Introduction

The City of New York, is the most populous city in the United States. It is diverse and is the financial capital of USA. It is multicultural. It provides lot of business oppourtunities and business friendly environment. It has attracted many different players into the market. It is a global hub of business and commerce. The city is a major center for banking and finance, retailing, world trade, transportation, tourism, real estate, new media, traditional media, **SPORTS**, advertising, legal services, accountancy, insurance, theater, fashion, and the arts in the United States.

This also means that the market is highly competitive. As it is highly developed city so cost of doing business is also one of the highest. Thus, any new business venture or expansion needs to be analysed carefully. The insights derived from analysis will give good understanding of the business environment which help in strategically targeting the market. This will help in reduction of risk. And the Return on Investment will be reasonable.

Business problem

The City of New York has a huge population and major part of the population is young people and most of them are intrested in playing sports .With the increase in health conciousness people are more intrested in playing sports .So there is a great need of sports facility in the city. This is a good investment if one establish the facility in the right place. So we will try to find a best place to establish a stadium or a sport facility.

Find a suitable location

To be a bussiness successful ,we have to find a best place to establish. Here we are trying to find the best palce for a stadium or a sport facilty .Establishing a sport facility in a place wherethere is a compition is very risky,and also we should look to find a best place to find a place where there is a demand for sports and a place to find a econmical and risk free place to find profits.

Target Audience:

My Client wants to open his business in NewYork , so I only focus on that borough during my analysis. The objective is to locate and recommend to the management which neighborhood of Newyork city will be best choice to start a sports facility.

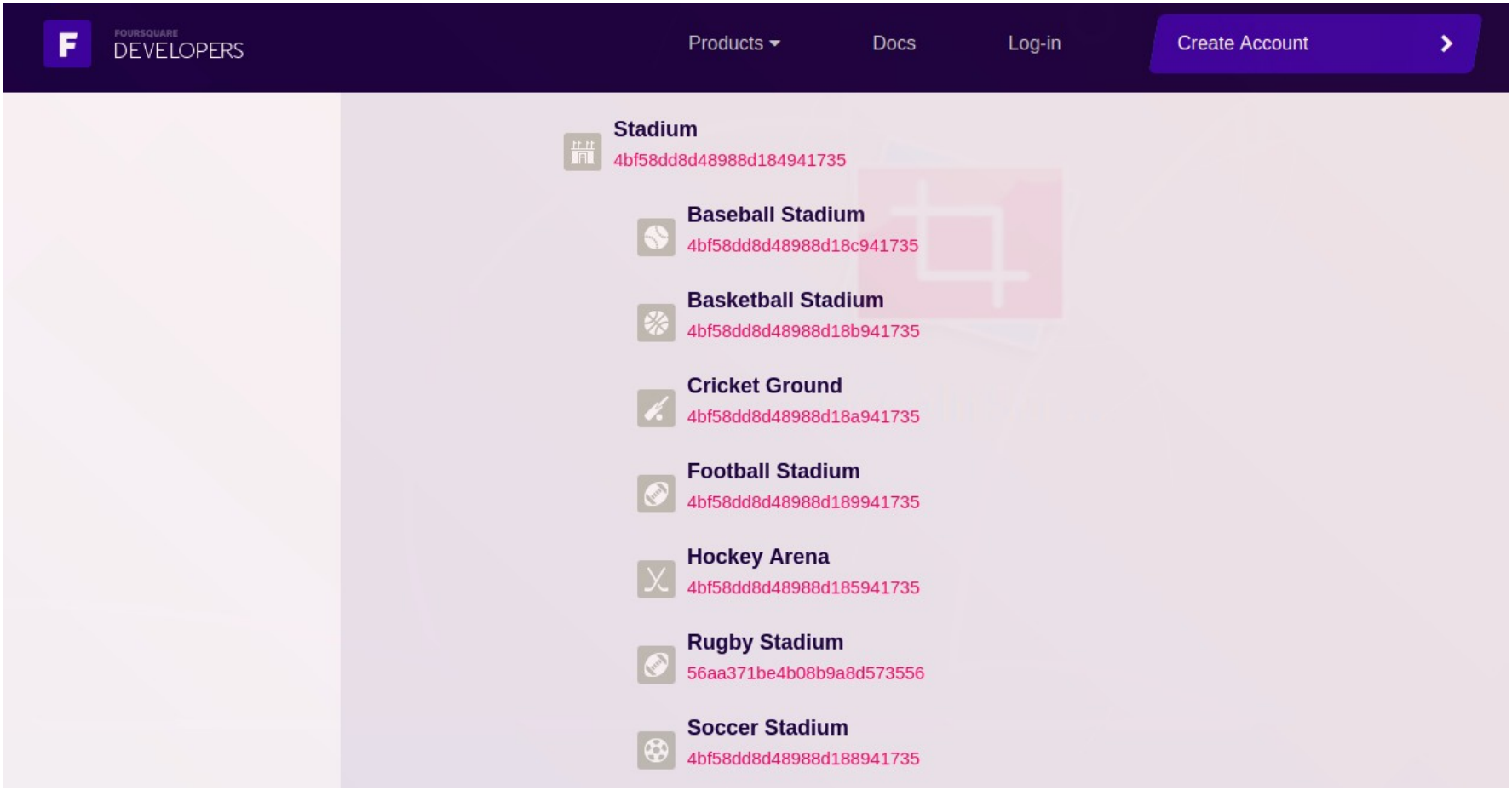
This would intrest people who wants to start a sports facility in NewYork.

