Abstract:

The goal of the project is to find how we can use the MTA data to improve ads at the stations on MTA subways...  
I used 3 months for this analysis, and I had some very interesting insights that may help companies to have better viewers ads on the subways or even to better negotiate with the MTA to put your ads.

Data:

the datasets from The Metropolitan Authority website ((<http://web.mta.info/developers/turnstile.html>) and they are publishing every week their turnstiles data it has 8 columns I’ve take 3 months data.

▪ C\_A: Control Area  
▪ UNIT: Remote Unit for a station  
▪ SCP: Subunit Channel Position represents a specific address for a device  
▪ STATION: Represents the station name the device is located at  
▪ LINE NAME: Represents all train lines that can be boarded at this station  
▪ DIVISION: Represents the Line originally the station belonged to  
▪ DATE: Represents the date in (MM-DD-YY) format  
▪ TIME: Represents the time (hh:mm:ss) for a scheduled audit event  
▪ DESC\_: Represent the "REGULAR" scheduled audit event  
▪ ENTRIES: The cumulative entry register value for a device  
▪ EXITS: The cumulative exit register value for a device

Algorithm:

find traffic for Top 5 stations

find traffic for Bottom 5 stations

find traffic for every station and Date

find traffic for every station and control area

Tools:

* NumPy and Pandas for data manipulation
* Seaborn for plotting
* Google cloud to create database in the cloud
* PowerPoint for presentation visualizations