Cyclistic

Google Data Analytics

Case Study

Apnatva Singh Rawat

Table of Contents

Aim of this Analysis

Increasing member percentage among the customers of Cyclistic.

- Analyse the data to find key differences between how Members and Casuals use the services provided by Cyclistic.

- Infer reasons to what habits may deter Casual Users from subscribing to the model.

- From the information learned, recommend steps to increase subscriptions among users.

Steps Involved

Collect the data (.CSV files) and accumulate them into a single SQL table.

Analyse and clean the data using SQL.

Run statistical analysis using Python.

Prep the data for visualisation in Excel and Tableau.

Compile data and visualisations.

Share insights and provide recommendations.

SQL Cleaning

SQL was used to quickly look through and clean the data. The steps are documented in the script provided to recreate the tables. Data between April '19 and April '20 was observed.

Overview:

- Data was imported from CSV files to an MSSQL Database.
- Removed NULL values, and repeated values.
- Corrected misspelt words.
- Removed names for stations since they are not integral to this analysis.
- Split data based on Trips Started and Trips Ended to visualise in Tableau
- This database was imported in Python to run statistical analysis.
- Exported results to an Excel Workbook 'fromSQL.xlsx' to visualise in Tableau.
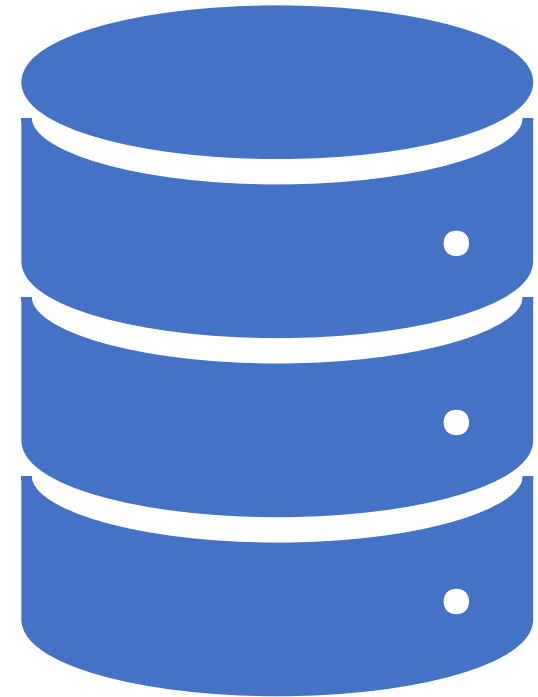
Analysis in Python

Statistics computed are categorised by Member / Casual status of users. Another table describes statistics for each day to find further insights into the habits of Cyclistic users.
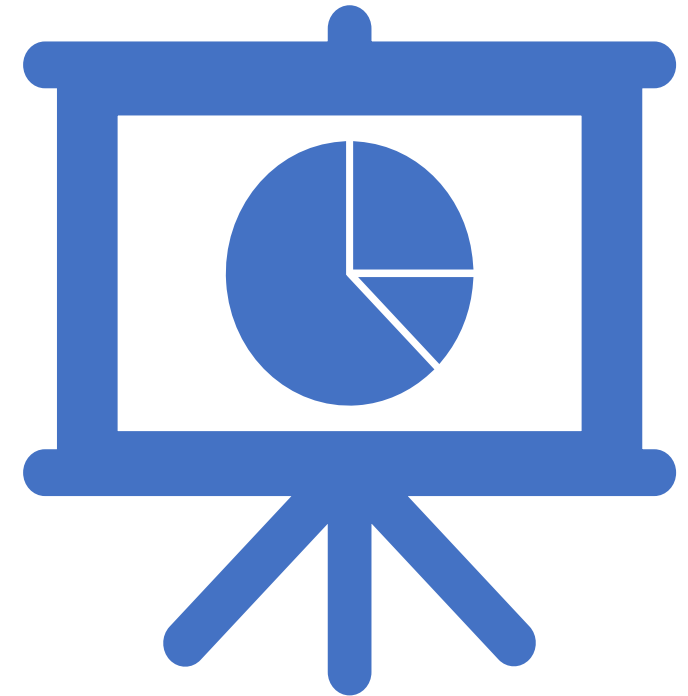
- Data was imported from SQL with pyODBC and stored in a Pandas dataframe.

