

Google Data Analytics Certificate

Capstone Project: Cyclistic Bikes

Project Report

Ask

Business Task Statement

The business task we will solve is how do annual members use Cyclistic bikes differently from casual riders; and subsequently, how do we convert casual riders into annual members? Our stakeholder is Lily Moreno, Director of Marketing at Cyclistic, and will oversee the implementation of the recommended strategy.

Prepare

Data Sources

Where is the data located? For the purpose of this exercise, the data has been sourced from Divvy Bikes, Chicago, all data has been sourced from this link:

<https://divvy-tripdata.s3.amazonaws.com/index.html> The data is now stored locally on my laptop, and on Google Cloud, for analysis in Google BigQuery. The aggregate data is also stored on Tableau Public for the purpose of creating visualisations.

Data organisation: Data collected from April 2020 onwards is organised into monthly packets. (Before April 2020 the data was organised into quarters, but this data does not form part of this analysis.)

Bias / Credibility of the data: the data is considered reliable as it has been collected by the company in question, not a third party. The data is original because it has not been tampered with before any processing within this analysis. The data is comprehensive because it covers a broad period of time, and no data has been omitted during capture. The data is as current as possible, considering we want to analyse a whole year of data. And the data can clearly be cited as collected by the company whom it is about.

Licencing: The data has been used under this licence:

<https://divvzbikes.com/data-license-agreement>

Privacy: The licence states that data must be anonymised. All personally identifiable information such as credit card details have already been removed.

Security: The raw data is open source, so there are no security concerns for this project.

Integrity: The data has integrity because it was sourced internally to the company which it is about (for the purpose of this exercise). I question the reliability of a small number of records, for example those with duplicate ride-IDs, or rides which appear to end before they begin. The duplicate records will be disregarded, to avoid skewing the analysis, these are due to a crossover of recording data between the May and June data sets . We have also been made aware that the rides that appear to end before they start are as a result of bikes being taken out of service for maintenance. These records will also be disregarded.

Answering the Question: The data will allow the question to be answered because I can separate Annual Members from Casual Users and look at how their behaviour tends to differ. I will then recommend a strategy to convert casual users to members.

Timeframe: The data timeframe we will be looking at is January - December 2024 only. No other period of time will be considered.

Code:

The SQL code used for cleaning and analysis can be found in a text file alongside this report.

Process

Cleaning and Manipulation of the Data

Tools: To clean and analyse the data, I have used SQL with Google BigQuery. This has allowed me to combine all months of data, and perform analysis on it as a whole, which would not be possible with the processing power of a local application such as Microsoft Excel.

Integrity: The data's integrity has been ensured by ensuring the total number of records in the combined dataset is equal to the sum of the monthly datasets, minus the number of records we have stated will be dropped.

Cleaning: As part of the cleaning process, I have:

- Combined all months of data into one dataset, and excluded duplicate Ride-IDs, as we are aware between May and June some rides were recorded into both monthly datasets.
- Calculated the length of each ride, and added a column for this information.
- Calculated the day of the week that each ride took place on, and added a column for this information.
- Converted the day of the week format to: 1 = Monday, 2 = Tuesday... 6 = Saturday, 7 = Sunday, so that a more user friendly visualisation will be produced.
- Deleted records where the ride length is less than or equal to 0, ie where the end time of the ride is before the start time, as we have been made aware that these records mean the bike was out of service for maintenance, so we do not want these records to form part of our analysis.

We know that data has now been cleaned because:

- The number of records in the combined data set is equal to the sum of the records in the monthly datasets, less the number that have been dropped due to duplicate ride-IDs, or ride lengths that are less than or equal to 0.
- The ride lengths are now all positive values.
- The day of the week value is now always an integer between 1-7. Spot checks on different days of the week have been carried out, against an alternative source, and these were found to be correct
- Rideable_type field has been checked and it contains one of three values: electric_bike, classic_bike, or electric_scooter. Field format is string.
- Member_casual field has been checked and it contains one of two values, member or casual. Field format is string.
- Month of ride, which is extracted from the started_at field, and is always an integer between 1-12.

Analyse

Organisation of Data: the data has now been organised and 5 key fields have been identified:

- Day_of_week
- Rideable_type
- Member_casual

- Ride_length_in_minutes
- Month (extracted from the “started_at” field)

Summary of Analysis

Trends / Relationships: The visualisations produced as part of this project, included below in this report, are recommended for showing the trends most effectively. However I will provide a summary here too.

