Restaurant Location Selection in New York City

IBM Data Science Professional Certificate Capstone

Charlie Kelley

1/23/2018

Table of Contents

Business Understanding

PROBLEM

I am opening a restaurant in New York City and want to find the most profitable neighborhood location. The restaurant will be fast casual sandwiches and the target audience is younger consumers who live in the neighborhood or will take public transit to get there.

My initial hypotheses are that:

- Selecting a neighborhood location that is around a high concentration of 25-34 year olds will help maximize revenue

- Selecting a neighborhood location that is around a high concentration of public transit stops will help maximize revenue

- Selecting a neighborhood location that is around a lower concentration of restaurants will help maximize revenue

OBJECTIVE

My objective is to find out what neighborhoods in NYC best meet the above criteria to help pick the optimal neighborhood location.

Data Understanding

I will use four datasets to help find the optimal location.

NYC NEIGHBORHOOD SHAPEFILE

This file will be used to group findings by neighborhood and display as a map. It contains geolocation data for all 195 neighborhoods in NYC.

https://data.cityofnewyork.us/City-Government/Neighborhood-Tabulation-Areas/cpf4-rkhq

NYC NTA DEMOGRAPHICS

This file contains the population of each neighborhood broken down by age group. It will be used to find population density for my desired age groups per neighborhood.

https://data.cityofnewyork.us/City-Government/Demographics-and-profiles-at-the-Neighborhood-Tabu/hyuz-tij8

MTA DATA

This file contains locations of NYC transit stops. I will use this data to find out how many stops are located in each neighborhood.

https://datamine.mta.info/

FOURSQUARE API

From this API I will find out how many restaurants are located in each neighborhood.

https://developer.foursquare.com