Analysing the Cost Of Living & Venue Data of New York City

A. Introduction:

New York City is located on the coast of the North-eastern United States which is composed of five boroughs. While Manhattan and Staten Island are islands, Brooklyn and Queens are geographically part of Long Island, and the Bronx is attached to the US mainland. Quit often comes in all our mind and populous city as well geographically located in well surrounded city is New York which I have chosen for my analysis of cost of living.

A.1. Description & Discussion of the Background

As we know that this city has more than 8 Million people live and famous restaurants, café and multi versatile region restaurants are available. Also people used to visit frequently from other places from both US and outside of US. When we think of this frequent movement of people as lot of industrial business is developed and out of US people were living as a bachelor for a long time especially from IT industry. So there is more demand of identify the rental stay, restaurant, transport and other utilities with different category of pricing in this city. Specifically sharing accommodation and food will be the major expenses for most of the people and they can't afford to invest to have own residential. However it is a challenge to get the information about the cheapest / medium cost of getting an accommodation & food to reduce their price or within the limit of their affordability along with which geographical location for their day to day business.

For addressing all of these problems, thought of providing information to utilize by them where they can choose the accommodation place of neighbourhood in New York City. Also analysing of an average living cost trend and identifying the fluctuation in each neighbourhood of New York City to compare with the cost of living index.

A.2. Data Collection & Cleansing:

- Foursquare API is used to get the most common venues of New York.
- I have extracted the geographical data of New York from Spatial Data Repository. The .json file has coordinates of the all city of New York. I cleaned the data for using to create choropleth map of plotting the neighbourhood of New York.
- I used Google Map, 'Search Nearby' option to get the centre coordinates of the each Borough.
- There are not too many public datasets related to cost of living calculator, indexes and parameters for the city of New York. Therefore you must set-up

your own data tables in most cases. In this case, I collected cost of living price per month dataset along with coordinates of each Borough of New York.

B. Audience

Most of the People are interested to fine tune and trying to reduce their expenses by staying sharing accommodation at New York City. And the below category of people are more rigorous to identify the accommodation and restaurant with affordable cost.

- Want to Staying at New York City for Business traveller
- Frequently travelling and interested to visit New York City
- Long terms business stayer

C. Testing Strategy

- I have taken a few random sample values of latitude and longitude to plot it into the map for making sure that it is working fine to confirm.
- And identified the missing values and followed the process of substituting mean value of cost of living. For the latitude and longitude vales are compared with the already extracted .json file value and substituted.
- Outliers are identified and adjusted the value to bring them into within range of cost of living for make sure that they are clustered. Also some of them are rejected because of avoiding variance in the prediction.

D. Methodology:

I used GitHub repository to store and maintaining the relevant datasets. The cost of living (COL) per month price dataset has been extracted and manipulated which has the main components Borough, Neighbourhood, COL Price, latitude and longitude of New York City. Cost of living per month value is converted for each Neighbourhood of New York City.

Below is the table of COL per month data frame values:

	Borough	Neighborhood	Latitude	Longitude	COL Per Month
0	Bronx	Wakefield	40.894705	-73.847201	947.05
1	Bronx	Co-op City	40.874294	-73.829939	742.94
2	Bronx	Eastchester	40.887556	-73.827806	875.56
3	Bronx	Fieldston	40.895437	-73.905643	954.37
4	Bronx	Riverdale	40.890834	-73.912585	908.34

