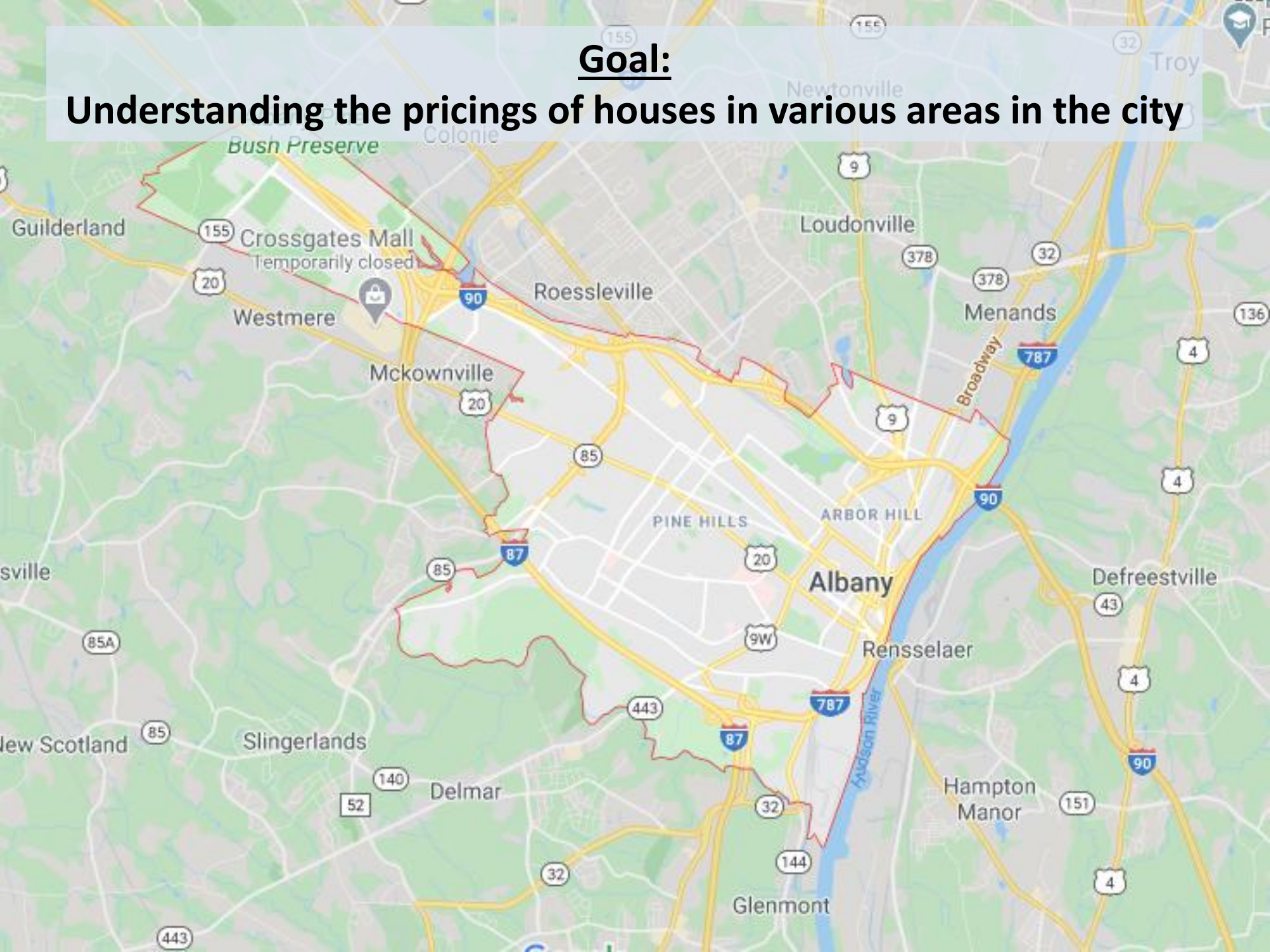# The City of Albany, NY

New York State capital city
Population: ~100,000
Economy: governmental centers,
        academy, industry & high-tech

