

CAPSTONE PROJECT REPORT
BATTLE OF NEIGHBORHOODS: NATURE'S PLATTER
Sowmya A

Table of Contents

1. Introduction
2. Data: Requirements & Collection
3. Methodology
4. Results
5. Discussion
6. Conclusion

1.Introduction to the Business Problem:

An exploratory analysis of Manhattan: for *Nature's Platter* to open its franchise.

Manhattan is the most densely populated city of New York; it is also described as the financial capital of New York and the world. As one can imagine, people in Manhattan have a demanding lifestyle, especially with the large section of people working with financial firms round the clock. With impending tasks at work and continuous time-crunch, it is difficult for this section of people to be particular and conscious of the food they have.

To aid people with such busy schedules in reducing the life-style issues by providing and encouraging them to go for better food choices, *Nature's Platter*, an organic eatery and grocery store is planning to open its branches in New York.

Nature's Platter currently focuses on targeting the locations around financial institutions to contribute and cater to the health needs of the people working in that sector. Hence, it wants to explore the Manhattan's financial firms' localities to study and examine the eateries around the firms, to recognize the commonly visited eateries and to find out if there are any other organic eateries already in place. The store's management also wants to visualize this data to enable them have a better understanding of their target areas and eateries around them.

Hence this study is for the *Nature's Platters'* management to help them with understanding the locations, eateries around the financial firms.

2.Data: Requirements, Collection & Preparation:

- Location Data of New York and Manhattan – To visualize and explore the locations and cluster them. We can get this from the json file from https://geo.nyu.edu/catalog/nyu_2451_34572.
- Venues data in Manhattan (like banks, financial firms and eateries around) – from Foursquare.

The New York neighborhood dataset is extracted as a json from the above location. The features key in the file has the relevant data, this data is extracted from the file. Next, this data is transformed into a pandas dataframe, i.e, transforming the nested dictionaries' data into a pandas dataframe, namely *neighborhoods*. The *neighborhoods* dataframe is created such that only the relevant location data i.e, Borough, Neighborhood, Latitude, and Longitude information is contained. From this neighborhood data, the data of Manhattan is extracted and a separate dataframe, is created with the neighborhood, latitude and longitude information. Thus, the data is cleaned and prepared for further procedure.

Borough	Neighborhood	Latitude	Longitude

Fig.2.1 Dataframe after data collection and cleaning

3.Methodology:

With the aforementioned business understanding, data analysis is carried out with the data science methodology in view. The stakeholder for this project (*Nature's Platter* team) is interested in searching for the financial firms in Manhattan and exploring the neighborhood eateries.

As for the *analytic approach* to solve this problem, segmenting and clustering the eateries around the financial firms can be done using the machine learning algorithm – *k-means clustering*.

Segmentation is the practice of partitioning data points of interest (eateries around financial firms) into groups of entities that have similar characteristics. It is a significant strategy as a business can target these specific groups for study, observation and effectively allocate marketing resources.

Clustering is segmentation of unsupervised data into clusters based on similarity of objects in the cluster and dissimilarity to data points in other clusters. The *k-means clustering*, despite its simplicity, is vastly used for clustering in many data science applications, especially useful if you need to quickly discover insights from unlabeled data.

3.1 Exploratory Data Analysis:

The geographical co-ordinates of Manhattan are found out using geopy library of Python. With the Manhattan location information, we plot the Manhattan map using Folium library. Next step in this analysis is to search and find out the financial firms in Manhattan, as *Nature's Platter* wants to target the people working in these firms as their customer base.

3.1.1 Extract the financial firms in Manhattan:

We use Foursquare API location provider to explore the financial firms in Manhattan's neighborhood. The list of financial firms is obtained as a json file, by providing the inputs as the search *url*, *foursquare credentials*, *radius* to look with in, as shown below in Fig 3.1.

```
url='https://api.foursquare.com/v2/venues/search?client_id={}&client_secret={}&ll={},{}&v={}&query={}&radius={}&limit={}'
CLIENT_ID, CLIENT_SECRET, latitude, longitude, VERSION, search_query, radius, LIMIT)

{'meta': {'code': 200, 'requestId': '5cb662691ed2196d9210b5fd'},
 'response': {'venues': [{ 'id': '4e5465301f6e7ab6b1b28def',
  'name': 'Fidelity Investment',
  'location': {'address': '1356 3rd Ave',
  'lat': 40.77302882997516,
  'lng': -73.95793558189008,
  'labeledLatLngs': [{ 'label': 'display',
  'lat': 40.77302882997516,
  'lng': -73.95793558189008}],
  'distance': 1905,
  'postalCode': '10075',
  'cc': 'US',
  'city': 'New York',
  'state': 'NY',
```

