

Google Data Analytics Case Study #1 Report

JianHui (Jake) Li

2022-07-26

Introduction

This is the report for Google Data Analytics case study #1 with the following deliverables:

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

Business Task:

Identify trends and differences between members and casual riders using Cyclistic historical bike trip data to help design marketing strategies for the Cyclistic marketing analytics team.

Data Source Description:

I downloaded the latest available 12 CSV files from **06/2021 to 05/2022** of the Cyclistic's historical trip data from here.

Each CSV file has **13 columns** which are:

- **ride_id:** Unique id per ride.
- **rideable_type:** Classic_bike, Electric_bike, or docked_bike
- **started_at:** Date and time when ride started
- **ended_at:** Date and time when ride ended
- **start_station_name:** The station name of where the ride started
- **start_station_id:** The station id of where the ride started
- **end_station_name:** The station name of where the ride ended
- **end_station_id:** The station id of where the ride ended
- **start_lat:** The latitude of where the ride started
- **start_lng:** The longitude of where the ride started
- **end_lat:** The latitude of where the ride ended
- **end_lng:** The longitude of where the ride ended
- **member_casual:** Member (who purchased annual memberships) or casual riders (who purchased single-ride or full-day passes)

Note: These datasets have a different name because Cyclistic is a fictional company, and the data has been made available by Motivate International Inc. under this license.

Cleaning and Manipulating Documentation:

- Merged all CSV files into one.
- Sorted data by started_at column descending.
- Removed irrelevant columns and rename columns for clarity.

- Added ride_length, day_of_week, month, season, and time_of_day columns.
- Fixed data types and categorical levels.
- Filter out rows that contain errors.
- Define outliers using the interquartile range method and percentiles. Then removed and added them to another dataframe.

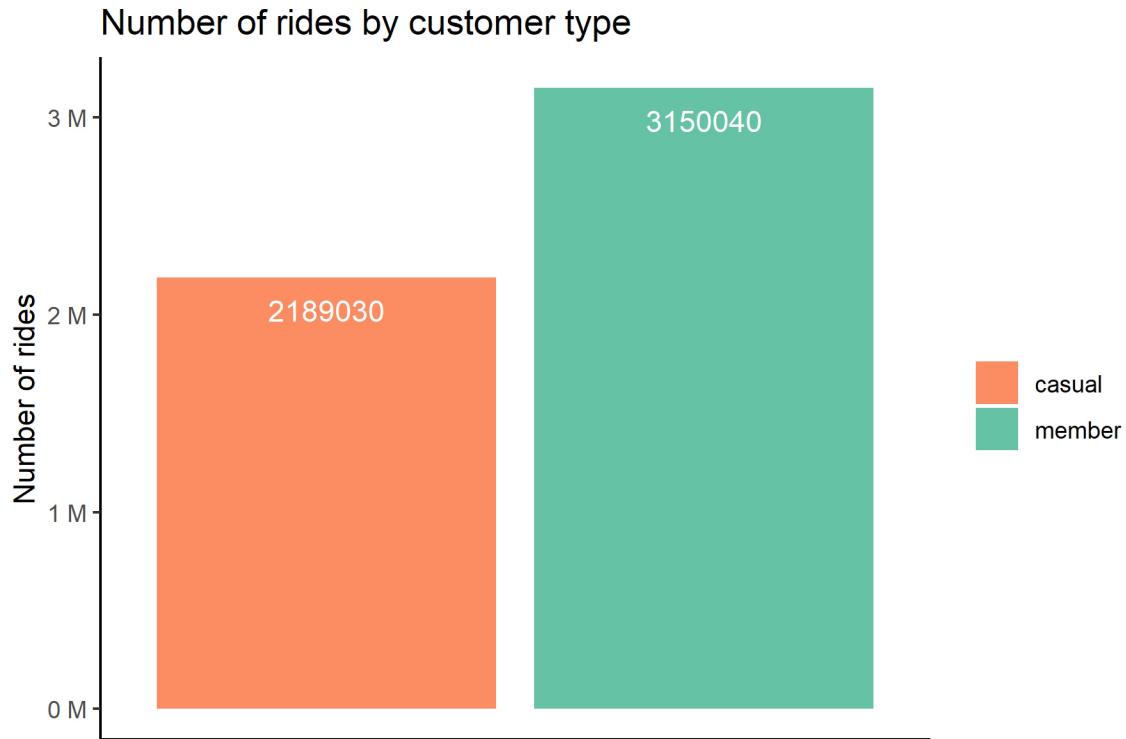
Analysis Summary:

