Project Proposal

MTA Turnstile Data

Feras Alyahya

**Problem Statement:**

For my project, I will analyze turnstile traffic data to discover patterns, such as rush and off-peak hours. Also, I would like to find out stations, which are busier than others. After, analyzing my data, I will try to explain why we have the problem of high traffic rate and discover its reasons, which will help me pursue a solution that will help us, society, to overcome the problem or reduce it.

**Question:**

The questions of my proposal would be:

* What are the most and least crowded stations?
* What is the date that had the maximum number of entries?
* What is the busiest time of the day? Ex. Morning, afternoon, evening, night, and after midnight?

**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