**CYCLISTIC CASE STUDY**

The process followed is the Google recommended 6-phase data analysis process: Ask, Prepare, Process, Analyze, Share & Act.

**ASK**

1. What is the problem you are trying to solve? How do annual members and casual riders use Cyclistic bikes differently?
2. Who are the stakeholders?

* Cyclistic Executive team - notoriously detail-oriented executive team will decide whether to approve the recommended marketing program.
* Cyclistic marketing analytics team
* Lily Moreno - The director of marketing and my manager

1. Clear statement of the business task: How annual members and casual riders use Cyclistic bikes differently.

**PREPARE**

The data was downloaded and stored in my local storage drive. Both the CSV and Excel formats were downloaded.

The data is in spreadsheet/table format.

Does the data ROCCC?

* Reliable: The dataset is reliable as it has been clearly stated in the exercise document that the dataset is appropriate for this exercise
* Original: The dataset is not original to Cyclistic since it is a fictional company, but this isn’t an issue for this exercise as indicated in the exercise document.
* Comprehensive: The dataset is comprehensive. It contains a lot of relevant information about the trips made by Cyclistic customers.
* Current: Data is current as it has the most recent 12-month data as of 02nd March 2022(Feb 2021 to Jan 2022)
* Cited: Data is cited. It has been provided by Motivate International under a license.

**PROCESS**

**Check the data for errors**

For the cleaning process I used Excel.

1. Checked for blanks: Some of the columns e.g., start\_station\_name had blanks, but I realized this won’t impact my analysis as these rides had both a start time and end time which were the columns needed for the analysis.
2. Duplicates: There were no duplicates found
3. Data format consistency: Ensured the consistency in the data formats for each of the columns.

**Data transformation**

Big query platform was chosen to do the transformation since some of the files are large and the process would be slow on Excel.

The steps taken for the data transformation were:

1. All the csv files were therefore uploaded for processing.
2. CSV files were merged to into a full 12-month view.
3. Created 2 new columns: “ride\_length” to calculate the length of each ride and day\_of\_the\_week to calculate the day of the week each ride started.
4. Some start/end station IDs had “TEST” in their names. I assumed those were test/maintenance rides and therefore deleted them to avoid any bias in the analysis.
5. Created “part\_of\_day” column to show what part of the day the trip occurred.

After transforming the data, I noticed something strange with some of the values:

1. Some ride lengths were 0 minutes long meaning no trip actually occurred.
2. Some maximum ride lengths had very high values e.g., 55944. After looking deeper into the matter, it became clear some casual riders returned the bikes several days after unlocking them.
3. Some ride lengths had negative values because end date was before start date. This was assumed to be a system error.

These trips were removed because they would skew the analysis.

**ANALYZE**

**Google Big Query**

For the analysis, Big Query platform was used because Excel can’t handle the data size (over 2M rows)

**12-Month Analysis**

Firstly, analysis was done for the 12-month period ranging from Feb 2021 to Jan 2022.

**Observations**

Total rides/trips

|  |  |
| --- | --- |
| member\_casual | total\_rides |
| casual | 2,504,678 |
| member | 3,036,527 |

* Casual riders took less trips than members.

Average ride length

|  |  |
| --- | --- |
| member\_casual | average\_ride\_length(min) |
| member | 13.68 |
| casual | 28.38 |

* Casual riders rode for longer. At least twice longer!

Average ride length day of the week

|  |  |  |
| --- | --- | --- |
| day\_of\_week | avg\_member | avg\_casual |
| Tuesday | 12.83 | 25.83 |
| Saturday | 15.35 | 30.72 |
| Monday | 13.26 | 28.79 |
| Sunday | 15.77 | 32.79 |
| Thursday | 12.80 | 24.57 |
| Friday | 13.38 | 26.45 |
| Wednesday | 12.81 | 24.59 |

* Casual riders rode for longer for each of the days.
* Average ride length is higher during the weekend for both types of riders.

Number of rides by day of the week

|  |  |  |  |
| --- | --- | --- | --- |
| day\_of\_week | total\_rides | member\_rides | casual\_rides |
| Friday | 800,080 | 439,995 | 360,156 |
| Monday | 697,680 | 413,741 | 283,993 |
| Saturday | 978,832 | 426,320 | 552,574 |
| Thursday | 732,024 | 448,560 | 283,507 |
| Sunday | 847,608 | 371,532 | 476,140 |
| Tuesday | 735,790 | 463,599 | 272,234 |
| Wednesday | 749,775 | 473,340 | 276,487 |

* Casual riders took more bike rides on weekends in comparison to members who preferred weekdays

Number of rides by rideable types

|  |  |  |  |
| --- | --- | --- | --- |
| member\_casual | classic\_bike\_rides | docked\_bike\_rides | electric\_bike\_rides |
| member | 1,958,989 | - | 1,078,098 |
| casual | 1,256,916 | 308,648 | 939,527 |

* Overall, both preferred the classic bike
* Casual riders used all the types of bikes unlike members who did not use docked bikes at all.
* Members used the electric & classic bikes more than the casual riders.

Number of rides by time of the day

|  |  |  |
| --- | --- | --- |
| part\_of\_day | casual | member |
| Morning | 461,243 | 842,908 |
| Afternoon | 898,391 | 980,268 |
| Evening | 732,118 | 909,815 |
| Night | 413,339 | 304,096 |

* For each of the time of day, member riders had more trips except at night.

**Seasonal Analysis**

The analysis was then done a seasonal basis to check for any trends. The four seasons were assumed to run in the following manner:

1. Winter – December to February
2. Spring – March to May
3. Summer – June to August
4. Autumn/Fall – September to November

**Seasonal Observations**

Total rides

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **member\_casual** | **Winter** | **Spring** | **Summer** | **Autumn** |
| casual | 96,966 | 473,520 | 1,212,414 | 722,010 |
| member | 298,177 | 613,347 | 1,116,535 | 1,008,779 |

* Members had more trips than casual riders except during the summer.

Average ride length

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **member\_casual** | **winter** | **spring** | **summer** | **autumn** |
| member | 12.08 | 14.65 | 14.35 | 12.81 |
| casual | 22.13 | 32.99 | 29.14 | 24.92 |

* Shows a similar pattern as the 12-month review; casual riders’ trips were almost twice as longer as members’ trips.
* Spring had the highest average ride length for both groups.

Mode of the day of the week

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| member\_casual | **winter** | **spring** | **summer** | **autumn** |
| casual | Saturday | Saturday | Saturday | Saturday |
| member | Thursday | Saturday | Tuesday | Wednesday |

* Casual riders preferred Saturday while members preferred weekdays except during Spring where the mode was Saturday.

**SHARE**

This part of the report has been developed in PowerPoint. The link to the PowerPoint has been provided on the GitHub platform.