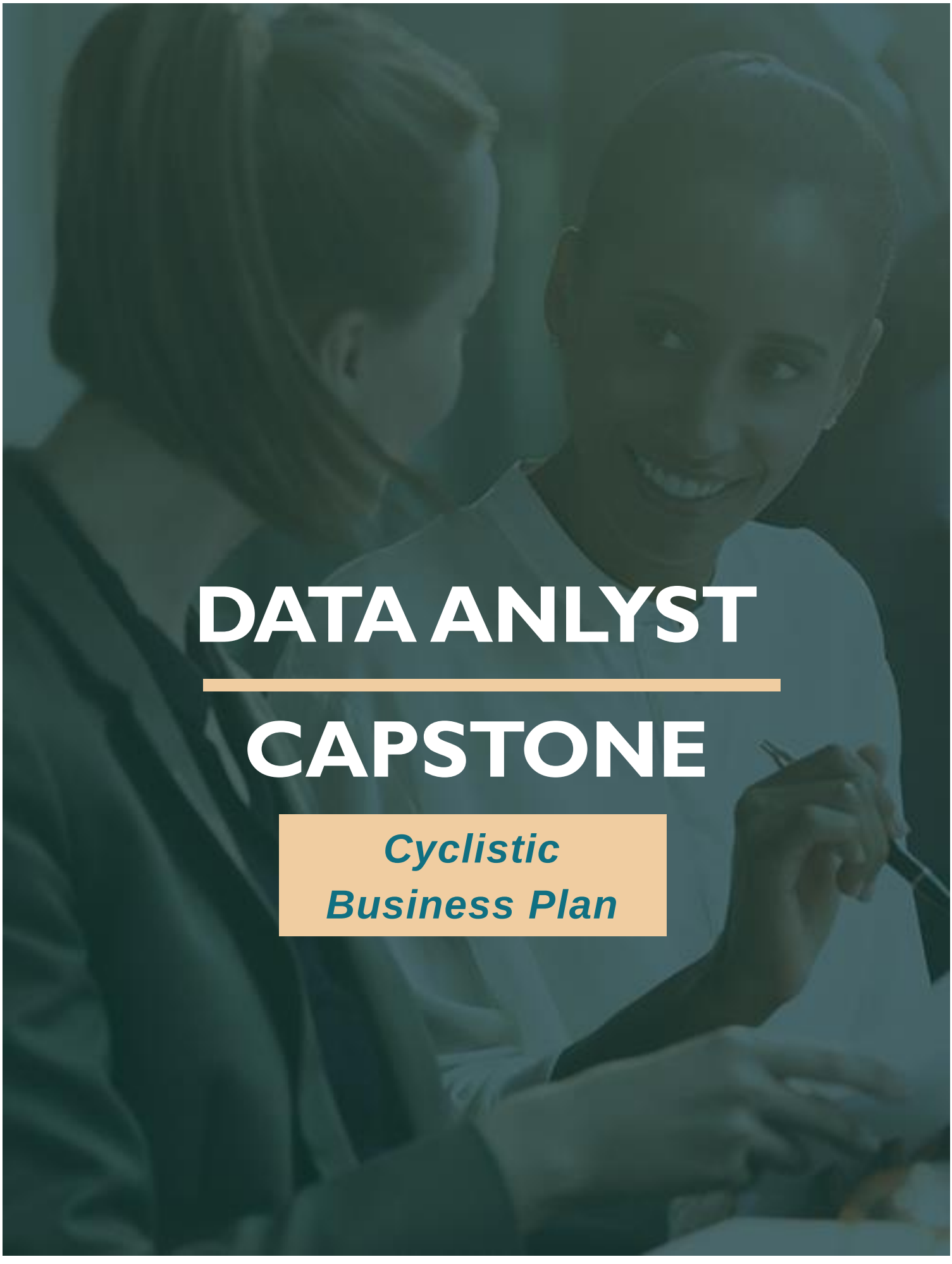


DATA ANLYST

CAPSTONE

Cyclistic Business Plan

TABLE OF CONTENTS

BUSINESS STATEMENT

Cyclistic is a successful bike-sharing company that wants to find a way to increase its revenue. One hypothesis brought up by Lily Moreno is that this can be achieved by increasing the number of annual members. Therefore, the main business statement is on how to convert more casual riders into annual members.

