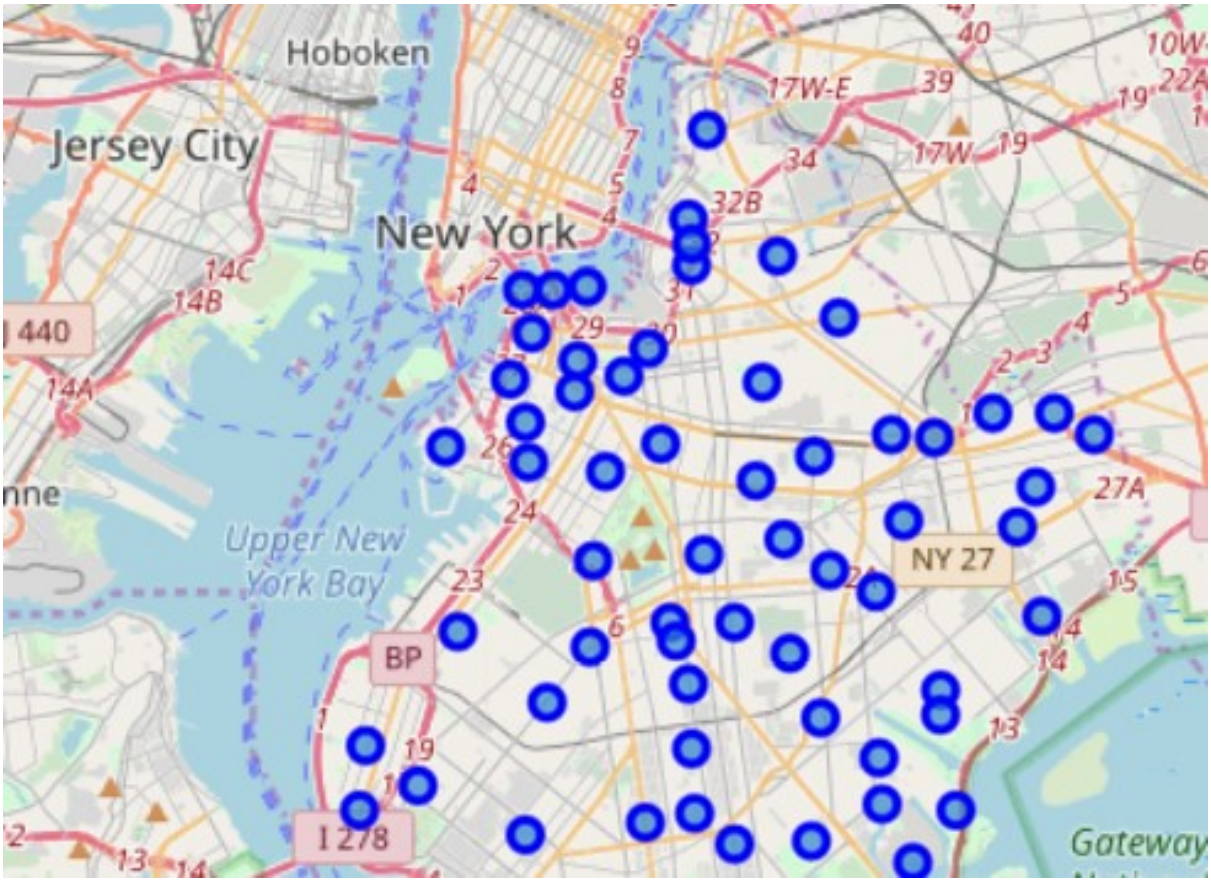
DATA SCIENCE

CAPSTONE PROJECT

THE BATTLE OF NEIGHBORHOOD

By – Ankush Bhatla


Hoboken
Jersey City
New York
Upper New York Bay
NY 27
BP
I 278
Gateway

Introduction

The **City of New York**, usually called either **New York City** (**NYC**) or simply **New York** (**NY**), is the most populous city in the United States. With an estimated 2018 population of 8,398,748 distributed over a land area of about 302.6 square miles (784 km^2), New York is also the most densely populated major city in the United States. Located at the southern tip of the state of New York, the city is the center of the New York metropolitan area, the largest metropolitan area in the world by urban landmass and one of the world's most populous megacities, with an estimated 19,979,477 people in its 2018 Metropolitan Statistical Area and 22,679,948 residents in its Combined Statistical Area. A global power city, New York City has been described as the cultural, financial, and media capital of the world, and exerts a significant impact upon commerce, entertainment, research, technology, education, politics, tourism, art, fashion, and sports. The city's fast pace has inspired the term *New York minute*. Home to the headquarters of the United Nations, New York is an important center for international diplomacy.