Goal:

Understanding the pricings of houses in various areas in the city

Database

Using the official database in the city website

|	tax_map_number	street_address	total_assessed_value	tax_roll_section	owner_name	company_address	city_state_zip_code
34	75.36-2-66	43 Academy Rd	170000.0	Taxable	Burke, Thomas E.	43 Academy Rd	Albany, NY 12208
93	75.36-2-79	3 Academy Rd	215000.0	Taxable	JAIN, ANITA	3 Academy Rd	Albany, NY 12208
94	75.36-2-78	5 Academy Rd	187100.0	Taxable	Erdman, Janice	5 Academy Rd	Albany, NY 12208
95	75.36-2-77	7 Academy Rd	250000.0	Taxable	SHARKEY, MATTHEW	7 Academy Rd	Albany, NY 12208
96	75.36-2-76	9 Academy Rd	210000.0	Taxable	CHRISTIE, JANE M.	9 Academy Rd	Albany, NY 12208

Data Cleaning

Attaching each address of private home a price and location in coordinates by using online services

|	street_address	total_assessed_value	city_state_zip_code	Latitude	Longitude
0	43 Academy Rd, Albany, New York, USA	170000.0	Albany, NY 12208	42.650718	-73.785333
1	3 Academy Rd, Albany, New York, USA	215000.0	Albany, NY 12208	42.652731	-73.785122
2	5 Academy Rd, Albany, New York, USA	187100.0	Albany, NY 12208	42.652630	-73.785131
3	7 Academy Rd, Albany, New York, USA	250000.0	Albany, NY 12208	42.652530	-73.785139
4	9 Academy Rd, Albany, New York, USA	210000.0	Albany, NY 12208	42.652429	-73.785148

