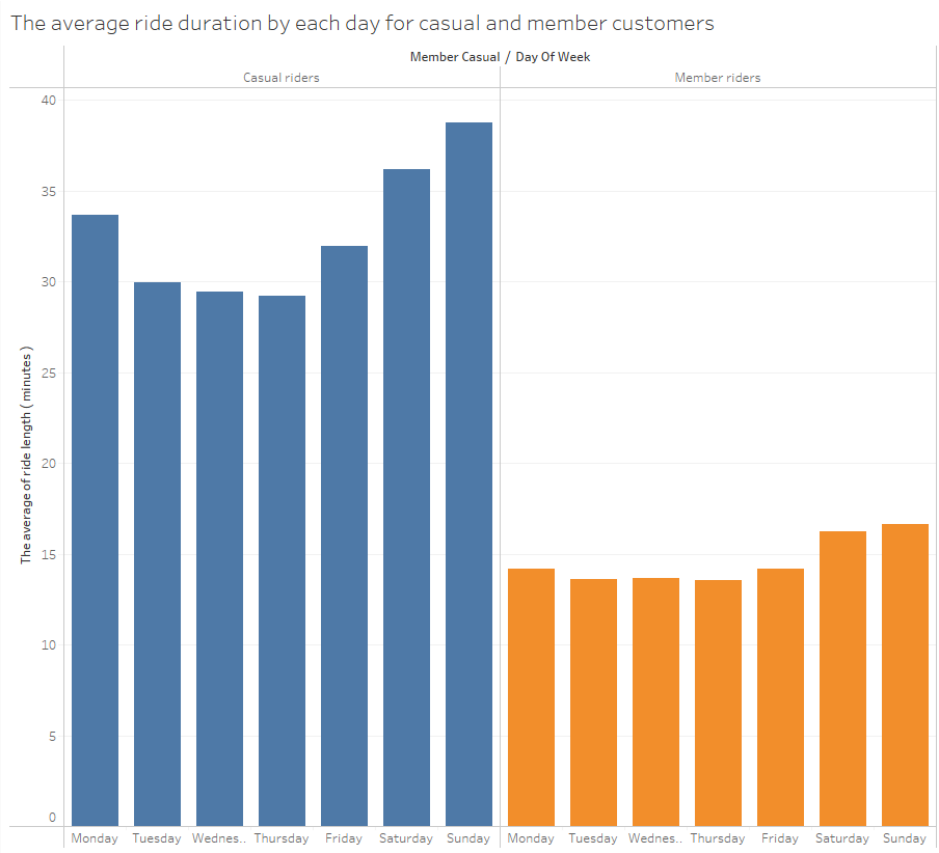

Figure 51. The average duration per day for two main types of users

The graph above captures the average hours spent using the Divvy service for two types of customers, demonstrating that demand for Divvy bikes is concentrated on weekends among casual riders, whereas official Divvy customers have a consistent time of about 15 minutes per day.

The number of rides by each day of week

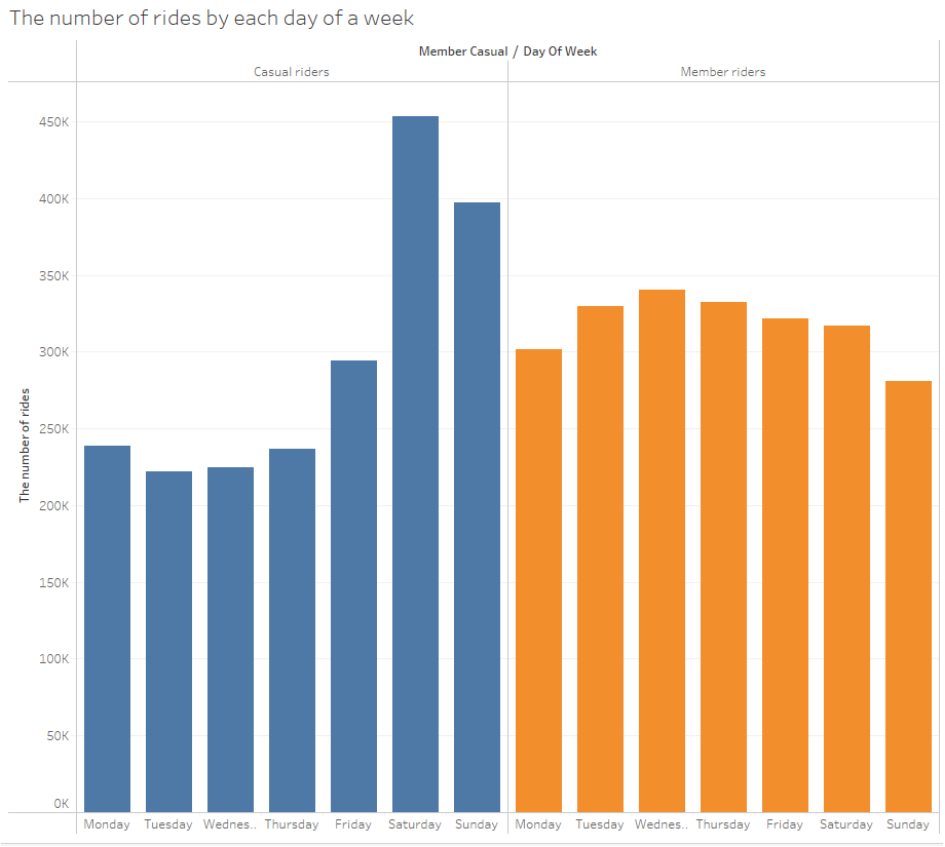

Figure 52. The number of rides by each day of week

The graph above depicts the time and number of times two types of customers used the Divvy service. It can be seen that Divvy bike usage is concentrated on weekends for casual riders, whereas official Divvy customers always have a stable number of Divvy app users.

The number of trips made per month

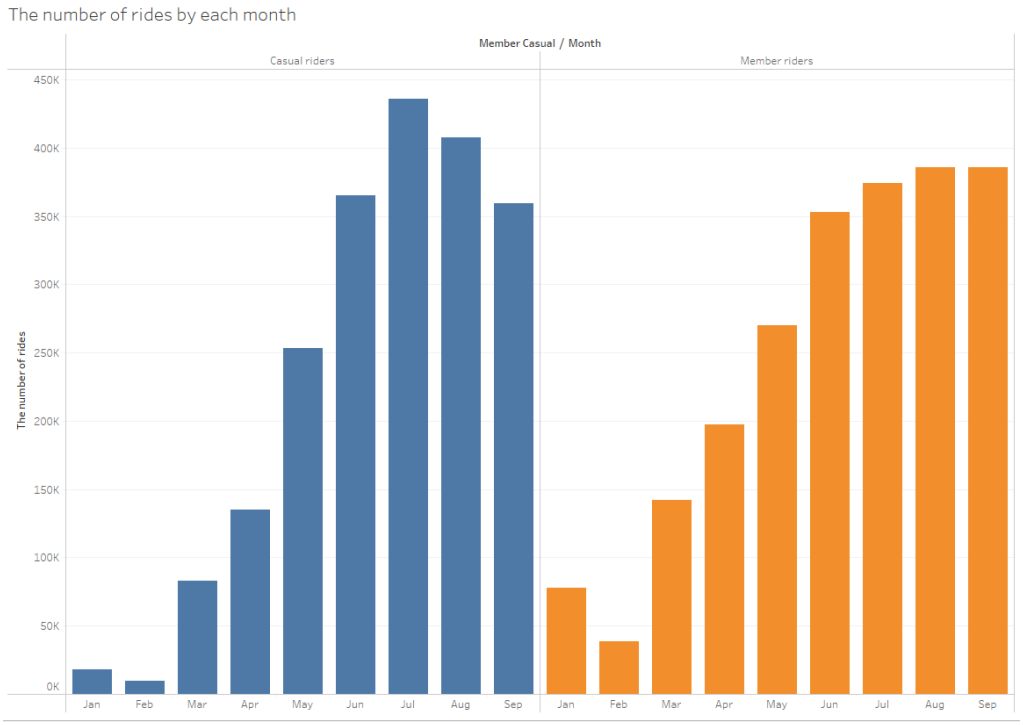

Figure 53. The number of rides monthly

The graph above shows the number of times of using Divvy service by two types of customers by month, accordingly it can be seen that Divvy bike usage is concentrated between June and September for both types of users.

