



Growing from Local to Global: A Recommender System for Restaurant Business Expansion

IBM Data Science Professional Capstone Project

Junkang Gu – 2020

jgu8@u.rochester.edu

University of Rochester, NY

I. Introduction

Business Problem



<https://www.nycgo.com/articles/new-york-city-in-three-days>

<https://www.nytimes.com/2012/01/17/nyregion/despite-long-slide-by-kodak-rochester-avoids-decay.html>

Target Audience



http://www.gramercyparkblockassociation.org/images/neighborhood_news_559_6_10_14.html

<https://www.jacksonvilleu.com/blog/business/business-analytics-market-research-analyst-job-description-salary/>

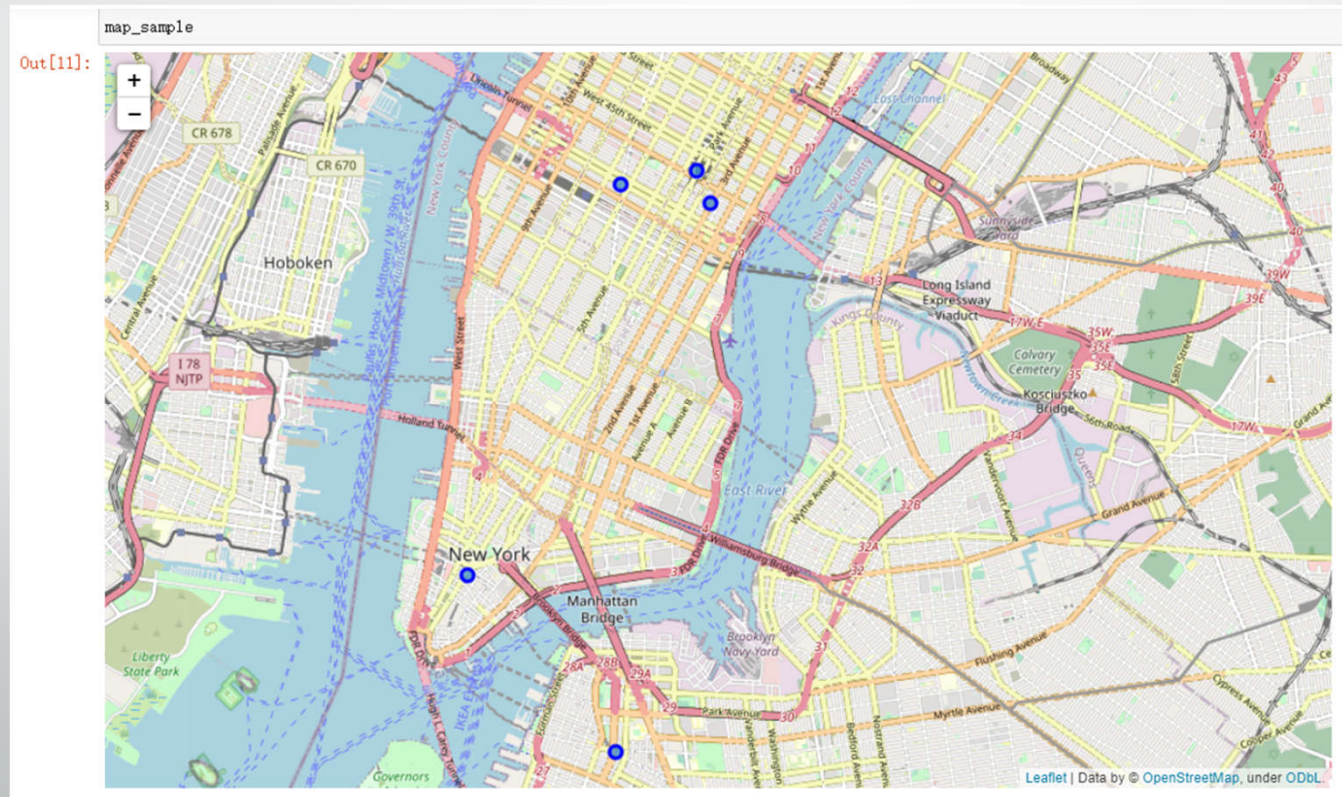
II. Data Description

Determine the Research Subjects



<http://www.sactownmag.com/Whats-Cooking/2019/Shake-Shack-Sacramento-Ice-Blocks/>