My task is to answer the following question:

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

1. ASK ~> PREPARE ~> PROCESS

The data sources I used to achieve this is the data from the last 12 months from the Cyclist datasets.

Ask

My initial question was the following:

1. Who is using our services?
    - Social-economic status
    - age
2. Why are they using our services?
    - For business or pleasure?
3. How are they using our services?
    - On what days, seasons of the year
    - How often?

All these basic questions will help me categorize the client. Segment them, find trends and understand their behavior in-order-to have our services apply to them much more than the current state.

Prepare

All 12 csv's are in the form of long data with each row containing one fulfilled ride. As in there's a start station with start coordinance and end station with end coordinance.

Because the data is so large one could take 1-2 months of data but that would cause bias for in the winter months are substantially different than the summer months. Therefore, I took all the data.

Process

I decided to use Power BI to handle **5.6M** rows of data.

Because the only data I have is one large flat long table, I decided to split the data into lookup tables. I added a calendar table, a time table and a tier table to group the rides by duration.

Changing names into ID

Attributes that can be linked to lookup tables were changed into numbers. I did this to these columns:

- ridable_type ~> ridable_id
- member_casual ~> member_casual_id

Cleaning the data

Looking at each attribute and knowing the behavior of the company, I can see that there are attributes are I can drop:

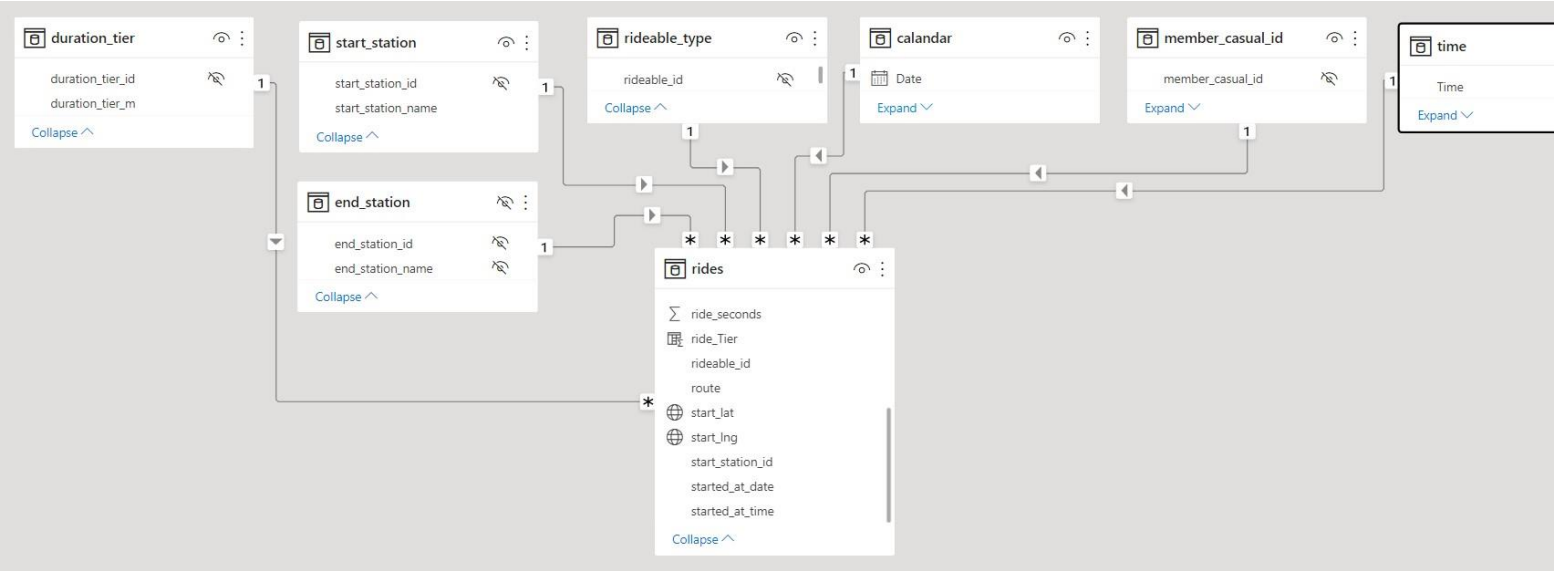
- Ride_id
- Ridable_type
- Memner_casual

I added a column “ride_duration” in minutes and a column “route” that is a concatenations of start station and end station

Even thought keeping NULL data is important to see defective data, this isn’t my objective therefore I removed NULL station ID as well as removing negative ride_durations. This could have been a bug in the system.

