

Capstone Project

The Battle of Neighborhoods (Week 2)

Opening a New Thai Restaurant in New York City

By: Toh Khai Chun

June 2021

Introduction

New York City (NYC) is the most populous city in the United States and is the largest metropolitan area in the world. It is the most ethnically diverse, religiously varied, commercially driven, famously congested urban centre in the country. The diverse and inclusive city provides a business friendly environment which allows different players from variety fields to develop their markets in New York City. The top industries are finance, trade, healthcare, real estate, media, publishing, manufacturing and information technology.

It represents that New York City is a globally competitive working city. Therefore, any new business venture and expansion should be planned with a secure and proper analysis. An insight with a better understanding will target market strategically in order to reduce potential risk and to have a stable return.

Business Problem

Restaurants are businesses which are providing cooked food and not raw goods to customers in return of money. Many cuisines of New York City are sourced various ethnic groups in this multicultural city. Almost all ethnic cuisines are well represented in New York City.

The aim of this capstone project is to analyse, explore and select the best location for opening a new Thai restaurant. So, this report is targeted specifically to investors or caterers who intended to start new business venture by run a Thai restaurant. The first move is always important so a choice of location is a critical consideration in the plan.

Target Audience of This Report

This project is useful for investors or caterers who are looking for an opportunity in opening a Thai Restaurant in New York City. Eater NY reported that more than one year after New York's first indoor dining shutdown, restaurants and bars continue to close their doors. At least 1,000 have closed since March 2020 due to the economic downturn caused by the coronavirus pandemic. The number of restaurant job losses to date is 122,400 or -48% from March 2020 to January 2021. Thus, placing business in a strategic location is a key ingredient in a business's success by attracting a customer base.

Data:

Based on the problem defined above, we will need the following data:

- List of the boroughs and neighbourhoods in New York City. It is used to understand the scope and size of the New York City.
- Latitude and longitude of each neighbourhood. It is purposed to get venue data and also to plot the map.
- Population density of each borough. It is to have a better understanding the future market demand.
- Number of Thai restaurants in each neighbourhood or each borough.
- Venue data related Thai restaurants only. They will be performed for the clustering on the neighbourhoods.

Following data sources will be extracted to get the required information:

1. Scrape borough data from https://en.wikipedia.org/wiki/New_York_City#Boroughs.

- The Wikipedia table's data will be used to determine the population and density of the New York City's five boroughs.

2. Download data from https://geo.nyu.edu/catalog/nyu_2451_34572.

- It will determine geographical coordinates of neighbourhoods:

Latitude : Latitude for Neighbourhood

Longitude : Longitude for Neighbourhood

Neighbourhood: Name of Neighbourhood

Borough: Name of Borough

3. Foursquare API

- It will be used to get locational data and to find restaurant information, particularly in Thai restaurants:

Neighbourhood : Name of the Neighbourhood

Neighbourhood Latitude : Latitude of the Neighbourhood

Neighbourhood Longitude : Longitude of the Neighbourhood

Venue : Name of the Venue

Venue Latitude : Latitude of Venue

Venue Longitude : Longitude of Venue

Venue Category : Category of Venue

After that, it will be worked as data cleaning, data wrangling, machine learning (K-Cluster), and map visualization (Folium).

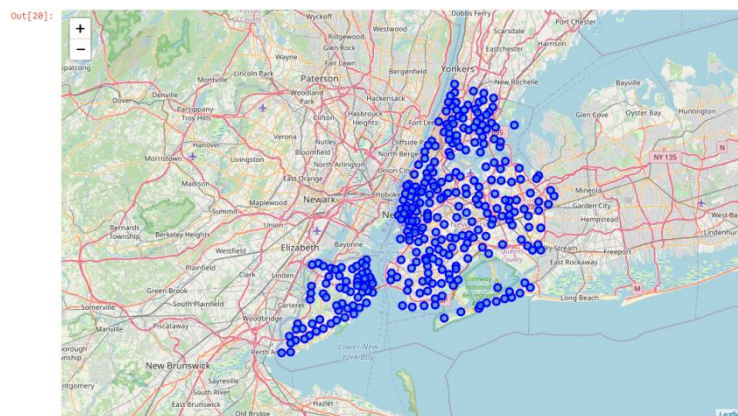
Methodology

First of all, we need to detect the population and density of boroughs in NYC so the list is available in https://en.wikipedia.org/wiki/New_York_City#Boroughs. We will request the Beautiful Soup package for web scraping and parsing data table from the Wikipedia page.

