

Battle of the Neighborhoods – Where to build a BBQ Restaurant in Manhattan, NY?

Introduction

Manhattan is the most densely populated of New York City's 5 boroughs. Among the world's major commercial, financial, and cultural centers, it is the heart of "The Big Apple." As it is highly developed city so cost of doing business is also of the highest. Restaurants are one of the most frequently started small businesses yet have one of the highest failure rates. In order to survive, a powerful strategic advantage is highly necessary: a sound business plan and feasibility study prior to opening. A business plan that precisely defines your business, identifies its goals, and serves as your firm's resume.

The basic components include a current and pro forma balance sheet, an income statement, and a cash flow analysis. It helps to allocate resources properly, handle unforeseen complications, and make overall good business decisions. Because it provides specific and organized information about your company and how you will repay borrowed money, a good business plan is a crucial part of any loan application. Additionally, it informs personnel, suppliers, and others about your operations and goals.

Of course, as with any business decision, opening a new barbecue restaurant requires serious consideration and is a lot more complicated than it may seem. Particularly, the location of the barbecue restaurant is one of the most important decisions that will determine whether the restaurant will be a success or a failure.

Business Problem

The objective of this project is to analyze and select the best locations within the Manhattan borough to open a barbecue restaurant. Using data science methodology and machine learning techniques (we will use k-means clustering), this project aims to provide solutions to answer the business question: In the city of Manhattan, New York, if a property developer is looking to open a new barbecue restaurant, where would you recommend that they open it? This project is timely as the novel coronavirus pandemic has forced many local restaurants have been forced to adapt by using new methods of food delivery and take out service.

Data

To solve this problem, we will need the following data:

- List of neighborhoods in Manhattan, New York. The following dataframe was produced from the publicly available list of all 5 New York City's boroughs and 306 neighborhoods: https://geo.nyu.edu/catalog/nyu_2451_34572. This defines the scope of the project which is confined to Manhattan, the most densely populated of New York City's 5 boroughs and latitude and longitude coordinates of those neighborhoods. This is required to plot a Folium map of neighborhoods and venues superimposed on it.

	Borough	Neighborhood	Latitude	Longitude
0	Manhattan	Marble Hill	40.876551	-73.910660
1	Manhattan	Chinatown	40.715618	-73.994279
2	Manhattan	Washington Heights	40.851903	-73.936900
3	Manhattan	Inwood	40.867684	-73.921210
4	Manhattan	Hamilton Heights	40.823604	-73.949688

- Venue data, particularly data related to the number of barbecue restaurants in each neighborhood (here is a few rows of that dataframe). We will later use this data frame to cluster the neighborhoods.

	# of Venues
Neighborhood	
Battery Park City	15
Carnegie Hill	10
Central Harlem	7
Chelsea	23
Chinatown	18
Civic Center	21
Clinton	28
East Harlem	5
East Village	18

- A data frame that shows all 698 barbecue (or similar) restaurants in Manhattan grouped by neighborhood. The full data frame can be found in this project's corresponding Jupyter Notebook, but below are a few rows of said data frame.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.910660	Sang's Barbecue Co	40.868831	-73.902634	BBQ Joint
1	Chinatown	40.715618	-73.994279	友情客串bbq	40.717550	-73.995060	Restaurant
2	Chinatown	40.715618	-73.994279	Best Wings	40.717716	-73.985140	Wings Joint
3	Chinatown	40.715618	-73.994279	Kimchi Smoke (Pop-up)	40.720341	-73.994049	BBQ Joint
4	Chinatown	40.715618	-73.994279	Pete's Smokehouse And Bbq	40.723947	-73.983926	BBQ Joint
5	Chinatown	40.715618	-73.994279	K&S BBQ Ramen	40.718332	-73.991525	BBQ Joint
6	Chinatown	40.715618	-73.994279	HNH BBQ	40.718364	-74.000785	BBQ Joint
7	Chinatown	40.715618	-73.994279	B&J BBQ	40.724093	-73.997439	BBQ Joint
8	Chinatown	40.715618	-73.994279	Flippin Chicken	40.721366	-74.005608	Fried Chicken Joint
9	Chinatown	40.715618	-73.994279	99 Orchard Rooftop Grill	40.718604	-73.990227	BBQ Joint
10	Chinatown	40.715618	-73.994279	Kimchi Smoke	40.720474	-73.994363	Korean Restaurant

- Finally, Foursquare has one of the largest databases of over 105 million places and is used by over 150,000 developers. Foursquare API will provide many categories of the venue data, of which we are particularly interested in the barbecue restaurant category to help us solve the business problem introduced. This is a project that will make use of many data science techniques including working with APIs (Foursquare), data cleaning, data wrangling, machine learning (k-means clustering) and map visualization (Folium).

