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

```
newyork_venues_bakery.shape
```
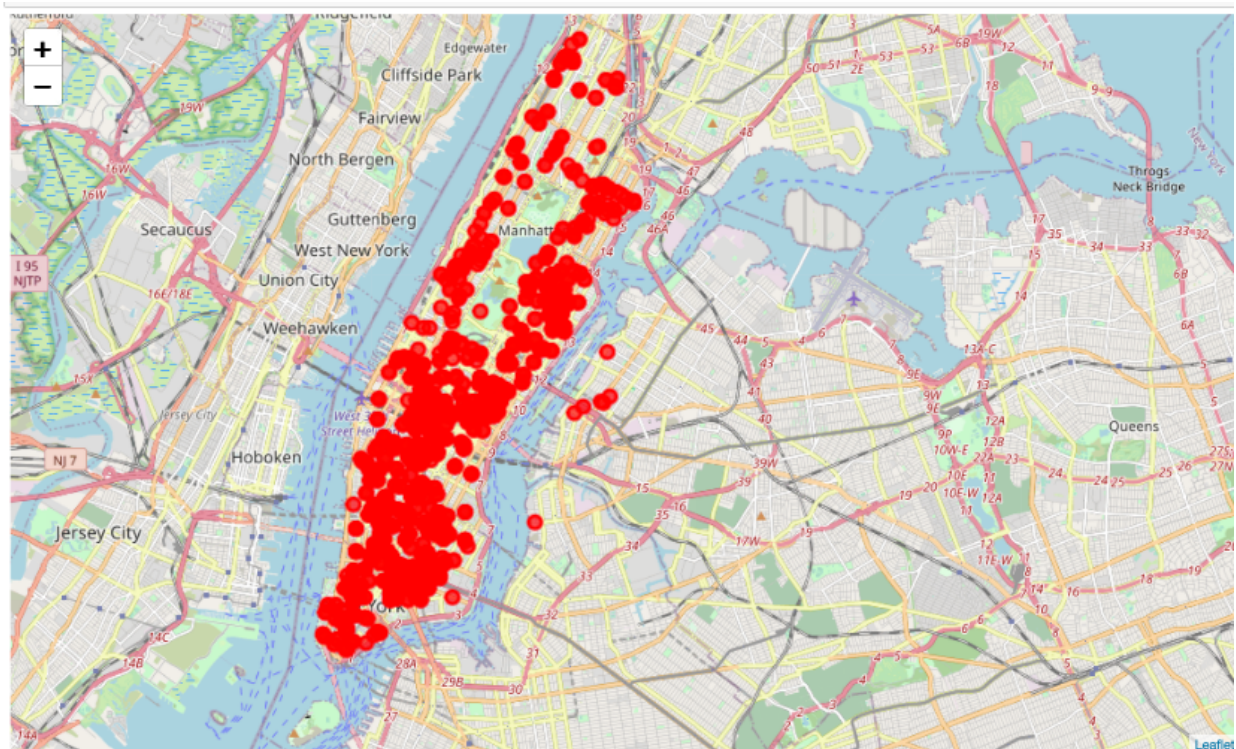```
(1575, 7)
```

<u>Geographic coordonates</u>

| | Neighborhood | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|---|---|---|---|---|---|---|
| 0 | Marble Hill | 40.876551 | -73.91066 | Auntie Anne's | 40.877400 | -73.904947 | Bakery |
| 1 | Marble Hill | 40.876551 | -73.91066 | Franco #2 Bakery | 40.874975 | -73.909381 | Food |
| 2 | Marble Hill | 40.876551 | -73.91066 | Rosarina Bakery (aka Franco Bakery) | 40.874870 | -73.909398 | Bakery |
| 3 | Marble Hill | 40.876551 | -73.91066 | Cafeccino Bakery | 40.880209 | -73.906884 | Bagel Shop |
| 4 | Marble Hill | 40.876551 | -73.91066 | Corner Cafe And Bakery | 40.887248 | -73.906822 | Bakery |