To make sure I’m taking reliable data, I also created a flag column flagging out rides that started and ended with the **exact same coordinance**. Something that seems extremally unlikely.

After the data was cleaned and my model is ready, the model looks like this:



A sample date of the rides table looks like this:

start_station_id	end_station_id	start_lat	start_lng	end_lat	end_lng	ride_duration	ride_seconds	ride_minutes	member_casual_id	rideable_id	ride_Tier	ride_flag	started_at_date
TA1307000128	TA1305000030	41.89884116	-87.62986116	41.8849131	-87.632358	0.0038425925925925	32	5	1	3	1	Approved	Monday, July 26, 2
13247	13196	41.895952	-87.66775266	41.8946288	-87.6536505	0.0038425925925925	32	5	1	3	1	Approved	Tuesday, July 6, 2
13427	KA1503000072	41.89047283	-87.62206733	41.8846403	-87.636561	0.0038425925925925	32	5	1	3	1	Approved	Tuesday, July 20, 2
13192	13154	41.919984	-87.64886533	41.9107195	-87.6532045	0.0038425925925925	32	5	1	3	1	Approved	Tuesday, July 13, 2
TA1307000107	TA1309000025	41.96155583	-87.65468183	41.9438193	-87.6490183	0.0038425925925925	32	5	1	3	1	Approved	Sunday, July 18, 2
13139	13307	41.8688275	-87.65907883	41.854805	-87.6636861	0.0038425925925925	32	5	1	3	1	Approved	Monday, July 26, 2
TA1307000117	TA1307000061	41.89137166	-87.62669916	41.8975695	-87.6284576	0.0038425925925925	32	5	1	3	1	Approved	Wednesday, July 21, 2
TA1309000033	TA1309000061	41.921714	-87.65381533	41.9291436	-87.6491161	0.0038425925925925	32	5	1	3	1	Approved	Thursday, July 22, 2
RN-	TA1305000003	41.89089033	-87.63527466	41.8916933	-87.6204641	0.0038425925925925	32	5	1	3	1	Approved	Saturday, July 10, 2
TA1307000129	KA1503000004	41.831065	-87.62687316	41.8312986	-87.6088843	0.0038425925925925	32	5	1	3	1	Approved	Monday, July 26, 2