I used python folium library to visualize geographic details of New York and its boroughs. And I created a geographical map of New York with boroughs on top with the standard background colour of blue. I used latitude and longitude values to get the visual as below:
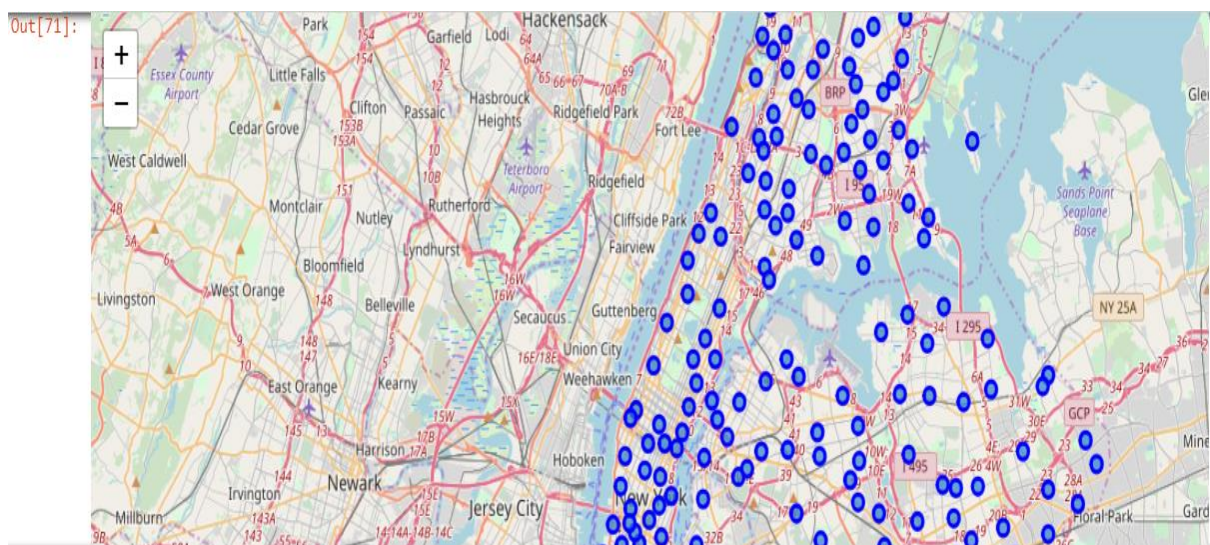
I utilized the Foursquare API to explore the boroughs and segment them. I designed the limit as 100 venues and the radius 500 meter for each borough from their given latitude and longitude information. Here is a head of the list Venues name, category and latitude and longitude information's from Foursquare API.

We have got 9 venues and 426 unique categories of venues from .json file. We have analysed each neighbourhood location of cost of living and grouping then with respect to category of important places.

Below table is showing count of venue category available for each neighbourhood.

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Allerton	27	27	27	27	27	27
Annadale	9	9	9	9	9	9
Arden Heights	5	5	5	5	5	5
Arlington	5	5	5	5	5	5
Arrochar	20	20	20	20	20	20
Arverne	15	15	15	15	15	15
Astoria	100	100	100	100	100	100
Astoria Heights	13	13	13	13	13	13
Auburndale	20	20	20	20	20	20
Bath Beach	46	46	46	46	46	46
Battery Park City	99	99	99	99	99	99

Below are the top 10 most common places are available in each neighbourhood of New York City. It will be useful to identify the restaurant, pizza place, café, coffee shop and geo-regional restaurant.
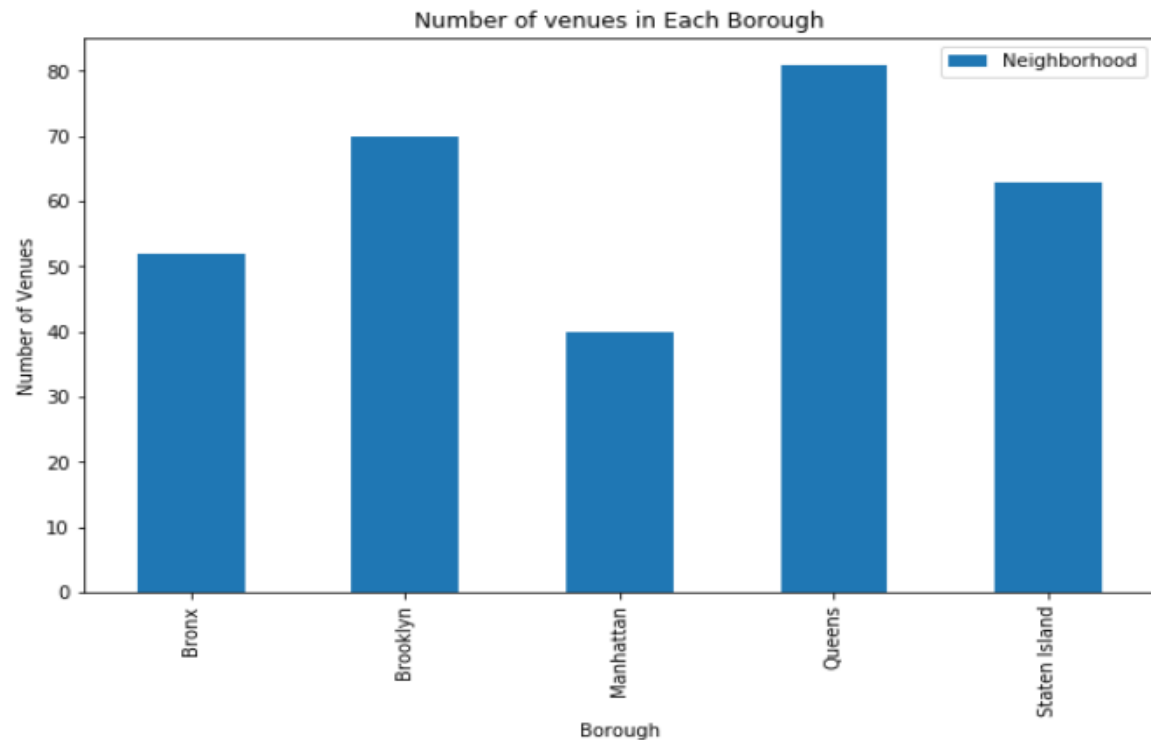
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Allerton	Supermarket	Bakery	Liquor Store	Spa	Pizza Place	Spanish Restaurant	Playground	Chinese Restaurant	Fast Food Restaurant	Martial Arts Dojo
1	Annadale	Pizza Place	American Restaurant	Sports Bar	Dance Studio	Restaurant	Park	Diner	Train Station	Farmers Market	Farm
2	Arden Heights	Deli / Bodega	Pizza Place	Coffee Shop	Bus Stop	Pharmacy	Women's Store	Field	Event Space	Exhibit	Eye Doctor
3	Arlington	Coffee Shop	American Restaurant	Boat or Ferry	Bus Stop	Grocery Store	Women's Store	Fish & Chips Shop	Eye Doctor	Factory	Falafel Restaurant
4	Arrochar	Bus Stop	Italian Restaurant	Pizza Place	Deli / Bodega	Mediterranean Restaurant	Liquor Store	Sandwich Place	Taco Place	Bagel Shop	Hotel

