Where to open Fitness Center

Deep Chatterjee

Introduction / Business Problem

- New York City was home to nearly 8.5 million people in 2014.
- The aim of this project is to discover where is the optimal place to build a gym in the New York City.
- To answer this question we will analyse which districts have a gym and if there are any similarities between them.

Data

- For this project we need the following data :
- New York City data that contains list Boroughs, Neighborhoods along with their latitude and longitude.
- Data source : https://cocl.us/new_york_dataset
- Description: This data set contains the required information. And we will use this data set to
- explore various neighbourhoods of new york city.
- Data source: Fousquare API:

Description: By using this api we will get all the venues in each neighbourhood. We can filter these venues to get only gym/fitness-center.

Methodology

- **Exploring and Cleaning Data**
- We need to know our data. We can do this through exploratory analysis. Once we know our data we can perform, if necessary, some cleaning (In our case we will see that it is necessary)

Analyze each neighborhood

In [48]:
```
# one hot encoding
manhattan_onehot = pd.get_dummies(newyork_venues_gym[['Venue Category']], prefix="", prefix_sep="")

# add neighborhood column back to dataframe
manhattan_onehot['Neighborhood'] = newyork_venues_gym['Neighborhood']

# move neighborhood column to the first column
fixed_columns = [manhattan_onehot.columns[-1]] + list(manhattan_onehot.columns[:-1])
manhattan_onehot = manhattan_onehot[fixed_columns]

manhattan_onehot.head()
```

Out[48]:

	Neighborhood	Athletics & Sports	Basketball Court	Beer Garden	Bike Shop	Boxing Gym	Building	Climbing Gym	Club House	Community Center	Corporate Amenity	Cultural Center	Cycle Studio	Dance Studio	Doctor's Office	Dog Run	E
0	Marble Hill	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	Marble Hill	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

Methodology

- **Sorting the neighborhood**
- We now sort the venues

```
neighborhoods_venues_sorted.head(10)
```

Out[51]:

	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	Battery Park City	Gym / Fitness Center	Gym	Boxing Gym	Yoga Studio	Gym Pool	Cycle Studio	Athletics & Sports	Corporate Amenity	Doctor's Office	Medical Center
1	Carnegie Hill	Gym / Fitness Center	Gym	Yoga Studio	Pool	Boxing Gym	Building	Climbing Gym	Community Center	Cycle Studio	Martial Arts Dojo
2	Central Harlem	Gym	Gym / Fitness Center	Yoga Studio	Cycle Studio	Martial Arts Dojo	Athletics & Sports	General College & University	Climbing Gym	Pilates Studio	Corporate Amenity
3	Chelsea	Gym / Fitness Center	Gym	Cycle Studio	Yoga Studio	Spa	Recreation Center	Bike Shop	Boxing Gym	Dance Studio	Basketball Court
4	Chinatown	Gym / Fitness Center	Gym	Yoga Studio	Pilates Studio	Boxing Gym	Martial Arts Dojo	Athletics & Sports	Cycle Studio	Office	Corporate Amenity
5	Civic Center	Gym	Gym / Fitness Center	Yoga Studio	Boxing Gym	Cycle Studio	Pilates Studio	Corporate Amenity	Gym Pool	Office	Martial Arts Dojo
6	Clinton	Gym	Gym / Fitness	Yoga Studio	Cycle Studio	Exhibit	Boxing Gym	Building	Medical Center	Residential Building (Apartment / Condo)	Track

Methodology

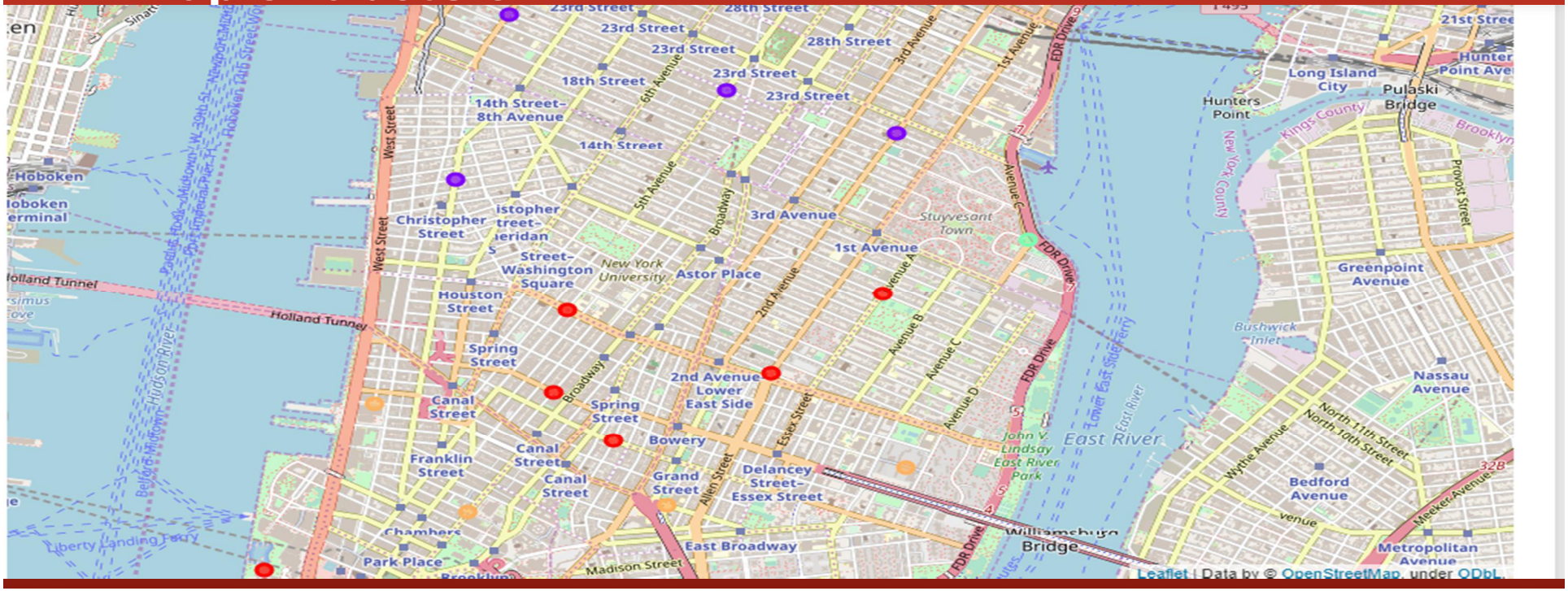
- **Clustering the venues**
- We now use Machine Learning algorithm K-Means Clustering to cluster out present fitness centers

