CYCLIST BIKE SHARE

Ajay Shankar A

2023-12-14

Table of contents

# 1. CYCLIST BIKE SHARE

## 1.1 Introduction

The analysis is done on **Cyclist Trip Data** obtained from *Coursera Google Data Analytics* course as part of Cap Stone Project.

The data contains month wise travel usage of bikes from the year of 2015-2023. We will be concentrating on data gathered in between *July-2022* to *June-2023* which will cover an entire year.

Let’s load the required packages first

* Loading the required packages i.e., pandas, numpy, matplotlib, seaborn and plotly.

import pandas as pd  
import numpy as np  
import matplotlib.pyplot as plt  
import seaborn as sns  
import plotly.express as px

### 1.1.1 Loading and Formatting Data

* Let’s look at the structure of the data in one of the downloaded .csv files.

trp\_data\_jul\_24 = pd.read\_csv("F:/Data\_Sci/Cap\_Stone\_Project/Cyclist\_trip\_data/202207-divvy-tripdata/202207-divvy-tripdata.csv")  
  
trp\_data\_jul\_24.info()

<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 823488 entries, 0 to 823487  
Data columns (total 13 columns):  
 # Column Non-Null Count Dtype   
--- ------ -------------- -----   
 0 ride\_id 823488 non-null object   
 1 rideable\_type 823488 non-null object   
 2 started\_at 823488 non-null object   
 3 ended\_at 823488 non-null object   
 4 start\_station\_name 711457 non-null object   
 5 start\_station\_id 711457 non-null object   
 6 end\_station\_name 702537 non-null object   
 7 end\_station\_id 702537 non-null object   
 8 start\_lat 823488 non-null float64  
 9 start\_lng 823488 non-null float64  
 10 end\_lat 822541 non-null float64  
 11 end\_lng 822541 non-null float64  
 12 member\_casual 823488 non-null object   
dtypes: float64(4), object(9)  
memory usage: 81.7+ MB

* Let’s look at the columns and try to understand what they represent
  + ride\_id is the unique identification token generated for each ride that was initiated.
  + rideable\_type indicates the type of bike used for the ride.
  + started\_at and ended\_at give us the time when the ride began and the ride ended respectively.
  + start\_station\_name and end\_station\_name give us the names of stations where ride began and ended respectively.
  + start\_station\_id and end\_station\_id are unique ID’s given to stations.
  + start\_lat and start\_lng represent co-ordinates where the ride began.
  + end\_lat and end\_lng represent co-ordinates where the ride stopped.
  + member\_casual identifies if the rider is a member or casual rider of the bike.

The trpdata\_july\_2022 contains r nrow(trpdata\_july\_2022) rows and r ncol(trpdata\_july\_2022) columns. In the results we can see all the columns and their data types.

* **Lets load data of remaining 11 months.**