Moving to London

Gennadiy Garshin, May 2020

Clustering segmentation for a daily routine

The family faced the question about moving to London. Having visited the city several times it was decided to choose the southern part of the capital. In this research I will try to find several options for possible neighborhoods to move to.

- The starting point of a comfortable stay will be the current place of living in Saint-Petersburg.

- The analysis will be made on the basis of open sources such as foursquare, wikipedia, and uk government.

- Clustering will help us to determine which borough is the most similar for a current place of living.

- To visualize result data we will see clusters on the map.

- To compare boroughs we will use the charts to see comparisons.

Data preparation

In the world of data there are a lot of open sources available for analysis. In this analysis we will use:

- Wikipedia pages,
- the official website of the United Kingdom government,
- the foursquare database to get geo-information using api requests.

Clustering Segmentation

Clustering segmentation will help us to segment up parts of London into several groups.

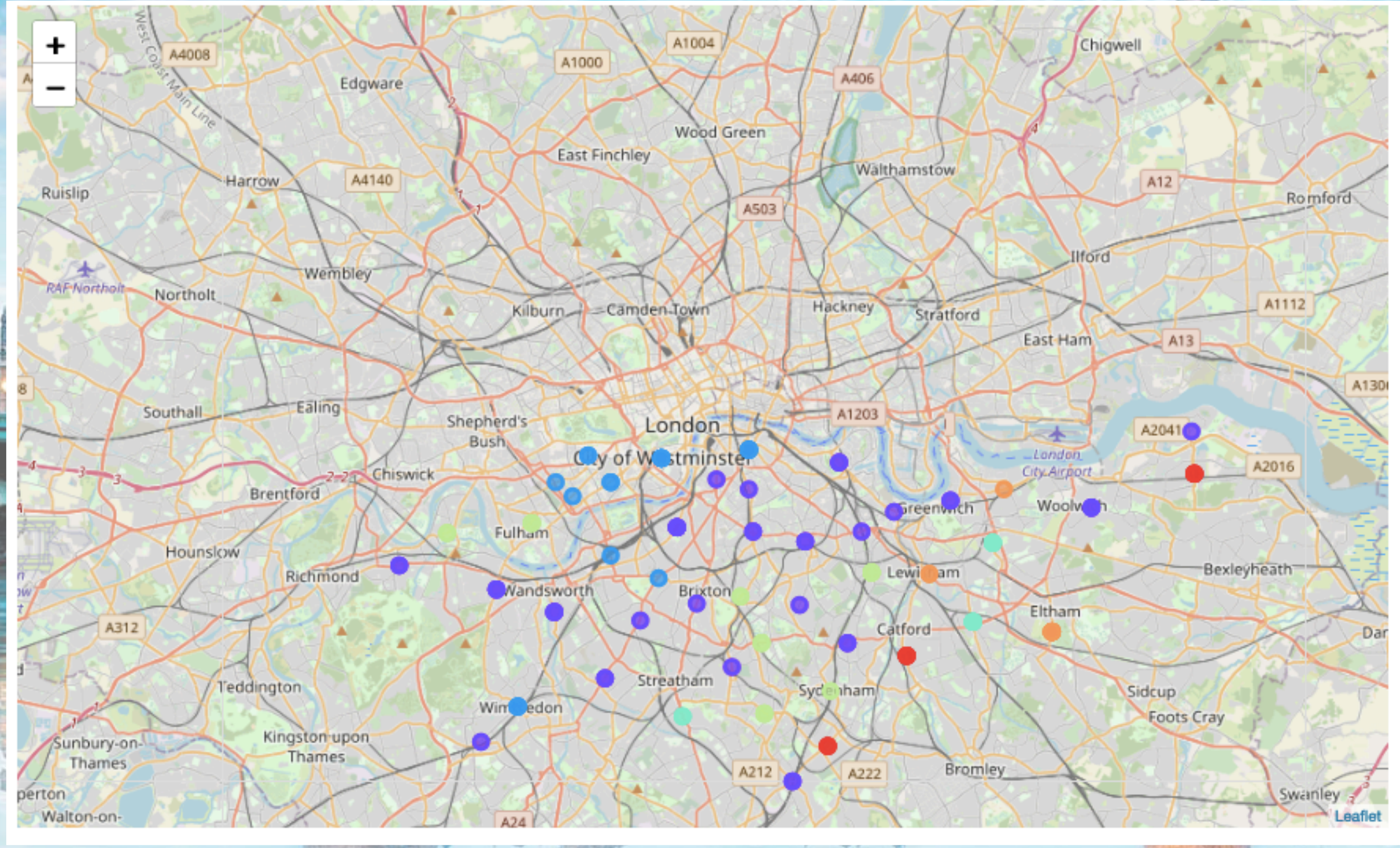
Then we will be able to find the current borough and view the cluster in more detail.

