Google Capstone Project

Cyclistic Case Study

Jan 2022 – Dec 2022

By Oliver Peel

Welcome!

My name is Oli Peel

I am an ex-auditor aspiring to be a data analyst

Email : oli-peel@hotmail.co.uk

LinkedIn: https://www.linkedin.com/in/oliver-peel-9a8025271/

Introduction to the Project

Cyclistic are a bike-share company

Bikes are available for users to ride between docking stations.

There are two different types of users:

Casual – Users who pay for each individual ride

and

Member – Users who pay an annual subscription fee for full access

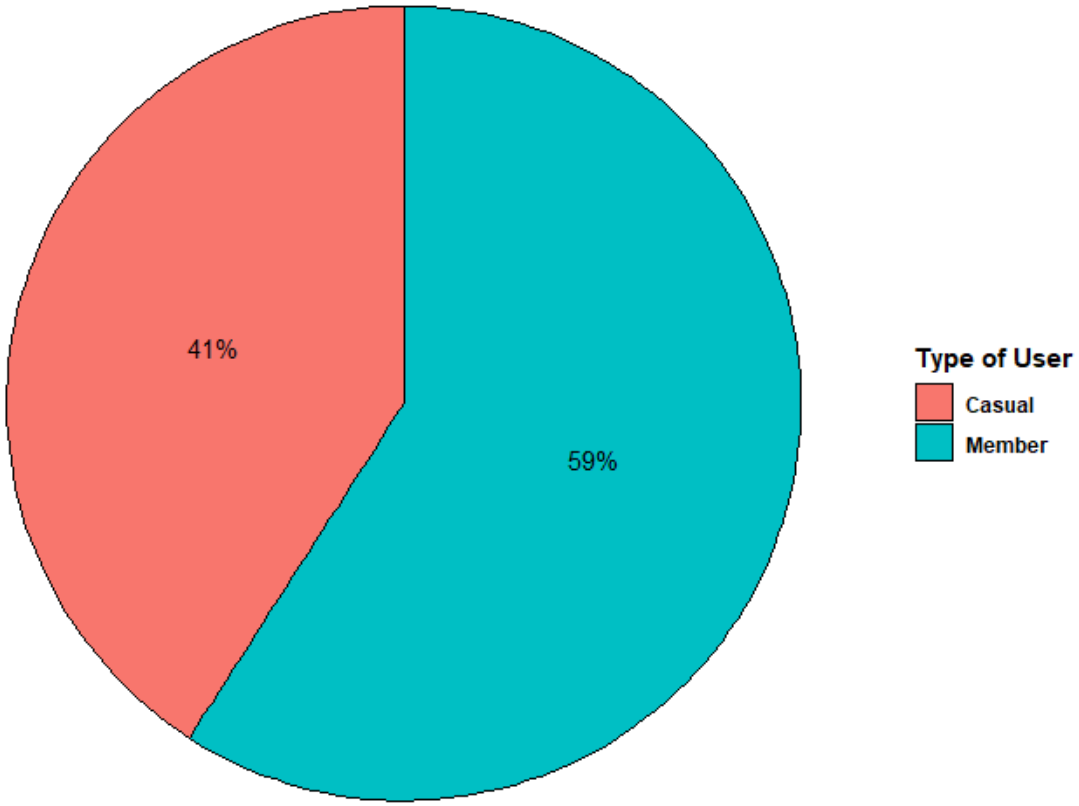
Business Task

Cyclistic wants to market to **casual** riders in order to convert them into annual **members** as part of their business strategy.

To understand whether or not this will be a good marketing strategy and to get insight into how to implement it our task is to determine:

How do annual **members** and **casual** riders use Cyclistic bikes differently?

This graph shows the ratio of journeys ridden by **members** and **casual** users

