Final Report On
Most Favorable Cities Near New York for Living

INTRODUCTION

The principle point of this Project is to help individuals in investigating better offices for living around New York urban communities. It will help individuals settling on shrewd and effective choice on choosing extraordinary spot out of quantities of different neighborhoods in New York, USA.

Loads of individuals are relocating to different conditions of America and required bunches of examination at great lodging costs and reputated schools for their kids. This undertaking is for those individuals who are searching for better areas. For simplicity of getting to Cafe, School, Super market, clinical shops, staple shops, shopping center, theater, clinic, similar individuals, and so on.

This Project mean to make an investigation of highlights for a people moving to New York to look through a best neighborhood as a similar examination between neighborhoods. The highlights incorporate middle lodging cost and better school as indicated by evaluations, crime percentages of that specific region, street network, climate conditions, great administration for crisis, water assets both freash and waste water and fertilizer passed on in sewers and recreational offices.

It will help individuals to get attention to the region and neighborhood prior to moving to another city, state, nation or spot for their work or to begin another new life.

COLLECT DATA

We will require information about various settings in various neighborhoods of that particular district. To pick up that data we will utilize "Foursquare" locational data. Foursquare is an area information supplier with data pretty much all way of settings and functions inside a region of interest. Such data incorporates scene names, areas, menus and even photographs. Accordingly, the foursquare area stage will be utilized as the sole information source since all the expressed required data can be acquired through the API.

Subsequent to finding the rundown of neighborhoods, we at that point associate with the Foursquare API to accumulate data about scenes inside every single area. For every area, we have picked the range to be 100 meter.

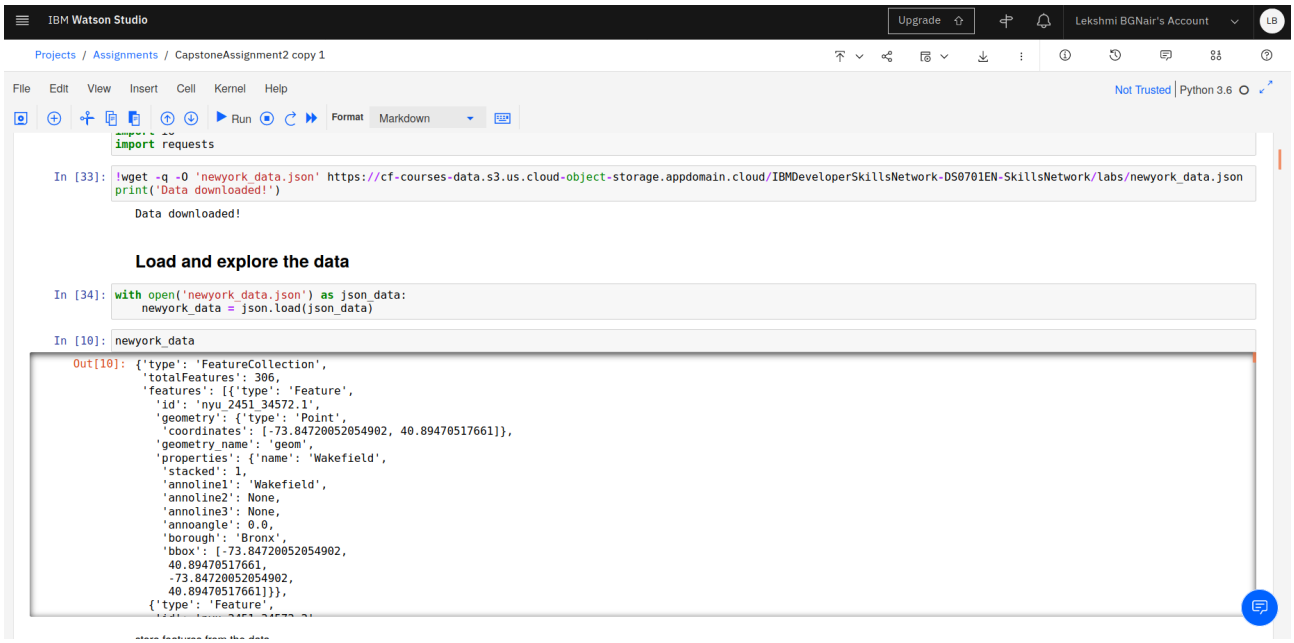
The information recovered from Foursquare contained data of scenes inside a predefined separation of the longitude and scope of the postcodes. The data acquired per setting as follows:

1. Neighborhood
2. Neighborhood Latitude
3. Neighborhood Longitude
4. Venue
5. Name of the venue e.g. the name of a store or restaurant
6. Venue Latitude
7. Venue Longitude
8. Venue Category

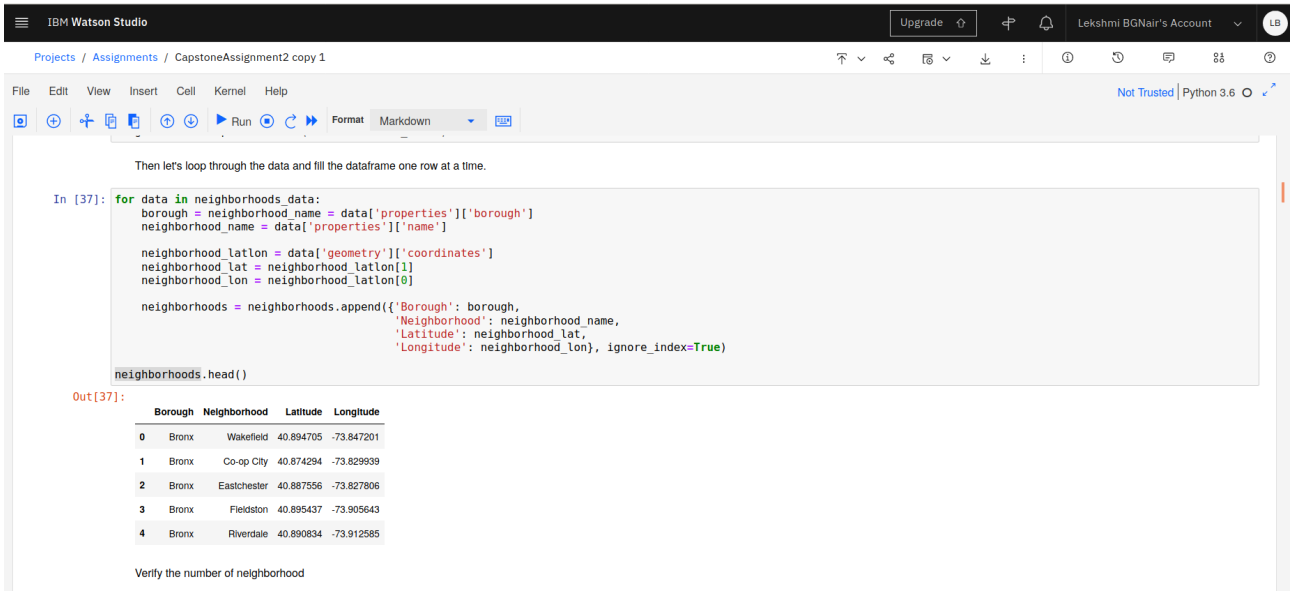
Methodology

To look at the likenesses of two urban communities, we chose to investigate neighborhoods, section them, and gathering them into groups to discover comparable neighborhoods in a major city near New York. To have the option to do that, we have to bunch information which is a type of unsupervised clustering algorithm: kNN clustering.