Situated on one of the world's largest natural harbors, New York City consists of five boroughs, each of which is a separate county of the State of New York. The five boroughs – Brooklyn, Queens, Manhattan, The Bronx, and Staten Island – were consolidated into a single city in 1898. The city and its metropolitan area constitute the premier gateway for legal immigration to the United States. As many as 800 languages are spoken in New York, making it the most linguistically diverse city in the world. New York City is home to more than 3.2 million residents born outside the United States, the largest foreign-born population of any city in the world. As of 2019, the New York metropolitan area is estimated to produce a gross metropolitan product (GMP) of US$1.9 trillion. If greater New York City were a sovereign state, it would have the 12th highest GDP in the world. New York is home to the highest number of billionaires of any city in the world.

Business Problem

The City of New York is famous for its excelllent cuisine. It's food culture includes an array of international cuisines influenced by the city's immigrant history. Sushi restaurants have become so popular in the United States now it seems that there is one on every corner, not only in major cities but also in smaller cities. Starting a sushi restaurant can be a great business opportunity, but you need to distinguish yourself from others to enjoy long-term success.

[5]:

	Borough	Neighborhood	Latitude	Longi
0	Bronx	Wakefield	40.894705	-73.84
1	Bronx	Co-op City	40.874294	-73.82
2	Bronx	Eastchester	40.887556	73.82
3	Bronx	Fieldston	40.895437	-73.90

If you plan a real restaurant that can demand higher prices for fresh fish, delivered daily from Japan, focus on neighborhoods and outlets that already attract a sophisticated Japanese client. If you plan a cheap buffet restaurant, points to the masses looking for affordable high-traffic locations with large shopping centers and other local points of interest.

My client wants to open his business in Brooklyn area, so I focus on that borough during my analysis. We define potential neighborhood based on the number of sushi bars which are operating right in each neighborhood. Brooklynhas full potential but also is a very challenging district to open a business because of high competition. New sushi bar should be open in an area that inadequate neighborhood in this way the bar 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.

Data

Data 1: Neighborhood has a total of 5 boroughs and 306 neighborhoods. In order to segement the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the the latitude and logitude coordinates of each neighborhood. This dataset exists for free on the web. Link to the dataset is: https://cocl.us/new_york_dataset

	Borough	Neighborhood	Latitude	Lo
0	Bronx	Wakefield	40.894705	-73
1	Bronx	Co-op City	40.874294	-73
2	Bronx	Eastchester	40.887556	-73
3	Bronx	Fieldston	40.895437	-73

Data2: Newyork city geographical coordinates data will be utilized as input for the Foursquare API, that will be leveraged to provision venues information for each neighborhood. We will use the Foursquare API to explore neighborhoods in New York City. The below is image of the Foursquare API data.

In addition, Sushi category Id 4bf58dd8d48988d1d2941735 is used for retrieving data from Foursquare API.

