Introduction

I have recently completed the Google Data Analytics Professional Certificate offered on Coursera. The course divides data analysis process into six stages (ask, prepare, process, analyze, share, & act).

In the last course of the certificate, I have completed my Capstone Project on the Cyclistic Case Study. Where I applied the six stages of data analysis to gather insights from a large amount of data.

I cleaned and analyzed the data using R, and I used Tableau to create insightful visualizations.

Background

Cyclistic is a fictional bike sharing company which has a fleet of 5,824 bicycles that are geo-tracked and locked into a network of 692 stations across Chicago. The bikes can be unlocked from one station and returned to any other station in the system anytime. Cyclistic offers flexible pricing plans: single-ride passes, full-day passes, and annual memberships. Customers who purchase single-ride or full-day passes are referred to as casual riders. Customers who purchase annual memberships are Cyclistic members.

1. Ask

Cyclistic's finance analysts have concluded that annual members are much more profitable than casual riders. And therefore, the director of marketing has set a clear goal to: design marketing strategies aimed at converting casual riders into annual members.

My Role in this scenario is to analyze the Cyclistic historical bike trip data to identify trends and answer the business question of: **How do annual members and casual riders use Cyclistic bikes differently?**

2. Prepare

To answer this question, I will be analyzing historical Cyclistic bike trip data for 12 months (September,2021 to August,2022). The data is reliable, free of any bias, and has been collected by Cyclistic and stored on the company's database separated by month in CSV format. For the purpose of this project, I have saved the 12 relevant CSV files in my local drive.

The data collection team at Cyclistic have outlined some key facts and constraints about the data:

1. Each month contains every single trip that took place during that period.

2. All personal customer information has been removed for privacy issues. Which will be a limiting factor to the analysis, since it won't be possible to look at individual customer history.

3. Classic bikes must start and end at a docking station, whereas electric bikes have a bike lock attached to them; thus, electric bikes can also start and end their trip locked up anywhere in the general vicinity of a docking station.

4. The data should have no trips shorter than 1 minute or longer than 1 day. Any data that does not fit these constraints should be removed as it is a maintenance trip carried out by the Cyclistic team, or the bike has been stolen.

3. Process

To combine and clean the data I used **RStudio**.

Below is an outline of my process (refer to analysis in R for details):

- Loaded the 12 separate csv files
- Checked column names and inspected data frames, to make sure that everything in aligned before combining the files
- Combined the 12 csv files into one file, and removed the individual month files

Here is a quick summary of my findings:

- **ride_id:** Has no duplicates, and each ride_id is exactly 16 characters long; no data cleaning is necessary. No leading or trailing spaces in this column as well as the rest of the columns.
- **rideable_type:** The data contains 3 types of bikes: classic, docked, and electric bikes.
- **started_at/ended_at:** these columns show the date and time that the bike trips started/ended. Will be used to remove rides less than 1 minute or more than 1 day in length.
- **start_station_name/end_station_name:** Some trips have null values in either the starting station name column or the ending station name, but we will not remove those rows since they still hold some valuable information.

- **start_station_id/end_station_id:** there are various inconsistencies in string length and these columns do not add value to our analysis; thus, we will remove the columns for the final cleaned table.

- **start_lat/end_lat & start_lng/end_lng:** these 4 columns show the starting and ending location of the bike trips. Which is unnecessary in our analysis and will be removed.

- **member_casual:** this column indicates if the customer of the trip was a casual customer or an annual member. I double-checked that 'casual' and 'member' are the only available strings in this column.