Overview

- Members took 72% more rides than Casuals.
- Casuals’ rides were on average 97% longer than Members’ rides.
- As a result, 14% more minutes of ride time were hired out to Casuals than Members.

Day of the Week

- Number of rides: Members make more journeys during the week than at the weekend, while the reverse is true for Casuals. That said, the number of weekend trips made by Casuals is still less than that of Members.
- Average length of each ride: The average length of Members’ rides stays fairly consistent throughout the week with a slight increase at the weekend, while the average length of casuals’ rides increases more substantially at the weekend.
- As a result, 85% more minutes of ride time are hired out to Casuals than Members at weekends.

Month of the Year

- Number of rides: Both Members and Casuals take substantially more rides in the summer months, with a sharp decline from October to December. The number of rides re-increases from January - September. Members take more rides than Casuals, in all months of the year.
- Total length of all rides: The trend is similar for the total length of all rides, the summer sees a large increase, while the winter sees a large drop. The summer sees Casuals overtake Members in this metric, and the winter sees Members re-overtake Casuals.
- Average length of each ride: Members’ average ride length stays similar throughout the year, while Casuals see a more defined change throughout the year: Casuals take the longest journeys in the summer, this declines as we move into winter. Surprisingly, October and February buck this trend, October more so than February.

Rideable Type

- Classic bikes are used by Members 81% more frequently than Casuals, however Casuals’ average ride lengths are 162% longer, resulting in the total ride time for Casuals being 45% more than Members.
- Electric Bikes are also used more frequently by Members than Casuals (73% more), while the average ride length for E-Bikes is only 29% more for Casuals.
- This means that, conversely to Classic bikes, electric bikes are ridden for 34% more minutes by Members, than Casuals.
- Electric Scooters are used substantially less than the other two rideable types (they account for just 1.48% of all ridden minutes), yet there are 44% more rides on electric scooters by Casuals, than Members.
- Casuals also unsurprisingly take longer journeys on e-scooters on average, resulting in the total number of minutes hired out to Casuals being 108% more than to Members.
- Overall, Classic Bikes account for 61.96% of all minutes hired out, E-Bikes account for 36.56% of all minutes hired out, and E-scooters account for 1.48% of minutes hired out.

How does this analysis answer the business question:

Business question 1: How do annual members use Cyclistic bikes differently from casual riders?

- Casuals tend to take fewer, longer journeys than Members; and Members take more, shorter journeys. Casuals account for 14% more ridden minutes than Members.
- Weekends are the most popular part of the week for Casuals, while weekdays are the most popular part of the week for Members.
- I would suggest this shows that Casuals tend to use Cyclistic for Leisure purposes where they take their time riding and use the service more at weekends, while Members tend to use the service for commuting purposes, as they tend to use the service more during the week, and tend to be in more of a hurry.
- Both Members and Casuals use Cyclistic more in the summer than the winter. This trend is more exaggerated for Casuals, resulting in Casuals overtaking Members on ridden minutes in the summer months. I suggest that this is due to Members and Casuals preferring to use traditional public transport in the winter, to keep warm and dry; and I would suggest this is more prevalent for Casuals (tourists) because Members (professionals) often don't have a choice but to get to work quickly, whereas tourists are more likely to be in a position to take their time and choose.
- Looking at Rideable type: Casuals' average journey lengths are longer on all rideable types, however, this is particularly the case on Classic Bikes, I suggest this is because Casuals tend to be less experienced riders so cannot cycle as fast as Members.
- Members have shorter average journey lengths, which shows (unsurprisingly) they tend to be more experienced riders. The trend also exists for E-bikes and E-scooters, however by nowhere near as much, likely because the speed of E-bikes and E-scooters is determined less by the experience level of the rider.
- E-scooters are more popular among Casuals, this could be explained by Casuals seeing E-scooters as a novelty mode of transport; Casuals may be in the city for leisure purposes and want to experience something new, whereas Members may want a faster and more reliable mode of transport such as a Classic Bike or E-bike.

Business Question 2 - How do we convert casual riders into annual members?

- This business question is answered in the "Act" section at the end of this report.

