Mitchell Brogan

Abstract:

The goal of this project was to use exploratory data analysis to discover what are the highest traffic metro stations in New York as well as their busiest hours. This information would prove useful to a street food vendor such as Lil’ Zeus Food Truck, as they would likely want to know the ideal time and place to station their food trucks for maximum sales to commuters. I worked with data from the publicly available MTA Turnstile data and used the sets spanning the first three months of 2022 to achieve results of the top 20 busiest stations as well as their busiest hours of operation.

Design:

The project was designed to retrieve the busiest metro station in New York. This was done by cross referencing the entries by the exits to determine how many people passed through the stations.

Data:

The dataset includes 2,728,513 rows of information as well as 11 columns relating to turnstile information, date/time and traffic in and out per turnstile.

Algorithms:

The only algorithms used were to create new columns of entries and exits and to combine them in such a way that would eliminate negative numbers.

Models:

The models were graphs representing the average daily foot traffic per station, and the average foot traffic per station at different times of the day. Each of these were selected to be on just the top 20 busiest stations.

Tools:

Pandas, Numpy, Matplotlib, Datetime, sqlite3, mpl\_toolkits, JupyterNotebook.

Communication:

Here is a link to my github repository containing the code as a ipynb file

<https://github.com/MitchellB9/MTA-Analysis>