The Battle of Neighborhoods

Report

Ahmed ElBashary

Capstone Final Project

IBM Data Science Specialization

Course 9

Data: New York Geographical Data

File Type: JSON File

API: Foursquare API

Language: Python

Tool: Jupyter Notebook

Introduction

People are usually on the go especially in a city like New York where everything is moving so fast. Also, the market is growing so fast and it is hard to know which business is the best to invest in. You can never know, but the least you could do is research the market and the neighborhoods, in order to be able to find the best location to open a new business and guarantee some profit. That is why we will be talking about the food that needs to be taken away quickly so people would not waste their time on waiting for food to be done and just take it as the go. This is a very good idea to start a business in a fast running life in New York City, especially Manhattan. Bakery Shops are one of the most efficient places to take food as you go because the food is usually pre-done before the customer enters the shop. All they have to do is just choose from the display and the employee hands them whatever they need. Moreover, everything is small and fairly cheap and easy to get as you are going.

Business Problem

All of the above takes us to the business problem. Starting a bakery shop business in New York will be very fruitful. You need to be unique as there are many bakeries. This report targets businessmen who want to open something common, but still want it to be successful and

want to try new areas to open something as common, fast, cheap and profitable. Also, the businessmen want to open the bakery shop in Manhattan which is a great place to start due to the high traffic of people in the area. We need to start finding the best neighborhood to open a bakery which will be next to places that have already succeed in similar areas, but still does not have as many bakeries as in the other crowded neighborhoods. The main objective of this analysis is to find the best suitable neighborhood in Manhattan for business where we can profit a lot from opening this bakery shop in that particular area. And still not be next to a lot of bakeries that will make competition even harder for a start up next to people that have been there for years.

Data Selection

The file that we will be using is a JSON file of the New York City geographical information. This file will be extracted from the web from a website called "**NYU Spatial Data Repository**" which has all the information we need and for free. The description of this file is "This New York City Neighborhood Names point file was created as a guide to New York City's neighborhoods that appear on the web resource, "New York: A City of Neighborhoods." It contains the boroughs, neighborhoods and the latitude and the longitude for every neighborhood. Then, using libraries from Jupyter Noteboook we will extract the latitude and longitude of New York City itself to use it with the Foursquare API, in order to get the nearby venues in every neighborhood. Foursquare API is a tool that will help us extract all the necessary information of the venues and bakeries that are open in every neighborhood.

Data Cleaning

First of all, we only needed 4 columns from the dataset which are boroughs, neighborhoods, latitude and longitude which we transformed into a pandas dataframe for easier data wrangling. Then, using geolocator we found the latitude and longitude of New York City. After that, Foursquare API was used to extract all nearby bakeries within

1000 meter radius from neighborhoods in Manhattan. This needed a category id for bakery which had to be retrieved from https://developer.foursquare.com/docs/resources/categories. This is how we got the id which was "4bf58dd8d48988d16a941735"