2. ANALYSE

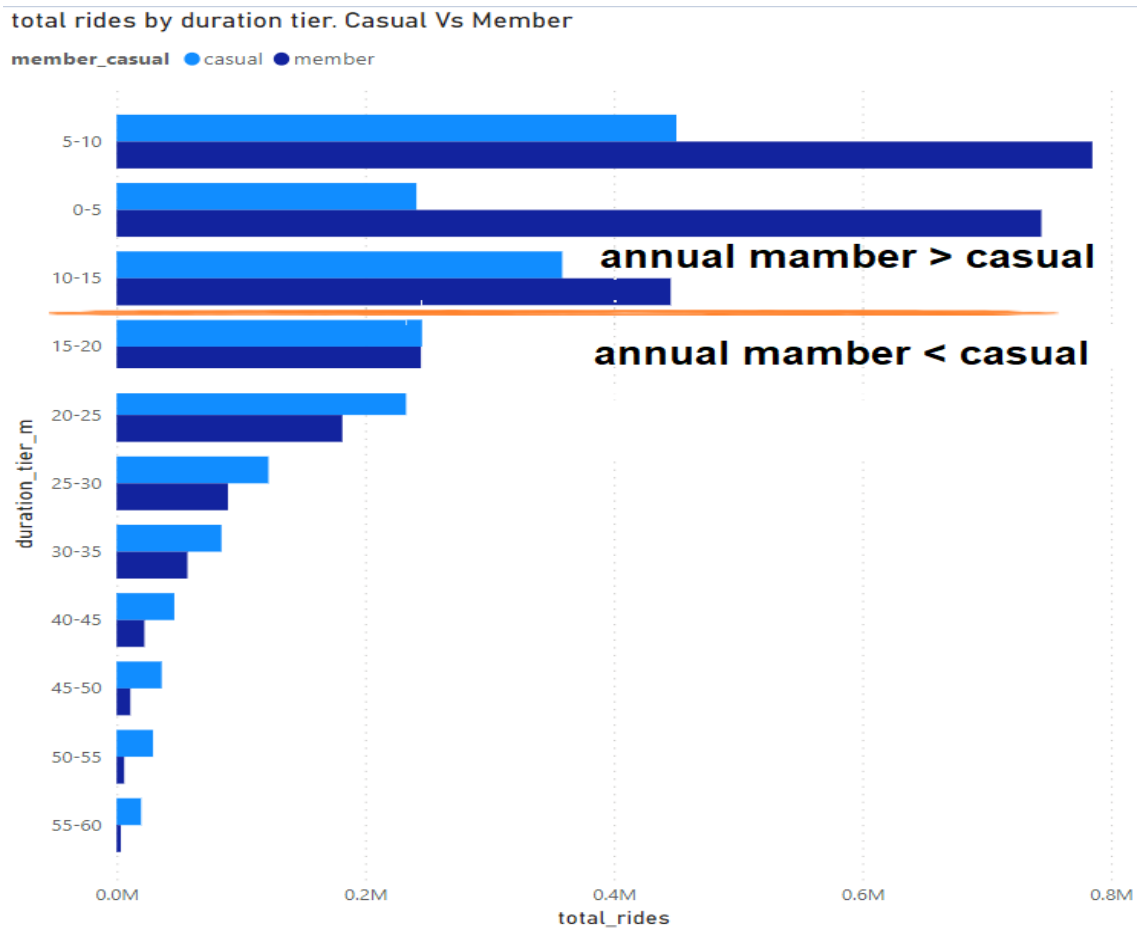
I want to analyze the behavior of Casual members Vs annual members

First, I look at plain simple stats such as total amount of rides and the durations of rides in minutes as well as average duration ride in minutes. And saw something interesting:

member_casual	Rides	%	avg duration ride in minutes
member	2,590,694	58.12%	11.50
casual	1,866,824	41.88%	17.30
Total	**4,457,518**	**100.00%**	**13.93**

Most of our rides are from our annual members with ~60% of total rides however our casual members ride for a longer duration of time (17 minutes). This immediately spiked my curiosity in knowing the purpose of their ride.

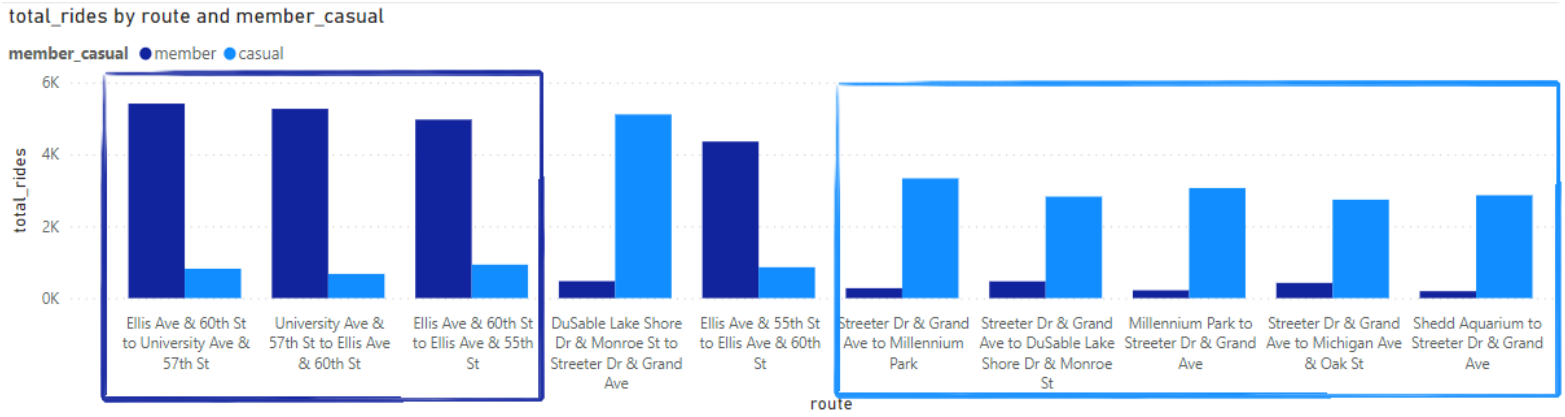
I then looked at the total rides in a tier format:



In the figure above we can see that rides up to 15 minutes are used more by our annual members and rides above 15 minutes are used more by our casual members.

Meaning our annual members are getting extra value from short distance / fast rides and our casual members require longer rides.

But to where? For what purpose? on what days? with what type of bike?



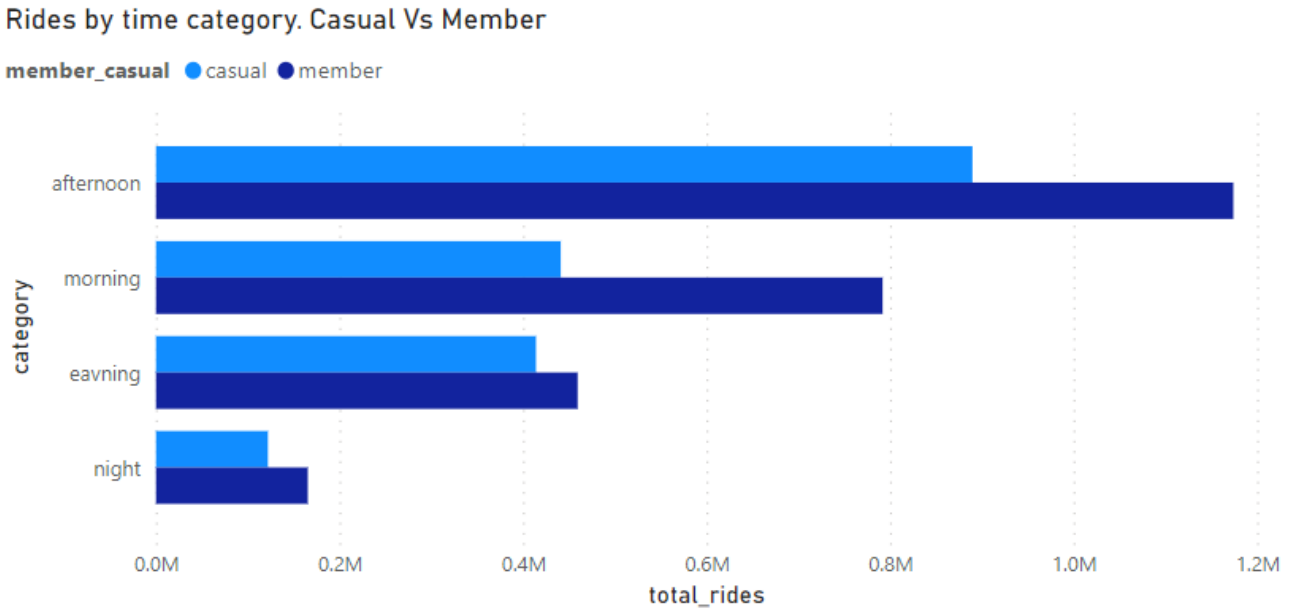
If we focus our attention on the left grouped columns, we can see where our annual members most popular routes are.

