Battle of Capitals

From Toronto to Belrin

Problem statement

- Relocation outside your home city can be stressful
- It is not straightforward to figure out which area is the best match, if you are not familiar with a new city

Key objective

Utilizing Foursquare location data, leveraged by clustering of venues to determine what might be the best neighborhood in Berlin to relocate from Crescent Town in Toronto

Stakeholders

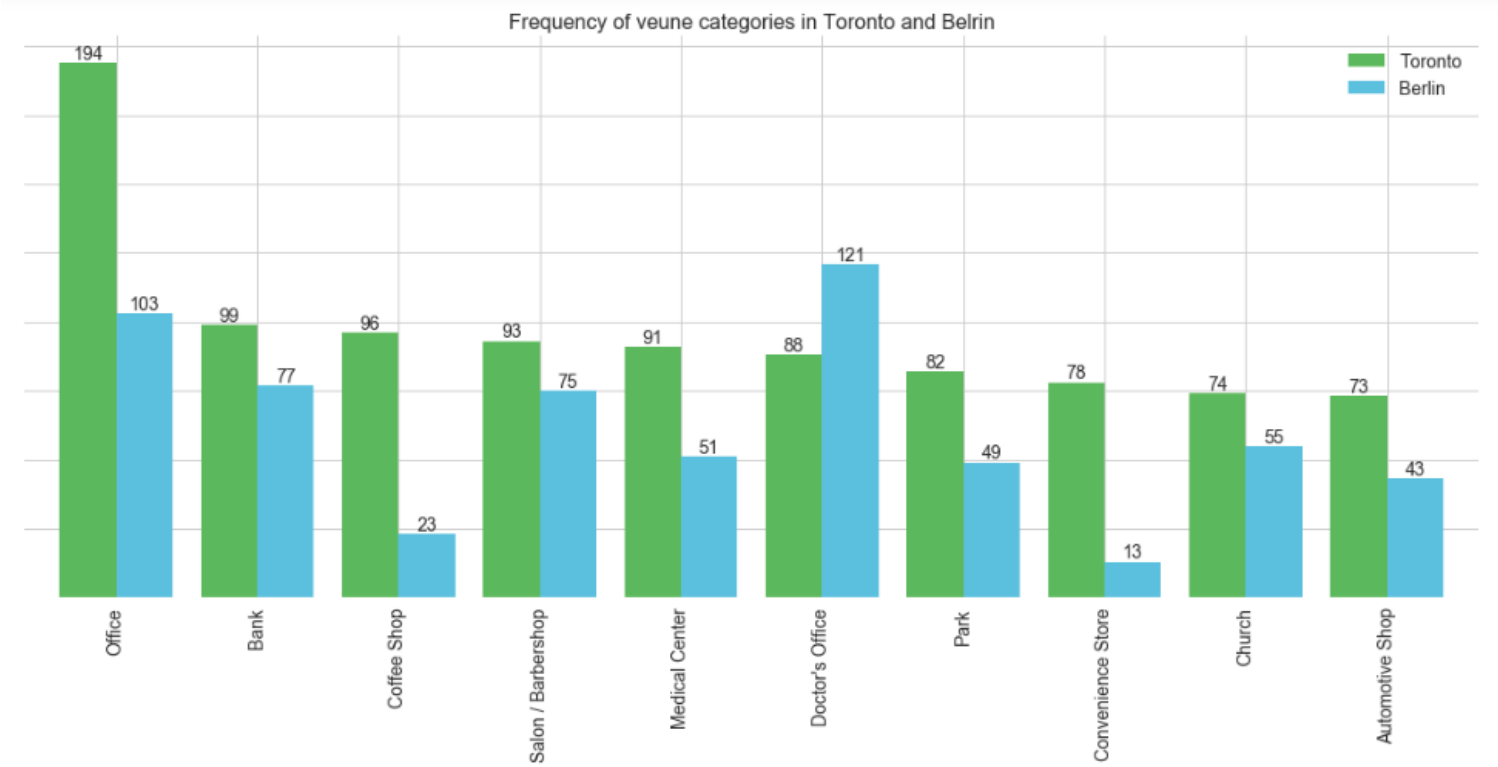
- Relocation agencies and real-estate agents, in order to provide better service and advice for their clients
- Business owners, willing to expand beyond their domiciles and need to obtain insights into new hoods upfront
- State authorities can explore how life quality and diversity vary in the cities within the country.