The average journey periods of users each month

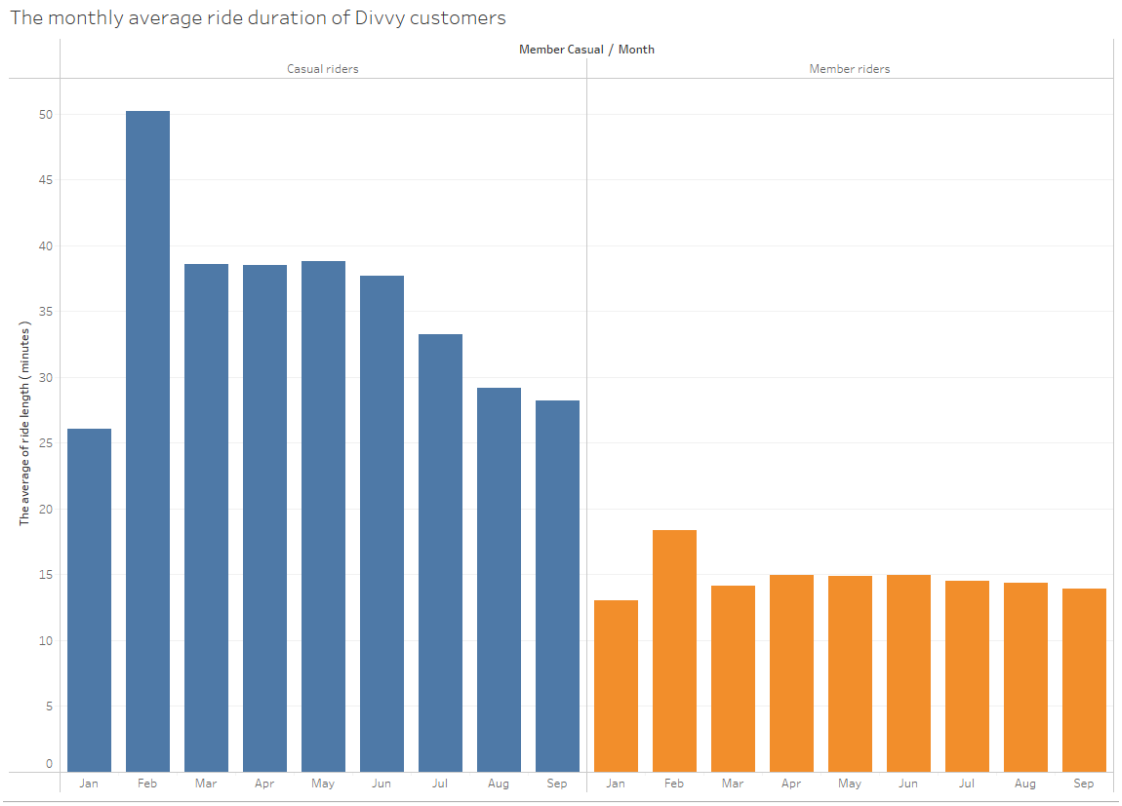

Figure 54. The monthly average duration of Divvy users

The graph above illustrates the average Divvy cycling time of the two types of customers by month. As can be seen, the average cycling time of the two members is fairly even, but it is quite high in February for casual members.

Top 5 starting station names

Top 5 Starting Stations (Jan 2021 - Sep 2021)

Starting Station Name	Number of Rides
Streeter Dr & Grand Ave	73118
Michigan Ave & Oak St	39271
Wells St & Concord Ln	35387
Millennium Park	35331
Theater on the Lake	33140

Figure 55. Top 5 starting stations

Top 5 ending station names

Top 5 Ending Stations (Jan 2021 - Sep 2021)

Ending Station Name	Number of Rides
Streeter Dr & Grand Ave	73287
Michigan Ave & Oak St	39572
Millennium Park	35749
Wells St & Concord Ln	35607
Theater on the Lake	33349

Figure 56. Top 5 ending stations

Top 5 starting and ending stations of annual participants

Annual Members, Top 5 Starting Station Names (Jan 2021 - Sep 2021)

Starting Station Name	Number of Rides
Clark St & Elm St	18836
Wells St & Concord Ln	18449
Kingsbury St & Kinzie St	16999
Wells St & Elm St	16224
Dearborn St & Erie St	14775

Figure 57. Top 5 starting stations of annual members

Annual Members, Top 5 Ending Station Names (Jan 2021 - Sep 2021)

Ending Station Name	Number of Rides
Clark St & Elm St	19036
Wells St & Concord Ln	18978
Kingsbury St & Kinzie St	17197
Wells St & Elm St	16791
Dearborn St & Erie St	15334

Figure 58. Top 5 ending stations of annual members

Top 5 starting and ending stations of casual users

Casual Users, Top 5 Starting Station Names (Jan 2021 - Sep 2021)

Starting Station Name	Number of Rides
Streeter Dr & Grand Ave	59193
Millennium Park	28683
Michigan Ave & Oak St	26941
Shedd Aquarium	19765
Lake Shore Dr & Monroe St	19617

Figure 59. Top 5 starting stations of casual users

Casual Users, Top 5 Ending Station Names(Jan 2021 - Sep 2021)

Ending Station Name	Number of Rides
Streeter Dr & Grand Ave	60832
Millennium Park	29318
Michigan Ave & Oak St	28056
Theater on the Lake	20725
Shedd Aquarium	18427

Figure 60. Top 5 ending stations of casual users

The number of rides by bike classification

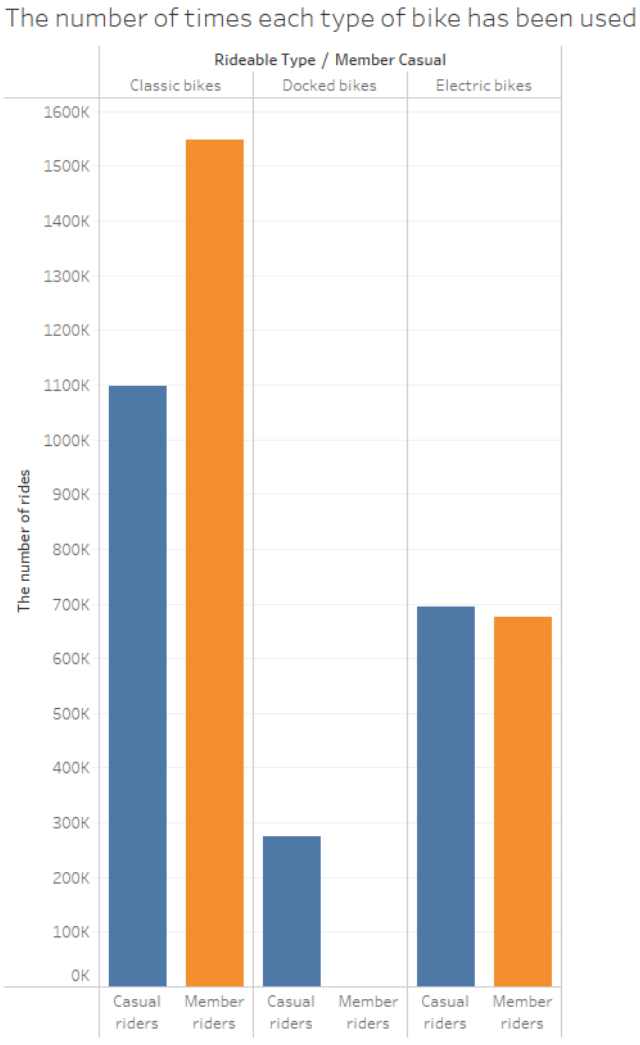

Figure 61. The number of rides by biker types

The graph above shows that the number of people using two core types of vehicles, which are classic and electric bikes, rather than docked bikes.

The average journey time of the two types of customers

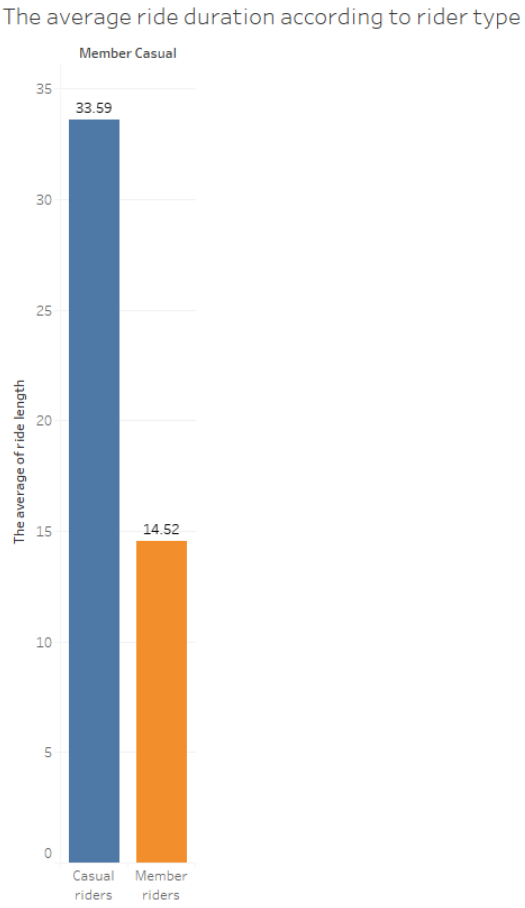

Figure 62. The average ride duration according to ride types

The graph above confirms that casual users frequently use Divvy bikes for seemingly twice as long as yearly members, specifically 33.59 minutes versus 14.52 minutes respectively.

