Coursera
IBM

The Battle of Neighborhoods

What is the best place where can I start my restaurant business in New York?

Author
Filali Taha

January 2, 2021

Contents

Chapter 1

Introduction

1.1 Background

New York City is the largest city by population in the United States, located in the state of New York. New York's population is similar to London in the United Kingdom with over 8 million people currently living in it, and over 22 million people live in the bigger New York metropolitan area. It is in the south end of the state of New York, which is in the northeastern United States. It is the financial capital of the US since it is home to the nation's stock market, Wall Street, and the One World Trade Center. A leading global city; New York exerts culture, media and capital internationally, as well as attracting great numbers of international travelers. It is also the home of the United Nations Headquarters.

1.2 Problem description

The cuisine of New York City comprises many cuisines belonging to various ethnic groups that have entered the United States through the city. Almost all ethnic cuisines are well represented in New York, both within and outside the various ethnic neighborhoods.[1] New York was also the founding city of New York Restaurant Week which has spread around the world due to the discounted prices that such a deal offers.[2] In New York there are over 12,000 bodegas, delis, and groceries, and many among them are open 24 hours a day, 7 days a week.

Food associated with or popularized in New York :

- Ashkenazi Jewish cuisine : celery soda, New York-style pastrami..
- Italian-American cuisine : Espresso, Cappuccino..
- Chino-Latino cuisine : fried rice, fried pork chop..
- Street food : arepas, calzones..

An X company needs to find the correct location to start.

The objective of this study is to locate and recommend to anyone who wants to open a new restaurant in New York City.

Chapter 2

The Data

In this project we will analyze the neighborhoods of the city New York City.

2.1 Data of Boroughs and neighborhoods in New York City

This data file includes all needed information to find out more about hosts, geographical availability, necessary metrics to make predictions and draw conclusions.

Link : https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data. So we took just $(\text{neighbourhood}_group, neighborhood, latitude, longitude) from the dataset$.

	neighbourhood_group	neighborhood	latitude	longitude
0	Brooklyn	Kensington	40.64749	-73.97237
1	Manhattan	Midtown	40.75362	-73.98377
2	Manhattan	Harlem	40.80902	-73.94190
3	Brooklyn	Clinton Hill	40.68514	-73.95976
4	Manhattan	East Harlem	40.79851	-73.94399

2.2 Restaurants in New York City:

This data will cover different restaurants that already exists in New York City and different food markets.

Analysis of localization of food markets permits to have an idea of location that will be near to fresh ingredients suppliers. This data will be provided by Foursquare API. Foursquare API permits to provide venues information for each neighborhood in New York City.

Table below gives some restaurant in New York City:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Restaurant type
4	Kensington	40.64749	-73.97237	Werkstatt	40.645252	-73.970341	Austrian Restaurant
6	Kensington	40.64749	-73.97237	La Loba Cantina	40.645657	-73.972814	Mexican Restaurant
7	Kensington	40.64749	-73.97237	Thai Farm Kitchen	40.644148	-73.976047	Thai Restaurant
11	Kensington	40.64749	-73.97237	Suzy's Roti Parlour	40.646618	-73.970269	Caribbean Restaurant
12	Kensington	40.64749	-73.97237	AM Thai Bistro	40.646810	-73.970162	Thai Restaurant

Chapter 3

Methodology

3.1 Business understanding:

The main goal of this project is to get the locations in NYC that will be suitable for starting a new restaurant business.

3.2 Analytic Approach:

NYC has 5 main counties and 221 neighborhood. In this project neighborhoods will be clustered following exploratory data that will be discovered in the next part.

3.3 Exploratory Data Analysis:

3.3.1 Data 1 : NYC counties and coordinates

This data file includes all needed information to find out more about hosts, geographi-calavailability, necessary metrics to make predictions and draw conclusions.