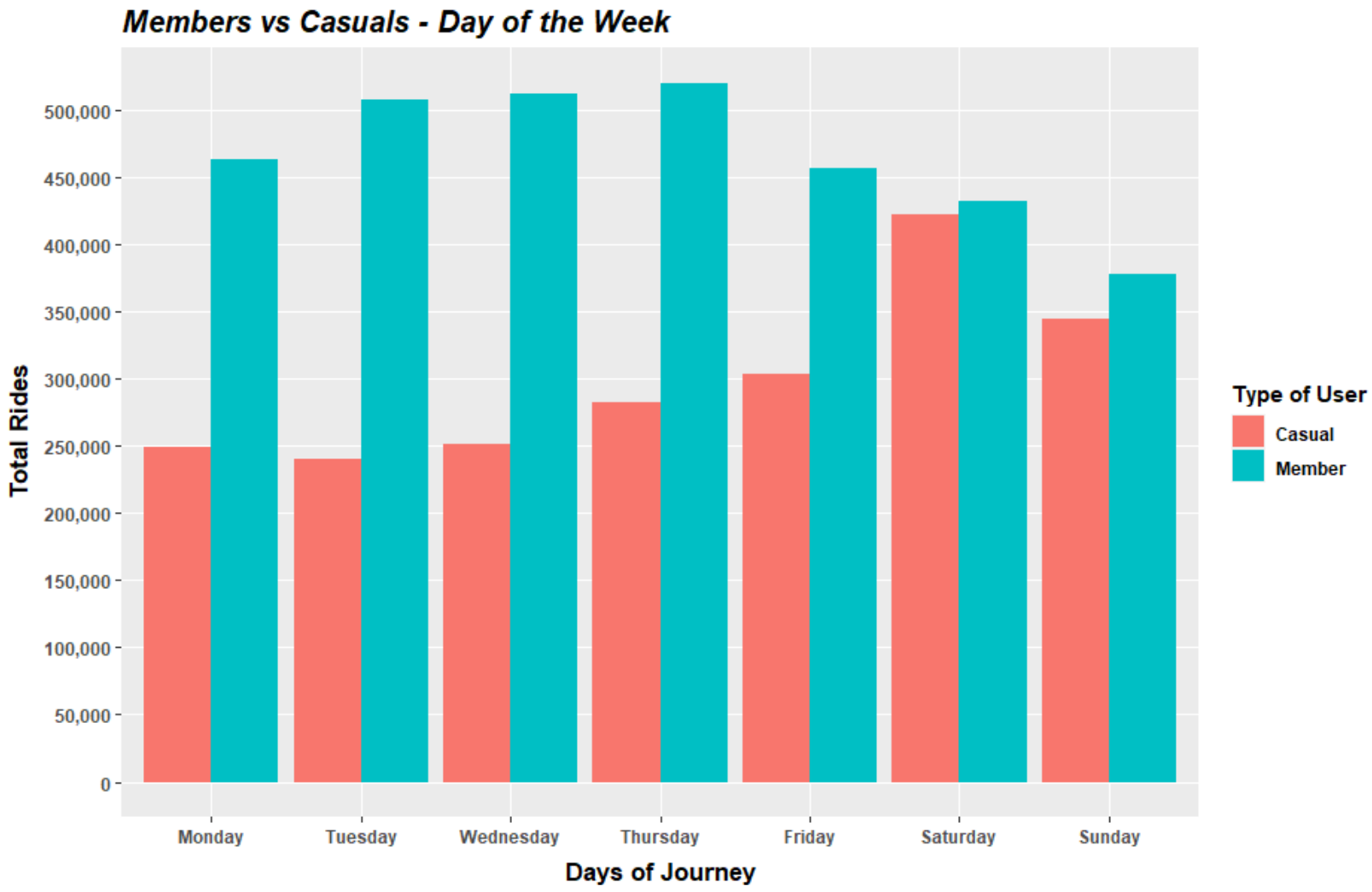


Note: Individual riders cannot be identified.

This graph does not show populations of the user types.