Geographical location of Divvy stations

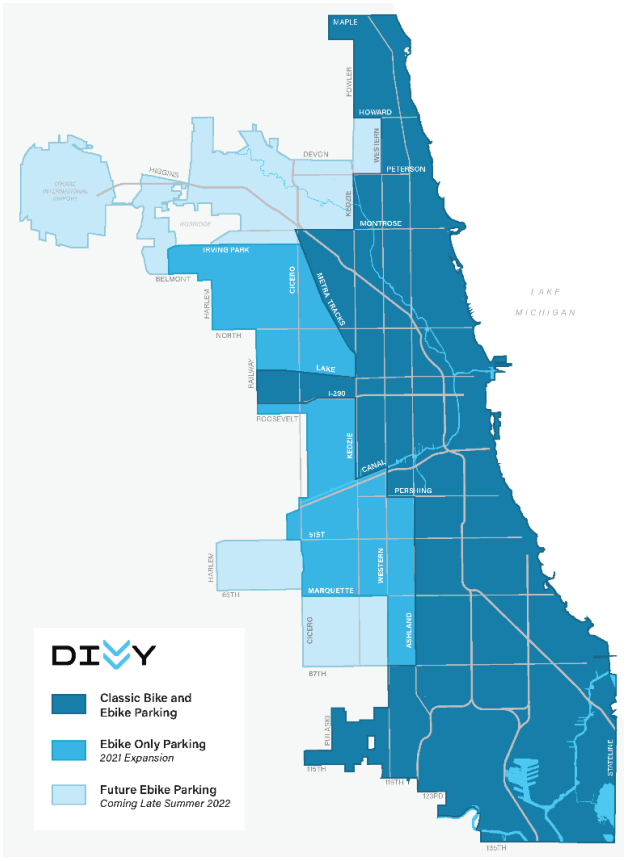

Figure 63. Divvy riding zones in Chicago (Divvy 2021j)

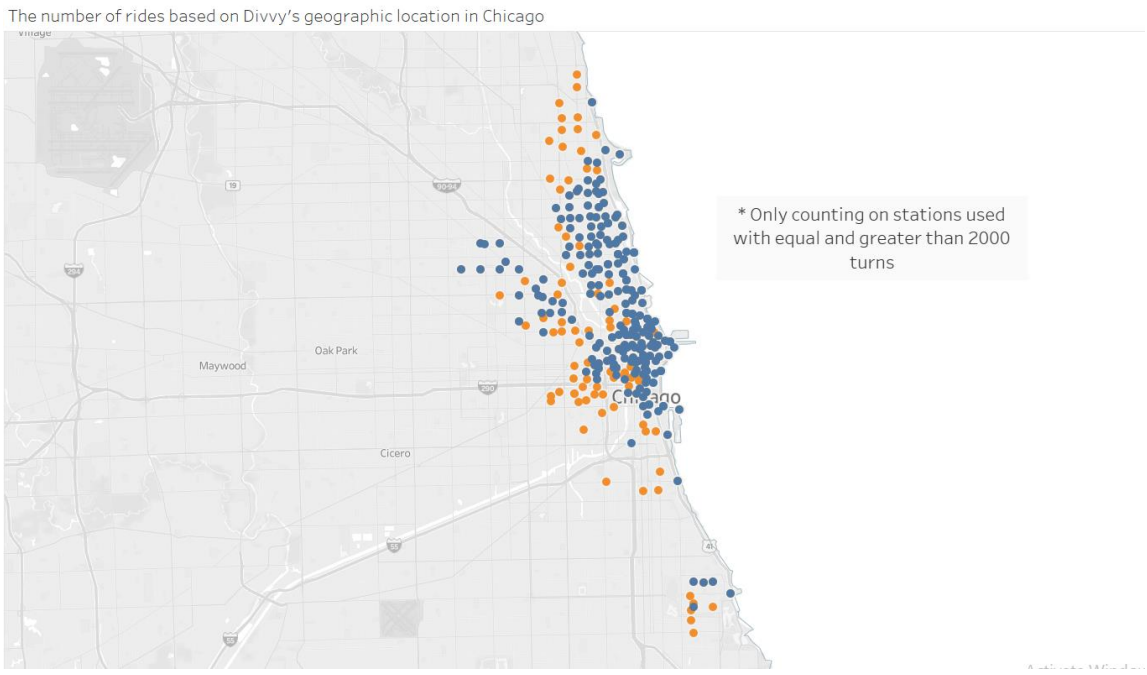

Figure 64. The number of selected rides based on the geographical location in Chicago of Divvy service

The graph above shows the areas of concentration of each type of customer, which allows Divvy station managers find the necessary areas for customer needs and cut unnecessary stations, allowing Divvy to enhance consumer retention and facilitate user movement.

Weather forecast in Chicago, USA

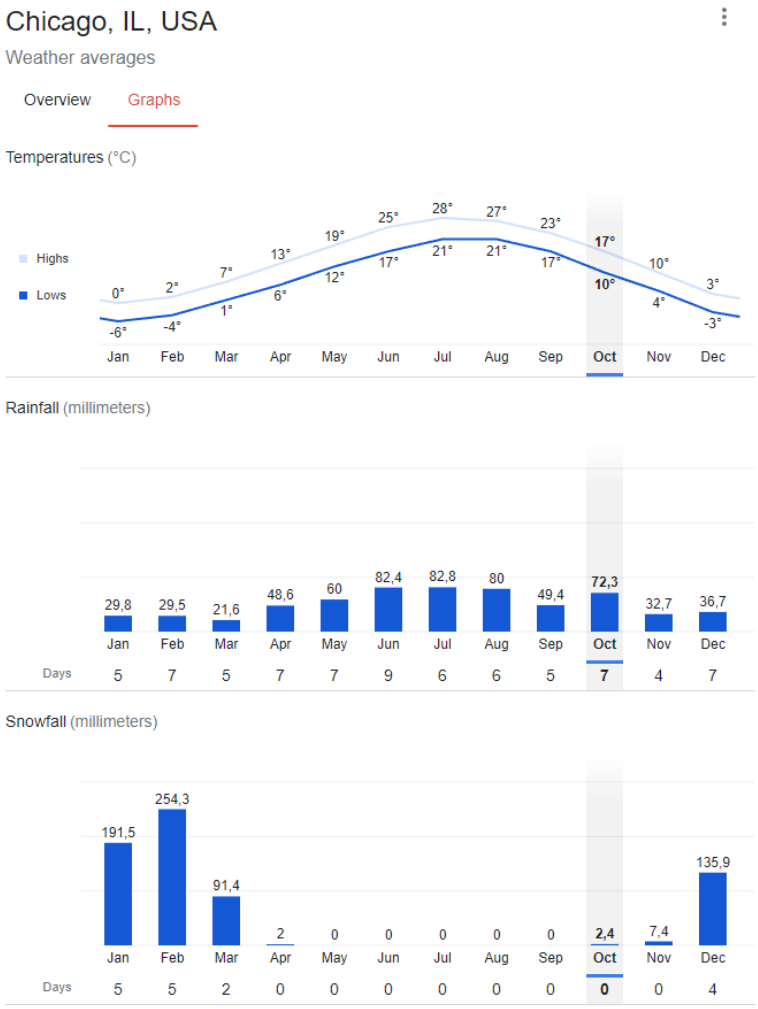

Figure 65. Weather information in Chicago, USA from January 2021 - September 2021 (NOAA 2021)

8.2 Recommendations

According to the graphs, casual users spent more time on using Divvy bikes than official Divvy members. Along with that, that chart show the number of casual riders with the highest demand for Divvy around the weekend, while member users usually maintain a level of usage in balance.