Methodology

New York City has a total of 5 boroughs and 306 neighborhoods. In order to segment the neighborhoods and explore them, we will need a dataset that contains those 5 boroughs and the neighborhoods that exist in each borough in addition to the latitude and longitude coordinates of each neighborhood. Fortunately, the link to the data set is available to the public (https://geo.nyu.edu/catalog/nyu_2451_34572).

We will web scrape from this website and clean the dataset to extract the list of neighborhoods data. After gathering the data, we will populate the data into a pandas dataframe and then visualize the neighborhoods in a map using the Folium package. This allows us to perform a sanity check to make sure that the geographical coordinates data returned by the Geocoder are correctly plotted in New York City. We then slice the original dataframe and create a new dataframe of the Manhattan data. We need to

get the geographical coordinates of the Manhattan borough in the form of latitude and longitude in order to be able to use Foursquare API. To do so, we will use the Geocoder package that will allow us to convert an address into geographical coordinates.

After gathering the data, we will populate the data into a pandas dataframe and then visualize the neighborhoods in a map once again using the Folium package. This allows us to perform another sanity check to make sure that the geographical coordinates data returned by the Geocoder are indeed correctly plotted for Manhattan. Next, we will use Foursquare API to get all the barbecue restaurants that are within a radius of 1000 meters. We need to register a Foursquare Developer Account in order to obtain our Foursquare ID and secret key.

We then make API calls to Foursquare passing in the geographical coordinates of the neighborhoods in a Python loop. Foursquare will return the venue data in JSON format and we will extract the venue name, venue category, and venue latitude and longitude. With the data, we can check how many venues were returned for each neighborhood and examine how many unique categories can be curated from all the returned venues. Then, we will analyze each neighborhood by grouping the rows by neighborhood and taking the mean of the frequency and occurrence of each venue category. By doing so, we are also preparing data for use in clustering. Since we are analyzing the “BBQ Restaurants” data, we will filter “BBQ Restaurant” as the venue category for the neighborhoods.

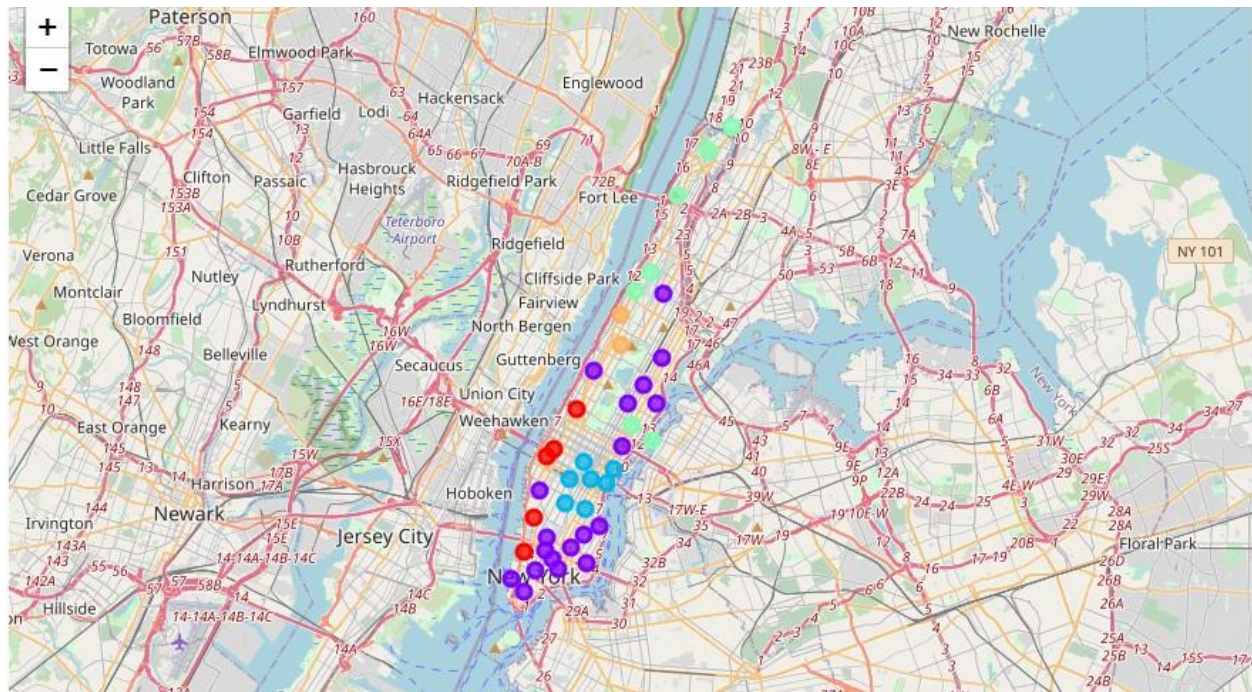
Lastly, we will perform clustering on the data by using k-means clustering. The k-means clustering algorithm identifies a certain number of centroids (represented by ‘k’), and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. It is one of the simplest and most popular unsupervised machine learning algorithms and is particularly suited to solve the problem presented in this project. We will cluster the neighborhoods into 5 clusters based on their frequency of occurrence for “BBQ Restaurant.” The results will allow us to identify which neighborhoods have a higher concentration of barbecue restaurants and which have a lower number of barbecue restaurants. Based on the occurrence of barbecue in different neighborhoods, it will help us answer the question of which neighborhoods are most suitable for opening a new barbecue restaurant.

