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.

Important Considerations 1

- Data was provided by Cyclistic but personal information which could be used to identify multiple rides by the same person was removed for privacy reasons.

- When bikes went out for repair, they were signed out which initiated a trip last even days. Such entries were identified and promptly removed.

- Some trips were started after they were ended. This was true for marginal cases and regarded as a logging mistake and start and end times were exchanged.

- The cost associated with a trip or any means to calculate such costs were not provided. Any recommendations regarding such were made with assumptions that the data suggested.

Important Considerations 2

- Rules regarding what a 'Single Day' or 'Single Ride' pass remain unclear, leaving room for assumptions.

- The geographical data collection did not point always to a docking station but seldom close by. This made it difficult to properly identify stations and visualise the data. To counter this the latitudes and longitudes were truncated to 4 decimal places, leaving the data accurate to 11 metres, and decreasing the number of data points for visualisation.

- The data was cleaned entirely in SQL, but it was also manipulated in Python for statistical analysis. Scripts for both are provided.

- Data for December '19 was found to be corrupted and hence was excluded from this analysis.

Steps Involved

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

Clean and analyse the data using SQL.

Run statistical analysis using Python.

Prep the data for visualisation in Excel and Tableau.

Compile data and visualisations in a report.

Share insights and provide recommendations.

SQL Cleaning

SQL was used to quickly look through and clean the data. Each query and its effect are detailed in the SQL script provided. Data between May '19 and April '20 was observed.

Major Steps:

- Data was imported from CSV files to an MSSQL Database.

- Removed NULL values, and repeated values.

- Corrected misspelt values in 'member_casual'. 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. Used the start_time and end_time columns to compute the duration of trips.

- Deleted rows where duration was greater than 20 hours as these bicycles. It was assumed that these bicycles were out for repair.

- Exported results to an Excel Workbook 'fromSQL.xlsx' to visualise in Tableau.
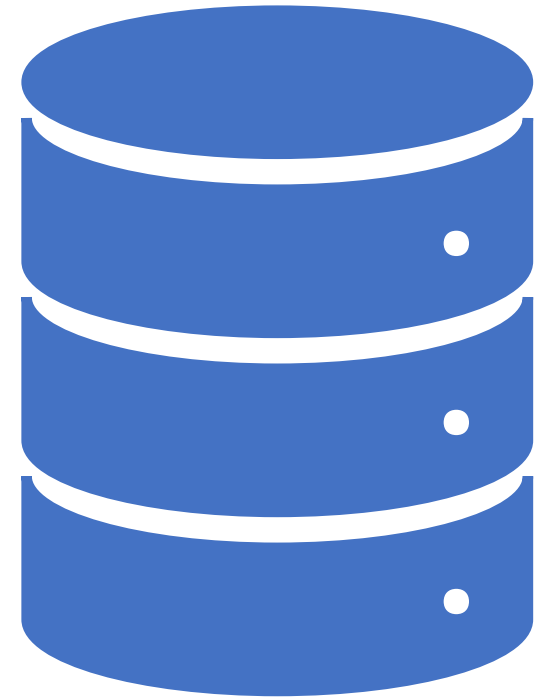
Analysis in Python

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.
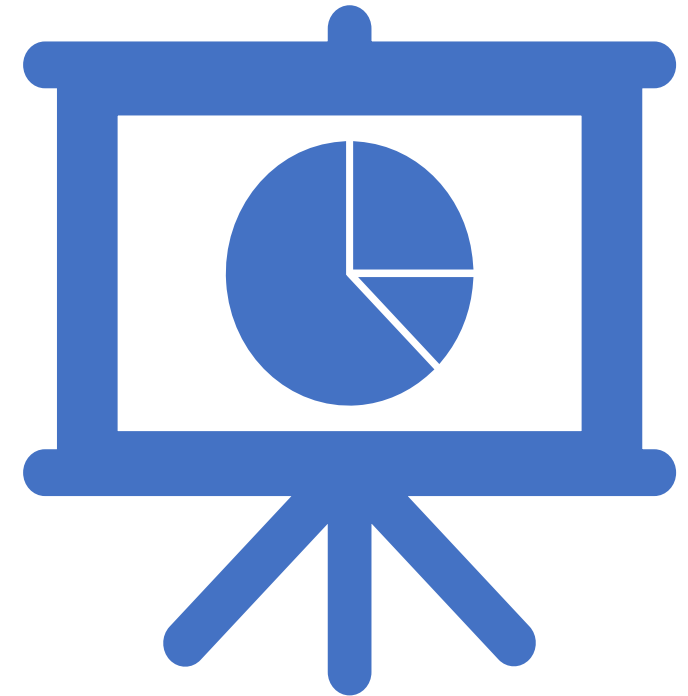
- Data was imported from SQL with pyODBC and stored in a Pandas dataframe.