# 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.

<u>Map of New York's bakery</u>

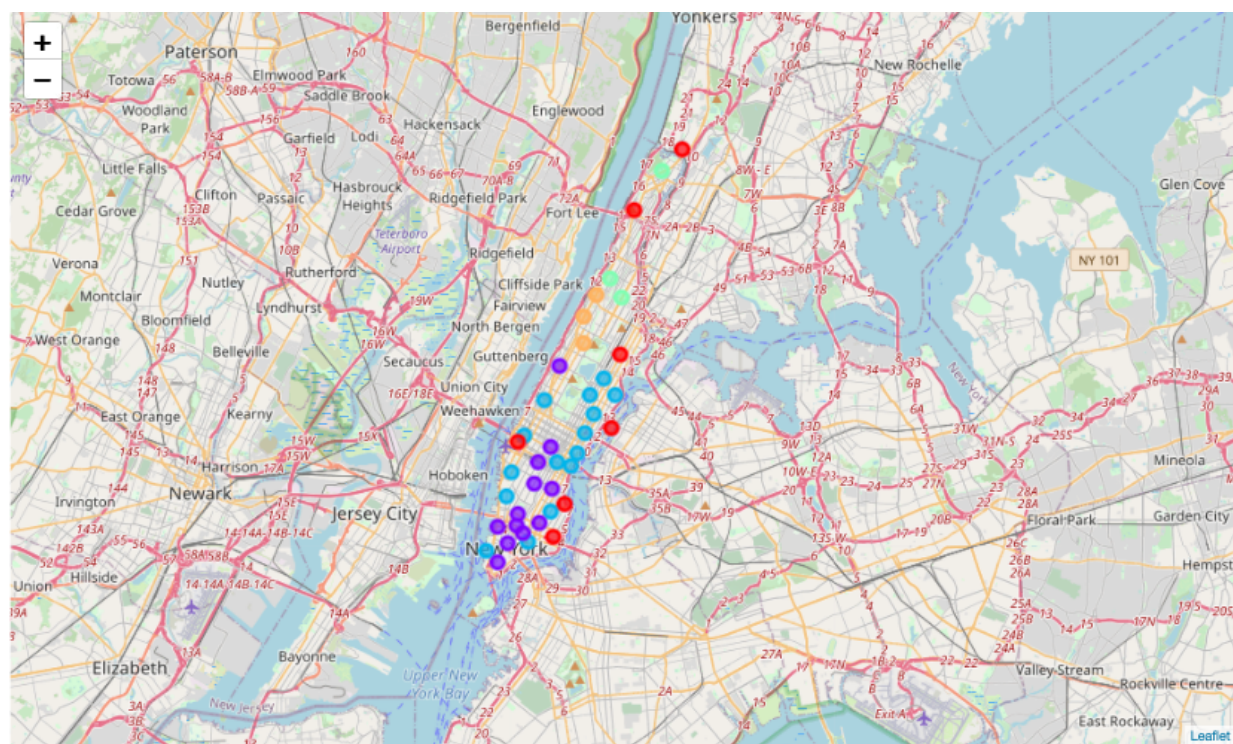Then using this feature to group the neighborhoods into clusters K-means.

Clustering algorithm will be use to complete this task. And also, the Folium library to visualize the neighborhoods in Manhattan and its emerging clusters.

