

New York - City Cost of Living Analysis

Analysis of 5 Borough and Neighborhood of New York City

Analysis for finding out the Cost of Living is more important for New York City

Cost of living (COL) can be categorized as ordered below:

The following levers are very important to estimate the Cost of living in any location

- Rent Per month - Accommodation
 - Food Expenses
 - Transport & Marketing
 - Clothing, Sports and other Utilities
-
- Rent and Food expenses are played major role in calculating the COL.
 - Identifying the nominal price for these levers especially rent will help to control our expenses and increase the saving.
 - To improve the confident level and indirectly motivate to invest to get reasonable returns.
 - People who are long terms stayer, Frequent traveler and Business people to that city is showing more interest to analysis

Data Collection and Cleansing

- Location data like Borough, longitude, latitude and neighborhood data (2018) from Foursquare API for categorizing location based on rental & restaurant.
- After cleansing, 29 features were considered for analysis.
- Cost of living dataset is referred from Kaggle data and manipulate it for New York city
- Near by locations are identified and grouped them.
- Totally 12400+ locations and 429 features are in the raw dataset
- Missing values, outliers and highly correlated values are handled.



Location Identification & Mapping

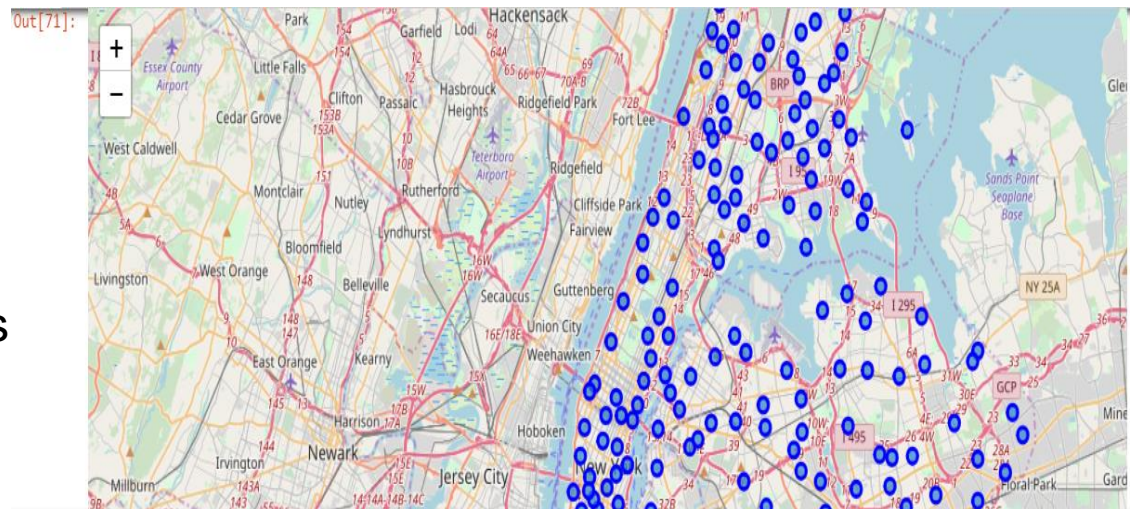
Location details along with COL and plot them in Geographical map

New York – Borough & Neighborhood

	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

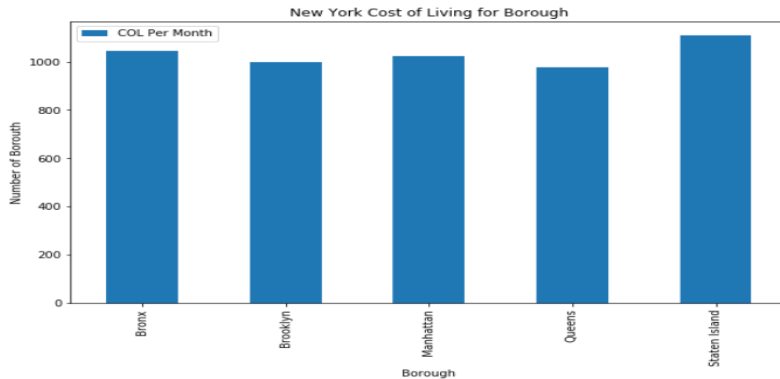
- Merging with the master dataset to cost of living.
- Plotted them in a map for New York city.

- 5 Borough and 305 neighborhood.
- Got 9 venues and 426 unique categories of venues from Foursquare API.



