**Intro**

This is a capstone analysis case study of a fictional company, Cyclistic, for the Google Data Analytics Professional Certificate.

The director of marketing believes the company’s future success depends on maximizing the number of annual memberships. To do this, the goal for the marketing analyst team is to design new marketing strategies to convert casual riders into members.

**Ask**

The marketing director has assigned me with the following analysis question to answer to help my team fulfill the objective mentioned above:

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

**Prepare**

The dataset was obtained from a link of the company's historical public data. The data has been made available by Motivate International Inc. under this [license](https://ride.divvybikes.com/data-license-agreement). Data-privacy issues have prohibited riders’ personally identifiable information from being included.

The dataset contained an index of several different zip files that ranged from 2013 to 2023. To identify the trends of my business task I will analyze the last 12 months of Cyclistic bike trip data, so March of 2022 to February of 2023 was downloaded and stored in a folder on a desktop.

**Process**

The zip files were too large to be used with spreadsheets so I used BigQuery SQL to process the data. The smallest file contained 190,301 rows and the largest file contained 823,488 rows.

The following are the steps I took to see if the data was clean and extract some data to analyze further in spreadsheets.

1. I checked to see if there were any misspellings in the member\_casual and rideable\_type columns. I also made sure there were not any duplicate trips recorded.

SELECT DISTINCT member\_casual

FROM `bike-case-study41223.bike\_data.bikes\_322`

SELECT DISTINCT ride\_id

FROM `bike-case-study41223.bike\_data.bikes\_322`

1. I made sure there were no null values in the member\_casual, rideable\_type, ride\_id, started\_at, and ended\_at columns.

SELECT member\_casual

FROM `bike-case-study41223.bike\_data.bikes\_322`

WHERE member\_casual is null

After doing some cleaning queries, I thought of some questions I wanted to answer to help me determine what kind of information I needed to analyze to complete my business task.

1. How many trips did members & casual riders make?

SELECT member\_casual,

COUNT(member\_casual) AS num\_of\_trips,

FROM `bike-case-study41223.bike\_data.bikes\_322`

GROUP BY member\_casual

\*Extracted results into March 2022 spreadsheet

1. Which bike type is most popular with each rider?

SELECT member\_casual, rideable\_type,

COUNT(rideable\_type) AS num\_of\_bikes\_used,

FROM `bike-case-study41223.bike\_data.bikes\_322`

GROUP BY member\_casual, rideable\_type

ORDER BY member\_casual

\*Extracted results into a new tab of March 2022 spreadsheet

1. Where do riders get their bikes from the most? Top 10 most visited locations?

There were visible null values in the start\_station\_name column but there was still sufficient information recorded in the other rows to be used for analysis.

SELECT member\_casual, start\_station\_name,

COUNT(start\_station\_name) AS trips\_from\_station

FROM `bike-case-study41223.bike\_data.bikes\_722`

WHERE start\_station\_name is not null AND member\_casual = "casual"

GROUP BY start\_station\_name, member\_casual

ORDER BY COUNT(start\_station\_name) DESC

LIMIT 10

\*Extracted results into a new tab of March 2022 spreadsheet

1. How long do the riders keep the bike?

SELECT member\_casual,

SUM(travel\_mins) AS total\_travel\_mins,

AVG(travel\_mins) AS avg\_travel\_mins

FROM `bike-case-study41223.bike\_data.With\_travel\_min\_322`

GROUP BY member\_casual

\*Extracted results into a new tab of March 2022 spreadsheet

1. How many rides were taken on each day of the week?

SELECT rideable\_type, started\_at, ended\_at, member\_casual,

EXTRACT(DAYOFWEEK from started\_at) AS day\_of\_week\_num

FROM `bike-case-study41223.bike\_data.bikes\_322`

\*Made a new table of the results to work with for the next step

SELECT \*,

CASE WHEN day\_of\_week\_num=1 THEN 'Sunday'

WHEN day\_of\_week\_num=2 THEN 'Monday'

WHEN day\_of\_week\_num=3 THEN 'Tuesday'

WHEN day\_of\_week\_num=4 THEN 'Wednesday'

WHEN day\_of\_week\_num=5 THEN 'Thursday'

WHEN day\_of\_week\_num=6 THEN 'Friday'

WHEN day\_of\_week\_num=7 THEN 'Saturday'

END AS day\_of\_the\_week

FROM `bike-case-study41223.bike\_data.day\_of\_week\_322`

\*Made a new table of the results to work with for the next step

SELECT member\_casual, day\_of\_the\_week,

COUNT(day\_of\_the\_week) AS bike\_trips

FROM `bike-case-study41223.bike\_data.Day\_of\_week\_322`

GROUP BY member\_casual, day\_of\_the\_week

ORDER BY member\_casual

\*Extracted results into a new tab of March 2022 spreadsheet

Once I extracted all of those data subsets from SQL, I repeated the same steps for the other 11 months. Then, I went through and formatted each spreadsheet. I added a month column, bold and capitalized titles, capitalized values, changed title names, formatted the numbers properly, and spell checked each sheet.

