

# Battle of the Neighborhoods

Coursera capstone project – IBM Data science

# Introduction and objective

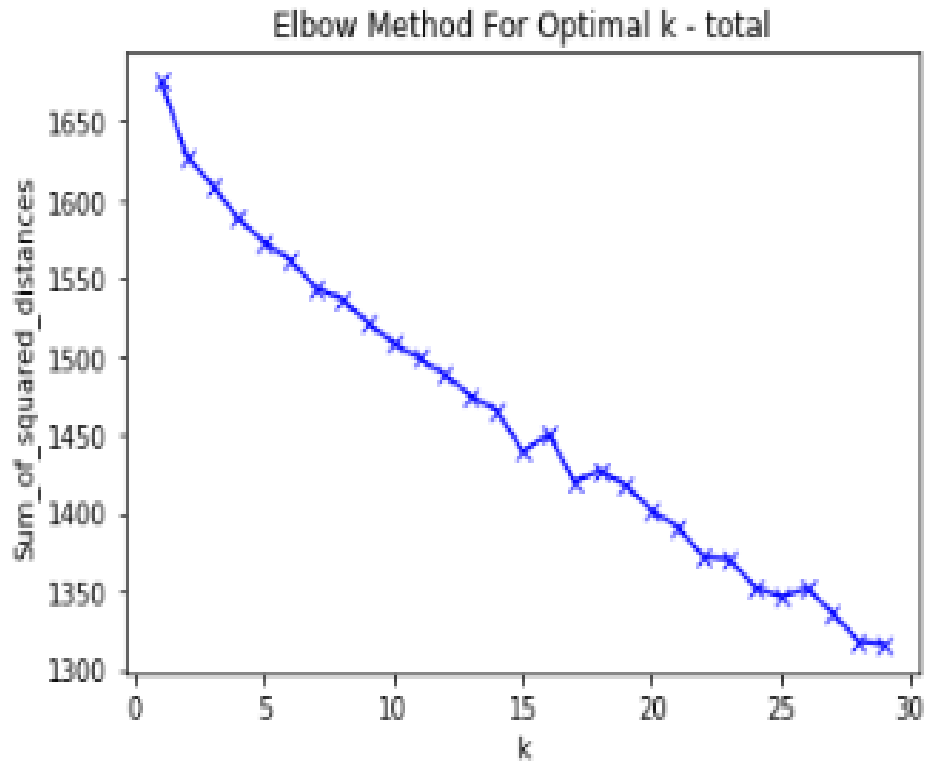
- A businessman has lived in New York, United states and now lives in Toronto, Canada. He has been asked to work for Fictive incorporation in Europe. They have offices in Paris and Berlin. He has a beautiful wife and two children. In New York and Toronto they have lived in nice neighborhoods. He does want to work in Europe, but in a similar Neighbourhood. In this project we will similarities between neighborhoods in the different cities.
- We are going to use the k-means clustering algorithm to spot (dis)similarities between neighborhoods in different cities. We are going to find venue data using the foursquare API.