Member vs Casual Journey's in 2022

Members ride more often on weekdays than weekends

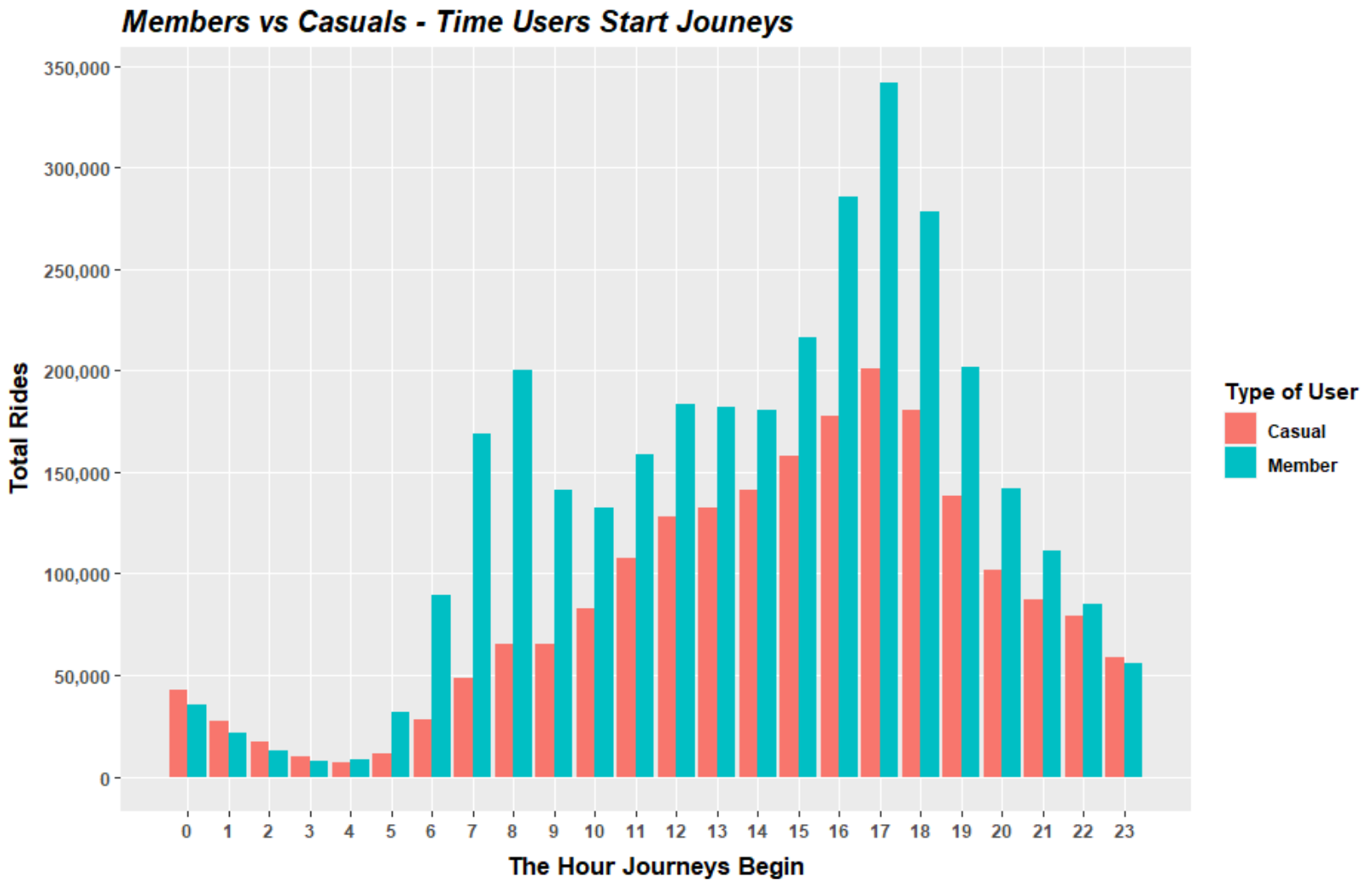


Casual users ride more often on weekends than weekdays

Rides by Day of the Week

Members start their rides around commuting hours most often.

The two peaks at 7am and 5pm is unique to **members**.



