AMJAD ALTHINYYAN

First Project Proposal

Overview

In this project, I will try to find the crowd times of people who use the MTA service in the days before and after Independence Day, to set up a movable booth to sell things related to the activities of this day in the crowded times. I will study the data for 6 years.

Goal

Find the crowd times of people who use the MTA service in the days before and after Independence Day.

Questions:

- What are the crowded times in MTA for the selected periods?
- Is there a difference in the crowded times for weekends or weekdays?
- Which station was the most crowded?

Dataset

The data is about Metropolitan Transportation Authority (MTA) Turnstile, and I obtained the data from http://web.mta.info/developers/turnstile.html. At the beginning, clean the data that tend to be noisy, incomplete, and inconsistent. Then, add the columns that I need when studying the data, after that, did the required analysis until get the expected results.

The data I have selected are 3 weeks for each year starting from 18 June 2016 until 6 July 2019 to get more accurate results. And I have avoided using the year 2020 because it is a special year due to covid19 and the results may not be accurate in this year.

In my case, I need to use station column, entry and also unit which will serve me in analyze the data.

I have added 2 more columns to get more accurate results and also to facilitate the analysis process, which are:

Name_of_day	The day's name extracted by the given date.
Entries_for_day	The entries registered per day

Tools

Technologies: Python, Jupyter notebook, SQL.
Libraries: Pandas, numpy, Matplotlib and Seaborn.

Conclusion

In conclusion, what I expected from studying this data is that the crowded times in the weekends in the days before and after Independence Day are more than weekdays, and also in weekdays