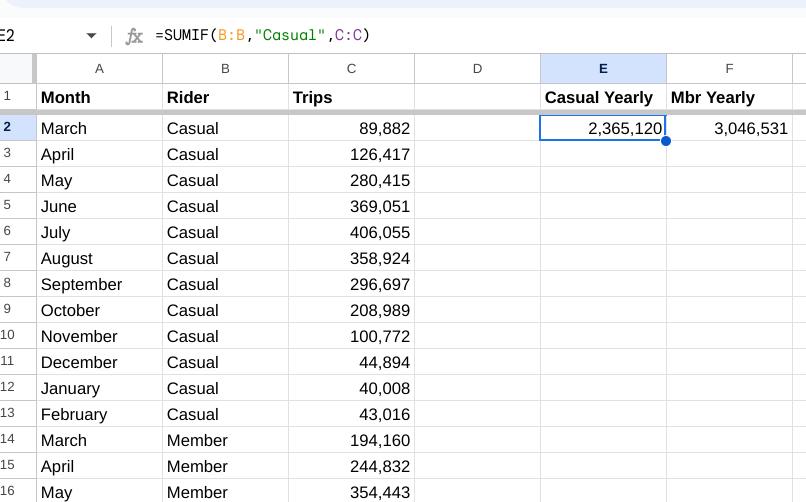
**Analyze**

I aggregated all of the information from all 12 spreadsheets into a year-review spreadsheet, formatted in the same way as the individual months spreadsheets, having a separate tab for each of the questions above.

* **Trips tab:**
  + sorted sheet by rider
  + manually arranged the rows from march to february
  + calculated the number of trips taken by members and casual riders for the year

=SUMIF(B:B,"Casual",C:C)

=SUMIF(B:B,"Member",C:C)

****

* **Minutes tab:**
  + sorted sheet by rider
  + sort range of average minutes traveled in descending order of each rider,
  + calculate total hours traveled
  + calculated the sum of the total hours traveled for each rider
  + calculated the average trip time of the year for each rider

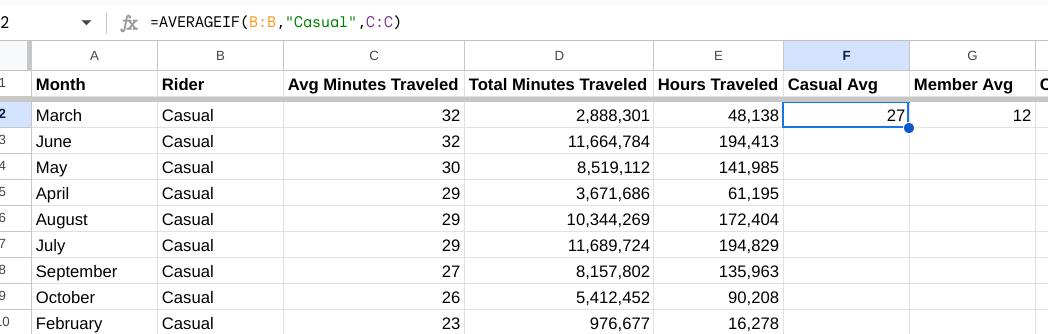
=D2/60

=SUMIF(B2:B25,"Casual",E2:E25)

=SUMIF(B:B,"Member",E:E)

=AVERAGEIF(B:B,"Casual",C:C)

=AVERAGEIF(B:B,"Member",C:C)

****

* **Bike type tab:**
  + sorted sheet by rider
  + calculated the sum of docked bikes used
  + calculated how many bikes of each type were used by each rider

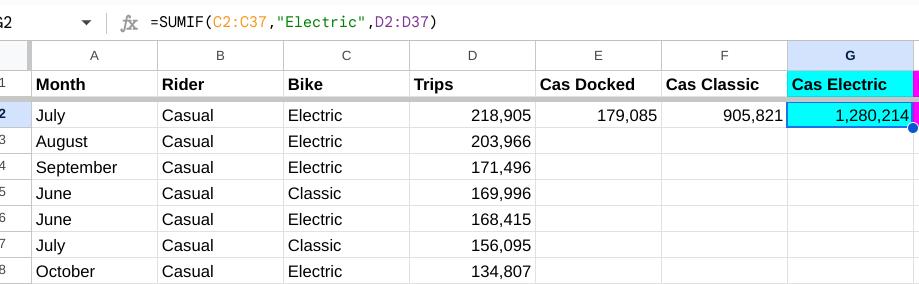
=SUMIF(C:C,"Docked",D:D)

=SUMIF(C2:C37,"Classic",D2:D37)

=SUMIF(C2:C37,"Electric",D2:D37)