JSON File - newyork_data.json

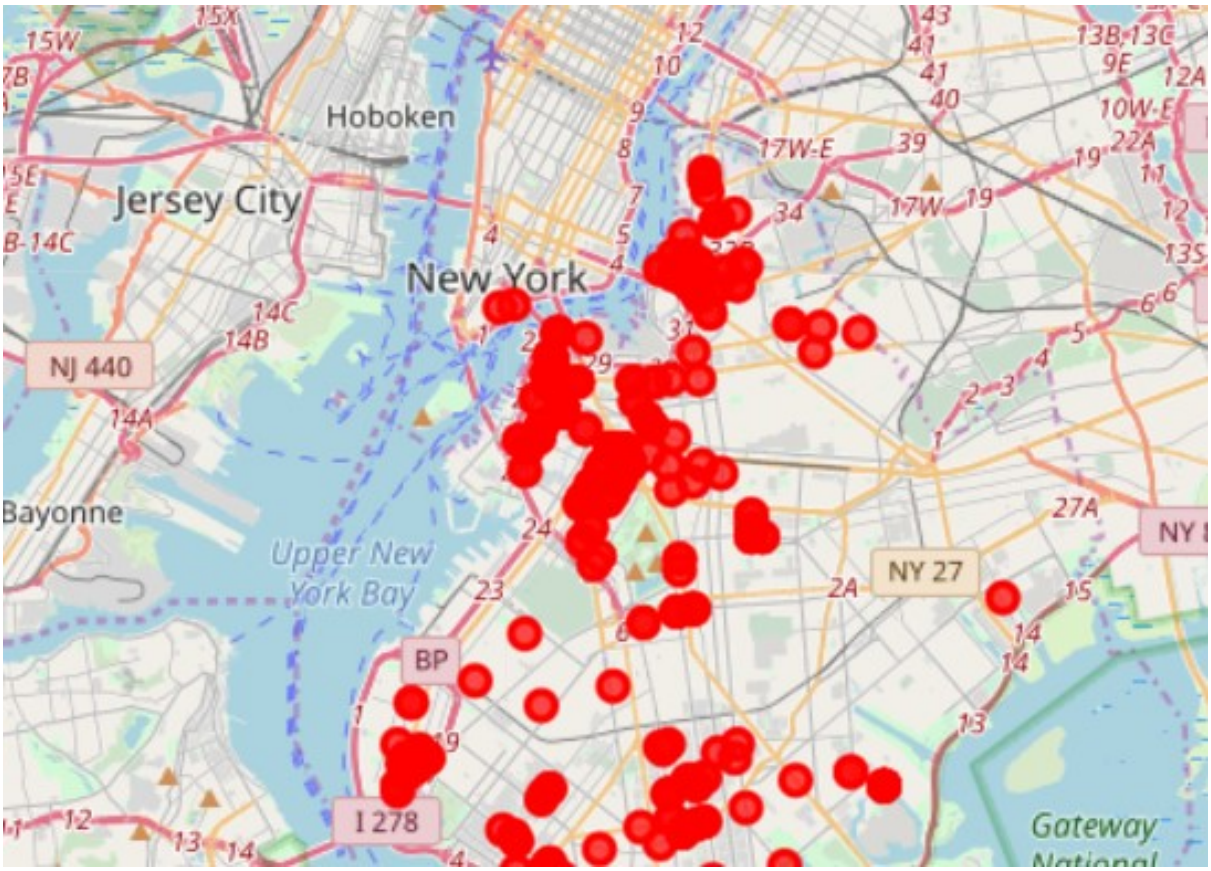
Methodology

In this project, I will use the basic methodology as taught in Week 3 lab.

	Borough	Neighborhood	Latitude	Lo
0	Bronx	Wakefield	40.894705	-73
1	Bronx	Co-op City	40.874294	-73
2	Bronx	Eastchester	40.887556	-73
3	Bronx	Fieldston	40.895437	-73

Above, I have done convert addresses into their equivalent latitude and longitude values. Then we will use the Foursquare API to explore neighborhoods in Brooklyn, New York. After that, explore function to get sushi restaurant categories in each neighborhood.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue L
0	Bay Ridge	40.625801	-74.030621	Inaka	40.625141	-7
1	Bay Ridge	40.625801	-74.030621	Sakana Sushi & Asian Bistro	40.623623	-7
2	Bay Ridge	40.625801	-74.030621	Omiya	40.622329	-7
3	Bay Ridge	40.625801	-74.030621	Sushi Yu III	40.632485	-7



Sushi bars in Brooklyn

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue L
0	Bay Ridge	40.625801	-74.030621	Inaka	40.625141	-
1	Bay Ridge	40.625801	-74.030621	Sakana Sushi & Asian Bistro	40.623623	-
2	Bay Ridge	40.625801	-74.030621	Sushi Yu III	40.632485	-
3	Bay Ridge	40.625801	-74.030621	Omiya	40.622329	-

Then use 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 Brooklyn and its emerging clusters.