Fig.3.1. List of financial firms using Foursquare API

The json file is cleaned and is structured into a dataframe with the name, category and geographic co-ordinates of the financial firms. In our dataset, around 50 firms are returned by Foursquare API within a radius of 10000 m. Now the financial firms are visualized on the Manhattan map using Folium library, as shown below in Fig. 3.2

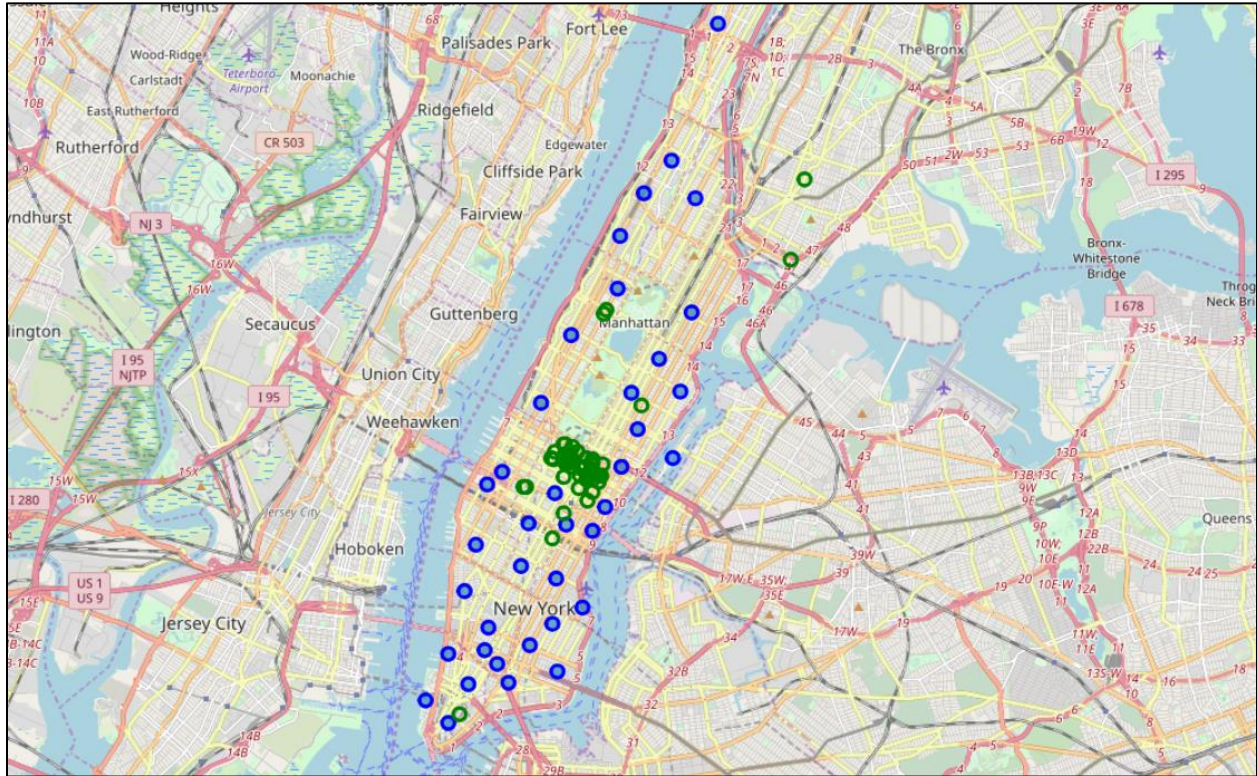


Fig.3.2. Financial Firms in Manhattan (the green circles) along with neighborhoods in Manhattan (blue circles)

The next step is to explore the eateries around these financial firms. A function is defined to extract the eateries around of the above firms along with the name of the eatery, eatery category, and its geographical co-ordinates. This data, along with corresponding financial firm's data is structured into a dataframe that looks as below.