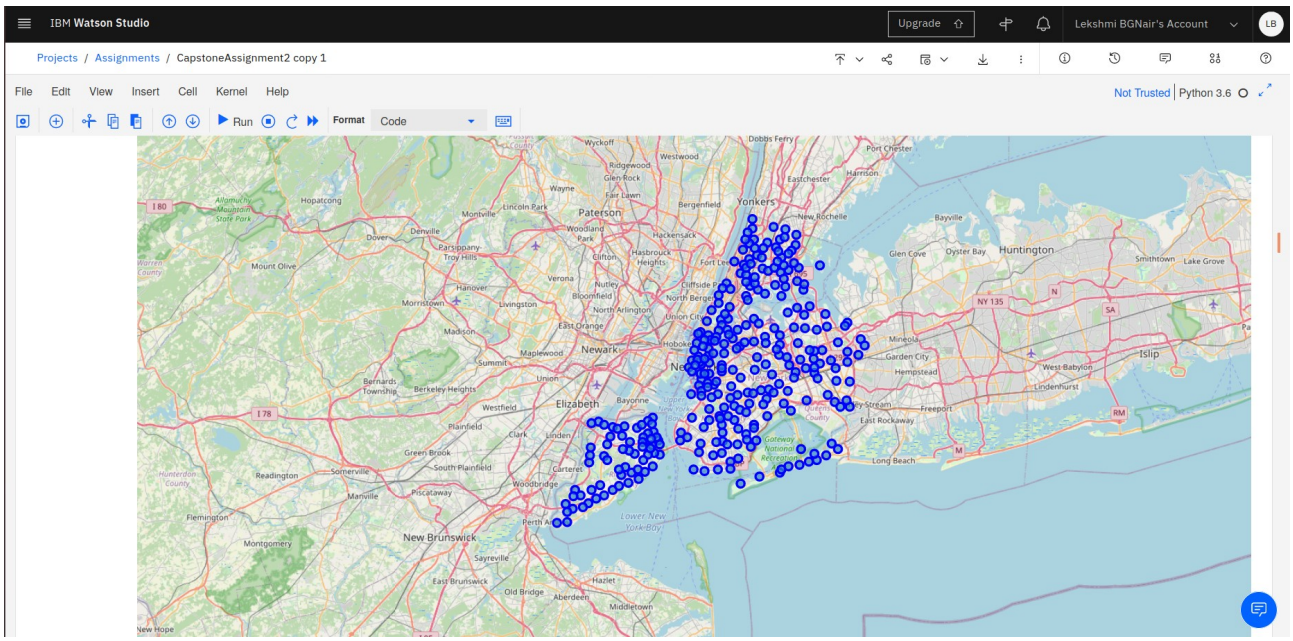
1.collect the data



2. Find the location parameters of the cities

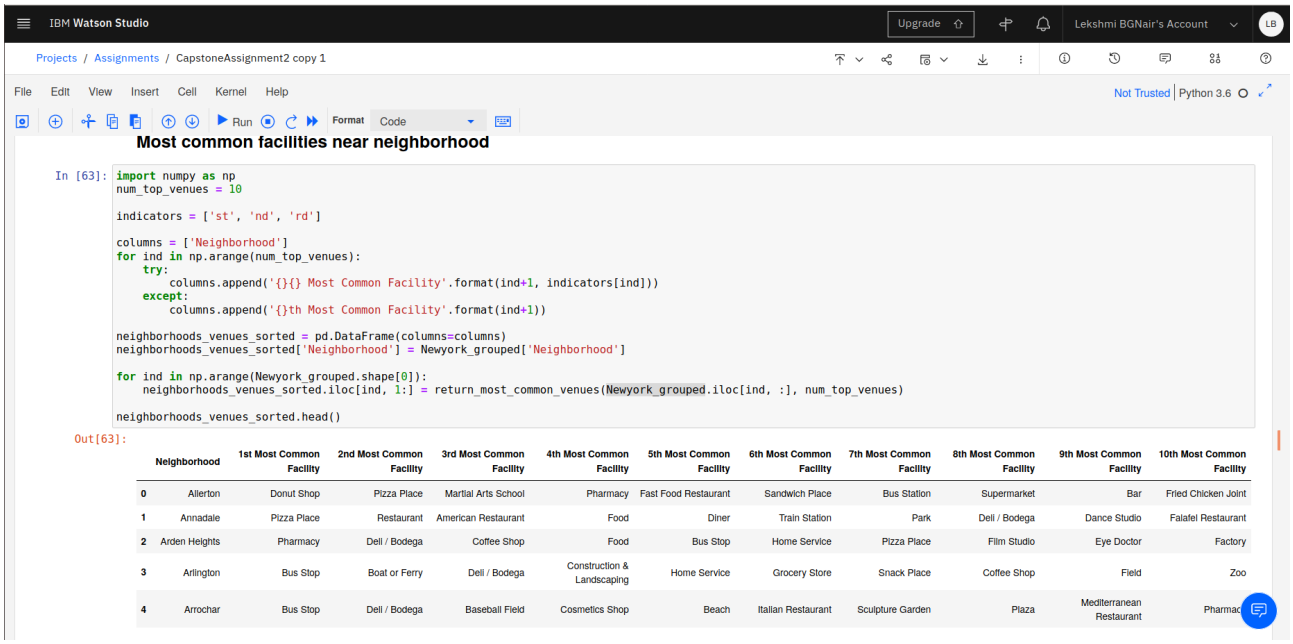


3.Find the location of New York city



4.Find the facilities available in all those cities

Using credentials of Foursquare API features of near-by places of the neighborhoods would be mined. Due to http request limitations the number of places per neighborhood parameter would reasonably be set to 100 and the radius parameter would be set to 500.



In [63]:
```
import numpy as np
num_top_venues = 10

indicators = ['st', 'nd', 'rd']

columns = ['Neighborhood']
for ind in np.arange(num_top_venues):
    try:
        columns.append('{}{} Most Common Facility'.format(ind+1, indicators[ind]))
    except:
        columns.append('{}th Most Common Facility'.format(ind+1))

neighborhoods_venues_sorted = pd.DataFrame(columns=columns)
neighborhoods_venues_sorted['Neighborhood'] = Newyork_grouped['Neighborhood']

for ind in np.arange(Newyork_grouped.shape[0]):