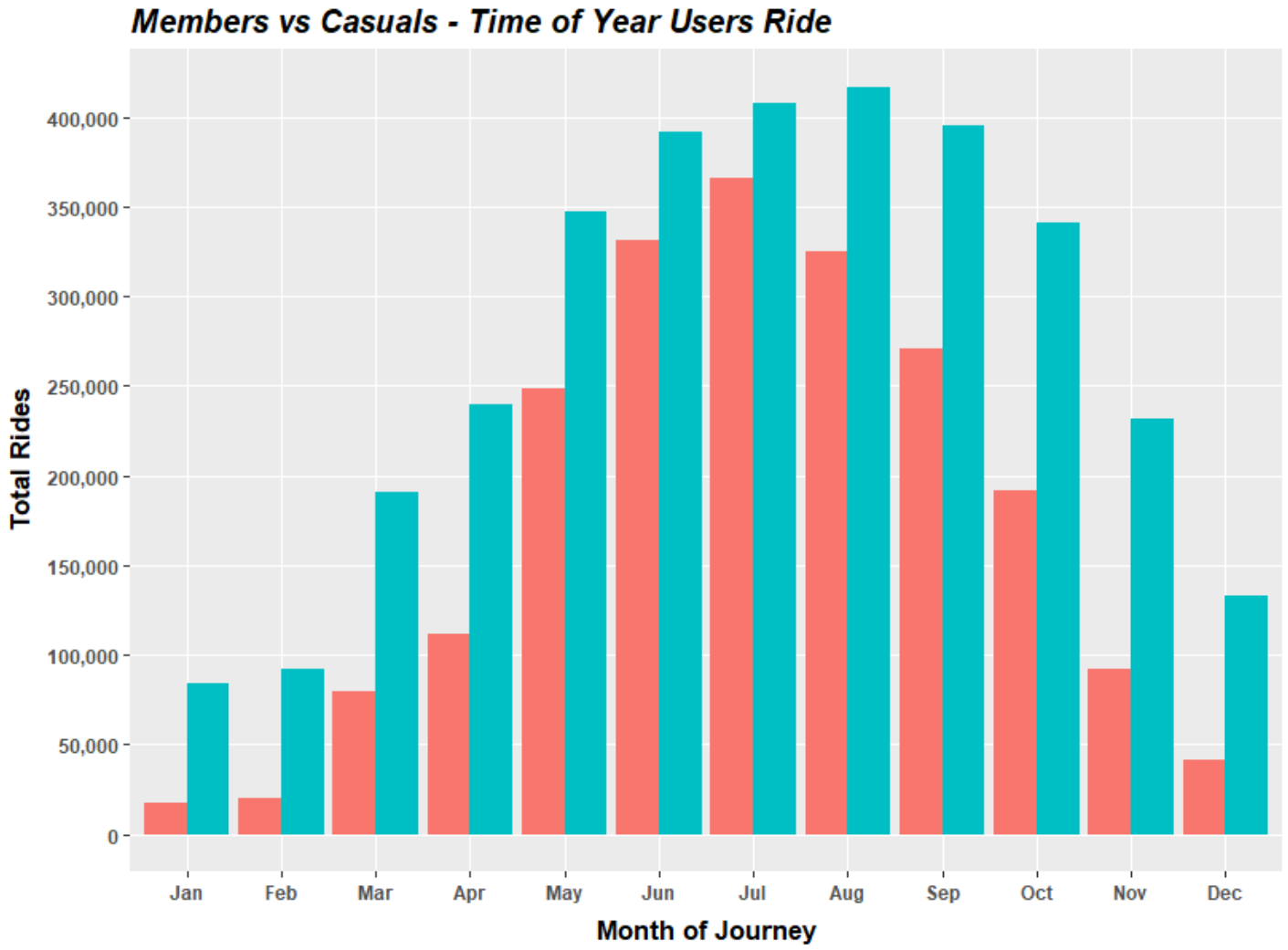
Casual users have a preference for riding in the afternoon.

Rides by Hour Journey's Begin

Members ride all year round.

They prefer the Spring through to Autumn.

Members make up a significant proportion of cold weather rides.

