TASK

I have been completing the Google Data Analytics Professional Certificate available on Coursera, and the final project is an optional case study to showcase skills learned. One of the options is to take on the role of a Junior Data Analyst to assist the fictional bike-sharing company Cyclistic in determining the differences in **how use differs** between casual riders and members. **The goal** is to be able to design a marketing strategy to turn these casual riders (customers) into members (subscribers).

DATA

Datasets

Data is from Divvy Trips (available under this license) to be used as Cyclistic's mock data spanning the last three quarters (roughly nine years). It is downloaded from the company's cloud server on S3 Amazon AWS https://divvy-tripdata.s3.amazonaws.com/index.html.

The files include 2019_Q3, 2019_Q4, and 2020_Q1, and each are set within their own spreadsheet. Each file provides 12 different fields to be prepared and processed.

Cleaning/Manipulation

The size of the files I used were suitable for using in **Google Sheets**, so I utilized functions to clean and manipulate the data as needed (sheets available within the repository).

Edited 2020_Q1 file

- Changed "casual" to "Customer" and "member" to "Subscriber" so that the usertypes match with the 2019_Q3 and _Q4 files.
- Changed a couple column titles so that they match the 2019 files
    - started_at to start_time
    - ended_at to end_time
    - start_station_id to from_station_it
    - start_station_name to from_station_name
    - end_station_id to_station_id
    - end_station_name to to_station_name
    - member_casual to usertype
- Assortment of rides are between the same "HQ QR" location and are recorded to typically last milliseconds.

    - Assumption: These occur during the company's quality control tests and provide no relevant data for analysis
    - Rows have been removed
    - Removed another row with missing ending station data but appears to be of the same "HQ QR" test ride

Edited 2019_Q3, 2019_Q4, and 2020_Q1

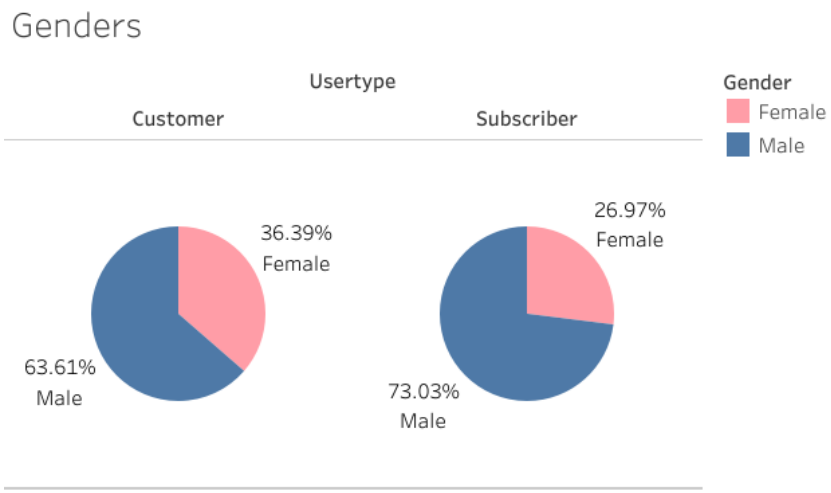
- Added ride_length column
    - Calculated as end_time - start_time
    - Formatted as Duration
    - Checked for negatives
        - Occurred during trips that extended multiple days: modified formula to correct for these instances
- Added day_of_week column
    - Calculated from the start_time (1=Sunday)
- Added start_date column
    - Calculated from extract the date from start_time

Along with these, I checked columns for any values outside of viable ranges, inconsistencies, and for excessive nulls that would affect the analysis.