On the other hand, the number of Divvy bicycle users is mostly concentrated from May to September, and from January to April, most of the customer demand is quite small. One of

the main reasons affecting demand for bicycle users is the weather. The frequency of excursions is reduced by high humidity, rain, and strong winds (Capo Velo 2017). According to National Centers for Environment Information (NOAA) above, during the first three months of the year, the weather is quite cold and snowy, which partly makes customers have no need to use bicycles. On account of this, Divvy should provide some strategies that as the following:

Through customer needs, from which Divvy can devise marketing strategies as well as maintain their vehicles, for example above, Divvy company may reduce the number of bicycles offered in the first 3 months of the year to repair, to maintain, and to develop bicycles to remain the best customer experience.

In addition, Divvy can switch from supplying classic bikes to all electric bikes during the first 3 months of the year which could reduce the negative impact on the environment, in that case, it will assist users in using the vehicle throughout winter season. (Capo Veli 2017). Casual users' Divvy vehicle usage time is consistent from January to March of the beginning of 2021, whereas member users' Divvy vehicle usage time is consistent for whole chosen months of 2021.

Marketing analysts can set up several banners or billboards about Divvy events and offers for consumers based on consumer arrivals and departures. Furthermore, Divvy can set up and edit stations based on the demands of users thanks to an analysis of the most frequently used arrival and departure locations. This helps to attract more promising customers, to increase customer retention, and to maintain a positive customer experience.

Based on the data visualizations display, the amount of riding time of casual ones is much more than that of annual users for the following reason: customers do not appreciate the option of using Divvy's annual membership. After reviewing the price as well as the payment method, it can be seen that it is less attractive and effective for users to spend money to buy a one-year package. Because each person's needs are distinct, for example, there are some customers who only desires to keep their membership for three months, nevertheless, it looks like to forces the customer to pay for 12-month subscription. Under those circumstances, Divvy should be divided into different small packages to motivate residents and tourists to be willing to pay for the service:

- Monthly plan: 12 dollars per month
- Quarterly package: 40 dollars per quarter
- Annual package: 108 dollars per year

Afterwards, according to the level of use of vehicles, people primarily selected classic and electric vehicles. The "number of uses by type of bike" chart indicates that classic bikes are the most commonly used. Furthermore, by taking into account the number of most frequently used arrival and departure stations based on geographical location, Divvy is up to form a business perspective and to offer centred marketing strategies on that area. Not only that, but Divvy can also eliminate unnecessary stations and concentrate on upgrading and repairing products to maintain its quality.

Besides that, one of the price campaigns that Divvy can implement is to provide a discount or trial riding time as a membership. It gives users more access to the product and makes them feel more at ease when making decisions. Based on user data from casual riders and previous research mentioned, Divvy should increase their advertising policies for casual riders in the future in order to convert them into annual members. In this context, the graphs above show that casual users frequently use the product on weekends and between summer and the end of autumn.

As a result, incentives such as “Free for First Time” promotion, which is conducting a short-term free first trip promotion for newcomers will have a significant effect on the success of Divvy services. It assists Divvy in both attracting more official customers and taking advantage of the product, avoiding the bicycle that has not been used for a long time.

Last but foremost, Divvy should create more competitive rankings for users, such as rating the travelled distance and awarding its achievements, which potentially maximizes the interaction between Divvy and end-users, giving Divvy organization more opportunities to grow their businesses.

8.2.1 Applications of 7P’s of marketing in Divvybikes

Based on the findings, the author uses the 7Ps of marketing to provide corresponding solutions for Divvy:

- Product: Based on historical data, we can see that Divvy should focus on developing classic and electric bikes rather than docked bikes, as this helps Divvy cut costs while also bringing the most appropriate products to customers.
- Price: In Divvy pricing offer, the price issue should be the primary focus. The division of prices into tiers creates conditions for customers to be flexible in payment as well as spend the appropriate amount of money.

- Place and Process: Using historical data, the arrangement of reasonable bike stations helps to increase product awareness for Divvy while also increasing convenience for customers when picking up or dropping off their bikes.
- Promotion: Divvy is able to design banners or advertising campaigns in the most user-friendly areas, attracting customers and increasing the effectiveness of the incentive strategy.
- Physical evidence: Divvy's bicycle products need to be improved and better developed for situations such as riding on uneven roads or in the winter.

8.2.2 Application of SWOT analysis

In terms of SWOT assessment in Divvy operations:

- Strength: The geographic location of Divvy in Chicago is concentrated near the coast, which is quite convenient because tourists frequently gather and visit here. It makes it simple to entice interest tourists and citizens. Furthermore, the quality of modern and reasonably reliable bicycles is a strong point of Divvy's service.
- Weaknesses: Customers may be discouraged from staying with Divvy for an extended period of time due to its pricing factors.
- Opportunity: Because travel demand in the United States positioned in a third place worldwide in 2020 (World Tourism Organization 2020). Thus, there is some potential for expanding the service to areas adjacent to Chicago.
- Threats: Based on the company's weaknesses, customers seem obligated to buy this package and it may cause customers to seek a more appropriate service, such as Uber, if Divvy does not work out and find other alternatives for providing price flexibility to consumers. This issue has an impact on the company's customer loyalty.

9 Discussion

The issues raised at the beginning of the research paper can be described thoroughly using the above analysis process.

Sub-question 1: How can author ensure data quality?

The data of the Divvy company is public on their official website, according to the information mentioned and clearly extracted about data collection in section 2.2 Data collection method. And, in compliance with the applicable patent law, the data can be accessed by everyone for the purpose of in-depth research.

Furthermore, data quality is guaranteed because this is secondary data derived from official sources, allowing the author to ensure the data's reliability and consistency.

Sub-question 2: How can useful information be acquired by taking advantages of historical data?

The application of data selection and analysis tools can gain valuable knowledge; in this thesis, the use of R programming and the data visualization application Tableau has yielded many potential insights for the Divvy company.

The charts in section 8.1 Findings accelerates for the audience to recognize all of the information obtained.

Sub-question 3: What are the results obtained for the marketing strategy?

Section 8.2 Recommendations interpreted and noted in detail the results and insights for Divvy's marketing strategy.

Main question: How can historical data help in providing a marketing strategy for Divvy bikes?

Historical data has completely figured out the concerns along with provided some very significant insights in promoting the Divvy brand. As well as it offered a number of improvements in the Divvy service, as a means to increase customer satisfaction and reach wider potential users, as mentioned in section 8.2 Recommendations.

10 Conclusion

The main goal of this thesis is to bring better marketing strategies to the company by using historical data collected in the past.

Undoubtedly, the analysis and insights provided by historical data demonstrate that the rational use of historical data brings immense value, not only for marketing but also for other aspects such as how to effectively manage service performance. The collected data is guaranteed for consistency and reliability by the Divvy Data License Agreement, which is discussed in chapter 1. Furthermore, useful information can be obtained by analyzing each representative value present in the data while performing data selection, classification, and analysis operations.

However, there are some limitations that remain after the research paper has been completed. This thesis has been somewhat successful with reference to marketing strategy solutions based on the 7P's of marketing framework, however, there is some aspects such as "People" element in 7P's still needs to be provided with more data as well as in-depth analysis to be able to bring about human-related strategies in Divvy's operating system. In addition, it is critical for the analyst to have an experience with data manipulation and tool utilization.

Because the thesis relies heavily on historical data to provide insights for the case company Divvy, the author has made some recommendations for future research. First, various areas apart from the 7P's of marketing mix can be thoroughly examined. Furthermore, another piece of advice is to use other types of data that are not historical data, such as financial figures, can still assist to deliver potential competitive advantages to a firm that have not been well researched. As a consequence, all of the above study directions choose to capitalize on big data's capability to add significant value to businesses.

Through the recommendations found based on analyzed historical data, Divvy can then understand the majority of the reasons why the number of Divvy bikes used is only concentrated in the summer and fall due to weather concerns; therefore, sensible strategies in winter times such as modifying the bike design for winter months or minimizing the operating area based on customer needs are required.

Furthermore, by assessing user needs, providing too few membership packages makes customers hesitant to use Divvy's service. Accordingly, Divvy should indeed evolve and increase service flexibility by offering more payment options to customers. Finally, some first-time incentives or direct advertising at the most strongly used docked stations will be good tactics.

References

Ahlemeyer-Stubbe, A. & Coleman, S. 2018. Monetizing Data: How to Uplift Your Business. John Wiley & Sons.