Data

Data 1: Neighborhood has a total of 5 boroughs and 306 neighborhoods. In order to segement the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the the latitude and logitude coordinates of each neighborhood. This dataset exists for free on the web. Link to the dataset is: https://geo.nyu.edu/catalog/nyu_2451_34572.

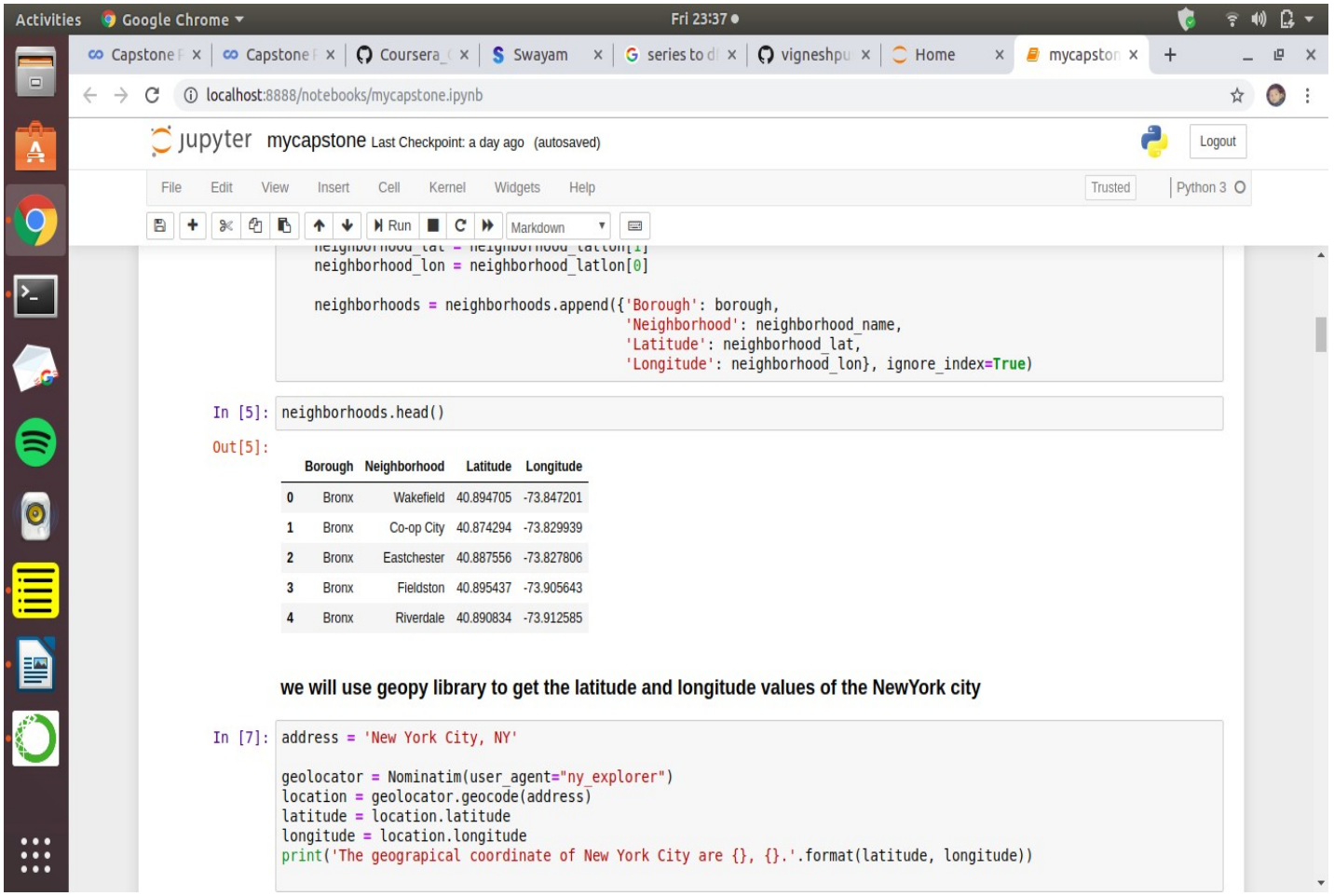
	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

Data2: Newyork city geographical coordinates data will be utilized as input for the Foursquare API, that will be leveraged to provision venues information for each neighborhood.We will use the Foursquare API to explore neighborhoods in New York City.

