

IBM Capstone Project - The Battle of the Neighborhoods

This project tends to solve a problem by applying data science methods on the location data gotten from FourSquare API

Description of the Project

1. Introduction

If you ever has come to New York, you will truly be attracted by its vigor atmosphere, and make a decision to stay and work there just like me. In this city, you will always have something to get fun because it never sleeps for 24 hours a day, 7 days a week. You won't have the opportunity to slow down your progress. Never-ending to-do list, booming business at every corners, and by the expertly-planned transport system which can take you everywhere in the city.

But when you live here, it's still have some problems to slove, one of the highest pressure of living just a part but most important of them. Just imagine finding a well payed job in New York, but a penthouse at the new Peter Marino-designed condo The Getty, just steps from the High Line, is sold for \$59 million. You must say "It can't be serious". Under the pressure, most people likely to choose the way such as working more time, they prefer to eat in the restaurant in order to adapt to the late off work time.

So in the project I decide to find what aspects of a neighborhood that can effect the choice of the restaurant. The taste of the food ? The distance between the restaurant and office? Can the venues surrounding an accommodation effect its price so it might me expensive for you to afford? People will choose a better evironment to enjoy the food even though they need to pay more money and wait for more time? All of these need data to prove.

2.The Project Needs to Solve:

In this project we will try to find an optimal location for a restaurant. This project will try to explore the neighborhoods of New York city to see:

- 1.How to guarantee the regular foot traffic for a restaurant?
- 2.What kinds of factors need to consider like particular cuisine style, price range, and desired target customers?

We will use our data science powers to generate a few most promising neighborhoods based on this criteria. Advantages of each area will then be clearly expressed and get the cluster character, so that best possible final location and restaurant type can be chosen by stakeholders. So, we want to explore the center candidate location that belongs to the restaurant type.

Specifically, this report will be targeted to stakeholders interested in opening an restaurant in New York, United States.

II. Description of the data

1. Data collection process:

To attempt a solution at this problem, we'll be using the Foursquare API to fetch venue information and analyse different neighbourhoods. To fulfil the requirements of the business problem stated, we'll specifically be looking for these types of information:

- Numbers of existing restaurants in the neighborhood (any type of restaurant).
- Number of and distance to different kinds of restaurants in the neighborhood, like any distance of neighborhood from city center number of school in the neighborhood.
- Data on the neighbourhoods in general, specifically the categories for each venue in each Manhattan neighbourhood.
- Popular days and hours for venues that serve our target customer, a popularity score for each of the target venues.

2. Using data to solve:

To attempt a solution at this problem, we'll be using the Foursquare API to fetch venue information and analyse different neighbourhoods. Following data sources will be needed to extract/generate the required information:

- New-York data set: https://cocl.us/new_york_dataset
- One is a free data source from NYU Spatial Data Repository - https://geo.nyu.edu/catalog/nyu_2451_34572. The neighborhood locations will be obtained from here.
- Number of restaurants, their type, location and rating will be obtained using Foursquare API and Zomato. <https://developer.foursquare.com/>

