Google Data Analytics Capstone Project: How Does a Bike-Share Navigate Speedy Success?

By: Khanh Pham

This case study is one of a part of the Google Data Analytics Professional Certificate program on Coursera. The analysis will follow the 6 phases of the Data Analysis process: Ask, Prepare, Process, Analyze, Share, and Act.

Ask/Identify the task

In this case, I am a data analyst working for Cyclistic, which is a fictional company in Chicago. Since its inception in 2016, Cyclistic has grown to a fleet of 5,824 bicycles that are geotracked and locked into a network of 692 stations across Chicago. The bikes can be unlocked from one station and returned to any other station in the system anytime. There are two types of customers: casual riders (Casual) and annual members (Member).

Lily Moreno, the director of marketing and your manager has set a clear goal: **Design marketing strategies aimed at converting casual riders into annual members.** I have been tasked to analyze how annual members and casual riders use Cyclistic bikes differently based on a bike-share company's data of its customer's trip details from Jan-Oct 2021 to help the company maximize the number of annual memberships. To tackle this assignment, I will use Excel, SQL, and Tableau to analyze the Cyclistic historical bike trip data to identify trends by answering 6 main questions below:

1. What is the total number of trips for users, and what percentage of total trips does each user represent?

2. What is the average riding time and are there particular days of the week?

3. What is the average number of trips taken every month?

4. What is the busiest hour of the day?

5. What are the top 5 common starting and ending stations for each?

6. Is there a preference for the rideable type for members and casual?

Prepare

The dataset used in this case study is actual public data made by Motivate International Inc. under this license. The data is located on the Cyclistic server and is stored by separated files for each month. Because this is a case study using public data, we are going to assume the data is credible. However, I found that start_station_name/end_station_name have multiple NULL values. I downloaded the data from the divvy trip data and saved it on the desktop. The data contains 13 columns.

ride_id: ID rider

ideable_type: the type of bicycle used

started_at: the date and time that the bicycle was checked out

ended_at: the date and time that the bicycle was checked in

start_station_name: where the trip started

start_station_id: a unique identifier for the start station

end_station_name: where the trip ended

end_station_id : a unique identifier for the end station

Start_lat: the latitude of the start

stationstart_lng: the longitude of the start station

end_lat: the latitude of the end station

end_lng: the longitude of the end station

member_casual: a field indicating whether it was taken by casual riders or annual member
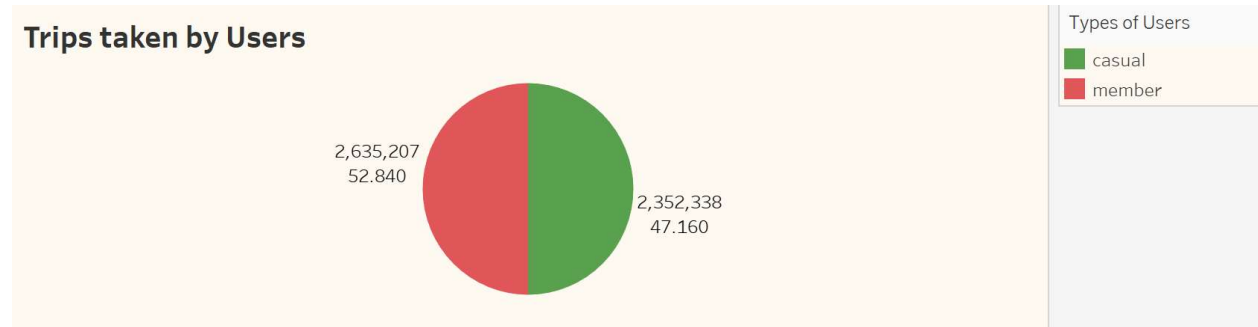
Process and Analyze

In this step, I am going to process the data and prepare it for the next step where I will find the answer to the questions. Tools that I am going to use are SQL for data cleaning and data analyzing and Tableau for data visualization. Please click the following link to see the dashboard that I created for this project.

First, after ensuring that all columns names are the same, then I merged 12 datasets into a unionized dataset by using SQL with Microsoft SQL on my laptop. I excluded 4 columns Start_lat, start_lng, end_lat, and end_lng because they are not necessarily used during the analyzing process.