To evaluate the correctness of the data after running the algorithms, we can view the result on the map and detailed data for each cluster in a table view.

For clustering we will use k-means algorithm.

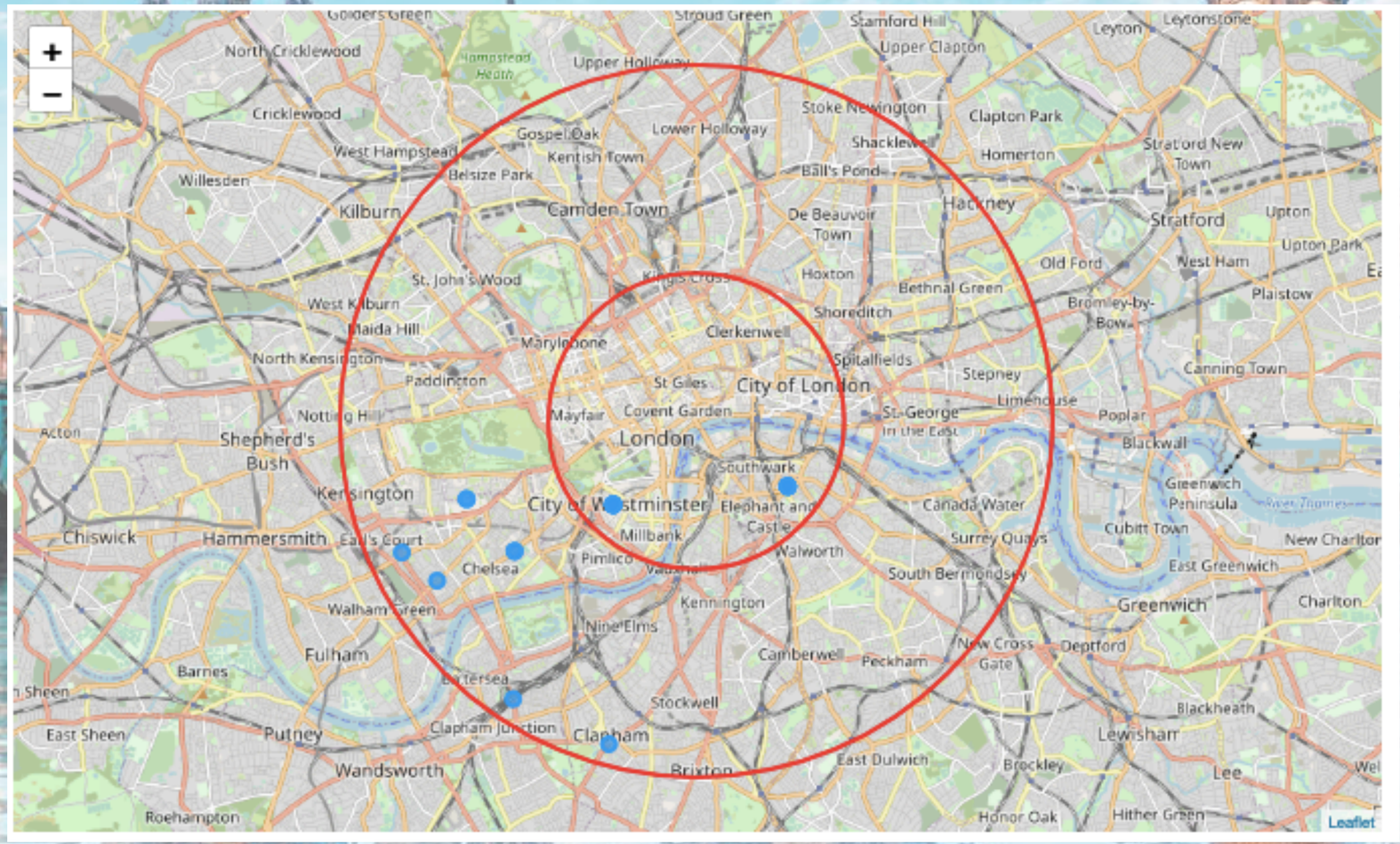
For showing result we will use the Folium Library.

