Capstone Project

The Battle of Neighborhoods 1



# **Introduction and Business Problem**

**Introduction**

There are about 6700 retail bakeries in the United States, that have a combined revenue of about $3,000,000,000. That puts the average revenue per bakery at about $450,000. The 50 largest players, however, earn about 20 percent of this income, so if we open a small retail bakery, we’ll probably take in less than the industry average – especially, while we're getting your enterprise off the ground.

Starting a bakery can be a beautiful business opportunity, but we need to distinguish ourself from others to enjoy long-term success.

**Business Problem**

A group of friends and I wants to open a little business in New York area, so as business analyst they ask to me to study the project by the angle of the location by using location data in addition to other datasets in the aim to identity the better NEW-YORK city borough for a new bakery.

We've define that this new bakery should be located in a well define neighborhood based on the number of bakery which are already operating right in each neighborhood. Manhattan has full potential but also is a very challenging district to open a business because of high competition. the new bakery should be open in an area which can attract a lot of people with different purpose ( tourist, employees, people living in the aera ....) that neighborhood defined like that can attract enough customers. Therefore, this analysis necessary to ensure that we have enough customers and that we are not so close to other bakery.

**Data Selection**

To identify the characteristics of our competitors' venues in New York we would first need to find out the number of bakeries in New-York currently and their location.

We then used a Map API to find their geographic coordinates based on their geographic coordonates.

we’ve retreive the geographic coordonates of 1575 bakeries splitted in all 7 neighborhood of Manhattan as we define that this is the borough with full potential. Here some extract of our notebook to illustrate our purpose.

Number of Bakery : 1575



Geographic coordonates

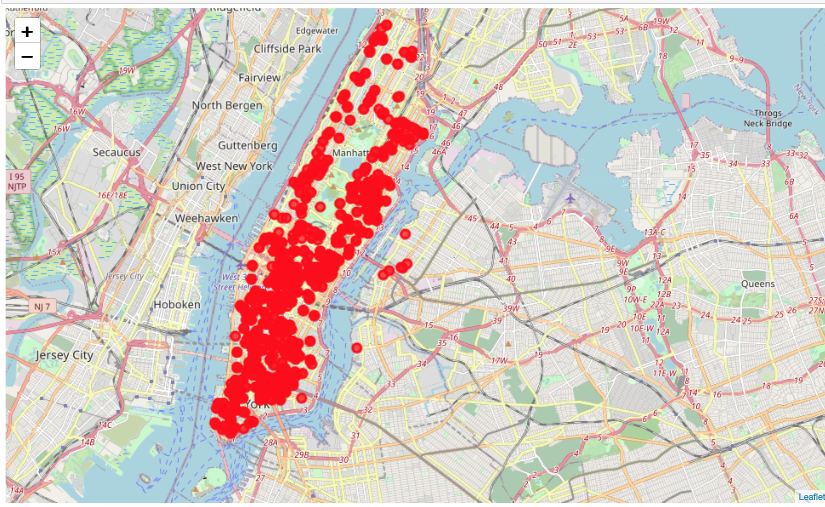


**Methodology**

We use Foursquare API to explore neighborhoods in Manhattan, New York.

Then we use a explore function to get bakery categories in each neighborhood.

Map of New York’s bakery



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

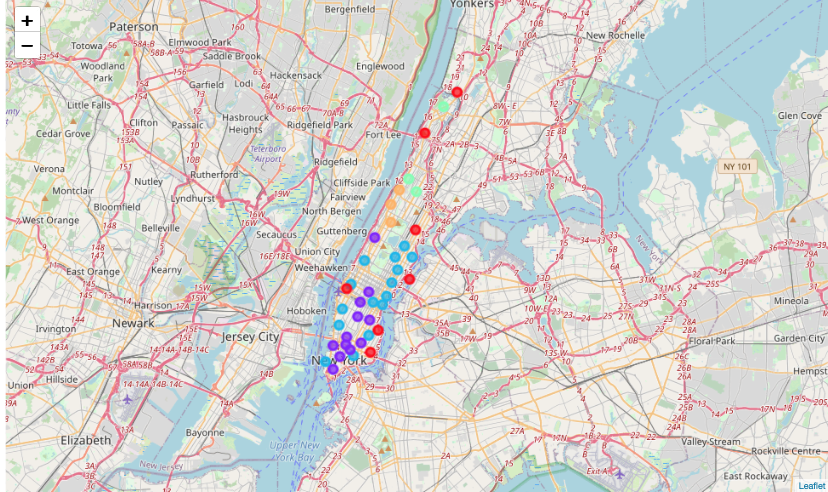
the neighborhoods in Manhattan and its emerging clusters.