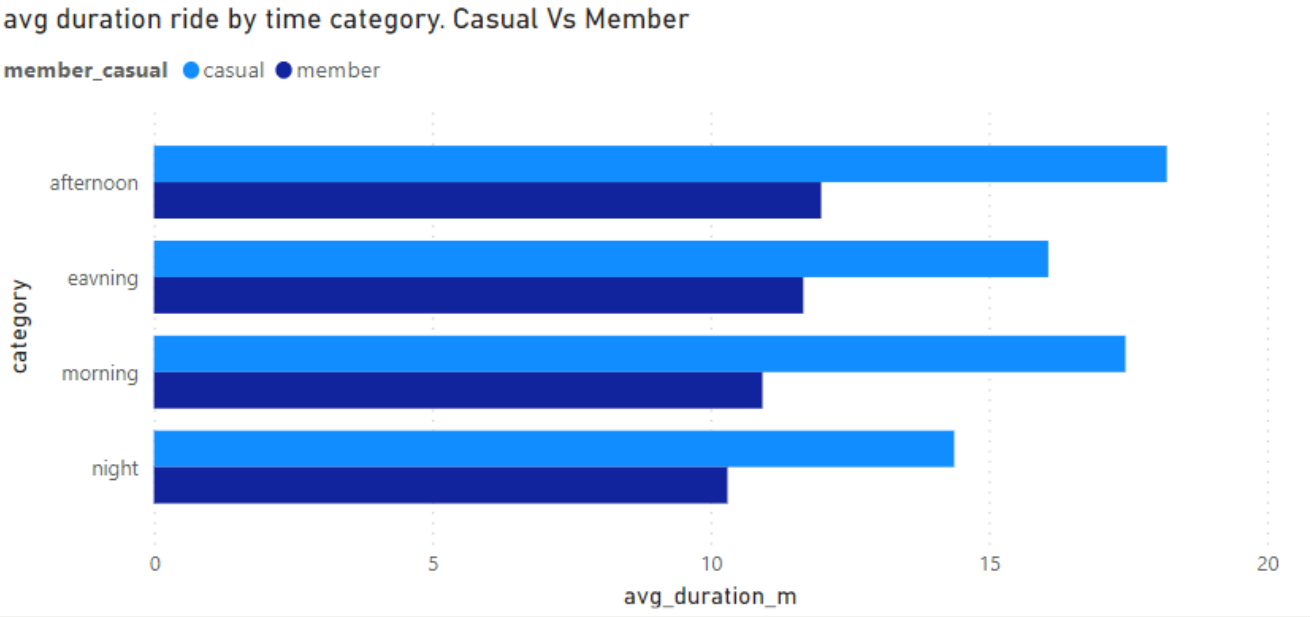
If we look at the right grouped columns, we can see the same for our casual members and I can identify a clear difference:

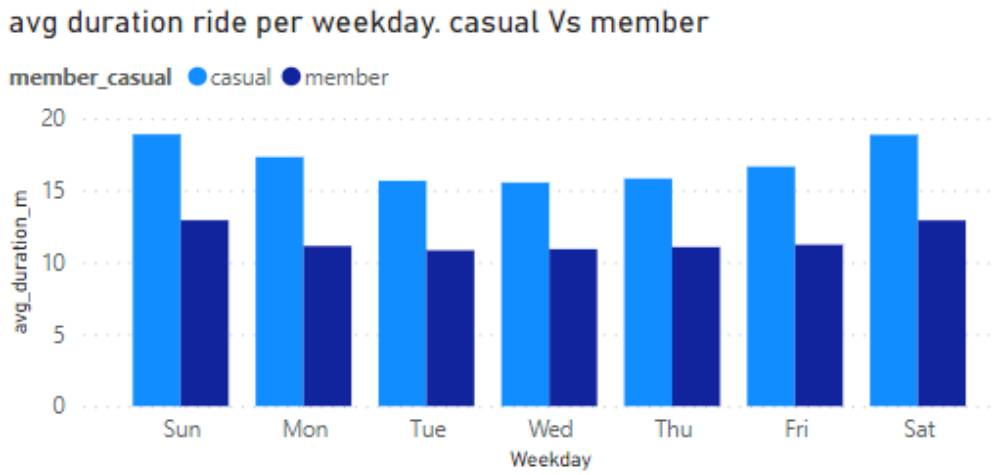
- Annual members use our services to go to places like work or university (with some outliers of our casual who do the same) having a fixed clear path they use every day
- Casual members use our services to go to recreational places such as parks, museums, and other touristy places and leisurely ride to explore the city. Making sense on why their avg duration is longer than our annual members.