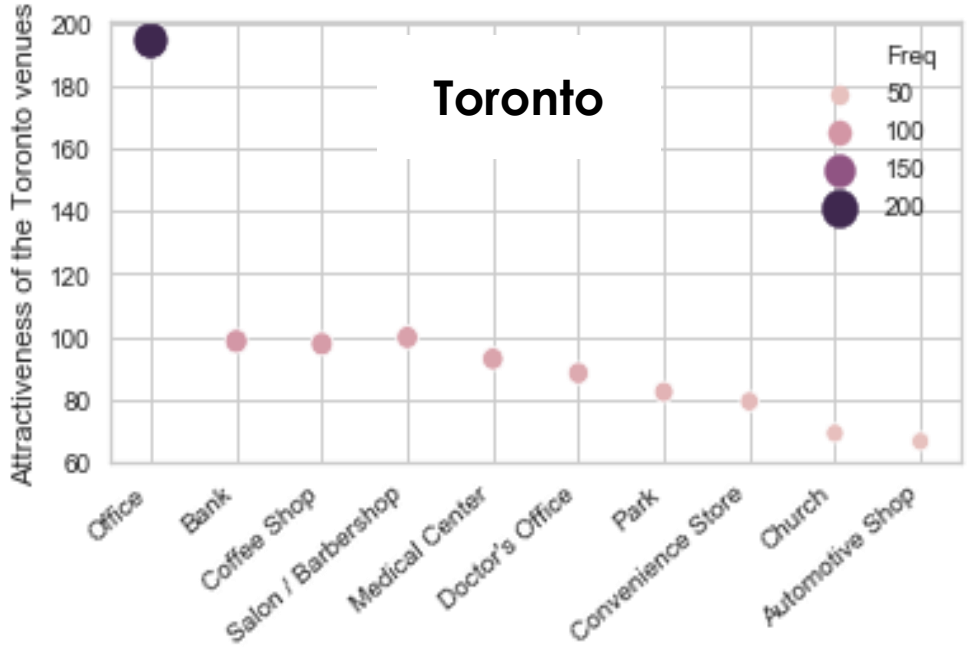
Data and Methodology

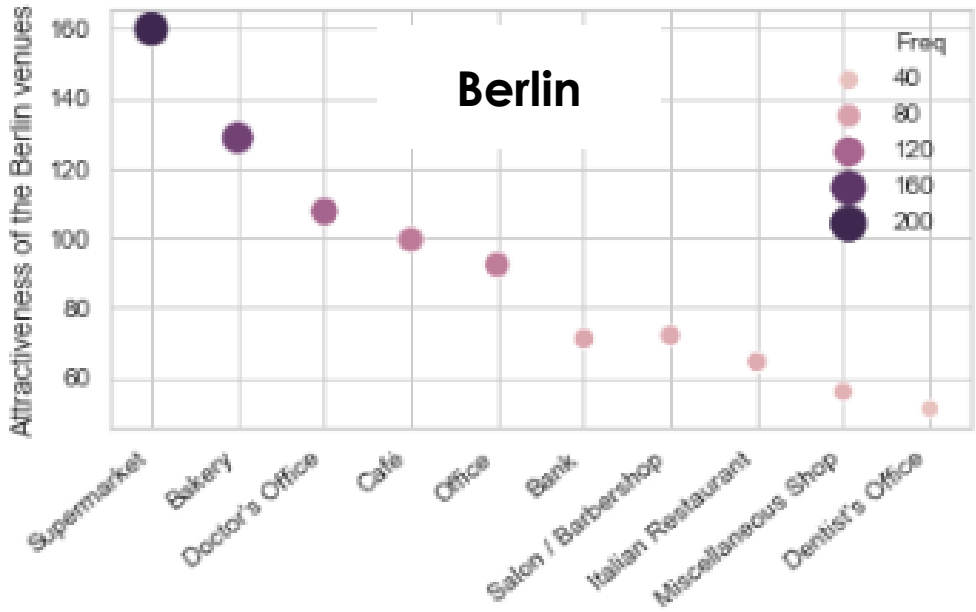
- Wiki pages for Toronto and Berlin hoods and boroughs
- After data cleaning the resulting data set has 8,711 venues and 310 unique venue categories
- Foursquare RESTful API calls to various endpoints to add venues attributes
- New features were added to capture user preferences regarding the venue and to account for the price
- Attractiveness of the area was quantified

