

# IBM Capstone Project Report

## Introduction

To determine the cities with the best quality of life, cities are often compared based on general metrics, such as population density, public transport system, green spaces, etc. For an individual, other factors might be more important when it comes to choosing a place to live. Here, I compared the 20 largest cities in Germany with respect to my specific interests, using the Foursquare venue location data. Specifically, I chose some types of restaurants I enjoy and venues related to my hobbies, which are dancing and theatre. In the notebook, those venue categories can be specified by the user, such that for a different individual, different important factors can be chosen. In addition, I took into account that the closer those venues of interest are, the more convenient it will be to travel within the city. Based on those features, I developed a score to rank the cities with a potentially good quality of life for the individual interests of a person.

## Data

The 20 largest cities of Germany, their population size and location, were gathered from this Wikipedia page [https://en.wikipedia.org/wiki/List\\_of\\_cities\\_in\\_Germany\\_by\\_population](https://en.wikipedia.org/wiki/List_of_cities_in_Germany_by_population) with the BeautifulSoup library.

The venue categories for the Foursquare API were chosen from this webpage:

<https://developer.foursquare.com/docs/build-with-foursquare/categories>

Finally, the venues within these categories in the specified cities were explored using the Foursquare API.

## Methodology

First, I used the beautifulsoup package to extract the location of the 20 largest German cities and their population. Using pandas and folium, I cleaned the data and visualized the cities (Figure 1). For each city, I defined a circular radius based on the city's area. The resulting dataframe is found Table 1 in the Appendix.

Next, I chose specific venue categories. The whole list of venue categories in the Foursquare API can be found on <https://developer.foursquare.com/docs/build-with-foursquare/categories>. In my case, those venue categories were: Jazz Club, Theatre, Dance studio, Costume Shop, Malay Restaurant, Bubble Tea Shop, Molecular Gastronomy Restaurant, Night club. I then explored all venues within the radius with the chosen categories. All resulting venues in all categories and all cities were stored in a pandas dataframe.

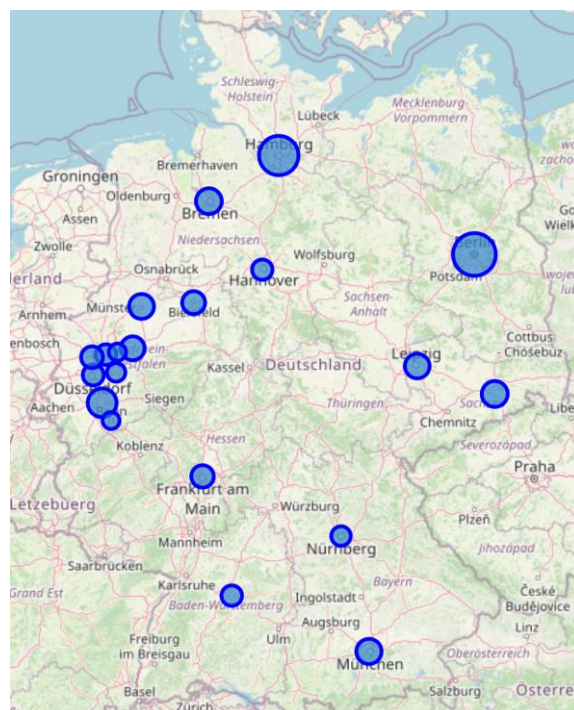


Figure 1 Map of Germany with 20 largest cities. Circle area corresponds to the population size.