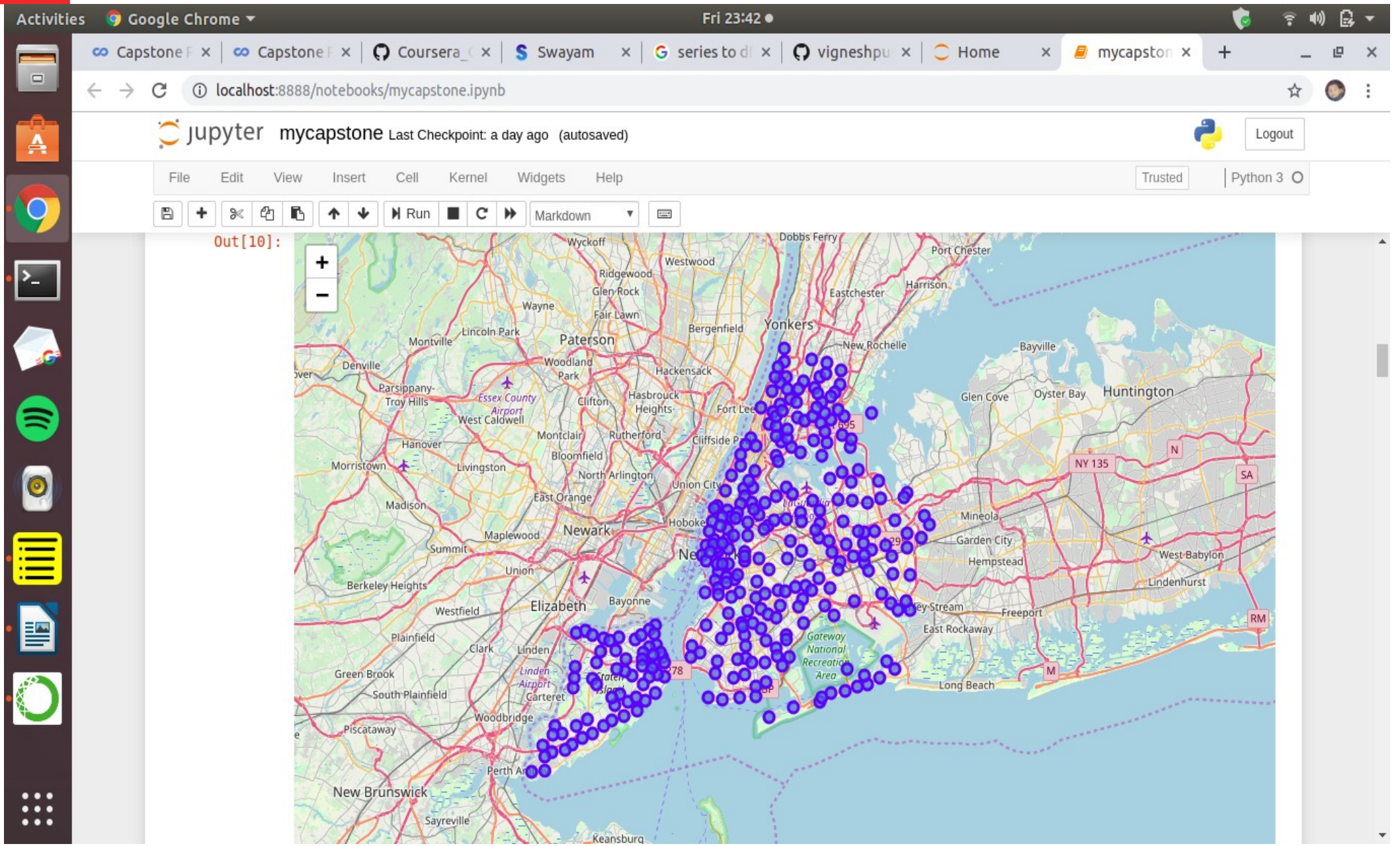

FOURSQUARE
DEVELOPERS
Products
Docs
Log-in
Create Account
Stadium
4bf58dd8d48988d184941735
Baseball Stadium
4bf58dd8d48988d18c941735
Basketball Stadium
4bf58dd8d48988d18b941735
Cricket Ground
4bf58dd8d48988d18a941735
Football Stadium
4bf58dd8d48988d189941735
Hockey Arena
4bf58dd8d48988d185941735
Rugby Stadium
56aa371be4b08b9a8d573556
Soccer Stadium
4bf58dd8d48988d188941735

Methodology:

First we will convert data from the various resources into a dataframe



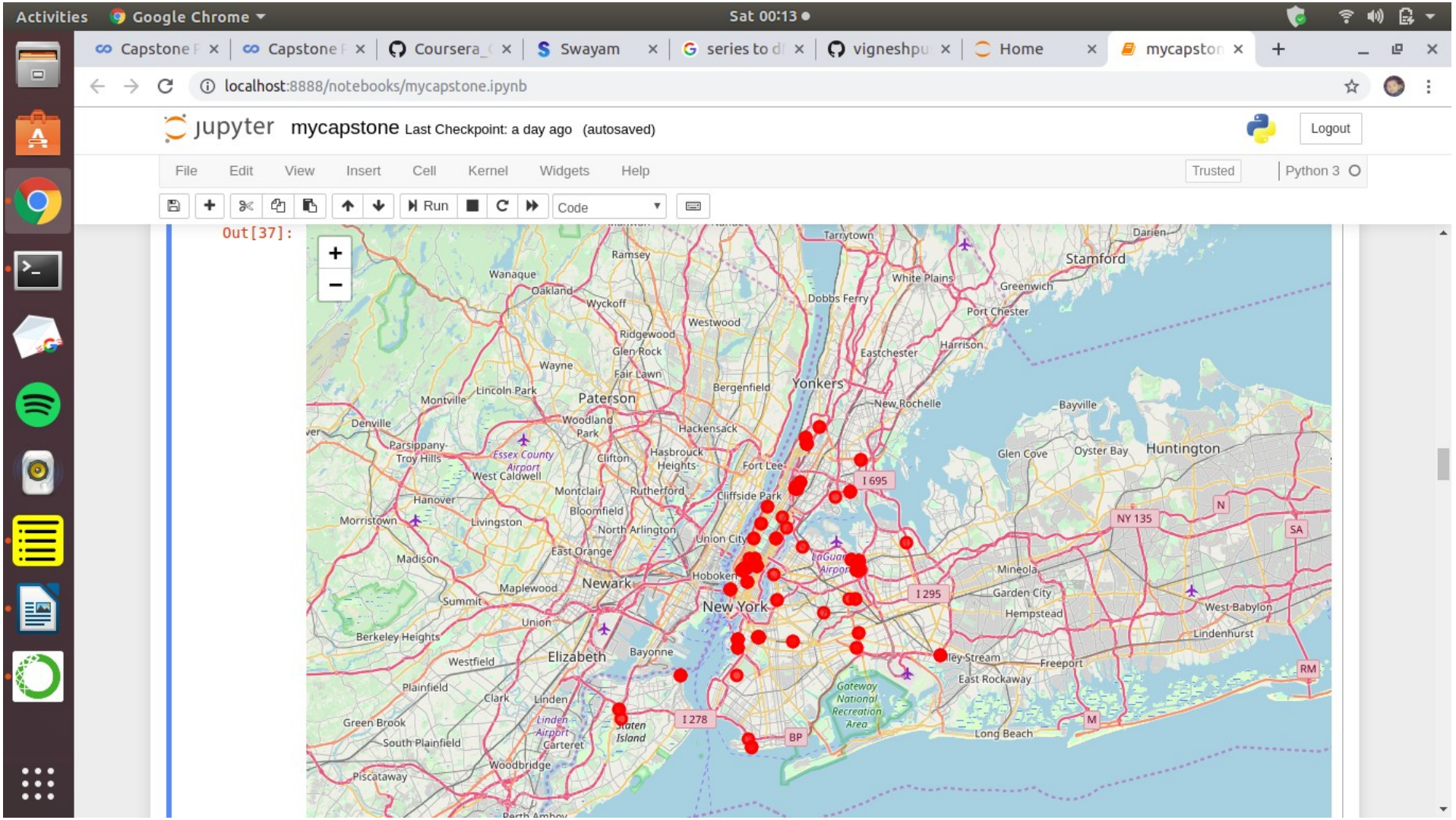
A MAP OF NEWYORK WITH NEIGHBORHOODS SUPER IMPOSED ON IT



FOURSQUARE

The data is taken from the foursqare.com
https://developer.foursquare.com/docs/resources/categories is the link to get the different catagories of resources in the newyork city
In this project we have taken Stadium as the categories and the Id of the Stadium is 4bf58dd8d48988d184941735
we retrive the data from the foursquare data using the Id and above data will be obtained.

This is the map of newyork superimposing the places where the sports facilities are present.



analyise each neighborhood

jupyter mycapstone Last Checkpoint: a day ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Code

Out[41]:

	Neighborhood	Baseball Field	Baseball Stadium	Basketball Stadium	College Gym	Conference Room	Food Service	Football Stadium	Historic Site	Hockey Arena	Soccer Field	Soccer Stadium	Sports Club	Stadium	Tennis Court	S
0	Fieldston	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
1	Kingsbridge	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
2	Kingsbridge	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3	Marble Hill	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
4	Marble Hill	0	0	0	0	0	0	0	0	0	0	1	0	0	0	
5	Pelham Parkway	0	0	0	0	0	0	1	0	0	0	0	0	0	0	
6	High Bridge	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
7	High Bridge	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
8	High Bridge	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
9	High Bridge	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
10	High Bridge	0	0	0	0	0	0	0	0	0	0	0	0	1	0	
11	High Bridge	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
12	High Bridge	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
13	High Bridge	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
14	High Bridge	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
15	High Bridge	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
16	High Bridge	0	0	0	0	0	0	0	0	0	0	0	0	1	0	
17	High Bridge	0	0	0	0	0	0	0	0	0	0	0	0	1	0	
18	High Bridge	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
19	High Bridge	0	0	0	0	0	0	1	0	0	0	0	0	0	0	
20	Soundview	0	0	0	0	0	0	1	0	0	0	0	0	0	0	
21	Soundview	0	0	0	0	0	0	1	0	0	0	0	0	0	0	
22	Spuyten Duyvil	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

K-means clustering (machine learning algorithm)

jupyter mycapstone Last Checkpoint: a day ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Code

clustering neighborhoods

In [87]:
```
# set number of clusters
kclusters = 5
ny_grouped_clustering = ny_grouped.drop('Neighborhood', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(ny_grouped_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]
```

Out[87]:
```
array([0, 0, 2, 4, 1, 4, 1, 0, 4, 4], dtype=int32)
```

In [88]:
```
ny_merged = neighborhoods
ny_merged = ny_merged.join(neighborhoods_venues_sorted.set_index('Neighborhood'), on='Neighborhood')

ny_merged
```

Out[88]:

	Borough	Neighborhood	Latitude	Longitude	Cluster Label	Cluster Labels	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 M Com Ve
0	Bronx	Wakefield	40.894705	-73.847201	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
1	Bronx	Co-op City	40.874294	-73.829939	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
2	Bronx	Eastchester	40.887556	-73.827806	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
3	Bronx	Fieldston	40.895437	-73.905643	1.0	1.0	Baseball Stadium	Track Stadium	Tennis Stadium	Tennis Court	Stadium	Sports Club	Soccer Stadium	Sc
4	Bronx	Riverdale	40.890834	-73.912585	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
5	Bronx	Kingsbridge	40.881687	-73.902818	3.0	3.0	Tennis Stadium	Baseball Stadium	Track Stadium	Tennis Court	Stadium	Sports Club	Soccer Stadium	Sc
6	Manhattan	Marble Hill	40.876551	-73.910660	3.0	3.0	Tennis	Soccer	Track	Tennis	Stadium	Sports	Soccer	Ho

Clustering

jupyter mycapstone Last Checkpoint: a day ago (autosaved)

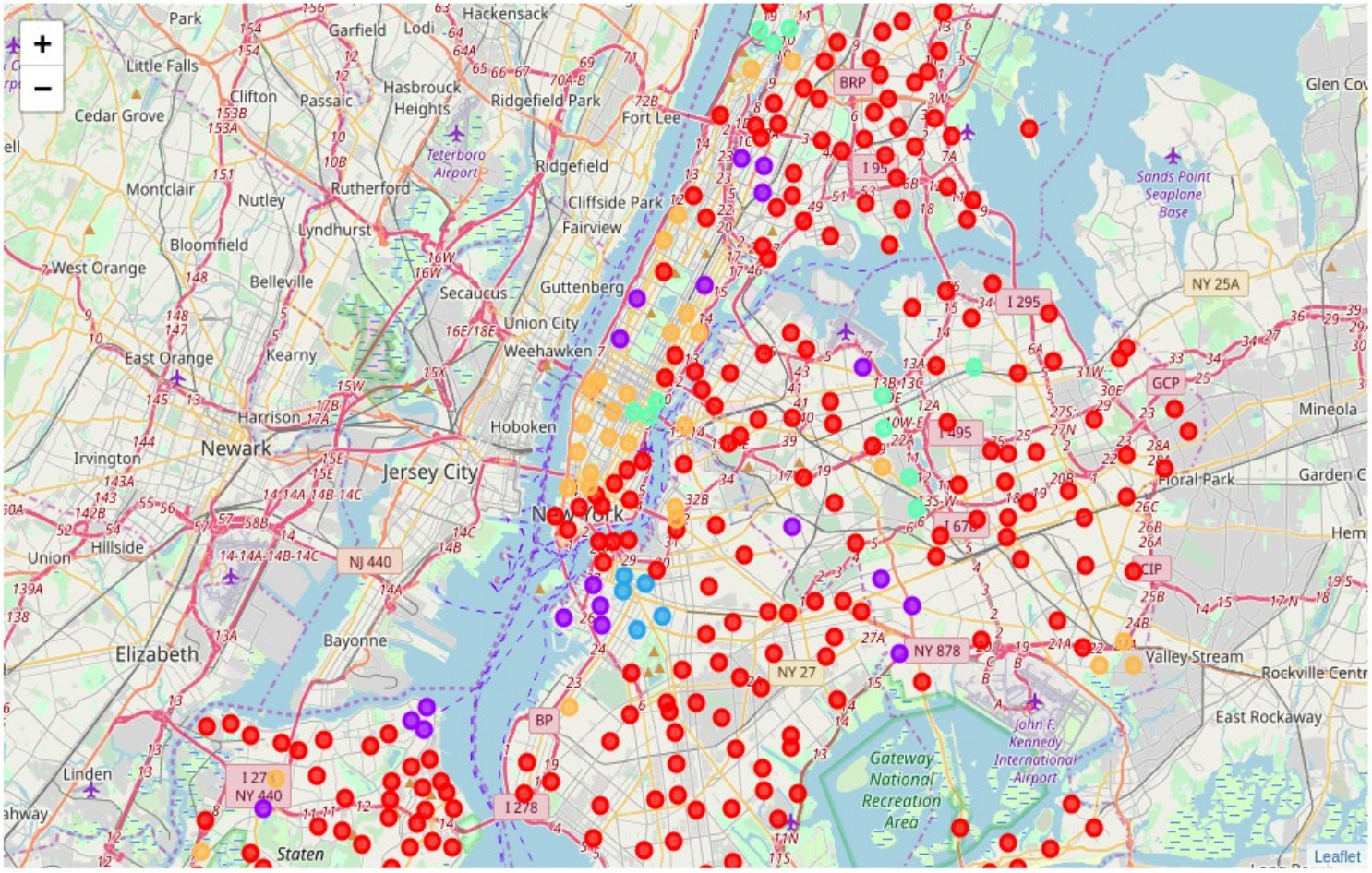
Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Code

```
map_clusters
```

Out[90]:



RESULTS

1st cluster

jupyter mycapstone Last Checkpoint: Yesterday at 01:14 (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3

Run Markdown

```
In [92]: ny_merged.loc[ny_merged['Cluster Labels'] == 1, ny_merged.columns[[1] + list(range(5, ny_merged.shape[1]))]]
```

Out[92]:

	Neighborhood	Cluster Labels	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
3	Fieldston	1.0	Baseball Stadium	Track Stadium	Tennis Stadium	Tennis Court	Stadium	Sports Club	Soccer Stadium	Soccer Field	Hockey Arena	Historic Site
19	High Bridge	1.0	Baseball Stadium	Stadium	Track Stadium	Football Stadium	Tennis Stadium	Tennis Court	Sports Club	Soccer Stadium	Soccer Field	Hockey Arena
43	Concourse	1.0	Baseball Stadium	Stadium	Football Stadium	Track Stadium	Tennis Stadium	Tennis Court	Sports Club	Soccer Stadium	Soccer Field	Hockey Arena
65	Cobble Hill	1.0	Baseball Stadium	Track Stadium	Tennis Stadium	Tennis Court	Stadium	Sports Club	Soccer Stadium	Soccer Field	Hockey Arena	Historic Site
66	Carroll Gardens	1.0	Baseball Stadium	Track Stadium	Tennis Stadium	Tennis Court	Stadium	Sports Club	Soccer Stadium	Soccer Field	Hockey Arena	Historic Site
67	Red Hook	1.0	Baseball Stadium	Track Stadium	Tennis Stadium	Tennis Court	Stadium	Sports Club	Soccer Stadium	Soccer Field	Hockey Arena	Historic Site
68	Gowanus	1.0	Baseball Stadium	Track Stadium	Tennis Stadium	Tennis Court	Stadium	Sports Club	Soccer Stadium	Soccer Field	Hockey Arena	Historic Site
78	Coney Island	1.0	Baseball Stadium	Track Stadium	Tennis Stadium	Tennis Court	Stadium	Sports Club	Soccer Stadium	Soccer Field	Hockey Arena	Historic Site
106	East Harlem	1.0	Baseball Stadium	Track Stadium	Tennis Stadium	Tennis Court	Stadium	Sports Club	Soccer Stadium	Soccer Field	Hockey Arena	Historic Site
111	Upper West Side	1.0	Football Stadium	Baseball Stadium	Track Stadium	Tennis Stadium	Tennis Court	Stadium	Sports Club	Soccer Stadium	Soccer Field	Hockey Arena
112	Lincoln Square	1.0	Baseball Stadium	Track Stadium	Tennis Stadium	Tennis Court	Stadium	Sports Club	Soccer Stadium	Soccer Field	Hockey Arena	Historic Site
141	East Elmhurst	1.0	Baseball Stadium	Track Stadium	Tennis Stadium	Tennis Court	Stadium	Sports Club	Soccer Stadium	Soccer Field	Hockey Arena	Historic Site
143	Ridgewood	1.0	Baseball Stadium	Track Stadium	Tennis Stadium	Tennis Court	Stadium	Sports Club	Soccer Stadium	Soccer Field	Hockey Arena	Historic Site
146	Woodhaven	1.0	Baseball Stadium	Track Stadium	Tennis Stadium	Tennis Court	Stadium	Sports Club	Soccer Stadium	Soccer Field	Hockey Arena	Historic Site

2nd cluster

jupyter mycapstone Last Checkpoint: Yesterday at 01:14 (unsaved changes)

Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Markdown

In [93]:
```
ny_merged.loc[ny_merged['Cluster Labels'] == 3, ny_merged.columns[[1] + list(range(5, ny_merged.shape[1]))]]
```

Out[93]:

	Neighborhood	Cluster Labels	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
5	Kingsbridge	3.0	Tennis Stadium	Baseball Stadium	Track Stadium	Tennis Court	Stadium	Sports Club	Soccer Stadium	Soccer Field	Hockey Arena	Historic Site
6	Marble Hill	3.0	Tennis Stadium	Soccer Stadium	Track Stadium	Tennis Court	Stadium	Sports Club	Soccer Field	Hockey Arena	Historic Site	Football Stadium
35	Spuyten Duyvil	3.0	Tennis Stadium	Soccer Stadium	Track Stadium	Tennis Court	Stadium	Sports Club	Soccer Field	Hockey Arena	Historic Site	Football Stadium
115	Murray Hill	3.0	Tennis Stadium	Stadium	Track Stadium	Tennis Court	Sports Club	Soccer Stadium	Soccer Field	Hockey Arena	Historic Site	Football Stadium
134	Corona	3.0	Tennis Stadium	Track Stadium	Tennis Court	Stadium	Sports Club	Soccer Stadium	Soccer Field	Hockey Arena	Historic Site	Football Stadium
135	Forest Hills	3.0	Tennis Stadium	Track Stadium	Tennis Court	Stadium	Sports Club	Soccer Stadium	Soccer Field	Hockey Arena	Historic Site	Football Stadium
180	Murray Hill	3.0	Tennis Stadium	Stadium	Track Stadium	Tennis Court	Sports Club	Soccer Stadium	Soccer Field	Hockey Arena	Historic Site	Football Stadium
195	North Corona	3.0	Tennis Stadium	Baseball Stadium	Stadium	Historic Site	Food Service	Track Stadium	Tennis Court	Sports Club	Soccer Stadium	Soccer Field
196	Forest Hills Gardens	3.0	Tennis Stadium	Track Stadium	Tennis Court	Stadium	Sports Club	Soccer Stadium	Soccer Field	Hockey Arena	Historic Site	Football Stadium
273	Turtle Bay	3.0	Tennis Stadium	Track Stadium	Tennis Court	Stadium	Sports Club	Soccer Stadium	Soccer Field	Hockey Arena	Historic Site	Football Stadium
274	Tudor City	3.0	Tennis Stadium	Track Stadium	Tennis Court	Stadium	Sports Club	Soccer Stadium	Soccer Field	Hockey Arena	Historic Site	Football Stadium