The below bar graph will shows the average cost of living based on the Borough in New York City. The range is between 970$ per month to ~1110 $ for Queens & Staten Island Borough respectively. Based on the average value between these Borough to choose the accommodation which resulted to optimize the cost of living of ~150$ per month.
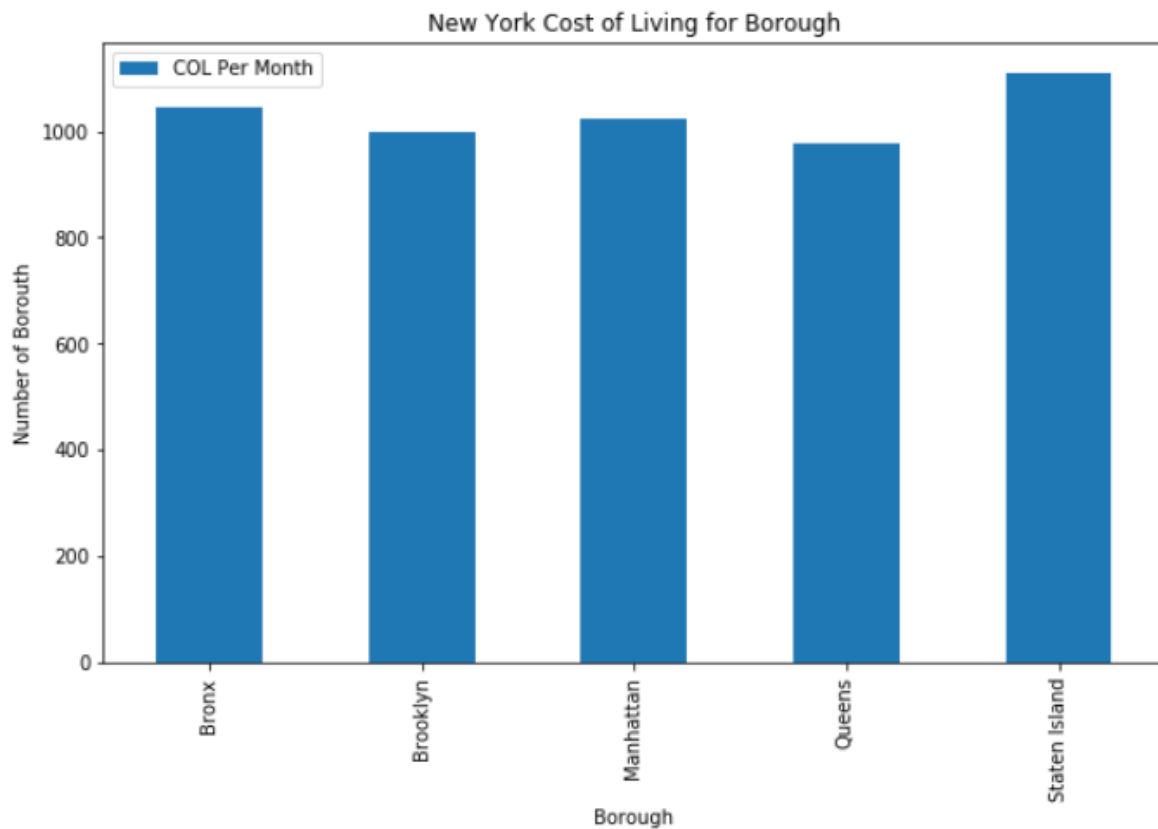
The below graph is showing the Brooklyn and Queens Borough has more than 70 venues and especially Manhattan Boroughs has around 40 venues in our given coordinates with Latitude and Longitude, in below graph.

