OCTOBER 2021

EDA writeup

WRITTEN BY

Rania Almneie

WRITTEN TO

Moburst Advertising Company

Abstract

The Moburst project's goal is to use the MTA data to recommend the Moburst company to place their animated billboards at the most three frequented stations. In order to attract the greatest number of entrants to the station to advertisements. In addition, they want to know the total entries for each day of the week during the period for which I analyzed to determine the price of the advertisements. After doing Exploratory Data Analysis (EDA) I plotted many plots to visualize and communicate my finding.

Design

Moburst Advertising has three billboards that they want to place them in the top three most-frequented stations. Moreover, they want to know the total entries for each day of the week so that the price of the advertisement on that day is based on it.

Data

To fulfill the project goal, I analyzed the data of the MTA, namely the following features (C/A, UNIT, SCP, STATION, DATE, TIME, ENTRIES) . Additionally, I obtained MTA data via the MTA website. I used data from July 03, 2021 to October 02, 2021, totaling 2934629 observations.

Algorithms

I followed four steps in the project to analyze the data:
1.Start with a questions : I asked myself several questions in order to better understand the goal.
2.Collect and clean data : I collected the data and stored it in Database using SQL and SQLAlchemy . Additionally I cleaned the data using a several methods in pandas.
3.Exploratory data analysis (EDA) : I've explored and analyzed the data using several pandas methods and Numpy such as describe method , isna , info , groupby , to_datetime , apply , sort_values and many more.
4.Communicate results : After discovering and analyzing the data, I plotted the result using Matplotlib and Seaborn .

Communication

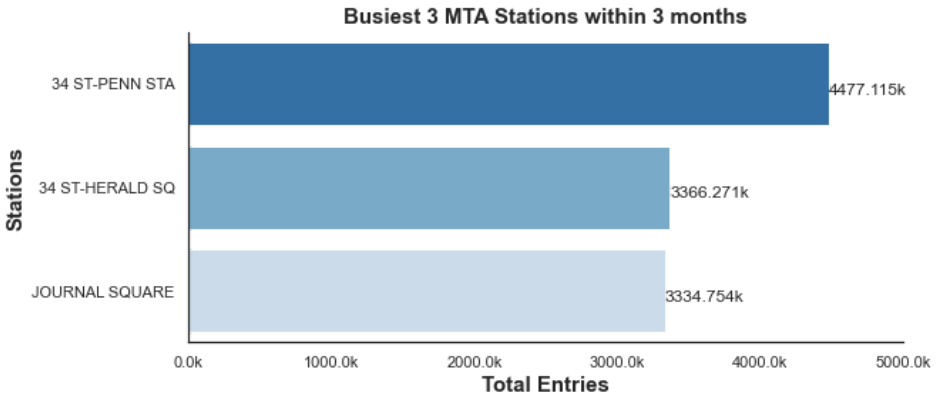

Figure 1

The outcome is depicted in Figure 1, where the bar chart displays the names of stations on the vertical axis and the number of entries within three months at these stations on the horizontal axis. Which displays the busiest 3 stations. This result after doing EDA on MTA data indicates that 24 ST-PENN STA is the most frequented station, followed by 34 ST-HERALD SQ, and then JOURNAL SQUARE. I recommend Moburst Advertising to place their billboards in these three stations.

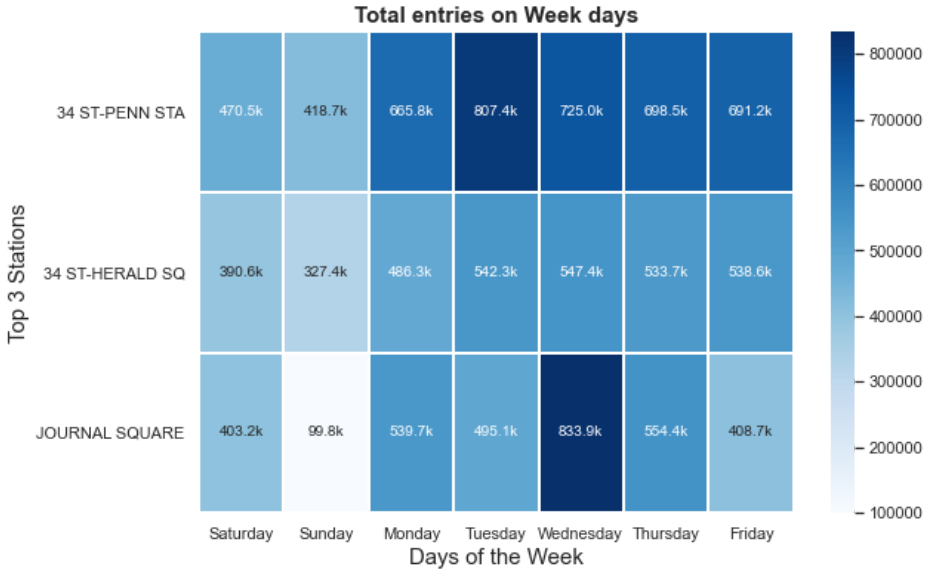

Figure 2

Figure 2 shows the total entries for each day of the week at the top three busiest stations.

Tools

- Jupyter & Python.
- SQLite & SQLAlchemy for data storing and retrieving .
- Pandas & Numpy for data manipulation.
- Matplotlib and Seaborn for plotting.