Clustering Neighborhoods in Berlin

A data analysis project to predict the best neighborhood for opening an Indian Restaurant

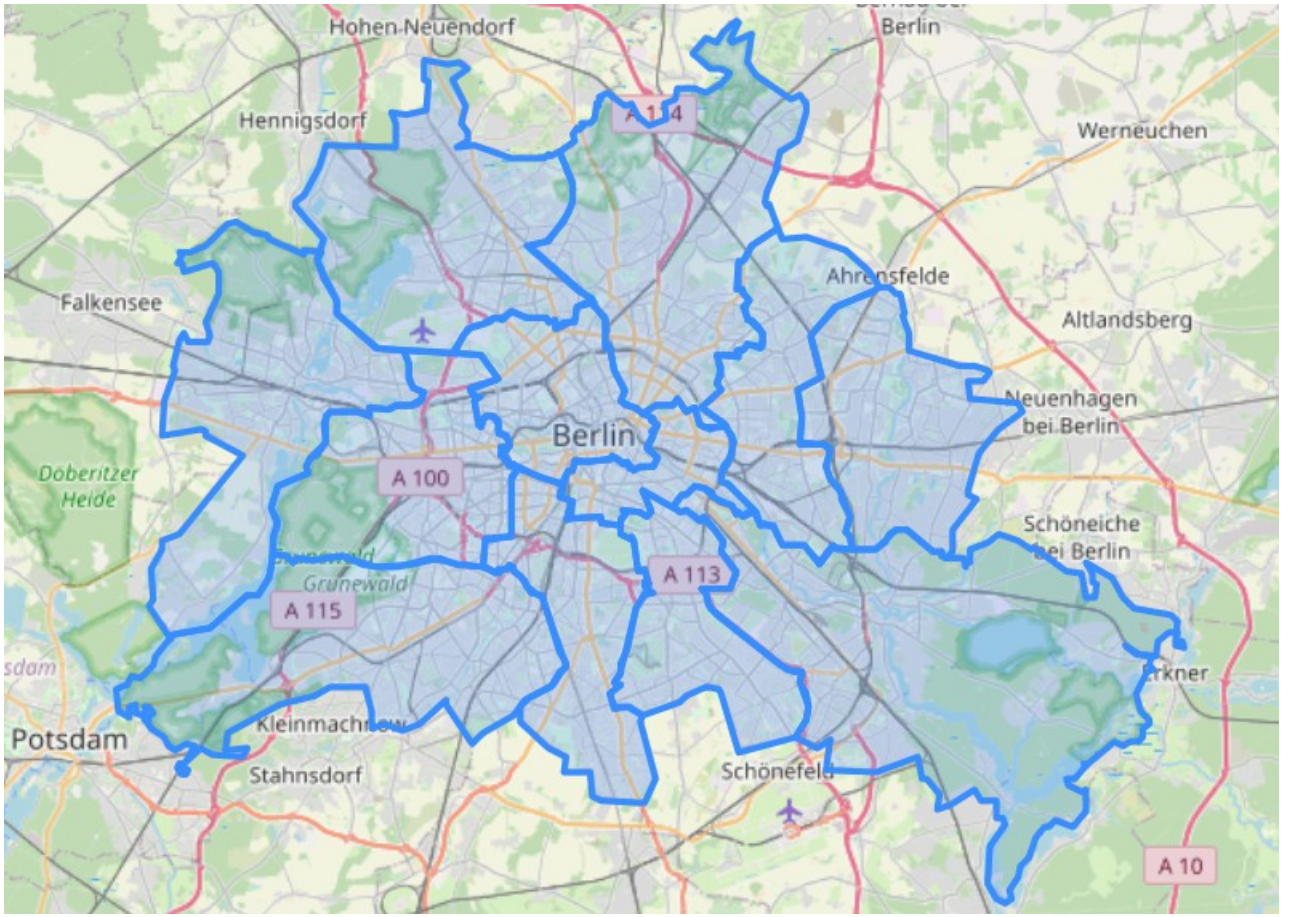
Why cluster analysis

- The cluster analysis will insure that we chose the suitable neighborhood
- It is not enough to chose a neighborhood which have no Indian restaurant. This neighborhood might be not suitable at all
- The cluster analysis allow us to chose the neighborhood where the market is not flooded with Indian restaurant, and the market is similar to the one where many Indian restaurant are open and successful