Calculated:

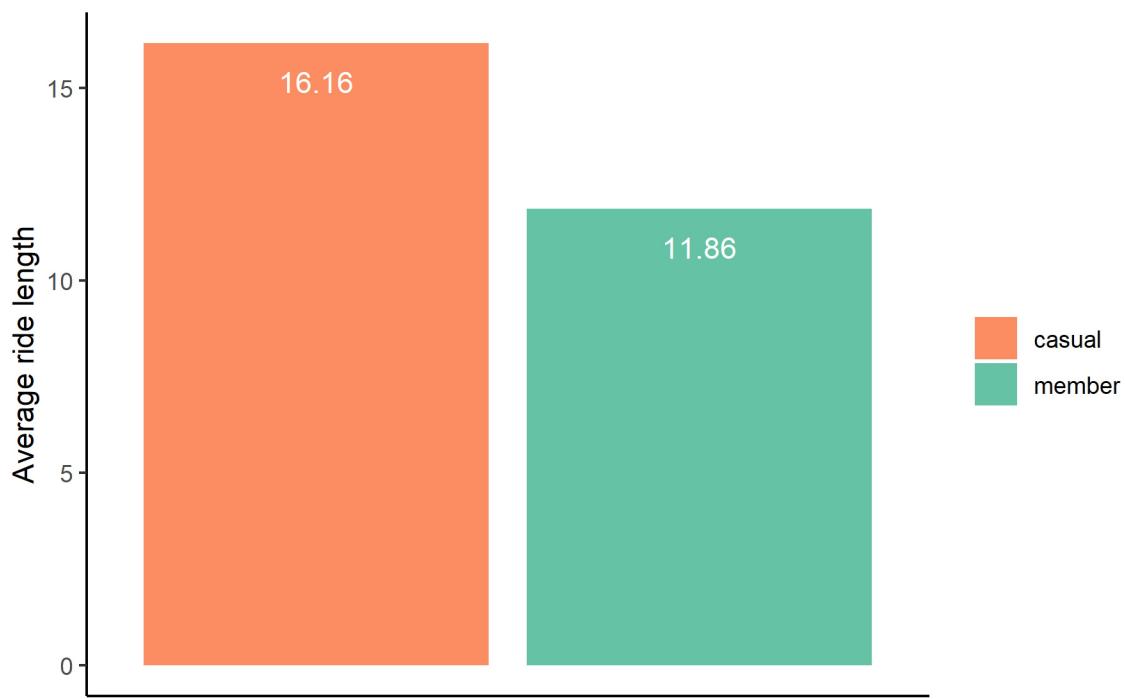
- 1) Number of rides, average, minimum, median, and maximum of ride length in minutes by customer type.
- 2) Number of rides, average, minimum, median, and maximum of ride length in minutes by customer type per season.
- 3) Number of rides and average ride length in minutes by customer type per month.
- 4) Number of rides and average ride length in minutes by customer type per day of week and time of day.
- 5) Rideable type frequency and its average ride length in minutes by customer type.

Key Findings and Supporting Visualizations

- 1) The average ride length for casual customers is always higher than members, but the number of rides is always less than members regardless of season and time of day. This suggests that casual customers take rides for long trips to make it worth the cost. In other words, casual customers use bikes for long trips, and members use them for shorter but more frequent trips.