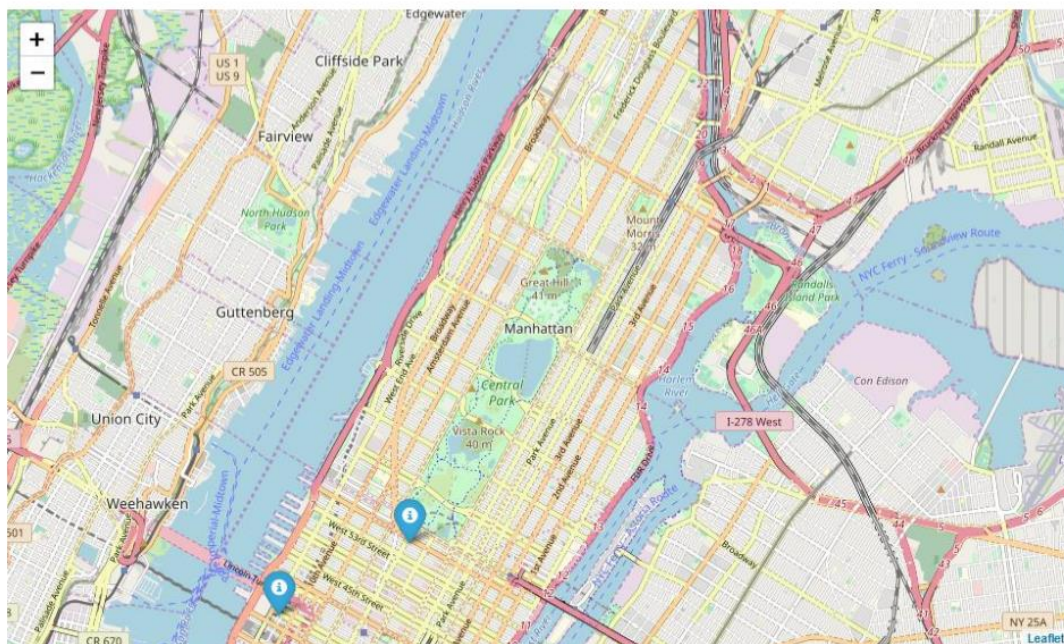
Methodology

Let's create latitude & longitude coordinates for centroids of our candidate neighborhoods. We will create a grid of cells covering our area of interest which is aprox. 12x12 killometers centered around Berlin city center.

After extraction, we dive into exploring the data. We use the PCA method to reduce dimensions and analyze the two main components KDE information. Finally, we use the KMeans method and Silhoutte metric to create the cluster. Caution, we preprocessing the feature Distance from center by StandardScaler method.

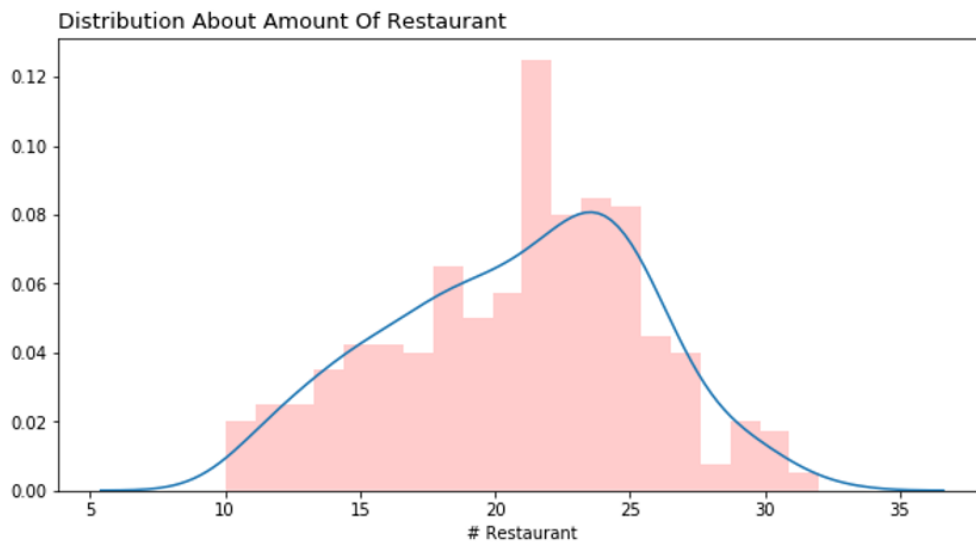
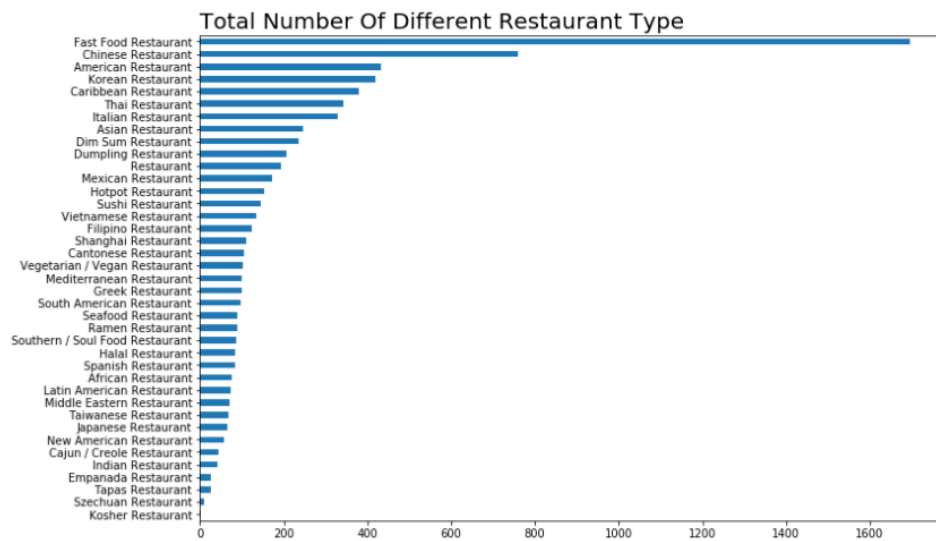
The NYU data source contains location information for all 5 NYC boroughs and 306 neighborhoods. Since we already know which neighborhoods we want to focus on we can simplify our data source from here.