- The script is provided to rerun any analysis to check for its legitimacy.

- The script generates 3 tables that are exported to an Excel workbook 'fromPython.xlsx' for visualisation.

- Instructions to run the code are provided on the script itself.
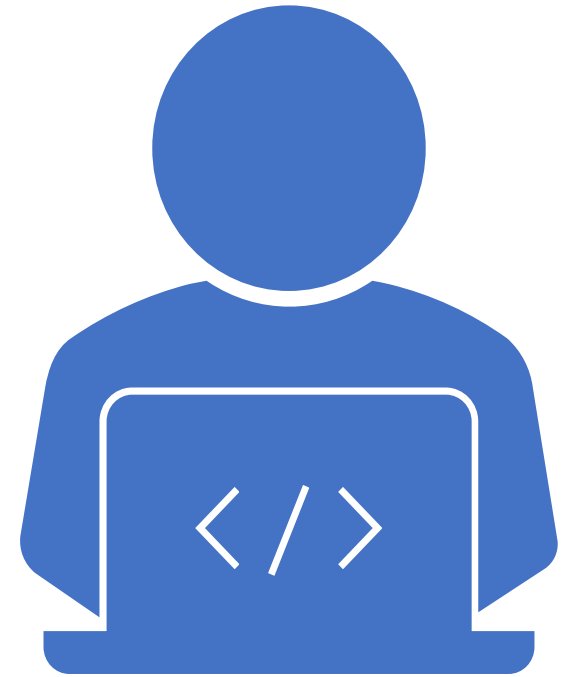
Visualising in Tableau and Excel

Data exported from SQL and Python was exported to CSVs or Excel Workbooks for visualisation.

- Geographical data was mapped in Tableau to represent the starting and ending locations.

- Statistical data was visualised in Python with % Bar Graphs, and Pie Charts. Some of these charts are included in this report.

- Other data which provided no useful insights into the analysis, was not included here. It can still be retrieved from the Excel workbooks.

INSIGHTS 1
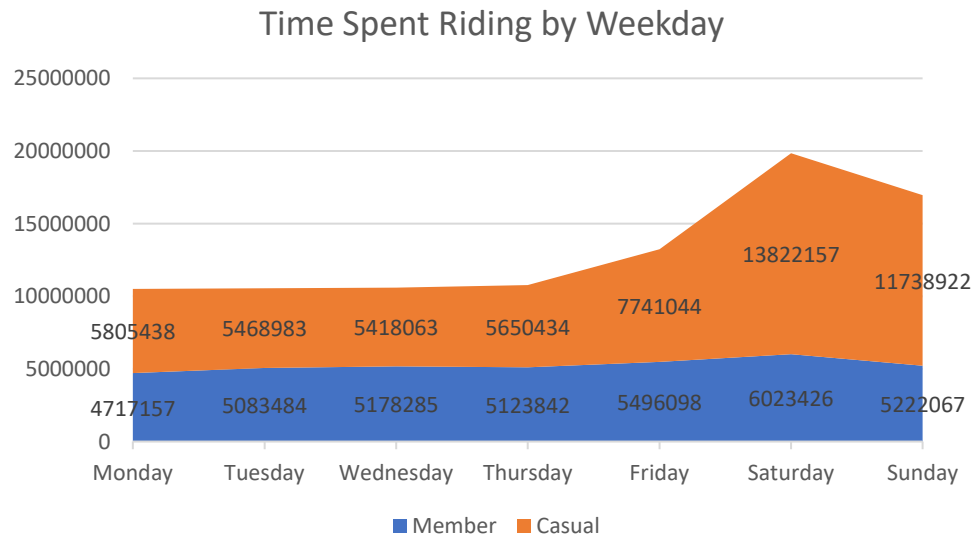
Member V Casual Comparisons

Category	Member	Casual
Median of Time Spent Riding (min)	11	21
Mode of Time Spent Riding (min)	6	10
Average Time Spent Riding (min)	14.99164205	35.59070875
Time Spent Riding (min)	36844359	55645041
No of Trips Started	2457660	1563471

