

Neighbourhood analysis of Frankfurt

Description of the problem and a discussion of the background

Frankfurt am Main is the fifth largest city in Germany, located in the centre of the country. With its vibrant city-life, a wide range of employers and renowned universities, the city is highly attractive for new citizens. Throughout the past decade, the city witnessed a steady increase in inhabitants leading to a higher demand of apartments and houses in all neighbourhoods.

(https://de.wikipedia.org/wiki/Frankfurt_am_Main)

However, the question for people who want to move to Frankfurt is which neighbourhood would be most suitable to live in.

Mostly, the decision for a place to live depends on many factors such as budget, distance to place of work or university etc. Often, a key decision criterion remains the number of coffee places and bars in the direct environment. This can be predominantly witnessed at students or young professionals.

The research will therefore focus on analysing the neighbourhoods in Frankfurt by answering the following research questions:

- **What are the neighbourhoods that show the highest density of venues?**
- **Using k-means clustering, in which neighbourhoods are predominantly bars, coffee places and restaurants located?**

Description of the data and how it will be used to solve the problem

A table with all neighbourhoods of Frankfurt is published on Wikipedia. Besides the name of the neighbourhood, additional relevant information such as citizens per square kilometre and the year since the neighbourhood belongs to the city Frankfurt is sourced. The table is scraped directly into a pandas data frame for further processing. (e.g. drop of columns, deleting irrelevant information in columns etc.). The cleaned table builds the foundation for the neighbourhood analysis and some descriptive statistics of the city Frankfurt.

(https://de.wikipedia.org/wiki/Liste_der_Stadtteile_von_Frankfurt_am_Main)

Geographic coordinates for each neighbourhood are sourced using the Nominatim Geocoding service. Latitudinal and longitudinal data is appended to the data frame in separate columns.

Geographical coordinates are plotted using the folium map.

Locational data for venues is sourced from Foursquare via the Foursquare API. Since the basic account is used, only 100 outputs can be received per request and geographical coordinate. The radius is set to 600 metre to cover the entire range of neighbourhoods. The locational data is used for clustering all neighbourhoods into distinctive clusters. The aim is to identify clusters that show a high rate of coffee places and bars – places likely visited by students and young professionals.

(<https://developer.foursquare.com/>)