Cyclistic Bike share Case study

Introduction

In this case study, I am going to perform the real-world tasks of a marketing junior data analyst of a fictional company, Cyclistic. Going to follow the six steps of the data analysis process: ask, prepare, process, analyse, share and act in order to answer the key business questions.

Scenario

Cyclistic offers a bike sharing program that features more than 5800 bicycles and 600 docking stations. It has two types of customers. Customers who purchase single-ride or full day passes are casual riders and customers who purchase annual memberships are Cyclistic members.

Cyclistic financial analysts' team have concluded that annual members are more profitable than casual riders. Lily Moreno, the director of marketing believes maximizing the number of annual members will be key to future growth and there is a very good chance to convert casual riders into annual members through marketing strategies.

Data Analysis Process

This report will have the following deliverables:

1. A clear statement of the business task (Ask)
2. A description of all data sources used (Prepare)
3. Documentation of any cleaning or manipulation of data (Process)
4. A summary of your analysis (Analyse)
5. Supporting visualizations and key findings (Share)
6. Your top three recommendations based on your analysis (Act)

Ask

Guiding questions

- What is the problem you are trying to solve?
  The three questions of the future marketing program are
  How do annual members and casual riders use Cyclistic bikes differently?
  Why would casual riders buy Cyclistic annual memberships?
  How can Cyclistic use digital media to influence casual riders to become members?

I have been assigned to answer the first question: How do annual members and casual riders use Cyclistic bikes differently.

- How can your insights drive business decisions?

Based on those insights, the team will design a new marketing strategy to convert casual riders into annual members.

Key tasks

1. Identify the business task

How do annual members and casual riders use Cyclistic bikes differently.

2. Consider key stakeholders

Key stakeholders are the director of marketing: Lily Moreno, marketing analytics team, who will design the new marketing strategy and executive team, who will approve the marketing program.

Prepare

Guiding questions

- Where is your data located?

  This is a public data made available by Motivate international Inc. The link is given by the course material.
- How is the data organized?

  The data has been organized in monthly format as .zip and .csv files.
- Are there issues with bias or credibility in this data? Does your data ROCCC?

  No issues with bias and credibility as it is collected directly by the company and it involves the whole population not just samples. And the data is Reliable, Original, Comprehensive, Current and Cited.
- How are you addressing licensing, privacy, security, and accessibility?

  These are data are licensed and doesn't have any personal identifiable information.
- How did you verify the data's integrity?

  As the data is collected from the original source, it is accurate and consistent hence maintaining the data integrity.
- How does it help you answer your question?

  The data contains the information about the usage of the Cyclistic bike by annual members and casual riders. Hence it will help to answer my question.
- Are there any problems with the data?

  There are some minor issues with the data like missing fields, duplicate records etc.

Deliverable

The data used in this analysis are past 12 months Cyclistic trip data (Oct 2021-Sep 2022) downloaded from the link given in the course material which are made available by Motivate International Inc.

Process

Guiding questions

- What tools are you choosing and why?

  I choose SQL for this step as data is huge.
- Have you ensured your data's integrity?

  As mentioned earlier, the data is collected from the original source, it is accurate and consistent hence maintaining the data integrity.
- What steps have you taken to ensure that your data is clean?

  Removed the duplicate entries of ride_id, removed the entries where ride_length_in_minutes are less than zero.

● How can you verify that your data is clean and ready to analyse?

Verified the data types of all column. Checked for null values and deleted the entries containing null value. Removed the duplicate entries. Removed the data which are unrealistic. Verified the strings for any error in the columns which can be grouped.

● Have you documented your cleaning process so you can review and share those results?

Yes.

Deliverable

Big query SQL is used for this step.