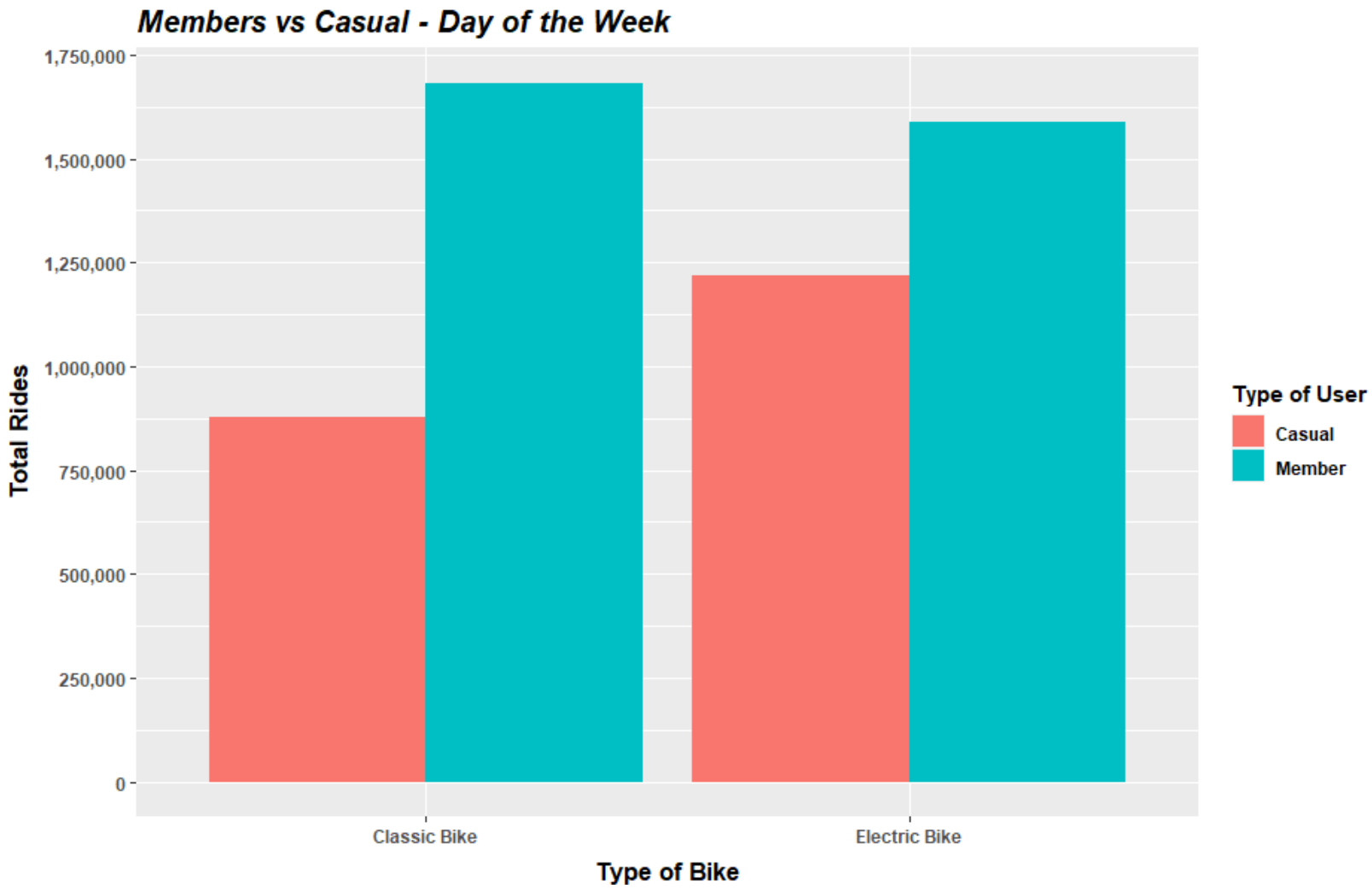


Casual users have a clear preference for the Summer months.

Casual users make up a large proportion of warm weather rides.