Data Cleaning

After locating the houses the data is presented using Folium package on the map
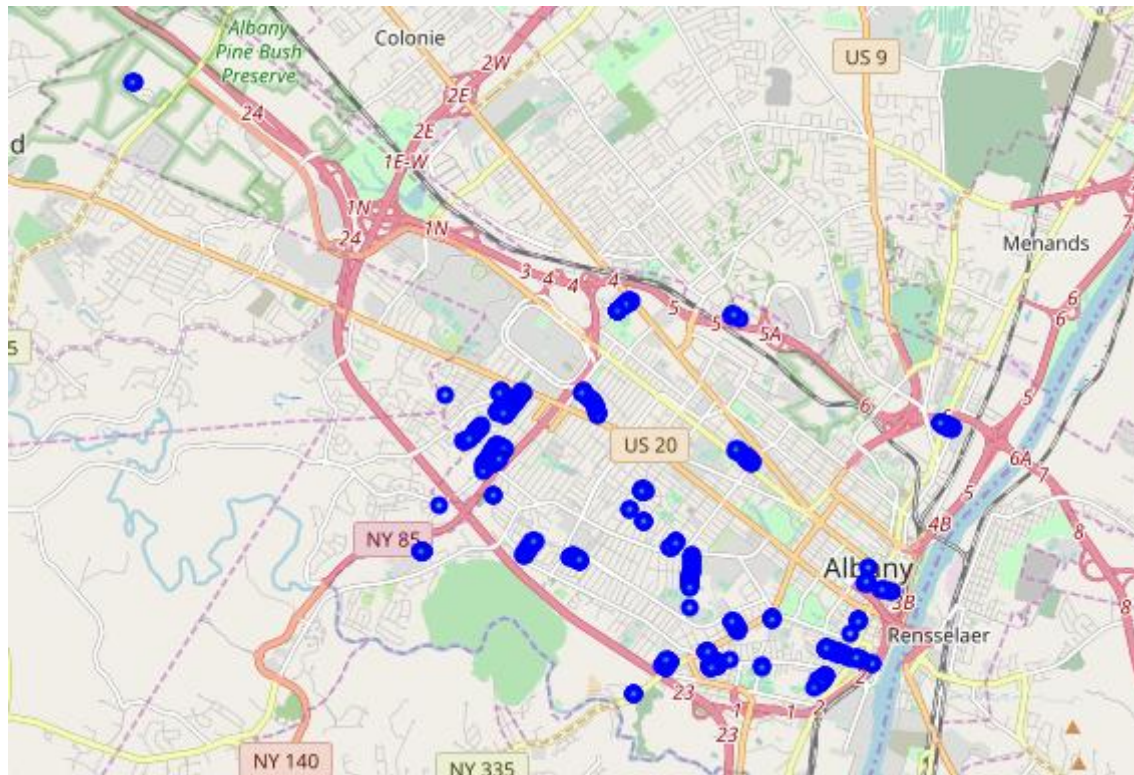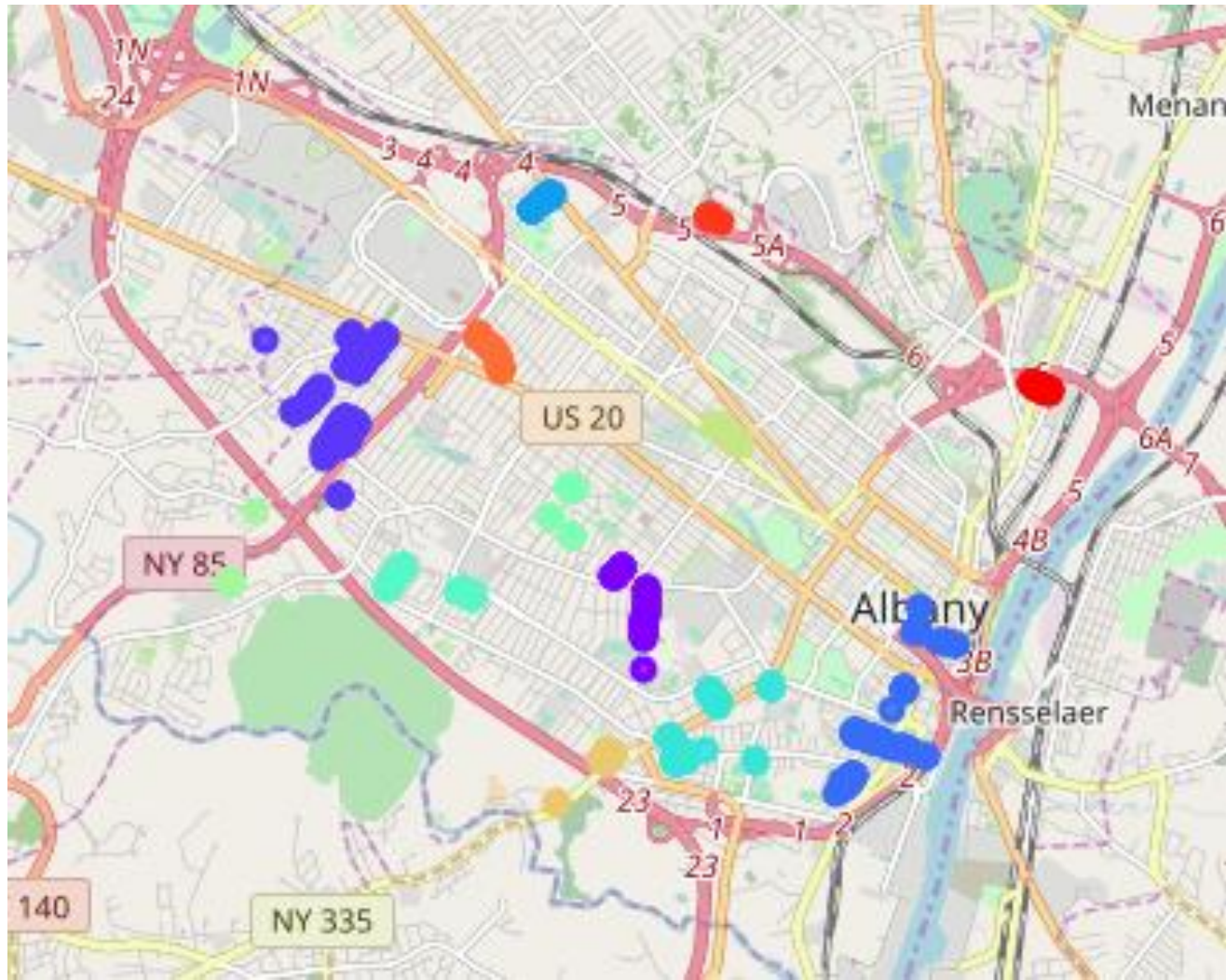
Clustering

