**Giving recommendations about moving company office from New York to Berlin**

**Boiarchenkov Ramil**

**April 08, 2020**

1. **Introduction**

Company office is located in one of the neighbourhood in New York. But because of business reasons stakeholders decided to move the office to Berlin. People are used to the infrastructure that they had in New York neighbourhood and they want to feel the same level of comfort in Berlin.

**The target audience** - company stakeholders and employees.

**The problem** - selecting the most similar neighbourhood to New York neighbourhood in Berlin.

**The main reason** - having the same infrastructure and the same level of comfort.

We will need to leverage the Foursquare location data for all neighbourhoods in both cities to make the right decision.

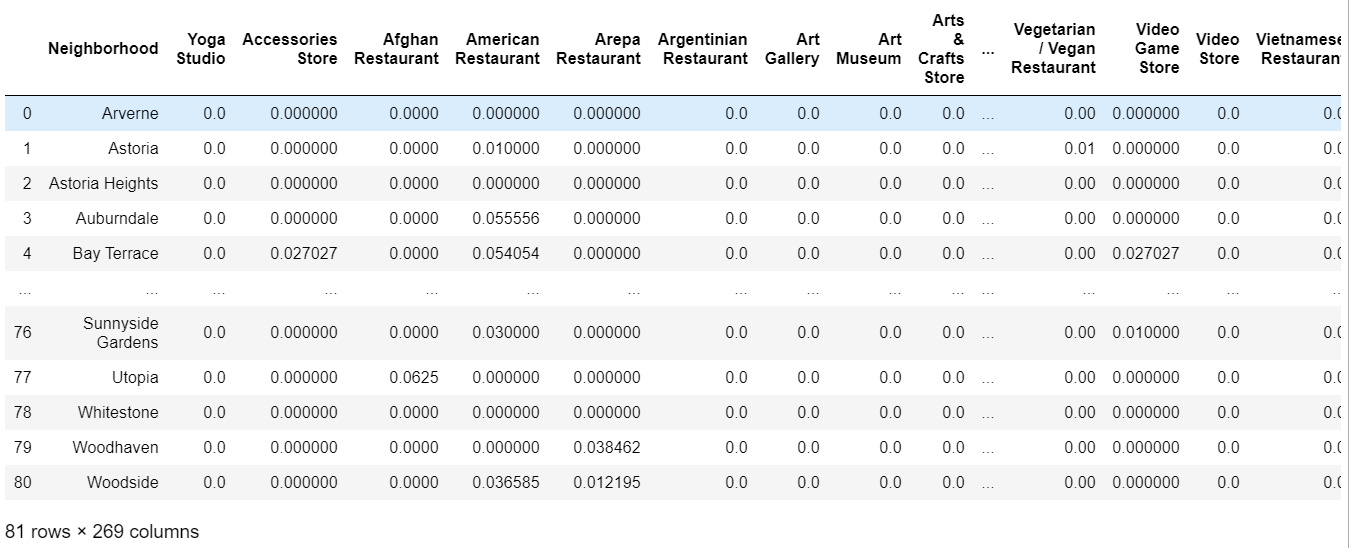
1. **Data**

We will take he Foursquare location data for all neighbourhoods in both cities. We will gather data on all vanues, preprocces it, so we have the mean amount of all vanues and cluster neighbourhoods in both cities.  
The final dataset will have the data on all vanues that are located in neighbourhoods. This will allow us to make a proper clustering.