| | 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 | Battery Park City | Bakery | Dessert Shop | Ice Cream Shop | Café | Coffee Shop | Gourmet Shop | American Restaurant | Sandwich Place | Snack Place | Breakfast Spot |
| 1 | Carnegie Hill | Bakery | Café | American Restaurant | Cupcake Shop | Coffee Shop | Gourmet Shop | Ice Cream Shop | Italian Restaurant | Dessert Shop | Deli / Bodega |
| 2 | Central Harlem | Bakery | Ice Cream Shop | Vegetarian / Vegan Restaurant | Dessert Shop | Arcade | Bagel Shop | Breakfast Spot | Bubble Tea Shop | Café | Caribbean Restaurant |
| 3 | Chelsea | Bakery | Café | Coffee Shop | Ice Cream Shop | Dessert Shop | Bagel Shop | Italian Restaurant | Caribbean Restaurant | Cupcake Shop | Vegetarian / Vegan Restaurant |
| 4 | Chinatown | Bakery | Dessert Shop | Bagel Shop | Donut Shop | Sandwich Place | Pie Shop | Breakfast Spot | Café | Chinese Restaurant | Cocktail Bar |

Using K-mean to clustering data area

| | Borough | Neighborhood | Latitude | Longitude | Cluster Labels | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Manhattan | Marble Hill | 40.876551 | -73.910660 | 0 | Bakery | Bagel Shop | Food | Vegetarian / Vegan Restaurant | Dessert Shop | Arcade | Breakfast Spot | Bubble Tea Shop | Café |
| 1 | Manhattan | Chinatown | 40.715618 | -73.994279 | 2 | Bakery | Dessert Shop | Bagel Shop | Donut Shop | Sandwich Place | Pie Shop | Breakfast Spot | Café | Chinese Restaurant |
| 2 | Manhattan | Washington Heights | 40.851903 | -73.936900 | 0 | Bakery | Ice Cream Shop | Vegetarian / Vegan Restaurant | Dessert Shop | Arcade | Bagel Shop | Breakfast Spot | Bubble Tea Shop | Café |
| 3 | Manhattan | Inwood | 40.867684 | -73.921210 | 3 | Bakery | Ice Cream Shop | Vegetarian / Vegan Restaurant | Dessert Shop | Arcade | Bagel Shop | Breakfast Spot | Bubble Tea Shop | Café |
| 4 | Manhattan | Hamilton Heights | 40.823604 | -73.949688 | 3 | Bakery | Ice Cream Shop | Coffee Shop | Vegetarian / Vegan Restaurant | Dessert Shop | Arcade | Bagel Shop | Breakfast Spot | Bubble Tea Shop |

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



## Cluster identified as 0

```
In [27]: manhattan_merged['Cluster Labels'] == 0, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1]))]]
```

Out[27]:

	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	Dessert Shop	Arcade	Bagel Shop	Breakfast Spot	Bubble Tea Shop
20	Lower East Side	Bakery	Bagel Shop	Donut Shop	Pie Shop	Cocktail Bar	Coffee Shop	Dessert Shop	Arcade	Breakfast Spot	Bubble Tea Shop
37	Stuyvesant Town	Bakery	Cupcake Shop	Dessert Shop	Vegetarian / Vegan Restaurant	Turkish Restaurant	Arcade	Bagel Shop	Breakfast Spot	Bubble Tea Shop	Café
39	Hudson Yards	Bakery	Coffee Shop	Donut Shop	Cupcake Shop	Chinese Restaurant	Dessert Shop	Pie Shop	Pizza Place	Snack Place	Bagel Shop