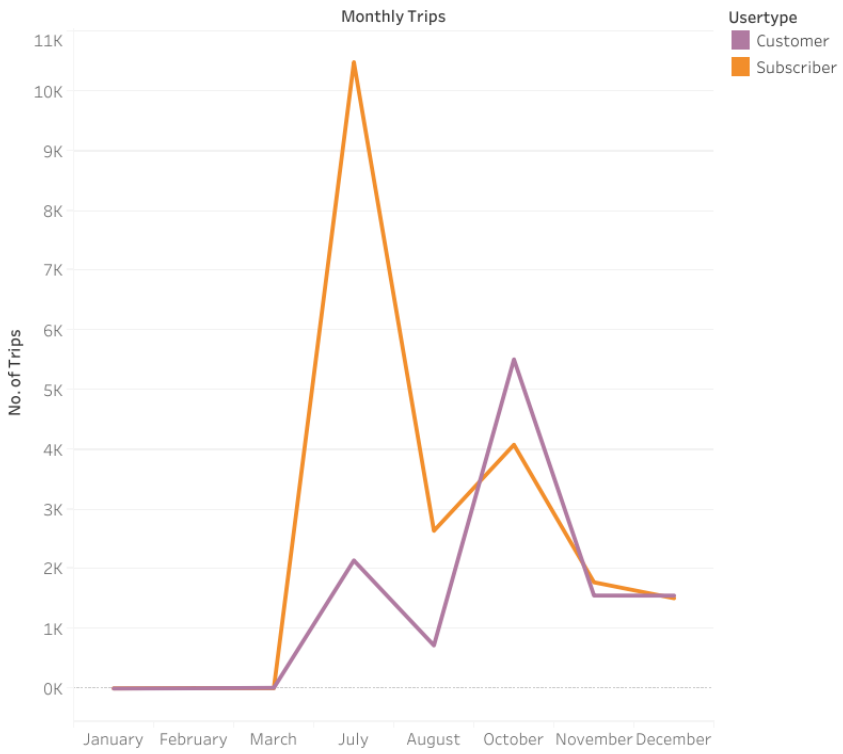
ANALYSIS

Visualizations of the data were done within Tableau, and all of the supportive diagrams can be found here.

The subscribers versus the customers are about a 65/35 split respectively with no significant difference between the amount of males and females in each usertype. There is, however, a large difference in the numbers of males and females themselves.

