Exploring New York Neighborhoods - to open a Vietnamese Restaurant

As a part of the IBM Data Science professional program Capstone Project, we worked on the real datasets to get an experience of what a data scientist goes through in real life. Main objectives of this project were to define a business problem, look for data in the web and, use Foursquare location data to compare different neighborhoods of New York city to figure out which neighborhood is suitable for starting a new restaurant business. In this project, we will go through all the process in a step by step manner from problem designing, data preparation to final analysis and finally will provide a conclusion that can be leveraged by the business stakeholders to make their decisions.

1. Description of the Business Problem & Discussion of the Background :

Problem Statement : Prospects of opening a Vietnamese Restaurant in New York city, US.

New York, the finance center of the world, is the most populous US city. Its diversity is reflected in New York's ethnic neighborhoods such as Marble Hill, Chinatown, Washington Heights, Hamilton Heights, Inwood. One of the most immigrant-friendly cities in North America, it is one of the best places to start a Vietnamese restaurant.

In this project we will go through step by step process to make a decision whether it is a good idea to open a Vietnamese restaurant. We analyze the neighborhoods in New York to identify the most profitable area since the success of the restaurant depends on the people and ambience. Since we already know that New York shelter a great number of Vietnamese in US, it is a good idea to start the restaurant here, but we just need to make sure whether it is a profitable idea or not. If so, where we can place it, so it yields more profit to the owner.

Target Audience

Who will be more interested in this project ? What type of clients or a group of people would be benefitted ?

1. Business personnel who wants to invest or open a Vietnamese restaurant in New York city. This analysis will be a comprehensive guide to start or expand restaurants targeting the Vietnamese crowd.
2. Freelancer who loves to have their own restaurant as a side business. This analysis will give an idea, how beneficial it is to open a restaurant and what are the pros and cons of this business.
3. Vietnamese crowd who wants to find neighborhoods with lots of option for Vietnamese restaurants.
4. Business Analyst or Data Scientists, who wish to analyze the neighborhoods of New York using Exploratory Data Analysis and other statistical & machine learning techniques to obtain all the necessary data, perform some operations on it and, finally be able to tell a story out of it.

2. Data acquisition and cleaning:

2.1 Data Sources

a) I'm using **https://cocl.us/new_york_dataset** to get all the information about the boroughs and the neighborhoods that exist in each borough as well as the the latitude and longitude coordinates of each neighborhood in New York.

b) To get information about the distribution of population by their ethnicity I'm using "Demographics of New York city" **https://en.wikipedia.org/wiki/Demographics_of_New_York_City** wiki page. Using this page I'm going to identify the neighborhoods which are densely populated with Vietnameses as it might be helpful in identifying the suitable neighborhood to open a new Vietnamese restaurant.

c) To get location and other information about various venues in New York city I'm using Foursquare's explore API. Using the Foursquare's explore API (which gives venues recommendations), I'm fetching details about the venues up present in New York and collected their names, categories and locations (latitude and longitude).

From Foursquare API (https://developer.foursquare.com/docs), I retrieved the following for each venue:

- Name: The name of the venue.
- Category: The category type as defined by the API.
- Latitude: The latitude value of the venue.
- Longitude: The longitude value of the venue.

2.2 Data Cleaning

a) Scrap the distribution of population from Wikipedia

Another factor that can help us in deciding which neighborhood would be best option to open a restaurant is, the distribution of population based on the ethnic diversity for each neighborhood. As

this helps us in identifying the neighborhoods which are densely populated with Vietnamese crowd since that neighborhood would be an ideal place to open a Vietnamese restaurant.

Scraped the following Wikipedia page, "Demographics of New York city" in order to obtain the data about the New York & the Neighborhoods in it. Compared to all the neighborhoods in New York below given neighborhoods only had considerable amount of Vietnamese crowd. We are examining those neighborhood's population to identify the densely populated neighborhoods with Vietnamese population.

b) Get location data using Foursquare

Foursquare API is very useful online application used my many developers & other applications like Uber etc. In this project I have used it to retrieve information about the places present in the neighborhoods of New York. The API returns a JSON file and we need to turn that into a data-frame. Here I've chosen 100 popular spots for each neighborhood within a radius of 1km.