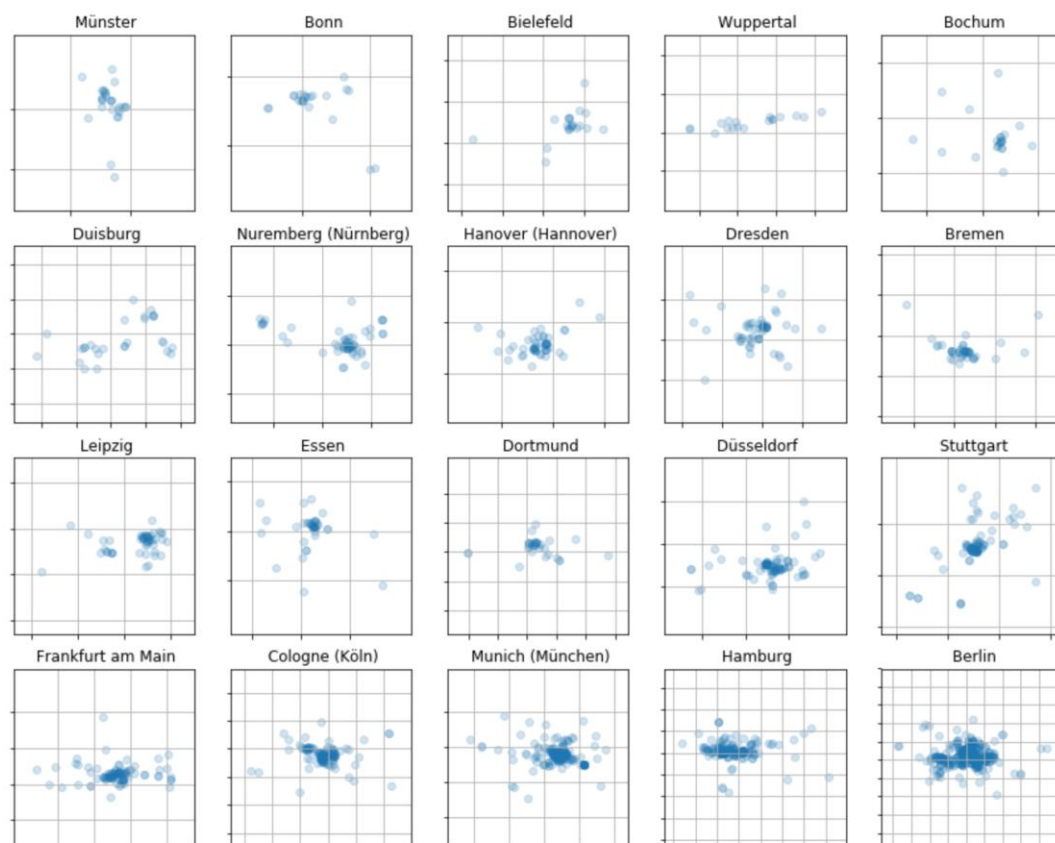
I then defined several features that were important to rank the cities: The overall number of venues of interest (VOI), The diversity of venues of interest (diversity factor, DF), The median distance of these venues from the city center (dist), the population of the city (pop). The diversity factor was determined as the total number of venue categories (nine in this case) minus the number of categories with zero venues in one city. I normalized those features and calculated a score that ranked the cities based on those factors.

$$\text{Score} = \text{VOI} + \text{DF} - \text{dist} - 0.2 * \text{pop}$$

Based on this score, I ranked the cities from the most to the least livable.

## Results

To gain an intuitive understanding of the cities layouts and any grouping of venues in specific areas of a city, I first visualized all venues of interest in all cities on a simple coordinate grid. The resulting plots are shown in Figure 2.



*Figure 2 Venue location overview. For each city, all venues of interest (all venues of the categories specified) are plotted as a marker on a coordinate grid of the city. One coordinate line corresponds to 0.05 degrees of latitude or longitude, respectively.*

I then extracted specific features of the data that I considered important for ranking the cities, as described in the methodology section. The resulting features, total number of venues, venue diversity and venue distance, I visualized in simple plots to gain an intuitive overview of the features (Appendix Table 2, Figure 3). In addition, I plotted the total number of venues against the venue distance (Figure 4). In this plot, potential cities with a high final score would localize in the bottom left corner (large number of venues with a small distance between them).

