Exploratory Data Analysis for Scooter Rental Application

Lubna Almuammar

Abstract

Scooter rental lets you rent an electric foot scooter for a quick errand, you rent the nearest available scooter, ride it to where you want to go, and leave it responsibly parked for the next person to ride. the Scooters will be distributed across the city, to make the most benefit of the application So where are the most suitable places to put the scooters? to answer this question, MTA dataset will be analyzed to figure out the list of Stations that have the most traffic to place the scooters.

Design:

Scooter rentals application, the application provides a shared vehicle the app provides the ability to share the devices using a short-term rental business model. To begin a rental, users unlock a scooter through the company's smartphone app. the Scooters will be distributed across the city, to make the most benefit of the application the scooters will distributed near the station especially the most crowded stations.

- problem is where to put the scooters in which station.
- value for the company to make more profit we must study the places where the scooters will be distributed.

Data:

The New York subway MTA turnstile data is a series of data files containing cumulative number of entries and exits by station, turnstile, date and time. Data files are produced weekly, data records are collected typically every 4 hours with some exceptions.

In this analysis a subset of data was used from the recent 4 months. Data size is over 3 million.

Algorithms:

- DATA COLLECTION :

  Gathered the datasets on single one database using sqlalcemy.

- DATA UNDERSTANDING :

  Understand the columns and what they represent.

- DATA CLEANING :

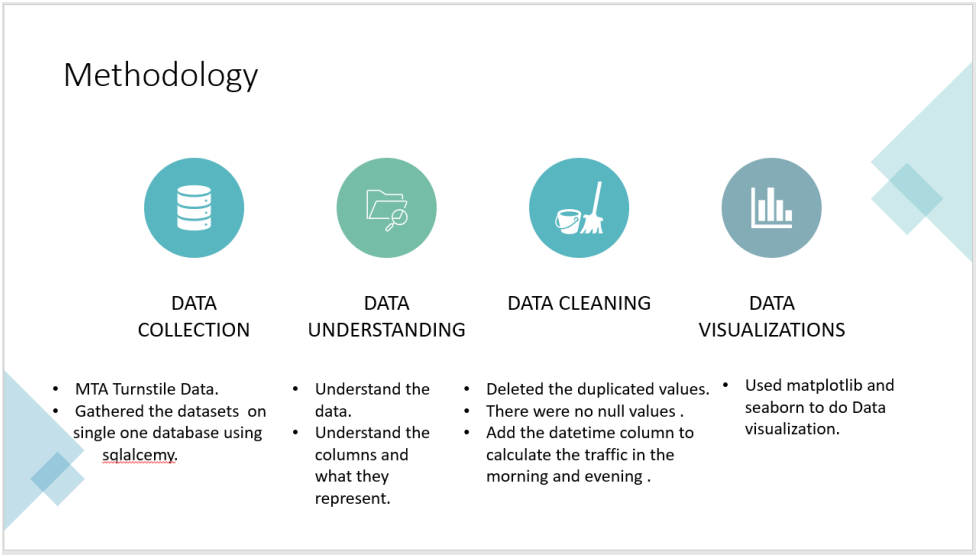
  Deleted the duplicated values.
  There were no null values.
  Add the datetime column to calculate the traffic in the morning and evening.

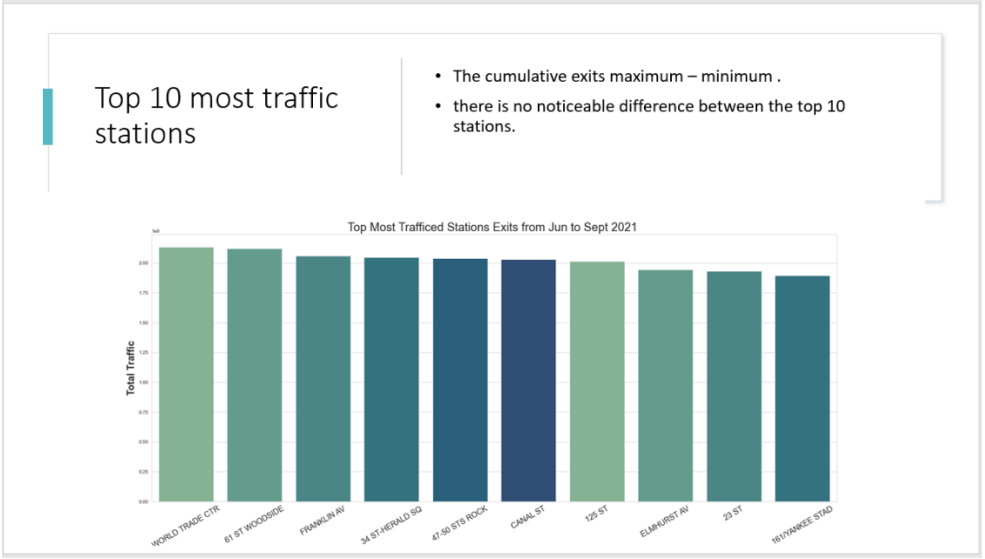
- DATA VISUALIZATIONS:
  Used matplotlib and seaborn to do Data visualization.

Tools:

- technologies: SQL, SQLite ,python, jupyter notebook.
- libraries: Numpy, Pandas, matplotlib.

Communication:





10 low traffic stations

- The cumulative exits maximum – minimum .
- There is a difference between the top 10 stations and the lowest trafficked stations.