**Google\_Data\_Analytics\_Capstone\_Case\_study**

**# A Case Study On Usage Of Bike By Subscriber And Non Subscriber:**

**Project Overview:**

Cyclistic Ltd. is bike-sharing company which has grown to a fleet of 5,824 bicycles into a network of 692 stations across Chicago.

**Ask**

**Key Questions:**

1. How do annual members and casual riders use Cyclistic bikes differently?
2. Why would casual riders buy Cyclistic annual memberships?
3. How can Cyclistic use digital media to influence casual riders to become members?

**Project Goal:**

Director assigned first question to answer: How do annual members and casual riders use Cyclistic bikes differently?

**Sources:**

Data : <https://divvy-tripdata.s3.amazonaws.com/index.html> License : <https://www.divvybikes.com/data-license-agreement> Note: The datasets have a different name because Cyclistic is a fictional company. For the purposes of this case study, the datasets are appropriate and will enable to answer the business questions.

**ROCC**

Reliability: The data came from the bicycles(tracker) Originality: The data came from the primary source (Motivate International Inc) Comprehensive: The data contains start time and end time, membership type, and starting station and ending station, this should be sufficient to identify the trends and conduct analyses. Current: The data is based on the most recent 12 month-period available. Cited: The data came from the primary source

**Data Cleaning:**

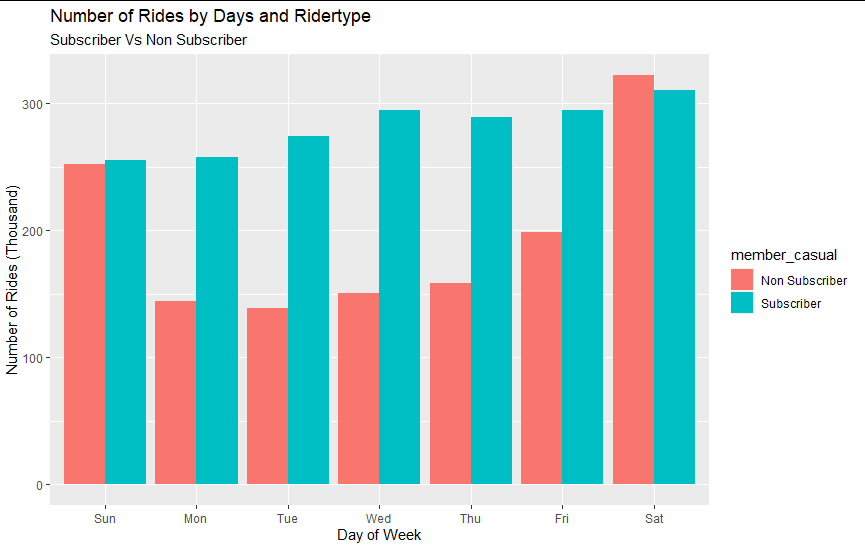
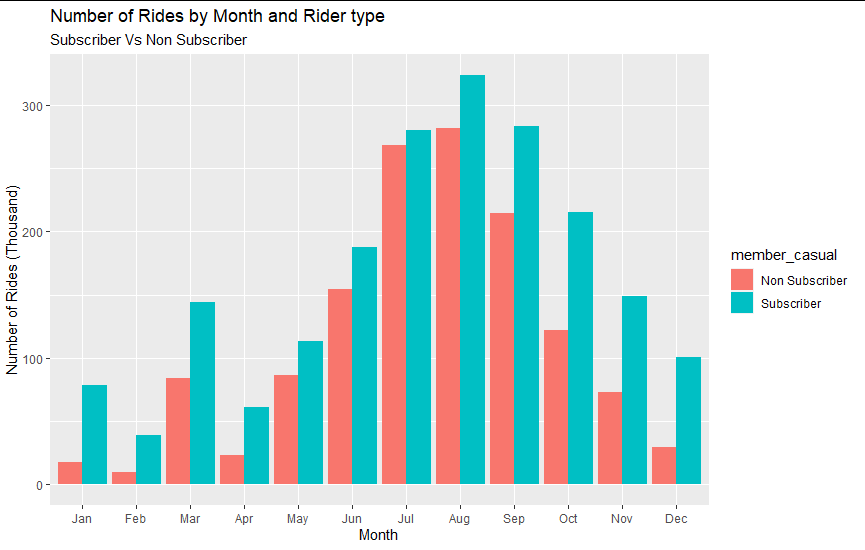
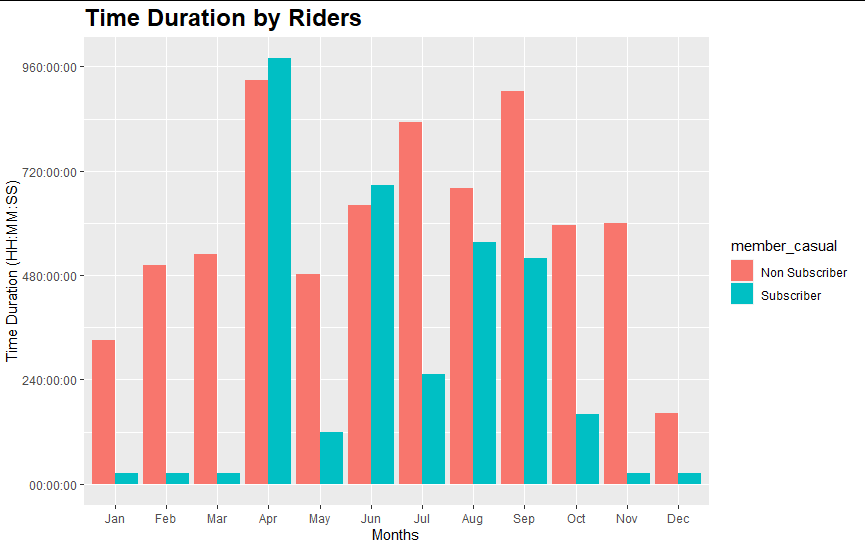
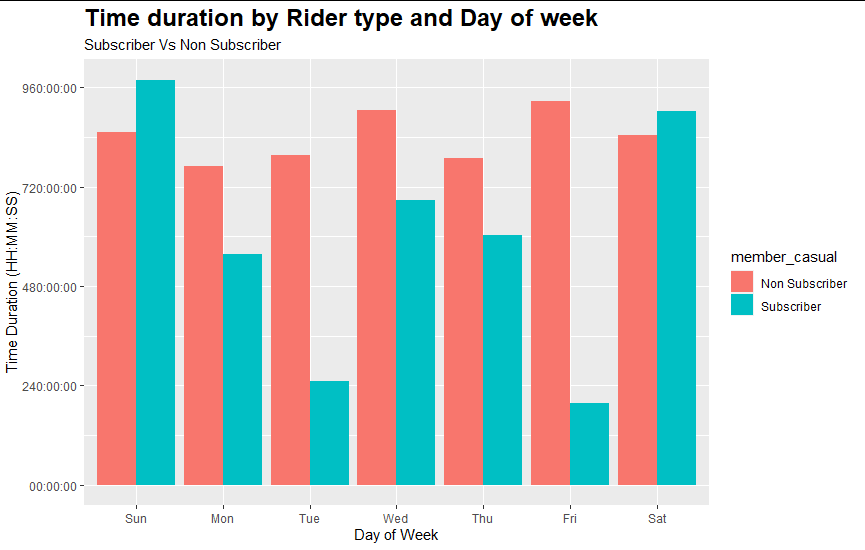
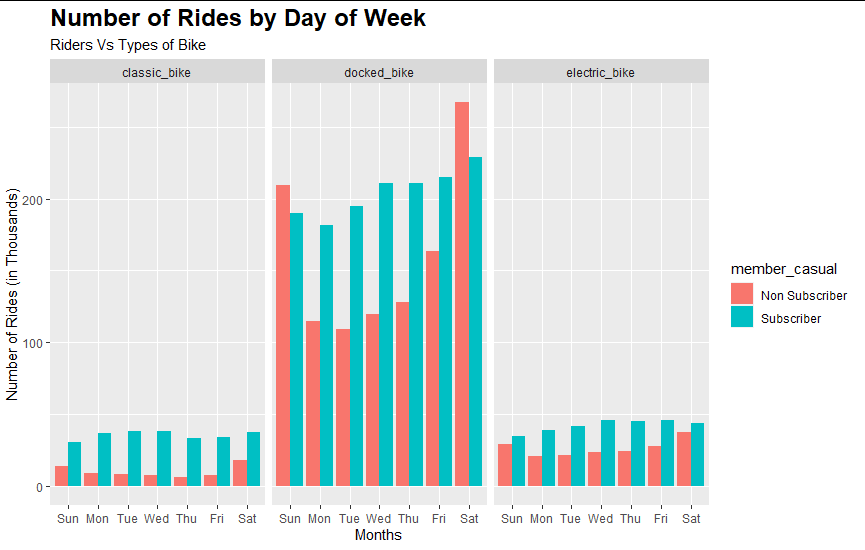
After data downloaded (Apr2020-Mar2021), following changes done in file:

1. Combined all 12 months data in one file.
2. Changed column "started\_at" and "end\_at" into date/time from characters.
3. Changed Column "member\_casual" data as "casual" to "Non Subscriber" and "member" to "Subscriber".
4. Analyzed on missing values and outliers.
5. Filtered "started\_at" greater then "ended\_at".
6. Removed unwanted columns.
7. Created new column "ride\_length" format "HH:MM:SS", "day\_of\_week" format as number or general and "month" format charcter.

**Analyze:**

1. Calculate mean of "ride\_length".
2. Calculate maximum of "ride\_length".
3. Caluclate mode of "day\_of\_week".
4. Calculate the average ride\_length for users by day\_of\_week.
5. Calculate the number of rides for users by day\_of\_week by adding Count of trip\_id to Values.

**Share:**

[](https://user-images.githubusercontent.com/72040187/132130483-22ef12e2-42a9-40e9-ada2-56e27f9b02e1.png) [](https://user-images.githubusercontent.com/72040187/132128799-1ae0aa9b-7895-4493-9d6e-9d0aac7e052a.png) [](https://user-images.githubusercontent.com/72040187/132128862-557a2366-902c-44e5-9cfd-bc000d77c22f.png) [](https://user-images.githubusercontent.com/72040187/132129019-08709c39-7e70-4ab2-8988-4628e55107e0.png) [](https://user-images.githubusercontent.com/72040187/132130239-c8a3bfe8-2251-4bcd-9b0f-c7fd2b2b74b9.png)

**Obseravtions:**

1. Time duration of Non subscriber riders are more then Subscriber riders.
2. Non subscribers riders rider's usually on weekdays.
3. Maximum numbers rides are taken in the month of June, July and August.
4. From three bike types ie "Classic Bike", "Docked Bike" and "Electric Bike", maximum number of riders rides on Docked Bike.