The result doesn't mean that inquiry run all the possible results in boroughs. Actually, it depends on given Latitude and Longitude information and here is we just run single Latitude and Longitude pair for each borough. We can increase the possibilities with Neighbourhood information with more Latitude and Longitude information.

Number of venues in Each Borough

We have some common venue categories in boroughs. In this reason I used unsupervised learning K-means algorithm to cluster the boroughs. K-Means algorithm is one of the most common cluster methods of unsupervised learning.

First, I will run K-Means to cluster the boroughs into 5 clusters because when I analyse the K-Means with elbow method it ensured me the 3 degree for optimum k of the K-Means. Here, I have included cluster labels for each borough for 305 venues. And the result has been merged with the Master table to include the Cost of Living value per month.

	Borough	Neighborhood	Latitude	Longitude	COL Per Month	ClusterLabels	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 Commo Venu
0	Bronx	Wakefield	40.894705	-73.847201	947.05	0	Sandwich Place	Caribbean Restaurant	Pharmacy	Laundromat	Ice Cream Shop	Dessert Shop	Foo
1	Bronx	Co-op City	40.874294	-73.829939	742.94	4	Bus Station	Mattress Store	Ice Cream Shop	Pharmacy	Park	Grocery Store	Pizza Plac
2	Bronx	Eastchester	40.887556	-73.827806	875.56	4	Caribbean Restaurant	Bus Station	Diner	Chinese Restaurant	Fast Food Restaurant	Deli / Bodega	Pizza Plac
3	Bronx	Fieldston	40.895437	-73.905643	954.37	0	River	Playground	Plaza	Women's Store	Field	Event Service	Event Spac
4	Bronx	Riverdale	40.890834	-73.912585	908.34	0	Park	Food Truck	Bus Station	Gym	Home Service	Playground	Ban
5	Bronx	Kingsbridge	40.881687	-73.902818	816.87	4	Pizza Place	Sandwich Place	Bar	Discount Store	Mexican Restaurant	Latin American Restaurant	Supermarke

When we examine above graph we can label each cluster as follows:

- Cluster 0 : "Clothing, Sports & Utilities Venues"
- Cluster 1 : "Social Entertainment Venues"
- Cluster 2 : "Accommodation & Rental Venues"

- Cluster 3 : "Restaurant Venues"
- Cluster 4 : "Transportation Services Venues"

We can also examine that what is the frequency of average cost of living prices in different ranges. Thus, histogram can help to visualization:

Histogram of Average COL for 5 Borough in New York
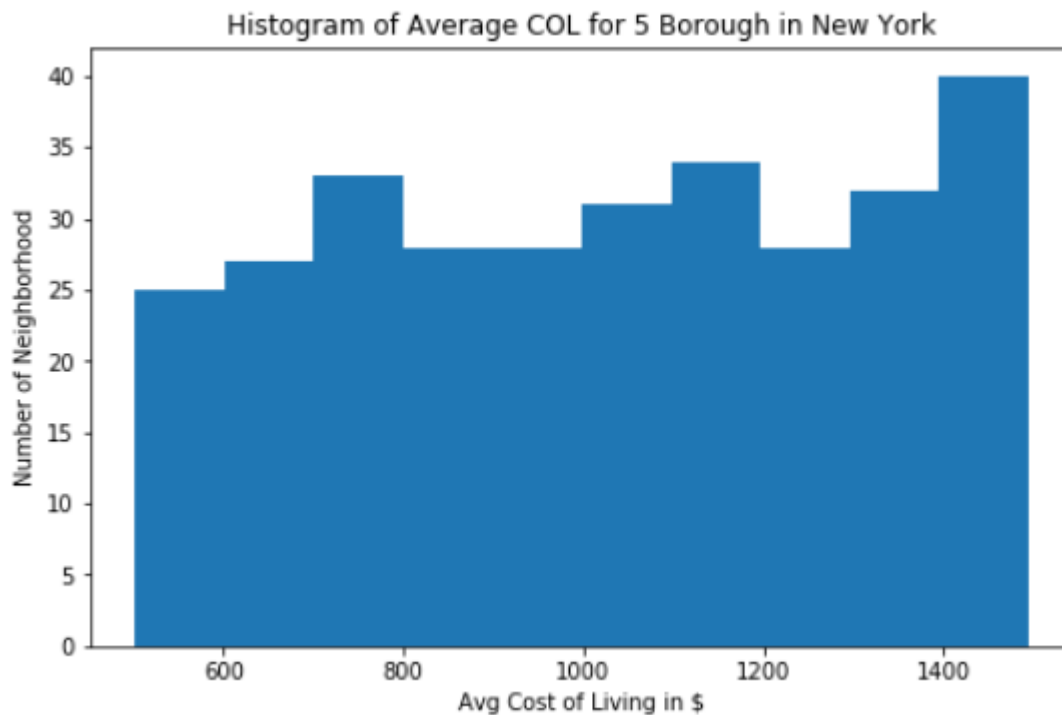
As it seems in above histogram, we can define the ranges as below: again, Cost of living range will be varying dynamically based on the location as an exception.

- 600 COL : "Low Level COL"
- 600–800 ACOL : "Avg-1 Level COL"
- 800–1000 ACOL : "Avg-2 Level COL"
- 1000–1200 HCOL : "High-1 Level COL"
- 1200–1400 HCOL : "High-2 Level COL"
- > 1400 COL : "High-3 Level COL"

Also plotted those cluster categories in different colour of circle marker in the below map.
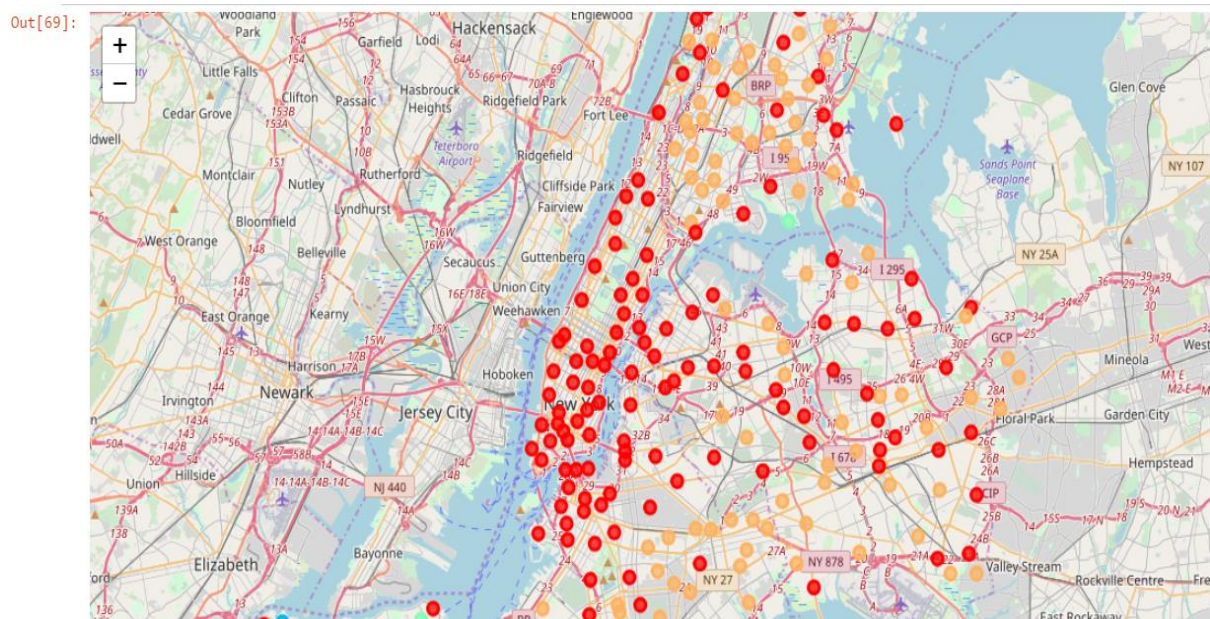
E. Results:

In summary, one of the goals was to visualize the location with average Cost of Living thru choropleth style map. Thus, first I downloaded a json file of the New York from Spatial Data Repository. I cleaned the json file and pull out only city of New York.

