**Cyclistic bike-share analysis case study:**

**Overview:**

Cyclistic is a bike-share company in Chicago, USA. The company has mainly 3 flexible pricing plans: single ride passes, full-day passes and annual memberships. Customer who purchases single-ride or full-day passes are referred to as casual riders whereas, customers who purchase annual memberships are Cyclistic members.

**Business Task:**

The company wants to maximize the number of annual memberships. So, as a data analyst our task was to understand how casual riders and annual members use Cyclistic bikes differently, which will help the Cyclistic Marketing Analytics team in designing new marketing strategies to convert casual riders into annual members.

This task involves:

* Gathering data from reliable sources.
* Preparing, cleaning and transformation of data.
* With the help of analyzing tools like Spreadsheet or SQL or programming language viz. R etc.
* Recommendations based on analysis.

**Description of all data sources used:**

We used Cyclistic’s historical data to analyze and identify trends. The data has been made available by Motivate International Inc. under this license (https://www.divvybikes.com/data-license-agreement). We used data from January 2020- December 2020 (12 months data) under this license. It contains public data about the different customer types using Cyclistic bikes.

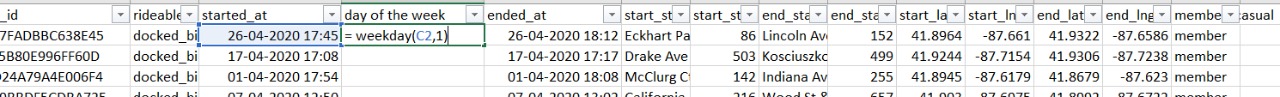
The data is organized in the CSV format. Data-privacy issues prohibit us from using rider’s personally identifiable information.

**1.Process**

After prepare phase next comes process stage, in which raw data was cleaned for missing and inconsistent values and then manipulated to gain more insights.

**1.1 Data Manipulation:**

* Day of the week column was created with the help of start date column.
* Trip duration column was formulated by subtracting end and start time column.



**1.2 Data cleaning:**

In data cleaning we had to make sure that data remains consistent throughout and all missing values were investigated properly.

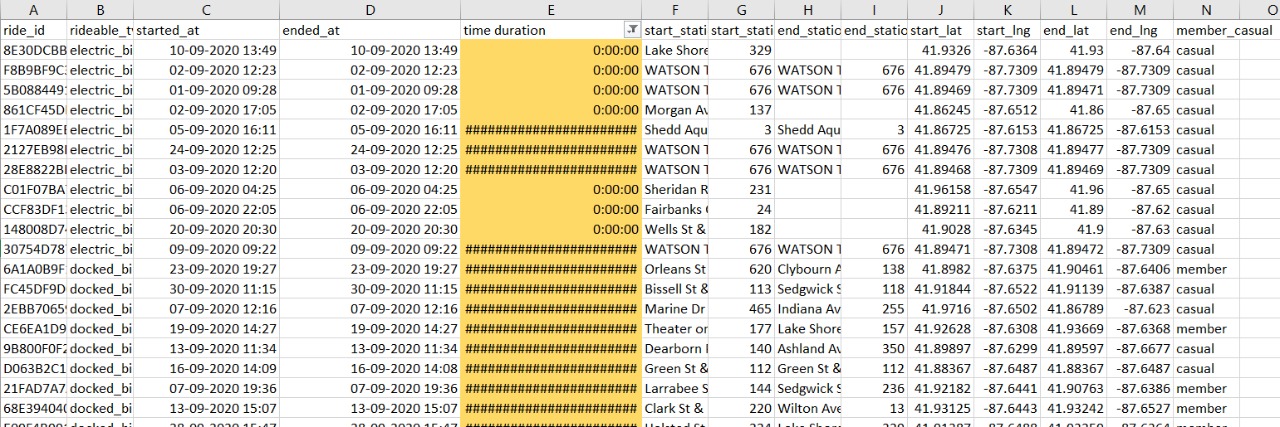
* Upon observing Ride-Id column, there were inconsistency with some rows containing improper id with different format, hence those rows were dropped.



* In most of monthly data, Station name column contained many missing values, as a result we had to remove those values due to lack of information.



* Latitude and longitude column values data format were inconsistent hence needed some fixing.
* In trip duration columns we dropped ######(negative values) and 0 values.