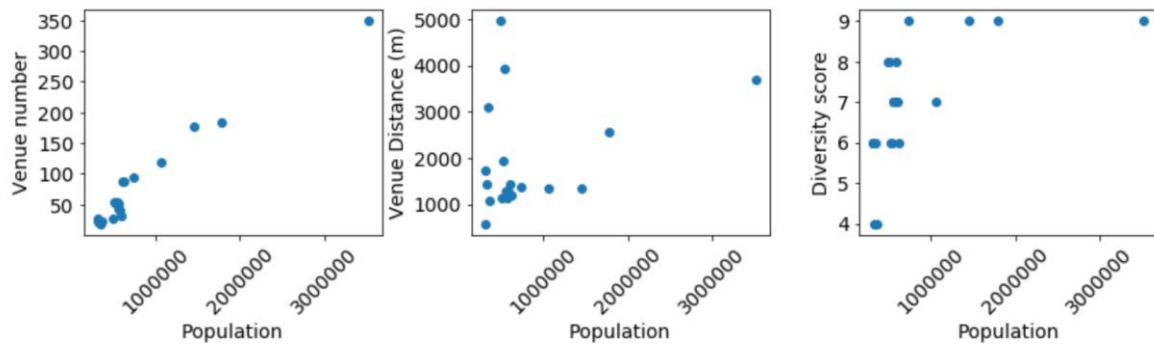


Figure 3 Venue features (venue number, venue distance and diversity score) plotted for each city, against the city populations.

I next calculated the final score based on those features, as described in the methodology section. The ranking of the cities based on the final score is shown in Figure 5 (Data in Appendix Table 2). The city of Munich overall reached the highest score in the city ranking.

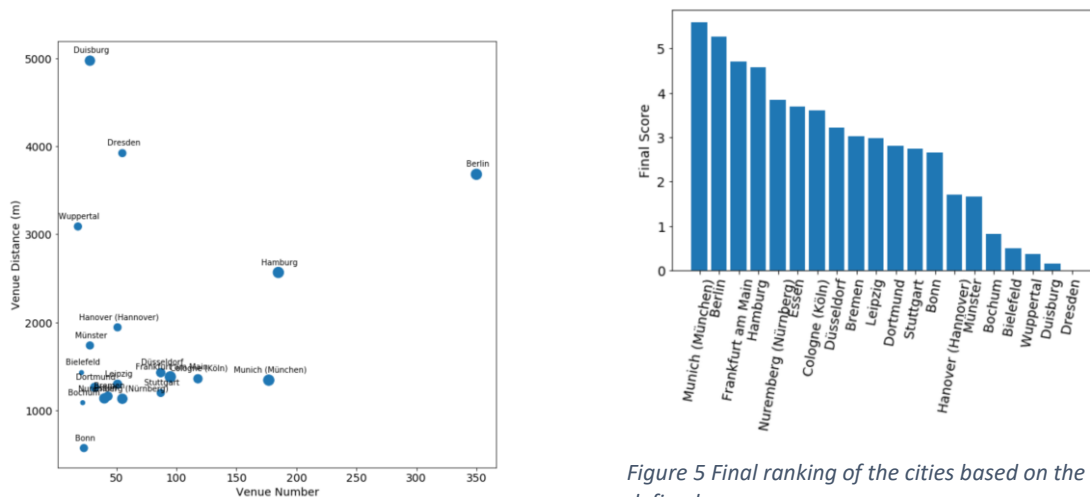


Figure 5 Final ranking of the cities based on the defined score.

Figure 4 Venue Number plotted against Venue Distance. The size of the circles corresponds to the Diversity Score. Cities with a large venue number, large diversity score and small venue distance will reach a large final score.

For the city of Munich, I then visualized all venues on a city map (Figure 6). The venues are colored by category, such that it is easy to search all venues of interest of a specific category.