Rides by Month of the Year

Members ride both electric and classic bikes with an even split

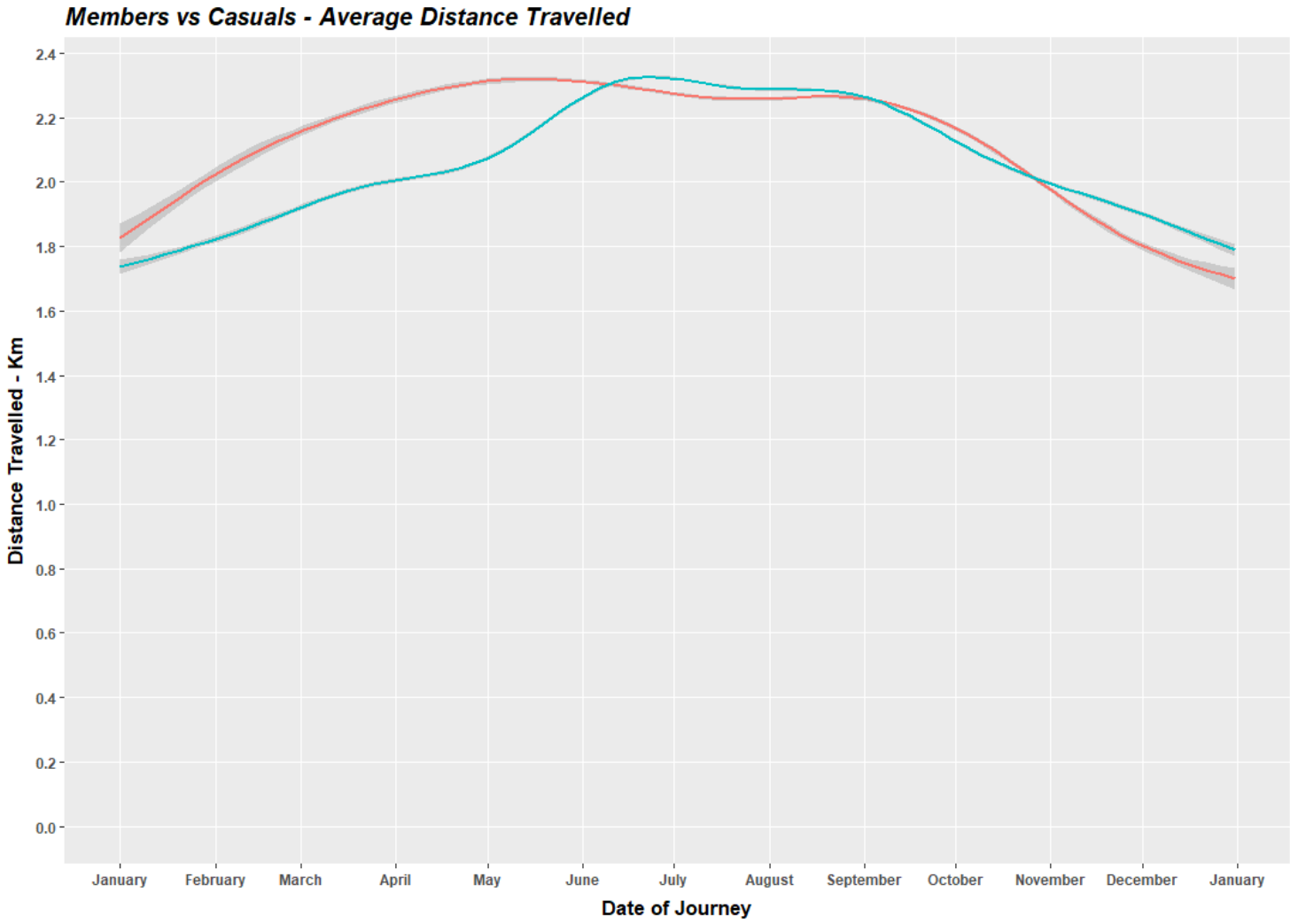


Casual users have a clear preference for electric bikes.

Roughly 60% of their journeys are on electric bikes.

Rides by Type of Bike

Members and **casual** users tend to ride very similar vectors.



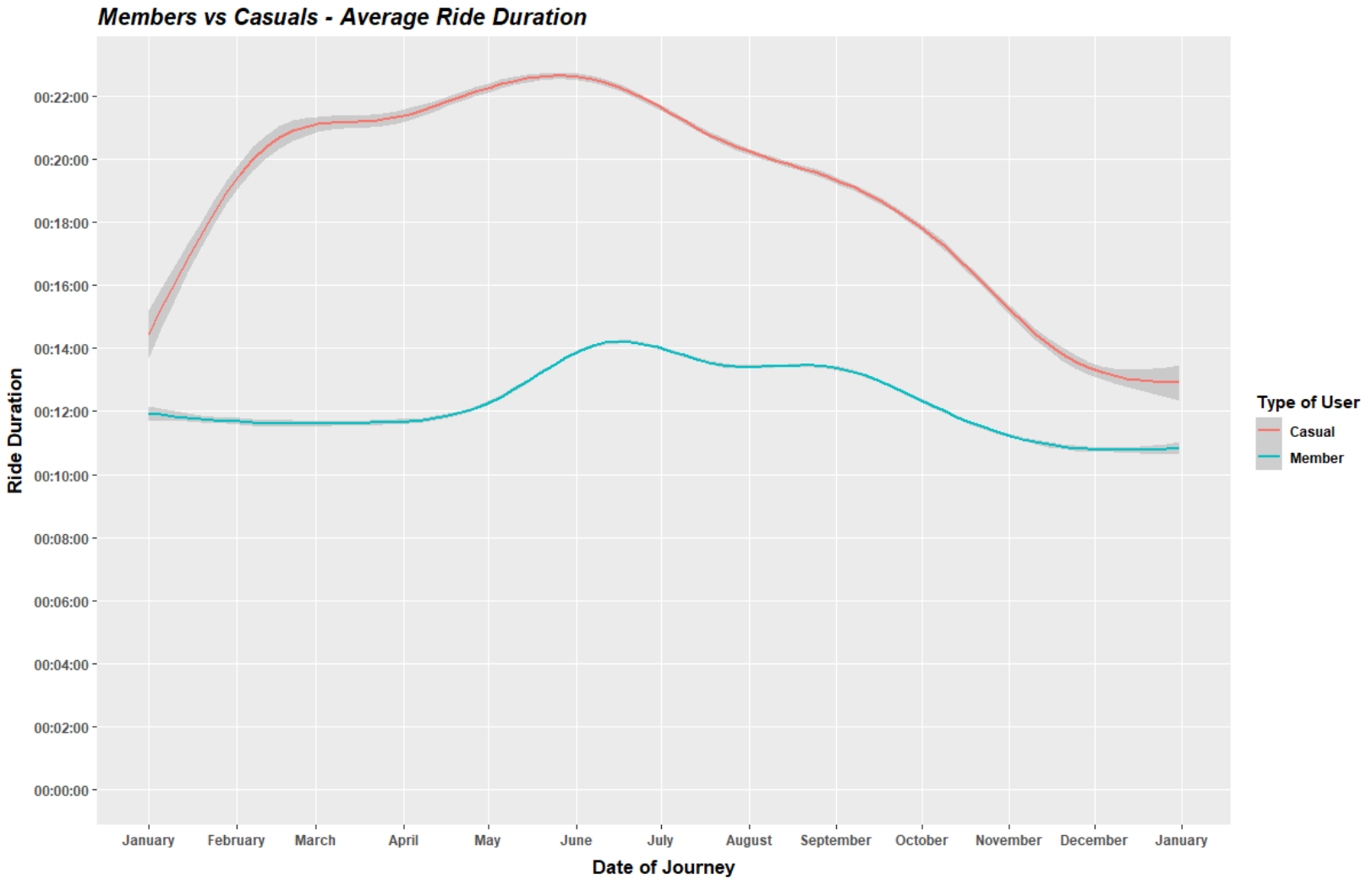
Note: This graph shows distance travelled as the crow flies.

