Google Capstone Project

This case study involves working as a junior data analyst at Cyclistic, a bike-share company. Our team is focused on understanding the different riders (casual, annual member), and their usage patterns. By identifying trends, patterns and by discovering insights, we will design a marketing strategy that will help convert more casual riders to members. Our strategy will be supported by our analysis within the data and will have data-driven recommendations.

Case Study 1:

How does a bike-share navigate speedy success?

Cyclistic is a growing company that has seen success within the bike-share industry in Chicago, and the company provides various pricing plans to accommodate consumers. Single-ride passes, full-day passes, and annual memberships are the options, but the company's financial analysts have given us insight that annual members are more profitable and a key factor for growth. Casual riders are aware of our company and the options we have, and there is potential to convert them into annual members if we can find the differences between the members. In this case study, we will explore the data to understand the different members and to find a data-driven strategy that can help us with marketing and assist in maximizing annual members.

In addition to the strategy, we will provide a report of the whole process including:

1. A clear statement of the business task

2. A description of all data sources used

3. Documentation of any cleaning or manipulation of data

4. A summary of your analysis

5. Supporting visualizations and key findings 6. Your top three recommendations based on your analysis

Ask:

To provide guidance for this case study, these questions will help align us to our goals and what we're trying to find.

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?

Prepare:

We will be using Cyclistic's historical data for our project and analysis. For compliance and security purposes, we will not be using any of the rider's PII, so we won't be able to know some factors such as residential areas of the rider's and if the rider had purchased single passes multiple times using the same credit card.

To ensure our data has no issues with privacy and security, we will store the dataset properly, organize for analysis, sort, and filter by following these guidelines:

1. The dataset is located in Index of bucket "divvy-tripdata", and includes many historical datasets from Cyclistic.
2. For our analysis, we will be downloading the datasets that correspond to 2021.01-2022.01. We are pulling data from these dates, so we have the right dataset for the year 2021 analysis.
3. We will be storing a copy of all the datasets in a safe, easily accessible location.
4. To ensure credibility, licensing, privacy, and security, we will be sorting and filtering out any information in the dataset that could lead to such issues. We will carefully examine the datasets and use sorting and filtering to find any issues within the dataset or any misleading data.

Process:

Before we begin our analysis, it is important for us to make sure the data has no errors, it is organized and filtered, ensures data integrity, the data is clean and ready for analysis.

For our tools, we will utilizing two different types of software: Excel and R

We are mainly using Excel for the beginning steps of the project; we will save the datasets and store an extra copy in a different folder for security and accessibility purposes. We will also be using Excel for creating some different columns that will aid our analysis.

We will be using the R software for our analysis; therefore, we will upload our datasets to R and make sure all datasets are consistent and attributes are the same in each dataset. We will perform our data cleaning in R, so we will merge all the datasets needed and inspect for any errors or problems to fix. We will perform the needed data cleaning, aggregating, conversions before we start our descriptive analysis. Data cleaning will also be documented with the necessary code.

Analyze:

For our analysis, we will need to concatenate all the files into one csv, so we can perform analysis on the entire timeline of the dataset. We will also format our dataset so that all columns have correct data types, and the dataset will not contain any null values.

When we start our analysis on R, we will dive into descriptive analysis where we will compare the members and casual users by comparing their mean, median, max, min. We will also see the average ride times for the different users and analyze the data by type and weekday to compare the ridership data. We will look for any trends and relationships by examining the analysis and performing visualizations.

These insights will allow us to see how the different users are utilizing our service and ultimately build profiles for members and casual users for more comparison and analysis. The visualizations will help us in our next step: Sharing.

Share:

With our analysis, we believe that we will be able to tell key differences between annual members and casual riders. We will share our data storytelling in the report after we have finished the previous steps/

Our audience for this project is the marketing analytics team of Cyclistic, and Lily Moreno, the director of marketing and manager of the team. We will create a PowerPoint presentation to show them our objectives, key findings, and recommendations. Through our findings, Lily Moreno will be able to give a full-depth analysis and recommendations, for our objective, to the executive team.

The data visualizations, we created, will be geared towards our objective, where we will be able to see the key findings and differences for the users. We will give more detail to the visualizations with labels that are easily readable and correspondent to the analysis. Also, the visualizations will be able to show the key highlights of our analysis by using distinguishable colors.

Act:

After completing the previous steps, we will be able to give a final conclusion on the different habits and usage between the members and casual riders.

Our insights can be used for the next marketing campaign or for a new, upcoming membership sale. Our key findings can be help guide the marketing.

We can dive into further analysis with different factors to see if there are other key findings, but through our project, the insights will definitely help guide the marketing campaign.

We could possibly add in weather data to see if it has an impact on who’s utilizing our service, and we can also compare on the distances the members and riders use the bike for to compare the distance biked for the type of users versus the average duration (our current analysis).