**Case Study: Report Cyclistic**

By: Priya Singh Date: 2022:04:28

**Tools:** Excel, My SQL, Tableau.

1. **A clear statement of the business task.**

* -Find out how do annual members and casual riders use Cyclistic bikes differently? By checking historical data.
* -Explain, why would casual riders buy Cyclistic annual memberships?

By checking price difference between annual and casual riders.

Find benefits of annual membership.

* -How can Cyclistic use digital media to influence casual riders to become members?
* Creating ADs and Campaigns.

**Stakeholders:**

Director of marketing and Executive team.

1. **A description of all data sources used.**

Cyclistic’s historical trip data available publicly.

Data available monthly, quarterly, and also for two quarters.

Data is unbiased.

More than 12 months data is provided, seems to be problem with data.

*I will ensure addressing licensing, privacy, security, and accessibility of data by understanding terms and conditions and following the same.*

**Description of all data sources used** - All data sources provided are unorganized. Also, there are few wrong data entries in this data set.

Total number of files 42.

Year 2013 – 1file with stations and trips.

Year 2014- 2 files for two quarters each Q1-Q2 and Q3-Q4.

Year 2015- 2 files for two quarters each Q1-Q2 and Q3-Q4.

Year 2016- 2 files for two quarters each Q1-Q2 and Q3-Q4.

Year 2017- 2 files for two quarters each Q1-Q2 and Q3-Q4.

Year 2018- 4 files for four quarters.

Year 2019- 4 files for four quarters.

Year 2020- 10 files where one Q1 file for one quarter and other nine files are for monthly data from April to December.

Year 2021- 12 files for each month from January to December.

Year 2022- 3 files for each month from January to march.

1. **Documentation of any cleaning or manipulation of data.**

**Excel-**

While running some functions in excel workbook, every function took too much time to find the result. I merged only 4 month’s data set in excel and I think the data set is to longer to perform in excel.

I wanted to analyze 1 complete year’s data but not able to merge 12 months file in excel because of its size. For that I cleaned, analyzed, and made pivot table of each month data separately. I found some important results while my analysis process like ***Mode of riders’ type, Ride length of every rider and which day maximum riders use bikes***

**MySQL-**

1. Imported 12-month data is MySQL database.
2. Merged all data by using UNION function and combined it into one single table.
3. There were few columns with missing data so by avoiding those columns, used SELECT and CREATE TABLE statements and made one new table for my further analysis process.
4. I’m getting few null values in my new table and not able to remove them with query.

Then I used ‘***Export***’ button in Results Pane and saved my other query result as

**.csv** file.

In my new table I used

¨ **UPDATE [table] SET [column]=0 WHERE [column] IS NULL¨ clause** to replace null values with ‘**0’ .**

1. As my final table after analysis process, I saved 11 columns like *wirelength*, *Day of* *week* ride started at and only *date* from column Started\_at and Ended\_at.
2. I used **Expor**t button again and saved this table as **.CSV** file for my visualization process.
3. **A summary of your analysis**

I started by downloading historical data provided to us for case study and for relevant analysis I used latest 12-month data for my analysis

**Ask:** I defined scope of my analysis by making an understanding of expectation from stakeholders.

**Prepare:** Nest I downloaded provided data and checked if data is correct and has correct content also, I shorted which file / data which are most relevant and useful out of provided data, during prepare phase I have also converted all the: CSV file to excel and stored the in specified folder structure for excel and .csv files**.** Also, I have made sure that licensing, privacy, security, and accessibility of data are as per by terms and conditions, and I followed he same.

**Process:** During process phase I started cleaning data in MS excel where I removed columns where data was not consistence and might not be useful during the analysis, I also checked rows with missing values since they are irrelevant and might affect analysis part.

**Analyze:** In MS excel I shorted data and created Pivot table to visualize pattern of casual users and Members based on ride length and week of the days. I also realize that since the data set is too big it’s quite difficult to handle such a big data in MS excel hence, I decided to use MY SQL for further data analysis. In My SQL started joining files and have used provided .CSV files and using queries I created tables as result of my analysis which I can further use in visualization tools.

**Share:** For creating visualized results from my analysis, I have imported .CSV files to Tableau and uploaded visual results on my tableau dashboard. showing users share between casual and members and frequencies for days of week to understand pattern for recommendation.

**Act:** Based on my analysis and visualized results I have given following recommendation to board of directors and marketing teams.

1. **Supporting visualizations and key findings**

Visualization done in Tableau.

One year data visualization [Link](https://public.tableau.com/views/CyclisticCaseStudyDataVisualization1Year/Dashboard1?:language=en-US&:display_count=n&:origin=viz_share_link)

One month data visualization. [Link](https://public.tableau.com/views/Cyclisticcasestudy_1monthvisualization/Dashboard1?:language=en-US&:display_count=n&:origin=viz_share_link)

After visualization process, I got to know few things.

1. Casual riders use every bike type like classic bikes, docked bikes and electric bikes were member riders uses only classic bikes and electric bikes.
2. Maximum riders use electric bikes.
3. Maximum *casual riders’* rides bike on *weekends*, mostly on *Saturday.* And maximum *member riders’* rides bike on *weekdays,* mostly on *Wednesday.*

4- This one-year data set is showing that casual riders are less than member riders.

1. **Your top three recommendations based on your analysis.**

Here are top three recommendations from this analysis:

1. As per results from analysis of data from last 12 month my recommendation to marketing team is to organize social media campaigns for casual riders of weekends since maximum casual riders are on weekend and promote benefit of yearly member ship over one time cost and benefit for users.
2. Marketing team can also mention that if a user uses rides all week the total cost will be less compared to paying per ride and can have more casual riders on weekdays and eventual have them convert to members
3. Marketing team can also advertise about environmental benefit of using bike riding compared to other transport and motivate casual riders to use bikes more often and have benefit of monthly or yearly membership.