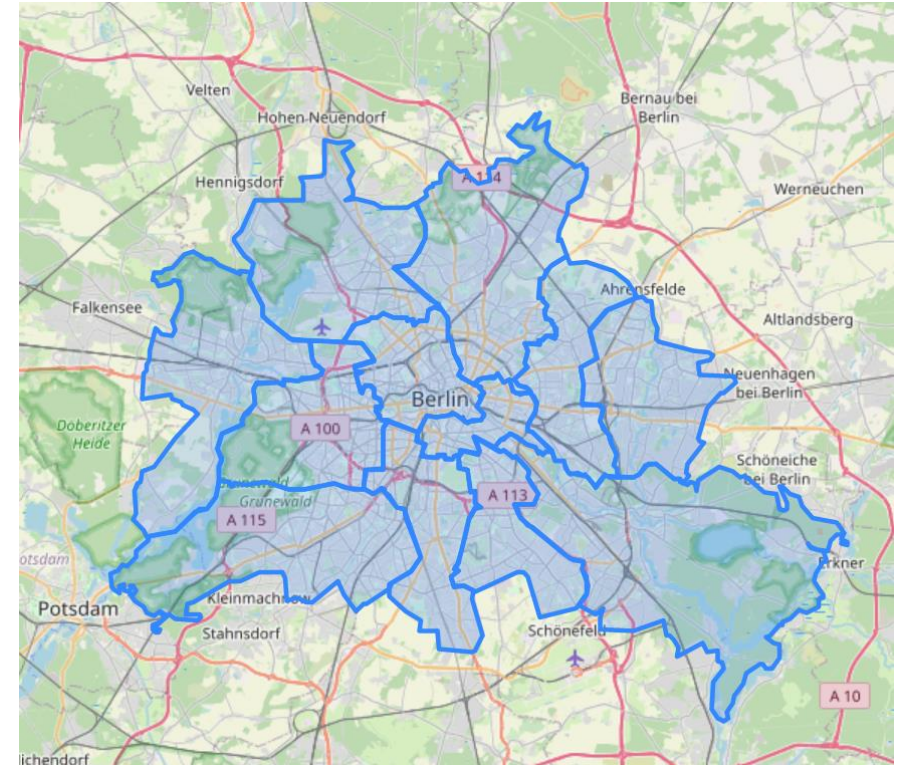# The Best District in Berlin

Assessing the best district in berlin

- The best district is evaluated based on the number of venues with respect to crimes in each district in Berlin

- The ratio of venues per crime is used as a definition of a so-called suitability index

Data acquisition and cleaning

- Berlin districts scraped from crime-reference:

  https://raw.githubusercontent.com/funkeinteraktiv/Berlin-Geodaten/master/berlin_bezirke.geojson

- Crime data for Berlin districts were acquired from Kaggle:

  https://www.kaggle.com/danilzyryanov/crime-in-berlin-2012-2019/version/4

- Venue data from Foursquare

- Datasets were merged and crime data for 2019 was selected for final analysis.