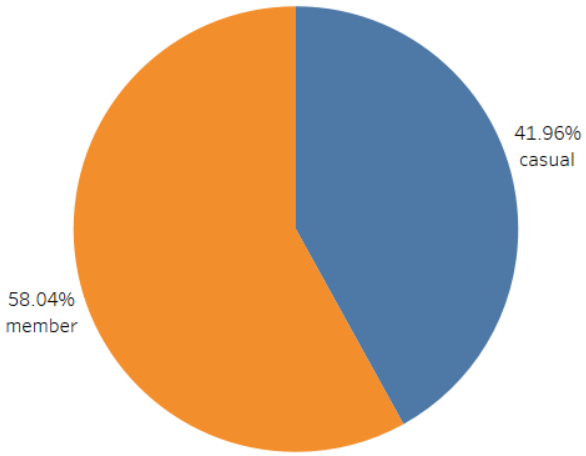
After carrying out the pre-cleaning data exploration process, I now know what data needs to be cleaned/removed and which columns can be created from the existing data to assist in our analysis. Below is a summary of the cleaning process

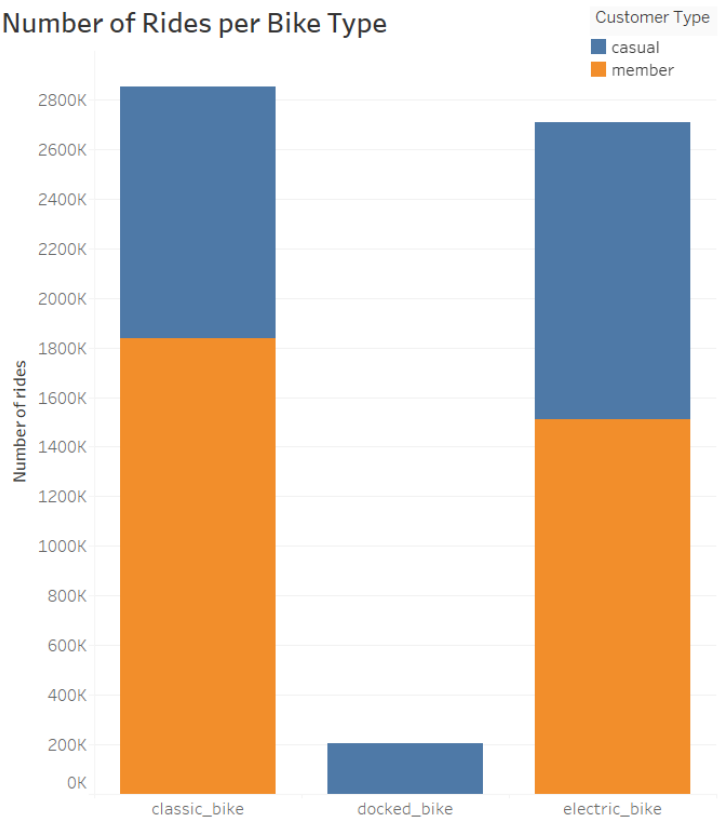
- Created a column for the ride duration
- Removed rides lasting less than one minute and rides lasting more than 1 day.
- Removed unnecessary columns.
- Created columns for: month-year, day of the week, time of the day for each ride start.

4. Analyze

Now that the data is completely clean it is time to analyze the data and answer the question: How do annual members and casual riders use Cyclistic bikes differently?
To analyze the data, I used **Tableau** to create insightful data visualizations.

a. First, I examined the customer type distribution and the bike type preferences between annual members and casual riders.



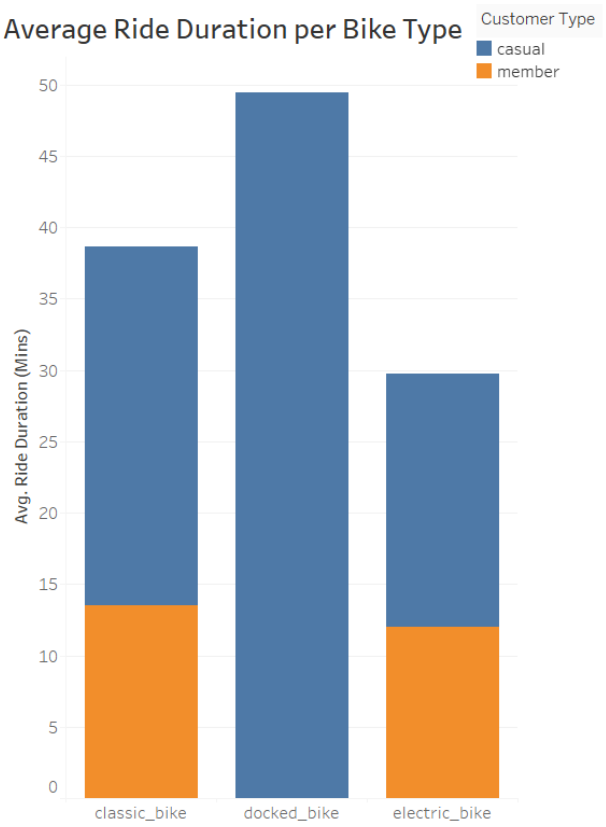


41.96% of the service users are casual riders while **58.04** are members. Which indicates that a big proportion of the service users can be targeted by the marketing campaign.