Methodology

Explanatory Analysis

Firstly, after extracting the data from the dataset, I had to look at the Manhattan borough because this is where we want to open the bakery shop. That is why we used folium library to map the Manhattan Borough neighborhoods, so we know where they are exactly located.
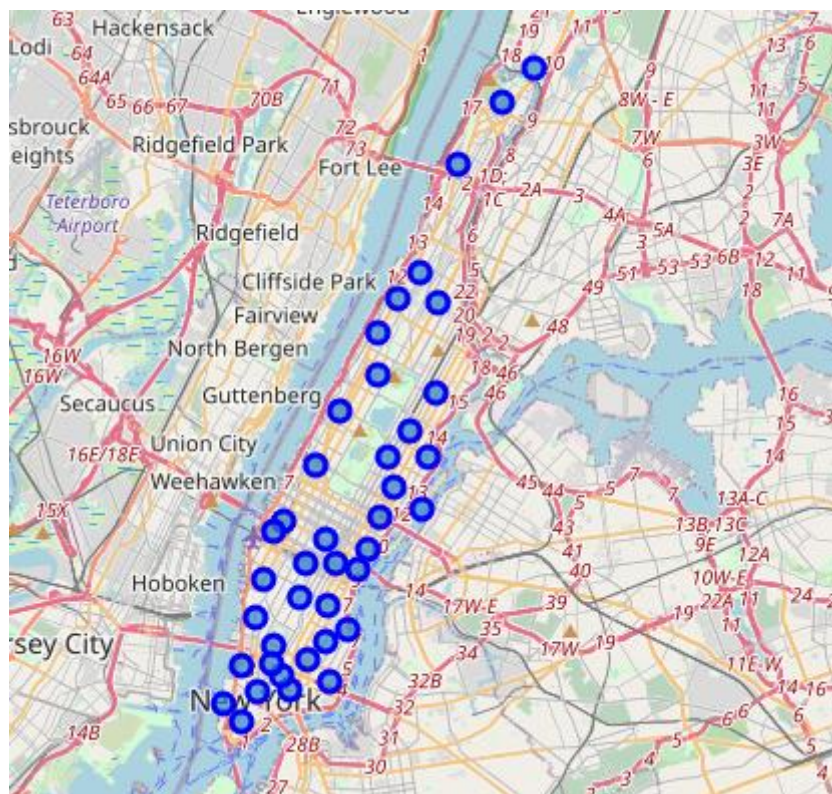
Then, using my client ID and client server, I extracted the necessary venues (bakery) in every neighborhood through Foursquare API. This is done through a URL "https://api.foursquare.com/v2/venues/search?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}" The client id is where you enter your client id and the same goes to the client secret. V stands for the version that you will be using, ll is the latitude and longitude of each neighborhood you are searching in and the radius is the maximum distance you want to search around. At this point you had to add the category id but under condition that it matches the one you got for the bakery shop. Then, we had to look at the map of the venues that we extracted from the tool. And this is how it looks like
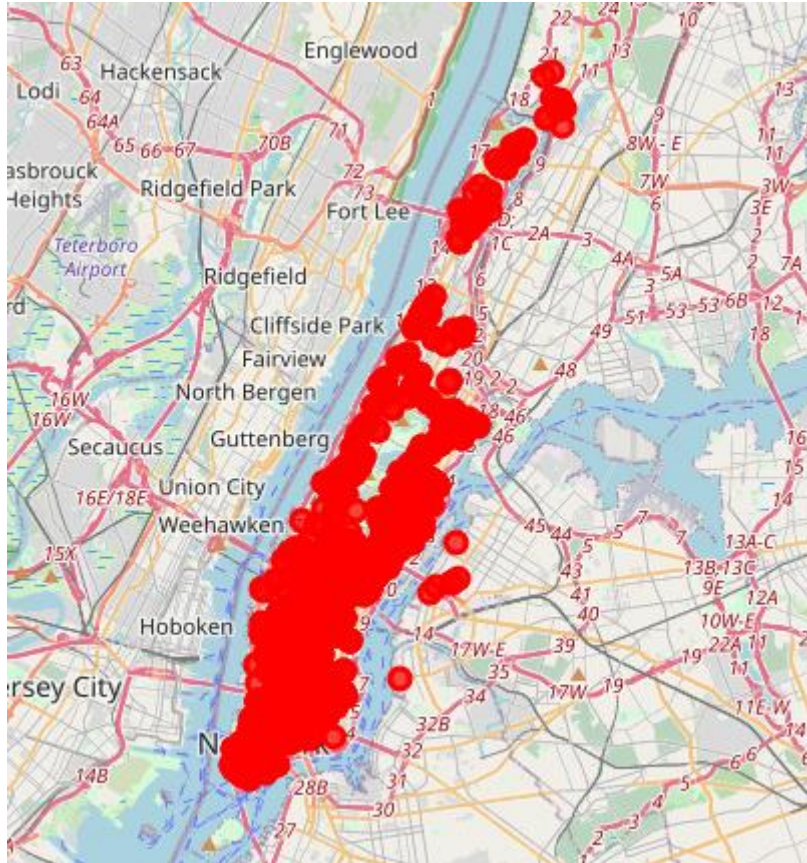
Machine Learning

In order to be able to do machine learning, we had to further analyze the neighborhoods that we got from the Foursquare API. First of all, we had to one hot encode the Venue Categories, so we can get how many venues are in every neighborhood that is so we can get to the best neighborhood to open our bakery shop at the end. Afterwards, we grouped them for every neighborhood then got the mean, so we know the frequency of each category in every neighborhood. As shown in the next figure.

