Determining the best location for opening a sushi bar in the Manhattan area of New York City

Introduction

New York…

is one of the few cities with an incredible mix of cuisines from around the world. Here, in every district, you can taste dishes of different nations.

In Little Italy, taste authentic Italian pizza, in Chinatown - duck soup and chicken with teriyaki sauce, in Harlem and the Bronx - Caribbean dishes. Well, and if you are bored of your usual food - welcome to Russian restaurants in Brighton Beach.

Most frequently "bought" dishes:

- local hot dogs. You can buy them literally at every step, and it will be a hearty and inexpensive snack;

- bagels. They are made from yeast dough, from which a ring is made, boiled, and then fried. Then add additional ingredients. For example, chocolate chips or sesame seeds. Bagels with cream cheese and a piece of salmon are very popular;

- the famous cheesecake. This New York symbol is prepared from cream cheese, vanilla, cream and cookies, and you can taste it at any local cafe.

From the above, we can conclude that the baking and fast food market in New York is the most oversaturated, in connection with which we will analyze the choice of a place in Manhattan to choose a place for a sushi bar.

Investors contacted us to invest in the restaurant business, in particular in the sushi bar. We will determine a potential location based on the number of sushi bars that operate in each area. Manhattan is relatively expensive and has a lot of potential. A new sushi bar should be opened in a Manhattan location so that the bar can attract a lot of customers. Hence, this analysis is necessary to make sure we have enough customers and that we are not that close to other sushi bars.

Data Selection

The dataset is available for free on the Internet at the dataset link: https://geo.nyu.edu/catalog/nyu_2451_34572

For convenience, the files have been downloaded and placed on the server, so you can simply run the wget command and access the data.

To determine the characteristics of our competitors establishments in Manhattan, we first need to find out the number of sushi bars in Manhattan, as well as their location. We then use the Google Map API to find their geographic coordinates based on their postal codes. Next, we use the Google Map API to find their geographic coordinates for several places (for example 3-5) shortlisted for our sushi bar.

The main libraries used for project development:

- Pandas: for creating and manipulating data frames.
- Scikit Learn: for importing k-means clustering.
- JSON: A library for handling JSON files.
- Beautiful Soup and Requests: Undo and a library for handling HTTP requests.

- Folium: A Python visualization library will be used to visualize the cluster distribution of neighborhoods using an interactive leaflet map.
- XML: for separating data from presentation, and XML stores data in text format.
- Geocoder: to get location data.
- Matplotlib: Python plotting module.

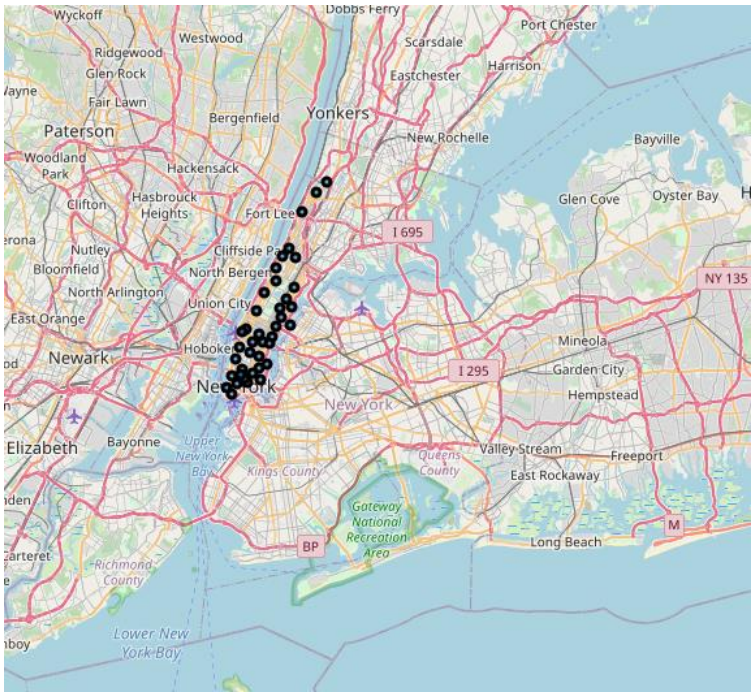
Methodology

Using geopy library to get the latitude and longitude values of New York City.

```
manhattan_data = neighborhoods[neighborhoods['Borough'] == 'Manhattan'].reset_index(drop=True)
manhattan_data.head()
```

