Google Data analytics Course Case Study

Leveraging Cyclistic bike-share Success

By: Amira Salama

Nov 2021

Contents

Google Data analytics Course Case Study

Summary :

Cyclistic, a bike-share company in **Chicago**, features more than 5,800 bicycles and +600 docking stations offering reclining bikes, hand tricycles, and cargo bikes, making bike-share more inclusive to people with disabilities and riders who can't use a standard two-wheeled bike.

Background on Data

- About 8% of riders use the assistive options
- About 30% use them to commute to work each day
- Lily Moreno: The director of marketing and your manager, believes the company's future success depends on **maximizing** the number of annual memberships
- Customers who purchase single-ride or full-day passes are referred to as **casual** riders
- Customers who purchase annual memberships are Cyclistic **members**
- pricing plans: single-ride passes, full-day passes, and annual memberships.

1 - Main Business task statement :

-Convert Casual Users/riders into Members by analyzing the behavioral differences between casual and member riders

Marketing program Purpose:

1. How do annual members and casual riders use Cyclistic bikes differently?
2. Why would casual riders buy Cyclistic annual memberships?
3. How can Cyclistic use digital media to influence casual riders to become members?

Google Data analytics Course Case Study

2- Data sources description :

- Downloaded 12 months of data started from January 2020 to December 2020
- Used Microsoft Excel to initially view the data
- RStudio to analyze , visualize and convert data
- Tableau software for dashboards and final visualizations
- Internet research for ridesharing market size and incentives to leverage its success

3 - Documentation of Cleaning or manipulation of data:

1 - combined all data of 2020 months files using code, that resulted in +3.5 million observations
(The data contained more than 3.542 million rows of observations + 13 variables/ columns)
2- This Data was Too big for Spreadsheets, so I used RStudio to import and analyze the data
2- checked data for null values, duplicates, and anomalies
3- Assessed data types integrity and made the necessary changes :

- Start date and end date was converted into date time data type
- Added columns that list the date, month, day, and year of each ride to better analyze the data at a granular level
- Added a calculated column (ride length) to calculate the duration of each ride by minutes
- Trimmed the data frame to exclude -ve rides, and trips starting at HQ QR (for maintenance purposes)
- Ordered the days of the week column to start at Sunday to compare average ride duration by day and total count of the rides
- Grouped data by month, days of the week to compare the different behavior of riders (members Vs casual riders)
- Grouped and concatenated the “start station name” column with the “end station name” then grouped them to query the most popular stations

4- Final export and visualization of clean data in Tableau

Google Data analytics Course Case Study

4- Summary of analysis :

1- Conducting descriptive statistics
Summary statistics of ride length in minutes

Min	Median	Average	Max
0.02	14.13	28.00	156,450.40

- Average overall ride length is 28 minutes
- Median 14 minutes

Comparing member and casual rider average ride duration

Member VS casual	Ride length minutes
casual	47.55651
member	15.74248

- Further breaking down of average ride length by member vs casual rider shows that the casual rider exceeded the average time of the member by 3 times

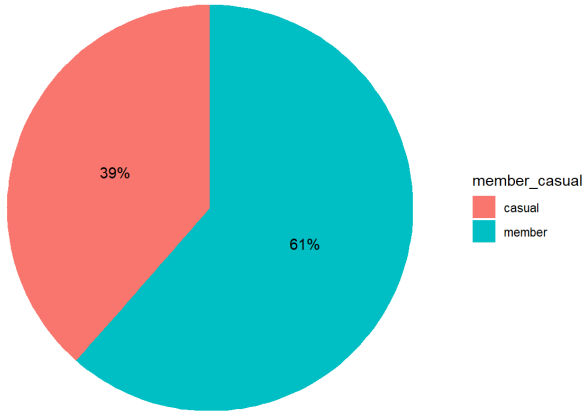
2- Total number of casual and member riders in 2020 :

Member Vs Casual	Count
casual	1,359,398
member	2,167,602
Total	**3,527,000**

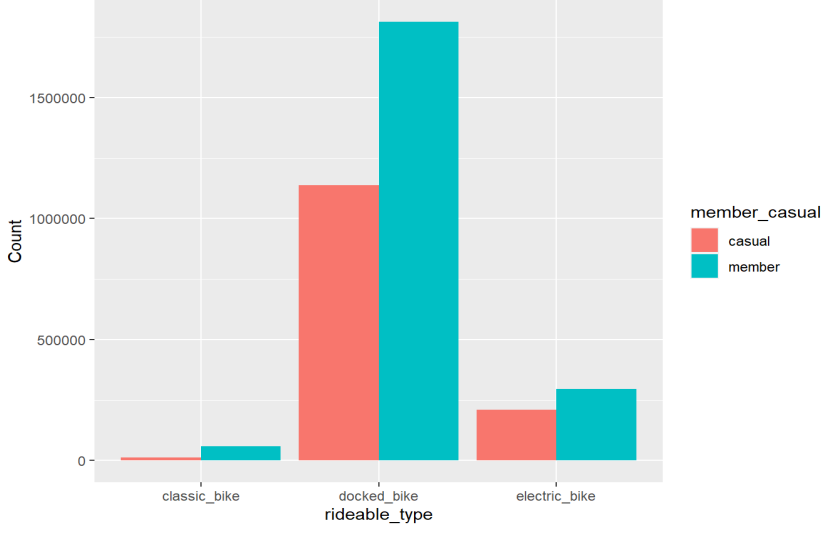
Google Data analytics Course Case Study

5- Supporting visualizations and key findings

- **Casual** Riders average ride length in minutes is about +3 times more than the **members**
- Members who took rides in 2020 made 61% of the total rides in 2020, casual riders were the other 39 %