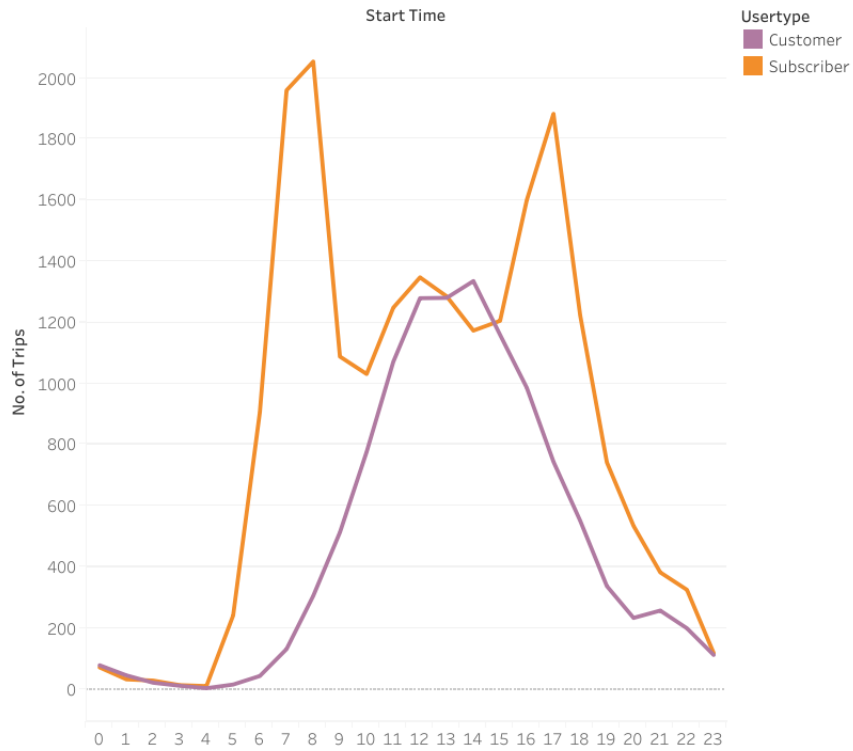


Total Rides per Month

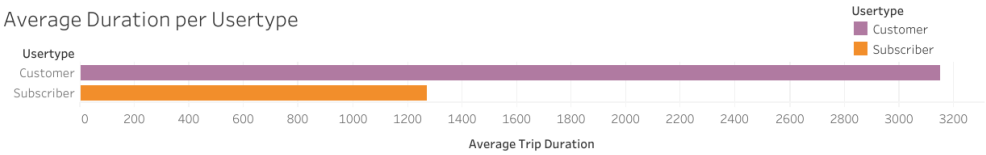


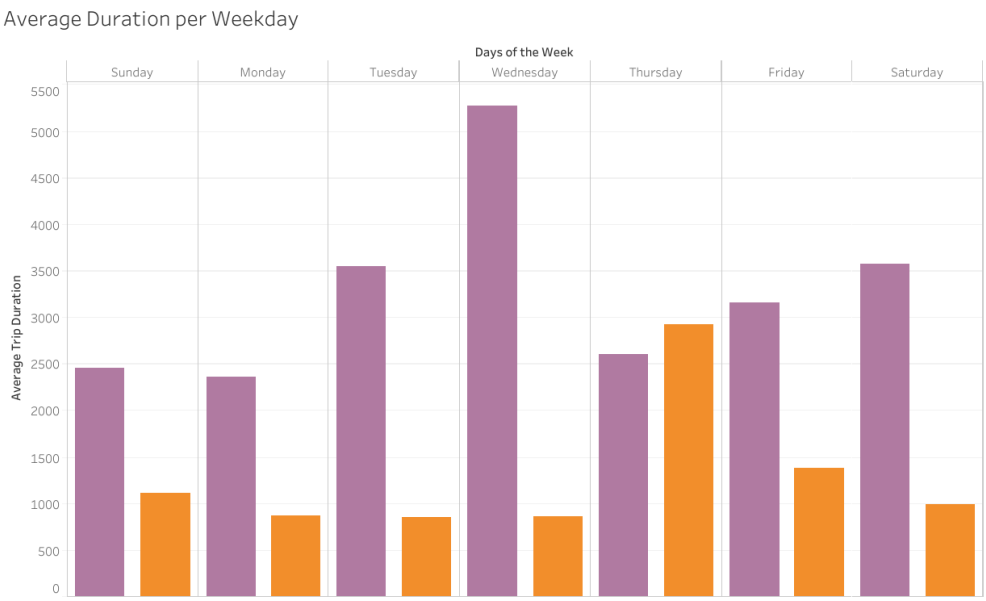
Throughout the year, bike rides **spiked** during the Summer months, particularly **July through October**.

Total Rides per Hour



During the day, trips typically start rising in volume around **5am** for **subscribers** while **customers** do so around **7am** before they **both** hit a decline around **5pm**.

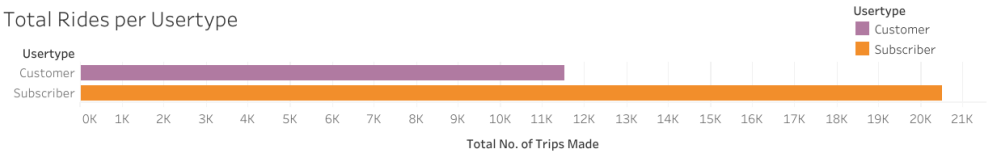


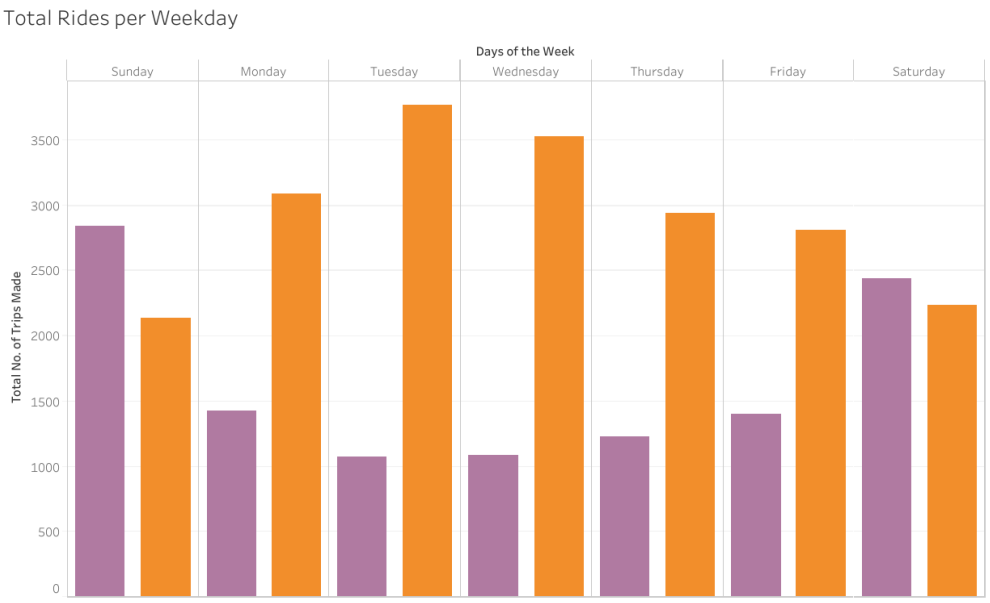


The average length of time spent riding was consistently **significantly higher** for the customers.

However, while the **subscribers** typically rode for **shorter lengths**, they had far more separate trips counted.

This makes sense as a subscriber wouldn't even need to think about grabbing a bike for a short ride, while a customer may be more particular, only wanting to ride for longer times when they have to pay for each individual trip.





Throughout the week, there were a lot more **subscribers** taking rides during the **work days** (possibly work commuters based on hourly trip rates), however customers still covered most of the distance covered on those days.

Recommendations

Based on the data presented, I can offer these possible suggestions to marketing:

- More ads could focus on **women** to bring them further into the market.
- Bike riding is at its highest during Summer, so advertisements will be **most effective** just before this time starting around **June through August**.
- Public ads (posters, flyers, etc.) are useful for customers and potential new users
- The company could provide **incentives** (such as discounts or rewards) for longer rides.
    - Can also promote current subscribers keeping their membership.