

Comparison of German Cities

Final project during Coursera course
„Applied Data Science Capstone“

The initial question

- ▶ Many cities in East and North German are suffering from a loss of population while the cities in West and South are gaining residents.
- ▶ Most likely the reasons for migration from East to West are to be found in the labor market
- ▶ *But: is there a difference in these cities that adds to this?*

Where do you go tonight?

- ▶ A city that offers many venues such as cafès, bars, cinemas, theaters etc. is more likely to hold its residents than a city that has a lower recreational value.

⇔ Examining the largest german cities if there are similarities or differences in public life.

It all comes with the (public) data

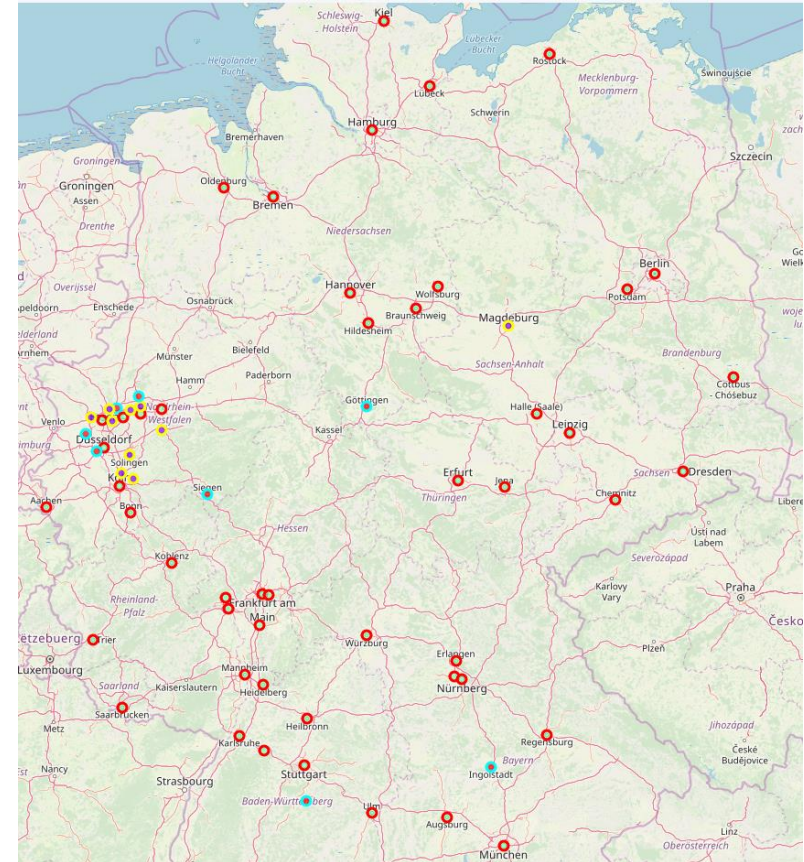
- ▶ A list of the largest german cities was generated based on a [Wikipedia article](#)
- ▶ Geocoordinates were added from [OpenStreetMap](#) using GeoPy
- ▶ Venue data was retrieved for each city from [Foursquare](#), using their public API

Methodology

- ▶ Generating a list of german cities with min 100.000 residents
- ▶ Retrieving max 100 venues per city from Foursquare API
- ▶ Grouping by venue type
- ▶ Removing cities with less than 15 venue types
- ▶ Creating „fingerprint“ by using just the 15 most occuring venue types for each city
- ▶ Machine learning: Clustering these cities using k-means algorithm

