**Abstract**

The goal of this project was to help Taylor Swift’s team to find a NYC subway station that she could perform live performance for 2-3 songs and meanwhile her team could film a scene for her documentary movie. I performed an exploratory data analysis on NYC MTA turnstile data(http://web.mta.info/developers/turnstile.html) to figure out which stations has the appropriate traffic and when is the best time for Taylor’s filming.

I first rank the stations by traffic density, then narrow down the candidate station list to 3 by setting ideal traffic range and looking at the station’s walkability on google photos. Then I perform a more detailed EDA on 3 stations specifically to select which timerange is best for this Taylor’s filming.

Final conclusion:

* Depends on Taylor’s schedule, the station choice could alter too:
* If July/August - Then Grand Central Station – 42 St is preferred, and filming time should be around 16:00-24:00 on weekdays to guarantee visitor flows
* If September – Then all three stations have relative high traffic, both weekdays 16:00-24:00 and weekends 20:00 – 24:00 could guarantee visitor flows

**Design**

First, I target the audience to be young adults who constitutes largest portion of Taylor’s fan. Then I researched for demographic of millennials in NYC and find out top 3 areas that has the largest share of millennials:

1. Downtown Manhattan

2. Garment District/ Murry Hill/ Kips Bay

3. Long Island City

Then I narrow down the candidate stations to 3 stations by setting ideal traffic flow and station walkability.

Then further EDA was performed on 3 stations data. Pandas functions was used to group data and calculate traffic, pivot table is also used for better presenting the data.

**Data**

The dataset contains 3 months of NYC MTA turnstiles data with 11 features for each. 4 features are used to identify a unique turnstile and 2 features could be used to get which lines this station could have access to. DATE and TIME features records the time whenever turnstile produce a record, ENTRIES and EXISTS gives the water-meter like counting on turnstile which could be used for computing daily/hourly traffic.

I used all the features except two: DIVISION and EXISTS

**Algorithms**

As this is an EDA project, no algorithms was used particularly. But below computation was performed for better visualization:

1. traffic was calculated by differencing entry numbers on turnstile

2. Traffic by station was calculated using group by function

3. Traffic by day of week by station was calculated using group by function

4. Traffic by hour by station was calculated using group by function

**Tools**

* Pandas for data manipulation
* Matplotlib for plotting

**Communication**

I’ll present the slides together with the visuals and other observations I have.