	Borough	Neighborhood	Latitude	Longitude
0	Manhattan	Hudson Yards	40.756658	-74.000111
1	Manhattan	Billionaires Row	40.76503	-73.979886



Result And Conclusion

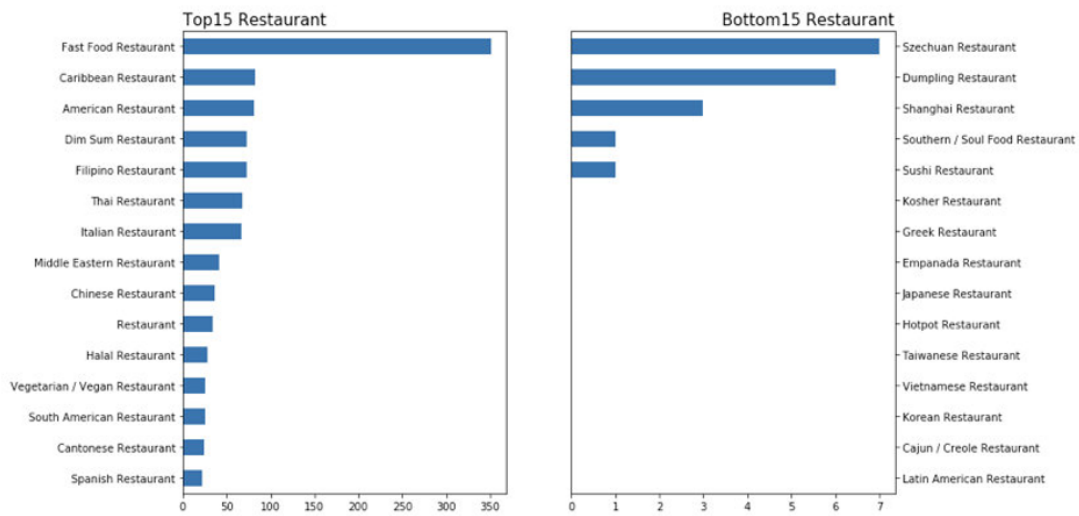
We can explore the cluster information, so that we can get the initial conclusion.



Cluster I Candidate Location

- Fast Restaurant is main type
- Western Restaurants have the largest market
- Eastern Restaurants have the few market

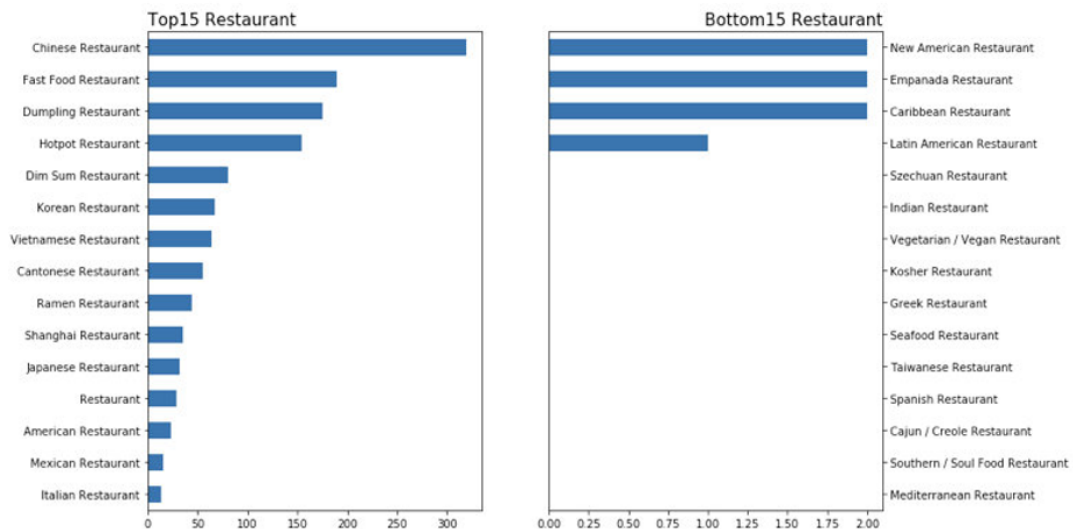
Total Number Of Different Restaurant Type In Cluster I