The bike preference shows the first difference between casual riders and members:

- For casual riders: **42%** of the rides used classic bikes, 8% used docked bikes and **50%** used electrical bikes.
- For members: **55%** of the rides used classic bikes while **45%** used electrical bikes. Note that members never use the docked bike type.
- Another way to look at it would be: for classic bikes **36%** of the rides were by casual riders, while **64%** of the rides were by members. And for electrical bikes **44%** of the rides were by casual riders, while **56%** of the rides were by members.

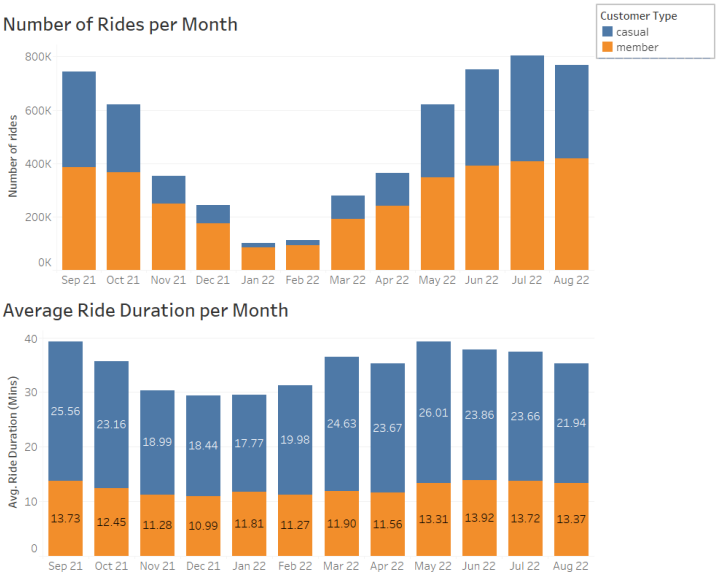
b. Next, I examined the average ride duration per customer type and bike type.



The ride duration shows the second difference between casual riders and users:

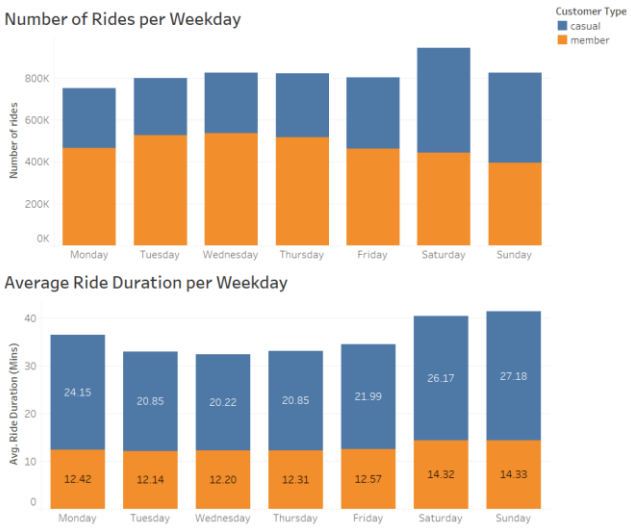
- Average ride duration is **23.5 minutes** for casual riders, and **12.8 minutes** for members
- Looking at the average ride duration per bike type we find the same trend of casual riders having longer ride duration than members.
- For casual riders, the average ride duration: on a classic bike is **25 minutes**, on a docked bike is **49.5 minutes** and on an electrical bike is **17.8 minutes**.
- For members, the average ride duration: on a classic bike is **13.5 minutes** and on an electrical bike is **12 minutes**.

c. Next, I examined the number of rides and the average ride duration per month, day of the week and time of the day for the two types of users.