	Neighborhood	American Restaurant	Arcade	Bagel Shop	Bakery	Breakfast Spot	Bubble Tea Shop	Café	Caribbean Restaurant	Chinese Restaurant
0	Battery Park City	0.026316	0.000000	0.000000	0.710526	0.00	0.000000	0.052632	0.000000	0.000000
1	Carnegie Hill	0.041667	0.000000	0.000000	0.770833	0.00	0.000000	0.083333	0.000000	0.000000
2	Central Harlem	0.000000	0.000000	0.000000	0.857143	0.00	0.000000	0.000000	0.000000	0.000000
3	Chelsea	0.000000	0.000000	0.020408	0.551020	0.00	0.000000	0.142857	0.020408	0.000000
4	Chinatown	0.000000	0.000000	0.020000	0.720000	0.02	0.000000	0.060000	0.000000	0.020000
5	Civic Center	0.020000	0.000000	0.000000	0.720000	0.02	0.000000	0.100000	0.000000	0.020000
6	Clinton	0.020408	0.000000	0.000000	0.693878	0.00	0.000000	0.081633	0.000000	0.000000
7	East Harlem	0.000000	0.000000	0.000000	0.847826	0.00	0.000000	0.043478	0.000000	0.000000
8	East Village	0.022727	0.000000	0.022727	0.704545	0.00	0.000000	0.022727	0.000000	0.000000
9	Financial District	0.026316	0.026316	0.000000	0.657895	0.00	0.000000	0.078947	0.000000	0.000000
10	Flatiron	0.020000	0.000000	0.020000	0.500000	0.00	0.000000	0.120000	0.000000	0.000000

And that takes us to the last step which is the machine learning technique that we used. We used k means neighbors clustering technique because it was the most suitable one to use. K-means clustering is "k-means clustering is a method of vector quantization, originally from signal processing, that aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean". In this project, we used 5 clusters as this was the best way to go with it. This takes us to the results that we reached due to this technique which will be discussed in the next section.

Results

First of all, we had to map the clusters to see where is the best place to open the bakery where we know bakeries are open there, but still it does not have that many bakeries so we can be able to predict profit as the competition will be minimal compared to the crowded areas. Also, the clusters help you know which other common venues next to the bakery are common for its success which differs according to the 5 different clusters that we created. So lets take a look at the map and then identify which area is the best area to open the bakery.

The circle is where we think will be the best area to open the bakery as it does not have that many bakeries, but still in the red cluster and the green cluster bakeries are the most common venues. Secondly, the clusters will be demonstrated separately for better understanding. The first cluster which is cluster 1 is colored in Red which looks like this.

:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Marble Hill	Bakery	Bagel Shop	Food	Vegetarian / Vegan Restaurant	Dessert Shop	Arcade	Breakfast Spot	Bubble Tea Shop	Café	Caribbean Restaurant
2	Washington Heights	Bakery	Ice Cream Shop	Vegetarian / Vegan Restaurant	Dessert Shop	Arcade	Bagel Shop	Breakfast Spot	Bubble Tea Shop	Café	Caribbean Restaurant
7	East Harlem	Bakery	Café	Ice Cream Shop	Mexican Restaurant	Italian Restaurant	Coffee Shop	Vegetarian / Vegan Restaurant	Deli / Bodega	Arcade	Bagel Shop
11	Roosevelt Island	Bakery	Ice Cream Shop	French Restaurant	Coffee Shop	Vegetarian / Vegan Restaurant	Deli / Bodega	Arcade	Bagel Shop	Breakfast Spot	Bubble Tea Shop
20	Lower East Side	Bakery	Bagel Shop	Donut Shop	Dessert Shop	Pie Shop	Café	Cocktail Bar	Coffee Shop	Arcade	Breakfast Spot
37	Stuyvesant Town	Bakery	Dessert Shop	Cupcake Shop	Vegetarian / Vegan Restaurant	Arcade	Bagel Shop	Breakfast Spot	Bubble Tea Shop	Café	Caribbean Restaurant