Cluster identified as 1

In [28]: manhattan_merged['Cluster Labels'] == 1, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1])))]]

Out[28]:

	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
12	Upper West Side	Bakery	Dessert Shop	Coffee Shop	Ice Cream Shop	Israeli Restaurant	Vegetarian / Vegan Restaurant	Pizza Place	Breakfast Spot	Bubble Tea Shop	Deli / Bodega
15	Midtown	Bakery	Coffee Shop	Café	Donut Shop	Japanese Restaurant	Cupcake Shop	Deli / Bodega	Dessert Shop	Pie Shop	Snack Place
18	Greenwich Village	Bakery	Coffee Shop	Café	Dessert Shop	Breakfast Spot	Cupcake Shop	French Restaurant	American Restaurant	Pie Shop	Bagel Shop
21	Tribeca	Bakery	Dessert Shop	Café	Breakfast Spot	Cupcake Shop	Coffee Shop	Ice Cream Shop	American Restaurant	Bagel Shop	Sandwich Place
22	Little Italy	Bakery	Dessert Shop	Café	Coffee Shop	Donut Shop	French Restaurant	Bagel Shop	Breakfast Spot	Chinese Restaurant	Cupcake Shop
23	Soho	Bakery	Dessert Shop	Café	Coffee Shop	Cupcake Shop	French Restaurant	Chinese Restaurant	American Restaurant	Breakfast Spot	Pie Shop
27	Gramercy	Bakery	Dessert Shop	Coffee Shop	Café	Cupcake Shop	American Restaurant	Japanese Restaurant	Italian Restaurant	Arcade	Bagel Shop
29	Financial District	Bakery	Café	Dessert Shop	Ice Cream Shop	Coffee Shop	Gourmet Shop	American Restaurant	Arcade	Sandwich Place	Snack Place
31	Noho	Bakery	Coffee Shop	Bagel Shop	Donut Shop	Café	Cupcake Shop	Cocktail Bar	French Restaurant	Bubble Tea Shop	Pie Shop
32	Civic Center	Bakery	Café	Dessert Shop	Coffee Shop	Cupcake Shop	Chinese Restaurant	Ice Cream Shop	American Restaurant	Breakfast Spot	Sandwich Place
33	Midtown South	Bakery	Café	Coffee Shop	Dessert Shop	Japanese Restaurant	Ice Cream Shop	Donut Shop	Bagel Shop	Snack Place	Breakfast Spot
38	Flatiron	Bakery	Coffee Shop	Dessert Shop	Café	Ice Cream Shop	Cupcake Shop	American Restaurant	Bagel Shop	Japanese Restaurant	Caribbean Restaurant

Cluster identified as 2

In [29]: manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 2, manhattan_merged.columns[[1] + list(range(5, manhatta

Out[29]:

	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	Dessert Shop	Bagel Shop	Donut Shop	Sandwich Place	Pie Shop	Breakfast Spot	Café	Chinese Restaurant	Cocktail Bar
8	Upper East Side	Bakery	Café	American Restaurant	Dessert Shop	Sandwich Place	Bagel Shop	Gourmet Shop	French Restaurant	Cupcake Shop	Arcade
9	Yorkville	Bakery	Café	Ice Cream Shop	Coffee Shop	French Restaurant	Gourmet Shop	American Restaurant	Sandwich Place	Dessert Shop	Bubble Tea Shop
10	Lenox Hill	Bakery	Dessert Shop	Cupcake Shop	Coffee Shop	French Restaurant	Gourmet Shop	Ice Cream Shop	American Restaurant	Bagel Shop	Sandwich Place
