Google Data Analytics Capstone Project

# Ask

How do annual members and casual riders use Cyclistic bikes differently? – business task? More detail?

Problem to solve: What strategies can be used to convert casual Cyclistic riders to annual members?

How can my insights drive business decisions? The ways casual riders use their bikes would be expected to be less costly for them. One question that follows is how pricing schemes for Cyclistic can be changed to make switching to an annual membership more advantageous, without Cyclistic losing revenue. Another question, probably the more important one given the stakeholders I’m working with, I need to work out what marketing strategy could be used to convince casual riders to buy annual memberships.

# Prepare

Data has been stored in csv files, one for each month of data. Each row in the file corresponds to one ride. There are over 500000 trips per month. The columns store information such as the start and times of each ride, bike type, start and end locations, and the rider’s membership status.

It is reliable, original (from rideshare source), comprehensive (has every ride over the month except trips that are taken by staff as they service and inspect the system; and any trips that were below 60 seconds in length), current (up to last month), cited (publicly data from Lyft <https://divvybikes-marketing-staging.lyft.net/system-data>).

Data privacy is addressed through anonymisation. This provides the limitation that we cannot attribute multiple trips to one user, so we’re limited in the knowledge of how users use the service. Most trips give a start and end station for the trip, but some have one or both details missing (possibly because it’s not necessary to dock at a station. The latitude and longitude data are inconsistent in precision. Some entries have these to two decimal places, while others have up to 10.

# Process

I’m using Excel, mostly to try it out, Google Sheets was having trouble loading the large datasets. It would be worth trying SQL or R for these steps as they might handle the multiple large datasets better.

Data is organised to show the ride length of each ride, an important feature in analysing how riders use the bikes. We create the feature for the day of the week, which could affect rider behaviour (e.g. members may ride to work on weekdays, casual riders more likely to ride on weekends).

Needed to convert data to correct datetime formats, and numeric values for the longitude and latitude data.

# Analyse

*2024/08 data*

Per pivot table of average time of riders: Casual riders ride for almost double the time of members (26:04 vs 13:25)

Per pivot table of average ride\_length for users by day\_of\_week: Nothing super insightful. Slightly longer ride times on weekends. Interestingly, big jump in casual ride times on Wednesdays.

With count of ride\_id: casual riders ride more Friday-Sunday. Less so for members, with more rides during the week, peaking on Thursdays