	Borough	Neighborhood	Latitude	Longitude
0	Manhattan	Marble Hill	40.876551	-73.910660
1	Manhattan	Chinatown	40.715618	-73.994279
2	Manhattan	Washington Heights	40.851903	-73.936900
3	Manhattan	Inwood	40.867684	-73.921210
4	Manhattan	Hamilton Heights	40.823604	-73.949688

Then we create a map of New York with the neighborhoods superimposed on top.



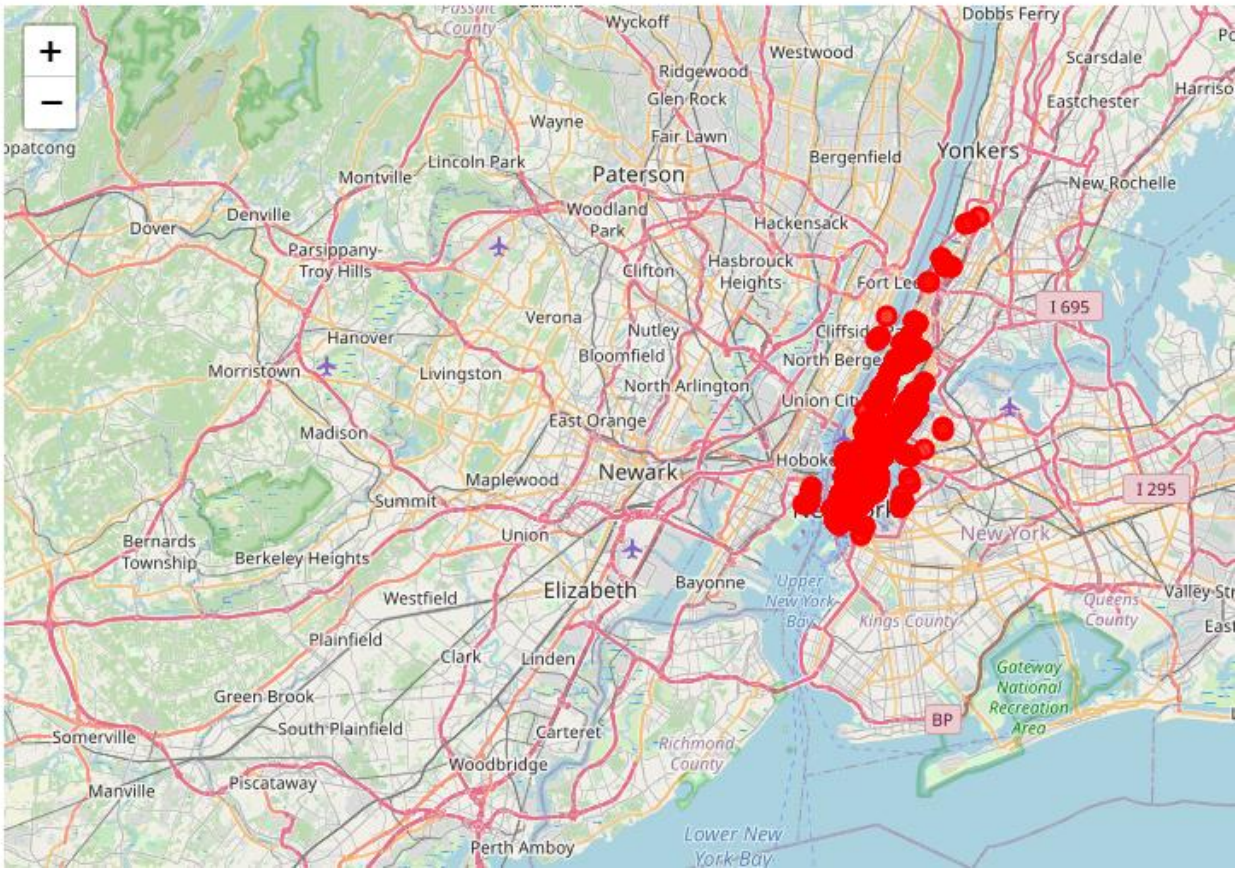
The Foursquare API is used to explore areas of Manhattan, New York. After that, explore the function to get the sushi restaurant categories in each area.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.91066	Planet Tokyo	40.886233	-73.909479	Sushi Restaurant
1	Marble Hill	40.876551	-73.91066	Yokohama	40.887214	-73.904708	Sushi Restaurant
2	Marble Hill	40.876551	-73.91066	Sushi Mambo	40.861031	-73.920435	Sushi Restaurant
3	Marble Hill	40.876551	-73.91066	Asian Tokyo	40.890839	-73.898335	Sushi Restaurant
4	Marble Hill	40.876551	-73.91066	Mama Sushi	40.866189	-73.927963	Sushi Restaurant

```
newyork_venues_sushi.shape
```

```
(1625, 7)
```