	Name	Firm Latitude	Firm Longitude	Eatery	Eatery Latitude	Eatery Longitude	Eatery Category
0	Fidelity Investment	40.773029	-73.957936	Ateaz Organic Coffee & Tea	40.775124	-73.953813	Café
1	Fidelity Investment	40.773029	-73.957936	Duke Eatery	40.756797	-73.975107	Food & Drink Shop
2	Fidelity Investment	40.773029	-73.957936	Cock & Bull British Pub and Eatery	40.755946	-73.980621	Pub

Fig.3.3. The *finfirms_eateries* dataframe with the eateries' details around each financial firm

We further explore the data, grouping by the financial firms, to find out the eateries around each firm. There are approximately 36 unique categories of eateries that we obtain from the above exercise.

3.1.2. Explore the eatery categories around each financial firm

Now analyze the *eatery categories* in the vicinity of each firm. This is done by converting the categorical variable data into numeric values by creating dummies that indicate the presence of eateries in the firms' vicinity. This data is further analyzed by determining the frequency of occurrence/presence

of each category of eatery around each firm. Thus, we obtain the frequency of the category for each firm. This is one of the parameters the Nature's Platter team is keen on determining for their study.

Furthermore, the details of the 5 most common eatery categories around each firm, based on the frequency, are obtained.

----1919 Investment Counsel----		
	venue	freq
0	Café	0.13
1	Restaurant	0.09
2	Mediterranean Restaurant	0.09
3	Deli / Bodega	0.09
4	Salad Place	0.09
----AC Investment Management----		
	venue	freq
0	Café	0.14
1	Restaurant	0.10
2	Mediterranean Restaurant	0.10
3	Deli / Bodega	0.10
4	Food & Drink Shop	0.05
----Al Rayyan Tourism Investment C.		
	venue	freq
0	Café	0.13
1	Restaurant	0.09
2	Mediterranean Restaurant	0.09
3	Deli / Bodega	0.09
4	Salad Place	0.09

Fig.3.4. Five most common eateries around each financial firm

The 10 most common eatery categories around each financial firm is also structured into a dataframe, as shown below, for further analysis.

	Name	1st Most Common Eatery	2nd Most Common Eatery	3rd Most Common Eatery	4th Most Common Eatery	5th Most Common Eatery	6th Most Common Eatery	7th Most Common Eatery	8th Most Common Eatery	9th Most Common Eatery	10th Most Common Eatery
0	1919 Investment Counsel	Café	Salad Place	Restaurant	Mediterranean Restaurant	Deli / Bodega	American Restaurant	Korean Restaurant	Food & Drink Shop	Food Truck	Health & Beauty Service
1	AC Investment Management	Café	Restaurant	Mediterranean Restaurant	Deli / Bodega	American Restaurant	Burger Joint	Food & Drink Shop	Food Truck	Health & Beauty Service	Korean Restaurant
2	Al Rayyan Tourism Investment Company	Café	Salad Place	Restaurant	Mediterranean Restaurant	Deli / Bodega	American Restaurant	Korean Restaurant	Food & Drink Shop	Food Truck	Health & Beauty Service
3	Apollos Partners NYC office (Executive Search ...	Food Truck	Fruit & Vegetable Store	Café	Market	Deli / Bodega	Asian Restaurant	BBQ Joint	Breakfast Spot	Burger Joint	Business Service
4	Aquamarine Investment Partners	Deli / Bodega	American Restaurant	Café	Mediterranean Restaurant	Food & Drink Shop	Food Truck	Health & Beauty Service	Korean Restaurant	Burger Joint	Breakfast Spot

Fig.3.5 Ten most common eatery categories around each financial firm

3.1.2. Explore the organic eateries around each financial firm

The organic eateries around the firms are extracted from the *finfirms_eateries* dataframe, and are explored similar to the process followed for the eatery categories in section 3.1.1. The most common organic eateries in each of the firm's neighborhood is obtained.