*2019Q1 and 2020Q1 combined data*

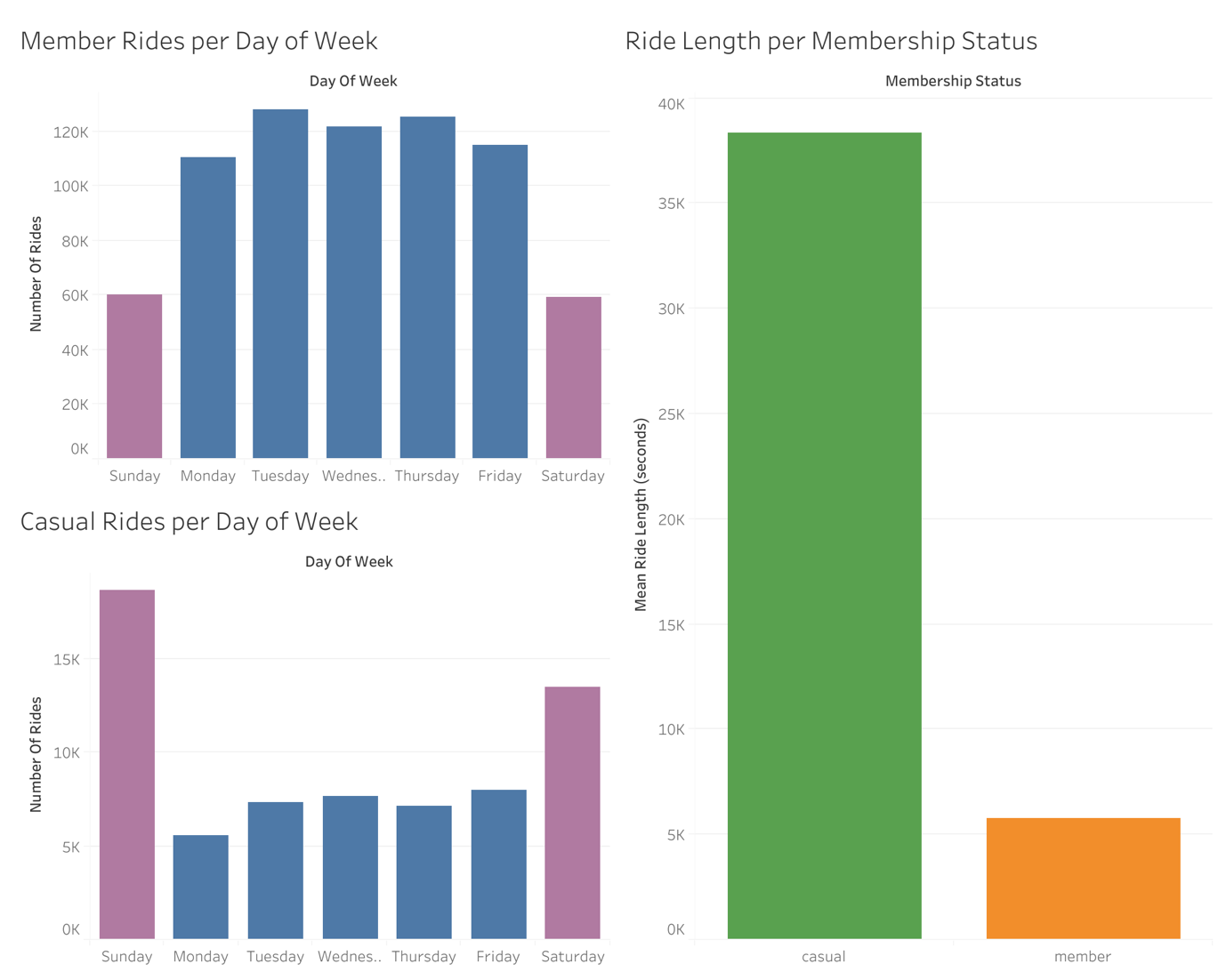
As shown in R template. Would be good to try with more recent data that spans a year. This will be difficult with an increased number of files, and larger file sizes, but hey this is the real world and I and want to do this work in the real world…

Observations are like those for the 2024/08 data, i.e. number of casual rides peaking on weekends and number of member rides peaking mid-week, average ride duration being higher for casuals (though ride times interestingly have a big jump on Thursdays). These differences between casuals and members seem to be much more pronounced in this earlier data, i.e. much more member rides in the earlier data, meaning Cyclistic need our help to gain members no much more than ever. And rather than casual rides being about double the length on average, they’re more than five times the length in this earlier data.

# Share

We have answered the question on how members and casual riders use Cyclistic bikes differently (particularly in use of different days, length of rides). See analyse for story.

Audience is the executive team. They’re probably not interested in the exact facts and figures, so best to give a more general overview of the story and the recommendations that come out of it. As always, visualisations should be easy to understand immediately. The visualisations should help with sharings findings, as seen in the R or excel analysis. We could even keep the bar graph as made, or lines over the days of the week may help with comparisons. Could be stacked for the number of rides, to help visualise the makeup of total rides better, though this probably detracts from the story. We actually only need to make the point that casual , we don’t need to directly compare their numbers, so having two separate bar graphs on a dashboard would be better for illustrating this. Can use slightly different colours for weekends and weekdays.



# Act

The key findings on the analysis are:

* Cyclistic members tend to ride about twice as much on weekdays as on weekends, whereas casual riders are the opposite. This would indicate that members are taking advantage of Cyclistic bikes for their daily commutes, whereas casual riders are more likely to use the bikes for leisure.
* Once again, this indicates members are more likely to use the bikes for convenience, particularly on their daily commute, whereas casual riders take longer, more leisurely rides.

Based on these findings, it is recommended to the executive team that marketing campaigns be utilised to highlight the convenience of having a membership for the daily commute, as well as for short trips in general. There can be emphasis on the time efficiency, cost savings and health benefits of relying on Cyclistic to make these short, regular trips on a membership.

Report

1. A clear statement of the business task
2. A description of all data sources used
3. Documentation of any cleaning or manipulation of data
4. A summary of your analysis
5. Supporting visualizations and key findings
6. Your top three recommendations based on your analysis