Clustering result



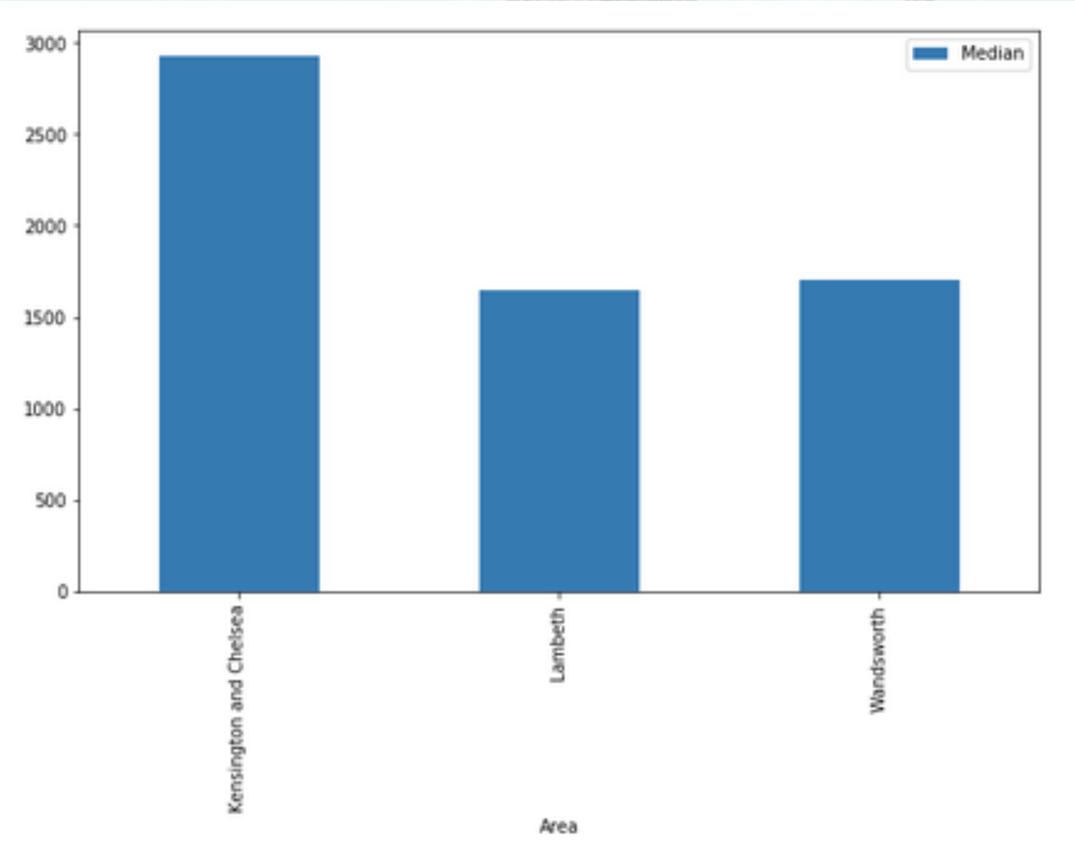
Playing with data

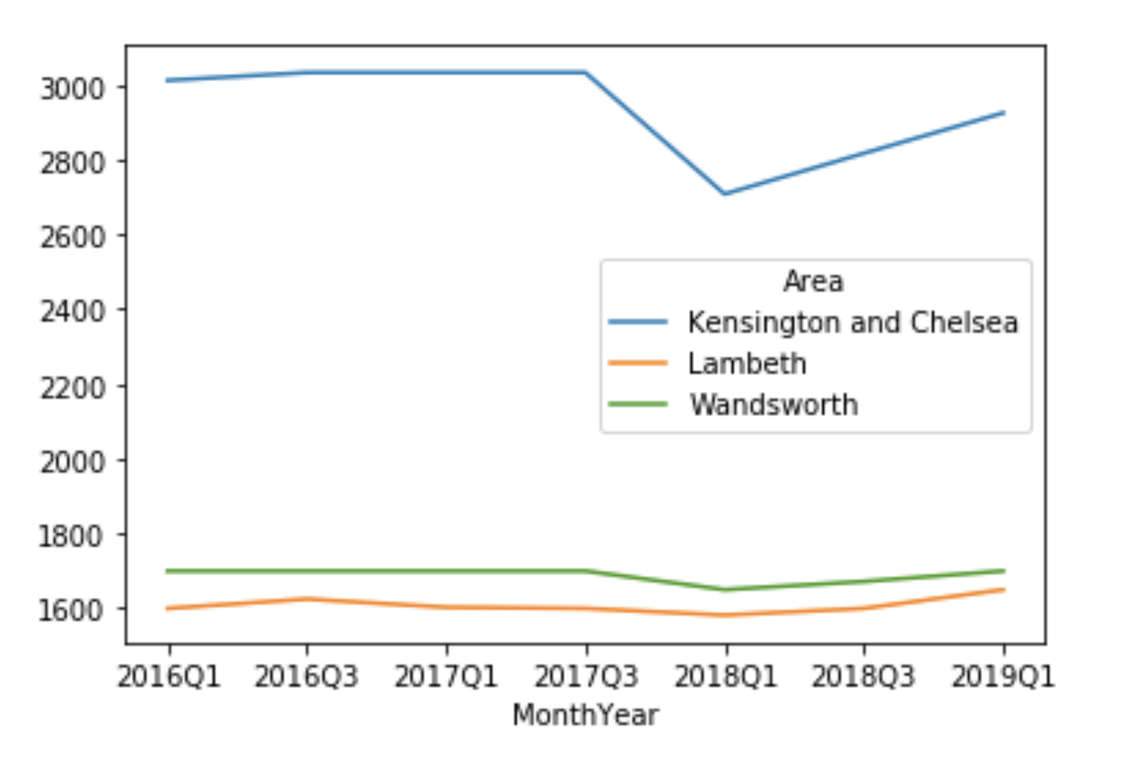
To see how many neighborhoods are included in chosen area, we can add two circles with different radius - 2.5km and 