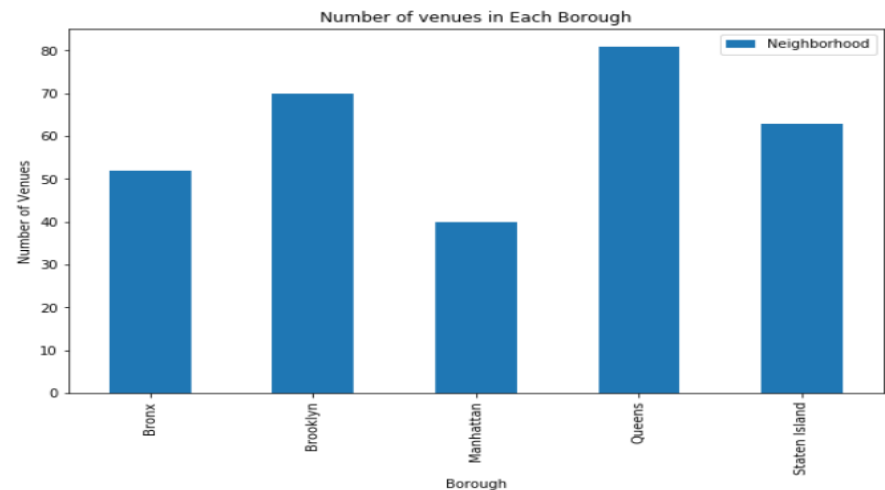
Cost of Living Vs. Borough

New York – Borough & Neighborhood



- All 5 Borough and neighborhood cities.
- Total 305 unique cities were considered and grouped them.

- Analysis of cost of living Vs. Borough.
- Range is between 970\$ & ~1110 \$ for Queens & Staten Island Borough .



Most Common Venues Vs. Cost of Living

Identified most common venues and Comparing cost of living for New York City

	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 Common Venue
0	Bronx	Wakefield	40.894705	-73.847201	947.05	0	Sandwich Place	Caribbean Restaurant	Pharmacy	Laundromat	Ice Cream Shop	Dessert Shop	Food
1	Bronx	Co-op City	40.874294	-73.829939	742.94	4	Bus Station	Mattress Store	Ice Cream Shop	Pharmacy	Park	Grocery Store	Pizza Place
2	Bronx	Eastchester	40.897556	-73.827806	875.56	4	Caribbean Restaurant	Bus Station	Diner	Chinese Restaurant	Fast Food Restaurant	Deli / Bodega	Pizza Place
3	Bronx	Fieldston	40.895437	-73.905643	954.37	0	River	Playground	Plaza	Women's Store	Field	Event Service	Event Space
4	Bronx	Riverdale	40.890834	-73.912585	908.34	0	Park	Food Truck	Bus Station	Gym	Home Service	Playground	Bar
5	Bronx	Kingsbridge	40.881687	-73.902818	816.87	4	Pizza Place	Sandwich Place	Bar	Discount Store	Mexican Restaurant	Latin American Restaurant	Supermarket

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

Cluster 0 : “Clothing, Sports & Utilities Venues”

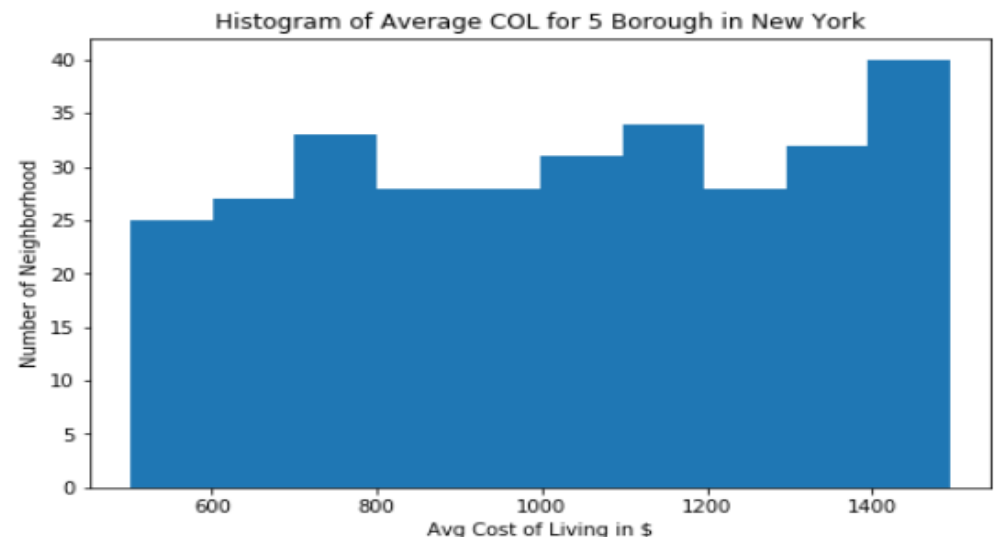
Cluster 1 : “Social Entertainment Venues”