	Neighborhood	Asian Restaurant	Bubble Tea Shop	Burger Joint	Café	Chinese Restaurant	Fish Market	Italian Restaurant	Japanese Restaurant	Kosher Restaurant
0	Bay Ridge	0	0	0	0	0	0	0	0	0
1	Bay Ridge	0	0	0	0	0	0	0	0	0
2	Bay Ridge	0	0	0	0	0	0	0	0	0
3	Bay Ridge	0	0	0	0	0	0	0	0	0

	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 Commo Venu
0	Bath Beach	Sushi Restaurant	Japanese Restaurant	Thai Restaurant	Snack Place	Poke Place	Kosher Restaurant	Italian Restaurant	Fish Marke
1	Bay Ridge	Sushi Restaurant	Japanese Restaurant	Asian Restaurant	Thai Restaurant	Snack Place	Poke Place	Kosher Restaurant	Italian Restaurar
2	Bensonhurst	Sushi Restaurant	Thai Restaurant	Snack Place	Poke Place	Kosher Restaurant	Japanese Restaurant	Italian Restaurant	Fish Marke
3	Bergen Beach	Sushi Restaurant	Japanese Restaurant	Thai Restaurant	Snack Place	Poke Place	Kosher Restaurant	Italian Restaurant	Fish Marke

Results

K-mean Cluster Using K-mean to clustering data area with less number of sushi bars

Cluster 0

[40]:

	Borough	Neighborhood	Latitude	Longitude	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	8
0	Brooklyn	Bay Ridge	40.625801	-74.030621	Sushi Restaurant	Japanese Restaurant	Asian Restaurant	Thai Restaurant	Snack Place	Poke Place	Kosher Restaurant	R
1	Brooklyn	Bensonhurst	40.611009	-73.995180	Sushi Restaurant	Thai Restaurant	Snack Place	Poke Place	Kosher Restaurant	Japanese Restaurant	Italian Restaurant	
2	Brooklyn	Sunset Park	40.645103	-74.010316	Sushi Restaurant	Asian Restaurant	Thai Restaurant	Snack Place	Poke Place	Kosher Restaurant	Japanese Restaurant	R
3	Brooklyn	Greenpoint	40.730201	-73.954241	Sushi Restaurant	Japanese Restaurant	Thai Restaurant	Snack Place	Poke Place	Kosher Restaurant	Italian Restaurant	

Cluster 1

```
manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 1, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1]))]]
```

	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
1	Chinatown	Sushi Restaurant	Japanese Restaurant	Vegetarian / Vegan Restaurant	Noodle House	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega
18	Greenwich Village	Sushi Restaurant	Japanese Restaurant	Sake Bar	Grocery Store	Vegetarian / Vegan Restaurant	Noodle House	Bakery	Chinese Restaurant
19	East Village	Sushi Restaurant	Japanese Restaurant	Vegetarian / Vegan Restaurant	Noodle House	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega
20	Lower East Side	Sushi Restaurant	Japanese Restaurant	Vegetarian / Vegan Restaurant	Noodle House	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega
21	Tribeca	Sushi Restaurant	Noodle House	Japanese Restaurant	Theme Restaurant	Vegetarian / Vegan Restaurant	Bakery	Chinese Restaurant	Cocktail Ba
22	Little Italy	Sushi Restaurant	Japanese Restaurant	Noodle House	Vegetarian / Vegan Restaurant	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega
23	Soho	Sushi Restaurant	Japanese Restaurant	Noodle House	Theme Restaurant	Vegetarian / Vegan Restaurant	Bakery	Chinese Restaurant	Cocktail Ba
31	Noho	Sushi Restaurant	Japanese Restaurant	Vegetarian / Vegan Restaurant	Noodle House	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega
32	Civic Center	Sushi Restaurant	Japanese Restaurant	Noodle House	Theme Restaurant	Vegetarian / Vegan Restaurant	Bakery	Chinese Restaurant	Cocktail Ba

