Ali Albayat

Backstory

A maintenance company ABC for the turnstile is planning to schedule a maintenance for New York City MTA (The Metropolitan Transportation Authority) stations in October 2019. Since NYC Subway is one of the world's largest underground lines, it is very important to do the maintenance with minimum effect on the station to prevent any inconvenient. Therefore, ABC company asked me as free lancer data analytics to determine the best 16 days in October and the best time of the day to do the maintenance knowing that ABC company will work 4 hours in each station but any time in the day.

My work in this project is to clean, explore, aggregate, and visualize the data of same month as previous years to determine the days where the stations have the lowest crowed as will as the time of the day. Therefore, from MTA turnstile data I will use the data of October 2018, 2017, and 2016 and my focus will be mainly on STATION, DATE, TIME, ENTRIES and EXITS columns from the data.

Question/need:

- What is the framing question of your analysis, or the purpose of the model/system you plan to build?

What are the best days and times to do maintenance for turnstile depending on previous data?

- Who benefits from exploring this question or building this model/system?

NYC MTA and ABC company.

Data Description:

- What dataset(s) do you plan to use, and how will you obtain the data?
  The data that will be used are:
  MTA turnstile data: October 2016, October 2017, and October 2018

- What is an individual sample/unit of analysis in this project? What characteristics/features do you expect to work with?
  Exploratory Data Analysis of the MTA turnstile data; clean, explore, aggregate, and visualize the data

Tools:

- How do you intend to meet the tools requirement of the project?
  Querying from that database into Python via SQLAlchemy.
  Exploratory data analysis in pandas.
  Python visualization libraries matplotlib and seaborn.