**Capstone Project - The Battle of Neighborhoods**

**Opening a Pizza Pizzerioa in New York**

**1. Introduction**

**1.1 Background**

New York City (NYC), often called simply New York, is the most populous city in the United States. With an estimated 2019 population of 8,336,817 distributed over about 302.6 square miles (784 km2), New York City is also the most densely populated major city in the United States. New York City comprises 5 boroughs sitting where the Hudson River meets the Atlantic Ocean. At its core is Manhattan, a densely populated borough that’s among the world’s major commercial, financial and cultural centers.

New York City's food culture includes an array of international cuisines influenced by the city's immigrant history. Central and Eastern European immigrants, especially Jewish immigrants from those regions, brought bagels, cheesecake, hot dogs, knishes, and delicatessens (or delis) to the city. Italian immigrants brought New York-style pizza and Italian cuisine into the city, while Jewish immigrants and Irish immigrants brought pastrami and corned beef, respectively. Chinese and other Asian restaurants, sandwich joints, trattorias, diners, and coffeehouses are ubiquitous throughout the city. Some 4,000 mobile food vendors licensed by the city, many immigrant-owned, have made Middle Eastern foods such as falafel and kebabs examples of modern New York street food. The city is home to "nearly one thousand of the finest and most diverse haute cuisine restaurants in the world", according to Michelin. The New York City Department of Health and Mental Hygiene assigns letter grades to the city's restaurants based upon their inspection results. As of 2019, there were 27,043 restaurants in the city, up from 24,865 in 2017. The Queens Night Market in Flushing Meadows–Corona Park attracts more than ten thousand people nightly to sample food from more than 85 countries.

**1.2. Problem**

In this project, I will investigate that if there is any good location in New York - Manhattan for opening a pizza restaurant.

A client wants to open his business in Manhattan area, so I focus on that borough during my analysis. We define potential neighborhood based on the number of pizza places which are operating right in each neighborhood. Manhattan has full potential but also is a very challenging district to open a business because of high competition. A new pizza place should be open in an area that inadequate neighborhood in this way the restaurant can attract more customers. Therefore, this analysis necessary to ensure that we have enough customers and that we are not so close to other sushi places.

**2. Data**

To consider the problem we can list the data as below:

* geospatial data for the city

<http://cocl.us/Geospatial_data>

* New York Dataset

<https://cocl.us/new_york_dataset>

* Foursquare API to download all venues information from all neighborhoods

<https://api.foursquare.com>

Pizza Place 4bf58dd8d48988d1ca941735

* New York Population: <https://data.cityofnewyork.us/City-Government/Manhattan-populations-by-neighborhood/8m6s-esnp/data>
* Python libraries:

Pandas: Creating and manipulating data frames

Numpy: Powerful n-dimensional arrays and numerical computing tools

Folium: Python visualization library would be used to visualize the neighborhoods cluster