Sample Stores Proven to be Successful



map_target

Out[19]:

Map of Rochester, New York, showing the city grid, major roads, and the Genesee River. Numerous blue circular markers are plotted across the city, representing data points. The map includes labels for streets, parks, and highways. A legend in the top left corner shows a plus sign and a minus sign, indicating zoom controls. The map is sourced from OpenStreetMap, as indicated by the text "Leaflet | Data by © OpenStreetMap, under ODbL." in the bottom right corner.

Venue Data

Out[8]:

	Name	Center Latitude	Center Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Shake Shack in NYC	40.71056	-74.009014	Shake Shack	40.710703	-74.009024	Burger Joint
1	Shake Shack in NYC	40.71056	-74.009014	Anthropologie	40.710618	-74.009661	Women's Store

Out[20]:

	Name	Center Latitude	Center Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
2	Shake Shack in NY						
3	Shake Shack in NY						
4	Shake Shack in NY						
0	Convention District	43.156389	-77.609167	Morton's The Steakhouse	43.156566	-77.608455	Steakhouse
1	Convention District	43.156389	-77.609167	Rochester Riverside Hotel	43.156826	-77.609954	Hotel
2	Convention District	43.156389	-77.609167	Starbucks	43.156616	-77.608549	Coffee Shop
3	Convention District	43.156389	-77.609167	Hyatt Regency Rochester	43.156469	-77.608561	Hotel
4	Convention District	43.156389	-77.609167	Hyatt Focus Lounge	43.156549	-77.608646	Bar

Data Cleaning and Preprocessing

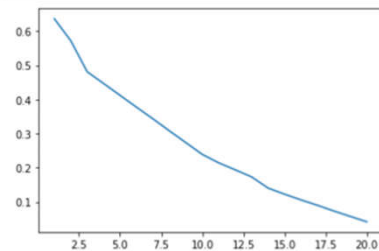
	Name	Airport Lounge	BBQ Joint	Bakery	Bar	Basketball Stadium	Bubble Tea Shop	Burger Joint	Café	Clothing Store
0	Shake Shack in NYC	0.013889	0.013889	0.013889	0.027778	0.013889	0.013889	0.138889	0.013889	0.013889
	Name	Airport Lounge	BBQ Joint	Bakery	Bar	Basketball Stadium	Bubble Tea Shop	Burger Joint	Café	Clothing Store
0	ABC Streets Neighborhood	0	0	0.000000	0.2	0	0	0.000000	0.000000	0
1	Bull's Head	0	0	0.000000	0.0	0	0	0.000000	0.000000	0
2	Changing of the Scenes	0	0	0.000000	0.0	0	0	0.000000	0.000000	0
3	College Town	0	0	0.076923	0.0	0	0	0.076923	0.076923	0
4	Convention District	0	0	0.000000	0.2	0	0	0.000000	0.000000	0

III. Methodology

Identify the Optimal K Value

```
In [31]: kmax = 20
errors = []
for k in range(1, kmax+1):
    error = 0
    kmeans = KMeans(n_clusters=k, random_state=0).fit(k_clustering)
    centroids = kmeans.cluster_centers_
    label = kmeans.labels_
    for i in range(0, len(k_clustering)):
        groupid = label[i]
        centroid = centroids[groupid]
        dist = linalg.norm(k_clustering.iloc[i].to_numpy() - centroid)
        error += dist
    errors.append(error/len(k_clustering))
print(errors)
```

```
In [32]: x_axis = range(1, kmax+1)
plt.plot(x_axis, errors)
plt.show()
```