This observation leads me to believe that our annual members are either student / Uni faculty or blue-collar workers just starting their carrier and our casual members are observers new to the tech and tourist.

Let's see if there's a difference in the date-time analysis:

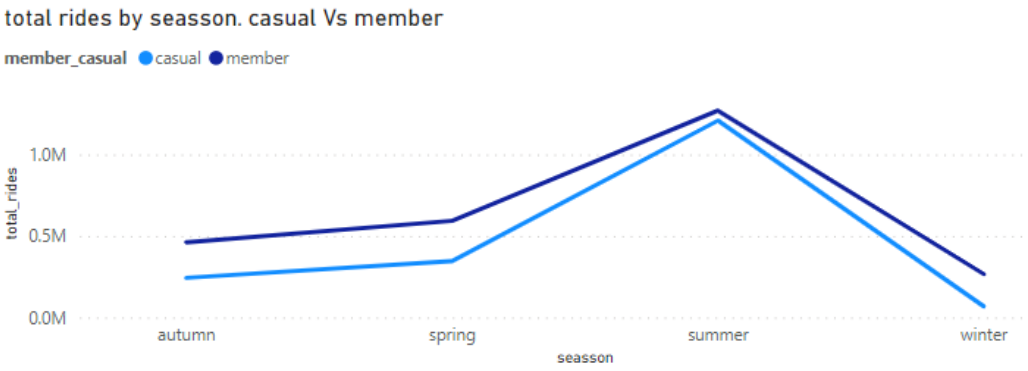

Rides by time category. Casual Vs Member
member_casual casual member
afternoon
morning
eavning
night
category
0.0M
0.2M
0.4M
0.6M
0.8M
1.0M
1.2M
total_rides


avg duration ride by time category. Casual Vs Member
member_casual casual member
afternoon
eavning
morning
night
category
0
5
10
15
20
avg_duration_m

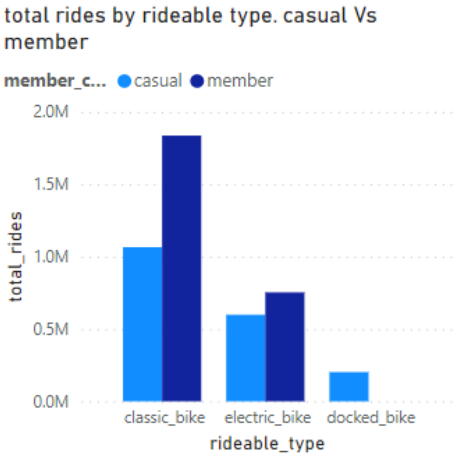


Looking at the average duration ride by weekday and by hour (three charts above), I can see that our annual members have a fixed route they use daily when our casual members vairy.

If we look at the combination of both weekday and time, we can see that our busiest time both types of members use our services is in the summertime weekends in the afternoon.



Overall, for both casual and annual members the prefered bike is the classic. It is worth noting that our annual members do not use the docked bikes at all.



3. INSIGHTS AND ACTIONS TO TAKE

As per my discoveries, I've seen that Cyclistic has two main types of members that use our services differently.

1. Annual members that are most likely students and/or new blue-collar workers trying to commute to work/Uni in a fast, accessible, and low-cost way.
2. Causal members use our services recreationally by traveling to popular touristy places. These members are most likely tourist, and our annual membership may be irrelevant to them.

- **There's an outlier of casual customers that aren't aware of our services but use them just like our annual members do, thus they can benefit by using said services.**
- **These are the prospectors I think we need to tackle to convert them into annual members. This is how I think we can do it:**

ACT

1. Collaborate with universities around university Ave and boost annual memberships for students as faculty. maybe make a deal with the Unis. The students will most likely keep using our services long after they finish their studies and faculty members will use our services as well.
2. Target users that ride under 15 minutes to raise awareness on the benefits on becoming an annual member.
3. Raise awareness during the weekends and during summer to boost the usage of our services during these "hot" times.

Thank you

Adam J Ben-Kiki

Data analyst