1. Imported all the past 12 months of data into big query.
2. Created a new table with the column similar to the individual monthly table.
3. Merged all 12 months data using and inserted into the new table.
4. Verified the new table with the monthly data by the count of total rows.
5. Calculated ride_length and added as a new column as asked in the assignment.
6. Calculated day of week and month and added as a new column.
7. Data cleaning:
    - Deletion of duplicate ride_ids
    - Deletion of unrealistic data i.e., ride_length <0
    - Verification of null values in columns
    - Verification of string value for any misspelling/typo error
    - Creation of new table deleting null values in station name for analysis involving station details

- The SQL queries are attached here: SQL queries

Analyse

Guiding questions

● How should you organize your data to perform analysis on it?

I have organized the whole data into a single table to perform analysis and cleaned the data for inaccuracies and inconsistencies.

● Has your data been properly formatted?

Yes.

● What surprises did you discover in the data?

No. of rides by the members are greater than that of the casual riders. But the ride_length and average ride_length is greater for the casual riders than that of the members.

● What trends or relationships did you find in the data?

1. No. of rides: Members are having more no of rides than casual riders.
2. Ride length: Casual riders are having more ride length than members.
3. No. of rides based on ride length: casual riders have more rides greater than 20 minutes And members have more rides between 5 to 10 minutes.

4. Usage of bike throughout the week: Usage of bike by members are more than that of the casual riders during weekdays and vice versa during weekends

5. Usage of bike throughout the year: Usage of bike is more during summer than winter for both riders.

6. Rideable types: Casual riders use more of electric bikes and members use more of classic bikes.

7.Round trips: No. of round trips by casual riders more than that of members.

8.Popular station/path: One start station or one path is more popular (having more no. of rides) among casual riders.

- How will these insights help answer your business questions
  These trends show how the members and casual riders use the bike differently.