Link : https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data. This data is also cleaned to extract the necessary of information and processe like this so we took just (neighbourhoodgroup,neighborhood,latitude,longitude) from the data set.

	neighbourhood_group	neighborhood	latitude	longitude
0	Brooklyn	Kensington	40.64749	-73.97237
1	Manhattan	Midtown	40.75362	-73.98377
2	Manhattan	Harlem	40.80902	-73.94190
3	Brooklyn	Clinton Hill	40.68514	-73.95976
4	Manhattan	East Harlem	40.79851	-73.94399

3.3.2 Data 2: NYC's restaurants

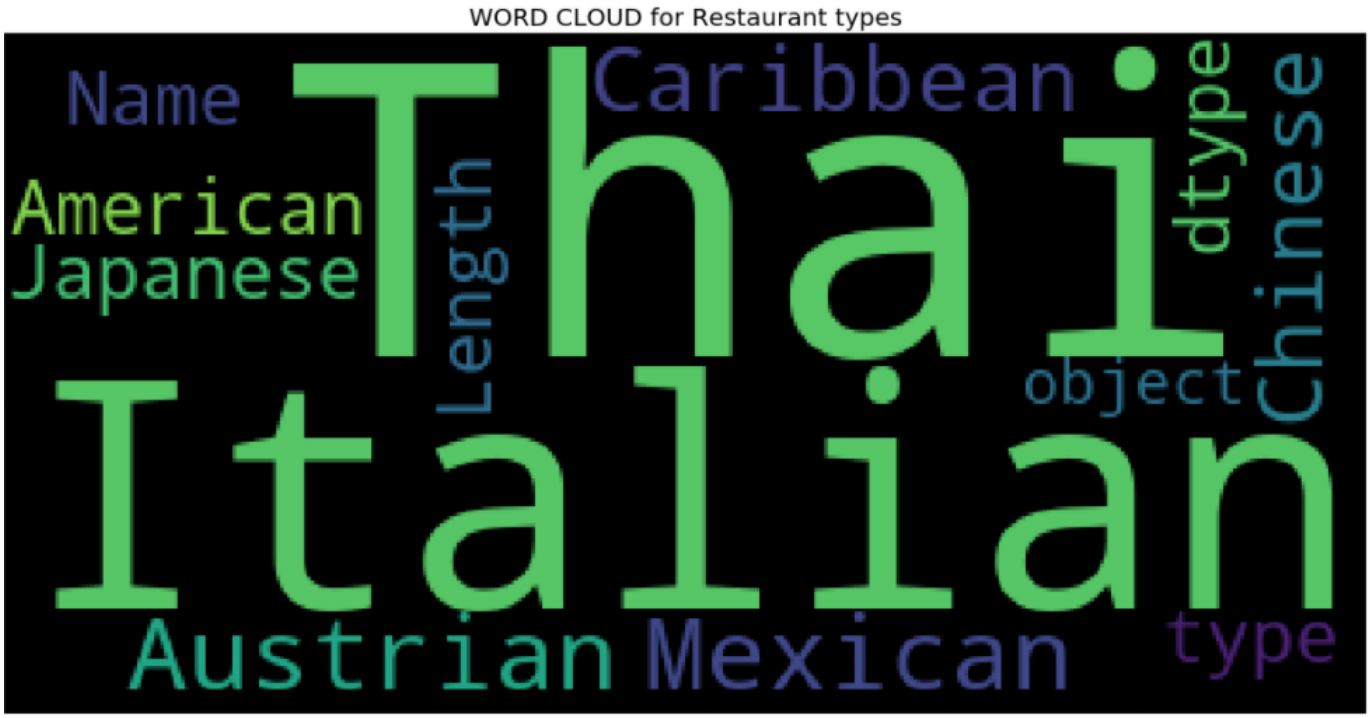
Data about restaurant in NYC is generated by Foursquare API by providing different restaurants that are near to a neighborhood or located in a certain county.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Restaurant type
4	Kensington	40.64749	-73.97237	Werkstatt	40.645252	-73.970341	Austrian Restaurant
6	Kensington	40.64749	-73.97237	La Loba Cantina	40.645657	-73.972814	Mexican Restaurant
7	Kensington	40.64749	-73.97237	Thai Farm Kitchen	40.644148	-73.976047	Thai Restaurant
11	Kensington	40.64749	-73.97237	Suzy's Roti Parlour	40.646618	-73.970269	Caribbean Restaurant
12	Kensington	40.64749	-73.97237	AM Thai Bistro	40.646810	-73.970162	Thai Restaurant

For example, that table above represent some restaurants that are near to Kensington neighborhood in NYC. This table contains information about name and localization of restaurant and restaurant category (type of restaurant e.g. Mexican).