Scikit Learn: for importing k-means clustering

JSON: Library to handle JSON files

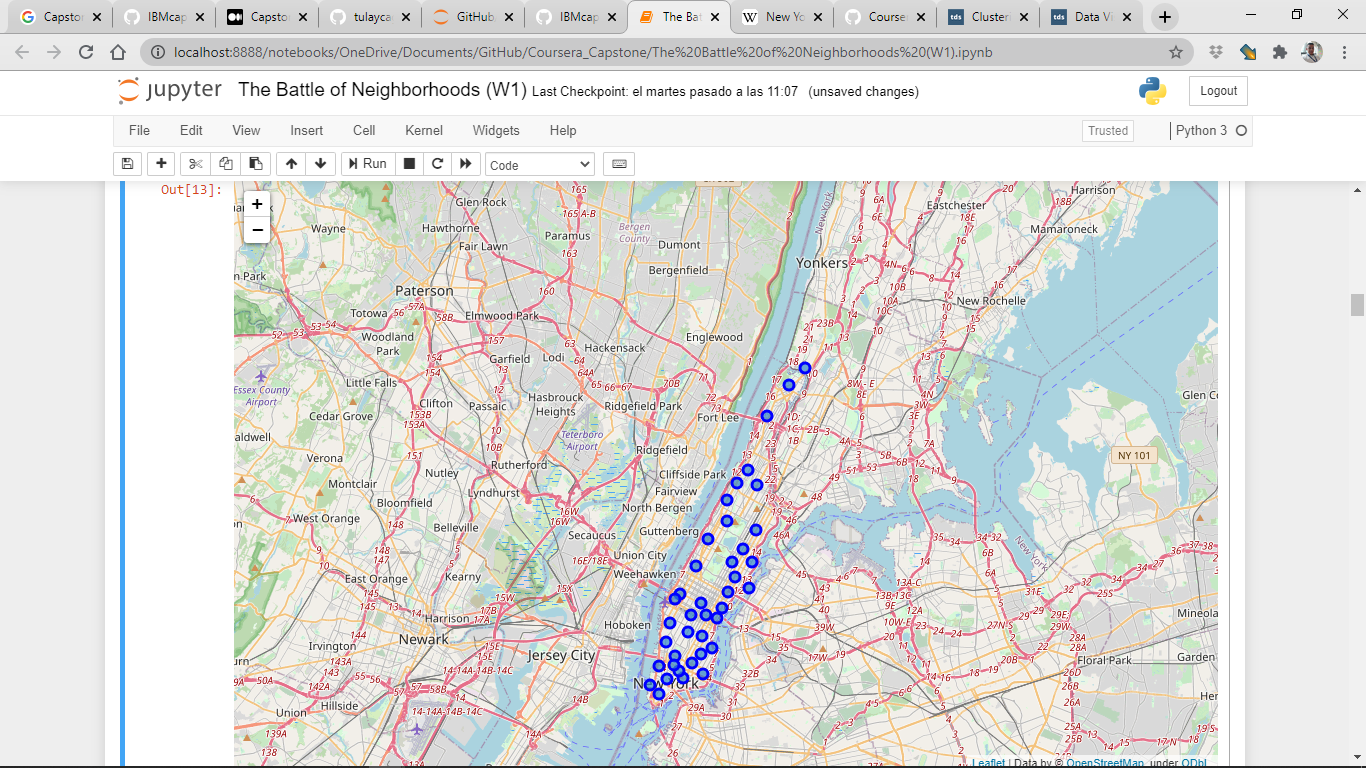
Geocoder: To retrieve Location Data

BeautifulSoup and Requests: To scrap and library to handle http requests

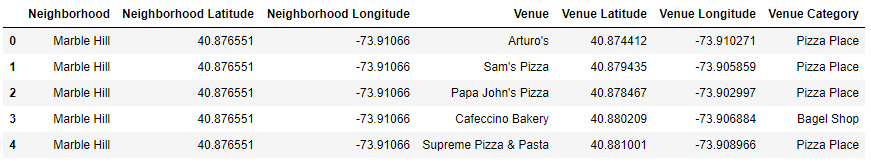
Matplotlib: Python Plotting Module

**3. Methodology**

I took neighborhood from wikipedia and put blue dots on New York map to see centers of neighborhoods. There are **40** neighborhoods in Manhattan



