The Battle of Neighbourhoods – Report

Introduction

This report is meant to help people in real-estate and home buyers when it comes to determining which neighbourhoods have the best home prices. This is also related to what facilities are near or available i.e. elementary schools, high schools, hospitals & grocery stores and recreational centres.

Business Problem

Can we use house price data to determine what neighbourhoods in New York City (NYC) have high prices thus helping agents who are starting business in real-estate give valuable advice to people looking to purchase houses? Can we use the Foursquare API to determine the neighbourhoods of interest?

To solve this business problem, we are going to cluster NYC neighbourhoods in order to recommend venues and the current average price of real estate where homebuyers can make a real estate investment. We will recommend profitable venues according to amenities and essential facilities surrounding such venues i.e. elementary schools, high schools, hospitals & supermarkets.

Data section

In this section we used the following sources for data:

- https://www.kaggle.com/new-york-city/nyc-property-sales#nyc-rolling-sales.csv) for house prices.
- The 2014 New York City Neighbourhood names https://geo.nyu.edu/catalog/nyu_2451_34572

These two sources were merged (using an INNER JOIN) and parsed to Foursquare using the neighbourhood names and location in terms of latitude and longitude.

Methodology section

The Methodology section will describe the main components of our analysis and predication system. The Methodology section comprises four stages:

1. Collect Data: As aforementioned, we use data from NYC properties along with 2014 Neighbourhood names data sets to merge our data to accurately pin point the neighbourhoods.
2. Data understanding and exploration
3. Data preparation and pre-processing
4. Modelling