Name	1st Most Common Org Eatery	2nd Most Common Org Eatery	3rd Most Common Org Eatery	4th Most Common Org Eatery	5th Most Common Org Eatery	6th Most Common Org Eatery	7th Most Common Org Eatery	8th Most Common Org Eatery	9th Most Common Org Eatery	10th Most Common Org Eatery
1919 Investment Counsel	Ateaz Organic Coffee & Tea	Bean & Bean Organic Coffee	Smokey Burger Organic	Chelsea Organic	Organic Mexican Food Truck	The Organic Salad	Vspot Organic	Jahlookova Natural Organic Health Mart	Ali Baba Organic Marketplace	Central Park Organic Deli Grocery
AC Investment Management	Ateaz Organic Coffee & Tea	Bean & Bean Organic Coffee	Smokey Burger Organic	Chelsea Organic	Organic Mexican Food Truck	The Organic Salad	Vspot Organic	Jahlookova Natural Organic Health Mart	Ali Baba Organic Marketplace	Central Park Organic Deli Grocery
Al Rayyan Tourism Investment Company	Ateaz Organic Coffee & Tea	Bean & Bean Organic Coffee	Smokey Burger Organic	Chelsea Organic	Organic Mexican Food Truck	The Organic Salad	Vspot Organic	Jahlookova Natural Organic Health Mart	Ali Baba Organic Marketplace	Central Park Organic Deli Grocery
Apollos Partners NYC office (Executive Search ...	NYC Healthy Bites - Organic Hot Dogs	Ateaz Organic Coffee & Tea	Central Park Organic Deli Grocery	Creative Organic Foods Store	Organic Fruit Shakes & Smoothies	Jahlookova Natural Organic Health Mart	Mom & Pop's Organic Market	Ali Baba Organic Marketplace	Bean & Bean Organic Coffee	Chelsea Organic
Aquamarine Investment Partners	Ateaz Organic Coffee & Tea	Bean & Bean Organic Coffee	Smokey Burger Organic	Chelsea Organic	Organic Mexican Food Truck	Olives Organic Market	The Organic Salad	Vspot Organic	Jahlookova Natural Organic Health Mart	Ali Baba Organic Marketplace

Fig. 3.6 Ten most common organic eateries around each financial firm

3.2 K-means Clustering:

Clustering is one of the unsupervised machine learning algorithms. Segmentation is one of the popular uses of clustering. Cluster is a group of objects that are similar to other objects in the group and dissimilar to the objects/data points in the clusters. Clustering is used mainly in exploratory data analysis, summary generation, outlier detection.

K-means clustering is one of the vastly used partition-based clustering algorithms. It can be used for medium and large databases. The main objective of k-means clustering is to minimize the distance of each datapoint to the centroid of its cluster and maximize the distance from other clusters. This divides data into k non-overlapping subsets without any internal cluster structure. The farther apart the clusters are placed, the better.

In our current application of the k-means clustering, we use the financial firms, along with frequency of each eatery category, as the data for clustering the firms. Here, we cluster the firms into 5 clusters. *K-means clustering* is done by using the *k-means* algorithm from the scikit-learn library of the Python libraries. Let the cluster labels generated be 0,1,2,3,4.

Now a new dataframe is created with the financial firms, their geographical co-ordinates, 10 most common eatery categories against each financial firm, and label of the cluster that each firm belongs to.

4. Results

Now that we have the clusters of the financial firms, we visualize these clusters on the map using folium library. To this map, we also add the locations of the organic eateries, that we earlier identified during our exploratory data analysis. The groups of circle markers in different colors as red, sear, purple, lightblue, lightorange depict the clustered financial firms. The scattered brown circles depict the organic eateries around the financial firms that we located in our study.