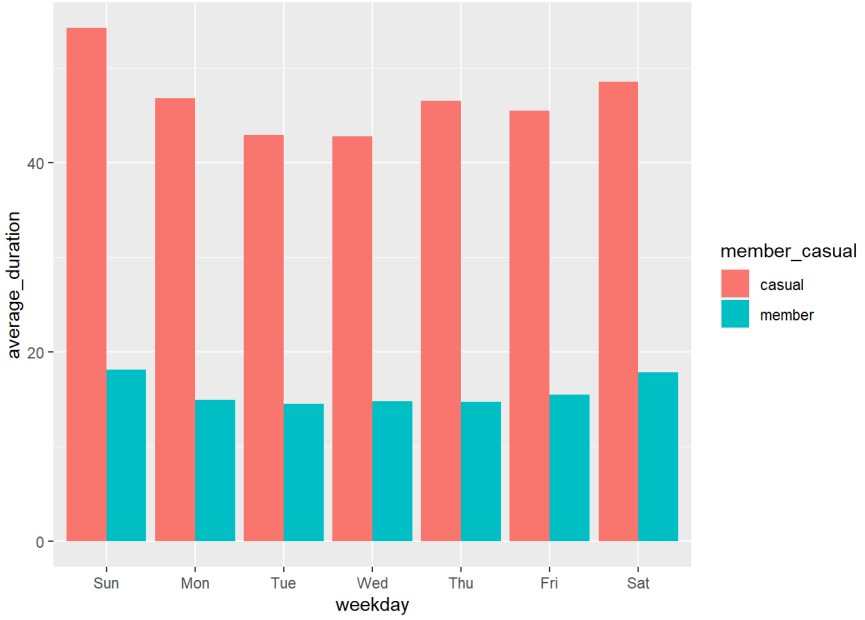


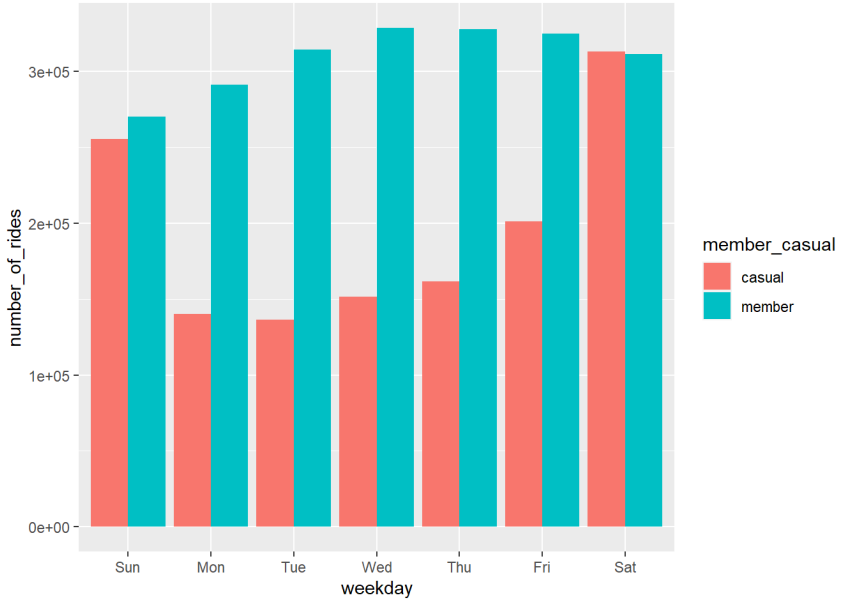
- Docked bikes are the most popular among riders also they are easier to find via smart phones mapping apps for members and casual, so **Docks** are special bike racks that lock the bike, and only release it by software control., whether casual or members then electric bikes and lastly classic bikes



