# DETAILED PROCESS

# DATA ANALYSIS PROCESS


Identify business task considering the key stakeholders

Source, organize, sort and filter data

Transform available data to usable form for analyse phase

Performing calculations and identifying trends & relations

Present data insights using compelling visualizations

Recommendations for further actions

# PROBLEM STATEMENT

*"Based on the past 12 months, how do casual riders differ from annual members in their usage of Cyclistic services ?"*

Scope : 12 months starting from August 2021 and ending on July 2022

Stakeholders : DOM, Cyclistic Executive team, Marketing Analyst team

Expected outcome : The results of the analysis would help the marketing team get approval from the executive team of Cyclistic for a marketing plan to convert more casual riders to annual memberships.

Definitions

1. Annual members (denoted as member) are those riders with annual memberships,
2. Casual riders (denoted as casual) are riders who pick a ride without an annual membership.

# DATA PREPARATION

Data source : Public data set made available by Motivate International Inc. under license. The original owners of the data is the City of Chicago accessible by all hence can be considered relevant, unbiased and current.

Data organization : 12 nos. of .xls files tagged by month and year

Data contents: : Data is structured in tabular form with 13 columns with each row showing details of a unique ride. Details include type of customer, type of bike chosen, starting and ending time, station details along with geographic co-ordinates.

This data is appropriate to answer the business task since it describes the rider activity for 12 months.

Restriction: Data privacy laws do not allow us to relate each ride to a particular person however, a unique ride id has been assigned to each ride.

# PROCESSING OF DATA

Tool chosen : Data cleaning done using SQL Server Management Studio 18

: Documentation done using Obsidian.

Steps taken to process data

1. Individual monthly excel files were joined to make one table (5,901,463 rows).
2. Data cleaning in each column involves checks for
    1. Unique / distinct values
    2. NULL values count.
    3. Any peculiarities in the column
    4. Check for spell errors and use of trim.
3. Check for relations between columns involve checking for inconsistencies.

Note: All assumptions made need to be clearly stated.

# PROCESSING OF DATA

## Assumptions made for cleaning process

1. Average latitude and average longitude is treated as a unique location point each station
2. 16 hours (960 minutes) of ride duration from ride start is only considered valid for the purpose of analysis.

## Results

- Type of bikes in the data include classic, electric and docked.
- There are 1243 unique station ids in total.
- The numeric coded stations with character length 3 and 4 are the original stations. Rest of the station ids are alphanumeric.
- There are stations with keywords such as 'test', 'charge' , 'repair', 'warehouse' which have not been deleted since they have been used by casual and members alike with valid start or end destination.
- Attempts have been made to clean NULLs in the data.
- End result table has been stored in SQL and labeled as cleaned data. Documentation also done.

# ANALYSIS OF CLEANED DATA

- Data queried using SQL to see trends in data
- Exported individual files to Excel in order to check conclusions.
- Made summary statements for results of the analysis including recommendations.

RESULT

- Casual riders have higher average duration of ride in comparison to annual members.
- Annual members have a higher number of rides than casual riders in any given day.
- Frequency of rides suggest that the majority of casual riders spend 60% more time than annual members each ride.
- Both casual and members show a higher preference for a classic bike, while docked bikes are exclusively used by casual riders for ride durations on average three times greater than classic or electric bikes.
- Casual riders have a high preference for stations closer to the harbor while annual members have a more uniform demand for stations closer to the business areas which are slightly away from the harbor.

# VISUALIZATIONS TO SEE TRENDS AND PRESENT

- Tools used: MS Excel, Tableau and MS Powerpoint
- Extracted data from SQL were displayed as dashboards in Tableau.

Based on the analysis, it was cross checked to verify that the results of the analysis are still in tune with the problem statement and not deviating.

Recommendations given for further study to make possible conversion of casual riders to annual members.