

The Battle of Neighbourhoods

Cluster analysis of New York Real-estate Market

Sibonelo Ngobese

Introduction

Real-estate market relies on the surrounding amenities and facilities such as:

- Schools
- Supermarkets
- Recreational areas
- Hospitals and Police stations.

Business Problem:

Using New York City (NYC) property data and the Foursquare API, can we help home buyers make better decisions thus lifting the real-estate housing market?

Data and Methodology Section

Merged data from two New York sources for a more accurate dataset.

- Kaggle data set for Housing property prices
- 2014 New York neighbourhood names

Methodology:

- Data collection
- Exploratory data analysis and understanding
- More cleaning and pre-processing
- Modelling using K-means

Results:

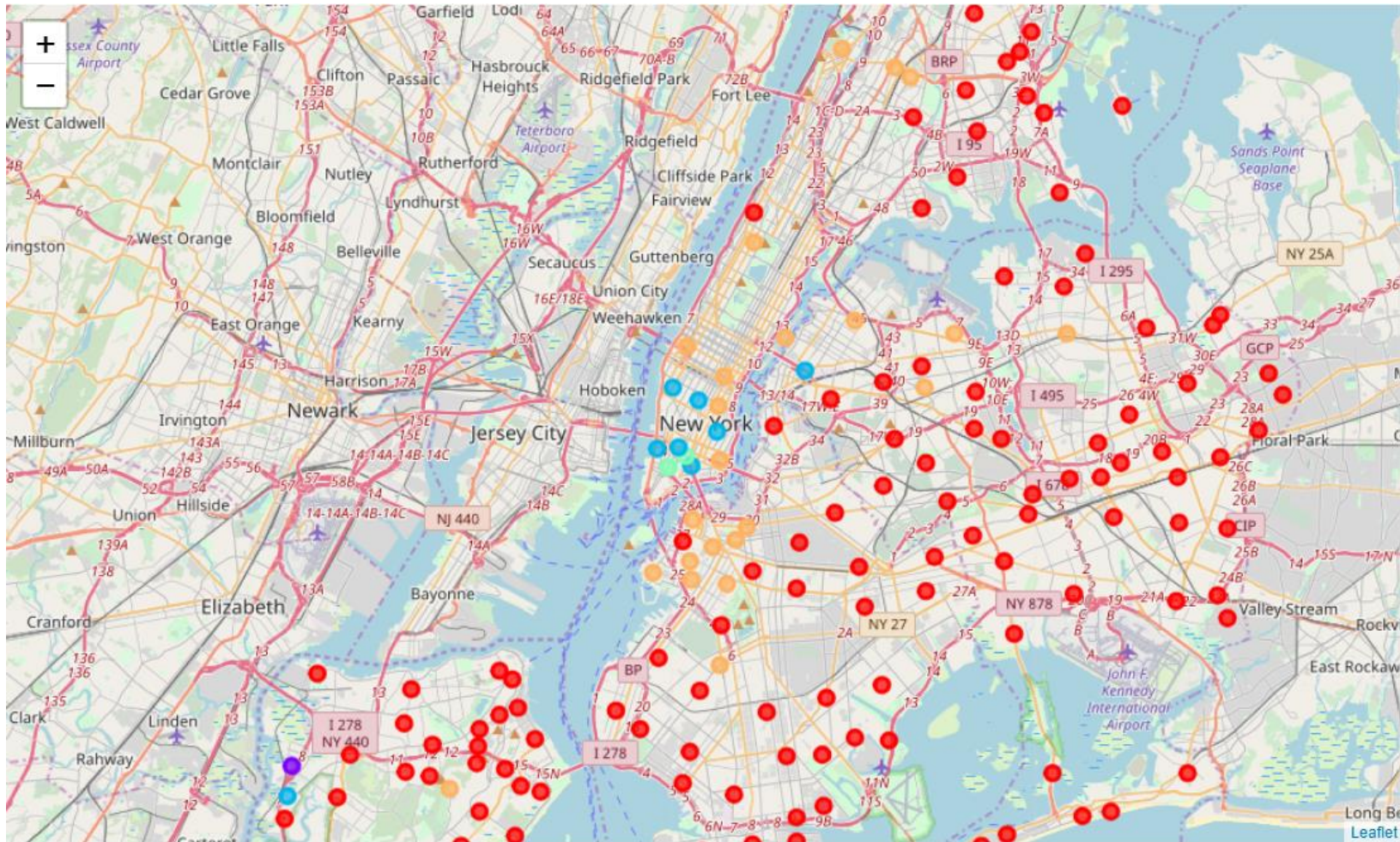
- Began by **averaging** the neighbourhoods by house price

	Neighborhood	Latitude	Longitude	PRICE
0	ANNADALE	40.538114	-74.178549	6.280466e+05
1	ARDEN HEIGHTS	40.549286	-74.185887	3.949565e+05
2	ARROCHAR	40.596313	-74.067124	5.700459e+05
3	ARVERNE	40.589144	-73.791992	3.794519e+05
4	ASTORIA	40.768509	-73.915654	1.399695e+06
5	BATH BEACH	40.599519	-73.998752	5.195168e+05
6	BAY RIDGE	40.625801	-74.030621	5.653236e+05
7	BAYCHESTER	40.866858	-73.835798	3.366427e+05
8	BAYSIDE	40.766041	-73.774274	6.825234e+05
9	BEDFORD STUYVESANT	40.687232	-73.941785	8.224549e+05



Results (Cont'd)

- Performed K-means clustering, unsupervised learning technique, $k=5$.



Conclusion

- The top areas/neighbourhoods that seem to be good for real-estate business are **Annadale, Aden Heights, Arrorchar, Arverne** and **Astoria**. Keep in mind these choices are due to the surrounding facilities that are mostly needed by any home buyer to be in the vicinity of the property.