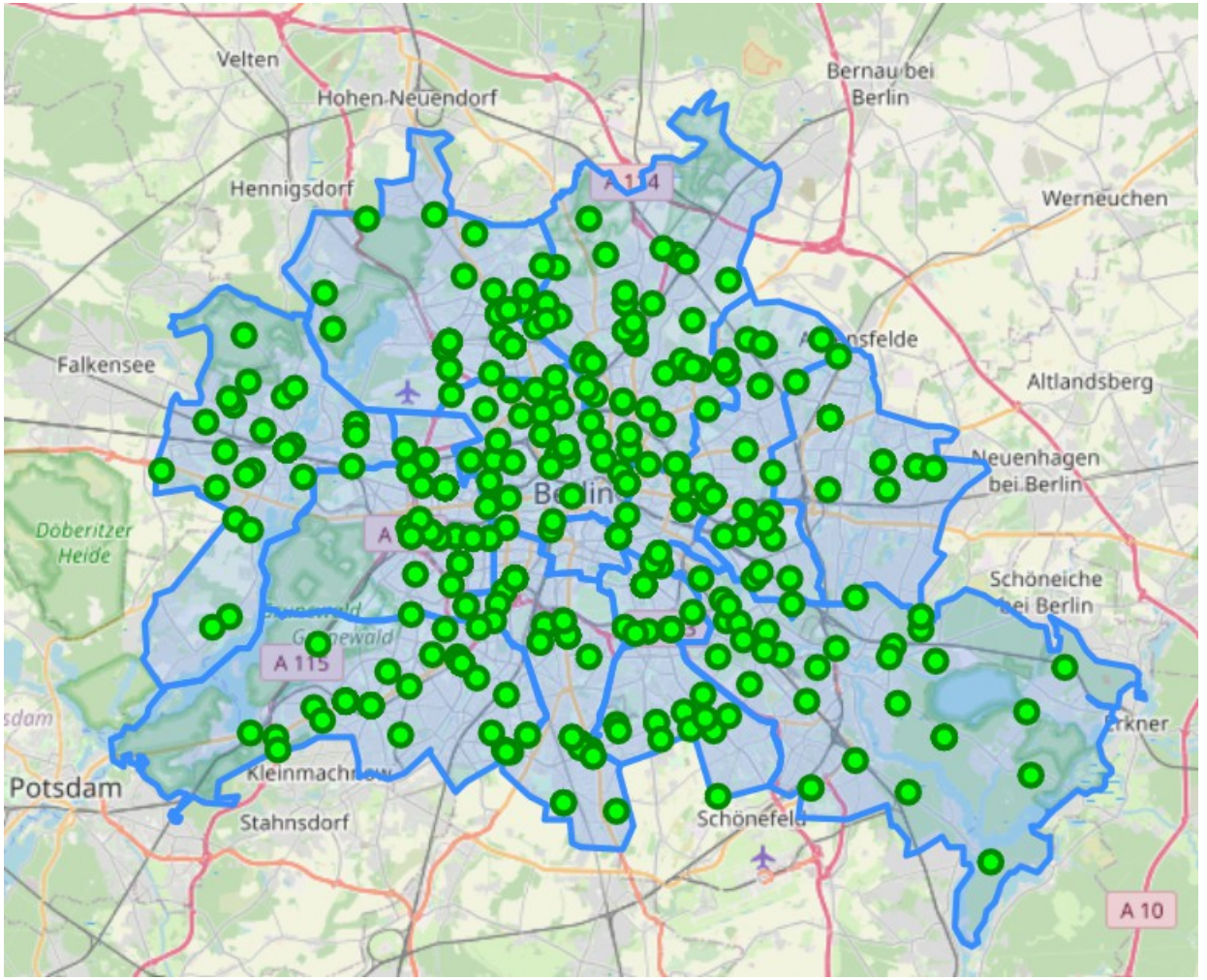
Data acquisition

- The Wikipidia page Verwaltungsgliederung Berlins provides a well ordered list of all the boroughs and there respictive neighboorhoods. This table will be extract using pandas library.
- The website Das Örtliche provides a full list of all Zip codes in Berlin which will be used later to get a better accuracy for addresses. the list will be extract using pandas
- Utilizing Geocoder the geographical coordinates of each zip code will be imported to serve as a center of neighborhood
- Using Openstreet maps from Google a geojson file for each Borough in Berlin as well as the whole city will be imported.
- Utilizing the Foursquare API all the venues will be imported for exploring the neighborhoods in Berlin and later with cluster analysis searching for the neighborhoods that are good for opining an Indian restaurant
- In total 12 Borough with 463 Neighborhoods with their zip codes are in the cleaned and prepared data frame