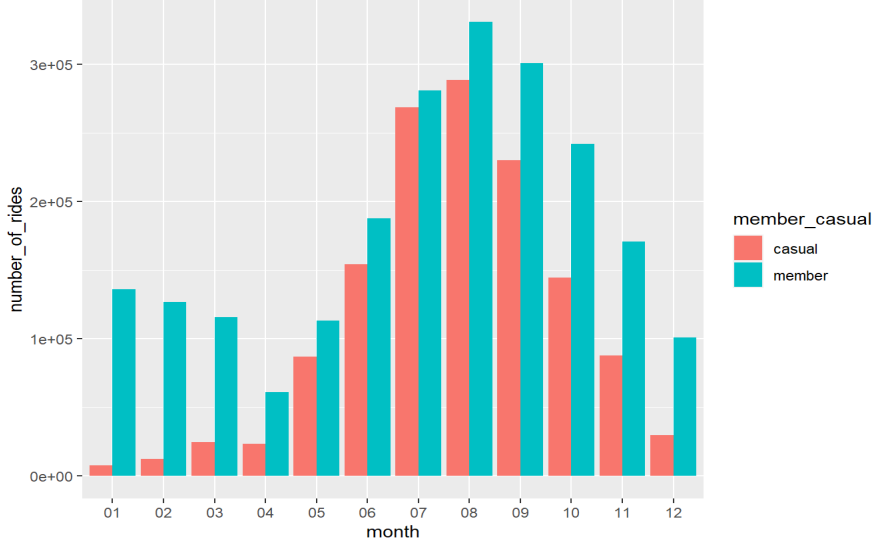
Google Data analytics Course Case Study

- Casual riders **exceeded** the average ride duration of members consistently during the whole week, also with the total number of rides, avg durations of rider are increase in the weekends (Saturday and Sunday)



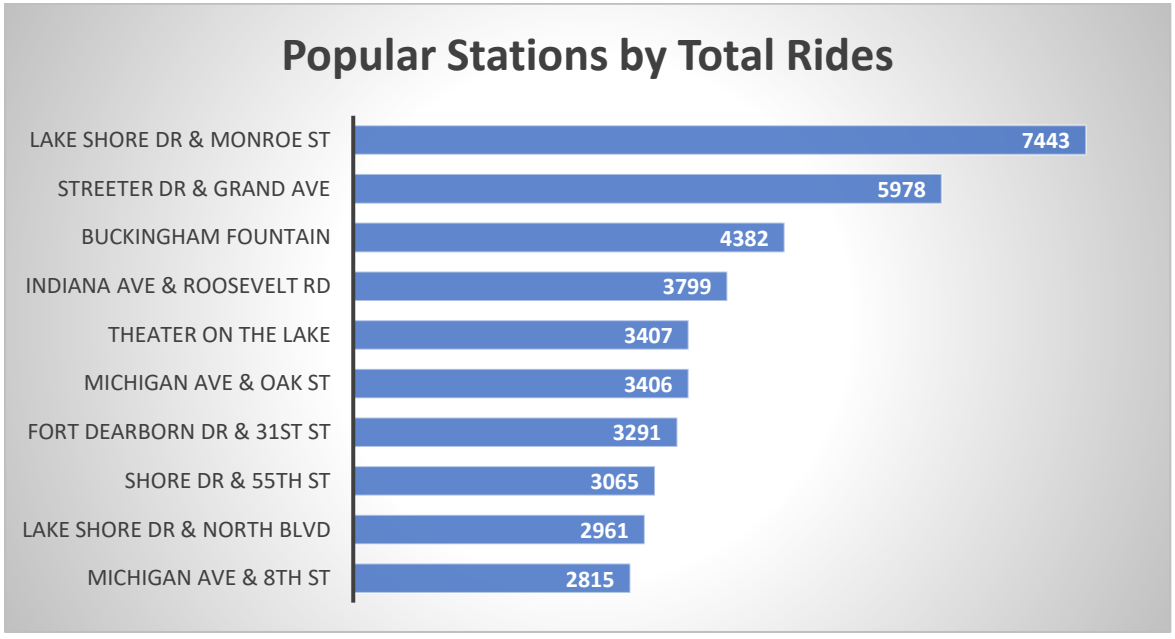


- Summer months are the most blooming months for ride sharing business particularly May till August, that's quite reasonable for a city like Chicago with cold winter weather with average daily highs are around 30 – 38°F (-1°C – 3°C) Source, however further drilling down of hours though each month shows a peak of average ride duration at the early morning that decreases throughout the day then rises again at night time (at hour 23)



Google Data analytics Course Case Study

- Top 10 Popular stations in 2020



6- Top three recommendations based on analysis:

1 - Best **seasons** to promote marketing campaign that targets **Casual riders** is during the spring and summer seasons (starting from April), with personalized marketing campaign that suits the client needs also it is important to mention that **docked bicycles** are the most favored among riders.

2- Best **Timing**/Hours are after work hours, for example 7 Pm till 11 Pm

3 – **Places** to promote the marketing campaign are around the Chicago harbor area, specifically Lake shore Dr & Monroe St. , and Streeter Dr & grand Ave that is where most exposure can be.

7- Data Limitation:

- The data does not contain personal member's (Id, age, status) , where further analysis can show their preferences and usage of the membership to personalize marketing (Due to data privacy these details were not given)
- The data does not contain pricing plans or other types of bikes for people with disabilities as the case stated, which would be helpful to asses the company's market share and their buyer's persona.