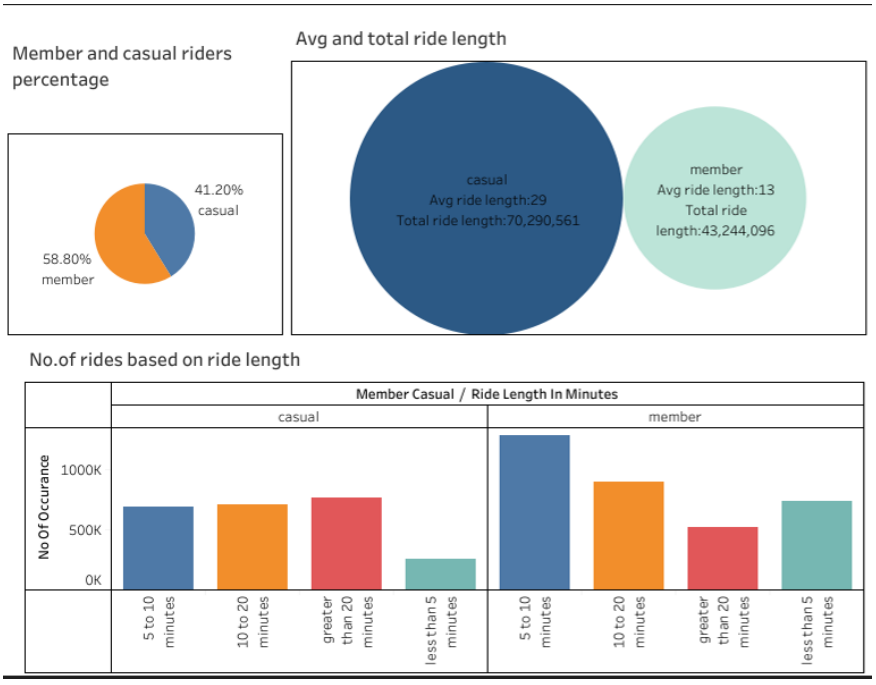
Share

Tableau dashboard is attached here: Visualizations in Tableau

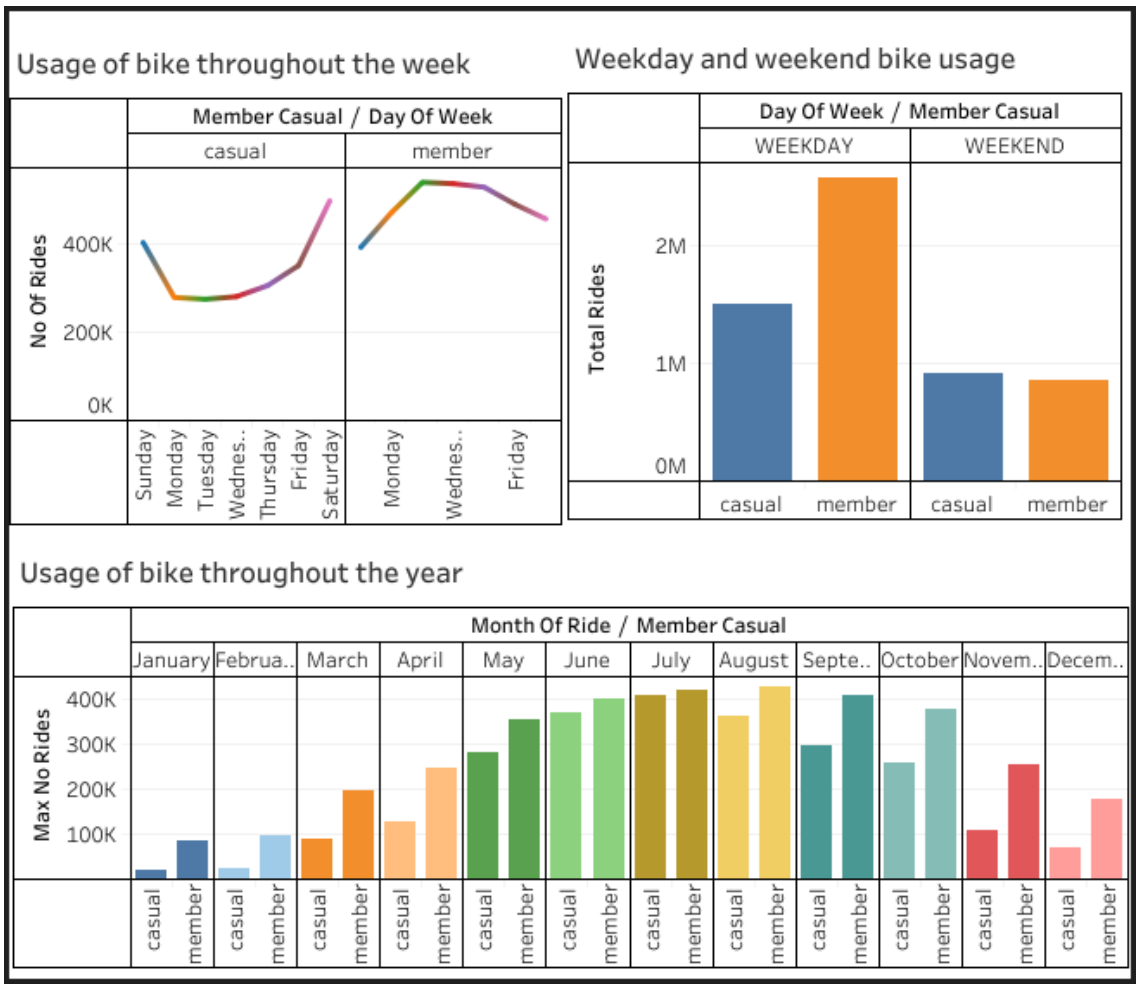
Guiding questions

- Were you able to answer the question of how annual members and casual riders use Cyclistic bikes differently?
  Yes.
- What story does your data tell?
  It shows how the members and casual riders are using the bike differently in many varied aspects like no of rides, ride length, rideable type, usage of bike throughout the week and year, round trips and most used station and path.
- How do your findings relate to your original question?
  The findings show how annual members and casual riders differ in the bike use in many different ways.
- Who is your audience? What is the best way to communicate with them?
  The audience are Lily Moreno, the director of marketing, Cyclistic marketing analytics team and Cyclistic executive team.
- Can data visualization help you share your findings?
  Yes. Data visualization are really helpful to understand and communicate the findings easily.
- Is your presentation accessible to your audience?
  The presentation is an interactive dashboard of tableau which is accessible by the audience.