6km. This distances are maximum and minimum distances from the city centre.



Borough comparison

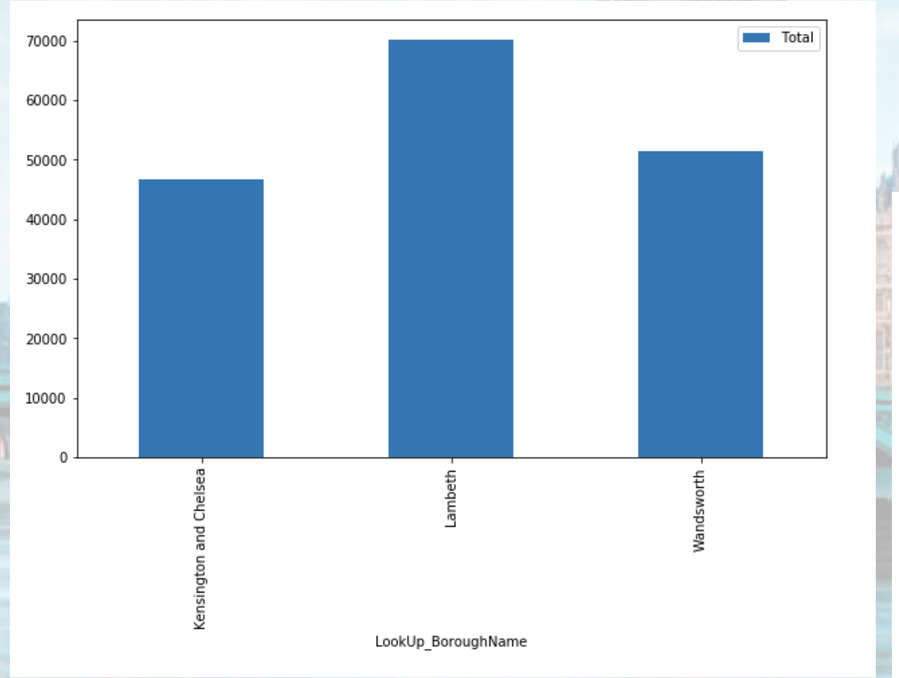
Once we have a list of boroughs, we can compare them by price characteristics and crime situation.