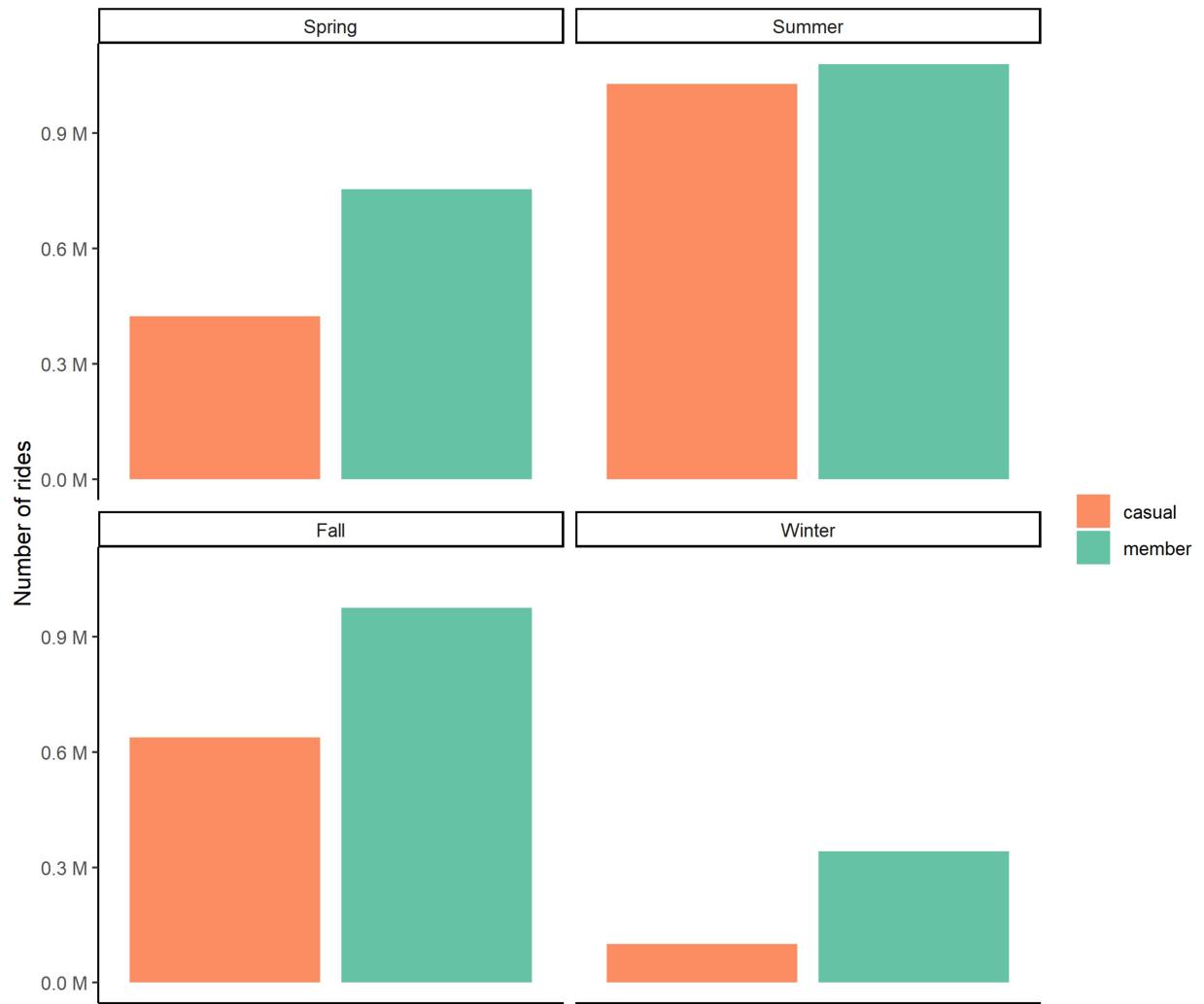


Average ride length in minutes by customer type



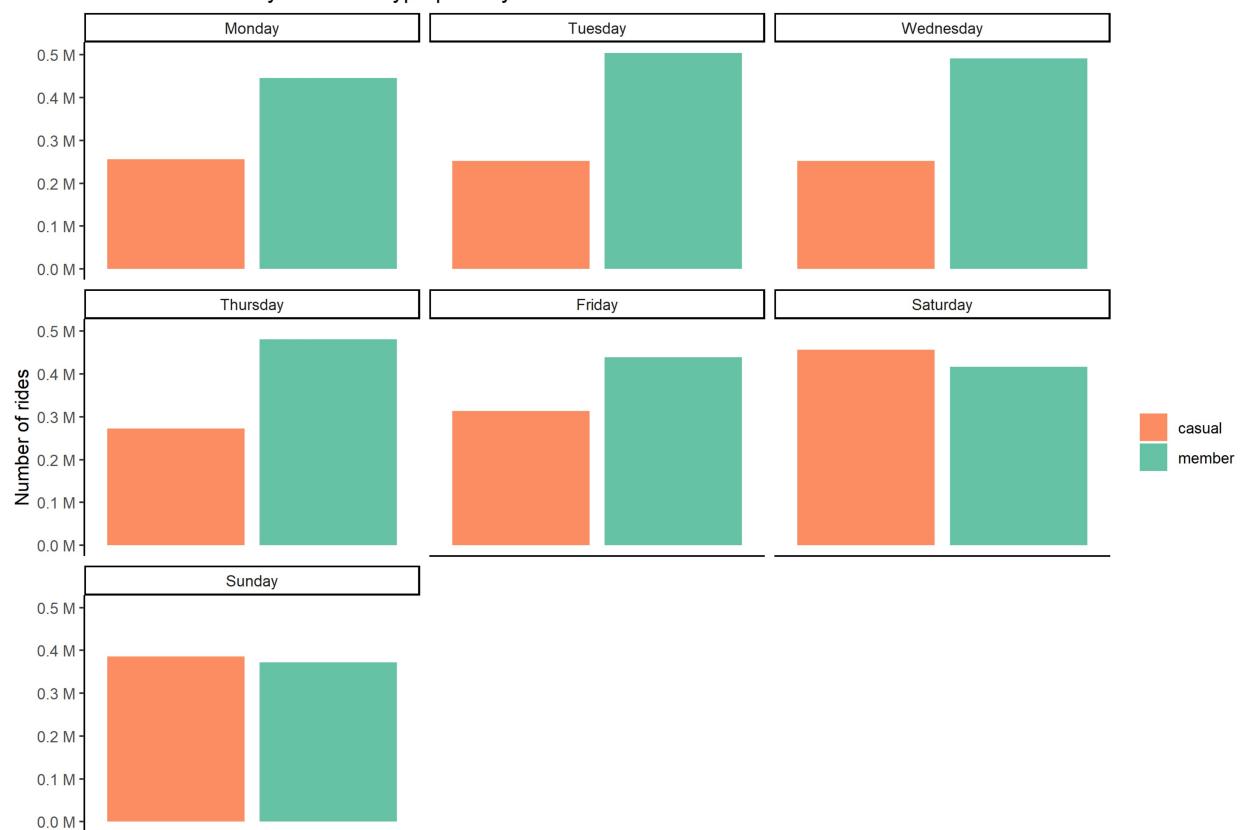
-
- 2) Casual customers take fewer rides in the spring, fall, and winter compared to members expect in the summer. Casual customers take nearly the same amount of rides as members during the summer. This suggests that more people are willing to pay to use the bike when temperatures are high.