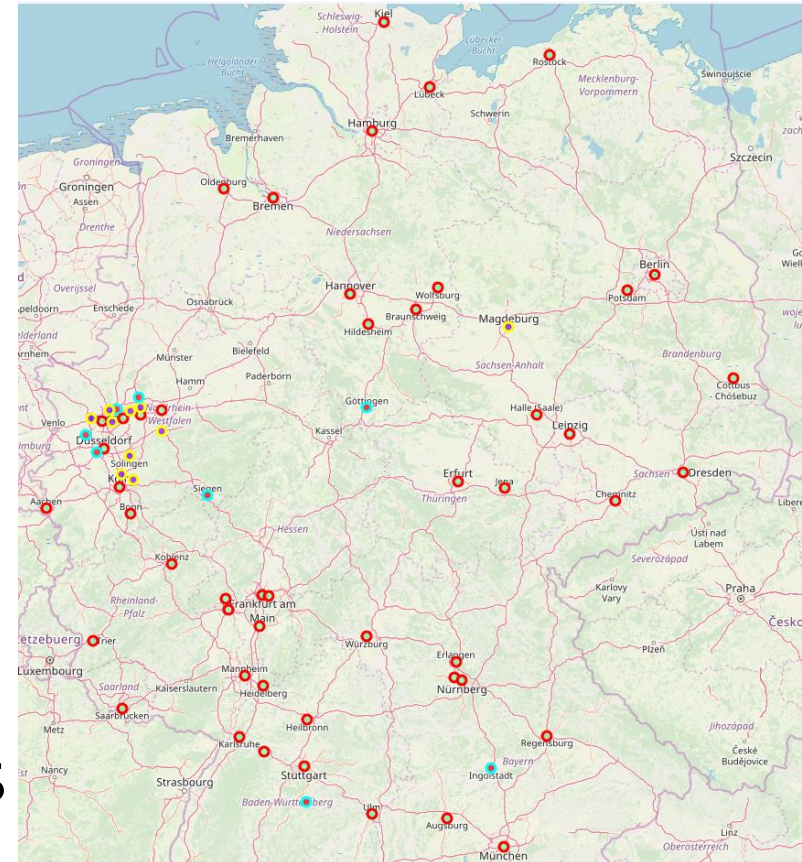
Results(1): all cities are equal

- ▶ Most cities (70%) are part of one cluster which is spread evenly across Germany (red markers)
 - There's no differences in east and west german cities
 - Indeed most german cities are similar



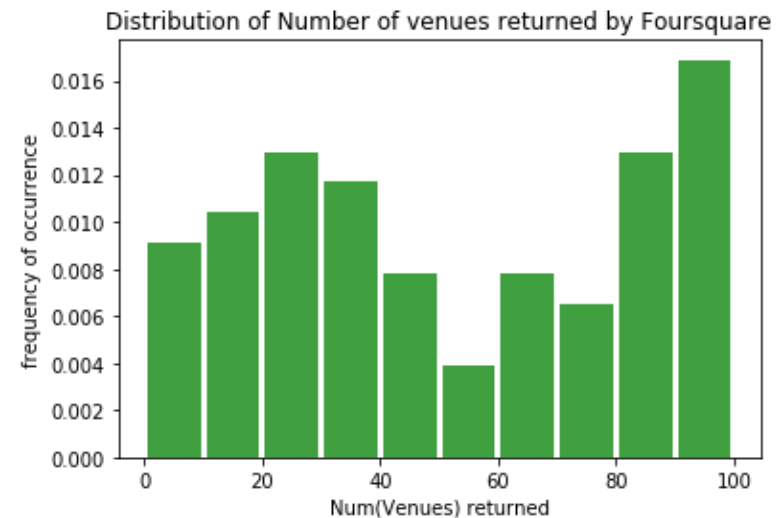
Results(2): but some are special

- ▶ There's one cluster with ten cities where nine are located in the Ruhr area (yellow markers)
 - This could be due to historic facts, this region being the industrial heart of Germany for decades
 - This should be investigated in a follow-up project
- ▶ The third cluster contains mainly „outliers“ (blue markers)



Results(3): Foursquare is American

- ▶ The Foursquare database contains only few data for most german cities
 - For many cities less than 50 venues could be retrieved
 - The Foursquare user base is mainly American
 - Due to small german user base the Foursquare data could lead to a certain bias



Author

Dr. Bernd Künnen

ARS Computer und Consulting GmbH
München

Email bernd.kuennen@ars.de