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%
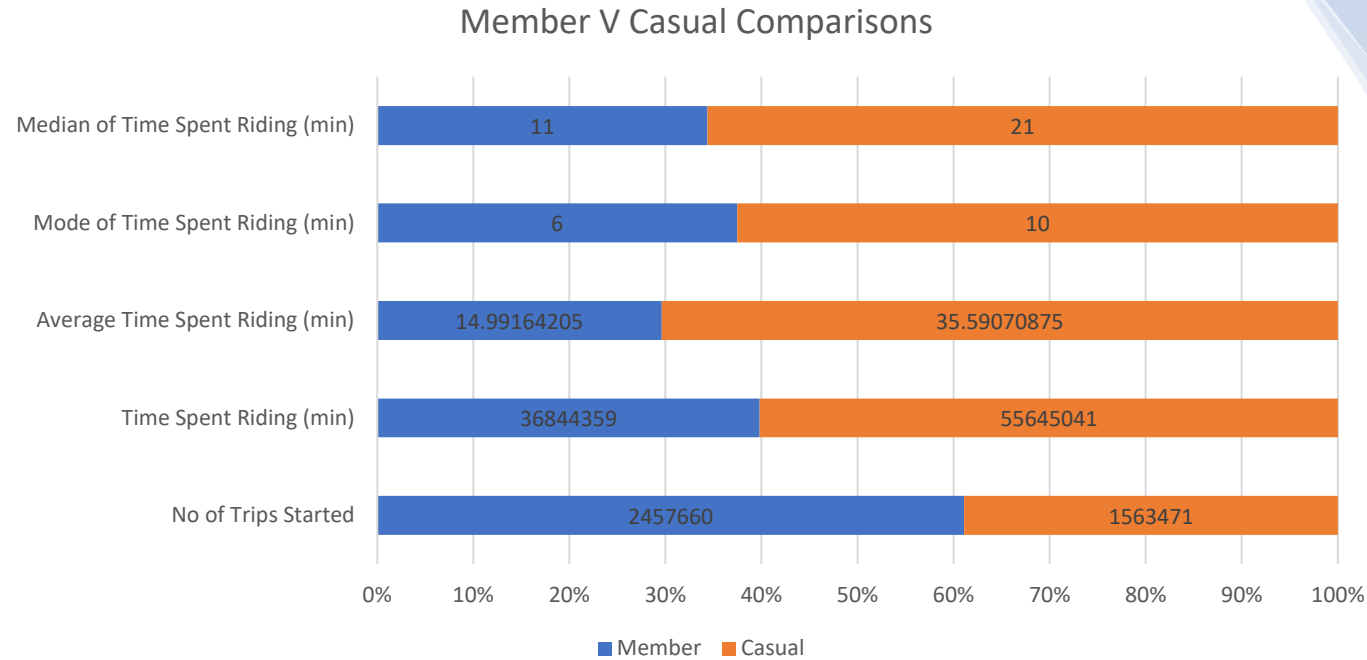
■ Member ■ Casual
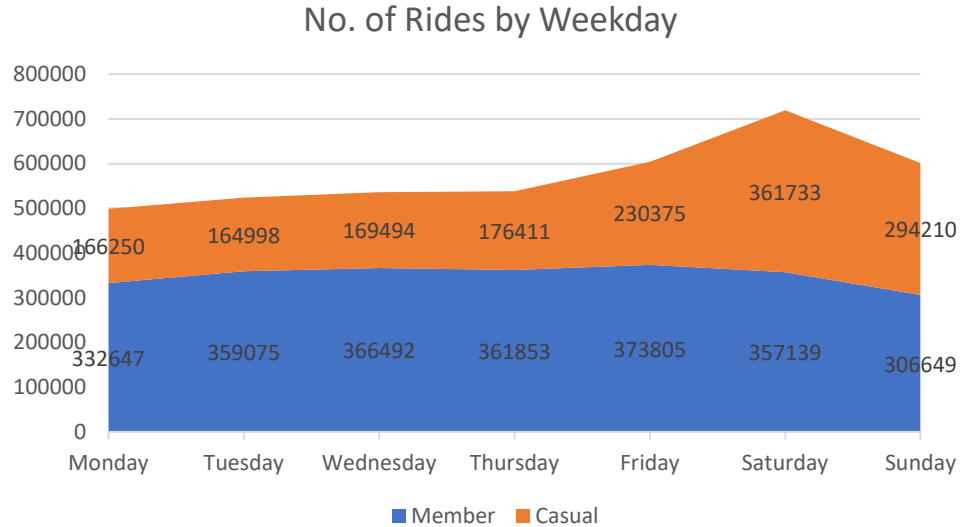
1. Members have made more trips than Casuals and can be called the 'majority users'.