Cluster 2

```
manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 2, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1]))]]
```

	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 Comm Ven
8	Upper East Side	Sushi Restaurant	Japanese Restaurant	Asian Restaurant	Grocery Store	Noodle House	Bakery	Chinese Restaurant	Cocktail B
9	Yorkville	Sushi Restaurant	Japanese Restaurant	Asian Restaurant	Chinese Restaurant	Noodle House	Bakery	Cocktail Bar	Deli / Bode
10	Lenox Hill	Sushi Restaurant	Asian Restaurant	Japanese Restaurant	Grocery Store	Noodle House	Bakery	Chinese Restaurant	Cocktail B
11	Roosevelt Island	Sushi Restaurant	Asian Restaurant	Japanese Restaurant	Noodle House	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bode
13	Lincoln Square	Sushi Restaurant	Japanese Restaurant	Smoothie Shop	Chinese Restaurant	Grocery Store	Vegetarian / Vegan Restaurant	Noodle House	Bake
14	Clinton	Sushi Restaurant	Japanese Restaurant	Poke Place	Chinese Restaurant	Cocktail Bar	Asian Restaurant	Seafood Restaurant	Sandwich Pla
15	Midtown	Sushi Restaurant	Asian Restaurant	Japanese Restaurant	Seafood Restaurant	Ramen Restaurant	Bakery	Vegetarian / Vegan Restaurant	Sandwich Pla
16	Murray Hill	Sushi Restaurant	Japanese Restaurant	Asian Restaurant	Restaurant	Bakery	Chinese Restaurant	Ramen Restaurant	Vegetarian / Veg Restaura
17	Chelsea	Sushi Restaurant	Japanese Restaurant	Asian Restaurant	Vegetarian / Vegan Restaurant	Smoothie Shop	Seafood Restaurant	Sandwich Place	Sake B
24	West Village	Sushi Restaurant	Japanese Restaurant	Vegetarian / Vegan Restaurant	Sake Bar	Grocery Store	Asian Restaurant	Sandwich Place	Seafood Restaura
25	Manhattan Valley	Sushi Restaurant	Hawaiian Restaurant	Japanese Restaurant	Vegetarian / Vegan Restaurant	Noodle House	Bakery	Chinese Restaurant	Cocktail B
27	Gramercy	Sushi Restaurant	Japanese Restaurant	Vegetarian / Vegan Restaurant	Chinese Restaurant	Noodle House	Bakery	Cocktail Bar	Deli / Bode
28	Battery Park City	Sushi Restaurant	Japanese Restaurant	Noodle House	Theme Restaurant	Vegetarian / Vegan Restaurant	Bakery	Chinese Restaurant	Cocktail B