American Marketing Association 2017. Definitions of marketing. Retrieved on 28th October 2021. Available at https://www.ama.org/the-definition-of-marketing-what-is-marketing/

BearingPoint 2021. Data & Analytics. Analytical insights to boost your business performance. Retrieved on 24th October 2021. Available at https://www.bearingpoint.com/files/Data_and_Analytics_boost_your_business_performance.pdf

Bhasin, H. 2021. Service Marketing Mix – 7 P's of marketing. Marketing91. Retrieved on 15th October 2021. Available at https://www.marketing91.com/service-marketing-mix/

Bryce, K. 2021.The many benefits of bike sharing programs. Retrieved on 13th October 2021. Available at https://www.commuteoptions.org/the-many-benefits-of-bike-sharing-programs/

Capo Velo 2017.What bike sharing programs need to succeed. Retrieved on 20th October 2021. Available at https://capovelo.com/bike-sharing-programs-need-succeed/

Divvy 2021a. About Divvy. Retrieved on 13th October 2021. Available at https://www.divvybikes.com/about

Divvy 2021b. Data License Agreement. Retrieved on 14th October 2021. Available at https://www.divvybikes.com/data-license-agreement

Divvy 2021c. How it works. Retrieved on 21st October 2021. Available at https://www.divvybikes.com/how-it-works

Divvy 2021d. Annual Memberships. Retrieved on 22nd October 2021. Available at https://www.divvybikes.com/pricing/annual

Divvy 2021e. Single Ride. Retrieved on 22nd October 2021. Available at https://www.divvybikes.com/pricing/single-ride

Divvy 2021f. Day Pass. Retrieved on 22nd October 2021. Available at https://www.divvybikes.com/pricing/day-pass

Divvy 2021g. Meet the bikes. Retrieved on 22nd October 2021. Available at https://www.divvybikes.com/how-it-works/meet-the-bikes

Divvy 2021h. Meet the Classic. Retrieved on 22nd October 2021. Available at https://www.divvybikes.com/how-it-works/classic

Divvy 2021i. Meet the Divvy ebike. Retrieved on 22nd October 2021. Available at https://www.divvybikes.com/how-it-works/ebike

Divvy 2021j. Ebike pricing & parking. Retrieved on 23rd October 2021. Available at https://www.divvybikes.com/pricing/ebikes

Divvy 2021k. Divvy Data. Retrieved on 10th November 2021. Available at https://www.divvybikes.com/system-data

European Cyclists' Federation 2021. Bike Share Schemes. Retrieved on 21st October 2021. Available at https://ecf.com/what-we-do/urban-mobility/bike-share-schemes-bss

Frankenfield, J. 2020a. Descriptive Analytics. Investopedia. Retrieved on 14th October 2021. Available at https://www.investopedia.com/terms/d/descriptive-analytics.asp

Frankenfield, J. 2021b. Data Analytics. Investopedia. Retrieved on 14th October 2021. Available at https://www.investopedia.com/terms/d/data-analytics.asp

Garcia, E. 2021. Cyclistic Case Study: Analysis & Visualizations. Kaggle. Retrieved on 14th November 2021. Available at https://www.kaggle.com/edgarcia1/cyclistic-case-study-analysis-visualizations

Halton, C. 2021. Predictive Analytics. Investopedia. Retrieved on 14th October 2021. Available at https://www.investopedia.com/terms/p/predictive-analytics.asp

Hanlon, A. 2021. How to use the 7Ps Marketing Mix. Smart Insights. Retrieved on 15th October 2021. Available at https://www.smartinsights.com/marketing-planning/marketing-models/how-to-use-the-7ps-marketing-mix/

I2tutorials 2020. Introduction to R programming language. Retrieved on 20th October 2021. Available at https://www.i2tutorials.com/introduction-to-r-programming-language/

Itl Education Solutions Limited 2010. Introduction to Database Systems. Pearson Education India.

Kathy 2020. Why is sustainable transportation important? Retrieved on 13th October 2021. Available at https://clarifygreen.com/why-is-sustainable-transportation-important/

Kenton, W. 2021. Strength, Weakness, Opportunity, and Threat (SWOT) Analysis. Investopedia. Retrieved on 15th October 2021. Available at https://www.investopedia.com/terms/s/swot.asp

Kotler, P., Keller, K. 2007. A Framework for Marketing Management 3rd edition. New Jersey: Prentice Hall.

Lutkevich, B. 2021. database (DB). TechTarget. Retrieved on 16th October 2021. Available at https://searchdatamanagement.techtarget.com/definition/database

Mailchimp 2021. Marketing Mix. Retrieved on 15th October 2021. Available at https://mailchimp.com/marketing-glossary/marketing-mix-7ps/

Marr, B. 2017. Data Strategy: How to Profit from a World of Big Data, Analytics and the Internet of Things. Kogan Page Publishers.

Martucci, B. 2020. What Is Bike Sharing? How Bike-Share Programs Work, Pros & Cons. Money Crashers. Retrieved on 16th October 2021. Available at https://www.moneycrashers.com/bike-sharing-best-bike-share-programs/

Mass Live Media 2021. Why you need the 7 P's of marketing. Retrieved on 15th October 2021. Available at https://www.masslivemedia.com/the-7-ps/

McLeod, S. 2019. What's the difference between qualitative and quantitative research? SimplyPsychology. Retrieved on 20th October 2021. Available at https://www.simplypsychology.org/qualitative-quantitative.html

Migrant & Seasonal Head Start Technical Assistance Center 2006. Introduction to Data Analysis Handbook. Academy for Educational Development.

Miller, S. 2021. What is R used for? Codecademy. Retrieved on 20th October 2021. Available at https://www.codecademy.com/resources/blog/what-is-r-used-for/

Moshirpour, M., Far, B. & Alhajj, R. 2017. Highlighting the Importance of Big Data Management and Analysis for Various Applications. Springer.

Nikitas, A. 2016. The global bike sharing boom – why cities love a cycling scheme. THE CONVERSATION. Retrieved on 13th October 2021. Available at https://theconversation.com/the-global-bike-sharing-boom-why-cities-love-a-cycling-scheme-53895

NOAA 2021. National Centers for Environmental Information. USA Government. Retrieved on 28th October 2021. Available at https://www.ncdc.noaa.gov/

Prachi, M. 2019. Service Marketing. The Investor Book. Retrieved on 15th October 2021. Available at https://theinvestorsbook.com/service-marketing.html

Queensland Government 2021. The 7 Ps of marketing. Retrieved on 15th October 2021. Available at https://www.business.qld.gov.au/running-business/marketing-sales/marketing-promotion/marketing-basics/seven-ps-marketing

R 2021a. The R Project for Statistical Computing. Retrieved on 16th October 2021. Available at https://www.r-project.org/

R 2021b. What is R? Retrieved on 16th October 2021. Available at https://www.r-project.org/about.html

Red Bike Marketing 2021. What is Physical Evidence? Retrieved on 15th October 2021. Available at https://redbikemarketing.com/marketing/physical-evidence/

Ruppenthal, A. 2019. Chicago's Divvy Divide: Convenient Bike-Share Program or Sign of Gentrification? Wttw. Retrieved on 20th October 2021. Available at https://news.wttw.com/2019/01/04/chicago-s-divvy-divide-convenient-bike-share-program-or-sign-gentrification

Russell, E. 2017. The Fundamentals of Marketing. Bloomsbury Publishing.

Saunders, M., Lewis, P. & Thornhill, A. 2016. Research Methods for Business Students. Seventh edition. Edinburgh: Pearson Education Limited.

Segal, T. 2021a. Prescriptive Analytics. Investopedia. Retrieved on 14th October 2021. Available at https://www.investopedia.com/terms/p/prescriptive-analytics.asp

Segal, T. 2021b. Big Data. Investopedia. Retrieved on 20th October 2021. Available at https://www.investopedia.com/terms/b/big-data.asp

Sekaran, U. & Bougie, R. 2016. Research Methods for Business: A Skill Building Approach. Seventh edition. Chichester, West Sussex, UK: John Wiley and Sons.