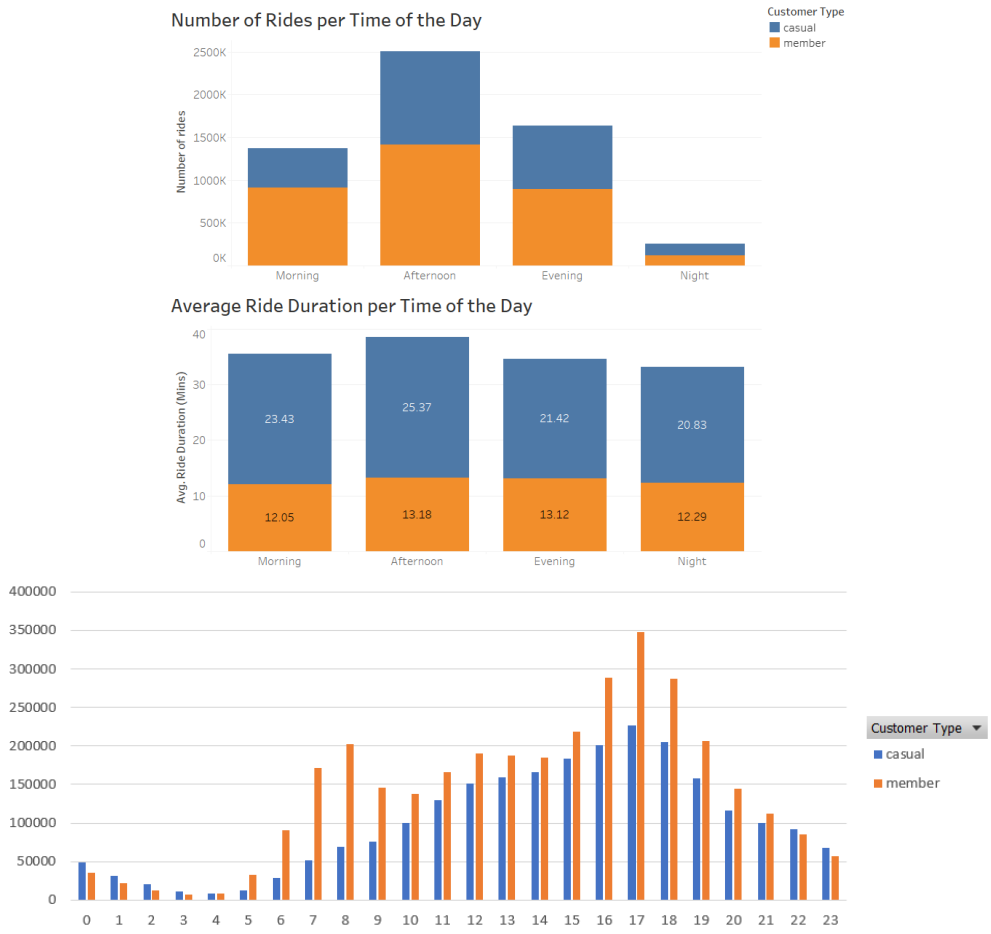
Both customer types show the same behavior of reduced number of rides in the colder months of the year (November to April), however for casual riders the reduction is far more severe as only 18% of the total yearly rides happens in the 6 colder months. In contrast, a considerable proportion of the annual members stay consistent in using the service even on those months, with 31% of the total yearly rides coming from the rides during the colder months.

Ride duration is also below average during those months, with the exception of March and April for casual riders as ride duration is back to normal.



- For casual riders: the number of rides and the ride duration is higher during the weekend.

- For members: the number of rides is lower during the weekend but the ride duration is higher.
- This is the third difference so far, and leads to the assumption that many members use the bike to commute to work.