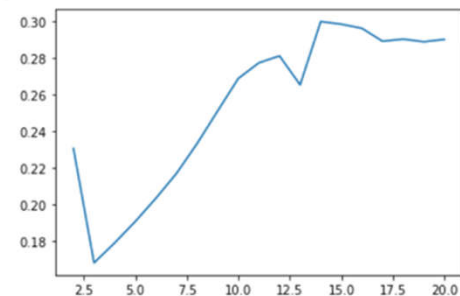


```
In [33]: from sklearn.metrics import silhouette_score

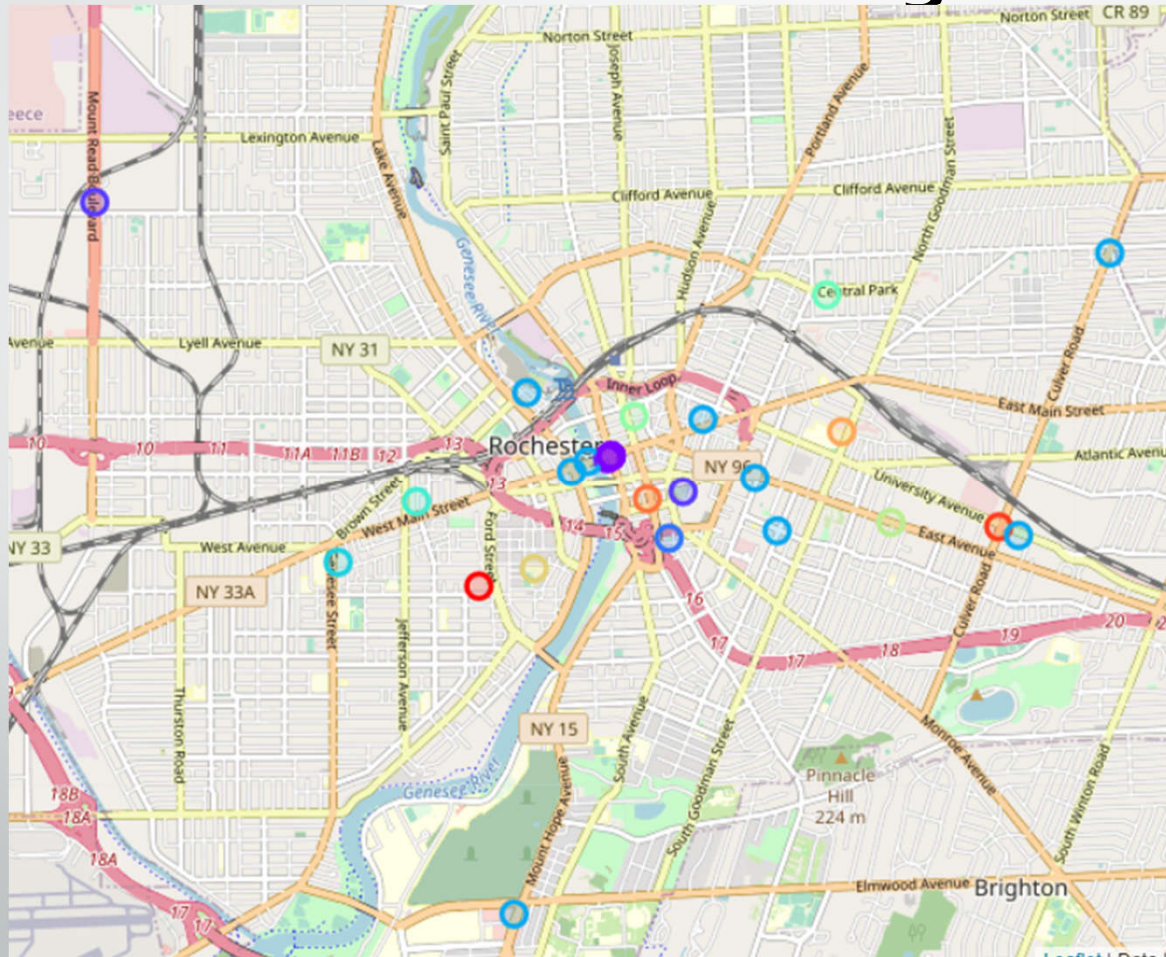
sil = []
kmax = 20

# dissimilarity would not be defined for a single cluster, thus, minimum number of cluster
for k in range(2, kmax+1):
    kmeans = KMeans(n_clusters=k, random_state=0).fit(k_clustering)
    labels = kmeans.labels_
    sil.append(silhouette_score(k_clustering, labels, metric = 'euclidean'))
```

```
In [34]: x_axis = range(2, kmax+1)
plt.plot(x_axis, sil)
plt.show()
```