    neighborhoods_venues_sorted.iloc[ind, 1:] = return_most_common_venues(Newyork_grouped.iloc[ind, :], num_top_venues)

neighborhoods_venues_sorted.head()
```

Out[63]:

	Neighborhood	1st Most Common Facility	2nd Most Common Facility	3rd Most Common Facility	4th Most Common Facility	5th Most Common Facility	6th Most Common Facility	7th Most Common Facility	8th Most Common Facility	9th Most Common Facility	10th Most Common Facility
0	Allerton	Donut Shop	Pizza Place	Martial Arts School	Pharmacy	Fast Food Restaurant	Sandwich Place	Bus Station	Supermarket	Bar	Fried Chicken Joint
1	Annadale	Pizza Place	Restaurant	American Restaurant	Food	Diner	Train Station	Park	Deli / Bodega	Dance Studio	Falafel Restaurant
2	Arden Heights	Pharmacy	Deli / Bodega	Coffee Shop	Food	Bus Stop	Home Service	Pizza Place	Film Studio	Eye Doctor	Factory
3	Arlington	Bus Stop	Boat or Ferry	Deli / Bodega	Construction & Landscaping	Home Service	Grocery Store	Snack Place	Coffee Shop	Field	Zoo
4	Arrochar	Bus Stop	Deli / Bodega	Baseball Field	Cosmetics Shop	Beach	Italian Restaurant	Sculpture Garden	Plaza	Mediterranean Restaurant	Pharmac

