Case Study – CYCLISTIC BIKE-SHARE

# ASK – Business Task

Identify the usage differences between Cyclist annual members and casual riders and establish digital media strategies for attracting casual riders into annual memberships.

# PREPARE – Data Preparation

Data obtained from 19 months of public Cyclist trip data.

### Data Description:

* Spans the following dates: 04/2020 – 10/2021
* File timespan: monthly
* File type: “.csv”
* Licensing: The data has been made available by Motivate International Inc. under this [license](https://www.divvybikes.com/data-license-agreement).

### Variables:

* ride\_id: ID attached to each trip taken
* rideable\_type: type of ride taken
* started\_at: day and time trip started, in CST
* ended\_at: day and time trip ended, in CST
* start\_station\_name: name of station where trip originated
* start\_station\_id: ID of station where trip originated
* end\_station\_name: name of station where trip terminated
* end\_station\_id: ID of station where trip terminated
* start\_lat: start station latitude
* start\_lng: start station longitude
* end\_lat: end station latitude
* end\_lng: end station longitude
* member\_casual: member type; "member" is an annual subscription user, "casual" is a single-ride or daily-pass user

# PROCESS – Cleaning and Data Manipulation

* Not all files had uniform column type, hence a forced column type structure was used to maintain integrity.
* There were 1304 observations in which the trip’s start time was after the end time, resulting in a negative total trip time, these observations were removed.
* There were several Station ID’s that seemed to be variations of an existing ID; however, this is a supposition as the available station meta data does not confirm this necessarily. Given the number of observations in which this occurs is large, I maintain them and assume they are existing stations.
* A new column for travel time and weekday was included in the final data frame.
* The final, aggregated, and cleaned data frame contains 8,090,879 observations for the time period.

# ANALYSIS

* When analyzing the trip length times, several outliers were observed. Thousands of observations contained either very short or very long trips. Given that these outliers can potentially influence the mean results, a limited time frame was considered for these trips: between 60 seconds and 2 hours.
  + This filter removed a total of 247,245 observations and the remaining count was of 7,843,634 total observations, which corresponded to a removal of about 3% of the total count.
  + The previous average ride time length was 1506 seconds (25 minutes) and the new average was 1145 (19 minutes). Showing that the few very large values for trip times were greatly skewing the average.
* The trip time length distribution looked like this:

Chart, histogram

Description automatically generated

* When adding a filter for usage between Members and Casuals, a clear difference can be seen, Casual riders have a much longer trip time average than Members. A few possible reasons for this difference can be hypothesized:
  + Annual members will be more prone to using the bicycles for more menial tasks as there is no additional cost in using them. Casual members might feel disinclined to using it for shorter distances as the idea of having a new cost consideration might weight against it. Casual users might want it to be more “worthwhile” to use it.
  + Shorter trips tend to be more utility oriented while longer trips might be an indication of enjoying time with the bicycle. This is further indicated with the difference in weekday usage.
* The following frequency distribution can be seen when filtering the type of member usage:

Chart

Description automatically generated

* When analyzing the weekday bicycle usage, there seems to be a similar percentage of usage during the week between members and casuals, however this proportion changes considerably during the weekend, especially on Saturday.

Chart, bar chart

Description automatically generated

* What is interesting to note is that usage among members does not really fluctuate during the week, except for the lower usage on Sunday and Monday, the number of trips is consistent.
* Casual users maintain a similar usage patter during the week, but on weekends, including Thursday, there is a spike in usage.
* This fits into the idea that there is a less utilitarian usage for casuals and a greater demand for enjoyment rides.
* A further detailing of this visualization can be obtained with a table demonstrating the change in member to casual usage ratio during the week:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Monday** | **Tuesday** | **Wednesday** | **Thursday** | **Friday** | **Saturday** | **Sunday** |
| **Ratio** | 1.483866 | 1.696857 | 1.703111 | 1.586477 | 1.256286 | 0.8234727 | 0.8430467 |

* Clearly the ratio of members to casuals is much smaller during the weekend, on Saturday and Sunday the ratio is such that there are more casuals than annual members.
* Graphically, this relationship is even more visible:

Chart, bar chart, histogram

Description automatically generated