Some notes about the data:

1. The dataset contains 4,987,545 rows.
2. 564,456 rows have null values in the start_station_name and 606,485 rows have null values in the end_station_name. These rows will be still be used in the overall analysis but will be filtered out when analyzing the top 5 start and end stations' names.
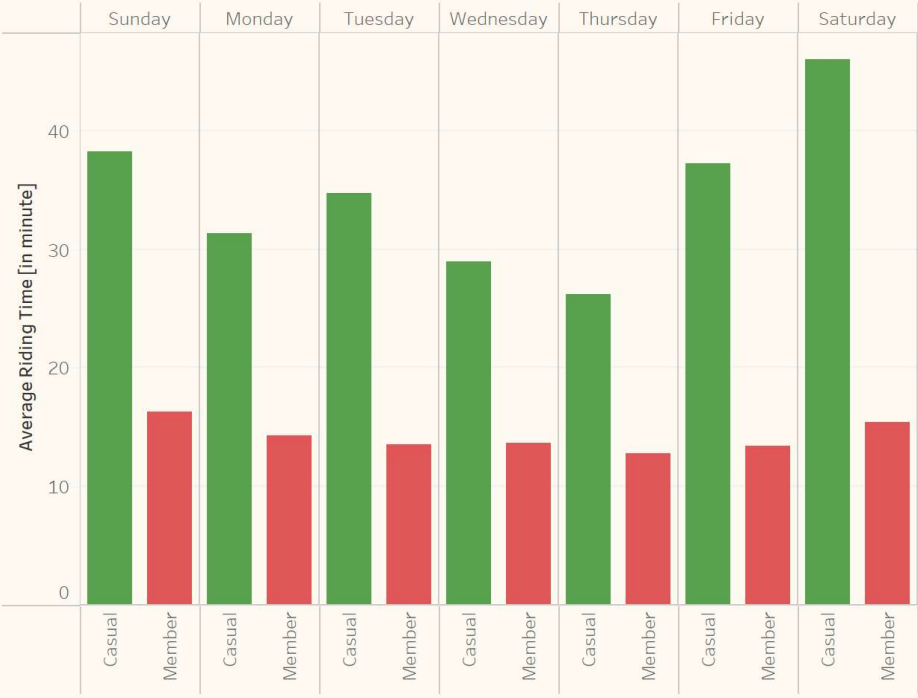
There are a total of 4,987,545 trips taken during the period Jan-Oct 2021. We can see the pie chart below showing the total number of trips taken by annual members is a little more than a half of the total number of trips taken by casual riders.

Trips taken by Users

Types of Users
- casual
- member

2,635,207
52.840

2,352,338
47.160

The observation here is that members tend to take short trips than casuals. In addition, casuals tend to rent the bikes mostly on weekends (Saturdays and Sundays) while members' usage is evenly distributed throughout the week.

There is a significant increase in trips taken during summertime when the weather is warmer.

Average Trip Duration

	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday

Average Riding Time [in minute]

40

30

20

10

0

Casual / Member (repeated for each day)