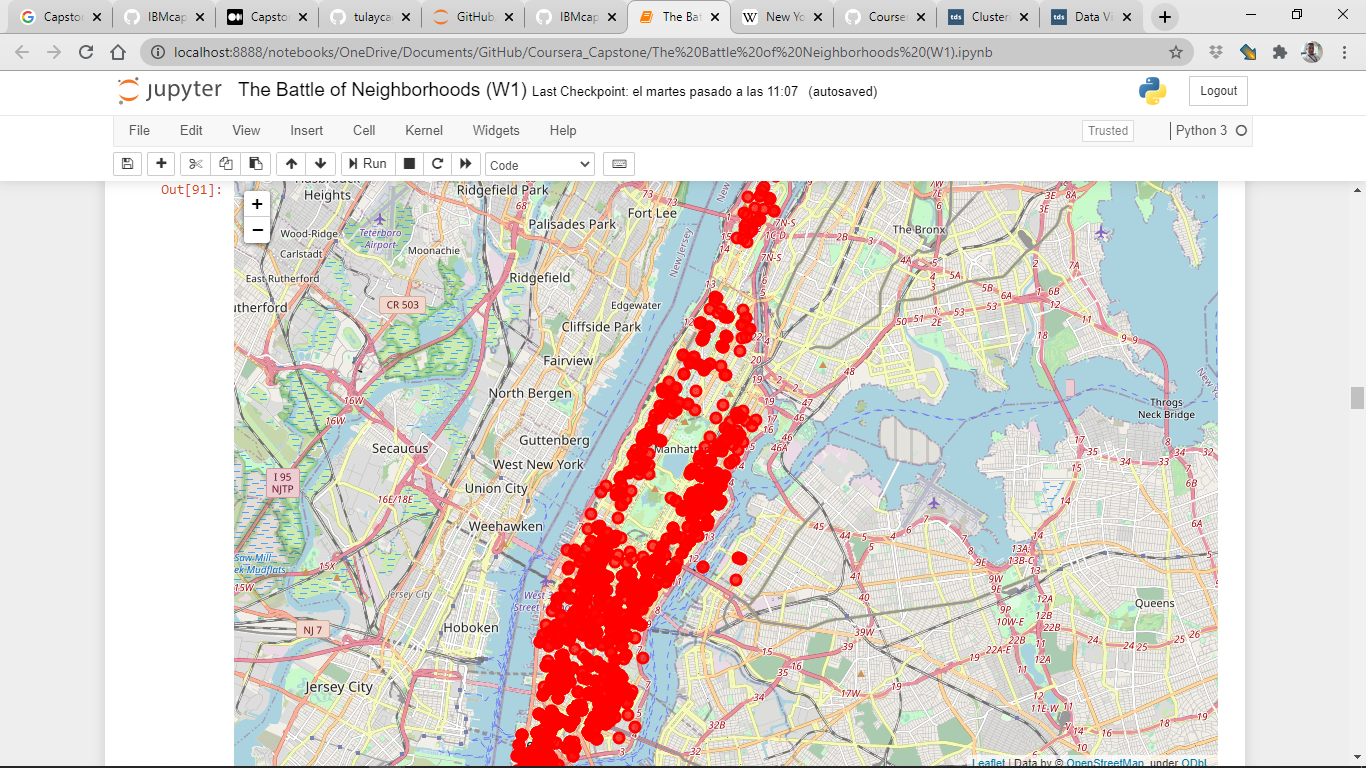
* Addresses are converted into their equivalent latitude and longitude values.
* Foursquare API is used to explore neighborhoods in Manhattan, New York.
* After that, explore function to get pizza places categories in each neighborhood.
* Next, we also used Google Map API to find their geographic coordinates of the 5 locations shortlisted for pizza places:



Calculating the Percentage of pizza restaurants:



Mappint pizza places in Manhattan



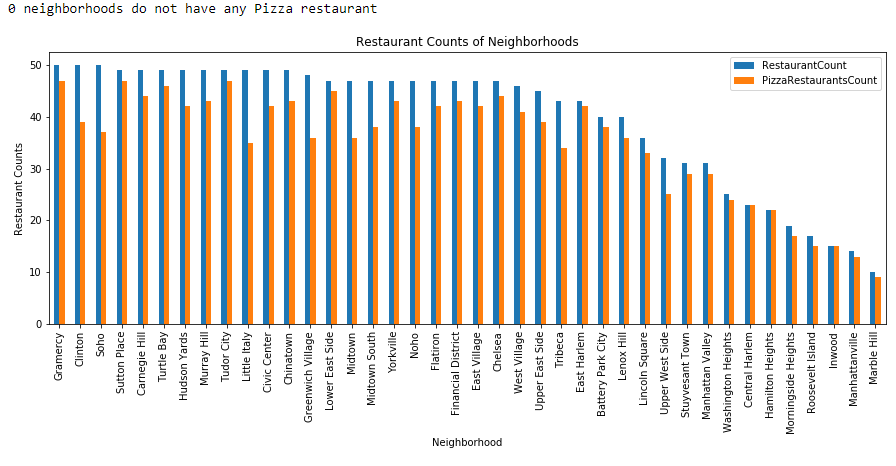


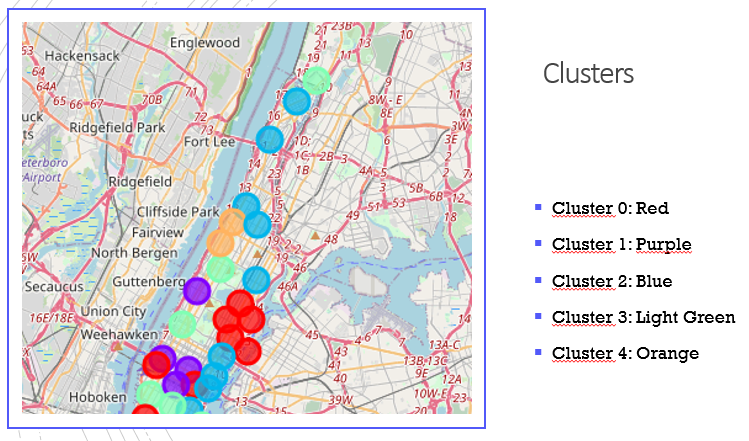
Then using this feature to group the neighborhoods into clusters K-means clustering algorithm will be use to complete this task. And also, the Folium library to visualize the neighborhoods in Manhattan and its emerging clusters.

**4. Result**

Getting the Restaurant Counts of Neighborhoods

Resutaruanst Vs Pizza Places

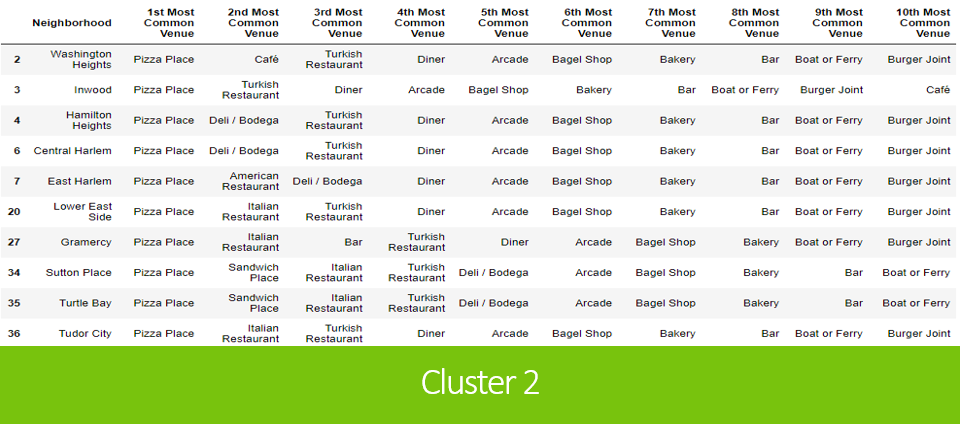




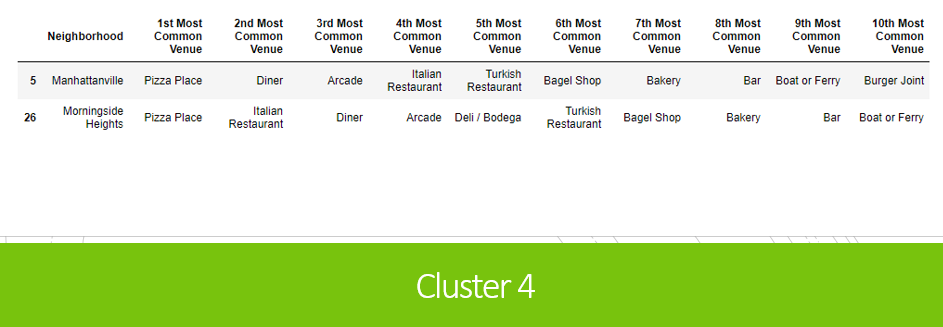
Using K-mean to clustering data











**5. Discussion**

0 neighborhoods do not have any Pizza restaurant

Cluster 4 It has fewer restaurants, but most of the restaurants are pizzerias

Cluster 0 and 1 especially likes Itallian tastes. We could have an opportunity there.

**6. Conclusion**

|  |  |
| --- | --- |
| • **Neighborhood : Hudson Yards , Cluster 0, population: 70,150**  I choosed Hudson Yards as first choice , since the Neighborhood of this cluster love italian tastes, the population and overall we still having profit margin for pizza places about the comparation restaurant / pizzerias. |  |
| • Neighborhood : Clinton , Cluster 1, population: 45,884  I choosed Clinton , since the Neighborhood of this cluster love italian tastes, and overall we still having profit margin for pizza places about the comparation restaurant / pizzerias. |
| • Neighborhood : Soho , Cluster 1, population: 42,742  I choosed Soho , since the Neighborhood of this cluster, love italian tastes, and overall we still having profit margin for pizza places about the comparation restaurant / pizzerias. |  |