

Comparing Berlin and Amsterdam neighborhoods

Krzysztof Zwolenik

1. Introduction

Berlin and Amsterdam are two big cities of Europe that could share some similarities between them.

There are people that know one city pretty well and are wondering which neighborhood of the other city is most similar to the one they like.

They could use this data to choose location of their hotel or apartment when visiting or moving to the other city.

2. Data

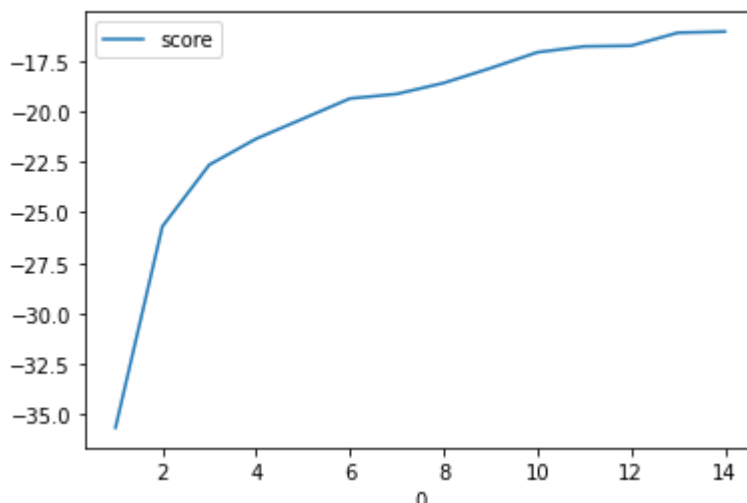
On the internet (website cybo.com) there are available basic information about the regions of both cities:

- postal code
- median age
- neighborhood names
- area
- population
- geo location

Additionally, it is possible to get venues and points of interest info about each region using foursquare API (foursquare.com), all that is needed is geolocation info from cybo.com

3. Methodology

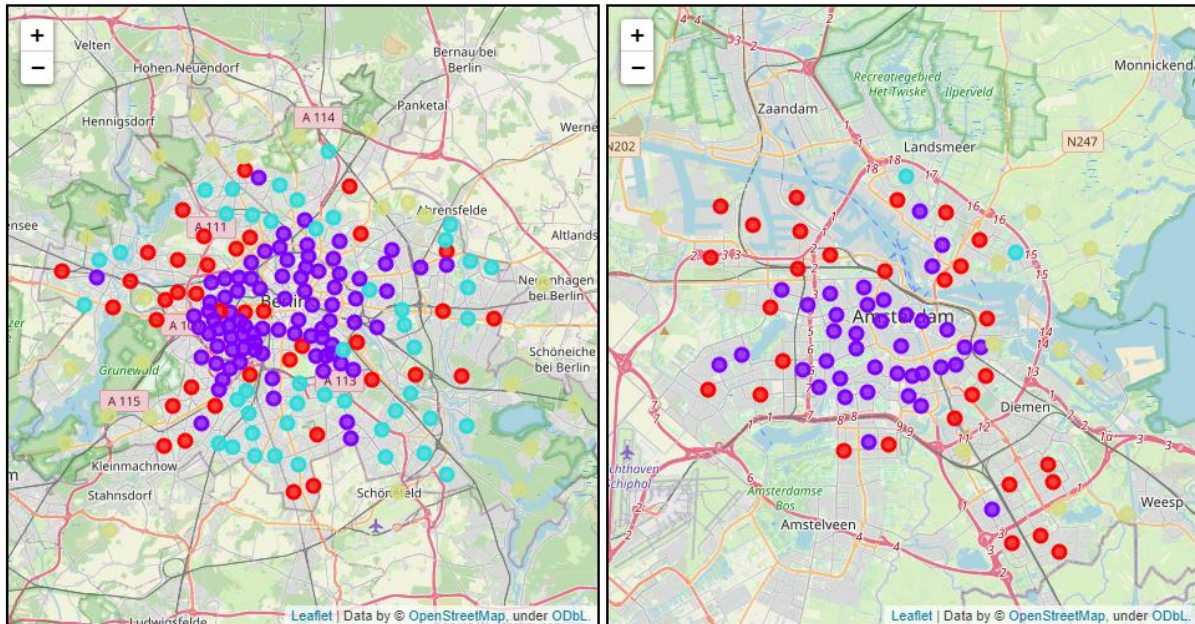
I will use the postal codes areas of both cities and I will get nearby venues in the radius based on area size. The density field is calculated for each postal code to represent the area and population relationship. The K-Means algorithm is going to be used. I will calculate the score for various k values to get optimal k value.



Based on calculations the best k is 4

Because the clustering algorithm is unsupervised, I will use map and top 10 venues for each cluster to determine properties of each cluster.

4. Result



Both cities were divided in 4 clusters, they are arranged in a shape that is similar to ring or circle, that is expected of cities, further from the center the type of neighborhood changes the most. It's clearly visible that cluster 1 represents city centre, cluster 0 outer part of the city, cluster 2 and 3 also are located on the most outer part of the city, I need more data on the venues to distinguish them better.

Cluster 0

Outer part of the city still close to center, this area still has plenty of food venues, but is characterized by more utility venues like supermarket and other stores, gyms.

Cluster 1

This cluster is located in the city center, the area is full of cafes, restaurants, bars, generally it's the center of entertainment. It's also a densely packed area. Most outer part of the city

Cluster 2

This cluster is dominated with supermarkets and other stores, bus and tram stops. There are way less restaurants, we can find also venues like warehouse stores. This cluster has a similar population density to cluster 0. There are only two places in Amsterdam within this cluster.

Cluster 3

This cluster is similar to above cluster but there are way less venues, there is presence of lakes. This cluster has the least dense population

5. Conclusion

This project intention was to find if the two cities share some similarities and if so, which part of the cities share these similarities. By clustering both cities as they were one entity, I was able to find similar areas in both cities. Tourist could use the information to know which zones could be of their interest, people who would like to move from one city to the other could use this information the choose better where to get an apartment based on their preferences.