- Month and Weekday columns constructed from the date on each trip.

- Data was divided into two dataframes for 'member' and 'casual' data.

- 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.

- Detailed descriptions are provided in the script with meaningful comments.
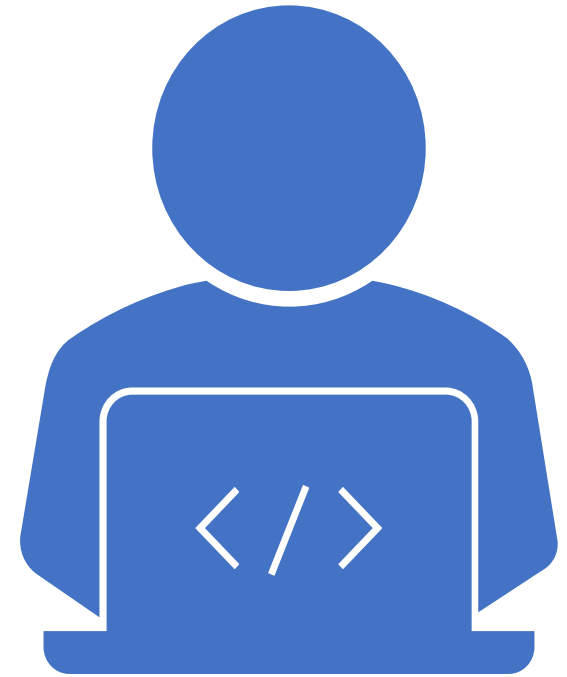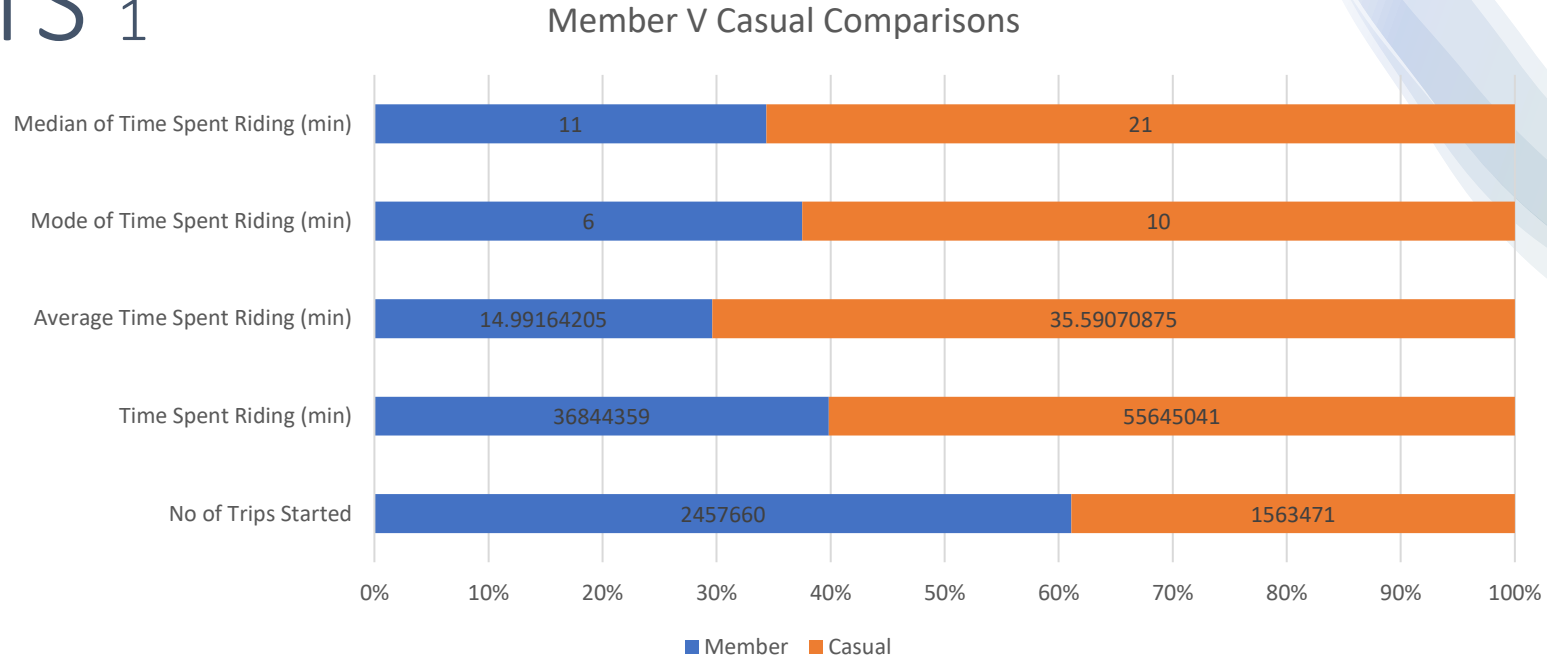
Visualising in Tableau and Excel