Most common venues

- Berlin has 2 times less offices that Toronto, but more doctor offices
- Toronto has more coffee shops and parks



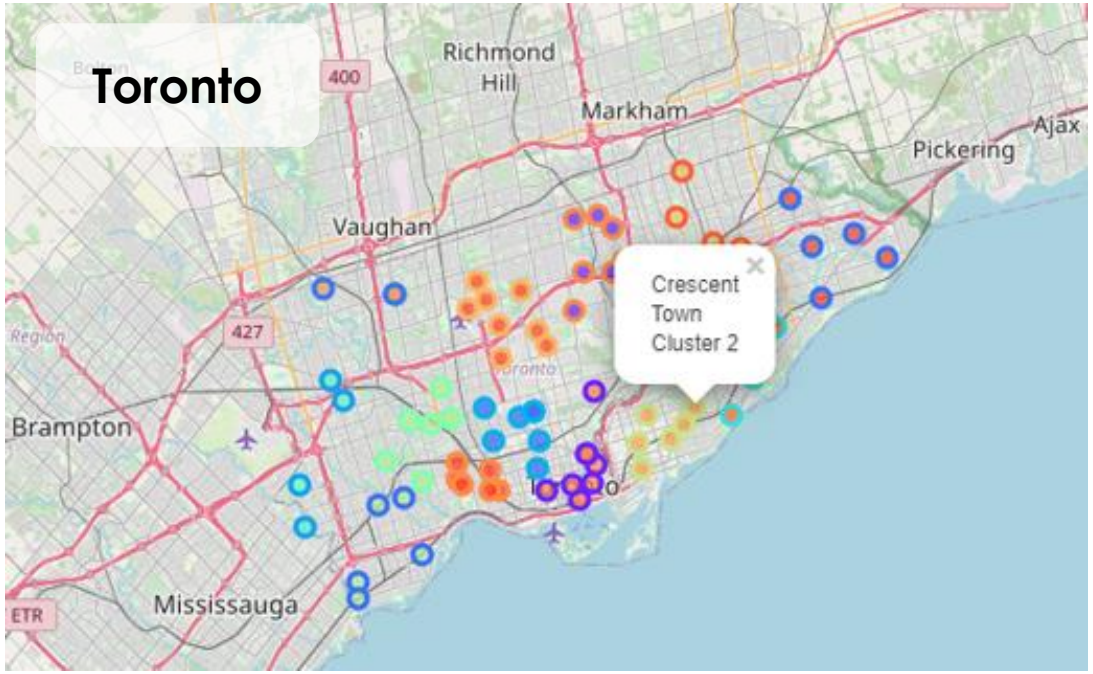


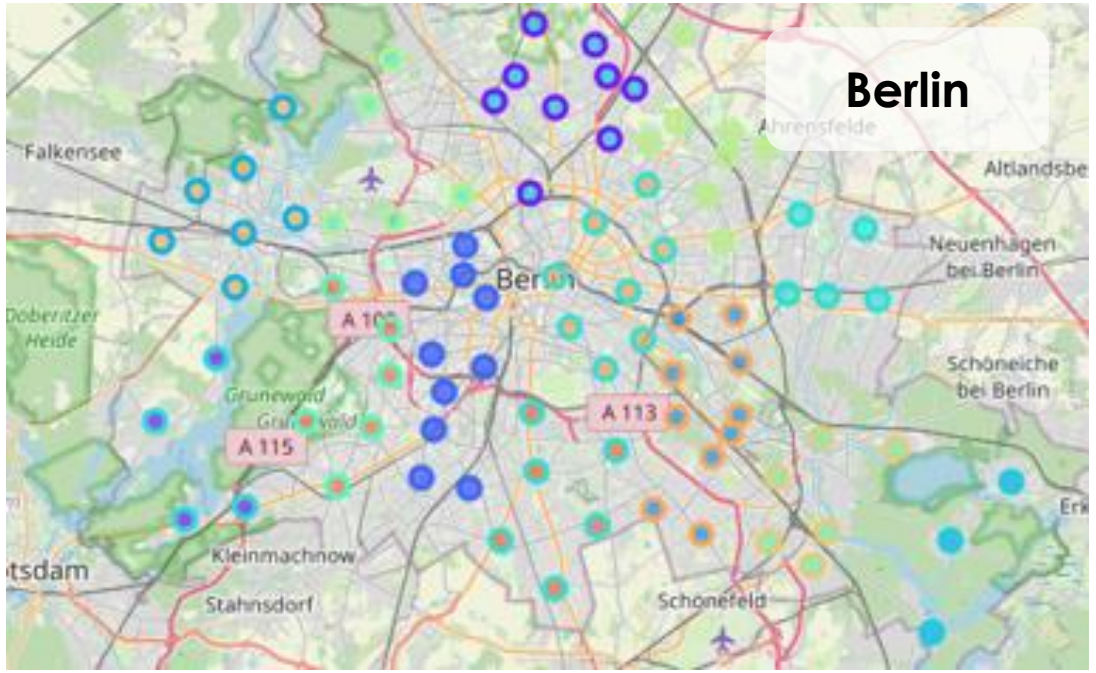


K-means clustering

- Silhouette method instead of Elbow Method to measure how well data points were classified
- **13 clusters** for Toronto and **14 for Berlin** were found though 40 runs of a loop over k from 10 to 39