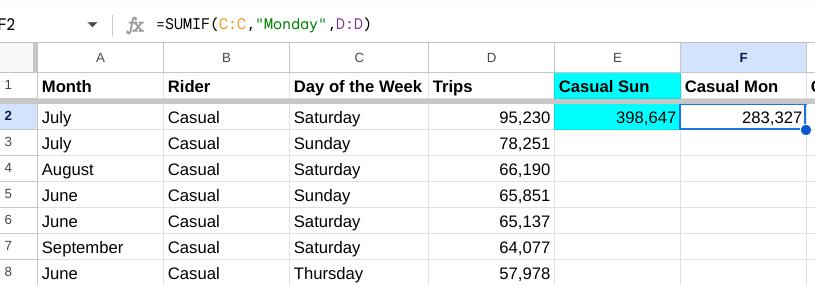
=SUMIF(C38:C61,"Classic",D38:D61)

=SUMIF(C38:C61,"Electric",D38:D61)

****

* **Casual days of the week tab & Member days of the week tab:**
  + calculated the sum of trips taken on each day of the week for each rider

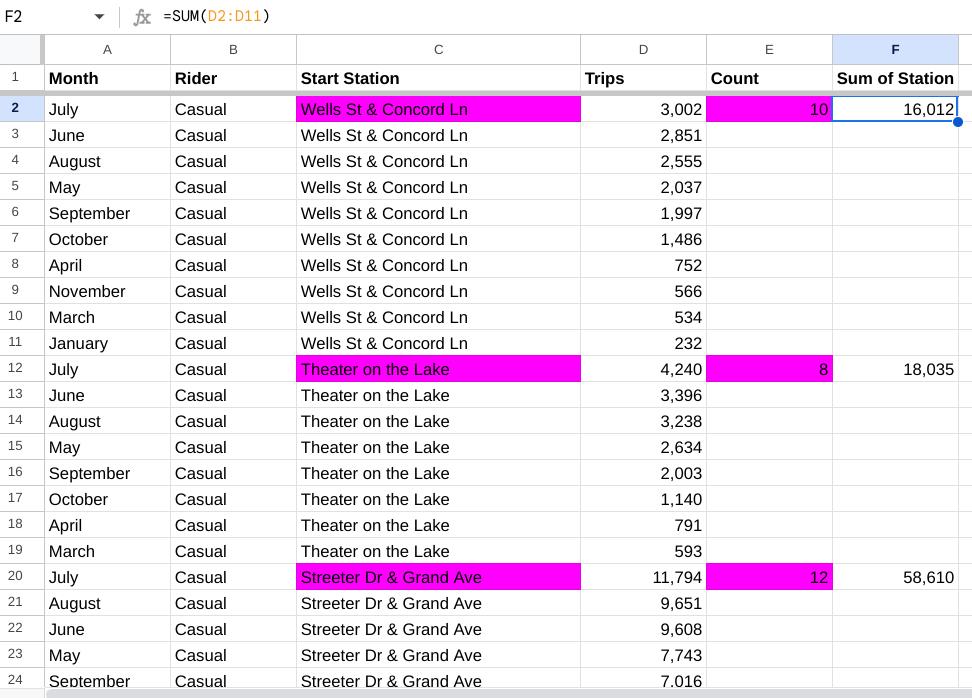
=SUMIF(C:C,"Sunday",D:D)

****

* **Casual locations tab & Member locations tab:**
  + sorted sheet by start station
  + count function of each station name
  + applied conditional formatting to the count function cells for greater than 7, then deleted all of the stations whose count was not highlighted
  + calculated sum of trips taken of each station left

=COUNTIFS(C:C,C2)

=SUM(D2:D11)



**Analysis Findings:**

Casual Riders

1. Use the bikes the most in the summer months (July, June, and August)
2. Use the bikes the primarily on the weekends (Saturday and Sunday and Fridays)
3. Most often use the electric bikes more than the classic bikes & only 7.6% of their total trips were taken with the docked bikes
4. Keep the bike for a longer period of time than members
5. Casual riders visit these stations the most to get the bikes
   1. Streeter Dr & Grand Ave
   2. DuSable Lake Shore Dr & Monroe St
   3. Millennium Park
   4. Michigan Ave & Oak St
   5. DuSable Lake Shore Dr & North Blvd

Members

1. Take more trips than casual riders every month
2. Primarily in the months ending summer leading into fall (August, July, and September)
3. Use the bikes primarily during the week (Thursday and Tuesdays and Wednesdays)
4. Most often use the classic bikes and they have not used the docked bikes at all in the last 12 months
5. Members visit these stations the most to get the bikes
   1. Kingsbury St & Kinzie St
   2. Clark St & Elm St
   3. Wells St & Concord Ln
   4. Ellis Ave & 60th St
   5. University Ave & 57th St

**Share**

Analysis visualizations were made using Tableau

**Act**

My top three recommendations based on my analysis would be:

1. Create a summer promotion of discounted bikes for new members
2. Discount the docked and electric bikes for members to get more casual riders to become members (and for more members to start using these bikes more)
3. Advertise all new promotions aimed at converting casual riders into members heavily in the areas where casual riders visit certain bike stations the most