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. Charts with meaningful inferences were 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.
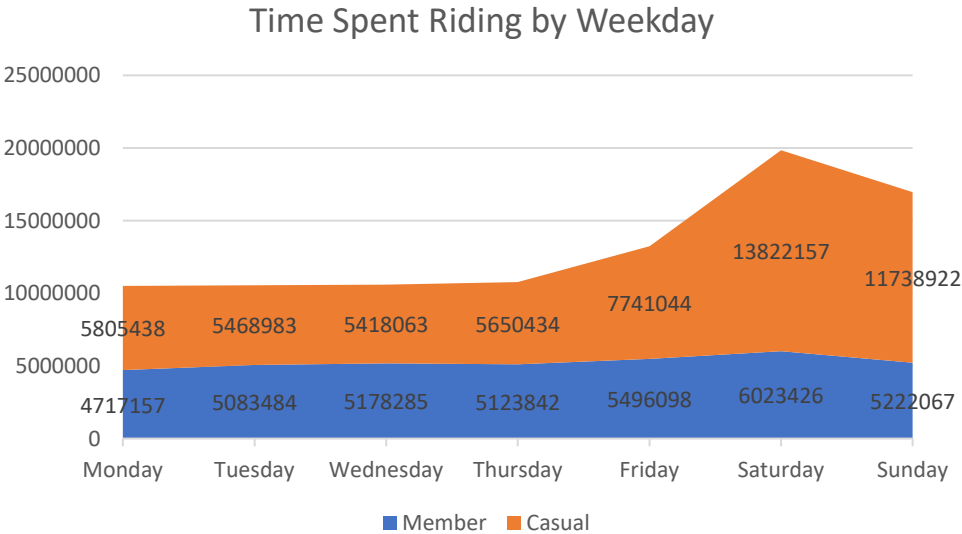
INSIGHTS 1