While the second cluster which is cluster 2 is colored in Purple which looks like this.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
22	Little Italy	Bakery	Café	Dessert Shop	Bagel Shop	American Restaurant	Cocktail Bar	Chinese Restaurant	French Restaurant	Ice Cream Shop	Coffee Shop
23	Soho	Bakery	Café	Dessert Shop	American Restaurant	Cocktail Bar	Chinese Restaurant	French Restaurant	Ice Cream Shop	Bubble Tea Shop	Cupcake Shop
25	Manhattan Valley	Bakery	Ice Cream Shop	Pizza Place	Chinese Restaurant	Coffee Shop	Dessert Shop	Vegetarian / Vegan Restaurant	Deli / Bodega	Arcade	Bagel Shop
26	Morningside Heights	Bakery	Ice Cream Shop	Chinese Restaurant	French Restaurant	Coffee Shop	Dessert Shop	Vegetarian / Vegan Restaurant	Deli / Bodega	Arcade	Bagel Shop
29	Financial District	Bakery	Café	Dessert Shop	Ice Cream Shop	American Restaurant	Gourmet Shop	Coffee Shop	Arcade	Sandwich Place	Snack Place
31	Noho	Bakery	Café	Coffee Shop	Dessert Shop	Bagel Shop	Cocktail Bar	French Restaurant	Ice Cream Shop	Donut Shop	Cupcake Shop

The third cluster which is cluster 3 is colored in light Blue which looks like this.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
15	Midtown	Bakery	Café	Coffee Shop	Cupcake Shop	Japanese Restaurant	Snack Place	Deli / Bodega	Dessert Shop	Pie Shop	Turkish Restaurant
17	Chelsea	Bakery	Café	Coffee Shop	Ice Cream Shop	French Restaurant	Cupcake Shop	Bagel Shop	Caribbean Restaurant	Dessert Shop	Vegetarian / Vegan Restaurant
18	Greenwich Village	Bakery	Coffee Shop	Café	Cupcake Shop	Dessert Shop	Cocktail Bar	French Restaurant	Ice Cream Shop	American Restaurant	Breakfast Spot
27	Gramercy	Bakery	Dessert Shop	Coffee Shop	Café	Ice Cream Shop	American Restaurant	Cupcake Shop	Japanese Restaurant	French Restaurant	Arcade
33	Midtown South	Bakery	Café	Coffee Shop	Dessert Shop	French Restaurant	Japanese Restaurant	Turkish Restaurant	Ice Cream Shop	Cupcake Shop	Bagel Shop
38	Flatiron	Bakery	Coffee Shop	Café	Dessert Shop	Cupcake Shop	French Restaurant	American Restaurant	Bagel Shop	Japanese Restaurant	Ice Cream Shop

The fourth cluster which is cluster 4 is colored in light Green which looks like this.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
3	Inwood	Bakery	Ice Cream Shop	Vegetarian / Vegan Restaurant	Dessert Shop	Arcade	Bagel Shop	Breakfast Spot	Bubble Tea Shop	Café	Caribbean Restaurant
4	Hamilton Heights	Bakery	Ice Cream Shop	Coffee Shop	Vegetarian / Vegan Restaurant	Dessert Shop	Arcade	Bagel Shop	Breakfast Spot	Bubble Tea Shop	Café
5	Manhattanville	Bakery	Ice Cream Shop	French Restaurant	Coffee Shop	Vegetarian / Vegan Restaurant	Deli / Bodega	Arcade	Bagel Shop	Breakfast Spot	Bubble Tea Shop
6	Central Harlem	Bakery	Ice Cream Shop	Vegetarian / Vegan Restaurant	Dessert Shop	Arcade	Bagel Shop	Breakfast Spot	Bubble Tea Shop	Café	Caribbean Restaurant