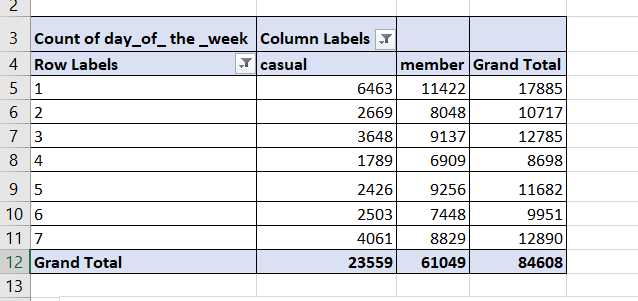


**2.ANALYZE:**

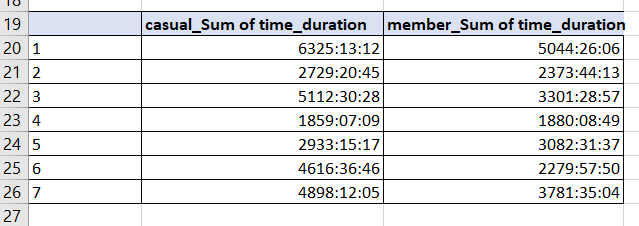
Next stage involves analyzing the cleaned data which was obtained after the process phase.

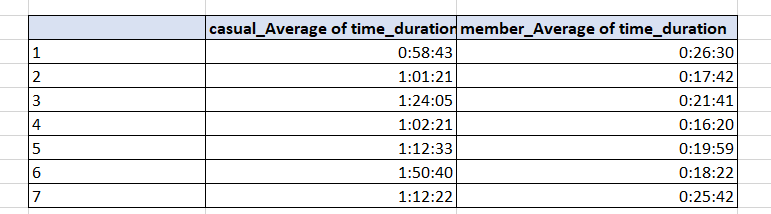
**2.1 Summary of analysis:**

For this step we made separate sheet for each month with pivot table containing the weekday column in the row and membership column in filter of the respective tables.

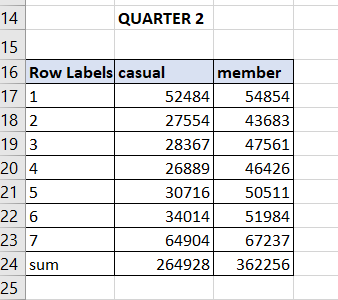


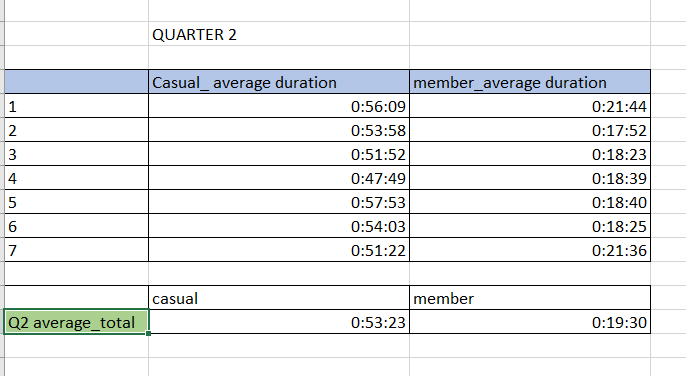
Another pivot containing trip duration in filter section (with average and max values) and weekday in row section was also made.





After analyzing each month, we grouped the months into 4 quarters to get total trips and average for each quarter which provided more information on how data changes in different quarters.





**3. Share:**

**3.1 Supporting visualizations and key findings for monthly data:-**

* Bar chart between total trips (casual and member) and day of the week for each month was drawn.

**3.1.1 Key findings and inference:**

There is a significant difference between the number of trips for members in weekdays and weekends as compared to casuals’ across all the months and all the quarters.

* Line chart between sum of trip duration and day of the week for both casual and member bikers were made. The same thing for average of trip duration and day of the week.

**3.1.2 Key findings:**

For the sum line chart, no noticeable inference could be made, while for the average graph ,we can see that the average time duration for casuals bikers were way more than member bikers across all seven days and also the average time duration for members remains constant throughout whole week while for casual there were concentration in weekends as compared to weekdays.

* Line chart between different rideable type for both casuals and members were made for average time duration for the last six months data of 2020.

**3.1.2 Key findings:**

While for members bikers, the average for all the types were almost same, but with casuals the case was different with average time duration for docked type way more than any other type.

**3.2 Supporting visualizations and key findings for quarterly data:-**

* Now for each quarter also the above-mentioned bar chart and line chart were made and analyzed.

**3.2.1 Key findings:**

Results obtained were same as that from the months.

**3.3 Supporting visualizations and key findings for yearly data:-**

* For a yearly analysis a line chart showing the average trip duration for each quarter was plotted.

**3.3.1 Key findings:**

For casual riders the average trip duration kept going downhill throughout the year , while it remains consistent for members.

**4. Key insights on the basis of analysis:**

Based on our findings through visualizing the data, the key insights are as follows:

* The average trip duration time for casual riders is always more than the member riders across all 12 months.
* Despite the introduction of electric bikes apart from existing docked bikes in the latter half of 2020, the average trip duration for docked bikes always remained higher than electric bikes for casual riders. For member riders, the average trip duration does not vary on the bike category as it is almost same for all bike types.
* The average trip duration for casual riders recedes from Q1 to Q4 of 2020. For member riders, it is stable throughout all four quarters.