Data Science Capstone Project week 1

Battle of the Neighborhoods

By Teresa Lau

Introduction/Business Problem

Background

There is a growing trend of plant-based diets in USA. New York City is one of the top cities in USA in terms of the number of plant-based businesses in the city . This presents a good opportunity for vegetarian and vegan restaurant's owner to expand their business.

Problem Description

This project will focus on finding the best neighborhood(s) in New York city to open up a new vegetarian/vegan restaurant.

We will collect information on popular venues from all neighborhoods, as well as demographics and economic information of residents in each neighborhood (population, age and income) to determine neighborhoods that are suitable for opening vegetarian/vegan restaurants
We will also analyze data of successful vegetarian/vegan restaurants in the city to get insights on important factors to consider when opening a new vegetarian/vegan restaurant .

Target Audience

The target audience is for someone who wants to open a vegetarian/vegan restaurant in New York City. It is also useful for company producing plant-based product to target their marketing to neighborhoods that has high demand for vegetarian food. It will be equally useful for other restaurants in a neighborhood to understand if they should offer more vegetarian-friendly dishes in their menu in order to increase their revenues.

Data Sources

Neighborhood Location Data

We will be using this data from NYU Spatial Data Repository to determine the neighborhoods and location that we are interested in.
https://geo.nyu.edu/catalog/nyu_2451_34572
Below is a sample of how the file looks like

Borough Type: String	Neighborhood Type: String	Latitude Type: String	Longitude Type: String
Manhattan	Marble Hill	40.87655077879964	-73.91065965862981
Manhattan	Chinatown	40.71561842231432	-73.99427936255978
Manhattan	Washington Heights	40.85190252555305	-73.93690027985234
Manhattan	Inwood	40.86768396449915	-73.92121042203897
Manhattan	Hamilton Heights	40.823604284811935	-73.94968791883366
Manhattan	Manhattanville	40.8169344294978	-73.9573853935188
Manhattan	Central Harlem	40.81597606742414	-73.94321112603905

Venue Data

Based on each neighborhood location, we will use Four Square crowdsource data to get
- Popular Venues by Neighborhood using Explore API

    o `GET https://api.foursquare.com/v2/venues/explore`
    o `Parameters: Section=food`

- Vegetarian / Vegan Venues by Neighborhood using Search API

    o `GET https://api.foursquare.com/v2/venues/search`
    o `Parameters:Category=` '4bf58dd8d48988d1d3941735' *# Vegatarian*

- For each. Vegetarian / Vegan Venues, we use the Venues Details endpoint to get Rating and Price info

    o `GET https://api.foursquare.com/v2/venues/VENUE_ID`

Neighborhood Demographics & Economic Data

Using NYC planning population fact finder, we will get information for each neighborhood on population, median age, median income etc. This data comes from the 2013-2017 American Community Survey (ACS)

https://popfactfinder.planning.nyc.gov/profile/12005/demographic

Using the Data Sources

Using the Four Square Venue data, we can get the list of popular venues for each neighborhood and use that as input to the machine learning model K-means clustering for unsupervised data and build clusters of similar neighborhood offering similar categories of restaurants.

Using rating information, we can locate where the most successful vegetarian/vegan restaurants are currently, and we will try to find neighborhoods that are in same clusters as them. This together, with demographics and economic info of neighborhoods will help us make our decision. The price range data will then be used to determine a price point for the new restaurant.