In [52]:
```
# set number of clusters
kclusters = 5

manhattan_grouped_clustering = manhattan_grouped.drop('Neighborhood', 1)

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

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

Out[52]:
```
array([0, 0, 3, 1, 4, 4, 3, 2, 0, 3])
```

Result

- After Clustering the venues we get the following map of clusters

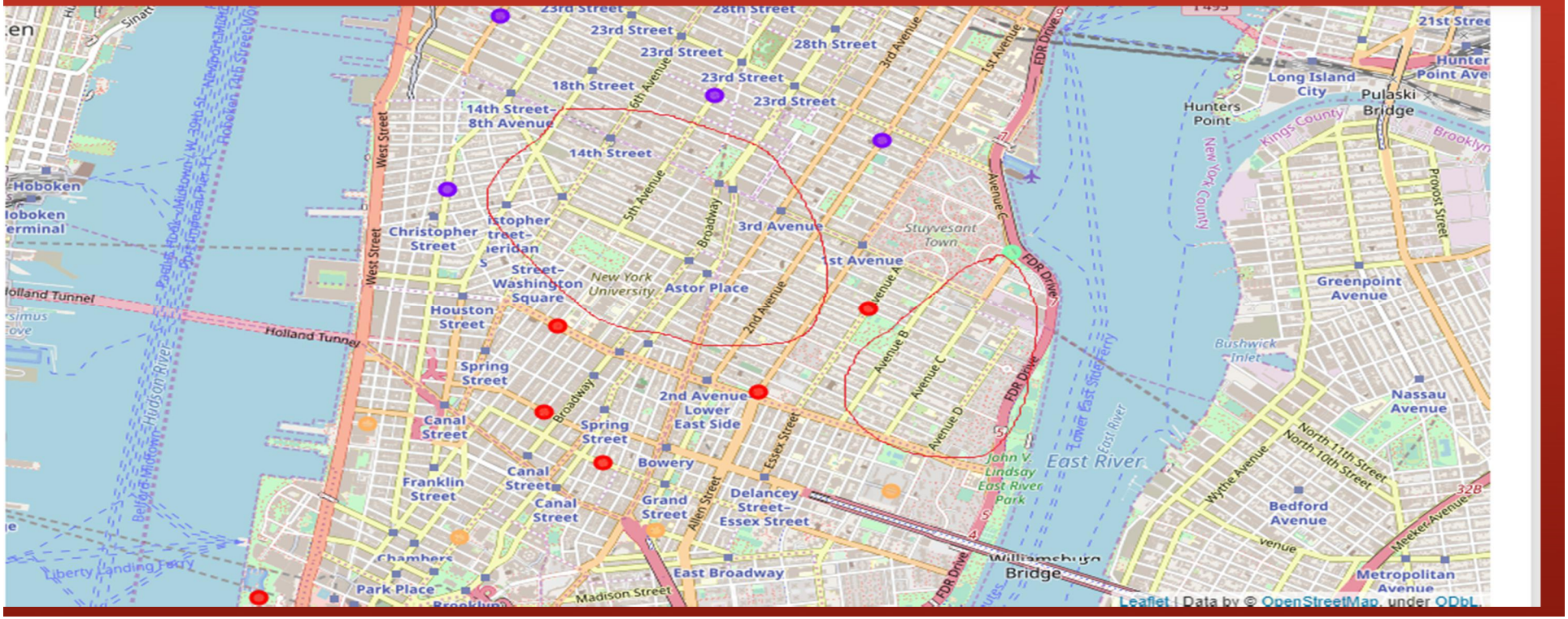


Discussion

- There is high competition in Lower Manhattan so it is very risky to open business in these areas.
- Soho and East Village have potential to open new centres.
- The above analysis is performed on limited data. Hence, it may be not very accurate. So, if good amount of data is available there is scope to come up with better results.

Conclusion

Below Circled areas can be considered as to open new Gym/Fitness Center



Thank you