Clustering results for the best k

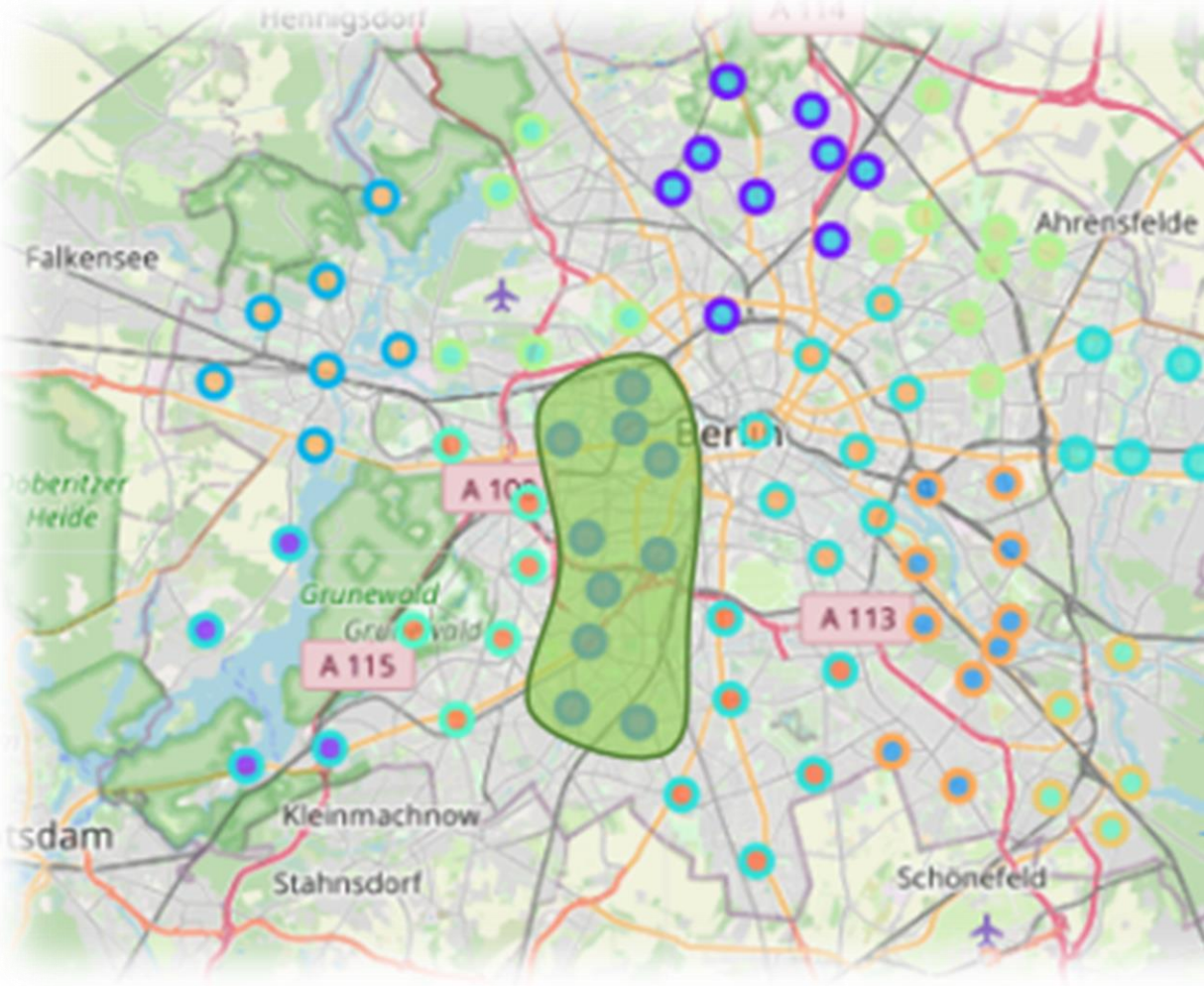




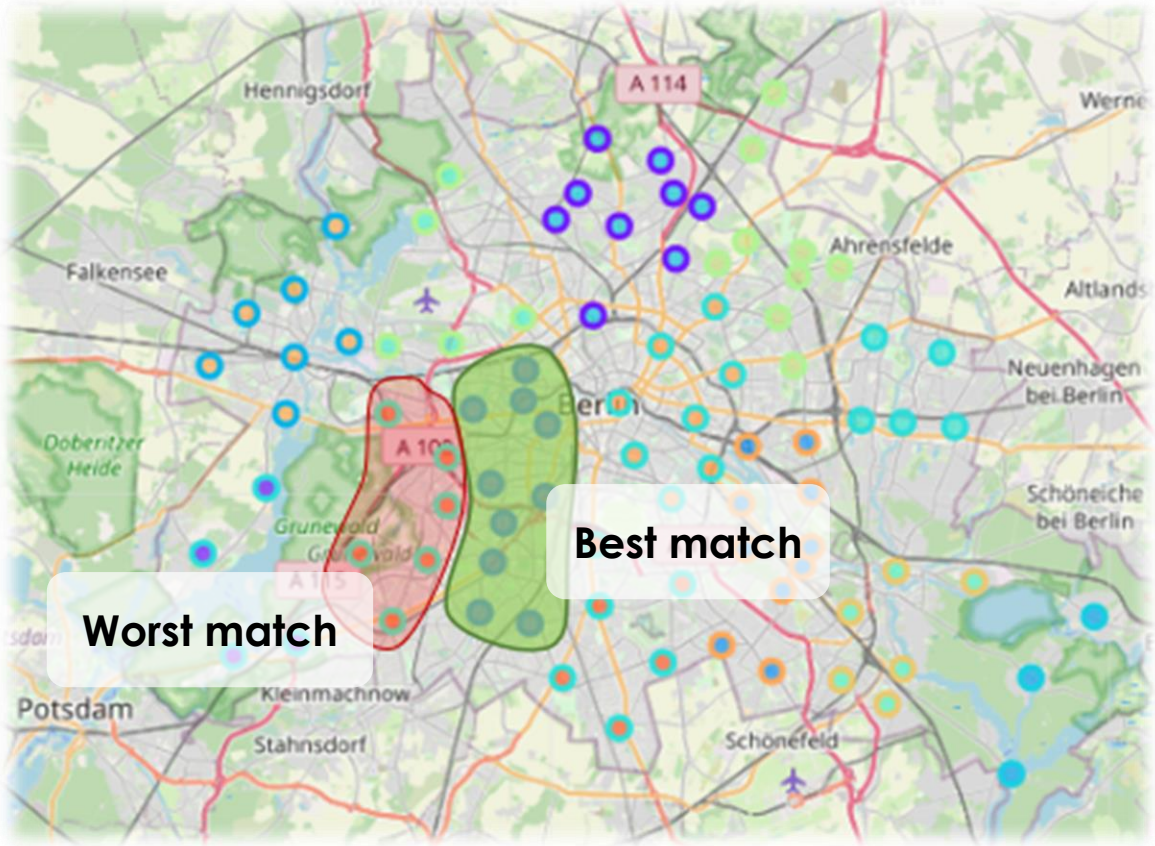
