**Cyclistic Bike-Sharing Company**

**Github link for queries:**

<https://github.com/mysto-007/PortfolioProjects>

**Documenting all the process:**

The data analysis process contains six steps which are the following:

* **Ask**
* **Prepare**
* **Process**
* **Analyze**
* **Share**
* **Act**

**Ask:**

The business task is the following:

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

**Prepare:**

We have used past 12 months of data of year 2021 from divvy\_trip data.We obtained the data from [**Index of bucket "divvy-tripdata"**](https://divvy-tripdata.s3.amazonaws.com/index.html) site. It is ROCCC (Reliable, Original, Comprehensive, Current, Cited). The data has integrity as all the data is accurate and consistent throughout all of the files.

**Process:**

As we explore the data initially in the Excel and checked for errors there were no errors although it has some null values but we will deal with it in a bit. We manipulated data a little for the further calculation purposes. We made a new column to calculate the ride\_length of each ride by simply subtracting the ended\_at from started\_at column values and then we made another column(day\_of\_week) to find the day at which day the ride started by simply using weekday() function.

After that we uploaded all of the files in the MS SQL Server for the analysis purposes.

**Analyze:**

This is the most important phase of the data analysis as we explore all the data and identify different trends and insights from the data.

All the data was uploaded to the MS SQL SERVER for analysis. Since we uploaded all the files separately so there were separate tables for every file so the first step we did was to take the union of all the table and saved the output in the new table called ‘cyclistic\_data’.

The query to take the union was this:

-- Taking Union of all tables and making a new table named cyclistic\_data

Select

\*

INTO cyclistic\_data

FROM

(

SELECT

\*

FROM

cyclistic..['2021-01-divvy-tripdata$']

UNION ALL

SELECT

\*

FROM

cyclistic..['2021-02-divvy-tripdata$']

UNION ALL

SELECT

\*

FROM

cyclistic..['2021-03-divvy-tripdata$']

UNION ALL

SELECT

\*

FROM

cyclistic..['2021-04-divvy-tripdata$']

UNION ALL

SELECT

\*

FROM

cyclistic..['2021-05-divvy-tripdata$']

UNION ALL

SELECT

\*

FROM

cyclistic..['2021-06-divvy-tripdata$']

UNION ALL

SELECT

\*

FROM

cyclistic..['2021-07-divvy-tripdata$']

UNION ALL

SELECT

\*

FROM

cyclistic..['2021-08-divvy-tripdata$']

UNION ALL

SELECT

\*

FROM

cyclistic..['2021-09-divvy-tripdata$']

UNION ALL

SELECT

\*

FROM

cyclistic..['2021-10-divvy-tripdata$']

UNION ALL

SELECT

\*

FROM

cyclistic..['2021-11-divvy-tripdata$']

UNION ALL

SELECT

\*

FROM

cyclistic..['2021-12-divvy-tripdata$']

) cyc

After that we checked the data to see the new table and there were **5595063 rows** in the dataset. But since there were some null values so we write query to clean the data so it won’t affect the analysis.

We run this query to clean the data from nulls.

SELECT

COUNT(\*)

FROM

cyclistic..cyclistic\_data

WHERE

start\_station\_id is not null

and

end\_station\_id is not null

The resulting dataset contains **3556178 rows** and we have cleaned **2038885 rows.**

Then we started exploring the datasets by writing some queries and you can see my queries in the link provided at the top.

We found some great insights from the data. The max average-rideLength of member-casuals per are as follows.

|  |  |
| --- | --- |
| Member | 15.28 |
| Casual | 33.09 |

Then we found the percentage each type of customer in the company.

|  |  |
| --- | --- |
| Member | 54.48% |
| Casual | 45.52 |

Then we found that there were 2 months were the rides of the casuals are greater than the members and these months are July and August.

The favorite bike for each type(member-casual) is **Classical bike.** It is the most popular bike among both.

The busiest time for both member and casual is below.

|  |  |
| --- | --- |
| Member | 17 (5 pm) |
| Casual | 17 (5 pm) |

Then we found top stations among both.

|  |  |
| --- | --- |
| Member | Dearborn st. & Erie St |
| Casual | Streeter Dr and Grand Ave |

**Share:**

Since no case study or analysis project is complete without a good dashboard which shows all the details in the appealing visualization. I have made all the visualizations in tableau. Visualizations will clearly communicate all of my finding.

To see the dashboard visit the link: <https://public.tableau.com/app/profile/hanan.saeed/viz/Cyclistic_bike-shareCapstoneProjectGoogleDataAnalytics/Dashboard1>

**Act:**

It is the time to give some recommendations based on the data.

* The casual members are more profitable in months like July and August as their rides are more than annual members. So, to convert them into annual members some packages should be given to them so they can take them and become annual members.
* The ride lengths of the casual members are greater than annual members so some incentive should be given to them for longer ride length and also some promotion should me done besides the incentive to convert them into annual members.
* Company can make promotion and include that ride can be started form any station and can be ended at any station, it is the plus point and it provide them the ease of mind as they don’t have to go to the same station to end the ride.