1. **Methology**

**3.1 Data cleaning and gathering**

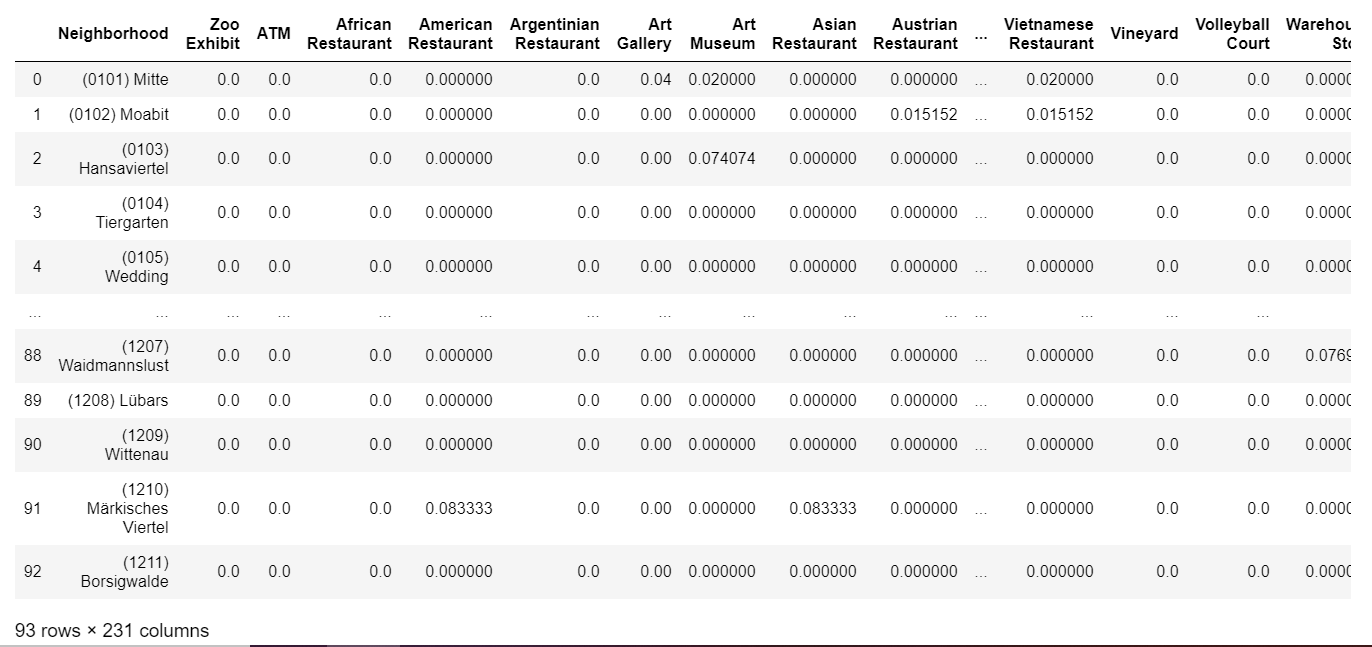
First, we repeat the same process for New York Queens as we did for Manhattan before. As a result we get the nest data frame:



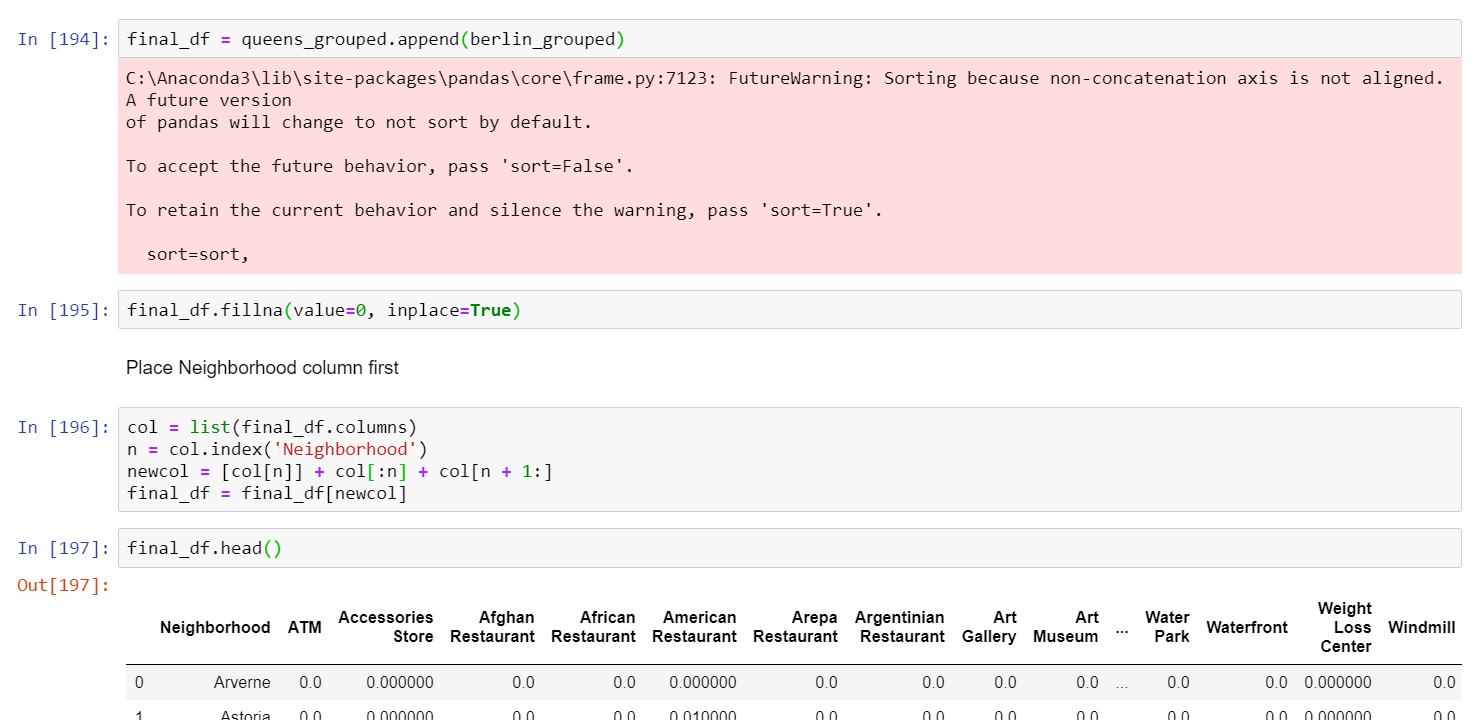
Next we will repeat the process for Berlin. First, we gather data about Berlin neighbourhoods from Wikipedia page ('https://en.wikipedia.org/wiki/Boroughs\_and\_neighborhoods\_of\_Berlin'):