Results

The results from k-means clustering show that all 5 clusters show a high frequency of barbecue joints with that category of venue being the 1st most common venue in each neighborhood. We can analyze further what some of the less frequent venues are in each cluster of neighborhoods to better determine what type of barbecue our restaurant should serve as to not further saturate the market. Here is a breakdown of the 5 clusters (the full list of clusters can be found in this project’s corresponding Jupyter notebook:

- Cluster 1: High frequency of BBQ joints, but a much lower frequency of wings joints
- Cluster 2: Much higher frequency of BBQ joints with a moderate amount of wings joints
- Cluster 3: Not as high frequency of BBQ joints as cluster 2, but still quite high, with a moderate amount of wings joints
- Clusters 4 & 5: High frequencies of both BBQ joints and wings joints

The results of the clustering are visualized in the map below with cluster 1 in red, cluster 2 in purple, cluster 3 in blue, cluster 4 in mint green, and cluster 5 in light orange.



Discussion

In the observations noted from the map in the 'results' section, we are going to focus our attention on cluster 1. While it had a high frequency of BBQ joints (as did all the other clusters with some having even higher frequencies), it had a much lower frequency of wings joints compared to the other clusters. The neighborhoods represented in cluster 1 are Lincoln Square, Clinton, Tribeca, West Village, and Hudson Yards. Of those 5 neighborhoods, the one with the fewest barbecue restaurants is Lincoln Square with 12 obtained by the chart we generated in the 'data' section. Additionally, there aren't even any wings joints in the top 10 most common venues in this neighborhood. All the other neighborhoods in this cluster (and all the others) have either a high frequency of BBQ joints and/or wings joints in their top 10 most frequent venues. This represents a great opportunity and high potential for there to be a BBQ restaurant that focuses on selling chicken wings as its main product.

Conclusion

We began this project asking the question: Which neighborhood in the Manhattan borough of New York City would be best for opening a new barbecue restaurant? As we analyzed the frequency of such restaurants in Manhattan, we quickly discovered that barbecue restaurants tend to be the most popular venue in each neighborhood. This is likely partially due to the fact the term 'BBQ Joint' is a very general restaurant category. Thus, we refined our focus and decided to look in a more specific direction in opening a barbecue restaurant that sells wings as its main product due to lower market saturation and competition. Through data analysis, data wrangling, and machine learning, we found that the neighborhood that displayed the lowest volume of barbecue restaurants and by far the lowest volume of wings restaurants, was Lincoln Square. Based on location data, this seems to be the optimal location in the Manhattan for building a new barbecue restaurant that sells wings as its main product. However, stakeholders should also obtain further information focusing on those neighborhoods examined in cluster 1 in the results section since that cluster of neighborhoods was by far the most promising. Information such as the attractiveness of each location (proximity to parks and water), socioeconomic data, and real estate prices should also be examined before developers move forward in making this decision.