Member V Casual Comparisons

Metric	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

■ Member ■ Casual

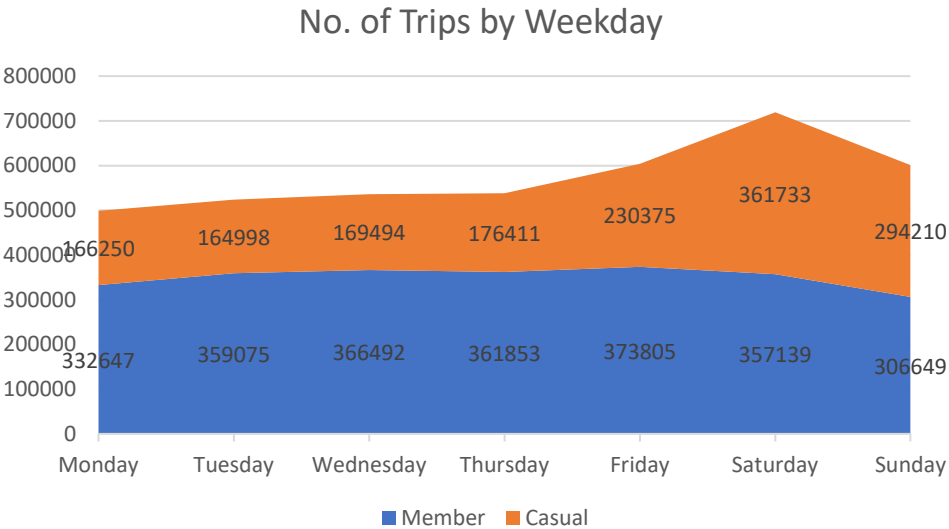1. Members have made more trips than Casuals and can be called the 'majority users', but Casuals have spent more time using Cyclistic's services.
2. The visualisation makes it clear that casuals travel infrequently and for longer, hinting that
   1. Cost of a single-ride/day pass is much lower than the price of those rides as a member.
   2. Membership is cheaper for users is they use the services frequently.

INSIGHTS 2

No. of Trips by Weekday

	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Casual	166250	164998	169494	176411	230375	361733	294210
Member	332647	359075	366492	361853	373805	357139	306649

■ Member ■ Casual

Time Spent Riding by Weekday

|	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Casual	5805438	5468983	5418063	5650434	7741044	13822157	11738922
Member	4717157	5083484	5178285	5123842	5496098	6023426	5222067

■ Member ■ Casual

1. Number of trips by members remains fairly consistent with a slight dip on the weekends, suggesting members commute to work using bicycles.