Best match

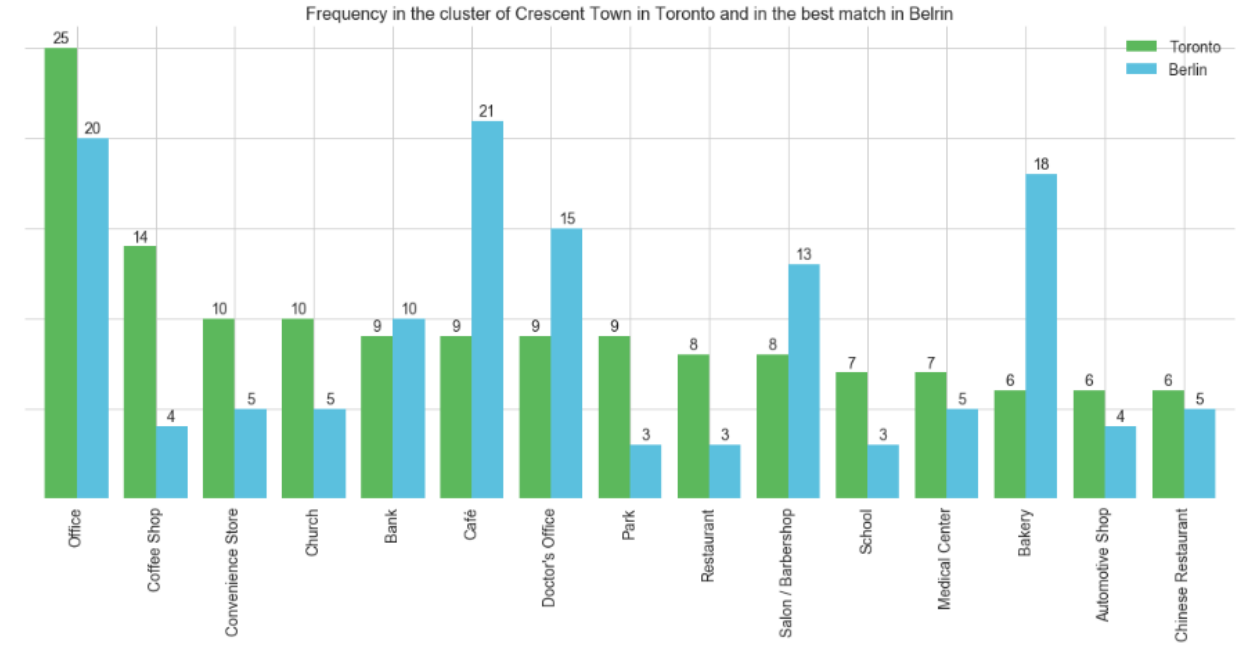
- Calculate Euclidean distance between Crescent Town cluster center and each cluster center in Berlin
- Cluster with a minimum distance would be the best match