Average Trips Taken Monthly

	January	February	March	April	May	June	July	August	Septem..	October

Number of Trips Taken

440K
400K
360K
320K
280K
240K
200K
160K
120K
80K
40K
0K

Casual / Member (repeated for each month)
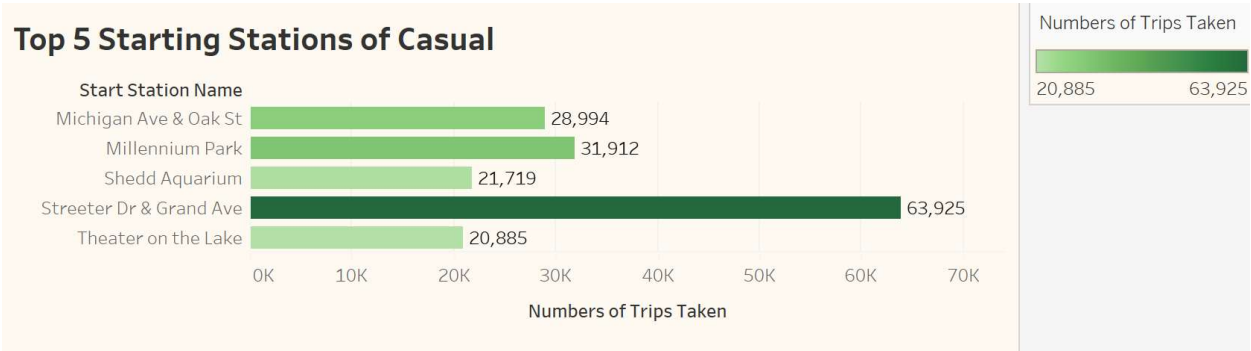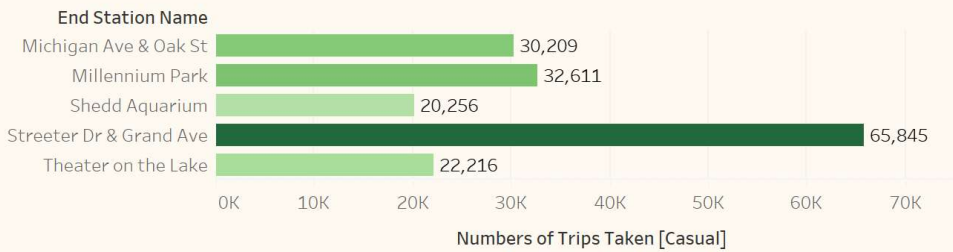
Both member and casual users' lines display peak hours during the day at 5:00 PM. The busy hours normally occur in the afternoon between 4:00 PM and 7:00 PM. The thickness of the lines increases when the number of trips increases.
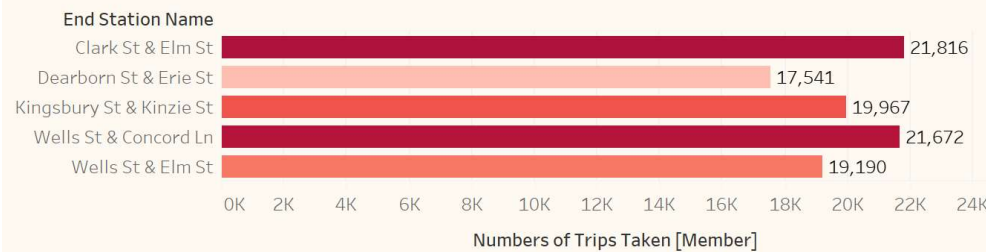
Hourly analysis

The four graphs below show the top 5 starting/ending stations of Casual and Member. The locations visited by casual users mostly are nearby downtown and are tourist locations.

Top 5 Starting Stations of Casual

Start Station Name	Numbers of Trips Taken
Michigan Ave & Oak St	28,994
Millennium Park	31,912
Shedd Aquarium	21,719
Streeter Dr & Grand Ave	63,925
Theater on the Lake	20,885

Numbers of Trips Taken: 20,885 — 63,925

Top 5 Starting Stations of Member

Start Station Name	Numbers of Trips Taken [Member]
Clark St & Elm St	21,631
Dearborn St & Erie St	17,016
Kingsbury St & Kinzie St	19,743
Wells St & Concord Ln	21,069
Wells St & Elm St	18,498

Numbers of Trips Taken: 17,016 — 21,631

Top 5 Ending Stations of Casual

Numbers of Trips Taken
20,256 — 65,845

End Station Name	Numbers of Trips Taken [Casual]
Michigan Ave & Oak St	30,209
Millennium Park	32,611
Shedd Aquarium	20,256
Streeter Dr & Grand Ave	65,845
Theater on the Lake	22,216