# Data

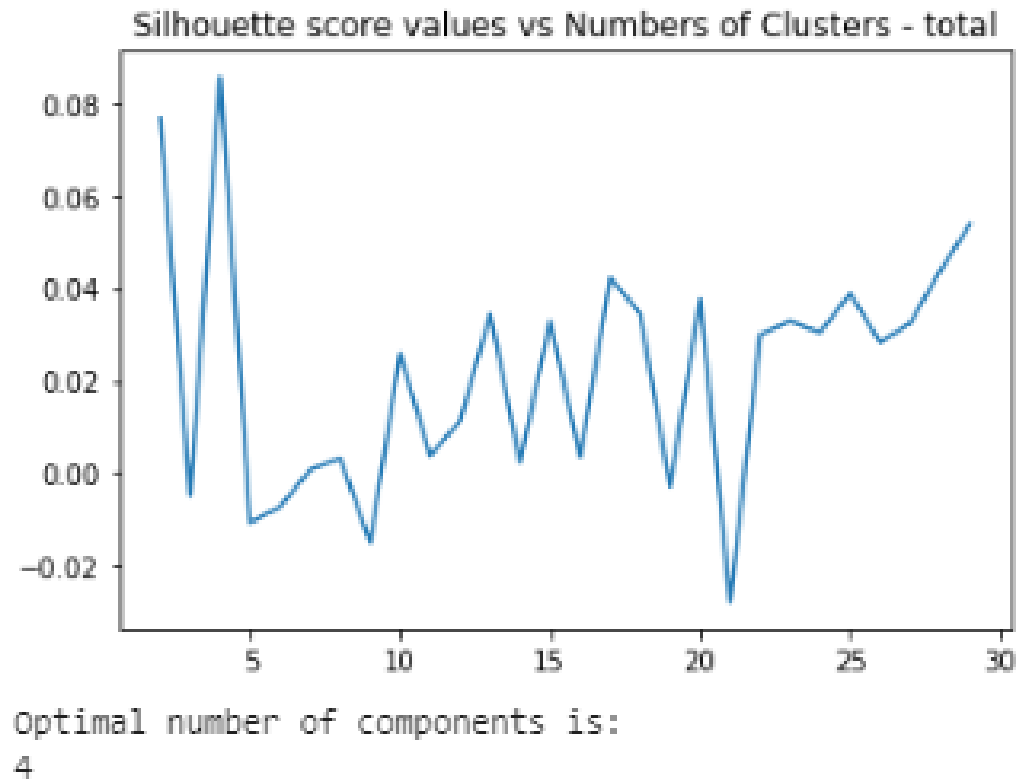
- Data containing boroughs of new York: [https://cocl.us/new\\_york\\_dataset](https://cocl.us/new_york_dataset)
- Data containing boroughs of Toronto:  
[https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)
- Data containing boroughs of Berlin:  
[https://en.wikipedia.org/wiki/Boroughs\\_and\\_neighborhoods\\_of\\_Berlin#External\\_links](https://en.wikipedia.org/wiki/Boroughs_and_neighborhoods_of_Berlin#External_links)
- Data containing boroughs of Paris:  
[https://en.wikipedia.org/wiki/Arrondissements\\_of\\_Paris](https://en.wikipedia.org/wiki/Arrondissements_of_Paris)
- Additional Data: Additional data will be gathered by using the Foursquare API.
- Use of data:
- In this project we will gather venue information about the different Boroughs and Neighbourhoods in the four different cities. This will be done by using the foursquare API. Some of the borough data has already been collected, other data will be collected by use of Beautifulsoup and/or pandas.