5.Next clustering neighbors with similar facilities

IBM Watson Studio

Upgrade | Lekshmi BGNair's Account

Projects / Assignments / CapstoneAssignment2 copy 1

File Edit View Insert Cell Kernel Help

Not Trusted | Python 3.6

Run | Format: Code

```
2, 0, 2, 0, 0, 2, 0, 2, 0, 2, 0, 0, 0, 2, 2, 2], dtype=int32)
```

In [75]:
```
Newyork_merged =neighborhoods.iloc[:16,:]

# merge toronto_grouped with toronto_data to add latitude/longitude for each neighborhood
Newyork_merged = Newyork_merged.join(neighborhoods_venues_sorted.set_index('Neighborhood'), on='Neighborhood')

Newyork_merged.head()
```

Out[75]:

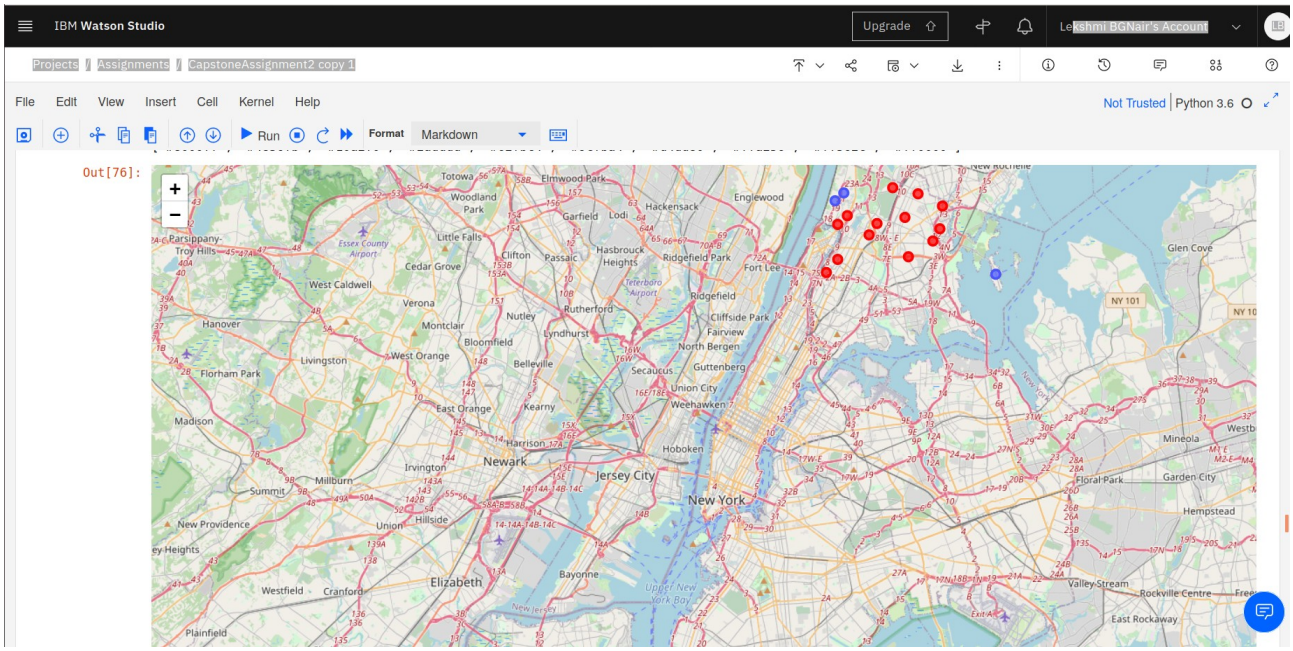
	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Facility	2nd Most Common Facility	3rd Most Common Facility	4th Most Common Facility	5th Most Common Facility	6th Most Common Facility	7th Most Common Facility	8th Most Common Facility	9th Most Common Facility	10th Most Common Facility
0	Bronx	Wakefield	40.894705	-73.847201	0	Gas Station	Pharmacy	Fried Chicken Joint	Caribbean Restaurant	Bakery	Fast Food Restaurant	Gift Shop	Pizza Place	Sandwich Place	Donut Shop
1	Bronx	Co-op City	40.874294	-73.829939	0	Accessories Store	Fast Food Restaurant	Bus Station	Print Shop	Men's Store	Bagel Shop	Harbor / Marina	Dumpling Restaurant	Trail	Grocery Store
2	Bronx	Eastchester	40.887556	-73.827806	0	Caribbean Restaurant	Fast Food Restaurant	Diner	Pizza Place	Burger Joint	Donut Shop	Shopping Mall	Grocery Store	Seafood Restaurant	Cocktail Bar
3	Bronx	Fieldston	40.895437	-73.905643	2	Plaza	Medical Supply Store	Home Service	Art Gallery	Moving Target	River	Park	Zoo	Fast Food Restaurant	Field
4	Bronx	Riverdale	40.890834	-73.912585	2	Bank	Park	Japanese Restaurant	Pizza Place	Mexican Restaurant	Playground	Italian Restaurant	Health & Beauty Service	Bagel Shop	Bar