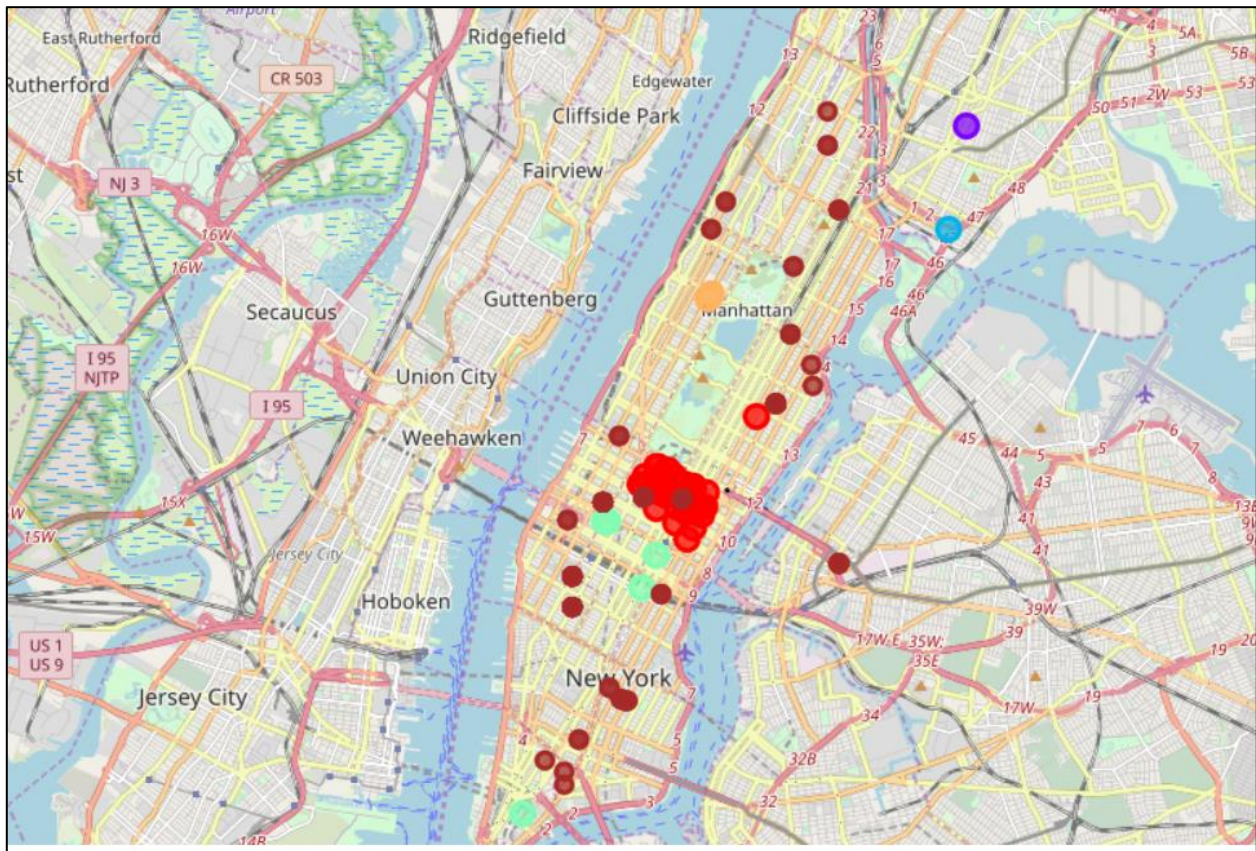


Fig. 4.1 Clusters of financial firms and the organic eateries (brown circles) around the firms.

5. Discussion

From the above data analysis exercise, we understand the distribution of the financial firms in Manhattan and realize the spread of different categories of eateries around each of the firms. It also gives a reasonable understanding of the most commonly visited eateries of each firm. The organic eateries around each firm are also identified, the most commonly visited organic eateries are identified for each firm. This is crucial in the study of the market area and the comprehension of the target audience preferences. This helps the Nature's Platter's team in their further market study and research.

As seen in the above map, we find the organic eateries are not mainly concentrated in the clustered firms' areas, but are scattered around. This enables the Nature's Platter team to decide on the location in which they can open their franchise. From the exploratory data analysis, we also find that the most commonly visited organic eateries are cafes, quick eat places like burger joints. This is also an

important piece of information that helps in deciding on the variety of menu items to be included. The other important observation is that around 20 organic eateries are presently located around the financial firms that we focused on. This means that the Nature's Platter team would need to be distinctive in its appeal and platter to appease the target audience.

6. Conclusion

In this study, I surveyed Manhattan for financial firms, explored eateries that are distributed around the firms. I further, specifically identified the organic eateries of the lot and investigated the most commonly visited organic eateries, along with their respective locations. The visualization of the clusters of financial firms and organic eateries on Manhattan map is further beneficial in giving the Nature's Platter team a quick and easy understanding of the data analysis. This is quite helpful to the Nature's Platter team in a number of ways. For example, in their market research and also enables them to decide on their business expansion strategy in Manhattan, targeting the financial firms.