# Methods for finding number of K's

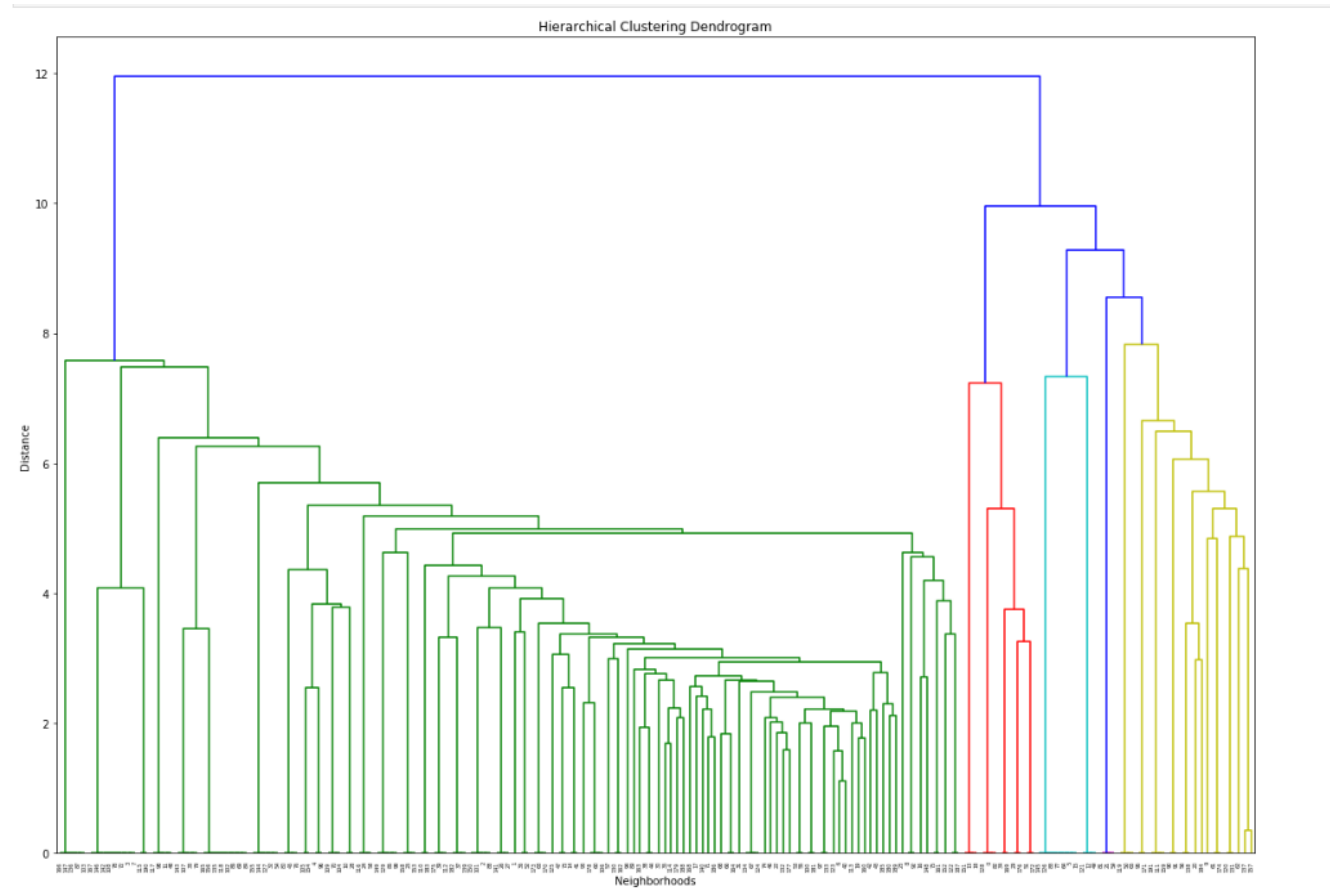
- Elbow method



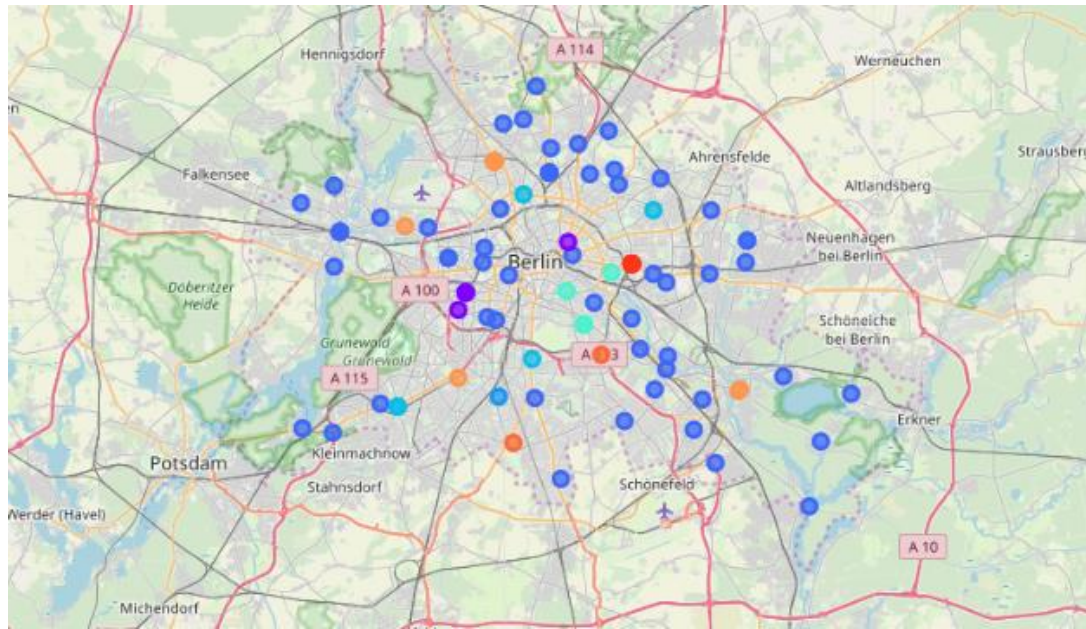
- silhouette



# Dendrogram

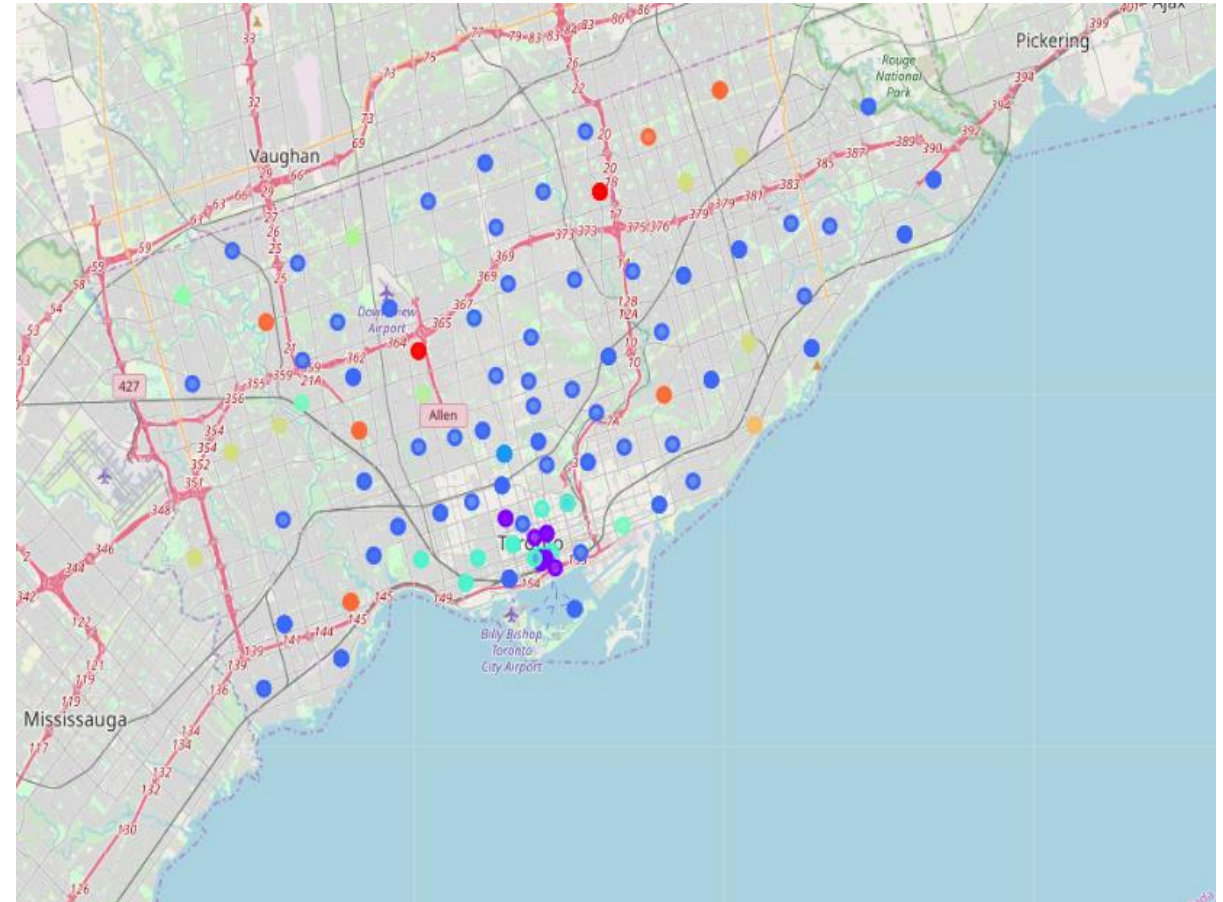
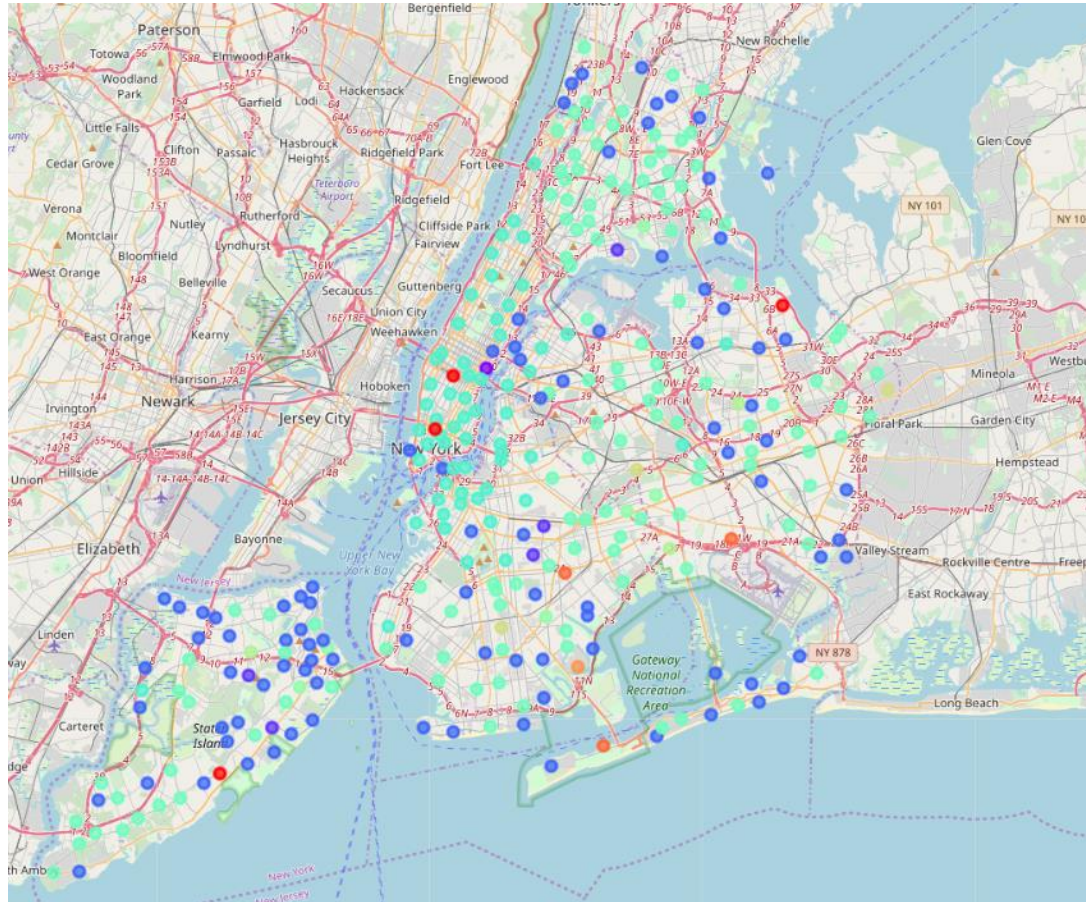


# Berlin and paris clustering





# New york and toronto



# Conclusion

- In general there are differences to be seen in the clusters. However, to really spot differences between neighborhoods in the different cities, the model has to be refined. More variables can be added. The foursquare data also has to be examined more closely. Example: sushi restaurant, ramen restaurant, udon restaurant can all be classified as Japanese restaurants. This will refine the model, so better differences can be spotted.
- Cluster 0: French dining area.
- This cluster contains mostly French restaurants in different cities.
- Cluster 2: Common area
- This cluster contains a mix of restaurants, bars, supermarkets and banks.
- Cluster 6: Food court
- This cluster contains mostly restaurants from different nationalities.
- Cluster 7: Traffic zone
- Lots of bus stops in this area.
- Cluster 13: Sports and events
- A lot of event places and gyms are found in this cluster.