Cluster 2 : “Accommodation & Rental Venues”

Cluster 3 : “Restaurant Venues”

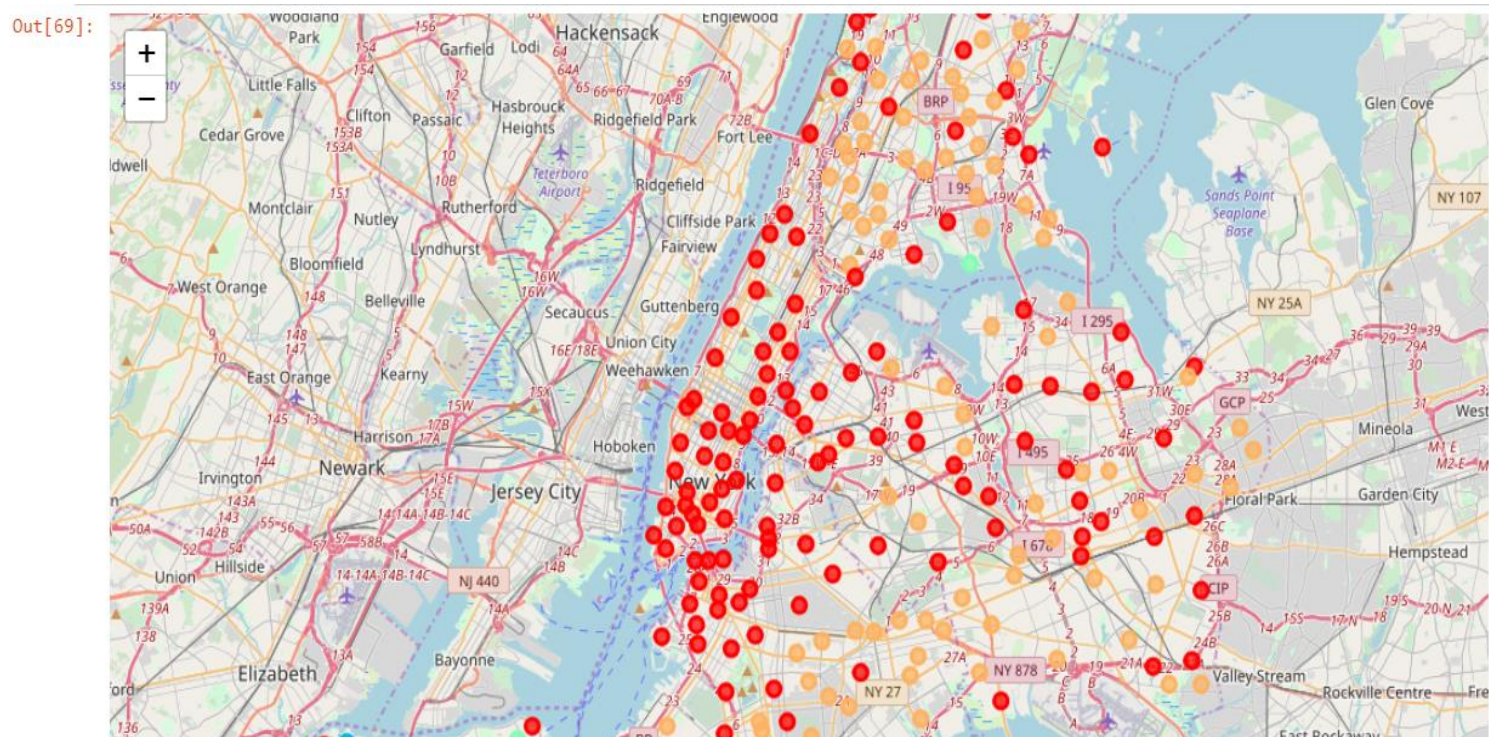
Cluster 4 : “Transportation Services Venues”

- Cost of living range will be varying dynamically based on the location
- 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”



Clustering and KMean is 5

5 Clustering of these locations and plot them into map with different color. It will be showing top 3 and bottom 3 cost living locations are identified by the people to choose based on their affordability.



Choropleth Map

Minimum and maximum cost of living price showing in the choropleth map for each clustering.



Conclusion & Future Enhancement

Conclusion

Future Enhancement

High Cost of living location

- People preferred who travelling frequently can afford and own transport

- Built models to include more parameters & volume of locations dataset

Medium Cost of living location

- People preferred staying with family & give importance in terms of security.

- Accuracy of the model for enhancement

Low Cost

- People who stay for employment, students and having less affordability.

- Financial condition of the people, nearest location and transport data