In [69]:
```
kclusters = 10
```

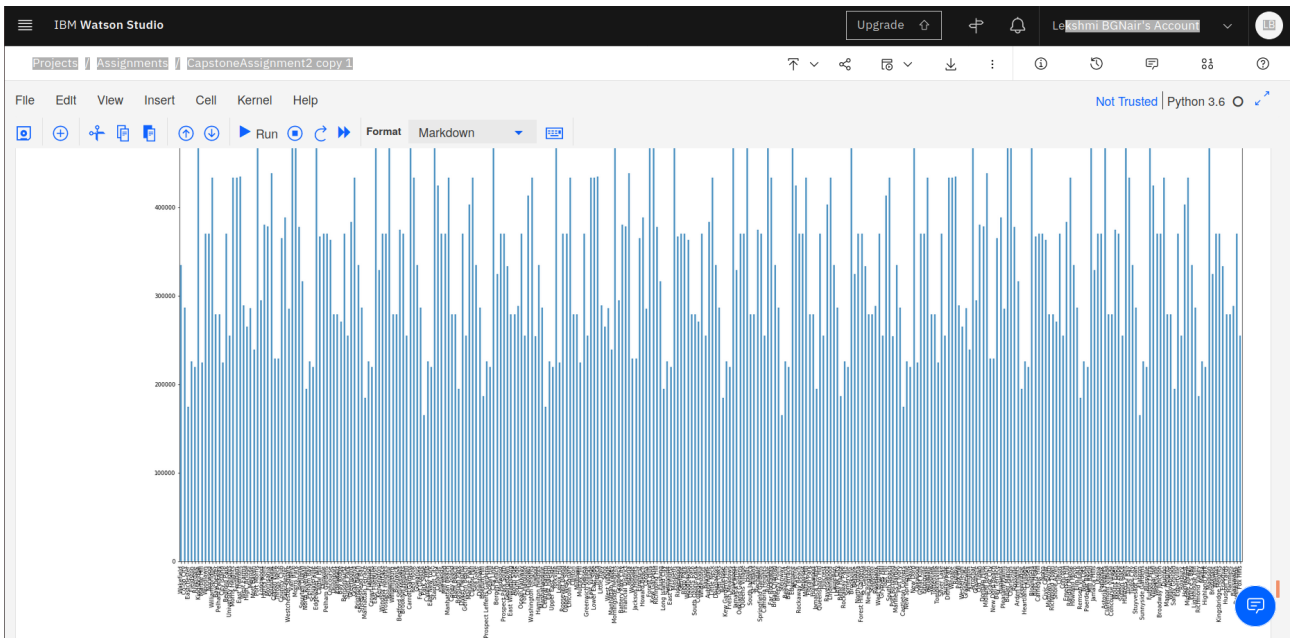
In [76]:
```
# create map
map_clusters = folium.Map(location=[latitude, longitude], zoom_start=11)

# set color scheme for the clusters
```