We can only measure the distance between the start and end points

Average Distance Travelled over 2022

Members' rides are consistent in average duration.

The average duration for **members** is also always faster than **casual** riders



Casual users spend longer cycling as the months get warmer

Average Ride Duration over 2022

Hypothesis

The data shows a difference in how **members** and **casual** riders use Cyclistic's service.

Based on the trends shown in the data I have a hypothesis that:

Members primarily use Cyclistic bikes for travel and commuting

Casual users primarily use Cyclistic bikes for leisure and exercise

Summary of Key Findings

- **Members** have a preference for weekday rides
- **Casuals** have a preference for weekend rides
- **Members** start their journeys around the beginning and end of normal work hours
- **Casuals** have a preference for summer rides
- **Members** travel to their destinations faster than Casuals despite travelling the same average distances.
- **Casuals** have a preference for electric bikes

Evidence is Inconclusive

The data justifies the hypothesis, however it does not provide enough evidence to reach a conclusion.

There is not enough data to be certain the hypothesis is correct.

Further analysis with additional data would be needed to gain additional certainty.

Currently there is a high risk that an overlap exists between how members and casual users utilise Cyclistic.

Without identifying the scope of the overlap we cannot gauge how successful future marketing schemes could be.

Next Steps

My recommendation is to gather additional detailed data and repeat the analysis.

This is a low risk approach as we can be more certain that any recommendations that come out of the further analysis will add meaningful value to the company.

The other option would be to carry on with the current recommendations based on the hypothesis. This is a higher risk strategy as we lack certainty around the hypothesis.

Gather Additional Data

To best accomplish the business task, additional data will be helpful.

Whilst seen through the current data that **members** and **casual** users use Cyclistic differently we haven't been able to conclude why; nor can we be certain how they differ.

By gathering additional data we can find out why **members** and **casual** users use Cyclistic differently; by learning this we can consider how to best improve the service to suit each groups needs.

Gather Additional Data

In order to clear up some of the issues I had with this round of analysis some specific additional data would be helpful.

Bike speedometer data would be useful. It would allow us to:

- Collect distance travelled by the user
- Collect max speed of the user
- Collect average speed of the user

This information will allow us to more accurately model how members and casual users use Cyclistic in terms of commuting, exercise and leisure.

Gather Additional Data

Two areas of sensitive user data that would be helpful to further improve my current analysis are GPS data and unique user IDs.

GPS data would be a luxury data source. It would add to the speedometer data suggestion to be able to map out where user have travelled.

This would further our understanding of how Cyclistic **members** and **casual** users differ.

Gather Additional Data

By adding a unique ID number for each user to their ride data we could use that to:

- Learn the number of **members** and **casual** users who made a journey.
- Find the average number of journeys made by **members** and **casuals** per year
- Split the users into categories based on how often and when they rode
- Analyse how many journeys are repeat journeys and how many are one-offs.

Gather Additional Data

It must be noted that using the data suggested will require caution.

The data must be gathered and utilised in such a way that individual users cannot be identified.

Ways to do so include:

- Generating a new unique user number for each user, then removing access to their current user ID
- Limiting the data connected to the GPS data to just member type, bike type and Speedometer data

Gather Additional Data

Through this quantitative data we could more confidently conclude how **casual** users and **members** differ in their use of Cyclistic.

By analysing the user ID data it may be possible to identify specific groups of **casual** users who use Cyclistic like current **members** do.

Such users would be good candidates to advertise **membership** to as they are likely to benefit from **membership**.

Gather Additional Data

Another approach that would work well in tandem to the quantitative data would be to gather new qualitative data.

By surveying a randomised group of **casual** users and **members** we could ask questions to support our understanding of how users use Cyclistic.

Also we can use the questionnaire to learn why users use Cyclistic as they do.