Using K-mean to clustering data area



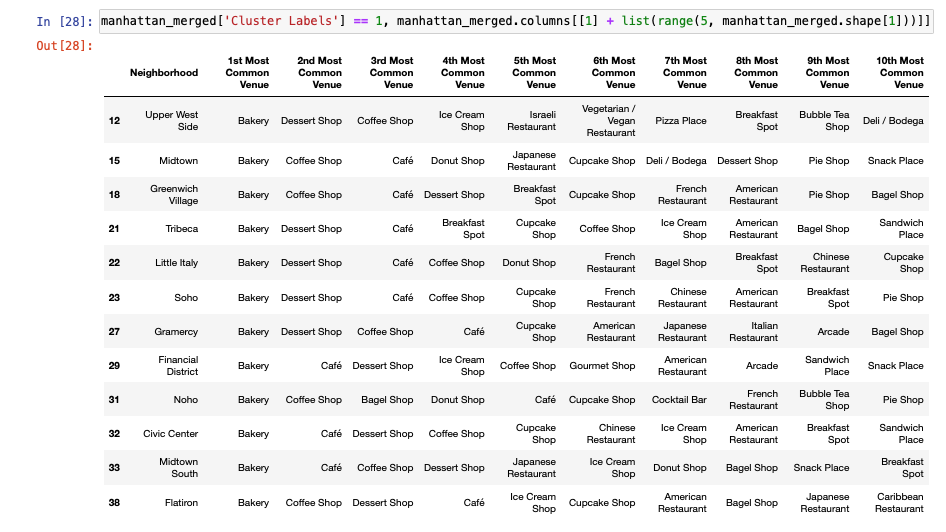
Then we create the map of with the clusters identified with the method Map of folium



