Abstract: The primary objective of this project is to decide where In-N-Out should build their new New York City location. In order to solve this problem, I used MTA turnstile data and built a dataset of coordinates of competitors. I used this data to identify which stations had the most exits during January of 2019-2019. Once I had the station totals, I created an interactive dashboard using Tableau.

Design: In-N-Out is looking to build a new restaurant in New York City and the purpose of this project is to give them insight for the location to open as well as to track the flow of potential customers throughout the week in order to make recommendations on staffing at the restaurant. Knowing the areas of New York that have the most foot traffic would allow In-N-Out to maximize potential customers. Pairing this information with the locations of competitors would allow In-N-Out to find areas with the most foot traffic that are in need of a fast food restaurant.

Data: In this analysis, I will be using the turnstile data from the month of January for the years 2016, 2017, 2018, 2019. I am sampling four different years to account for any changes over time, and are picking years that are pre-COVID because they will be a better representation of future foot traffic than numbers post-COVID. I have also built a dataset that includes the locations of various fast food restaurants that are close by the subway stations that we are honing in on in our analysis.

Algorithms: In terms of feature engineering, I added a few columns that are related to time. I first created the datetime column by adding the DATE and TIME columns together. From the column, I extracted the year value and the month value so that I could filter on just data from January, and so I could group the data by year. I also added a day of the week column in order to make the series. In order to get the time of day onto the time series, I built a function that would map the listed time with an associated fraction to represent how far along in the day it is. I created that tables that I would use in tableau by splitting the longitudes and latitudes provided from the table and mapping in the totals that we discovered.

Tools

- SQL for data loading
- NumPy and Pandas for data manipulation
- Matplotlib and Seaborn for initial visualization
- Tableau for geographic visualizations

Communication: These are the visuals that are in my presentation and I have added a video to demonstrate the Tableau visualization.