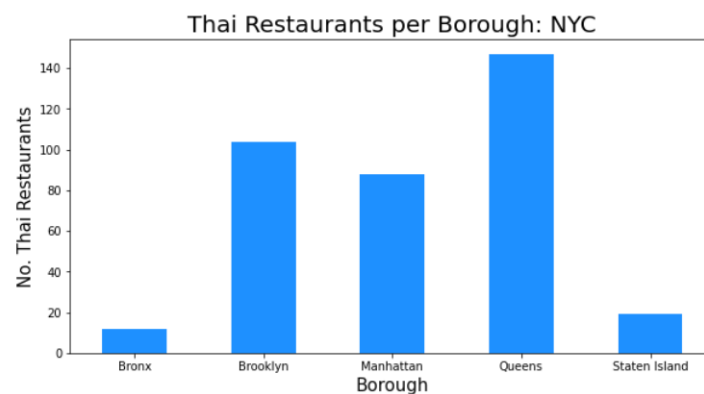
Out[5]:

	Borough	County	Estimate (2019)	billions(2012 US\$)	square miles	squarekm	persons /mi2	persons /km2
0	The Bronx	Bronx	1,418,207	42.695	42.10	109.04	33,867	13,006
1	Brooklyn	Kings	2,559,903	91.559	70.82	183.42	36,147	13,957
2	Manhattan	New York	1,628,706	600.244	22.83	59.13	71,341	27,544
3	Queens	Queens	2,253,858	93.310	108.53	281.09	20,767	8,018
4	Staten Island	Richmond	476,143	14.514	58.37	151.18	8,157	3,150

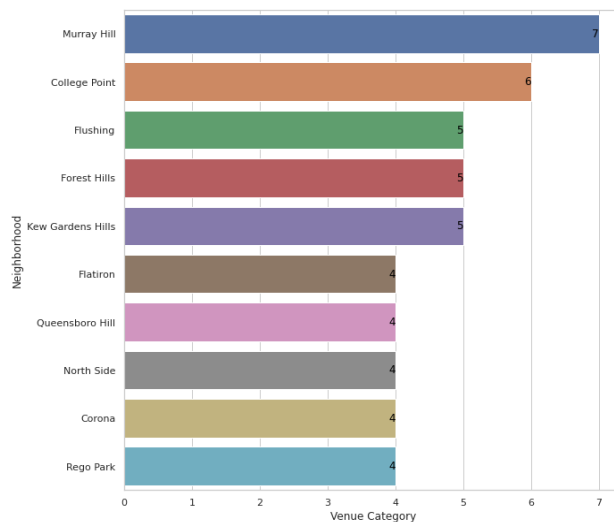
Besides that, we also need to get the NYC geographic coordinate data so `newyork_data.json` will be loaded to get data about latitude and longitude of each neighbourhood. The data will be transferred into pandas frame and will be visualised into a map by using Folium package.



Next step, we will explore restaurant across neighborhoods of NYC within a radius of 10000 meters. We have to register a Foursquare Developer Account so that we can obtain the Foursquare ID and Foursquare secure key. Venue data will be returned after an API call has been made. Venue data will be extracted in venue name, venue category, venue latitude, and venue longitude. With the data, we can check the number of Thai restaurants for each neighborhood and each borough.



As per table above, Queens has the largest number of Thai Restaurants.



As per table above, College Point of the Queens and Murray Hill of the Manhattan are the neighborhoods in all of New York City with the most Thai Restaurants.

Besides that, we also can examine the total of each neighborhood and the number of unique categories from the venue data. Then, we will analyze each neighborhood by grouping the rows of neighborhoods and taking mean of the occurrence frequency of each venue category. Other than that, a data will be prepared for the purpose of analyzing clustering. “Thai Restaurants” will be filtered out from venue categories for neighborhoods’ data.