Sisense 2021. Diagnostic Analytics. Retrieved on 14th October 2021. Available at https://www.sisense.com/glossary/diagnostic-analytics/

Spinu, V., Grolemund, G. & Wickham, H. 2021. Overview of lubridate. Lubridate. Retrieved on 10th November 2021. Available at https://lubridate.tidyverse.org/

Tableau 2021a. What is Tableau? Tableau. Retrieved on 20th October 2021. Available at https://www.tableau.com/why-tableau/what-is-tableau

Tableau 2021b. Why choose Tableau? Tableau. Retrieved on 20th October 2021. Available at https://www.tableau.com/why-tableau

Techopedia 2021. Analytics. Retrieved on 14th October 2021. Available at https://www.techopedia.com/definition/30296/analytics

The Chartered Institute of Marketing 2015. Marketing and the 7Ps: A brief summary of marketing and how it works. Retrieved on 15th October 2021. Available at https://www.cim.co.uk/media/4772/7ps.pdf

Twin, A. 2021. The 4 Ps. Investopedia. Retrieved on 14th October 2021. Available at https://www.investopedia.com/terms/f/four-ps.asp

Vidhya, A. 2019. A Beginner's Guide to Tidyverse – The Most Powerful Collection of R Packages for Data Science. Analytics Vidhya. Retrieved on 10th November 2021. Available at https://www.analyticsvidhya.com/blog/2019/05/beginner-guide-tidyverse-most-powerful-collection-r-packages-data-science/

World Tourism Organization 2020. World Tourism Barometer. Retrieved on 14th November 2021. Available at https://www.e-unwto.org/doi/epdf/10.18111/wtobarometereng.2020.18.1.5

Appendix 1. R notebook

Code

Providing marketing mix strategies through historical data of Divvy company

Introduction

In this notebook will show the process of collecting, converting, organizing and analyzing the available data of Divvy company, thereby giving important insights for their upcoming marketing strategies.

Description of data origin utilized

Divvy's data is collected primarily from their operations in the Chicago location. Current data is from January 2021 to the end of September 2021, thus, it includes 9 .csv files. The detailed information table below will show each data type and the meaning of each data type:

Column Name	Data Type	Description
ride_id	STRING	Unique Ride ID Number
rideable_type	STRING	Bike Type (Classic Bike, Docked Bike, Electric Bike)
started_at	TIMESTAMP	Ride Start Date (YYYY-MM-DD HH:MM:SS UTC)
ended_at	TIMESTAMP	Ride End Date (YYYY-MM-DD HH:MM:SS UTC)
start_station_name	STRING	Ride's Starting Station Name
start_station_id	STRING (Data from January - September 2021, INTEGER)	Starting Station ID Number
end_station_name	STRING	Ride's Ending Station Name
end_station_id	STRING (Data from January - September 2021, INTEGER)	Ending Station ID Number
start_lat	FLOAT	Starting Station Coordinates (Latitude)

Column Name	Data Type	Description
start_lng	FLOAT	Starting Station Coordinates (Longitude)
end_lat	FLOAT	Ending Station Coordinates (Latitude)
end_lng	FLOAT	Ending Station Coordinates (Longitude)
member_casual	STRING	User Type (Annual Member or Casual User)
ride_length	TIME	The numbers of times that user rided
day_of_week	INTEGER	The day that the customer used services , sorted by number from 1 to 7

P/S: Day_of_week explanation:

Day_of_week	Number
Monday	2
Tuesday	3
Wednesday	4
Thursday	5
Friday	6
Saturday	7
Sunday	1

Motivate International Inc. makes the data available under the license, which is located in https://www.divvybikes.com/data-license-agreement (https://www.divvybikes.com/data-license-agreement). The dataset contains no member-identifying information, thus privacy is not a concern (no names or credit card info). The data looks to be reliable and consistent throughout, and Motivate's license provides preliminary security that their data gathering and organization is trustworthy.

Document of data cleaning and manipulation

R programming is mainly chosen during data import, processing, cleaning and analysis, all of which are covered below:

Set up the environment

1.Install and load the packages

Hide

```
install.packages("tidyverse")
```

```
WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate
version of Rtools before proceeding:

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into ‘C:/Users/ThuBy/OneDrive/Documents/R/win-library/4.1’
(as ‘lib’ is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/tidyverse_1.3.1.zip'
Content type 'application/zip' length 430311 bytes (420 KB)
downloaded 420 KB
```

```
package ‘tidyverse’ successfully unpacked and MD5 sums checked

The downloaded binary packages are in
    C:\Users\ThuBy\AppData\Local\Temp\RtmpgHOQnf\downloaded_packages
```

Hide

```
install.packages("lubridate")
```

```
WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate
version of Rtools before proceeding:

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into ‘C:/Users/ThuBy/OneDrive/Documents/R/win-library/4.1’
(as ‘lib’ is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/lubridate_1.8.0.zip'
Content type 'application/zip' length 1716189 bytes (1.6 MB)
downloaded 1.6 MB
```

```
package 'lubridate' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
    C:\Users\ThuBy\AppData\Local\Temp\RtmpgHOQnf\downloaded_packages
```

Hide

```
library(tidyverse)
```

```
Registered S3 methods overwritten by 'dbplyr':
  method         from
  print.tbl_lazy
  print.tbl_sql
-- Attaching packages --------------------------------------------------------- tidyverse 1.3.1 --
v ggplot2 3.3.5     v purrr   0.3.4
v tibble  3.1.5     v dplyr   1.0.7
v tidyr   1.1.4     v stringr 1.4.0
v readr   2.0.2     v forcats 0.5.1
-- Conflicts ------------------------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
```

Hide

```
library(ggplot2)
library(readr)
```

2.Setup my working directory

Hide

```
getwd()
```

```
[1] "D:/R thesis file/thesisDivvy"
```

3.Now our directory is the right place, we can easily access to targeted data

Import the files relating to the past 9 months of data in 2021

Hide

```
t1_2021 <- read_csv('202101-divvy-tripdata.csv')
```

```
Rows: 96834 Columns: 13
-- Column specification --------------------------------------------------------------------------
Delimiter: ","
chr  (7): ride_id, rideable_type, start_station_name, start_station_id, end_station_name, end_statio...
dbl  (4): start_lat, start_lng, end_lat, end_lng
dttm (2): started_at, ended_at

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Hide

```
t2_2021 <- read_csv('202102-divvy-tripdata.csv')
```

```
Rows: 49622 Columns: 13
-- Column specification --------------------------------------------------------------------------
Delimiter: ","
chr  (7): ride_id, rideable_type, start_station_name, start_station_id, end_station_name, end_statio...
dbl  (4): start_lat, start_lng, end_lat, end_lng
dttm (2): started_at, ended_at

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Hide

```
t3_2021 <- read_csv('202103-divvy-tripdata.csv')
```

```
Rows: 228496 Columns: 13
-- Column specification --------------------------------------------------------------------------------
Delimiter: ","
chr  (7): ride_id, rideable_type, start_station_name, start_station_id, end_station_name, end_statio...
dbl  (4): start_lat, start_lng, end_lat, end_lng
dttm (2): started_at, ended_at

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Hide

```
t4_2021 <- read_csv('202104-divvy-tripdata.csv')
```

```
Rows: 337230 Columns: 13
-- Column specification --------------------------------------------------------------------------------
Delimiter: ","
chr  (7): ride_id, rideable_type, start_station_name, start_station_id, end_station_name, end_statio...
dbl  (4): start_lat, start_lng, end_lat, end_lng
dttm (2): started_at, ended_at

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Hide

```
t5_2021 <- read_csv('202105-divvy-tripdata.csv')
```

```
Rows: 531633 Columns: 13
-- Column specification --------------------------------------------------------------------------------
Delimiter: ","
chr  (7): ride_id, rideable_type, start_station_name, start_station_id, end_station_name, end_statio...
dbl  (4): start_lat, start_lng, end_lat, end_lng
dttm (2): started_at, ended_at

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Hide

```
t6_2021 <- read_csv('202106-divvy-tripdata.csv')
```

```
Rows: 729595 Columns: 13
-- Column specification --------------------------------------------------------------------------------
Delimiter: ","
chr  (7): ride_id, rideable_type, start_station_name, start_station_id, end_station_name, end_statio...
dbl  (4): start_lat, start_lng, end_lat, end_lng
dttm (2): started_at, ended_at

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Hide

```
t7_2021 <- read_csv('202107-divvy-tripdata.csv')
```

```
Rows: 822410 Columns: 13
-- Column specification --------------------------------------------------------------------------------
Delimiter: ","
chr  (7): ride_id, rideable_type, start_station_name, start_station_id, end_station_name, end_statio...
dbl  (4): start_lat, start_lng, end_lat, end_lng
dttm (2): started_at, ended_at

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Hide

```
t8_2021 <- read_csv('202108-divvy-tripdata.csv')
```

```
Rows: 804352 Columns: 13
-- Column specification --------------------------------------------------------------------------------
Delimiter: ","
chr  (7): ride_id, rideable_type, start_station_name, start_station_id, end_station_name, end_statio...
dbl  (4): start_lat, start_lng, end_lat, end_lng
dttm (2): started_at, ended_at