People can visualize and choose the affordable price of spending per month can be achievable using this information. In final section, I created choropleth map which also has the below information for each borough:

- Borough name,
- Cluster name,
- Cost of Living (COL) Levels,
- Top 3 number of venue
- Bottom 3 number of venue
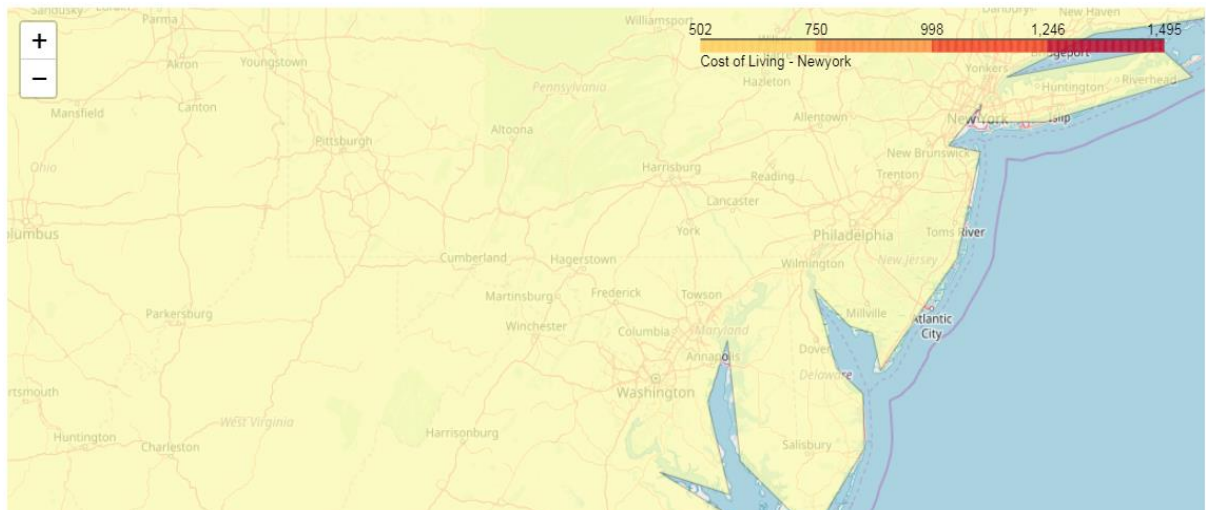
F. Discussion:

As I mentioned before, New York is North-eastern United States which is composed of five boroughs. I have approached very differently in clustering and classification studies at best. Moreover, it is obvious that not every classification method can yield the same high quality results for this New York City.

I used the KMeans algorithm as part of this clustering study. When I tested the Elbow method, I set the optimum k value to 5. However, only 305 different neighbourhood coordinates were used. For more detailed and accurate guidance, the data set can be expanded and the details of the neighbourhood or street can also be drilled.

I also performed data analysis through this information by adding the coordinates of districts and cost of living averages as static data on GitHub. In future studies, these data can also be accessed dynamically from specific platforms or packages.

I have learned strategy of study by visualizing the data and clustering information on the New York map. In future studies, we will use the distance as measures to identify shortest nearby location where can we get the rental services, restaurant with the affordable price of cost of living to web or telephone applications can be carried out to direct visitors, frequent travellers and other business purposes.

G. Conclusion:

As a result, people are frequently travelling to any big cities to choose the affordable place of rental accommodation, transport for day to day business or work and restaurants such as mandatory for them. For this reason, people can achieve better outcomes through their access to the platforms where such information is provided.

Not only for visitors but also long term stayers and citizenships can manage the city more regularly by using similar data analysis types or platforms.

H. References:

- New York — Wikipedia
- Foursquare API
- Google Map
- Captured the below thru Google about New York
    - Cost of living index in New York is approx. 27% higher.
    - Cost of living rank 12th out of 400+ cities in the world.
    - New York has a cost of living index of 100.00

Distribution of expenses using our statistical model

10.5%
25.6%
48.1%

- Transportation
- Clothing And Shoes
- Sports And Leisure
- Markets
- Utilities (Monthly)
- Rent Per Month
- Restaurants