2. The average time spent using the services provided by Cyclistic is higher for Casuals.

3. This clearly tells us that Members make trips more frequently but Casuals are riding longer.

INSIGHTS 2

No. of Rides by Weekday

Member: Monday 332647, Tuesday 359075, Wednesday 366492, Thursday 361853, Friday 373805, Saturday 357139, Sunday 306649

Casual: Monday 166250, Tuesday 164998, Wednesday 169494, Thursday 176411, Friday 230375, Saturday 361733, Sunday 294210

Time Spent Riding by Weekday

Member: Monday 4717157, Tuesday 5083484, Wednesday 5178285, Thursday 5123842, Friday 5496098, Saturday 6023426, Sunday 5222067

Casual: Monday 5805438, Tuesday 5468983, Wednesday 5418063, Thursday 5650434, Friday 7741044, Saturday 13822157, Sunday 11738922

1. Number of trips by members remains fairly consistent with a slight dip on the weekends.

2. Number of trips by casuals has an exponential increase over the weekends.

3. Consistent with the earlier graph, Casuals spend more time riding the bicycles and almost thrice as much as members on Saturday.

INSIGHTS 3

Duration of Trips by Month with Temp
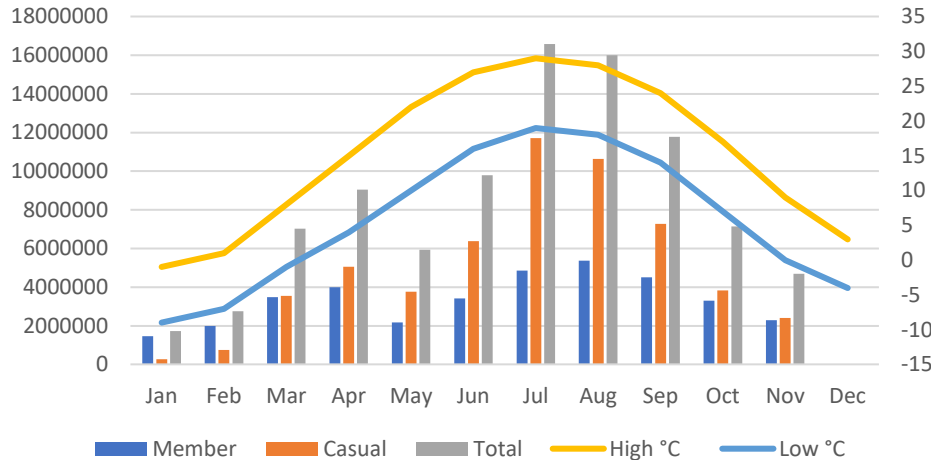

Number of Trips by Month with Temp
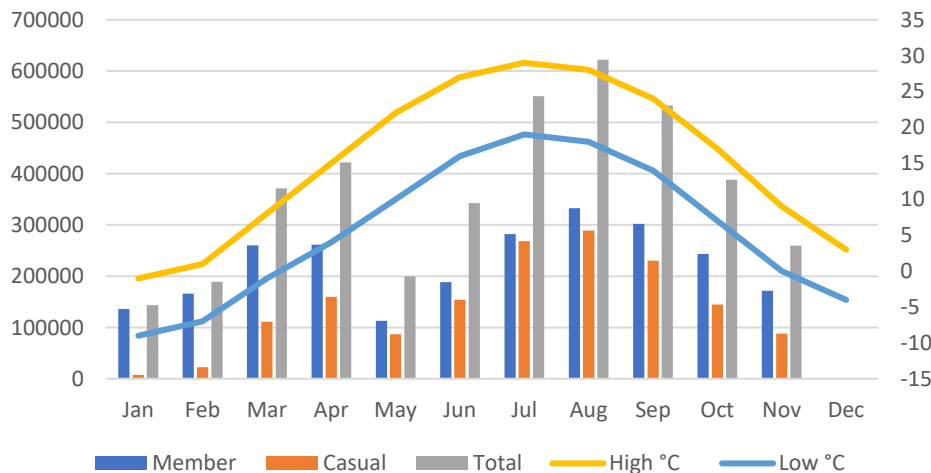
1. Data for December was missing at sources and hence is missing from this analysis.