Number of rides by customer type per season

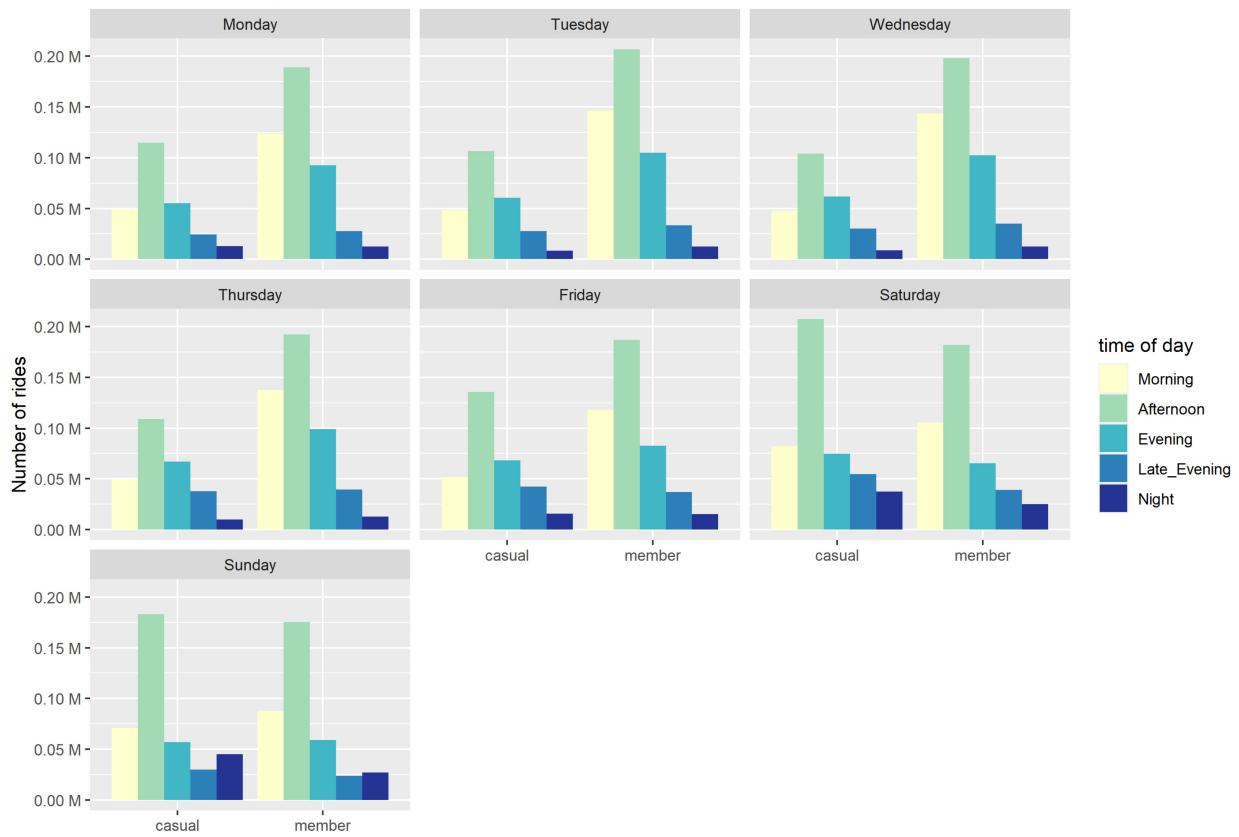


-
- 3) Members take more rides on weekdays than casual customers. This suggests that members are using bikes to get to work/school. In addition, the most common time of day of the ride takes place in the morning and afternoon for members. On the other hand, for casual customers, the most common time of day of the ride is in the afternoon and evening. This further confirms that members are taking bikes to work or routine related.

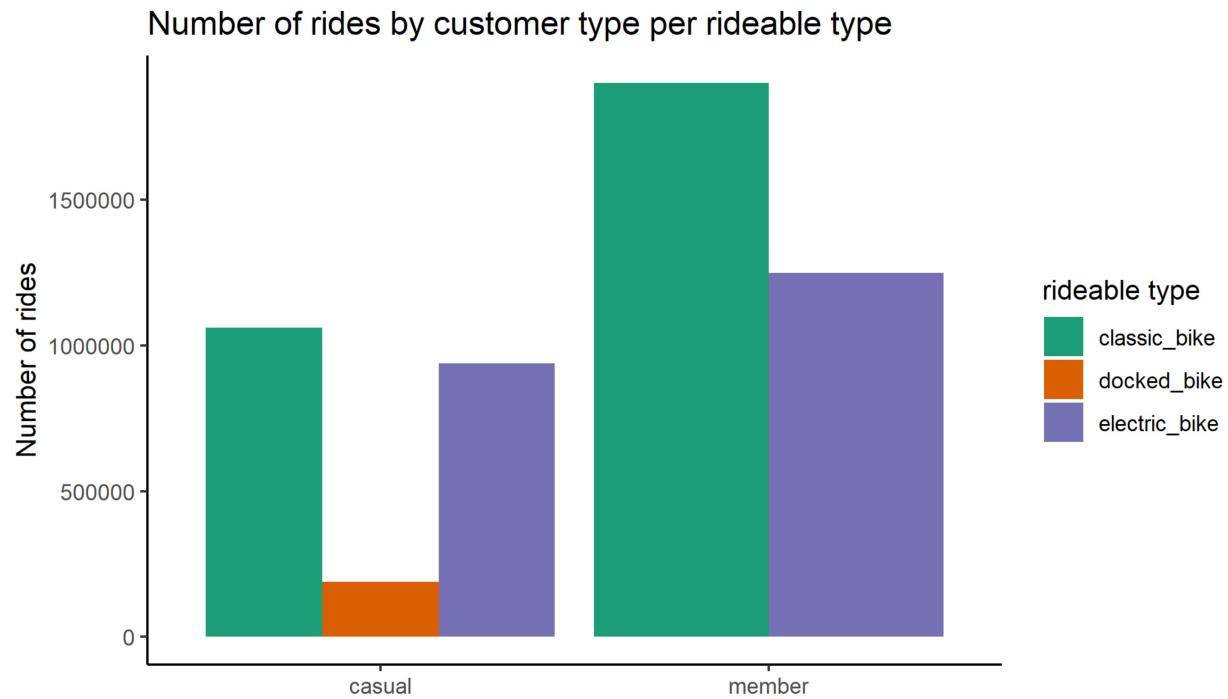
Number of rides by customer type per day of week



Number of rides by customer type per day and time



-
- 4) Casual riders use electric bikes almost as often as classic bikes even though nonmembers need to pay an extra fee to unlock electric bikes. This suggests that the ebike is a popular choice among casual riders even though it cost more.



Top Three Recommendations

- 1) Emphasize membership's affordability and its access to unlimited 45-minute rides on classic bikes or electric bikes. Also mention it's an affordable and convenient way to get to work, school, gym, leisure, etc. This attracts frequent casual riders or riders who need to commute every day.
- 2) Since there are significantly more casual riders in summer, releasing advertisements near/during the summer will help attract casual riders who seek affordable transportation during high temperatures.
- 3) Make a point that members don't have to pay a fee to unlock an electric bike. This targets casual riders who prefer to use electric bikes over classic bikes.