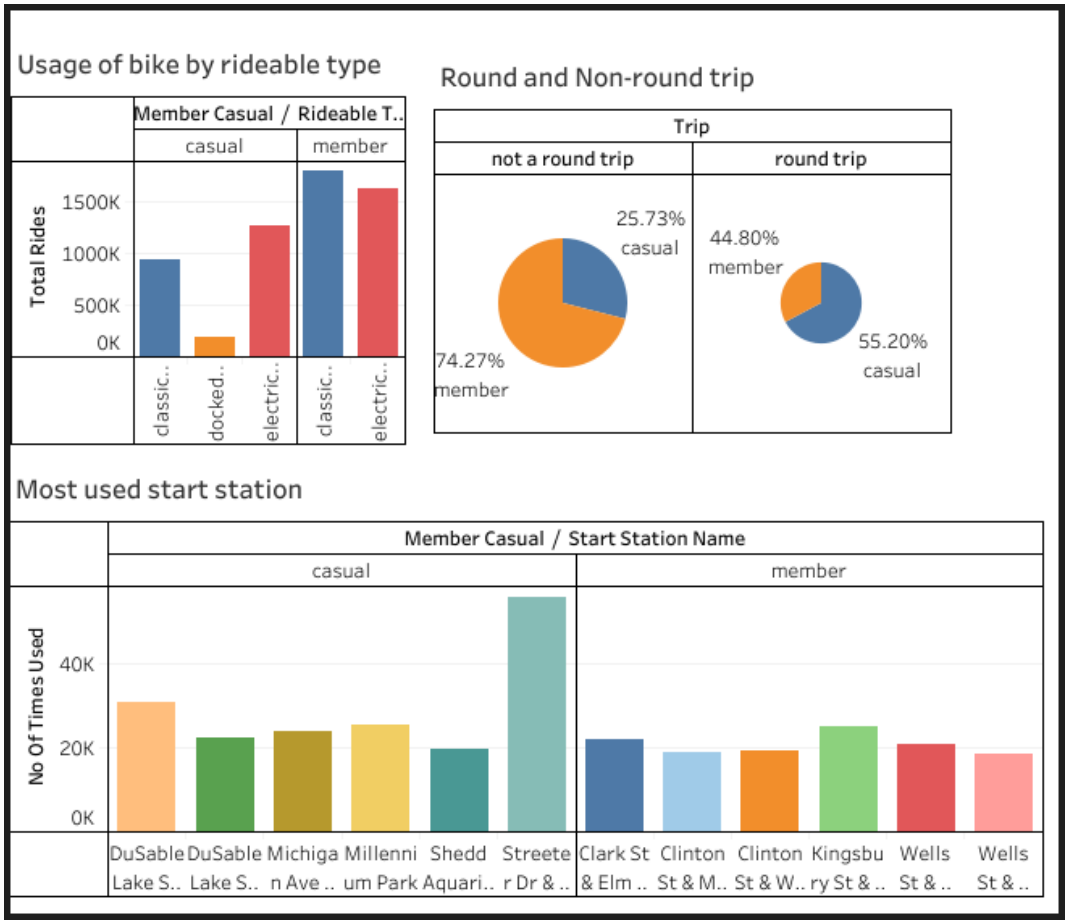
Deliverable



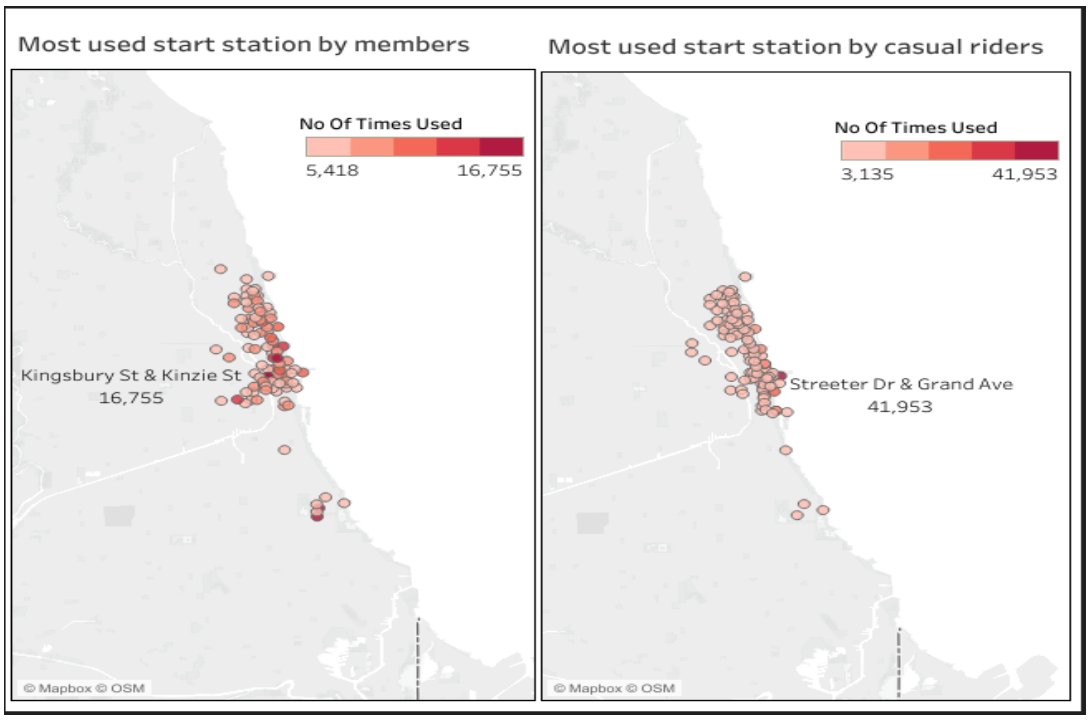
1.Total no. of rides by members account for 58.80% while casual riders account for 41.20%.
2.Total ride length of casual riders is around 70M and of member is 43M. Average ride length of casual riders is 29 minutes and of member is 13 minutes.
3. Casual riders usually have more rides in 'greater than 20 minutes' category. Members have more rides in '5 -10 minutes' category.