Share

Data Visualisations

All visualisations show 3 key metrics; which when combined give a good picture of customer behaviour. These metrics are:

- 1 - number of rides
- 2 - average length of ride
- 3 - total length of all rides

Data is always broken down by customer type, as this analysis is all about identifying differences in behaviour between Members and Casuals.

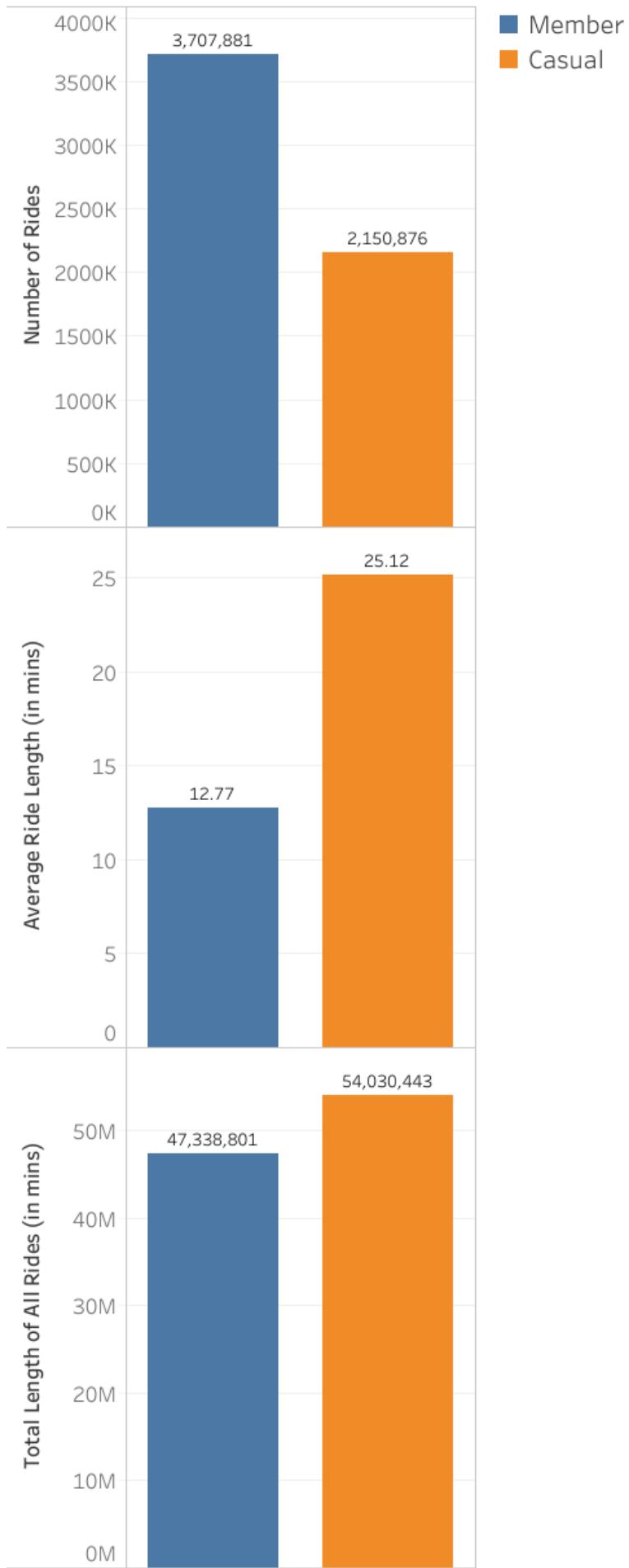
I have created 4 visualisations:

- 1 - Summary - shows how behaviour of Members and Casuals is different on average across the year.
- 2 - Day - how day of the week affects behaviour of Members and Casuals differently.
- 3 - Month - how month of the year affects behaviour of Members and Casuals differently.