Top 5 Ending Stations of Member

Numbers of Trips Taken
17,541 — 21,816

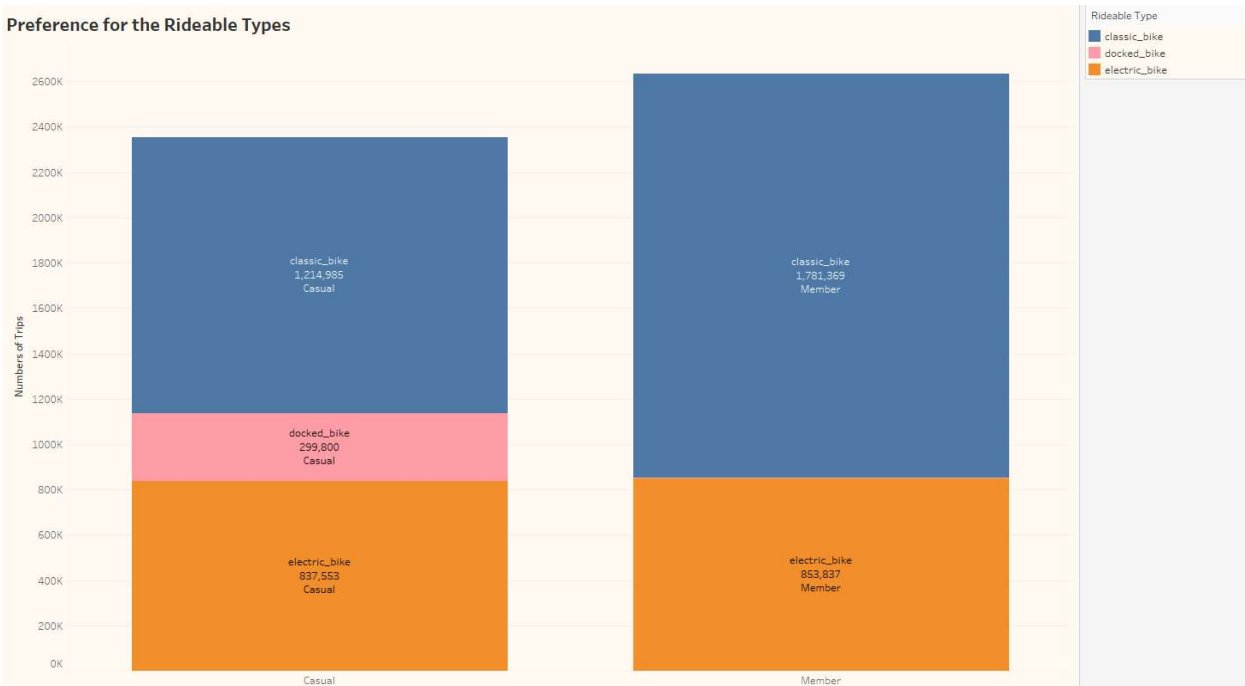| End Station Name	Numbers of Trips Taken [Member]
Clark St & Elm St	21,816
Dearborn St & Erie St	17,541
Kingsbury St & Kinzie St	19,967
Wells St & Concord Ln	21,672
Wells St & Elm St	19,190

Now, let's plot a bar graph that shows the likely distribution of the member and casual users with bike types. The classic bike is the most popular one among the 3 types. In contrast, the docked bike is the least popular one.

Preference for the Rideable Types

Rideable Type
classic_bike
docked_bike
electric_bike

classic_bike 1,214,985 Casual
docked_bike 299,800 Casual
electric_bike 837,553 Casual

classic_bike 1,781,369 Member
electric_bike 853,837 Member

Share and Act

After using SQL to communicate and analyze the dataset and Tableau for data storytelling, it is time to share the result and answer the questions "How can we convert casuals to members?"

1. We see that the season affects the total trip duration for casual riders and annual members. The temperature is very low during the winter season, people are not willing to go out as normal as spring or summer, and people who need to commute to work would rather choose other public transports like bus or train. Therefore, the idea of lowering the cost of the annual membership plan during winters will lure the riders to use for bike share over public transport.

2. Members tend to take short trips than casuals. However, casuals tend to rent the bikes mostly on weekends (Saturdays and Sundays) while members' usage is evenly distributed throughout the week. Therefore, we can increase renting a bike for the weekend to target casual riders into having a membership, also we should launch more special deals for annual members only to motivate casual riders to switch their plans like members have the option to book weekday rides at a lower rate.

3. Both annual member and casual riders prefer classic bikes. The docked bike is the least popular one, especially there is only one docked bike rented during all the period of analysis.

4. We know the most popular start/end station names and routes for casual users, we should put banners or special discount advertisements in those areas or routes that would target casual users.

5. The company can also run a survey to collect the data from current casual riders to understand what kind of benefits they look for when considering subscribing to annual members, and then create a plan to act based on the result of the survey.

For interactive Tableau Dashboard, please click here.

For SQL queries, please click here.