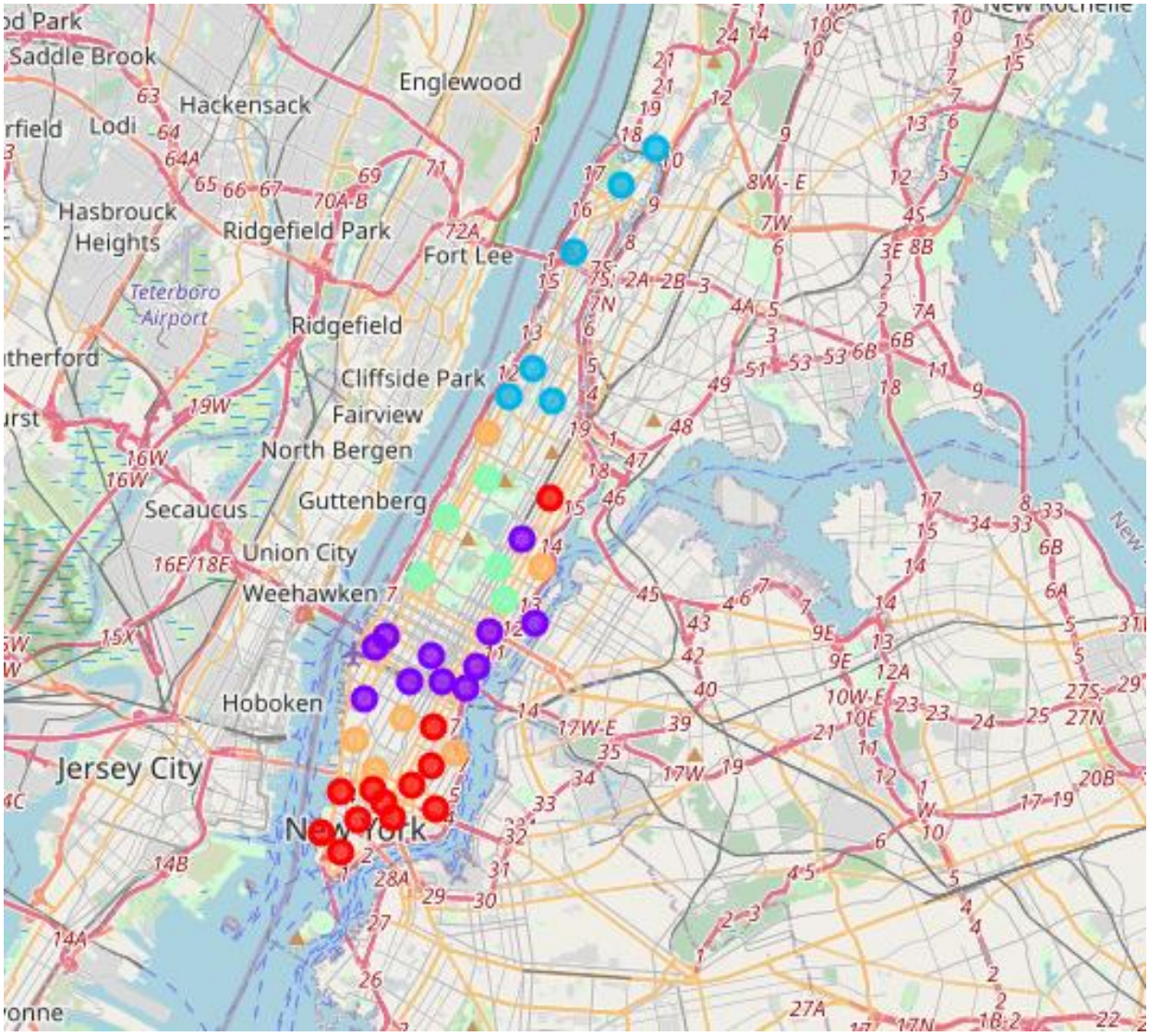
```
map_newyork_sushi = folium.Map(location=[latitude, longitude], zoom_start=10)
addToMap(newyork_venues_sushi, 'red', map_newyork_sushi)

map_newyork_sushi
```