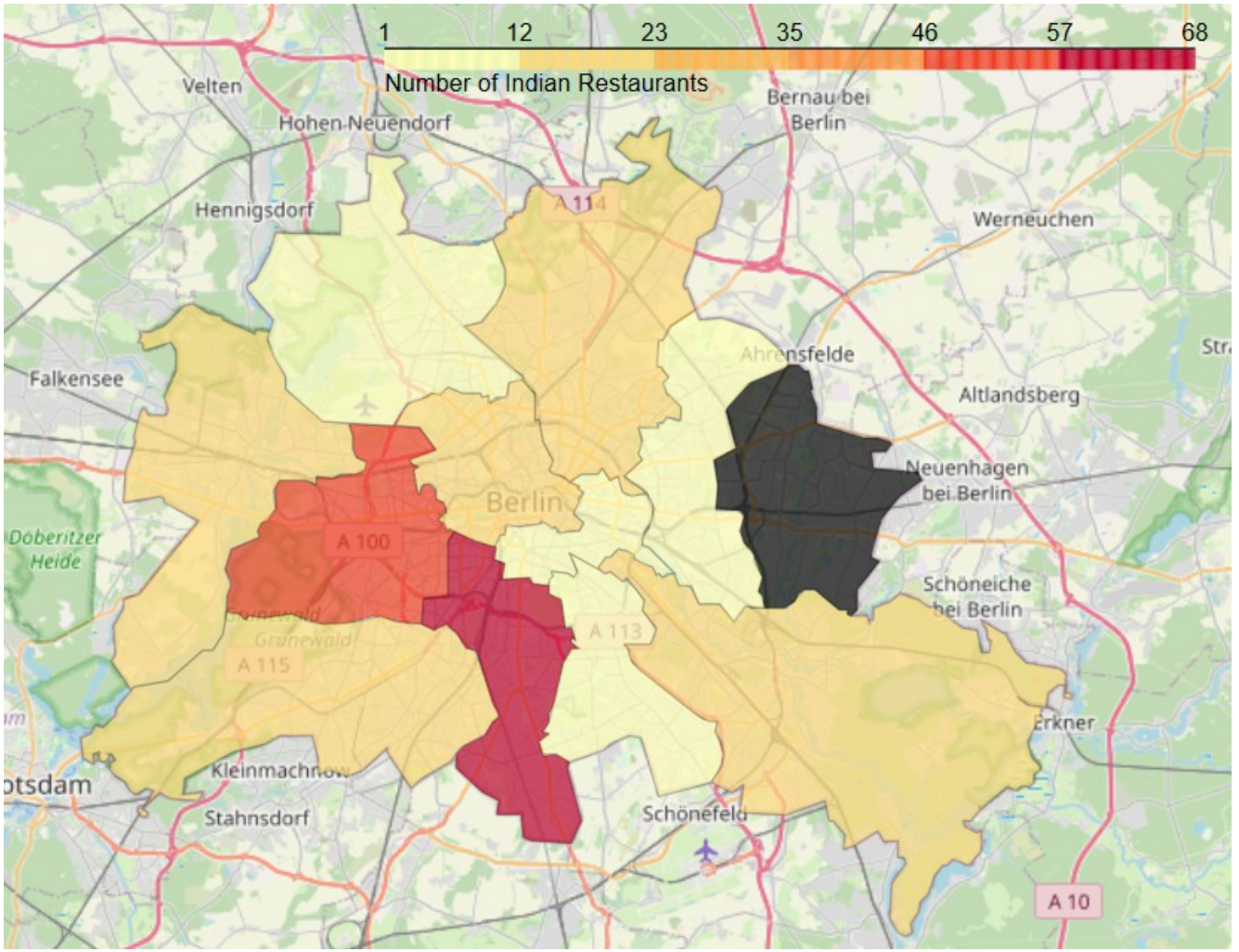
Berlin map with all borough



