MTA Analysis - Part 1: Proposal

Our client Sephora is going to have a surprise sale to boost sales and reward to their customers, the duration will be 2 weeks. They want to place advertising boards inside of the train station where everyone can see the information when they pass by. Our clients want to know which several stops would be the best options if they can only set up at 5 locations due to the limited budget. Please also indicate what time slot, and which day of the week there's more traffic so that our clients can have some employees to send out coupons/samples for the non-selected stops during a specific time slot.

Question/need:

- What is the framing question of your analysis, or the purpose of the model/system you plan to build?

Purpose: For this model, we need to find out where to place the advertisement in an effective way for our client, so that they can boost sales by bringing in more customer through advertising

- Who benefits from exploring this question or building this model/system?

Our clients - more profit from regular customers and new customers.
Customers of the clients - receiving discounts, spending less money for the products.
Our company - gain more trust from our clients. (It's important for long term relationship)

Data Description:

- What dataset(s) do you plan to use, and how will you obtain the data?

MTA turnstile data, we can know the number of people getting on/off the train so that identify the amount of traffic. From the time people show up, we can know the busiest time.

- What is an individual sample/unit of analysis in this project? What characteristics/features do you expect to work with?

Features: DATE(extra the day of the week), TIME, Station, LINENAME, ENTRIESn, exit

- If modeling, what will you predict as your target?

Target - The increase in "total guests count" compared to the same data prior years as the sales may be seasonal. (Compare to "Increase in total profit during the sale", "total guests count" is a better metric as we just want to see if there's dramatic increase on orders)

Tools:

- How do you intend to meet the tools requirement of the project?

I will use SQLite to check the data to get an intuition, then using Python Pandas & Matplotlib & Seaborn to check out the distribution of data, as well as the relationships between features.

- Are you planning in advance to need or use additional tools beyond those required?

Possibly Postgres with PostGIS for demographic analysis.

MVP Goal:

- What would a minimum viable product (MVP) look like for this project?

Help our clients to find out 5 stops/exits for placing advertisements (Further details will be provided in next submission)