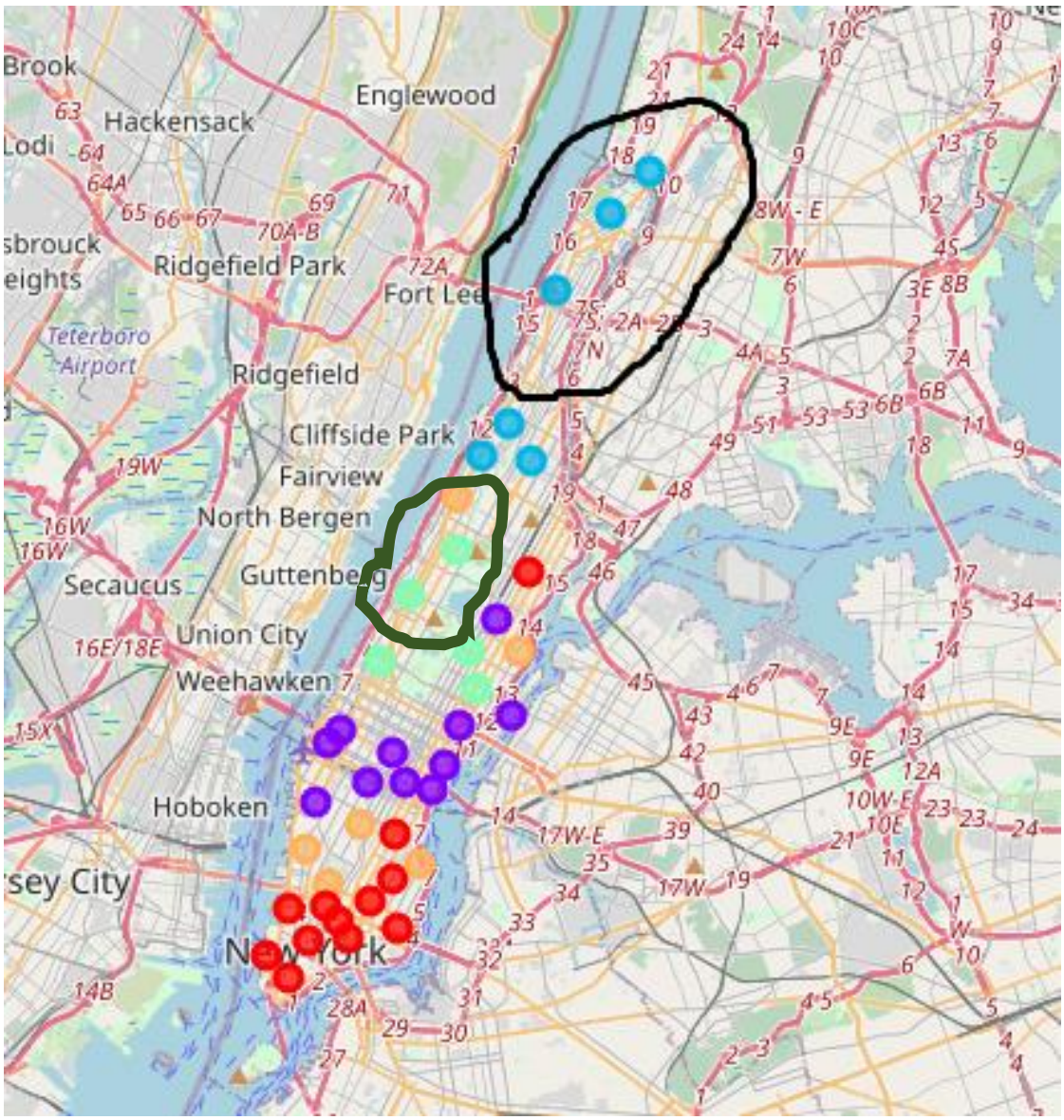
We analyze each area in Manhattan and after writing the necessary functions (in more detail in the file ipynb), we display the list in a tabular version.