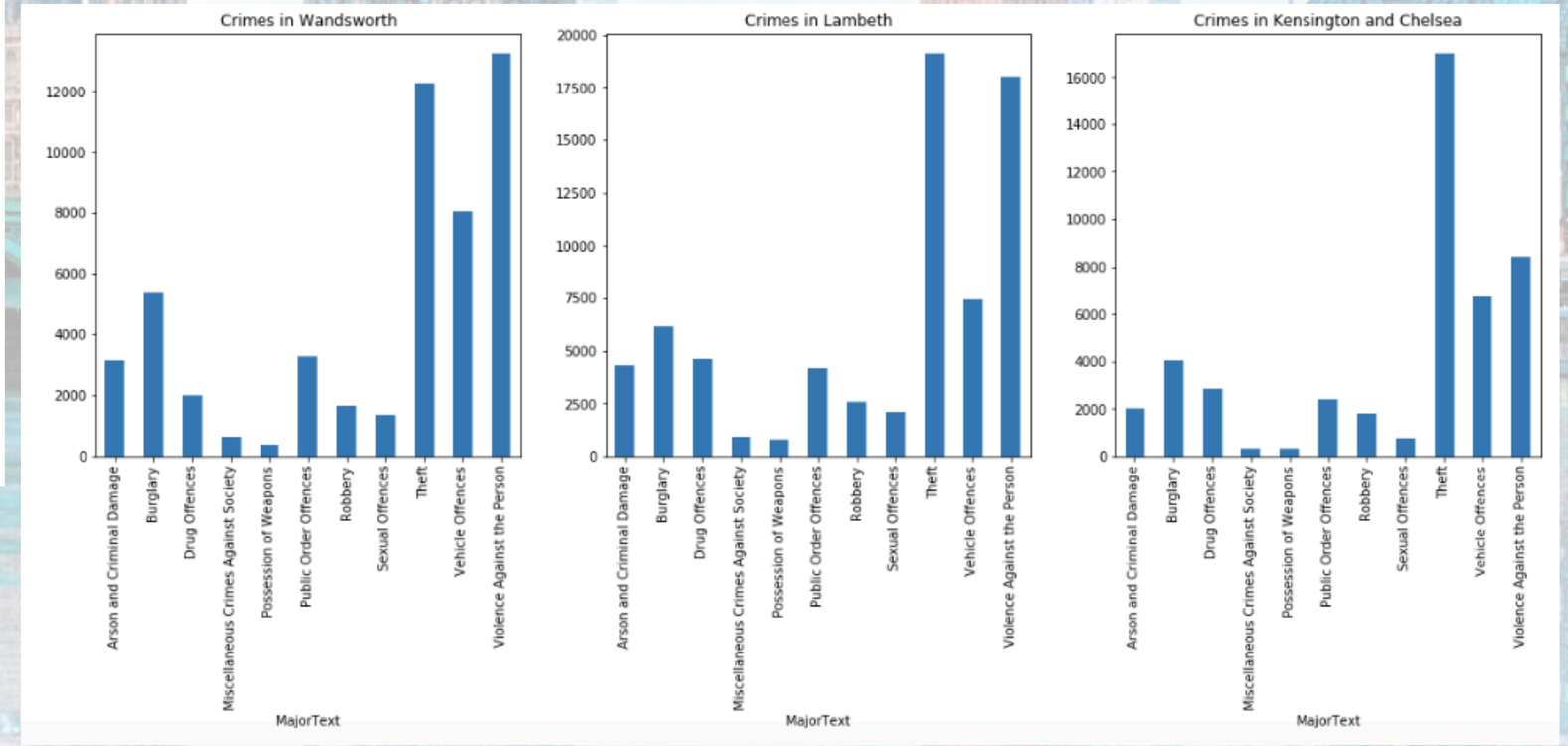




Crime situation

On those charts we can see number of incidents and number of crimes by Crime Types for each borough.





Conclusion

In this report we made some work to identify similar neighborhoods in two different cities. Cluster segmentation was selected as the definition method. As a result, we have received a list of boroughs that meet the conditions. Then we analyzed the received boroughs and looked at them from different angles. Based on the data obtained, we have prepared a list of priority boroughs.