2. From examining the trends, it can be seen that Cyclistic Users travel more in July and August than any other month. These are also the hottest months. The number of rides is fairly consistent with the temperature, except for May and June.

3. The number of trips for casuals was almost non-existent in the coldest months, which adds more weight to the fact that they choose to use Cyclistic's services for leisure.

INSIGHTS 4

This is the geographical data which was plotted in Tableau. The precision of each location was reduced to offer fewer data points and make the visualisation less crowded. This does not affect the credibility of the map in small regions with a datapoint being off by 11 metres. This is only a snapshot of the visualisation in Tableau, the interactive version can be accessed here.

1. A larger circle means more trips.
2. The labels depict % of Members and % of Casuals in that location.
3. The majority of the data points seem to be blue but on zooming in it is observed that Members have more data points. They are not as concentrated in a single area as Casuals.
4. Casuals have more trips nearer to coastline while members are in the majority further from the shore.
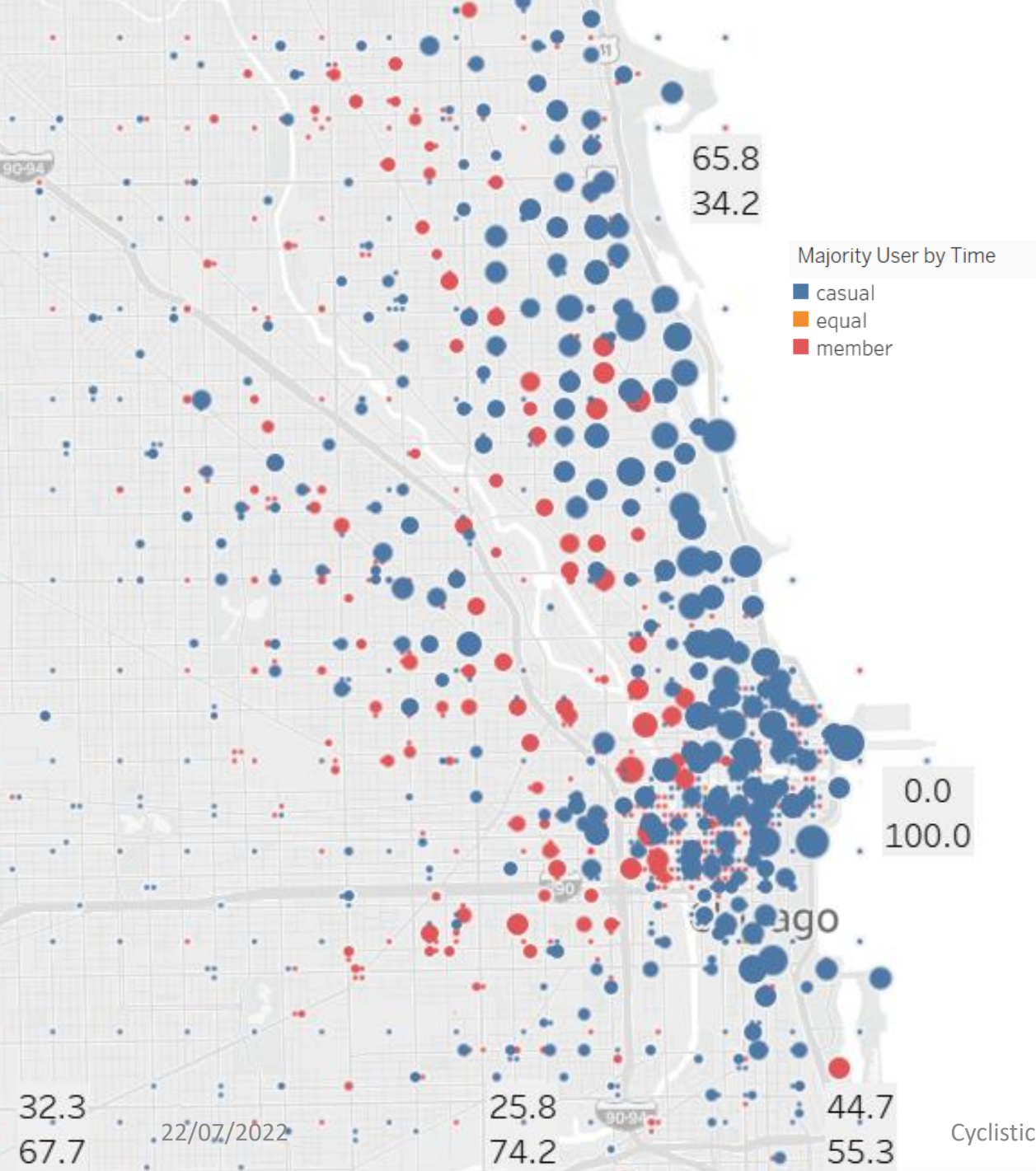
Key Differences