- Suitability indices were computed for each district

# Using suitability index as a measure for best district

- Highes suitability index for Reinickendorf district

- Reinickendorf is therefore given as a recommendation for a best district
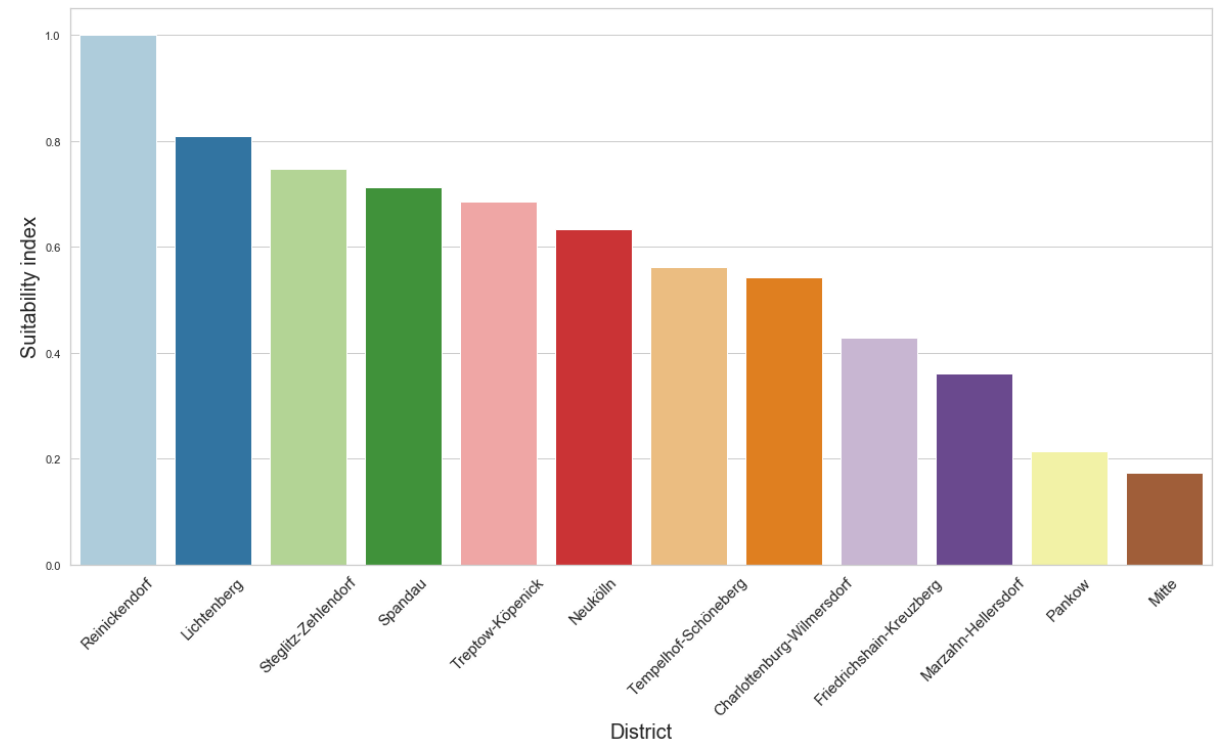
Data clustering using K-means

- Clustering of crime and venue counts groups similar districts

- 6 districts with similar characteristics, defined as districts with modest to high venue count and low crime rate

	Cluster Labels	Year	District	crime_count	venue_count	suitability_index
63	1	2019	Reinickendorf	27846	32	1.000000
23	1	2019	Lichtenberg	32239	30	0.809753
79	1	2019	Steglitz-Zehlendorf	29080	25	0.748098
71	1	2019	Spandau	26872	22	0.712419
95	1	2019	Treptow-Köpenick	30422	24	0.686493
31	1	2019	Marzahn-Hellersdorf	26447	11	0.361934
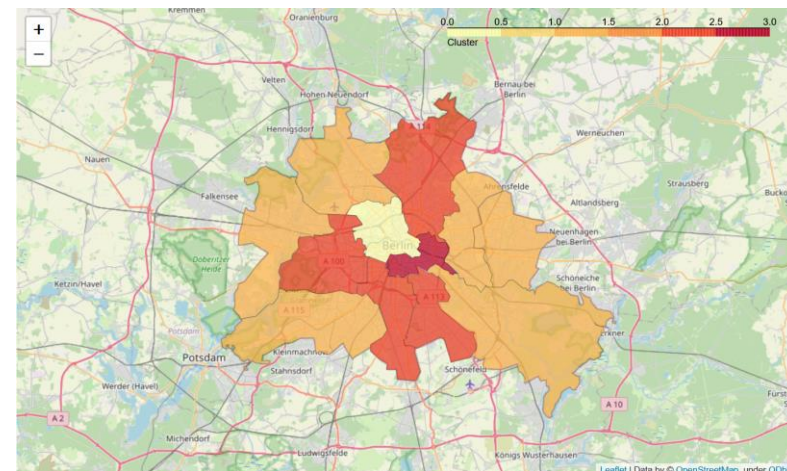
Conclusions

- **Renickendorf-district** has the highest suitability index and is recommended for the client

- Following districts fall in the same group and could be considered as optional "best district" candidates:

  - Lichtenberg

  - Steglitz-Zehlendorf

  - Spandau

  - Treptow-Köpenick