Cluster II Candidate Location

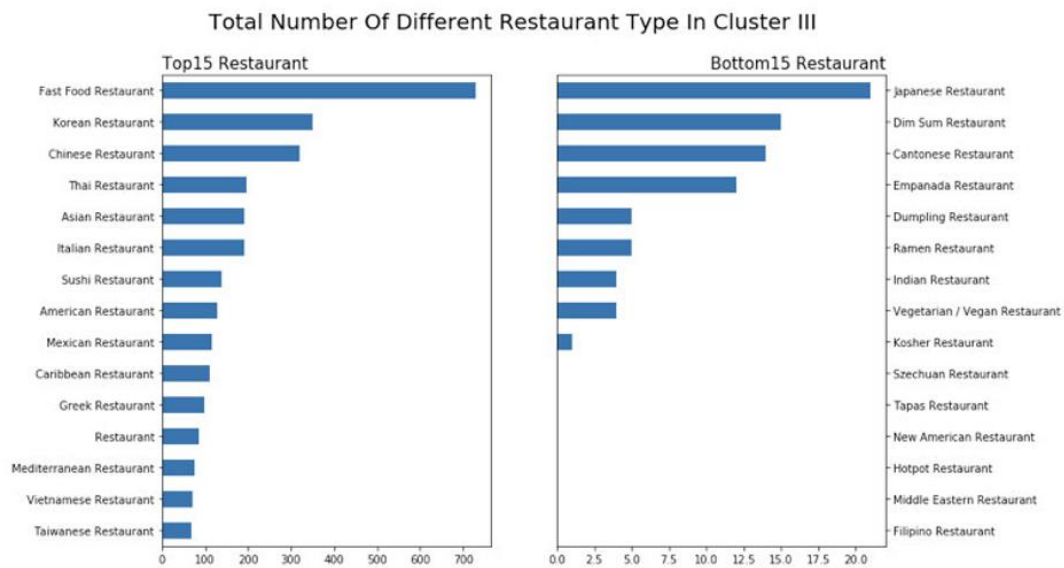
- Chinese Restaurant is main type
- Asian Restaurants have the largest market, like Chinese Breakfast restaurant
- Western Restaurants have the few market