The fifth cluster which is cluster 5 is colored in Orange which looks like this.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	Chinatown	Bakery	Café	Dessert Shop	Donut Shop	Sandwich Place	Bagel Shop	Pie Shop	Breakfast Spot	Chinese Restaurant	Cocktail Bar
8	Upper East Side	Bakery	Café	American Restaurant	Dessert Shop	Cupcake Shop	Sandwich Place	Bagel Shop	Gourmet Shop	French Restaurant	Coffee Shop
9	Yorkville	Bakery	Café	Ice Cream Shop	Dessert Shop	French Restaurant	Gourmet Shop	American Restaurant	Sandwich Place	Coffee Shop	Bubble Tea Shop
10	Lenox Hill	Bakery	Coffee Shop	Cupcake Shop	Dessert Shop	French Restaurant	Gourmet Shop	Ice Cream Shop	American Restaurant	Bagel Shop	Sandwich Place
12	Upper West Side	Bakery	Dessert Shop	Vegetarian / Vegan Restaurant	Pizza Place	Israeli Restaurant	Ice Cream Shop	Coffee Shop	Cupcake Shop	Arcade	Bagel Shop
13	Lincoln Square	Bakery	Café	Dessert Shop	Donut Shop	French Restaurant	Coffee Shop	Cupcake Shop	American Restaurant	Italian Restaurant	Cocktail Bar
14	Clinton	Bakery	Café	Coffee Shop	Donut Shop	Ice Cream Shop	Cupcake Shop	French Restaurant	American Restaurant	Pizza Place	Snack Place
16	Murray Hill	Bakery	Café	Coffee Shop	Japanese Restaurant	Cupcake Shop	Turkish Restaurant	Dessert Shop	Vegetarian / Vegan Restaurant	Deli / Bodega	Arcade
19	East Village	Bakery	Coffee Shop	Dessert Shop	Cupcake Shop	American Restaurant	Bagel Shop	Café	Ice Cream Shop	French Restaurant	Deli / Bodega
21	Tribeca	Bakery	Café	American Restaurant	Dessert Shop	Sandwich Place	Bagel Shop	Breakfast Spot	Coffee Shop	Cupcake Shop	Arcade
24	West Village	Bakery	Coffee Shop	Dessert Shop	Café	Bagel Shop	Bubble Tea Shop	Caribbean Restaurant	Ice Cream Shop	Cupcake Shop	Vegetarian / Vegan Restaurant
28	Battery Park City	Bakery	Dessert Shop	Ice Cream Shop	Café	American Restaurant	Gourmet Shop	Coffee Shop	Sandwich Place	Snack Place	Breakfast Spot
30	Carnegie Hill	Bakery	Café	American Restaurant	Ice Cream Shop	Gourmet Shop	Coffee Shop	Cupcake Shop	Dessert Shop	Deli / Bodega	Arcade

Discussion

As you can see from the results section that we reached a conclusion that the best neighborhood to open the bakery shop is upper west side of Manhattan which was circled in the map figure in the previous section. The only 2 clusters that were inside the zone that we chose are clusters 1 and 4 and from these clusters we observed that whenever bakery is the most common venue, the second most common venue is either ice cream shop or bagel shop. This is why we recommend that anyone who wants to open a bakery shop in Manhattan it should be in the Upper West area where there are ice cream shops around it. This is the best possible option for a successful bakery business.

Conclusion

In conclusion, I think we reached a very good estimate to the best possible neighborhood to open a bakery shop in Manhattan, New York by examining all the possible neighborhoods in Manhattan. This required a lot of machine learning and analyzing to the venues around the neighborhoods to see which one is the perfect scenario for opening a successfully bakery shop. Thank you for taking the time to read this report and I hope it helps you make a business decision that will be fruitful and successful.