Members

Shorter Trips

More Frequent Trips

Use services majorly as a form of commuting

Casuals

Longer Trips

Less Frequent Trips

Use services majorly as a leisure activity

What does the Data say?

These are inferences supported by facts from the analysis.

1. Biking is not only a form of commuting but also a leisure activity. Casuals ride for longer durations although these trips are less frequent. This is suggestive of Casuals tending to ride bikes more for leisure. This assumption is promptly supported by two facts:
   - The average riding time for Casuals is much higher than that for Members.
   - During the weekends there is a sudden spike in the number of trips by Casuals, even surpassing those of Members on Saturdays.

2. Casuals are also repeat customers for Cyclistic but they still choose to not become members.
   - It can be inferred from the tableau visualisations that casuals tend to have larger data points. A customer who rides for leisure will start their trip from the same biking station routinely, driving up the average for that location.

3. Members use bikes as a form of commuting, especially to work.
   - This assumption is backed up by the fact that members had more trips during weekdays with a significant dip on Sunday. A small inconsistency can be observed for Monday, where the time spent riding is the lowest. This can be attributed to some of the Members also riding for leisure on the weekends but choose not to on Mondays.

Recommendations

Possible routes to increase the member percentage among Cyclistic customers.

1. Revise the subscription model to incentivize more time spent travelling over more trips.
    *   This will promote subscription among casuals who do not use the services frequently.
    *   It might impact the current subscribers since they might feel that the older model was more beneficial for them.
2. Offer two subscription models, one for commuters and the other for leisure riders.
    *   Users utilise Cyclistic's services for either of the above-mentioned reasons. If the subscription model caters more personally to their requirements, customers will be more likely to subscribe.
3. Promote Cyclistic not as a Bike Share Company but as a Biking Community.
    *   This will help move the company's image from a service provider to a community that promotes biking, a healthy and fun activity.
    *   It will also increase engagement from non-frequent leisure riders as the community grows and attracts customers for reasons beyond renting a bike. a

Going Forward

- Survey the customers to gain more insights into how they use the services provided by Cyclistic.
  - The survey would aim to confirm certain assumptions that were made in this analysis before the company starts reconstructing its subscription model. This may also help gain additional insight which could not be inferred from the data.

- Study the current model to find if it suits the habits of Members and Casuals both.
  - Consider revising the current model or offering several plans depending on the market research from the previous note.
  - Understanding users' needs will help the company offer better, more personalized services.

- Better Geographical Data collection will allow for better visualisation of the dataset.
  - From observing the data it can be seen that it takes the user's approximate location as the pick-up and drop-off location. Data points are approximately close but a more organized dataset offers better insight when visualizing the trips on a map.
  - It should simply log the data for the docking station where the trips started/ended. This will make it easier to see where the trips are most concentrated and find popular destinations.

- Use Social Media to establish a Cyclistic Biking Community to better integrate with society.
  - Engaging with customers as a social service by making subscriptions commonplace even for entire households. This could be a great opportunity for the company to promote biking as a pollution-free mode of transport and a great exercise.
  - Social work can lead the way for Cyclistic to be more than a Bike Share company. It can offer users a healthy exercise program akin to a gym membership for better health. Cyclistic has the infrastructure and the user base to go forward with this.

Thank You