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Hide

```
t9_2021 <- read_csv('202109-divvy-tripdata.csv')
```

```
Rows: 756147 Columns: 13
-- Column specification --------------------------------------------------------------------------------
Delimiter: ","
chr  (7): ride_id, rideable_type, start_station_name, start_station_id, end_station_name, end_statio...
dbl  (4): start_lat, start_lng, end_lat, end_lng
dttm (2): started_at, ended_at

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Convert ride_id and rideable_type to character so that they can stack correctly

Hide

```
t1_2021 <- mutate(t1_2021,ride_id = as.character(ride_id),rideable_type = as.character(rideable_type))
t2_2021 <- mutate(t2_2021,ride_id = as.character(ride_id),rideable_type = as.character(rideable_type))
t3_2021 <- mutate(t3_2021,ride_id = as.character(ride_id),rideable_type = as.character(rideable_type))
t4_2021 <- mutate(t4_2021,ride_id = as.character(ride_id),rideable_type = as.character(rideable_type))
t5_2021 <- mutate(t5_2021,ride_id = as.character(ride_id),rideable_type = as.character(rideable_type))
t6_2021 <- mutate(t6_2021,ride_id = as.character(ride_id),rideable_type = as.character(rideable_type))
t7_2021 <- mutate(t7_2021,ride_id = as.character(ride_id),rideable_type = as.character(rideable_type))
t8_2021 <- mutate(t8_2021,ride_id = as.character(ride_id),rideable_type = as.character(rideable_type))
t9_2021 <- mutate(t9_2021,ride_id = as.character(ride_id),rideable_type = as.character(rideable_type))
```

Stack individual quarter's data frames into one big data frame

Hide

```
all_2021 <- bind_rows(t1_2021,t2_2021,t3_2021,t4_2021,t5_2021,t6_2021,t7_2021,t8_2021,t9_2021)
```

Check the columns of the dataset

Hide

```
colnames(all_2021)
```

```
 [1] "ride_id"            "rideable_type"      "started_at"         "ended_at"
 [5] "start_station_name" "start_station_id"   "end_station_name"   "end_station_id"
 [9] "start_lat"          "start_lng"          "end_lat"            "end_lng"
[13] "member_casual"
```

Add a new column : "ride_length" (in minutes)

Hide

```
all_2021$ride_length <- difftime(all_2021$ended_at,all_2021$started_at, units = "mins")
```

Do some summary check about our new dataset : all_2021

Hide

```
summary(all_2021)
```

```
   ride_id          rideable_type        started_at                    ended_at
 Length:4356319     Length:4356319     Min.   :2021-01-01 00:02:05   Min.   :2021-01-01 00:08:39
 Class :character   Class :character   1st Qu.:2021-05-23 16:11:48   1st Qu.:2021-05-23 16:42:05
 Mode  :character   Mode  :character   Median :2021-07-08 17:25:10   Median :2021-07-08 17:43:02
                                       Mean   :2021-07-01 07:12:39   Mean   :2021-07-01 07:36:01
                                       3rd Qu.:2021-08-18 21:33:27   3rd Qu.:2021-08-18 21:55:50
                                       Max.   :2021-09-30 23:59:48   Max.   :2021-10-01 22:55:35

 start_station_name start_station_id   end_station_name   end_station_id       start_lat
 Length:4356319     Length:4356319     Length:4356319     Length:4356319     Min.   :41.64
 Class :character   Class :character   Class :character   Class :character   1st Qu.:41.88
 Mode  :character   Mode  :character   Mode  :character   Mode  :character   Median :41.90
                                                                            Mean   :41.90
                                                                            3rd Qu.:41.93
                                                                            Max.   :42.07

   start_lng         end_lat         end_lng       member_casual      ride_length
 Min.   :-87.84   Min.   :41.51   Min.   :-88.07   Length:4356319     Length:4356319
 1st Qu.:-87.66   1st Qu.:41.88   1st Qu.:-87.66   Class :character   Class :difftime
 Median :-87.64   Median :41.90   Median :-87.64   Mode  :character   Mode  :numeric
 Mean   :-87.65   Mean   :41.90   Mean   :-87.65
 3rd Qu.:-87.63   3rd Qu.:41.93   3rd Qu.:-87.63
 Max.   :-87.52   Max.   :42.17   Max.   :-87.49
                  NA's   :3952    NA's   :3952
```

As we can see that, in the summary of overall Divvy data from the beginning of January 2021 to the end of September 2021, there are some main data we should focus on, particularly member_casual and ride_length. Therefore, we will learn more about the two data columns above. However, in ride_length column, it is necessary to do a convertion for the calculation.

Data check for specific columns

Hide

```
table(all_2021$member_casual)
```

```
 casual  member
2095096 2261223
```

Convert “ride_length” from factor to numeric so we can run calculations on the data

Hide

```
is.factor(all_2021$ride_length)
```

```
[1] FALSE
```

Hide

```
all_2021$ride_length <- as.numeric(as.character(all_2021$ride_length))
```

Check it again about its mode

Hide

```
is.numeric(all_2021$ride_length)
```

```
[1] TRUE
```

Data check for ride_length

Hide

```
summary(all_2021$ride_length)
```

```
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
   -7.05     7.25    12.85    23.36    23.22 55944.15
```

The above result shows the inconsistency of the ride_length data, because all Divvy bike rides must be greater than or = 0, so all numbers below zero (< 0) need to be removed to keep the data consistent.

Data cleaning : eliminate some negative values and any trips that were below 60 seconds in length should be removed because the possible incorrect starts or people attempting to re-dock a bike to assure its security

Hide

```
all_2021a <- all_2021[!(all_2021$ride_length<1),]
```

Check the data cleaning process

Hide

```
summary(all_2021a$ride_length)
```

```
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max. 
    1.00     7.47    13.07    23.71    23.45 55944.15 
```

As demonstrated above, the data has been cleaned.

Check the data again for the steady data before analysis

Hide

```
summary(all_2021a)
```

```
  ride_id          rideable_type        started_at                   ended_at                  
 Length:4290839     Length:4290839     Min.   :2021-01-01 00:02:05   Min.   :2021-01-01 00:08:39  
 Class :character   Class :character   1st Qu.:2021-05-23 15:58:44   1st Qu.:2021-05-23 16:30:17  
 Mode  :character   Mode  :character   Median :2021-07-08 17:49:36   Median :2021-07-08 18:07:24  
                                       Mean   :2021-07-01 07:27:01   Mean   :2021-07-01 07:50:43  
                                       3rd Qu.:2021-08-18 22:01:42   3rd Qu.:2021-08-18 22:23:05  
                                       Max.   :2021-09-30 23:59:48   Max.   :2021-10-01 22:55:35  

 start_station_name start_station_id   end_station_name   end_station_id       start_lat     
 Length:4290839     Length:4290839     Length:4290839     Length:4290839     Min.   :41.64  
 Class :character   Class :character   Class :character   Class :character   1st Qu.:41.88  
 Mode  :character   Mode  :character   Mode  :character   Mode  :character   Median :41.90  
                                                                            Mean   :41.90  
                                                                            3rd Qu.:41.93  
                                                                            Max.   :42.07  

   start_lng         end_lat         end_lng       member_casual       ride_length       
 Min.   :-87.84   Min.   :41.51   Min.   :-88.07   Length:4290839     Min.   :    1.00  
 1st Qu.:-87.66   1st Qu.:41.88   1st Qu.:-87.66   Class :character   1st Qu.:    7.47  
 Median :-87.64   Median :41.90   Median :-87.64   Mode  :character   Median :   13.07  
 Mean   :-87.65   Mean   :41.90   Mean   :-87.65                      Mean   :   23.71  
 3rd Qu.:-87.63   3rd Qu.:41.93   3rd Qu.:-87.63                      3rd Qu.:   23.45  
 Max.   :-87.52   Max.   :42.17   Max.   :-87.49                      Max.   :55944.15  
                  NA's   :3923    NA's   :3923                                          
```

Create more date types columns for further assessment

Hide

```
all_2021a$date <- as.Date(all_2021a$started_at) #The default format is yyyy-mm-dd
all_2021a$month <- format(as.Date(all_2021a$date), "%m")
all_2021a$day <- format(as.Date(all_2021a$date), "%d")
all_2021a$year <- format(as.Date(all_2021a$date), "%Y")
all_2021a$day_of_week <- format(as.Date(all_2021a$date), "%A")
```

Check these columns

Hide

```
table(all_2021a$month)
```

```
   01    02    03    04    05    06    07    08    09
95415 48649 225523 332422 523243 717699 809760 792943 745185
```

As it shows, month columns are displayed with numbers, instead of letters. So it needs editing. ### Modify the day_of_week column to the correct letter

Hide

