Introduction:

In this project, I will be doing EDA for the MTA dataset in New York. I plane to find the best control aria to put the Food Truck beside it. And the best unit to put the announcement in it.

Backstory:

Adel is my client. He has a Food Truck, He wants to place it beside the best station and control aria. Also, he wants to put an announcement beside the best unit in that control Aria. He hired me to so.

Dataset:

I selected 3 months March, April and May 2020 of data obtained from <http://web.mta.info/developers/turnstile.html>.

Algorithms:

For data cleaning and pre-processing, I started by deleting the duplicate records and check if there are any null values then drop them. Also Then, I added new column that will have the exact number of entries at each turnstile at control area by calculating the difference from the previous device. I visualized the results.

Tools:

I used kaggle, Excel, sqlalchemy, Numpy and Pandas for data manipulation and Matplotlib and Seaborn for plotting

Conclusion:

At the end, I determined the best control aria for the food truck after EDA, and the best unit for the announcement.