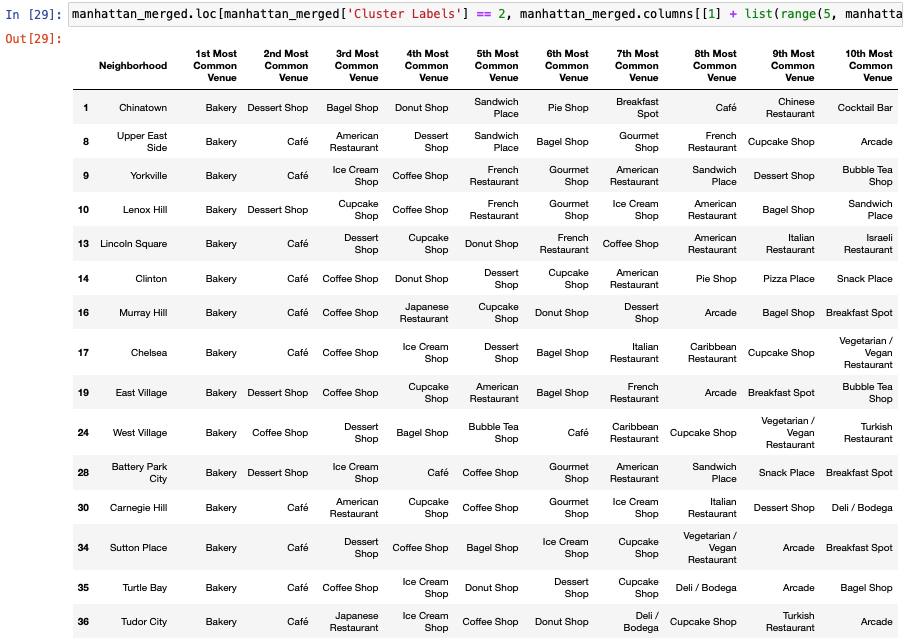
**Cluster identified as 0**



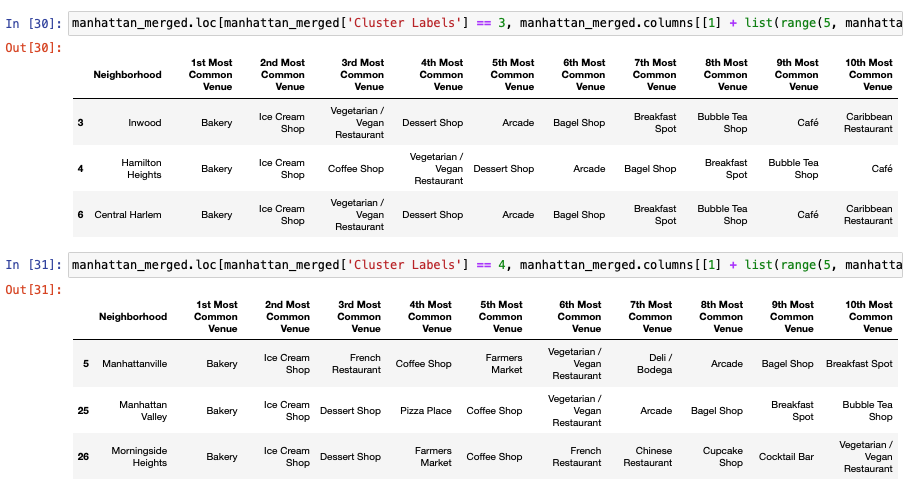
**Cluster identified as 1**

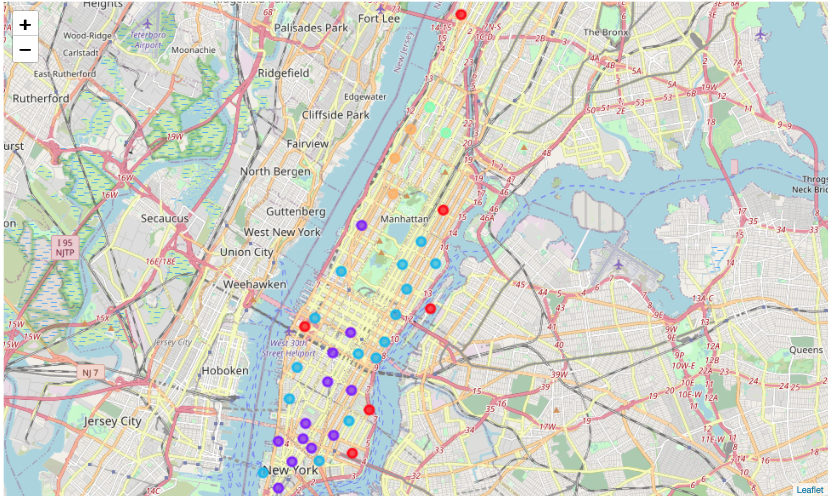


**Cluster identified as 2**



**Cluster identified as 3 and 4**





Based on the Dataframe analysis we can conclude that Cluster 3 (Inwood, Hamilton high, Central Harlem ) and Cluster 4 (Manhattanville, Manhattanvalley, Morningside Heights) areas are the best places to open a new bakery.

**Conclusions.**

* This analysis is performed on limited data. Which by consequences limit the value of these conclusions. But with good amount of data is available there is scope to come up with better results.
* There is high competition in Midtown and lower Manhattan so it is very risky to open business in these part of the town.
* It can be done more detailed analysis by adding other factors such as transportations, touristic frequentation, attractions such as museums, theaters etc....