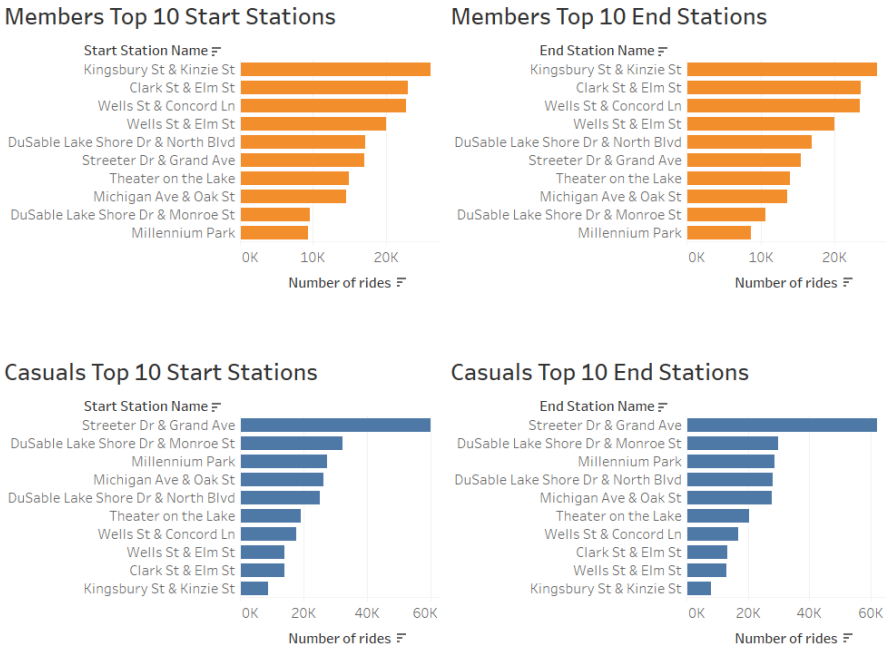


Both customer types show the same behavior of taking more rides in the afternoon (12pm to 5pm) with slightly higher than average ride duration. And also, both customer types had very small number of rides in the night (12am to 5am)

Annual members number of rides level rose significantly between 6am and 8am and peak between 4pm and 6pm, before dropping considerably after 7pm, which confirms the assumption that many members use the bike to commute to work.

Casual riders' number of rides started rising gradually from 7am and peaked at 5pm and then started dropping.

d. Next, I examined the top ten start and end stations for the two types of users



- Members have the same ten station as the top starting and ending station with the same order, which indicates that members tend to have fixed routes. This goes in line with our previous assumption that many members use the bike to commute.
- Casual riders also have the same ten stations as the top starting and ending station but with minor differences in the order.
- As we can see above, the top 5 stations for casual riders are generally near the water and big parks; in contrast, annual members usually start their trips more inland, near Chicago's financial district, office buildings, and apartment buildings.
- Further, casual riders' top 5 most visited stations are the same for both starting and ending stations. These include Streeter Dr & Grand Ave (the pier), Millennium Park, Theater on the Lake, and Shedd Aquarium. All very big tourist attractions and recreational areas. None of these stations are in the top 5 most visited stations for annual members with only one exception with the pier station which comes as the 5th most popular station for members, indicating that recreation and sightseeing are not as big of a priority.,

5. Share

Summary of Insights

Casual Riders:

- Tend to use the Cyclistic bikes for leisure; preferring to ride during the midday hours on the weekend, concentrated in the late spring and early summer months.
- Their average ride time (23.5 minutes) is around 2x longer than annual members.
- Tend to start and end their trips near the water, city parks, and other tourist/recreational attractions.

Annual Members:

- Tend to use the Cyclistic bikes for commuting; preferring to ride all-year-round on the weekdays, with a drop off in the winter months, and during working hours (9am/3pm).
- Their average ride time (12.8 minutes) is around half the average trip time of casual riders.
- Tend to start and end their trips in the city, near office and apartment buildings.

6. Act/Recommendations

Having answered the question of how do annual members and casual riders use Cyclistic bikes differently, the marketing team can start to develop their strategy for converting casual riders to annual members. One limiting factor to consider is that, a large proportion of the casual riders' segment are most likely tourists visiting. However, for converting Chicago residents who are casual members into annual members, I recommend that the advertising campaigns should be carried out during the times with peak service usage (during the weekend on the warmer months), and near the most popular stations for casual riders.