K-means Clustering



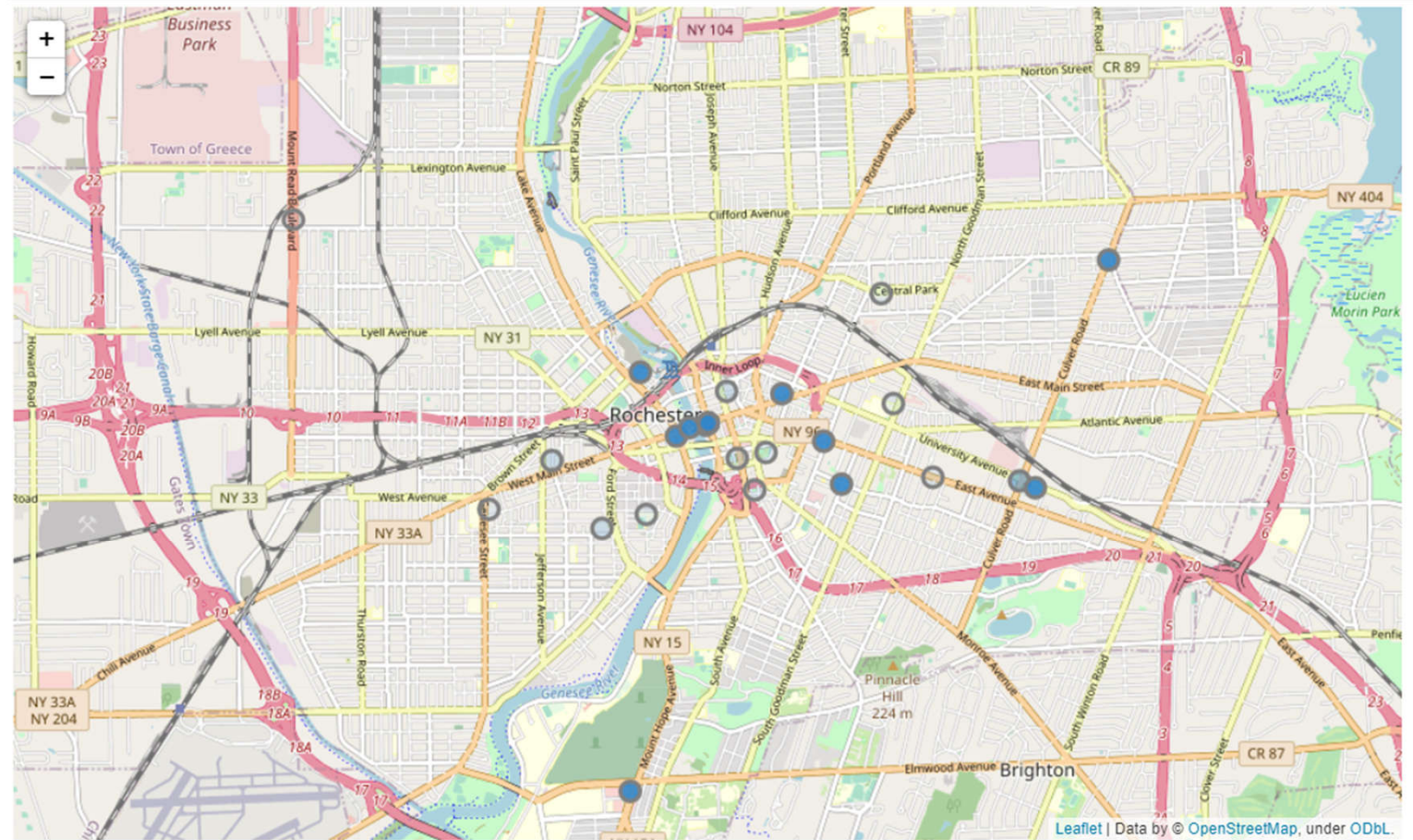
Recommendation

```
Out[43]: {4: 'Top Recommended',  
          1: 'Strongly Recommended',  
          13: 'Strongly Recommended',  
          6: 'Weakly Recommended',  
          8: 'Weakly Recommended',  
          12: 'Weakly Recommended',  
          0: 'Weakly Recommended',  
          2: 'Not Recommended',  
          3: 'Not Recommended',  
          5: 'Not Recommended',  
          7: 'Not Recommended',  
          9: 'Not Recommended',  
          10: 'Not Recommended',  
          11: 'Not Recommended' }
```

