Aiman Chughtai Project Proposal- MTA Turnstile data

Question/need:

The framing question for my model is whether MTA turnstile data can be used to guide Starbucks in their opening of a hypothetical new curb-side location. If this model were to yield conclusive data to pinpoint a location, it would help the Starbucks franchise expand their customer reach.

Data Description:

To determine options for Starbucks' new location, I will be using MTA turnstile data to determine foot-traffic in New York City's most popular train stations. Starbucks is interested in opening their newest location somewhere that is densely populated by their target customers: young people (ages ~25-37) from wealthy households (with median salary of ~65K or above), with an undergrad degree or beyond . This demographic data can be obtained using The U.S Census Bureau data.

Specifically, the features I expect to work with are MTA turnstile entry/exit data in peak season months of January, April, July, and October in 2019, during their peak hours of 9am-2pm. The entry/exit data will allow me to quantify foot-traffic in order to determine the popularity of a particular station.

The target variable for this model is between 100,000 - 200,000 daily turnstile entries and/or exits. According to MTA data on average weekday ridership (source), the most popular subway station is Times Square-42nd St, which had a daily ridership of about 155,000.

Tools:

In order to conduct this analysis I plan on using python, SQLAlchemy, pandas, matplotlib and seaborn.

MVP Goal:

An MVP for this project would be a graph of stations with the greatest foot traffic.