Then we get coordinates as we did before and add information about venues:

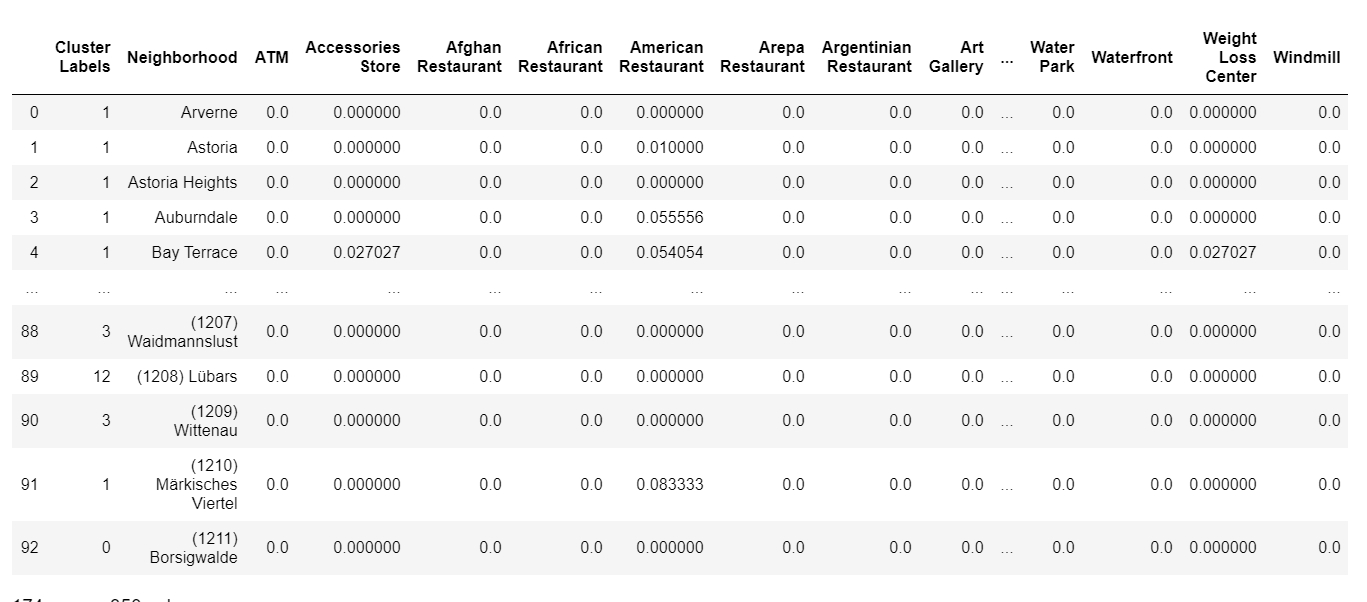


And finally, before clustering we gather data frames:



* 1. **Clustering**

We use machine learning algorithm of KMeans to cluster gathered neighbourhoods. After applying the method we get the next result, data frame with cluster labels:



Now we can recommend some pattern of behaviour for any given Neighborhood. But if company is located in 'Cambria Heights', it can move to any Berlin Neighborhood cause there is no any similar:



If company is located in South Ozone Park it can move to (0605) Dahlem Berlin Neighborhood. And it's one to one match.

1. **Results**

As a result of this project we achieved the next:  
1) We gathered the data on Berlin and New York Neighborhoods from Foursquare;  
2) We clustered Neighborhoods according to characteristics;  
3) We can say if there is a similar Neighborhood in Berlin to a given Neighborhood in New York.

1. **Discussion**

We noticed next interesting things:  
1) If New York Neighborhood is from the first cluster then it is very easy to find a matching Neighborhood in Berlin;  
2) Some Berlin Neighborhoods are really different and therefor they produce the third cluster;  
3) Even though cities are really different we can find some similar Neighborhoods.

We recommend:  
1) If company is located at 'Cambria Heights' to move to any Berlin Neighborhood cause there is no any similar;  
2) If company is located at 'South Ozone Park' to move to (0605) Dahlem Berlin Neighborhood. And it's one to one match;

We also can recommend some pattern of behavior for any given Neighborhood

1. **Conclusion**

In this project we clustered Neighborhoods of two different cities in order to recommend some Berlin Neighborhood for a company to move to.  
We gathered the full peacture of the situation and now can recommend some pattern of behaviour for any given Neighborhood in New York or Berlin.

Further developement can be next:

1) Gather more information about Neighborhood to make clustering more accurate;  
2) Gat new complex features from gathered ones for the same reason.