Now I grouped the houses in the city by their location in order to understand how their location change their price. In order to understand how many groups should be clustered I checked the accuracy by SSD algorithm

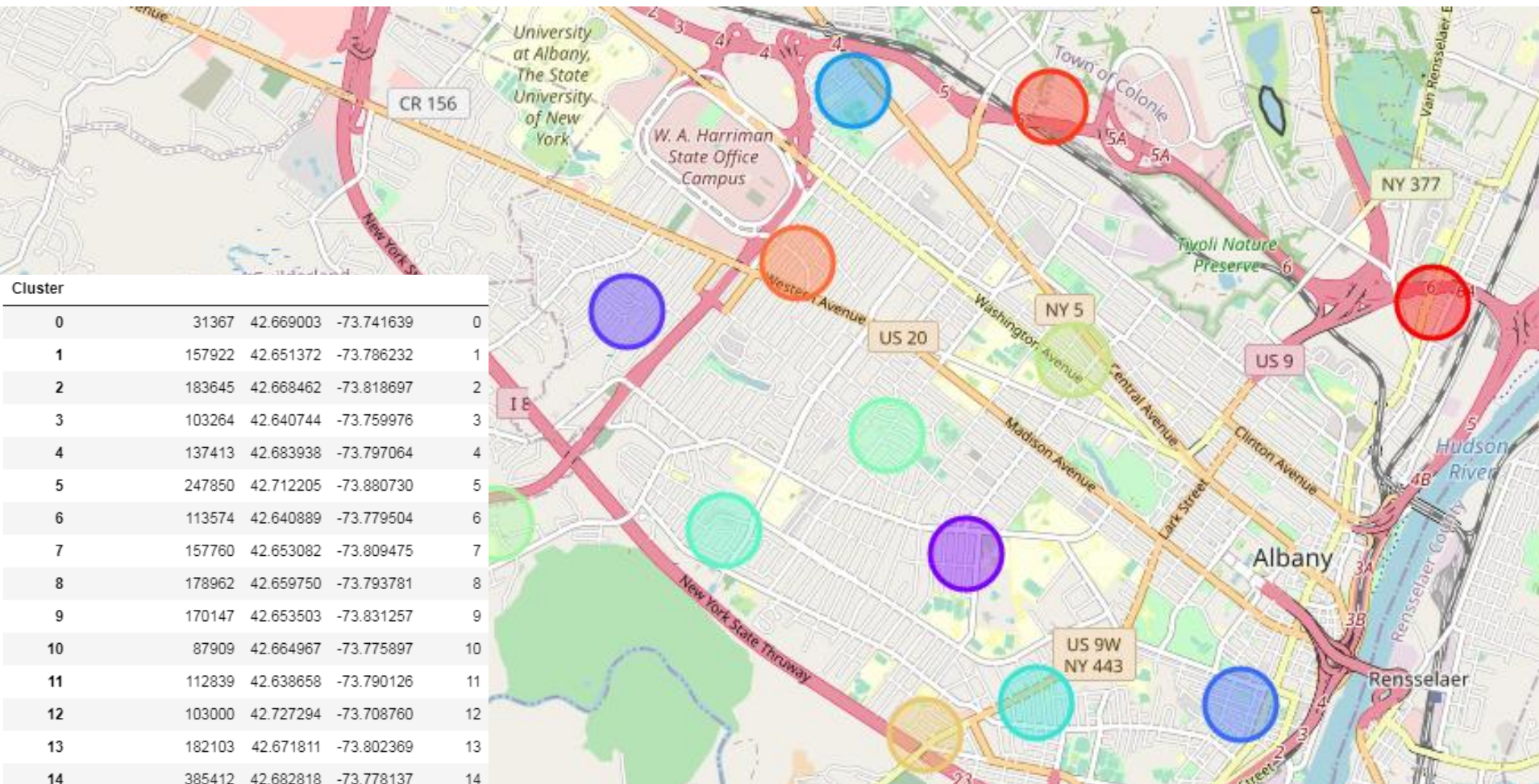15 clusters is chosen as seen to be good enough
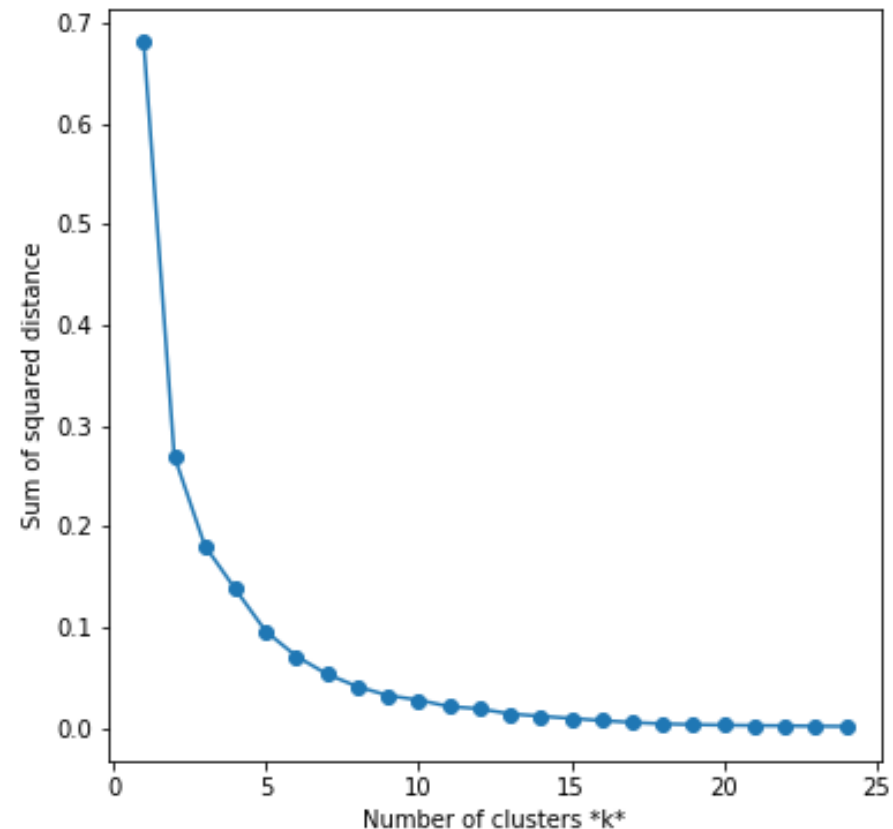
Clustering
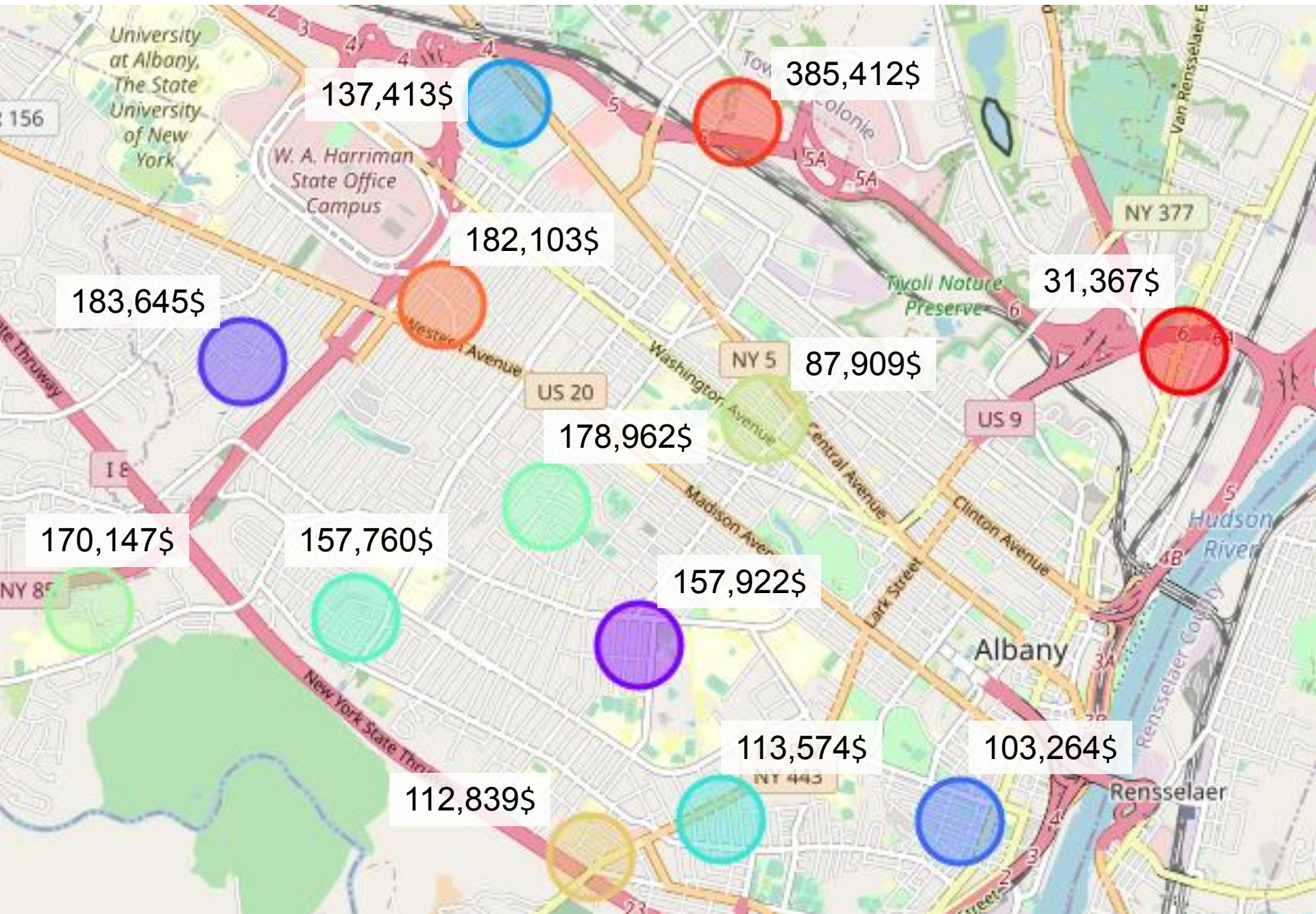
Clusters presented on map by Folium

Clustering

Then for each cluster housing prices were averaged and put on map

Cluster			
0	31367	42.669003	-73.741639	0
1	157922	42.651372	-73.786232	1
2	183645	42.668462	-73.818697	2
3	103264	42.640744	-73.759976	3
4	137413	42.683938	-73.797064	4
5	247850	42.712205	-73.880730	5
6	113574	42.640889	-73.779504	6
7	157760	42.653082	-73.809475	7
8	178962	42.659750	-73.793781	8
9	170147	42.653503	-73.831257	9
10	87909	42.664967	-73.775897	10
11	112839	42.638658	-73.790126	11
12	103000	42.727294	-73.708760	12
13	182103	42.671811	-73.802369	13
14	385412	42.682818	-73.778137	14

Clustering

Conclusions

- Houses closer to city center have higher prices
- Houses closer to highways have higher prices
- The highest prices are at the northern side of the city