	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	Sushi Restaurant	Japanese Restaurant	Noodle House	Asian Restaurant	Bakery	Chinese Restaurant	Deli / Bodega	Fish Market	Grocery Store	Hawaiian Restaurant
1	Carnegie Hill	Sushi Restaurant	Japanese Restaurant	Asian Restaurant	Fish Market	Indian Chinese Restaurant	Vegetarian / Vegan Restaurant	Bakery	Chinese Restaurant	Deli / Bodega	Grocery Store
2	Central Harlem	Sushi Restaurant	Asian Restaurant	Bakery	Chinese Restaurant	Deli / Bodega	Fish Market	Grocery Store	Hawaiian Restaurant	Indian Chinese Restaurant	Japanese Restaurant
3	Chelsea	Sushi Restaurant	Japanese Restaurant	Vegetarian / Vegan Restaurant	Asian Restaurant	Bakery	Fish Market	Ramen Restaurant	Chinese Restaurant	Deli / Bodega	Grocery Store
4	Chinatown	Sushi Restaurant	Japanese Restaurant	Noodle House	Asian Restaurant	Bakery	Chinese Restaurant	Deli / Bodega	Fish Market	Grocery Store	Hawaiian Restaurant

We cluster the neighborhoods, as previously mentioned, 3-5 clusters were planned. Let's take 5 and also visually create a map with cluster information overlaid (in different colors).



We display information on 5 clusters in the form of tables. And based on the analysis of the data above, cluster 2 is the best place to open a new sushi bar (circled in black).



Discussion

This analysis is performed on limited data. This may not be 100% certain, but if there is enough data, more accurate results can be obtained.

Clusters 0 and 1 have the highest competition, roughly speaking there are sushi bars at every turn, so opening a business in these areas is very risky.

Clusters 3 and 4 might be interesting if you consider doing a more detailed analysis, adding factors such as: population demographics, transportation, what kind of infrastructure, salary and weather statistics. Then, if the analysis shows a

sufficiently high probability of success, then it is possible to open a sushi bar only in that part of the map that is circled in dark green.

Conclusion

Despite the fact that all the goals of this project have been achieved, there is definitely room for further improvement and development of the project.

Thus, the project objectives have been achieved, and after some additional work, it can be developed into a full-fledged application, thanks to which it is possible to support the discovery of a business idea in an unknown location.