Gather Additional Data

Some quantitative data can also be gathered from a survey. Data points such as:

- Age
- Number of journeys per day / week
- Disposable income
- Weight

These can be complemented with qualitative data such as:

- Is membership (too expensive, fine, cheap)
- Reasons you cycle (commute, leisure, exercise, travel)

Gather Additional Data

Additionally, questions can be added to get opinions from individual users.

A question could ask about local issues. Are there enough bikes locally?

Members could be asked why they decided to become a **member**? Which benefits of **membership** do they appreciate the most?

Casual users could be asked questions about why they haven't become **members**? What additional benefits would they want to consider becoming **members**? At what price would they become **members**?

Gather Additional Data

The surveys would help build an understanding of the customer base; why people become **members** and what is holding **casual** users back.

With an understanding of the benefits **members** appreciate the most you can improve advertising to focus on these benefits.

With an understanding of issues preventing **casuals** from upgrading to **members** you can consider changes to alleviate these issues.

All of this additional analysis should provide clarity on how best to achieve the business task of converting **casual** users into **members**.

Current Recommendations

If you wish to move on without pursuing further analysis here are my recommendations.

Consider issues that may be stopping **casual** users from converting to **members**.

For example:

- Lack of availability of electric bikes at a **casual** users local docking station
- General lack of bikes available during peak usage hours at a **casual** users local docking station

Current Recommendations

Utilise internal user data to identify individual **casual** users who's behaviour pattens match the established pattern of **members**.

For example:

- **Casual** users who cycle in the winter months
- **Casual** users who ride multiple times per week
- **Casual** users who ride twice in a day during commuting hours

Utilise targeted advertising to show these users how they'd benefit from converting to a **members** account.

Current Recommendations

By utilising localised data at docking stations the company can identify local issues which may deter willing **casuals** to convert to **members**.

Consider adding functionality to the Cyclistic app to show **members** the local availability of different bikes at local docking stations.

Also consider giving the option of **members** to reserve a bike to be used at a certain time as an additional benefit.

Current Recommendations

Consider adding specialised **membership** types to cater for the needs of **casual** users.

For example:

- Summer **membership** passes
  - Catered to the needs of **causal** users who wish to cycle in the warmer months of the year.
- Weekend **membership** passes
  - Catered to the needs of **casual** users who like to cycle but don't have the time to or means to on weekdays.

Conclusion

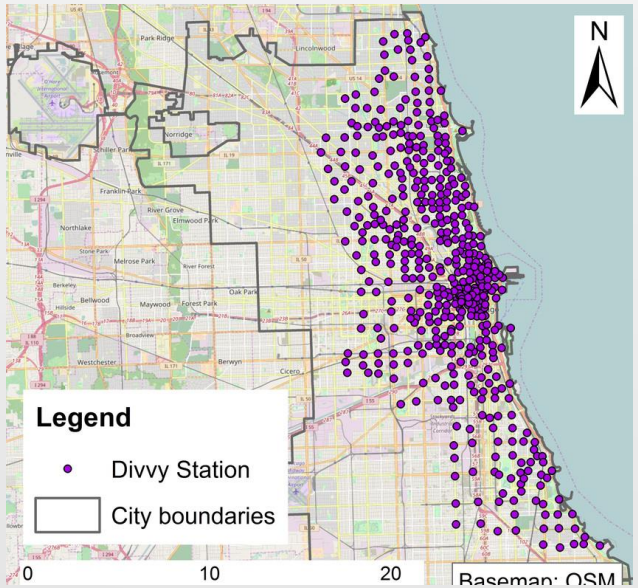
In conclusion, we have seen that **members** and **casual** users do have different behaviours but can't conclude why.

Further important data can be gathered to greatly improve the accuracy of our observations and reduce the risk in our recommendations.

Moving forward is currently a risk without the additional data.

Data Sources

- Cyclistic Monthly Cycle Ride Data January – December 2022
- Map of Cyclistic Chicago Docking stations



Data Cleaning & Manipulation

I completed the data analysis with the R programming language.

Here is a list of the work I did before creating the graphs:

- Transformed the dates into Datetime formats and then split into Date and Time
- Separately created fields for the start hour, start month and for the week day of each ride
- Checked to remove any empty columns. (0 in total)
- Removed any rows with N/a values (5858 in total)
- Created a field for ride duration by subtracting the start time from the end time.
- Converted the latitude and longitude into radians, then used the Haversine formula to calculate the distance between a ride's start and end point in kilometres.
- Removed the columns for the station name's and station IDs from the data as many entries were empty
- Removed rows with "Docked Bikes" as these are under quality control / repair by cyclistic (174585 in total)
- Removed rows where the ride duration was negative. (120755 in total)
- Removed rows where the distance travelled was over 50km as this is impossible (8 in total)

END