**Cyclist bike-share Analysis**

**Problem**

**Cyclistic**, a bike-share company in Chicago that features more than 5,800 bicycles and 600 docking stations have two different riders which include the **Casual riders** and **Member riders.** The company believes that its future success depends on maximizing the number of annual memberships. Therefore, wants to understand how casual riders and annual members use Cyclistic bikes differently and know how to design a new marketing strategy to convert casual riders into annual members from the insights generated from analyzing the Cyclistic historical bike trip data and also identify trends.

**Data Source**

The Cyclistic’s historical trip data I’ll be using to analyze and identify trends in this project is a public dataset

Download the previous 12 months of Cyclistic trip data here. [[Index of bucket "divvy-tripdata"](https://divvy-tripdata.s3.amazonaws.com/index.html)] (Note: The datasets have a different name because Cyclistic is a fictional company).

I’ll be making use of two datasets in this project which include the Divvy\_trips\_2019\_Q1 and the Divvy\_trips\_2020\_Q1 dataset

**Data transformation and cleaning**

* Loaded Divvy\_trips\_2019\_Q1 and Divvy\_trips\_2020\_Q1 into the Excel power query editor through the Get data option.
* Removed trip\_duration, gender and birth \_year column from Divvy\_trips\_2019\_Q1.
* I noticed the column names of the 2019\_Q1 dataset aren’t consistent with the 2020\_Q1 dataset
* Changed some columns name of the 2019\_Q1 dataset to be consistent with the 2020\_Q1 dataset: trip\_id to ride, bike\_id to rideable\_type, start\_time to started\_at, end\_time to ended\_at, from\_station\_id to start\_station\_id, from\_station\_name to start\_station\_name, to\_station\_id to end\_station\_id, to\_station\_name to end\_station\_name, user\_type to member\_casual.
* Replaced ‘Subscriber’ under the member\_casual column with ‘member’ and also replaced ‘Customer’ with ‘casual’ to make the column be in line with the 2020 dataset as the organization change the name of its membership type over the year.
* Removed start\_lat, start\_lng, end\_lat, end\_lng columns from Divvy\_trips\_2020\_Q1 dataset as they aren’t present in the Divvy\_trips\_2019 dataset and won’t be needed in the analysis.
* Now that the two datasets are consistent, I can easily combine the two datasets into one table using the append option under the combine option in the power query home.
* Added 2 new column [“ride\_length” and “day\_of\_week”] to the newly combined table with Excel formula: ride\_length = ended\_at – started\_at and day\_of\_week=weekday(started\_at)
* I was able to find all the empty cells with the help of the Find and Select option on the home tab and further delete the rows that contains those empty cells with the Delete option.
* I noticed some cells of the ride\_length columns are showing negative sign which wasn’t meant to be maybe this was due to some downtime of the bike which they needed to be taken for repair or maintainaince which doesn’t necessarily add up to the ride length of the customers.
  + -So I filtered the ride\_length column to show only rows from 00:00:00 minute and above.
* I look for the average ride\_length and the maximum ride\_length using excel “AVERAGE” and “MAX” function respectively, I also get to know the most busy day\_of\_week with excel “MODE” function.

**Descriptive Analysis and visualization with Pivot table and charts**

* Now that the data is clean and consistent, with the columns needed for further analysis created, I then proceed to doing some descriptive analysis by grouping and aggregating the dataset with a pivot table and further visualizing the data with a pivot chart.
* I Calculated the average ride\_length for members and casual riders with rows = member\_casual and Values = Average of ride\_length in the pivot table field and then visualize the result using a stacked column chart in the pivot chart option.

* Calculated the average ride\_length per day for each user with Rows = Day\_of\_week, Column = Member\_casual and Values = average ride\_length in the pivot table field. Then I visualize the data with a line chart to easily make visible the trend of average ride length for each day of the week using the pivot chart.
* I calculated the users rides per day with a filter to help know how each users make use of the bike during the course of the week with Rows = Day\_of\_week, Values = count of ride\_id and filter = Member\_casual in the pivot table field, I went further by filtering the users rides per day by members only and I visualize the data using a line chart, I also filtered the data by casuals only and also visualize it with a line chart to show the trend of how users make use of bikes during the weekdays
* I further Calculated the Users total rides with Rows = Member\_casual and Values = count of ride\_id in the pivot table field and then used a bar chart for visualization in the pivot chart.
* Finally, get to know the total number of days users make use of bikes with Rows = daw\_of\_week and Values = count of ride\_id in the pivot table field, then I visualize the data using a bar chart also in the pivot chart.

**Data Insights**

* The average ride\_length by users over the year stands at 00:19:44 hours.
* The maximum ride\_length by users is 01:20:22 hours.
* Tuesday is the day of the week which has the most riders with 7,985 casual riders and 127,974 member riders giving a total of 135,959 riders.
* Looking at the average ride length of the two users:
* Casual users make use of bikes for a longer time than the Member users.
* While comparing the average ride length for the two users per day of the week, I noticed:
* Casual riders also make use of the bikes for longer hours each day of the week compared to Member riders.