Total Number Of Different Restaurant Type In Cluster II



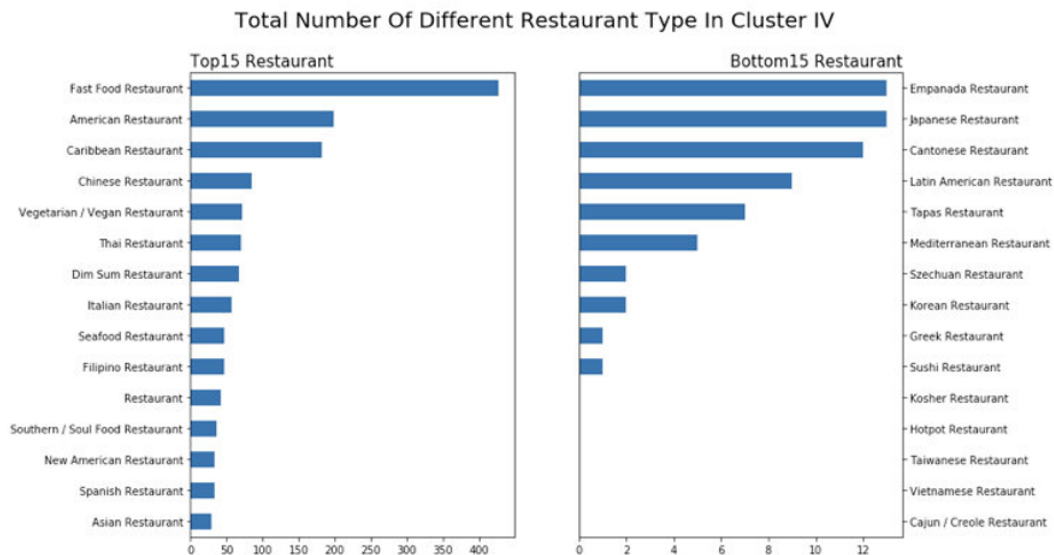
Cluster III Candidate Location

- Fast Food Restaurant is main type.
- The candidate location has mixture restaurant type. Asian Restaurants and western restaurant are same important type the Asian restaurant is lack of variety.
- The candidate location is a good choice to open an Asian restaurant. Maybe the Chinese Breakfast restaurant is a good idea

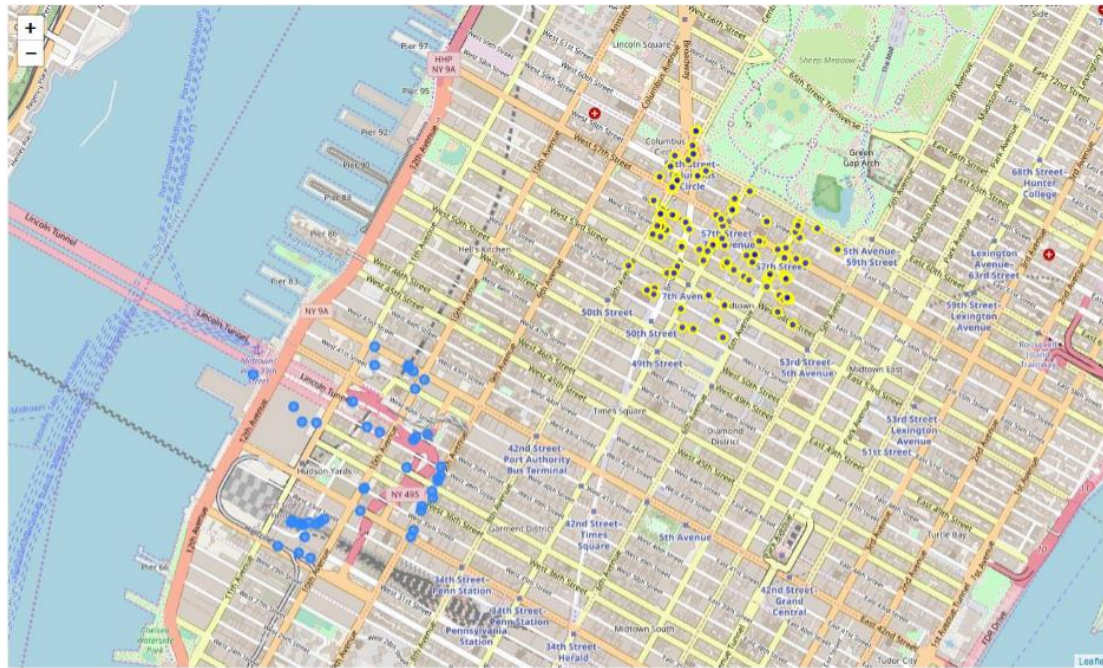


Cluster IV Candidate Location

- Fast Food Restaurant is main type.
- The candidate location has mixture restaurant type.
- Asian Restaurants and western restaurant are same important type the Asian restaurant is lack of variety. The candidate location is a good choice to open an Asian restaurant. But the southeastern Asian restaurant is not a good idea



Map of restaurant location in respective neighborhoods.



Final Discussion

1. We can make a choice that choose the appropriate restaurant type at appropriate candidate location. Like that the Chinese Breakfast restaurant is good choice at Cluster III candidate location.
2. Our analysis shows that although there is a great number of restaurants in New York City (~2000 in our initial area of interest which was 12x12km around Queens), the fast food restaurant is main type close to city center. The eastern Queens has much more food restaurant than the western Queens center, while the school is opposite. So, if we want to choose a appropriate location to open a new restaurant, we choose from the target.