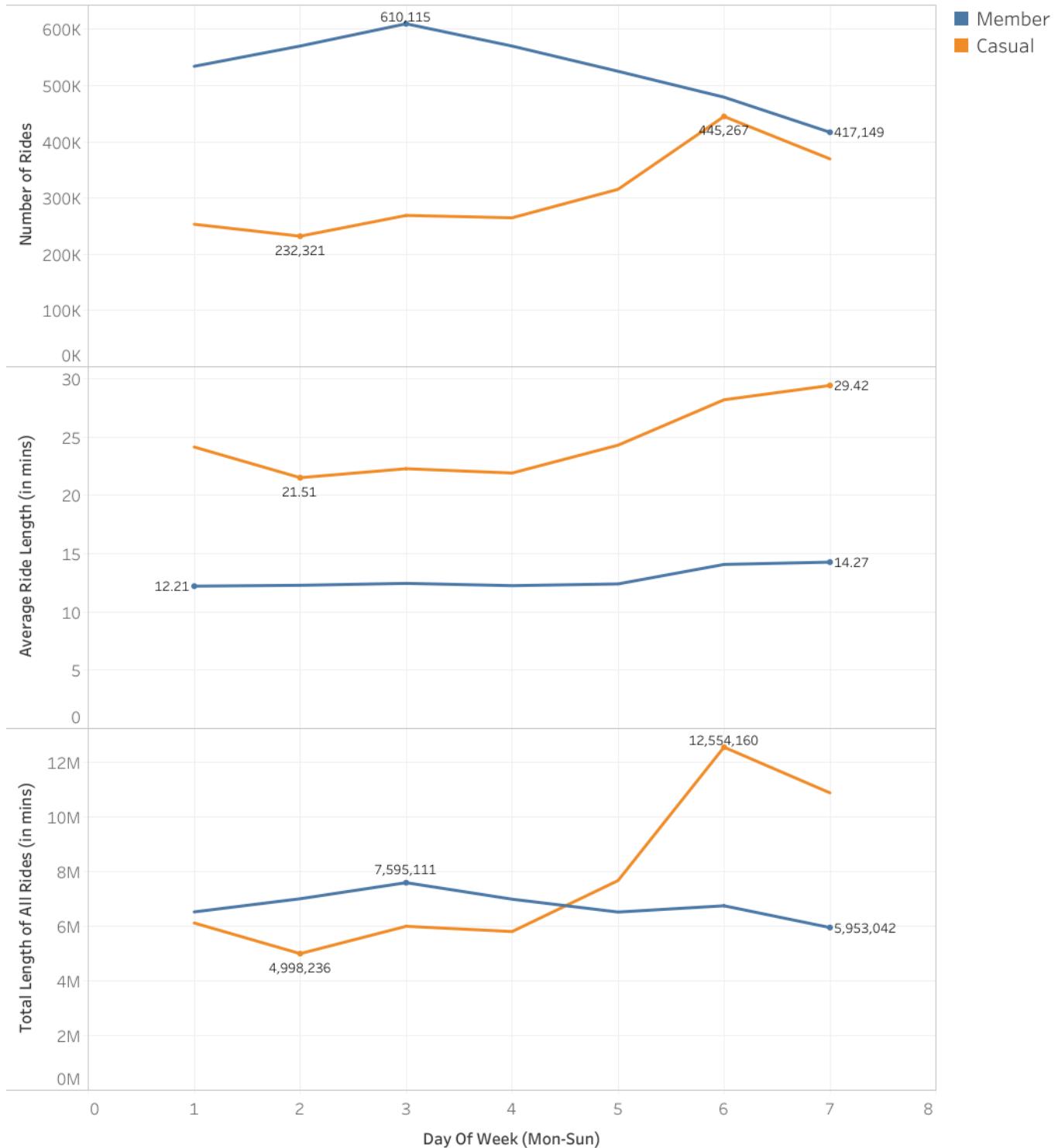
- 4 - Rideable type - how type of rideable affects behaviour of Members and Casuals differently.

The 4 visualisations are included below:

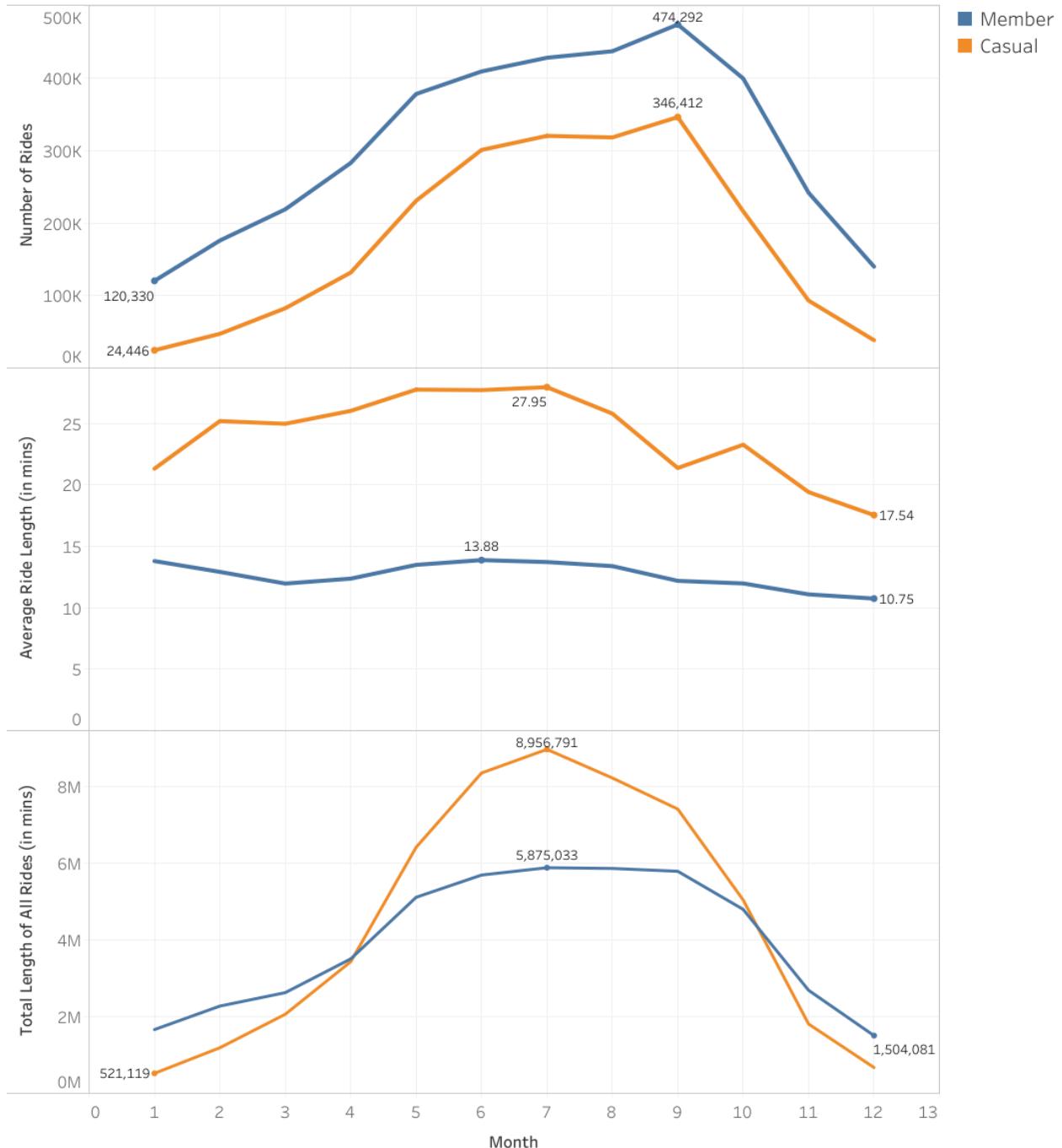
Summary - Whole of 2024



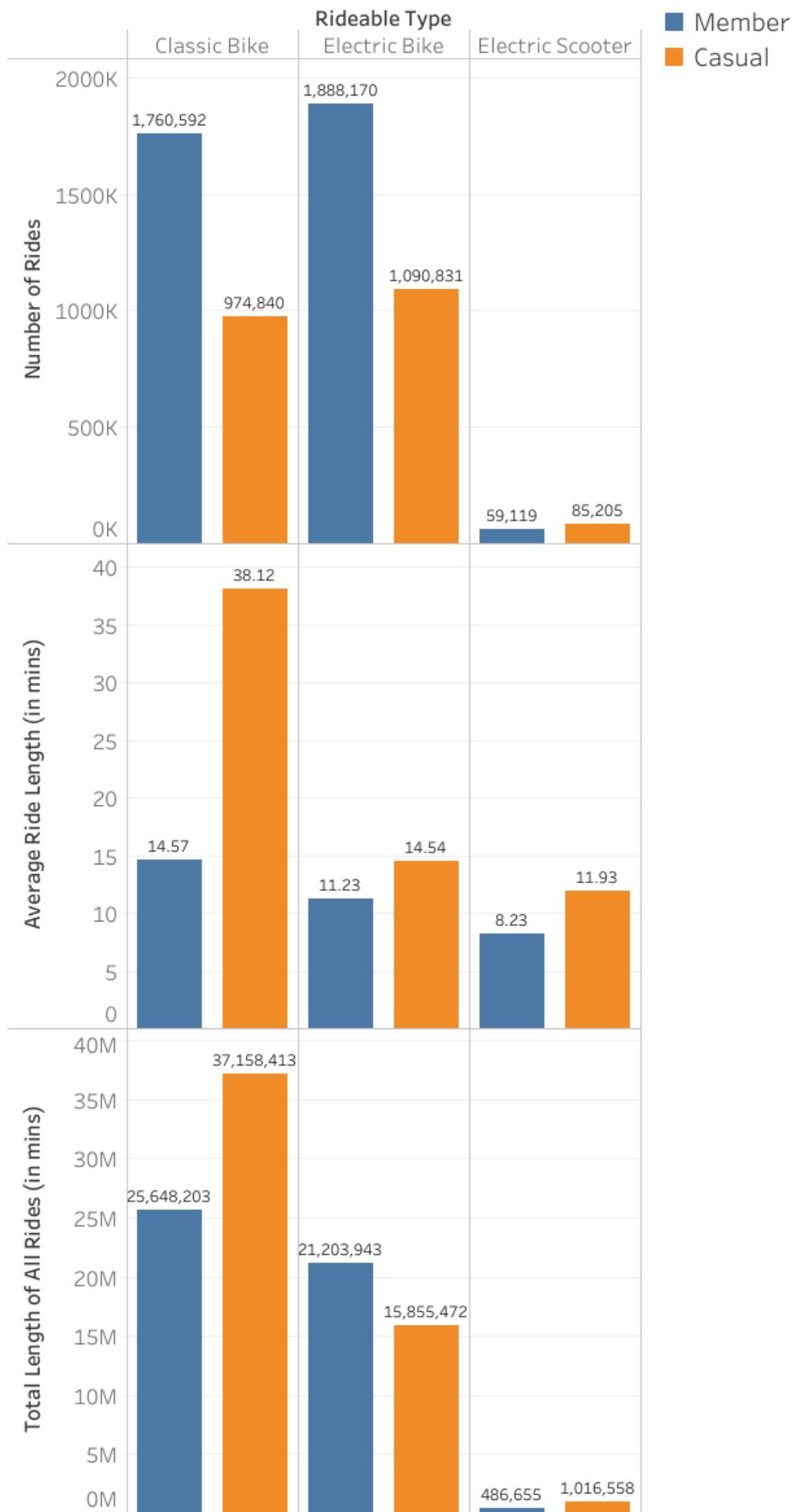
Breakdown by Day-of-the-Week



Breakdown by Month



Breakdown by Type-of-Rideable



Act

Report Recommendations

The data shows that Casuals and Members use Cyclistic for different purposes, and different variables affect Casuals and Members behaviour in different ways. Below I recommend a promotion that takes advantage of what we have learned from the data in this analysis, in order to achieve our goal.

I recommend a targeted marketing campaign offering temporary discounts to casual riders if they use the service little-and-often during the working week, for example for commuting to work, or for completing their commute from a railway station to their office. Another example is for professionals traveling around the city to business meetings, or out to lunch. These journeys would also open up the possibility of a return journey using Cyclistic. This will prompt a change in habit, in that these customers now habitually use Cyclistic in the style of Members. When the promotion is removed, it will then be worthwhile to the customer to switch to a Member subscription, at which point our goal has been achieved.

Casual weekend riders are a source of revenue that is essential to retain, in fact Casuals account for 14% more ridden minutes than Members. Therefore, I suggest that the terms of the promotion are set so that the discount is only available if customers keep up their existing patronage at weekends as well.

I also recommend that the promotion is introduced gradually, and monitored closely - Casuals provide more revenue than Members; while our aim is to increase patronage, it is important to retain this revenue from current-Casuals even when they become Members. In the unlikely event that the strategy is having the wrong effect, then a gradual introduction of the promotion will reduce losses, and an alternative strategy can then be devised.

The promotion can also aim to increase Casuals' use of Classic Bikes, this will improve their fitness, resulting in a faster ride in future, resulting in happier customers, resulting in customers who want to return to Cyclistic more often.

If successful, this strategy will lead to a long term increase in company revenue.