Introduction:

In this project, I will be doing EDA for the MTA dataset in New York. I plane to find the best stations to put the Food Track beside it.

Backstory:

I'm a data analyst, My client 'Adel' told me that he has a food truck, and he want to put it in the best place in MTA. I will help him to do that.

Dataset:

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

Algorithms:

For data cleaning and pre-processing, I will start by deleting the duplicate records and check if there are any null values then drop them. Also, I will add new column that will have the date and time and change it is type to datetime to allow me to do time-based operations. Then, I will add 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 will visualize the results and try to find any correlations such as a relationship between the date and time and the number of entries.

Tools:

I will use Jupyter notebook, Excel and SQL

Conclusion:

At the end, I will have determined the best place for the food truck after EDA.