```
all_2021a$day_of_week <- ordered(all_2021a$day_of_week, levels=c("Sunday", "Monday",
"Tuesday", "Wednesday", "Thursday", "Friday", "Saturday"))
```

Modify year column to complete words

Hide

```
all_2021a <- all_2021a %>% mutate(month = recode(month
                                                 ,"01"="Jan"
                                                 ,"02"="Feb"
                                                 ,"03"="Mar"
                                                 ,"04"="Apr"
                                                 ,"05"="May"
                                                 ,"06"="Jun"
                                                 ,"07"="Jul"
                                                 ,"08"="Aug"
                                                 ,"09"="Sep"))
```

Carry out some checks

Hide

```
table(all_2021a$month)
```

```
  Apr   Aug   Feb   Jan   Jul   Jun   Mar   May   Sep
332422 792943  48649  95415 809760 717699 225523 523243 745185
```

Hide

```
table(all_2021a$day_of_week)
```

```
  Sunday    Monday   Tuesday Wednesday  Thursday    Friday  Saturday
  677953    540502    551494    565154    569377    616149    770210
```

Modify the order of the month

Hide

```
all_2021a$month <- ordered(all_2021a$month, levels=c("Jan", "Feb", "Mar", "Apr", "May", "Jun", "Jul", "Aug", "Sep"))
```

###Re-check the month column

Hide

```
table(all_2021a$month)
```

```
  Jan   Feb   Mar   Apr   May   Jun   Jul   Aug   Sep
95415 48649 225523 332422 523243 717699 809760 792943 745185
```

All data appears to have been edited and sorted in preparation for analysis.

Data analysis

Check rideable_type

Hide

```
table(all_2021a$rideable_type)
```

```
 classic_bike   docked_bike electric_bike
      2645236        275190       1370413
```

There are 3 classes of bikes : classic, docked and electric ones. As the number illustrates, people mainly used classic and electric ones.

Execute some analysis related to ride length of users

Hide

```
summary(all_2021a$ride_length)
```

```
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
    1.00     7.47    13.07    23.71    23.45 55944.15
```

Take a ride length column with decreasing values

Hide

```
dv_ridelength <- order(all_2021a$ride_length, decreasing = T)
```

Check top 10 rows with decreasing value of ride length

Hide

```
head(all_2021a[dv_ridelength,],10)
```

ride_id <chr>	rideable_type <chr>	started_at <S3: POSIXct>	ended_at <S3: POSIXct>
F043F0F6A1AA4F85	docked_bike	2021-06-05 02:27:26	2021-07-13 22:51:35
7F0578ABF030FC83	docked_bike	2021-06-04 22:03:33	2021-07-13 14:15:14
BDA1217EC8532C7B	docked_bike	2021-05-02 02:56:07	2021-06-08 13:37:43

ride_id <chr>	rideable_type <chr>	started_at <S3: POSIXct>	ended_at <S3: POSIXct>
E6E1E5BFFD72ECEC	docked_bike	2021-06-05 23:33:51	2021-07-12 13:55:14
99415B003B7E7EEE	docked_bike	2021-07-08 19:29:49	2021-08-11 21:56:58
E84DF812305C9C9F	docked_bike	2021-04-02 17:50:00	2021-05-05 22:06:42
8016DC50082D9E93	docked_bike	2021-06-05 21:47:40	2021-07-08 13:18:31
89F33397EABB55E7	docked_bike	2021-07-08 15:13:08	2021-08-06 13:18:39
E25E58F5E94EE351	docked_bike	2021-08-01 18:53:10	2021-08-30 16:42:20
D4E35EE211346690	docked_bike	2021-07-10 15:59:21	2021-08-07 22:43:57

1-10 of 10 rows | 1-4 of 19 columns

Top 5 starting station names where riders start their trips

Hide

```
top_5_start_station_names <- sort(table(all_2021a$start_station_name), decreasing=TRUE)[1:5]
knitr::kable(top_5_start_station_names,
             col.names = c("Starting Station Name", "Number of Rides"),
             caption = "Top 5 Starting Stations (Jan 2021 - Sep 2021)")
```

```
Registered S3 methods overwritten by 'htmltools':
  method               from
  print.html           tools:rstudio
  print.shiny.tag      tools:rstudio
  print.shiny.tag.list tools:rstudio
```

Top 5 Starting Stations (Jan 2021 - Sep 2021)

Starting Station Name	Number of Rides
Streeter Dr & Grand Ave	72052
Michigan Ave & Oak St	38709

Starting Station Name	Number of Rides
Wells St & Concord Ln	34975
Millennium Park	34851
Theater on the Lake	32700

Top 5 ending station names where riders stop their trips

Hide

```
top_5_ending_station_names <- sort(table(all_2021a$end_station_name), decreasing=TRUE)[1:5]
knitr::kable(top_5_ending_station_names,
             col.names = c("Ending Station Name", "Number of Rides"),
             caption = "Top 5 Ending Stations (Jan 2021 - Sep 2021)")
```

Top 5 Ending Stations (Jan 2021 - Sep 2021)

Ending Station Name	Number of Rides
Streeter Dr & Grand Ave	72276
Michigan Ave & Oak St	39032
Millennium Park	35312
Wells St & Concord Ln	35219
Theater on the Lake	32937

Check top 5 station names via form of users

Hide

```
annual_member <- all_2021a[!(all_2021a$member_casual == "casual"),]
casual_user <- all_2021a[!(all_2021a$member_casual == "member"),]
top_5_member_start <- sort(table(annual_member$start_station_name), decreasing=TRUE)[1:5]
top_5_member_end <- sort(table(annual_member$end_station_name), decreasing=TRUE)[1:5]
top_5_casual_start <- sort(table(casual_user$start_station_name), decreasing=TRUE)[1:5]
top_5_casual_end <- sort(table(casual_user$end_station_name), decreasing=TRUE)[1:5]
knitr::kable(top_5_member_start,
             col.names = c("Starting Station Name", "Number of Rides"),
             caption = "Annual Members, Top 5 Starting Station Names (Jan 2021 - Sep 2021)")
```

Annual Members, Top 5 Starting Station Names (Jan 2021 - Sep 2021)

Starting Station Name	Number of Rides
Clark St & Elm St	18535
Wells St & Concord Ln	18206
Kingsbury St & Kinzie St	16775
Wells St & Elm St	16012
Dearborn St & Erie St	14551

Hide

```
knitr::kable(top_5_member_end,
             col.names = c("Ending Station Name", "Number of Rides"),
             caption = "Annual Members, Top 5 Ending Station Names (Jan 2021 - Sep 2021)")
```

Annual Members, Top 5 Ending Station Names (Jan 2021 - Sep 2021)

Ending Station Name	Number of Rides
Wells St & Concord Ln	18746
Clark St & Elm St	18738
Kingsbury St & Kinzie St	16976
Wells St & Elm St	16585

Ending Station Name	Number of Rides
Dearborn St & Erie St	15122

Hide

```
knitr::kable(top_5_casual_start,
             col.names = c("Starting Station Name", "Number of Rides"),
             caption = "Casual Users, Top 5 Starting Station Names (Jan 2021 - Sep 2021)")
```

Casual Users, Top 5 Starting Station Names (Jan 2021 - Sep 2021)

Starting Station Name	Number of Rides
Streeter Dr & Grand Ave	58472
Millennium Park	28346
Michigan Ave & Oak St	26628
Shedd Aquarium	19519
Lake Shore Dr & Monroe St	19413

Hide

```
knitr::kable(top_5_casual_end,
             col.names = c("Ending Station Name", "Number of Rides"),
             caption = "Casual Users, Top 5 Ending Station Names(Jan 2021 - Sep 2021)")
```

Casual Users, Top 5 Ending Station Names(Jan 2021 - Sep 2021)

Ending Station Name	Number of Rides
Streeter Dr & Grand Ave	60153
Millennium Park	29009
Michigan Ave & Oak St	27766
Theater on the Lake	20508

Ending Station Name	Number of Rides
Shedd Aquarium	18205

Compare members and casual users by ride length

Hide

```
aggregate(all_2021a$ride_length ~ all_2021a$member_casual, FUN = mean)
```

all_2021a$member_casual <chr>	all_2021a$ride_length <dbl>
casual	33.59217
member	14.52029

2 rows

Extract data to .csv file for additional exploration

Hide

```
write.csv(all_2021a,"D:\\R thesis file\\thesisDivvy\\all2021data.csv")
```

Finish the progress.