Our property data averaged 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

The map of New York City
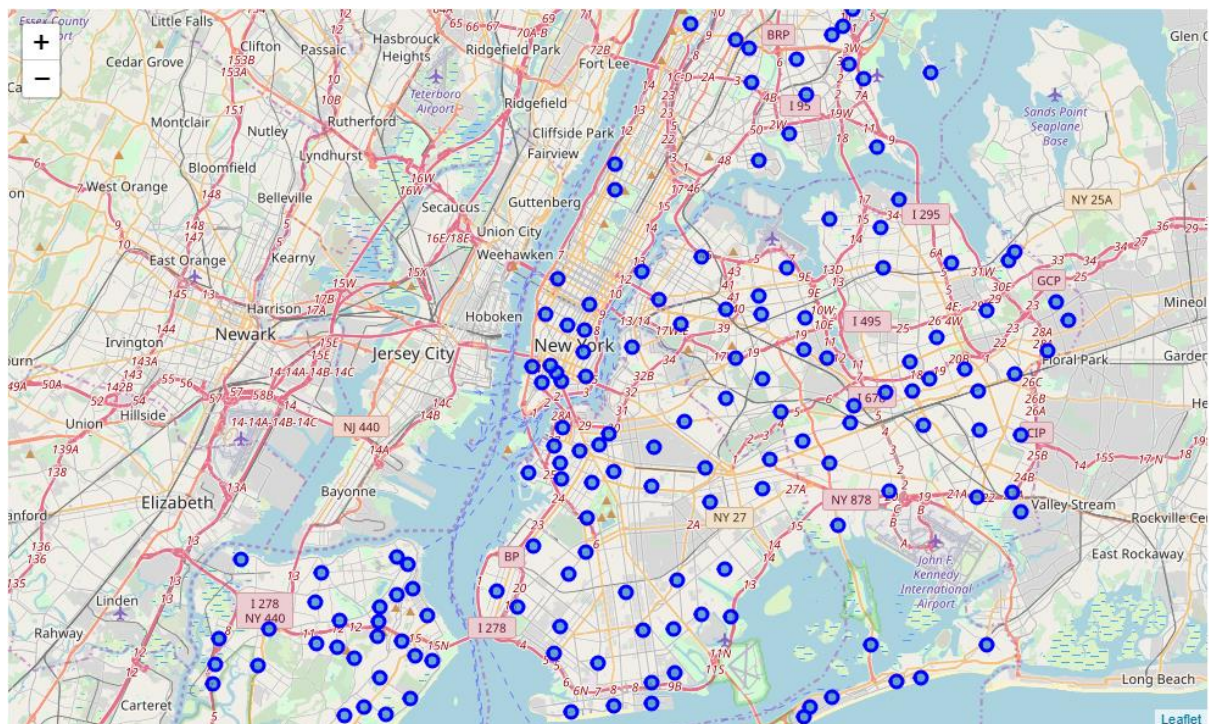
Map after K-Means Clustering
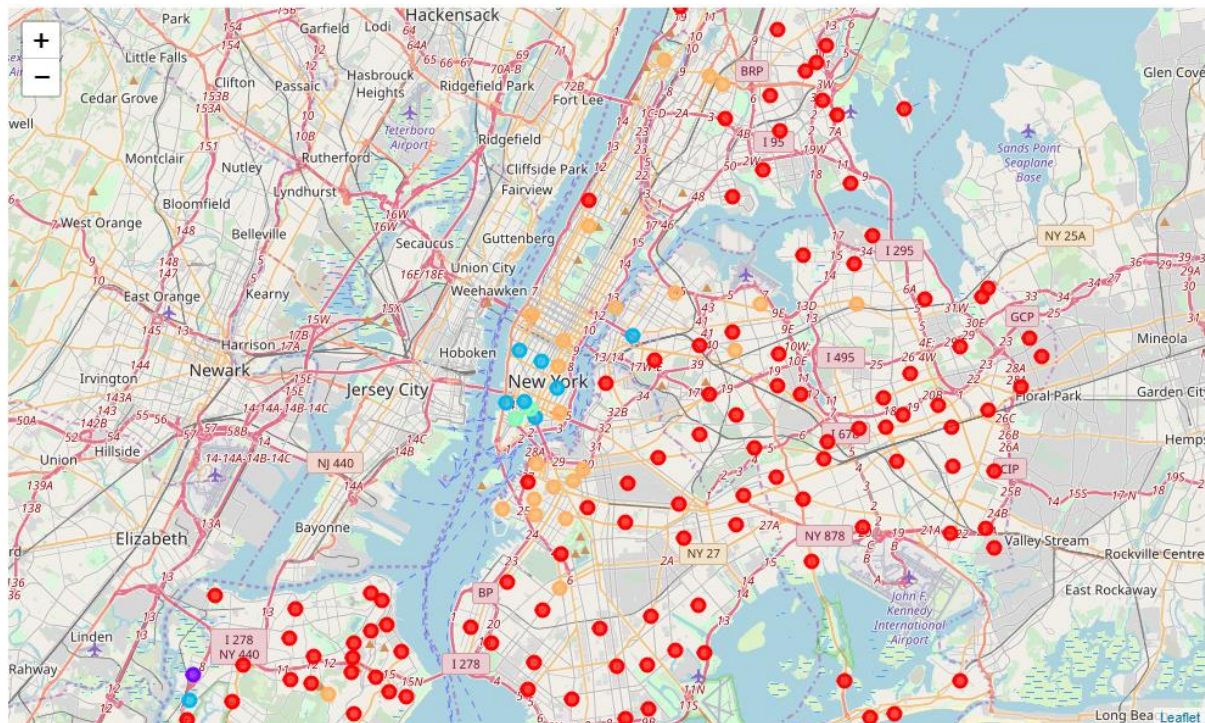
Map of NYC after K-means clustering has been applied. 5 clusters where chosen for the technique.

Results – Discussion

We performed a clustering technique on New York neighbourhoods in order to recommend venues and the current average price of real estate where home-buyers can make a real estate investment. We recommended profitable venues according to amenities and essential facilities surrounding such venues i.e. elementary schools, high schools, hospitals & grocery stores.

We started by collecting New York city rolling prices of properties data from Kaggle (https://www.kaggle.com/new-york-city/nyc-property-sales#nyc-rolling-sales.csv) and the relative price paid data were extracted from . Moreover, to explore and target recommended locations across different venues according to the presence of amenities and essential facilities, we accessed data through Foursquare API interface and arranged them as a data frame for visualization. Integrating data from Foursquare API and the New York property we suggest which neighbourhoods are great for real-estate investment business based on the surrounding facilities and amenities.

The Methodology section can be broken down into four stages:

1. Collect Data

2. Explore and Understand Data

3. Data preparation and preprocessing

4. Modelling

In the modelling section, we used the k-means clustering technique as it is fast and efficient in terms of computational cost, it is highly flexible to account for mutations in real estate market in New York and it is accurate.

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 trending and well placed in terms of availability.