NYC Subway Traffic and Walk-up Vaccine Sites

By: Miho Hunter

ABSTRACT: The goal of this project was to use exploratory data analysis to identify the areas in New York City that have a low number of Covid-19 and flu vaccination sites relative to heavy commuter traffic.

DESIGN: The client is New York City's health department. They are interested in finding out if the current placement of the walk-up vaccination sites makes sense in relation to the volume of subway traffic within the city.

DATA: The primary data is the 12- week MTA turnstile data between 12/11/21 and 2/26/22. The data was merged with the station location data and also with the NYC walk-up vaccine site data in order to find the high traffic zip codes and the number of vaccine sites per zip code.

ALGORITHMS: I performed an exploratory data analysis of the MTA turnstile data. I cleaned, explored, aggregated, and visualized the data in order to address the client's problem. I first found the total traffic for each station over the 12 week period and then by zip code. I merged the data with the location of vaccine sites and found high traffic areas and their count of available walk-up vaccine sites.

TOOLS: SQL was used for data ingestion, storage, and extraction into Python via SQLAlchemy. Pandas was used for exploratory data analysis, and matlibplot, plotly express, and folium were used for visualization. Beautiful Soup was used to extract data from HTML. Geopy was used to get the zip code from latitude and longitude.

COMMUNICATION: All slides, visuals, codes, and data are available on my github site.

CONCLUSIONS: Manhattan had the highest traffic over the 12-week period, which was expected. All top 5 zip codes were in Manhattan. The walk-up vaccine sites are available throughout the city, but it might help to add more sites to the busy Manhattan area. The zip codes that have vaccination sites available do not necessarily have a high vaccination rate.