MTA Turnstile Data Analysis for Cleaning Services

By: Mohammed Abduldaim

Table of Content

BMS Janitorial Services

- BMS is a janitorial and cleaning services provider based in New York. The company has entered a contract with MTA to handle the subway station janitorial and cleaning tasks. So they have approached me to analyze the stations traffic to figure out the best time of day to perform the cleaning services for the stations.

Problem Description

- BMS is responsible for cleaning the stations however they only perform their services once a day. So, they would like to find the busiest time of day and start operating afterward.
- This would allow BMS to target the optimal time of day to perform and ensure client satisfaction which can assist their growth in the industry and becoming more reputable.

Data Set

- The MTA data set contain information pertaining the turnstile gate for every station. Spanning over a decade, the data records every entry and exist for each gate cumulatively every four hours. It also includes the gate id, where it's station date and time of records and more.
- The sample I've chosen is 3 month long (MAY – August) 2021, as it is the newest set of data.
- The data has 2513079 rows and 11 columns

Data Set

- C/A: Control Area
- UNIT: Remote Unit for a station
- SCP: Subunit Channel Position represents a specific address for a device
- STATION: Represents the station name the device is located at
- LINENAME: Represents all train lines that can be boarded at this station
- DIVISION: Represents the Line originally the station belonged to BMT, IRT, or IND
- DATE: Represents the date
- TIME: Represents the time for a scheduled audit event
- DESC: Represent the "REGULAR" scheduled audit event
- ENTRIES: The cumulative entry register value for a device
- EXITS: The cumulative exit register value for a device

Tools

Technologies:

- **Jupyter Notebook**
- Python
- SQL
- SQLite

Libraries:

- Pandas
- Numpy
- Seaborn
- Requests
- beautifulSoup