Cluster 3

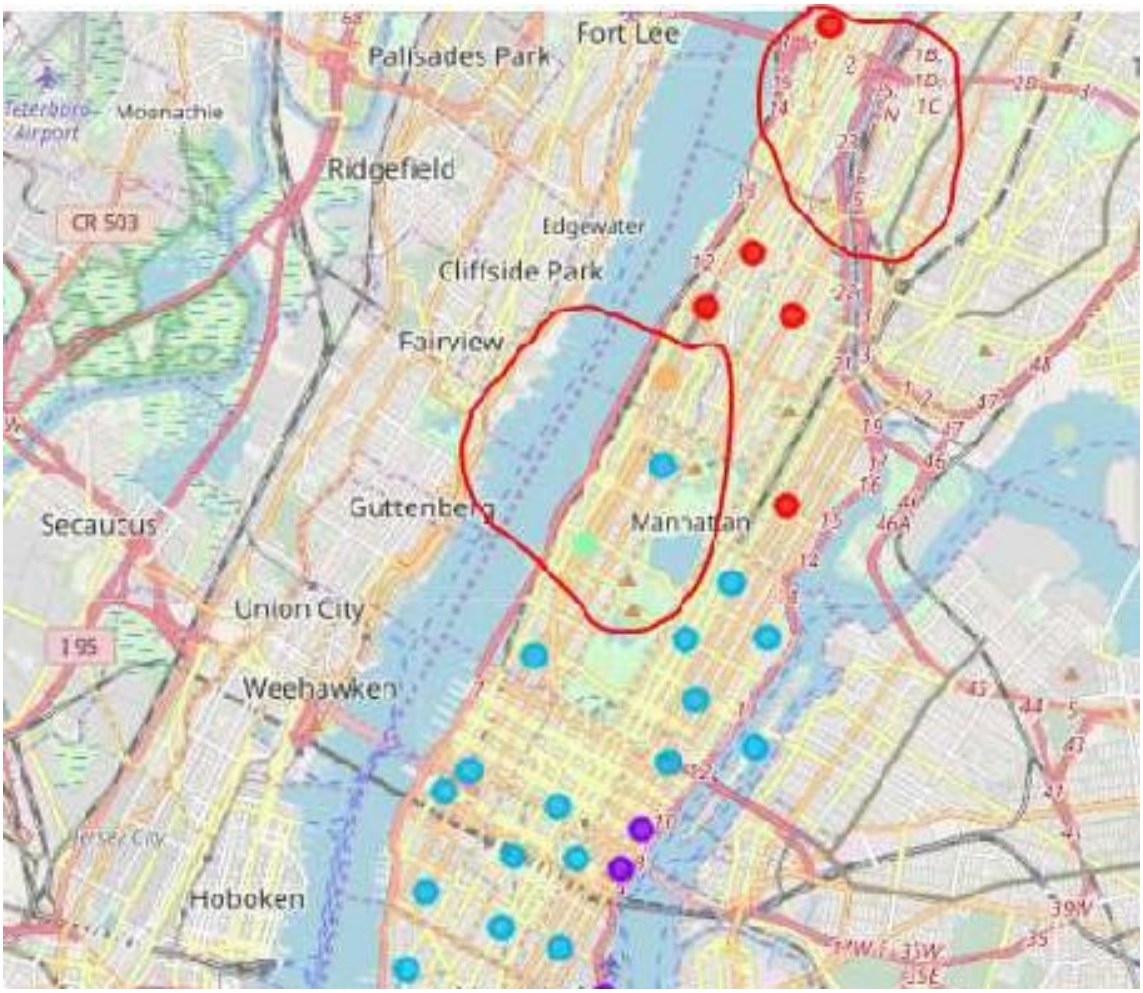
```
manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 3, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1
```

	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
12	Upper West	Sushi Restaurant	Japanese	Asian Restaurant	Grocery Store	Noodle House	Bakery	Chinese	Cocktail Bar

Cluster 4

```
manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 4, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[
```

	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
26	Morningside Heights	Sushi Restaurant	Hawaiian Restaurant	Vegetarian / Vegan Restaurant	Noodle House	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega



Based on dataframe analysis above Cluster 3 (Upper West Side) and Cluster 4 (Morningside Heights) areas are the best places to open a new sushi bar business.

Discussion

In this section, I would be discussing the observations I have noted and the recommendation that I can make based on the results.

This analysis is performed on limited data. This may be right or may be wrong. But if good amount of data is available there is scope to come up with better results.

·There is high competition in Midtown and Soho so it is very risky to open business in these areas.

·Central Harlem has also potential where closes to Morningside Heights area.

·It can be done more detailed analysis by adding other factors such as transportation, demographics of inhabitants.

Finally, FourSquare proved to be a good source of data but frustrating at times. Despite having a Developer account I regularly exceeded my hourly limit locking me out for the day.

Conclusion

Although all of the goals of this project were met there is definitely room for further improvement and development as noted below. However, the goals of the project were met and, with some more work, could easily be devleoped into a fully phledged application that could support the opening a business idea in an unknown location.

As per the neighbourhood or restaurant type mentioned like Sushi restaurants analysis can be checked. A venue with lowest risk and competition can be identified.