IV. Results

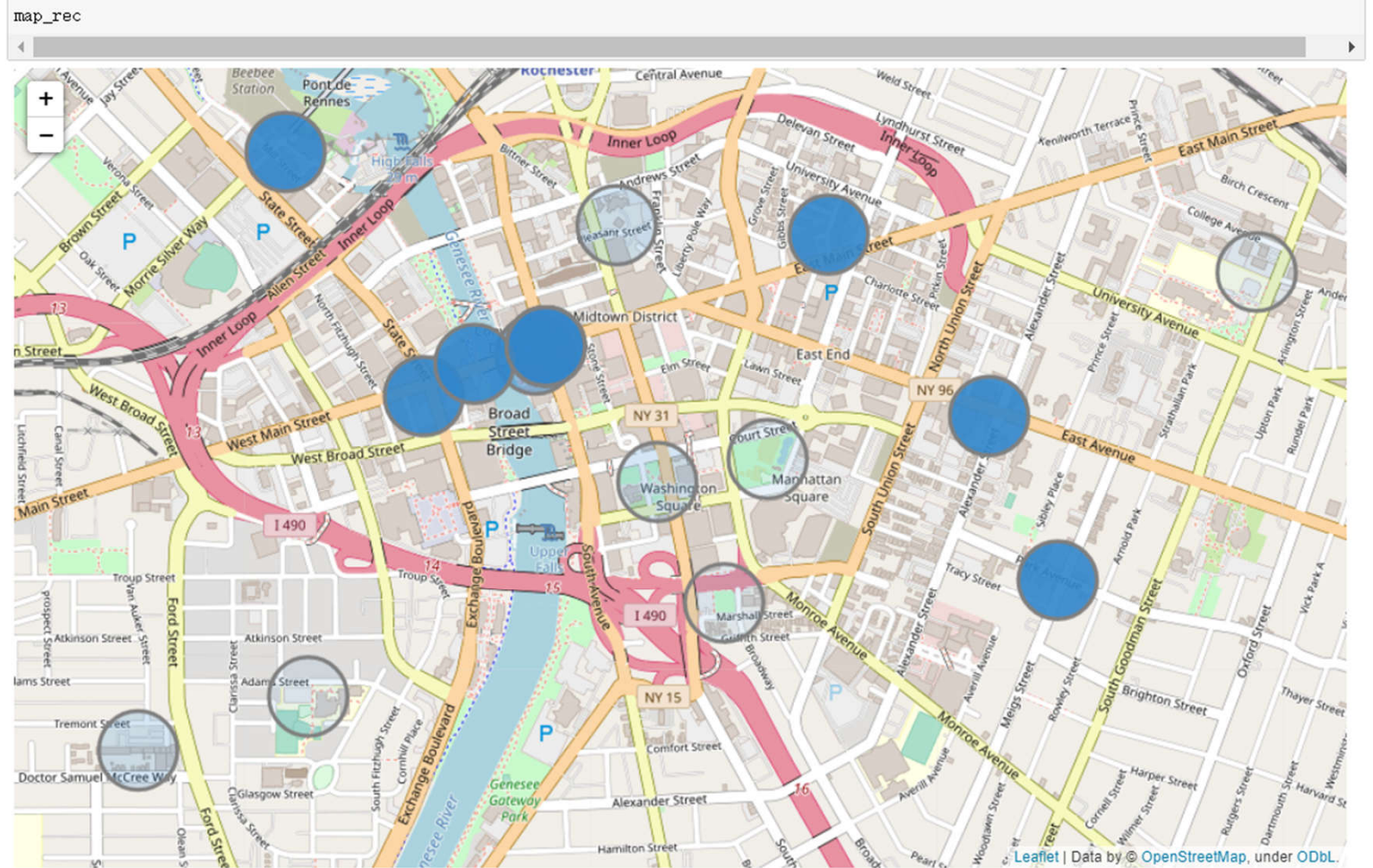
map_rec

Out[45]:



V. Discussion

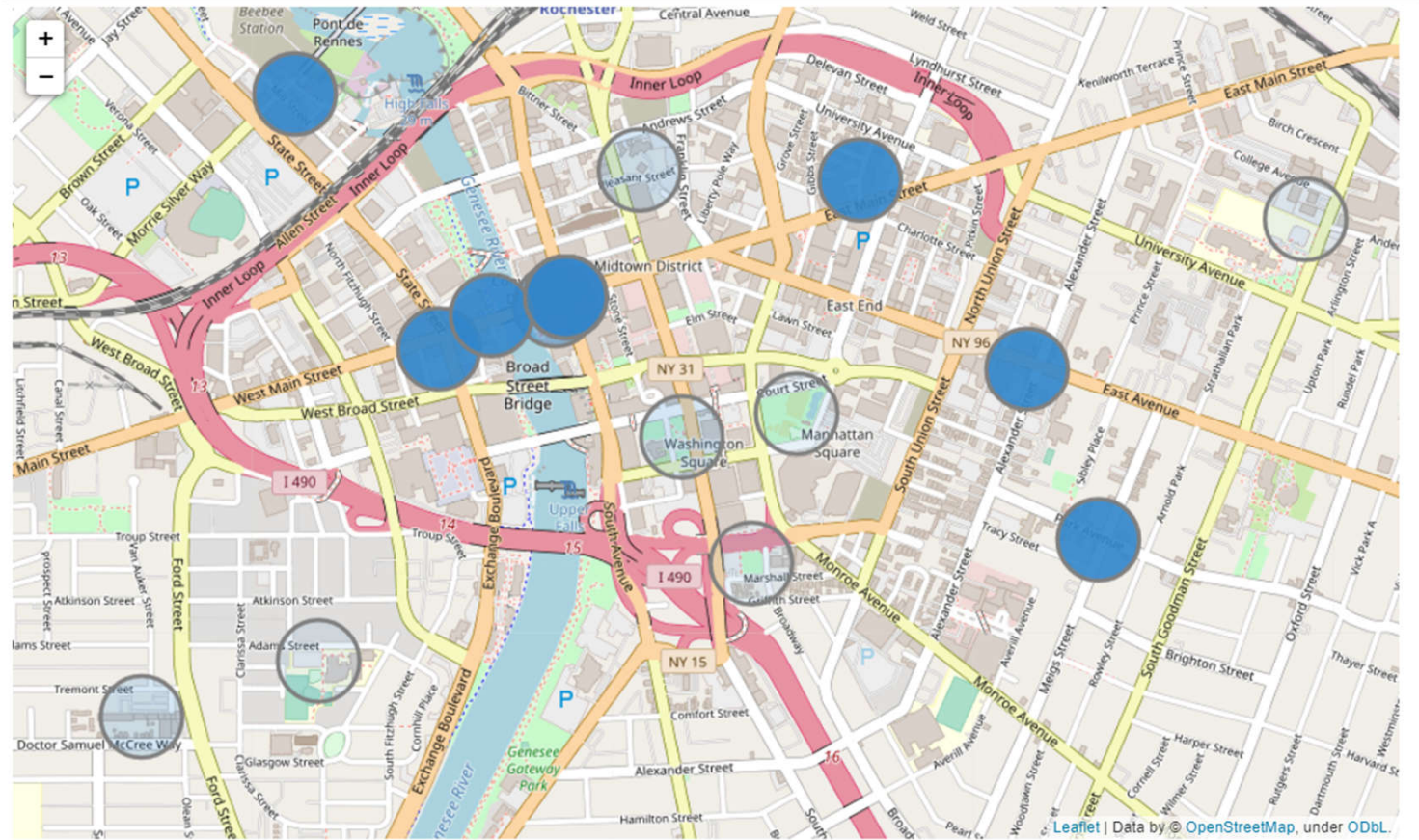
Out[45]:



VI. Conclusion

Out[45]:

map_rec



Leaflet | Data by © OpenStreetMap, under ODbL.