2. Number of trips by casuals has an exponential increase over the weekends, suggesting casuals bike for leisure.

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

INSIGHTS 3

No. of Trips by Month with Temp
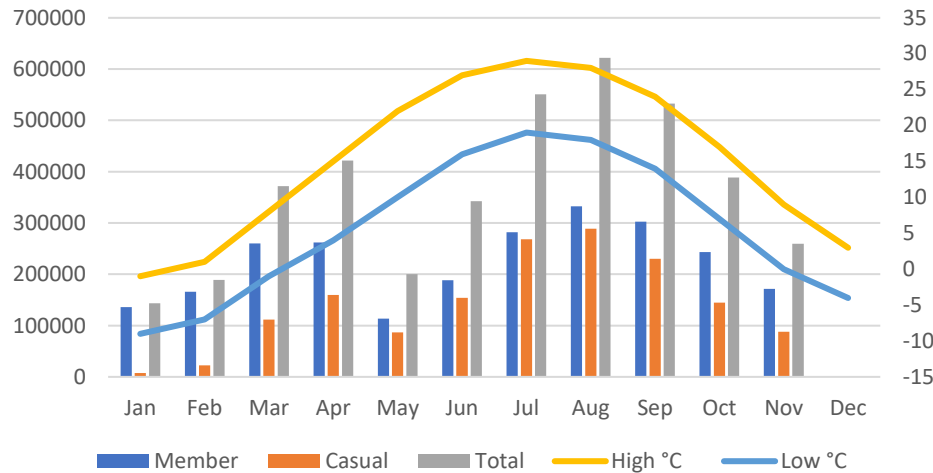

Duration of Trips by Month with Temp
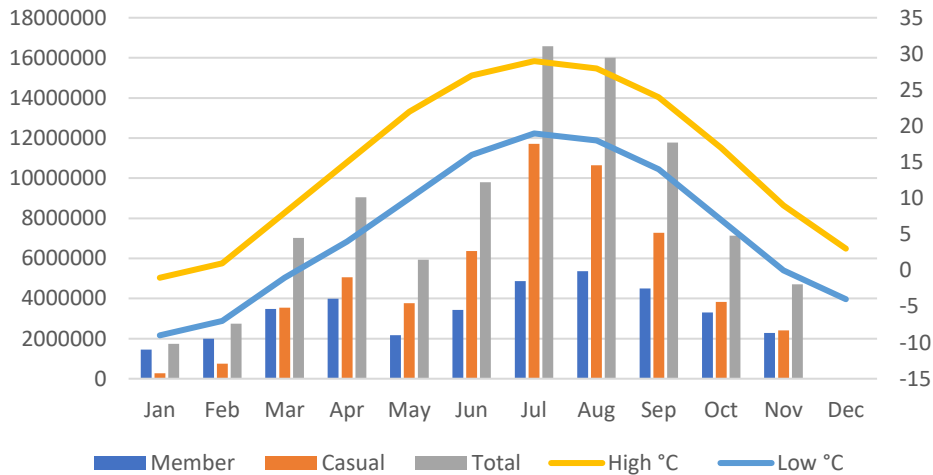
1. Data for December was corrupted and thus, is missing from this analysis. Temperate data from [here](#).

2. From examining the trends, it can be seen that Cyclistic Users travel more in July and August than any other month. The number of rides is fairly consistent with the temperature, except for May and June.
   - It is important to note that the data is between May '19 and April '20, the lower number of trips could be attributed to the company having fewer customers earlier.

3. The duration of trips for casuals has high variance and drops significantly 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.
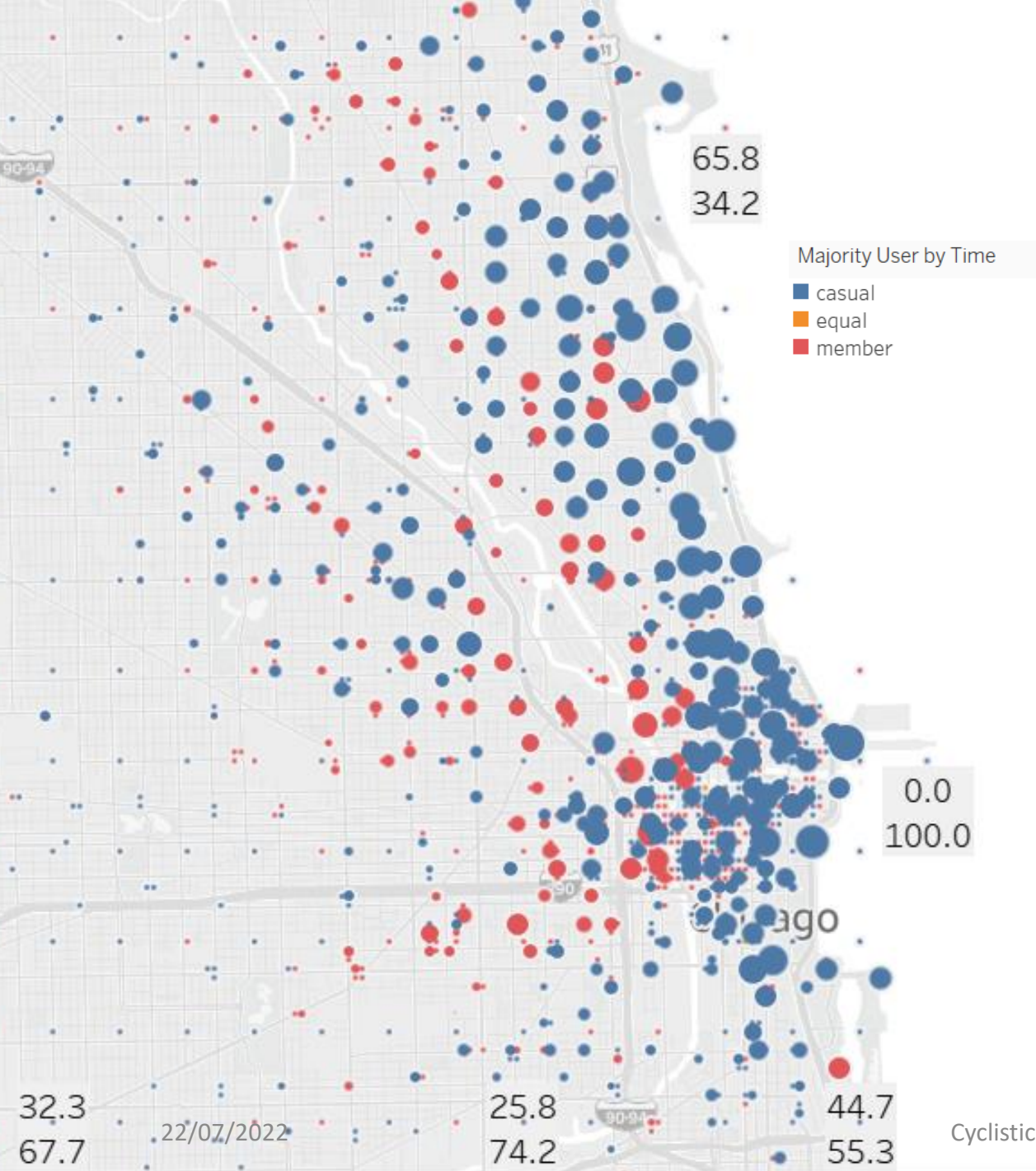
Key Differences