This data will help us to understand what is the type of restauration that have the domination on each borough. For this task, we will use word cloud. Word cloud is a good visual tool that permit to have a global idea of domination of a category over others. Word cloud is an image composed of words used in a particular text or subject, in which the size of each word indicates its frequency or importance

For example, we built word cloud for all categories of restaurants in NYC and it gives:



3.4 Machine learning uses:

For segmenting neighborhood in NYC based on existed restaurant and localization of restaurants, and finding recommendation for a good start of restaurant business, clustering algorithm will be used especially K-mean method. This method is used as unsupervised algorithm. It doesn't need previous recommendations to build a model. K-mean method is good for segmentation. It divides the data into clusters without any cluster-internal structures or label.

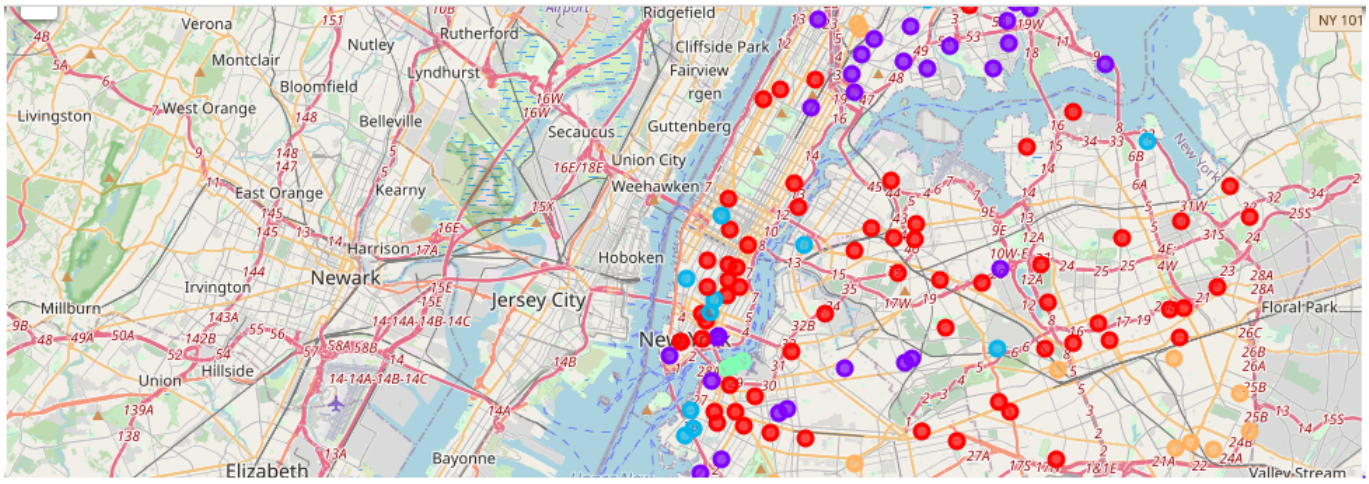
Chapter 4

Results

From venues data, dataset is filtered and used only data where venues are restaurant as we focused only on restaurant business.

4.1 Neighborhood K-Means clustering based on mean occurrence of restaurant type

To cluster the neighborhoods into five clusters we used K-Means clustering Algorithm. K-Mean clustering aims to partition n observation into k clusters in which each observation belongs to the cluster with the nearest mean. It uses iterative refinement approach. In this project, 5 clusters are chosen. The map below, that is constructed using Geopy and Folium libraries, shows clustering of neighborhoods in Paris into five clusters.



Each Cluster is marked by colored circle markers. It is about neighborhoods where restauration type is dominated and we can start here a restaurant business.

Chapter 5

Conclusion

As we can see from data and map of clustered neighborhoods, it is more preferable to start restaurant business neighborhoods in periphery of NYC.

In this project, we have used data available in the net about counties and already existed restaurant in NYC. The K-mean algorithm, that is a clustering algorithm for unsupervised learning, is used to show us what are neighborhoods that are good places where we can start restaurant business.