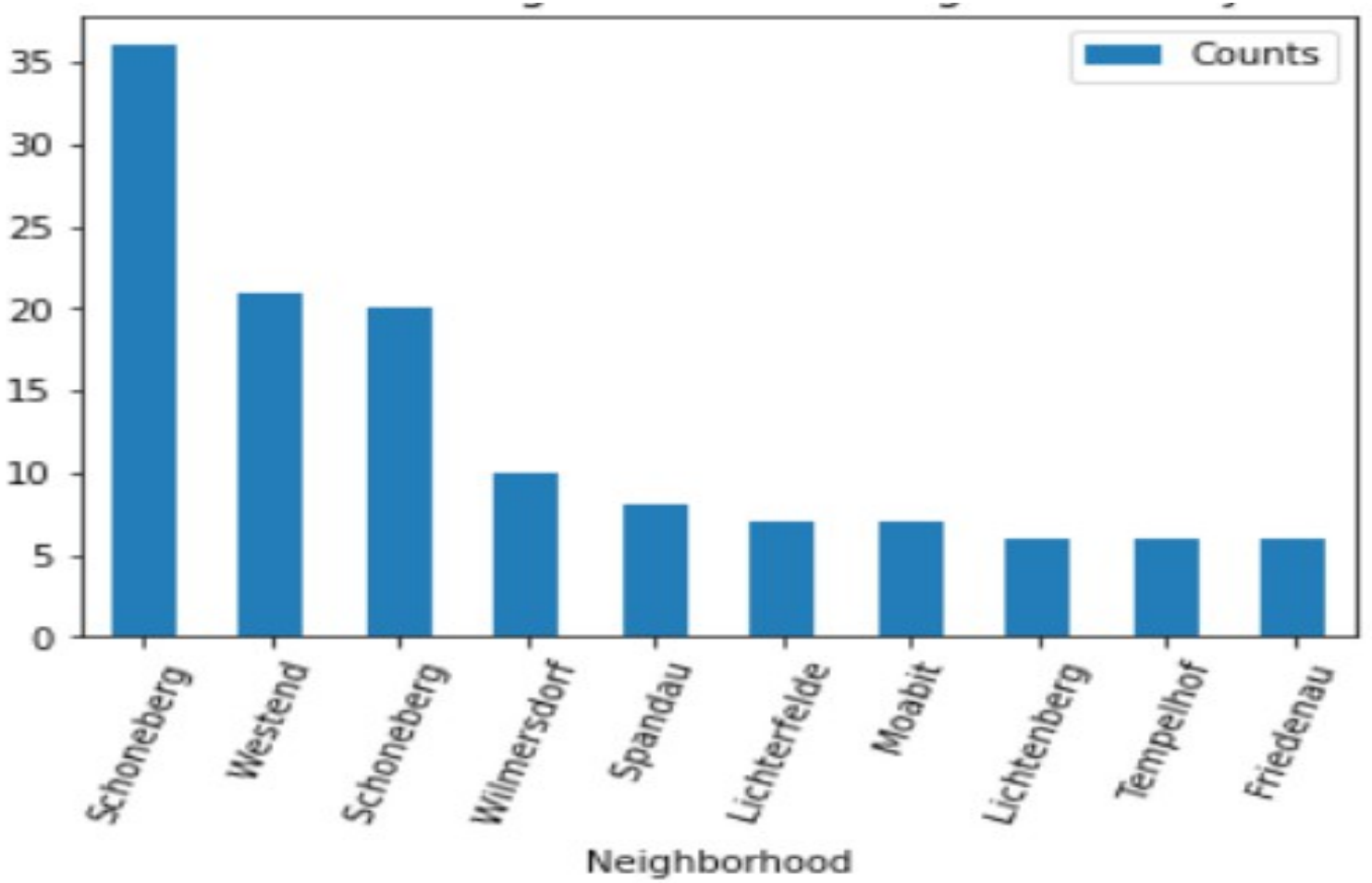
Berlin map with all borough & centers of Neighborhoods