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? [1]

These are inferences supported by facts from the analysis.

1. Casuals ride for longer durations although these trips are less frequent. This is suggestive of Casuals tending to ride bikes more for leisure.

    - 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 Casuals trips, 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 routinely start their trip from the same biking station, driving up the average for that location.

What does the Data say? ₂

These are inferences supported by facts from the analysis (continued).

3. Members use bikes as a form of commuting.
    - This assumption is backed up by the fact that members had more trips during weekdays with a significant dip on Sundays while the trend was the opposite for Casuals. This commute could be to work/school/college etc.
    - A small inconsistency can be observed on Monday. The time spent riding is the lowest but the number of trips remains close to average.

4. The subscription model is biased.
    - Casuals rent out bicycles infrequently, while members use the same services frequently tells that the current subscription model tends to favour only members. Single-Ride/Day passes are thus more beneficial for such customers of Cyclistic.
    - A subscription model that is based on the number of trips rather than time spent riding is at fault. It incentivises frequent usage of services while providing zero benefits for longer rides.

Steps to Consider [1]

Possible routes to increase the member percentage among Cyclistic customers.

1. Revise the subscription model to incentivize time spent travelling over more trips.
    - The cost of making 10 trips of 3 minutes each should be the same as that of a single 30-minute trip. Charging customers for the duration of the trip is fair and unbiased.
    - This will promote subscription among repeating casual customers who use the services infrequently.
    - 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.
    - This would also minimise losing older members.

Steps to Consider ₂

Possible routes to increase the member percentage among Cyclistic customers (continued).

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.

4. Replace Single-Ride/Day passes with time-based/hourly passes.
    - Single-Ride passes are too ambiguous since they do not specify the length of the ride. This makes them a better alternative for infrequent travellers.
    - Introduce time-based passes with a discounted rate for members.

Going Forward 1

1. Survey users to gain 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 regarding the habits of a casual/member. Understanding users' needs will help the company offer better, more personalized services.

2. Reconsider costs for services.
   - On moving toward a time-based model Cyclistic is sure to observe a change in revenue. Estimate and compare revenue from the original and newer subscription model to set new appropriate prices.
   - Even if the company decides to not establish a new model, it is evident that Single-Ride/Day passes are too cheaply priced. Cyclistic should consider either increasing the cost or adding time constraints to these passes.

3. Consider partnering with Workplaces/Educational Institutions etc.
   - A considerable number of casual use the services every day but do not buy a subscription.
   - Partnering would allow Cyclistic to add more members while offering reserved bikes for said any member belonging to such partners. The company could also consider providing private docking stations.

Going Forward ₂

4. 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 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.

5. Use Social Media to establish a Cyclistic Biking Community.

   - Engaging with customers as a community could make subscriptions commonplace, even for entire households.

   - 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.

   - The company could also promote biking as a pollution-free mode of transport appealing to the climate-conscious members of society.

Fin.

Completed by

Apnatva Singh Rawat
5th August 2022.

Google Data Analytics Case Study