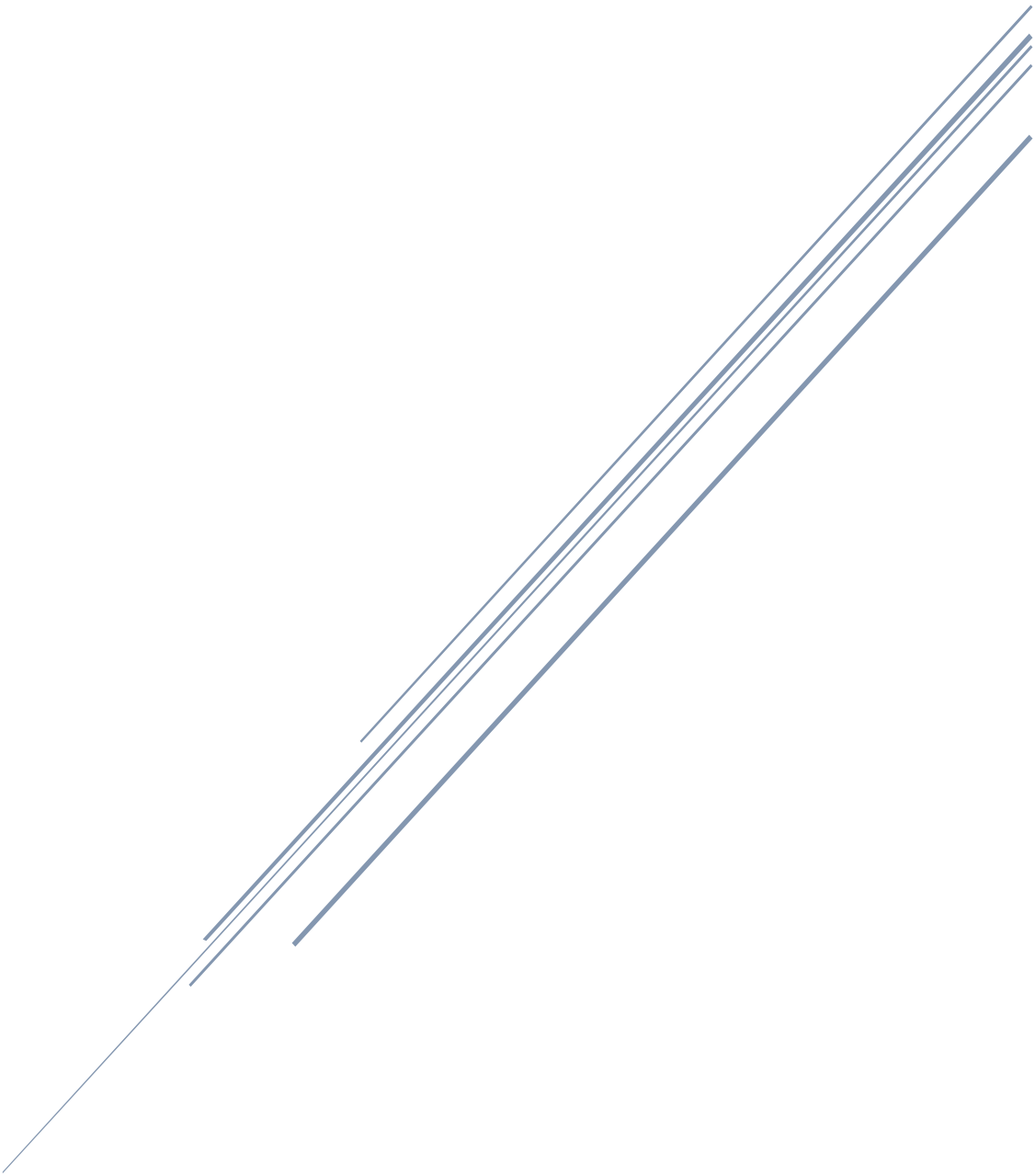
PROJECT PROPOSAL

MTA Turnstile Data



Feras Alyahya

Problem Statement:

For my project, I will be analyzing turnstile traffic data to discover patterns, such as peak hours and off-peak hours. I would also like to know which stations are busier than others. After analyzing my data, I will try to solve the problem by sending (assigning) more employees at the busiest stations and reducing the number of employees at the least crowded stations.

Question:

The questions of my proposal would be:

- What are the 10 most crowded stations?
- What is the busiest day? From one of the 10 most crowded stations, such as ST. George.
- What is the busiest time of the day using 5 stations? Ex. Before and after midday (AM / PM)

Data Description:

Field Name	Description
C/A	Control Area (A002)
UNIT	Remote Unit for a station (R051)
SCP	Subunit Channel Position represents an specific address for a device (02-00-00)
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 (MM-DD-YY)
TIME	Represents the time (hh:mm:ss) for a scheduled audit event
DESC	Represent the "REGULAR" scheduled audit event (Normally occurs every 4 hours)
ENTRIES	The comulative entry register value for a device
EXITS	The cumulative exit register value for a device

Tools:

- Python
- Pandas
- Matplotlib