In [94]:
```
ny_merged.loc[ny_merged['Cluster Labels'] == 4, ny_merged.columns[[1] + list(range(5, ny_merged.shape[1]))]]
```

Out[94]:

	Neighborhood	Cluster Labels	1st Most Common	2nd Most Common	3rd Most Common	4th Most Common	5th Most Common	6th Most Common	7th Most Common	8th Most Common	9th Most Common	10th Most Common

3rd cluster

jupyter mycapstone Last Checkpoint: Yesterday at 01:14 (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

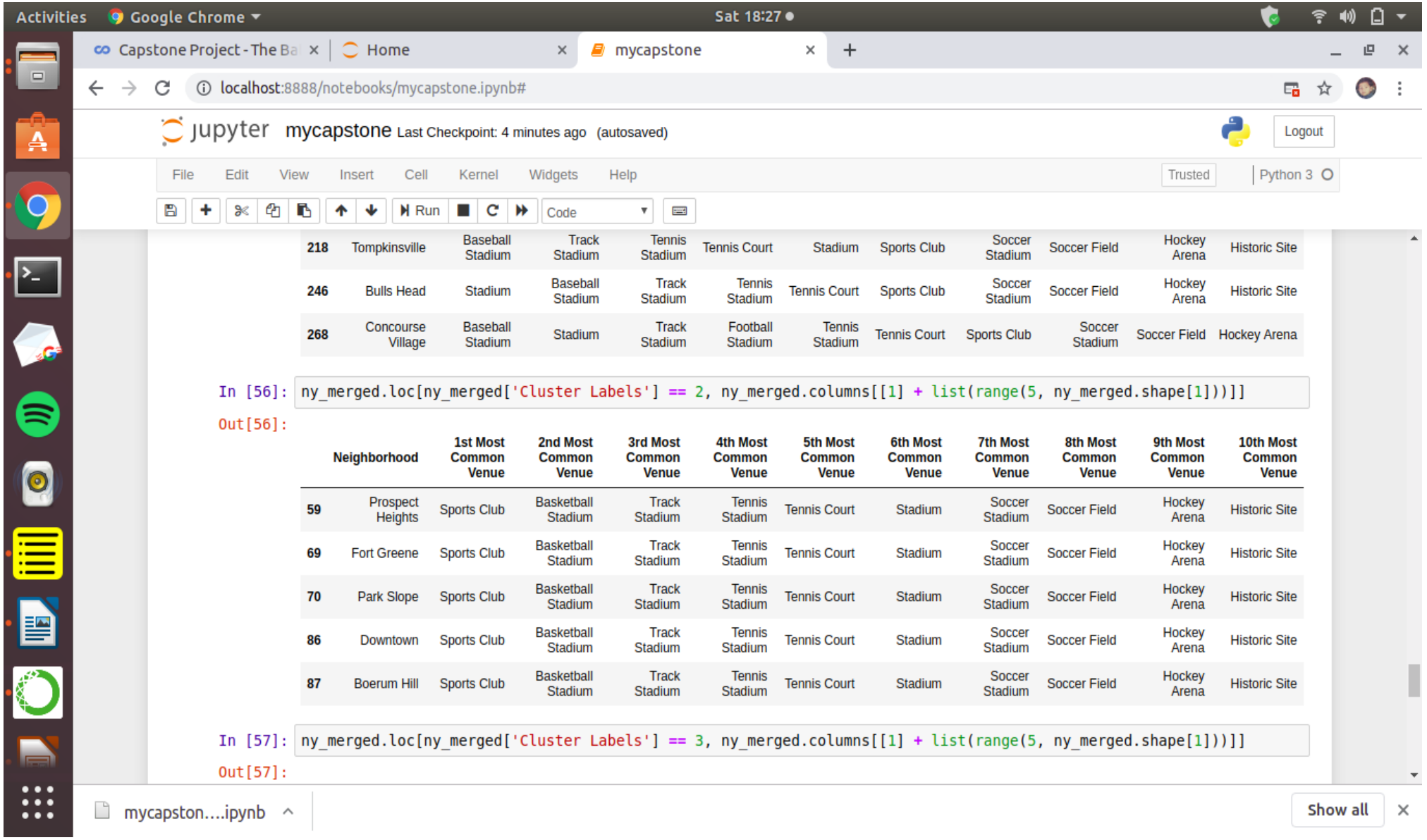
Run Markdown

In [94]:
```
ny_merged.loc[ny_merged['Cluster Labels'] == 4, ny_merged.columns[[1] + list(range(5, ny_merged.shape[1]))]]
```

Out[94]:

	Neighborhood	Cluster Labels	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
48	Sunset Park	4.0	Stadium	Track Stadium	Tennis Stadium	Tennis Court	Sports Club	Soccer Stadium	Soccer Field	Hockey Arena	Historic Site	Football Stadium
96	North Side	4.0	Track Stadium	Tennis Stadium	Tennis Court	Stadium	Sports Club	Soccer Stadium	Soccer Field	Hockey Arena	Historic Site	Football Stadium
97	South Side	4.0	Track Stadium	Tennis Stadium	Tennis Court	Stadium	Sports Club	Soccer Stadium	Soccer Field	Hockey Arena	Historic Site	Football Stadium
102	Inwood	4.0	Soccer Stadium	Track Stadium	Tennis Stadium	Tennis Court	Stadium	Sports Club	Soccer Field	Hockey Arena	Historic Site	Football Stadium
104	Manhattanville	4.0	Tennis Court	Track Stadium	Tennis Stadium	Stadium	Sports Club	Soccer Stadium	Soccer Field	Hockey Arena	Historic Site	Football Stadium
107	Upper East Side	4.0	College Gym	Track Stadium	Tennis Stadium	Tennis Court	Stadium	Sports Club	Soccer Stadium	Soccer Field	Hockey Arena	Historic Site
108	Yorkville	4.0	College Gym	Track Stadium	Tennis Stadium	Tennis Court	Stadium	Sports Club	Soccer Stadium	Soccer Field	Hockey Arena	Historic Site
113	Clinton	4.0	Stadium	Football Stadium	Conference Room	Basketball Stadium	Baseball Stadium	Track Stadium	Tennis Stadium	Tennis Court	Sports Club	Soccer Stadium
114	Midtown	4.0	Stadium	Tennis Stadium	Conference Room	Basketball Stadium	Baseball Stadium	Track Stadium	Tennis Court	Sports Club	Soccer Stadium	Soccer Field
116	Chelsea	4.0	Stadium	Football Stadium	Track Stadium	Tennis Stadium	Tennis Court	Sports Club	Soccer Stadium	Soccer Field	Hockey Arena	Historic Site
117	Greenwich Village	4.0	Soccer Field	Track Stadium	Tennis Stadium	Tennis Court	Stadium	Sports Club	Soccer Stadium	Hockey Arena	Historic Site	Football Stadium
120	Tribeca	4.0	Soccer Field	Track Stadium	Tennis Stadium	Tennis Court	Stadium	Sports Club	Soccer Stadium	Hockey Arena	Historic Site	Football Stadium
122	Soho	4.0	Soccer Field	Track Stadium	Tennis Stadium	Tennis Court	Stadium	Sports Club	Soccer Stadium	Hockey Arena	Historic Site	Football Stadium
123	West Village	4.0	Soccer Field	Track Stadium	Tennis Stadium	Tennis Court	Stadium	Sports Club	Soccer Stadium	Hockey Arena	Historic Site	Football Stadium

4th cluster



218	Tompkinsville	Baseball Stadium	Track Stadium	Tennis Stadium	Tennis Court	Stadium	Sports Club	Soccer Stadium	Soccer Field	Hockey Arena	Historic Site
246	Bulls Head	Stadium	Baseball Stadium	Track Stadium	Tennis Stadium	Tennis Court	Sports Club	Soccer Stadium	Soccer Field	Hockey Arena	Historic Site
268	Concourse Village	Baseball Stadium	Stadium	Track Stadium	Football Stadium	Tennis Stadium	Tennis Court	Sports Club	Soccer Stadium	Soccer Field	Hockey Arena

```
In [56]: ny_merged.loc[ny_merged['Cluster Labels'] == 2, ny_merged.columns[[1] + list(range(5, ny_merged.shape[1]))]]
```

Out[56]:

	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
59	Prospect Heights	Sports Club	Basketball Stadium	Track Stadium	Tennis Stadium	Tennis Court	Stadium	Soccer Stadium	Soccer Field	Hockey Arena	Historic Site
69	Fort Greene	Sports Club	Basketball Stadium	Track Stadium	Tennis Stadium	Tennis Court	Stadium	Soccer Stadium	Soccer Field	Hockey Arena	Historic Site
70	Park Slope	Sports Club	Basketball Stadium	Track Stadium	Tennis Stadium	Tennis Court	Stadium	Soccer Stadium	Soccer Field	Hockey Arena	Historic Site
86	Downtown	Sports Club	Basketball Stadium	Track Stadium	Tennis Stadium	Tennis Court	Stadium	Soccer Stadium	Soccer Field	Hockey Arena	Historic Site
87	Boerum Hill	Sports Club	Basketball Stadium	Track Stadium	Tennis Stadium	Tennis Court	Stadium	Soccer Stadium	Soccer Field	Hockey Arena	Historic Site

```
In [57]: ny_merged.loc[ny_merged['Cluster Labels'] == 3, ny_merged.columns[[1] + list(range(5, ny_merged.shape[1]))]]
```

Out[57]:

Final observation

Here we can observe the different common venues for the different clusters.By observing the clusters.

If my client want to build a football stadium then cluster 2 and 3 will be the best places where there is less competiton for the football stadium and there people are very intrested in participating in sports activities.

Discussion

In this section, I would be discussing the observations I have noted and the recommendation that I can make based on the results. This analysis is performed on limited data. This may be right or may be wrong. But if good amount of data is available there is scope to come up with better results.

There is high competition in Midtown and Soho so it is very risky to open business in these areas.
Central Harlem has also potential where closes to Morningside Heights area.
It can be done more detailed analysis by adding other factors such as transportation, demographics of inhabitants.

Conclusion

Although all of the goals of this project were met there is definitely room for further improvement and development as noted below. However, the goals of the project were met and, with some more work, could easily be devleoped into a fully phledged application that could support the opening a business idea in an unknown location.