It’s evident that Casual riders do ride the bike for a longer hour than the member riders, I need to know how often casual and member riders make use of Cyclistic bikes to further understand the reason for casual riders longer hours of ride.

* After aggregating the trip\_id to each Users to know how often users make use of of Cyclistic bikes, I noticed:
* Member riders do ride the bikes more often than Casual ridersMember riders have the highest number of rides with a greater margin than the Casual riders
* Aggregating trip\_id to days of the week to know the total number of days users spend on trips:
* Member riders use more days riding Cyclistic bikes than the Casual riders.

There seems to be a contrast, Member riders make use of Cyclistic bikes more often and they also use more days riding the bikes than casual users but the member riders still have an average ride length far less than the casual riders and also the Casual riders have a longer riding hour each day despite Member riders making use of the bikes more each days.

There is a need to understand the trend of how users make use of Cyclistic bikes for each day of the week to solve the current puzzle.

* By comparing and studying the trend of how users make use of Cyclistic bikes, I came to the understanding that:

**N.B:** 1-7 represents the Days of the week, with 1 representing Sunday and 7 representing Saturday

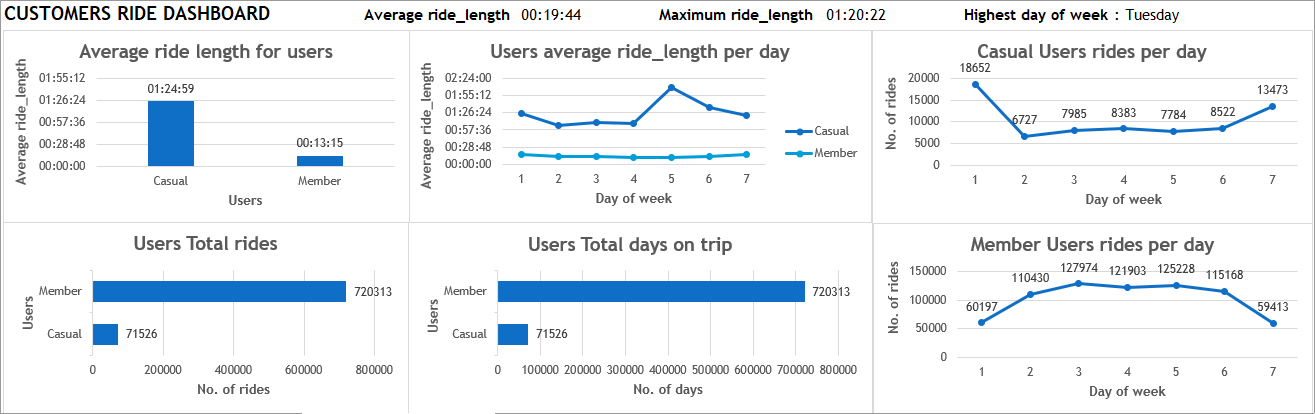
* Member riders make use of bikes more on working days than any other day.
* It’s evident that they ride bikes to commute to work mostly than for pleasure.
* They use shorter time commuting to work which explains the reason for their smaller average ride length.
* They make use of the bikes more often because the go to work majority of the days.
* Casual riders make use of bikes more during the weekends than the working days.
* Casual riders are more of pleasure riders than member riders.
* They spend more time riding for pleasure and touring the city which explains their higher average ride length.

**Summary**

Member riders make use of bikes more often and for most days of the week than Casual riders but have a smaller average ride length compared to Casual riders who make use of bikes mostly on weekends and less often during the working days.

Member riders make use of bikes mostly to commute to work which only takes a shorter ride length compared to Casual riders who make use of bikes mostly for pleasure with a longer average ride length.

Casual riders make use of bikes less often but with longer ride length while Member riders make use of bikes more often but with shorter ride length.

**Users Dashboard**

**Recommendations**

* Cyclistic should create targeted marketing campaigns that highlight the benefits of annual membership. For example, emphasize cost savings, convenience, and exclusive perks for members through the use of digital media channels with the contact details provided by the casual users.
* The company should offer discounted monthly or seasonal memberships specifically aimed at casual riders and Promote these discounts during seasons when ridership is lower (e.g., winter months. ). By experiencing the benefits of membership during these periods, casual riders may be more inclined to upgrade to a full-year membership.
* How annual memberships benefit daily commuters should be highlighted. Since member riders use bikes more often, the convenience of having a membership for flexibility and daily work commutes should be emphasized. Annual memberships should be positioned as a cost-effective solution for regular commuting for casual users who need reliable and flexible transportation.

**THANK YOU**