**Case Study: Cyclistic Bike-Share Study**

***Scenario***

I’m a junior data analyst working in the marketing analyst team at Cyclistic, a bike-share company in Chicago. The director of marketing believes the company’s future success depends on maximizing the number of annual memberships. Therefore, my team wants to understand how casual riders and annual members use Cyclistic bikes differently. From these insights, my team will design a new marketing strategy to convert casual riders into annual members. But first, Cyclistic executives must approve my recommendations, so they must be backed up with compelling data insights and professional data visualizations.

***Data analysis process phase#1: Ask***

Ask Three questions will guide the future marketing program:

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

I was assigned the first question to answer

**Business task**

The company wants to increase profit of the company. The director of marketing believes that increasing the number of annual memberships will reinforce the company profit. Rather than creating a marketing campaign that targets all-new customers, The director of marking believes it is easier to convert casual riders into members. In order to design marketing strategy to convince casual riders to subscribe annual membership, the director of marketing would like to know how annual members and casual riders use Cyclistic bikes differently.

**Analytic tools**

The tools used for data analysis in this project are MS SQL Server for data cleaning and excel for data visualization

***Data analysis process phase#2: Prepare***

**Description of data sources used**

1. The historical trip data was used in this study. The data has been publicly made available by Motivate International Inc. under license.
2. The bike trip data was collected starting from March 2021 to February 2022. Data is organized in rows and columns. Each table contains 13 fileds:

ride\_id,

bike\_type,

date/time start

date/time end

start\_station

end\_station

start\_latitude

start\_lontitude

end\_latitude

end\_longtitude

type of rider (member/ casual user)

1. Twelve data files were downloaded in CSV file and initially stored in the folder name ‘DivvyTripData\_original’. Each file name contains name ‘divyy-tripdata’, the date and version.

202103-divvy-tripdata

202104-divvy-tripdata

202105-divvy-tripdata

202106-divvy-tripdata

202107-divvy-tripdata

202108-divvy-tripdata

202109-divvy-tripdata

202110-divvy-tripdata

202111-divvy-tripdata

202112-divvy-tripdata

202201-divvy-tripdata

202202-divvy-tripdata

1. Each cell of data contains data from an individual trip. In order to reveal the different riding behaviors of member riders and causal riders, the data was used to analyze to see the trends of month of year preference, day of week preference, time of day preference and ride duration of both rider type.

***Data analysis process phase#3: Process***

All data files were then uploaded to MS SQL server for further analysis.To ensure data integrity, after uploading data into SQL server, I have combined all 12 tables into a single table. Then check if

1. The number of total rows and columns are correct
2. Data has a consistent format.
3. There is any duplicates data.
4. Data has a correct data type

To confirm if the data was clean, I rechecked it again.

***Data analysis process phase#4: Analyze***

To answer a question, how do annual members and casual riders use Cyclistic bikes differently?

Data exploration has been done to reveal the different behaviors of annual and casual rider based on day of week, month of year, time of day preference and ride duration,

Here is the summary of my analysis (please see file ‘BikeShareSQLQuery’ for the SQL script)

STEP#1 Add a ride length column to a table and calculate the ride length for all individuals.

STEP#2 Add day of week column to a table.

STEP#3 Calculate total number of rides by day of week for each type of rider.

STEP#4 Calculate the average, max and min of the ride length (duration) by day of week for each type of rider (member, casual riders).

STEP#5 Calculate total number of rides by month of year for each type of rider.

STEP#6 Calculate total number of rides by time of day for each type of rider.

**Finding**

|  |
| --- |
| 1.The number of casual rides is raised on weekend whereas, the number of member rides on weekday is slightly higher than that of on weekend.  2. The number of casual rides dramatically increases in the afternoon on weekend whereas the number of member rides is high in the morning and the evening on weekday.  2. Surprisingly, even the number of member rides is higher than that of casual rides overall, casual rides spent longer time on rides on every day of week and overall.  3. The number of rides dramatically increases in summer in both casual riders and member riders and drop in winter. |

***Data analysis process phase#5: Share***

(Please see file ‘BikeShareDataVisualization’ for a dashboard)

The result reflects the fact that many casual riders use bike for leisure whereas member riders use bike to commute to work or school on weekday.

***Data analysis process phase#6: Act***

Recommendations

1. Since the casual riders tend to spend longer time per trip than the member riders. In order to persuade the casual riders to become member riders, I would recommend extend the limit ride length per trip for annual membership
2. Launch campaign to encourage locals to use bike to commute to work or school would be another way to convert the casual riders to become member riders.
3. Offer membership discount in low season.