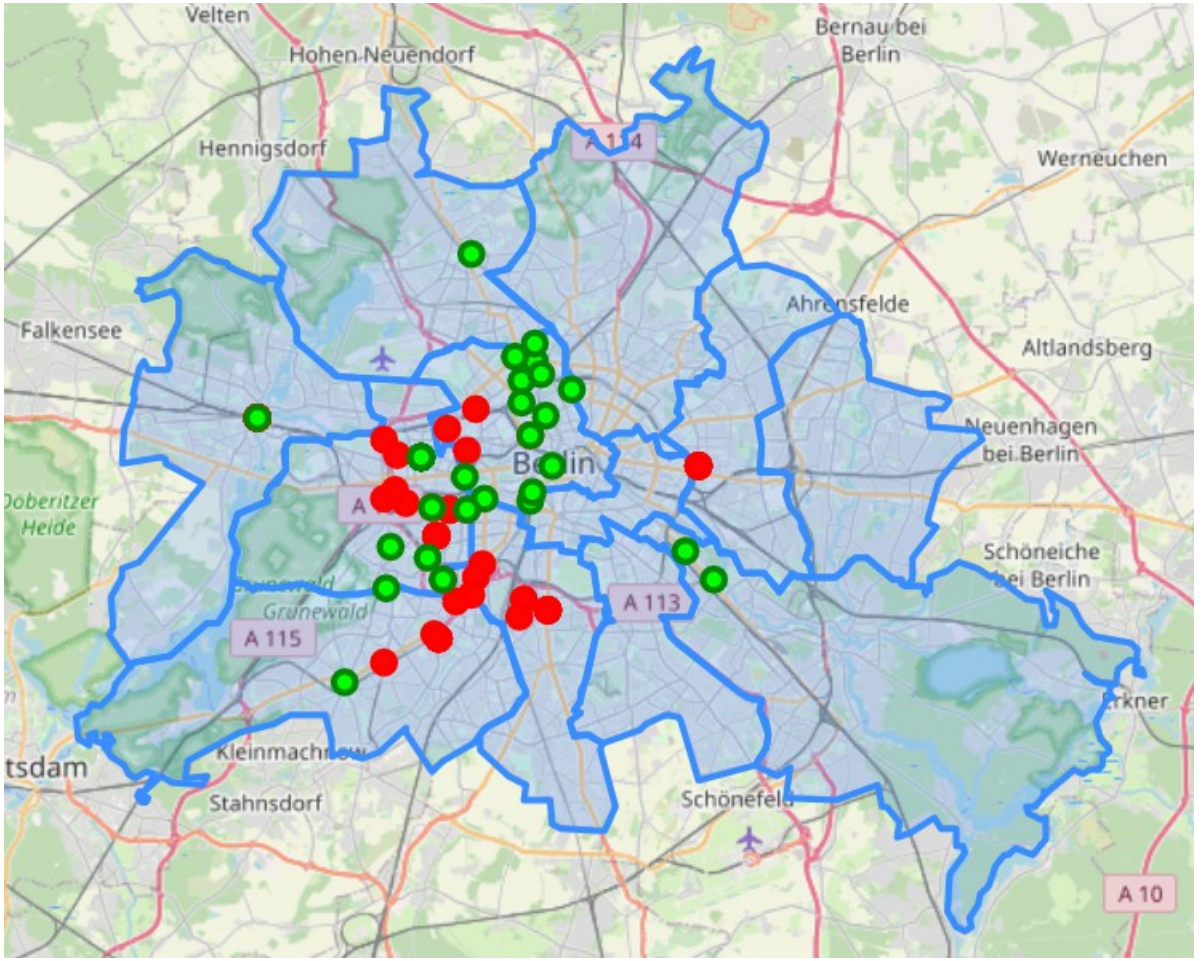
The distribution of the Indian restaurants in boroughs of Berlin



10 most crowded neighborhoods with Indian restaurants



Map of Berlin with analysis results



- Red dots are the center of the Neighborhoods with the highest number of Indian restaurants
- The green dots shows the suggested neighborhoods for opening an Indian restaurant

Conclusion and future directions

- We built a useful model to cluster and segment the neighborhoods in Berlin
- The model provide realistic recommendations for neighborhoods that might host a new Indian restaurant
- Accuracy of the models has room for improvement
- Future Idea:
  - Is to replace the addresses used in this Model with a uniformly distributed centers over the city of Berlin
  - Include more parameters in the model like the number of residence in each area and how busy the streets are