13	Lincoln Square	Bakery	Café	Dessert Shop	Cupcake Shop	Donut Shop	French Restaurant	Coffee Shop	American Restaurant	Italian Restaurant	Israeli Restaurant
14	Clinton	Bakery	Café	Coffee Shop	Donut Shop	Dessert Shop	Cupcake Shop	American Restaurant	Pie Shop	Pizza Place	Snack Place
16	Murray Hill	Bakery	Café	Coffee Shop	Japanese Restaurant	Cupcake Shop	Donut Shop	Dessert Shop	Arcade	Bagel Shop	Breakfast Spot
17	Chelsea	Bakery	Café	Coffee Shop	Ice Cream Shop	Dessert Shop	Bagel Shop	Italian Restaurant	Caribbean Restaurant	Cupcake Shop	Vegetarian / Vegan Restaurant
19	East Village	Bakery	Dessert Shop	Coffee Shop	Cupcake Shop	American Restaurant	Bagel Shop	French Restaurant	Arcade	Breakfast Spot	Bubble Tea Shop
24	West Village	Bakery	Coffee Shop	Dessert Shop	Bagel Shop	Bubble Tea Shop	Café	Caribbean Restaurant	Cupcake Shop	Vegetarian / Vegan Restaurant	Turkish Restaurant
28	Battery Park City	Bakery	Dessert Shop	Ice Cream Shop	Café	Coffee Shop	Gourmet Shop	American Restaurant	Sandwich Place	Snack Place	Breakfast Spot
30	Carnegie Hill	Bakery	Café	American Restaurant	Cupcake Shop	Coffee Shop	Gourmet Shop	Ice Cream Shop	Italian Restaurant	Dessert Shop	Deli / Bodega
34	Sutton Place	Bakery	Café	Dessert Shop	Coffee Shop	Bagel Shop	Ice Cream Shop	Cupcake Shop	Vegetarian / Vegan Restaurant	Arcade	Breakfast Spot
35	Turtle Bay	Bakery	Café	Coffee Shop	Ice Cream Shop	Donut Shop	Dessert Shop	Cupcake Shop	Deli / Bodega	Arcade	Bagel Shop
36	Tudor City	Bakery	Café	Japanese Restaurant	Ice Cream Shop	Coffee Shop	Donut Shop	Deli / Bodega	Cupcake Shop	Turkish Restaurant	Arcade

Cluster identified as 3 and 4

```
In [30]: manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 3, manhattan_merged.columns[[1] + list(range(5, manhatta
```
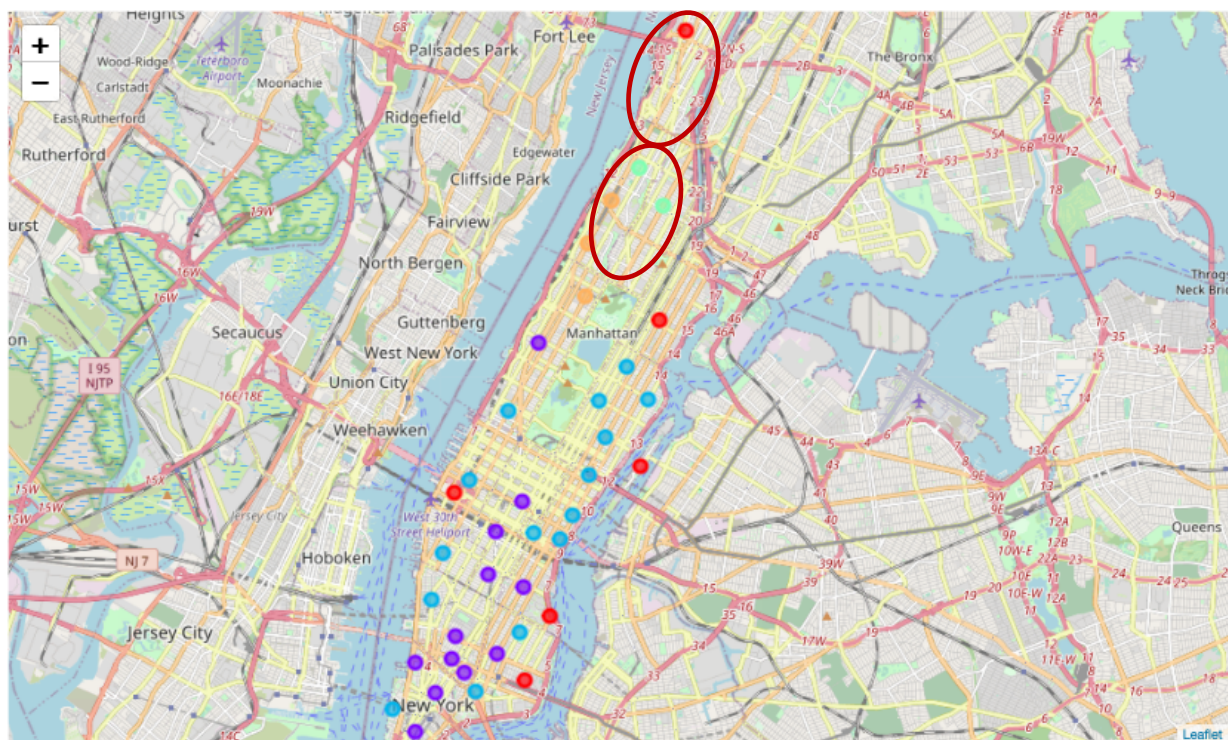Out[30]:

	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
6	Central Harlem	Bakery	Ice Cream Shop	Vegetarian / Vegan Restaurant	Dessert Shop	Arcade	Bagel Shop	Breakfast Spot	Bubble Tea Shop	Café	Caribbean Restaurant

```
In [31]: manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 4, manhattan_merged.columns[[1] + list(range(5, manhatta
```
Out[31]:

	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
5	Manhattanville	Bakery	Ice Cream Shop	French Restaurant	Coffee Shop	Farmers Market	Vegetarian / Vegan Restaurant	Deli / Bodega	Arcade	Bagel Shop	Breakfast Spot
25	Manhattan Valley	Bakery	Ice Cream Shop	Dessert Shop	Pizza Place	Coffee Shop	Vegetarian / Vegan Restaurant	Arcade	Bagel Shop	Breakfast Spot	Bubble Tea Shop
26	Morningside Heights	Bakery	Ice Cream Shop	Dessert Shop	Farmers Market	Coffee Shop	French Restaurant	Chinese Restaurant	Cupcake Shop	Cocktail Bar	Vegetarian / Vegan Restaurant

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....