Determining the best locations for promotion and advertising

Abstract

The goal of this project is to use Exploratory Data Analysis to get the best locations that help in the success of the advertising campaigns that the client wants to do, so I analyzed the dataset of the Metropolitan Transportation Authority and found the busiest stations and located them so that the client can start his campaigns around those locations.

Data

The data set contains 379 stations, each with a number of turnstiles. It also contains the number of entrants and exits defined by date and time. Therefore, I chose three months and analyzed them to find out what stations include the largest number of people and what its times.

Algorithms

1- Data Cleaning started by deleting duplicate rows and check for any null values to delete it also.
2-Add new columns one of them for time operations which contain date and time, and the other to gather all the turnstile information to make it easier to access to.
3- By using the difference I get the exact number for the entries from the previous device.
4- Use Visualizations of data distributions and correlation to show the relationships in my results such as: relationship between date and number of entries.

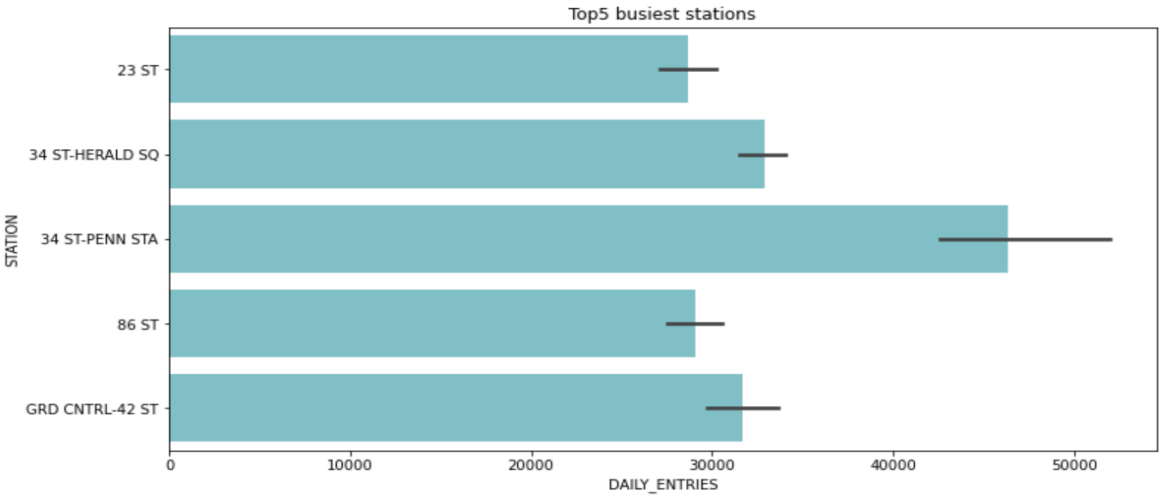
Tools

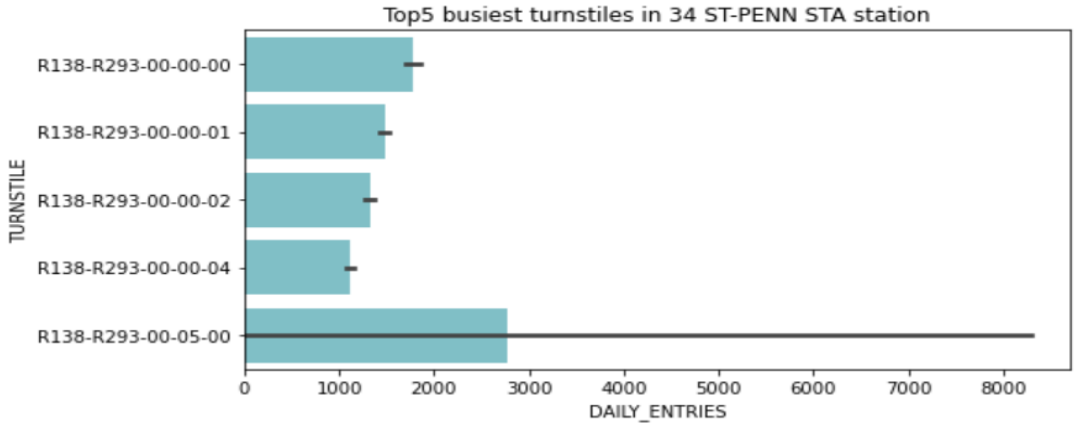
- **SQLAlchemy** for ingesting the raw data into a SQL database and querying from that database into Python.
- **Exploratory data analysis** in pandas**.**
- Python visualization libraries (such as **matplotlib** and **seaborn**)

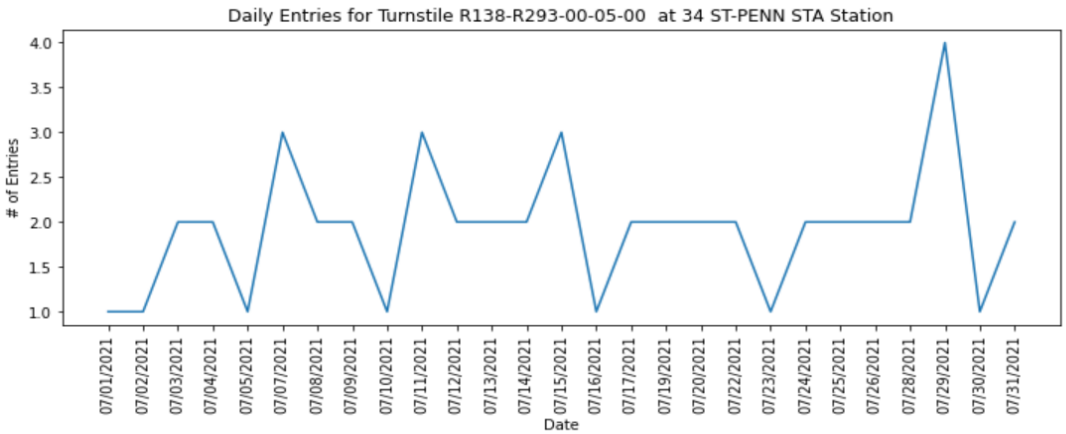
Communication

I will show the charts here

I determined what are the busiest stations, then I found the busiest turnstiles for that station, and finally I show the number of entries for that turnstile in July to be more specific







We note here the decrease in the number of entrants on weekend days.