RESULTS

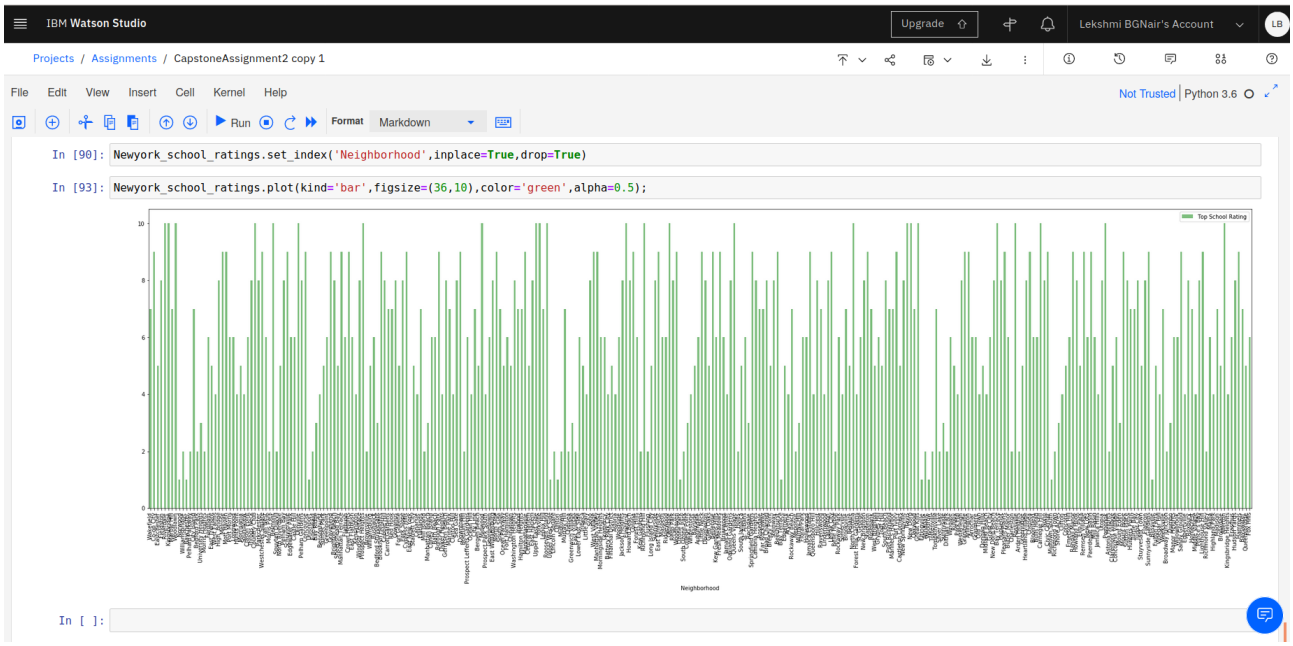
Map of clusters of NewYork cities are shown below



Average housing price per cities



School rating by clusters



Conclusion

In this report, utilizing k-means clustering isolated the neighborhood into 100 various groups and for 306 distinctive lattitude and logitude from dataset, which have fundamentally the same as neighborhoods around them. Utilizing the graphs above outcomes introduced to a specific area dependt on normal house costs and school rating have been made.

This task has indicated me a useful application to determine a genuine circumstance that has affecting individual and budgetary effect utilizing Data Science devices. The planning with Folium is a ground-breaking procedure to solidify data and settle on the investigation and choice better with certainty.

Future Works:

This task can be proceeded for making it more exact in wording to discover best house in NewYork. Best methods based on all required things(daily requires or things we have to carry on with a superior life) around and furthermore as far as savvy.