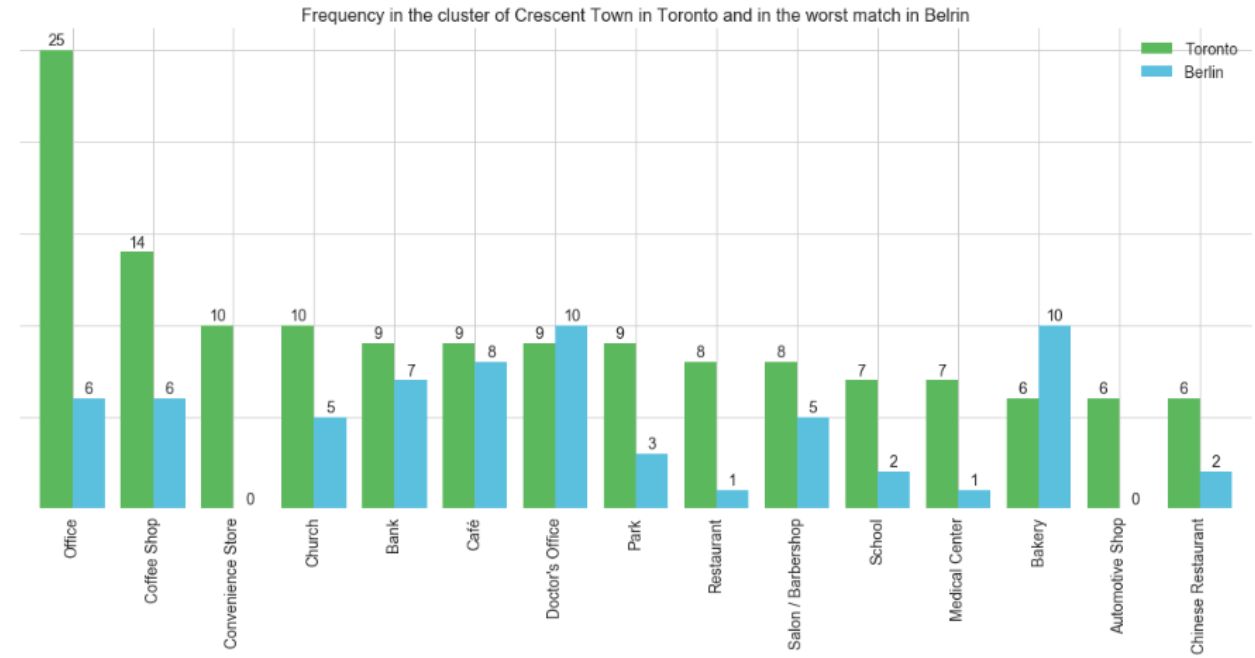
Conclusion:
Pack your things to relocate to one these nice Berlin areas: *Hansaviertel, Wilmersdorf, Friedenau, Schöneberg, Steglitz, Charlottenburg, Moabit, Lichterfelde, Tiergarten, Lankwitz*



Best vs worst match







Conclusion and further directions

- Clustering allows to group unlabeled high dimensional datasets, helping to understand and structure data
- Toronto was split in13 disjoint groups based on area attractiveness and habits
- Crescent Town area was compared against Berlin cluster and the best match was found
- Crescent Town and Berlin hood similarity was analyzed

- The results of the study can be improved if using broader range of location data providers
- Propagation of the same idea on other cities
- Suggestions to business owners, willing to grow abroad
- Support relocation service providers, real-estate companies and real-estate developers, or any other party willing to invest it the new area with valuable insights
- Better understanding of life quality and similarity within the country.

References

Functions that pullout data from Foursquare were taken from Applied Data Science Capstone' course labs, although slightly adopted for my own needs

Special thanks goes to Alex Aklson and Polong Lin, who created those notebooks