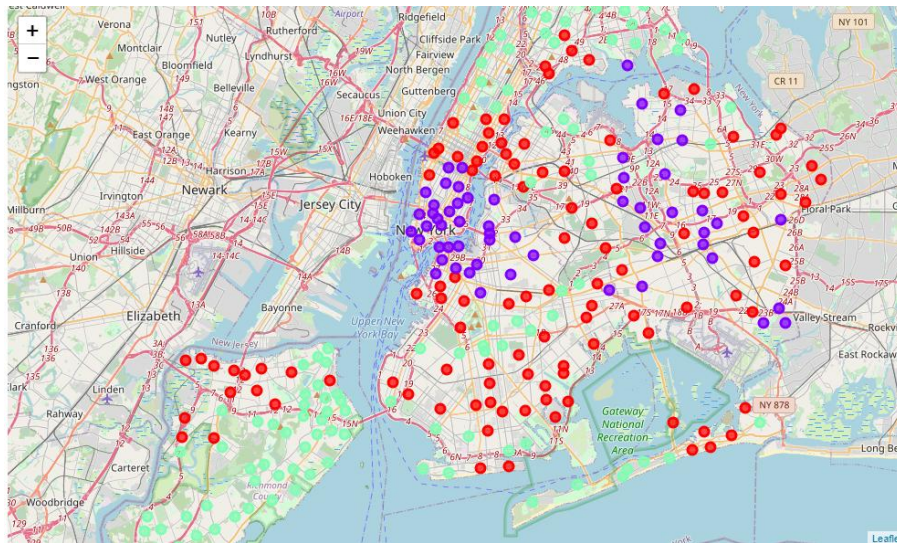
In final step, we have to focus the promised area by creating location clusters so k-mean clustering will be performed. K-mean clustering is one of the simplest and popular unsupervised machine learning algorithms. It identifies k number of centroids, and then allocates every data point to the nearest cluster. The neighborhoods will be clustered into 3 clusters based on their occurrence frequency so that it is able to find out the lowest or highest number of Thai restaurants for each cluster. After that, it helps to answer the best location of opening a new Thai restaurant.

Result

The result of k-mean clustering shows that we can classify 3 clusters based on occurrence frequency of Thai Restaurants:

- Cluster 0: Neighborhoods with moderate number of Thai restaurants
- Cluster 1: Neighborhoods with high density of Thai restaurants
- Cluster 2: Neighborhoods with the rarest Thai restaurants

The result is visualized in the map below with cluster 0 in red colour, cluster 1 in purple colour, cluster 2 in green colour.



Discussion and Recommendation

As per result of the displayed map above, we can see that most of the Thai restaurants are concentrated in the central area and east area of New York City, with the highest number in cluster 1 and moderate number in cluster 0. Besides that, cluster 2 has very low number to almost none of Thai restaurants in those neighborhoods. It represents a huge opportunity and great potential areas to open new Thai restaurants as there is very little to no competition from existing restaurants. Meanwhile, Thai restaurants in cluster 1 are suffering from intense competition due to oversupply of Thai restaurants and the market is saturated.

Therefore, this project will recommend stakeholders to open new Thai Restaurant in The Bronx because it is one of the boroughs in cluster 2 which has least supply of Thai food. Other than that, The Bronx is the borough that has least number of Thai restaurants and the third most densely populated county in the New York City (NYC).

Conclusion

This project is conducted with limited data so it may lead a wrong or right result. If a good amount of data then will come up a better result.

The purpose of this project is to identify neighborhoods of New York City with a low presence of Thai restaurants by specifying required data, extracting and preparing data, performing clustering analysis based on their similarity, and providing best option to investors or caterers.

First of all, we analyzed the population amount of boroughs in the NYC. After that, we proceed to find out the Thai restaurant number for each neighborhood and each borough. Then, identifying Thai restaurant presence in each neighborhood from Foursquare data we have clustered.

The final decision about the best location for opening a new Thai Restaurant will be decided by investors or caterers after the option has been recommended in order to help them in avoiding from overcrowding areas of Thai restaurants and identifying great opportunity on high potential areas.