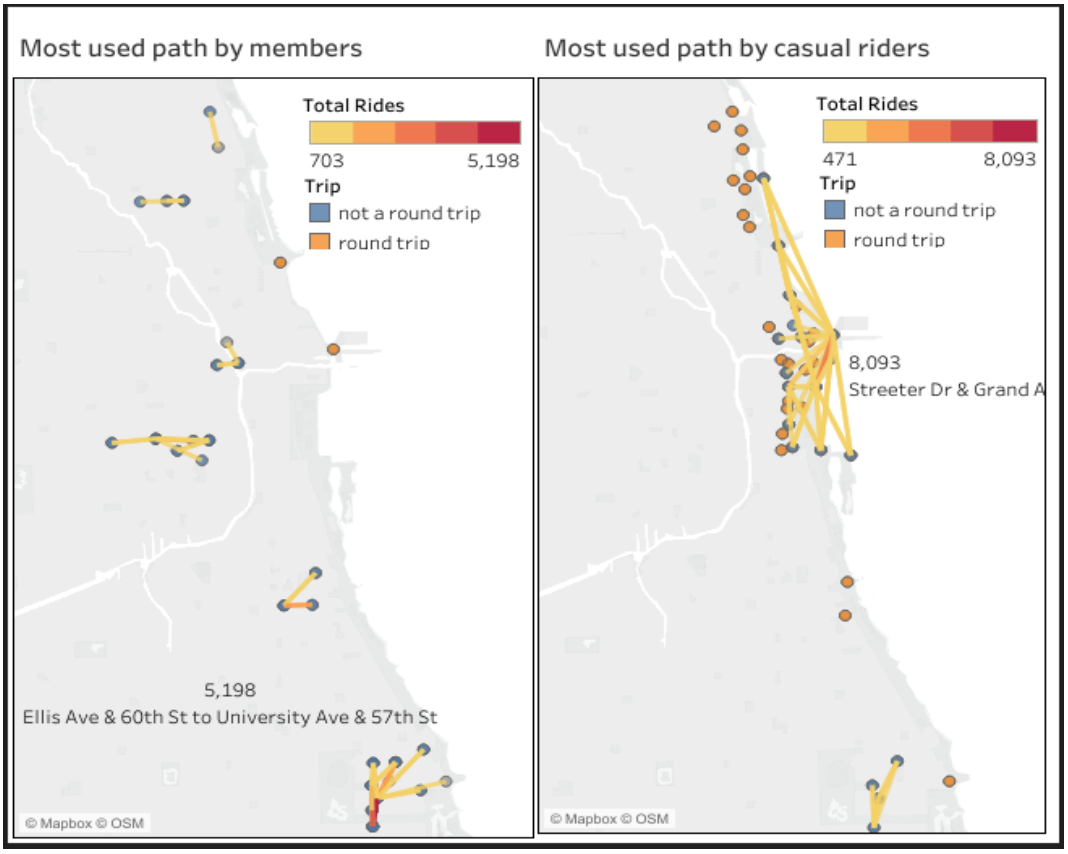
4.Members use the bike more on weekdays while casual riders use the bike more on weekends.
5.The usage of bike by both members and casual riders are more during the summer season than on the winter season.



6.Casual riders use the electric bikes more than the classic bikes. Members use the classic bikes more than the electric bikes.
7.Members take more of 'not a round trips'(one way trip) while casual riders take more of 'round trips'.



8.The most used start station of casual riders ‘Streeter Dr & Grand Ave’ accounts for around 42k rides while that of member ‘Kingsbury St & Kinzie St’ accounts for around 17k rides.



9.The most used path by casual riders is a round trip to and from ‘Streeter Dr & Grand Ave’ with around 8K rides. And the most used path by members is from ‘Ellis Ave & 60th St to University Ave & 57th St’ with around 5k rides.

Act

Guiding Questions:

● What is your final conclusion based on your analysis?

Members and casual riders use the bike differently in many ways. Members are usually taking short rides which are more of ‘not a round trip’. They are using the bike throughout the week especially during weekdays. Hence, members use the bike as a transportation for the office trips.

Where as casual riders are usually taking long rides and ‘round trip’. They use the bike mainly during the weekends and most of them are using a single station/path for their rides. Hence, casual riders use the bike for leisure.

● How could your team and business apply your insights?

By knowing how the users use the bike differently according to their needs, it will be helpful to make the decisions on the marketing strategies and target audience.

● What next steps would you or your stakeholders take based on your findings?

The team will decide on the marketing strategies and the ways to implement it.

● Is there additional data you could use to expand on your findings?

The answer and analysis for the other two questions of the director of marketing would help in the final conclusion.

Deliverable:

The major difference between members and casual riders in bike usage are

1. Members take a greater number of rides though the ride length is short.
2. Members use the bike almost equally the throughout the week i.e., they are consistent in using the bike.
3. The usage of station and path by members are wide spread whereas casual riders use one station/path more compared to other stations/paths.