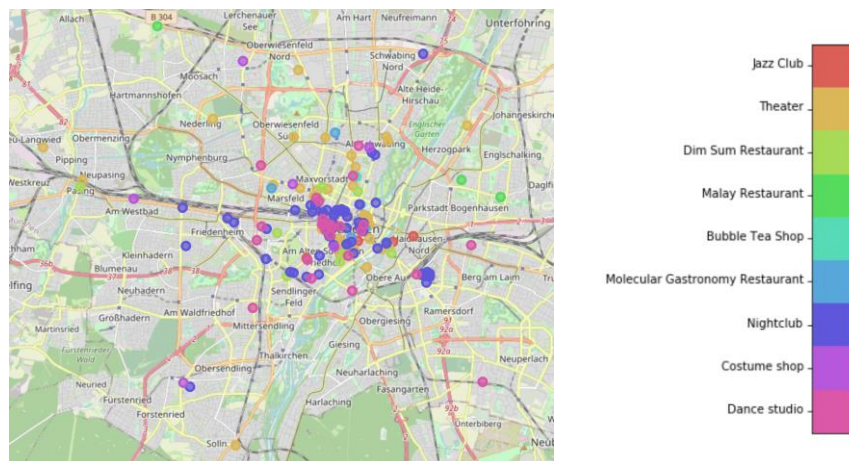


Figure 6 Map of Munich with venues of interest labelled in different colors.

## Discussion

Overall, cities with a higher population tend to have more venues that are interesting. However, consequently, those cities are also larger and therefore the distance between venues will be higher. As a consequence, the city that scored the highest final rank was not the city with the largest population, but still a fairly large city with a reasonable density of venues in the center. Contrary to my expectations, in the visualization of venues within a city, different neighborhoods (clusters of venues), were not readily visible, even though I did not quantify this observation formally.

## Conclusion

In this project, I analyzed the twenty largest cities of Germany, and quantified their livability with respect to my individual interests. This is useful since general indices of 'quality of life' only take into account general measures of livability. Depending on the individual interests of a user, venue categories were explored with the Foursquare API, and the prevalence of interesting venues together with their distribution across the city were used as features to calculate an individual livability score.

## Appendix

The appendix lists the summarized data of this project.

|                             | 2015 pop  | area_km2 | Latitude | Longitude | Radius       |
|-----------------------------|-----------|----------|----------|-----------|--------------|
| City                        |           |          |          |           |              |
| <b>Berlin</b>               | 3520031.0 | 891.68   | 52.517   | 13.383    | 16847.271569 |
| <b>Hamburg</b>              | 1787408.0 | 755.30   | 53.550   | 10.000    | 15505.465392 |
| <b>Munich (München)</b>     | 1450381.0 | 310.70   | 48.133   | 11.567    | 9944.791684  |
| <b>Cologne (Köln)</b>       | 1060582.0 | 405.02   | 50.933   | 6.950     | 11354.376694 |
| <b>Frankfurt am Main</b>    | 732688.0  | 248.31   | 50.117   | 8.683     | 8890.417754  |
| <b>Stuttgart</b>            | 623738.0  | 207.35   | 48.783   | 9.183     | 8124.134102  |
| <b>Düsseldorf</b>           | 612178.0  | 217.41   | 51.233   | 6.783     | 8318.879273  |
| <b>Dortmund</b>             | 586181.0  | 280.71   | 51.517   | 7.467     | 9452.659316  |
| <b>Essen</b>                | 582624.0  | 210.34   | 51.450   | 7.017     | 8182.499707  |
| <b>Leipzig</b>              | 560472.0  | 297.80   | 51.333   | 12.383    | 9736.153455  |
| <b>Bremen</b>               | 557464.0  | 326.18   | 53.083   | 8.800     | 10189.520041 |
| <b>Dresden</b>              | 543825.0  | 328.48   | 51.033   | 13.733    | 10225.381725 |
| <b>Hanover (Hannover)</b>   | 532163.0  | 204.14   | 52.367   | 9.717     | 8061.003670  |
| <b>Nuremberg (Nürnberg)</b> | 509975.0  | 186.38   | 49.450   | 11.083    | 7702.376035  |
| <b>Duisburg</b>             | 491231.0  | 232.80   | 51.433   | 6.767     | 8608.283308  |
| <b>Bochum</b>               | 364742.0  | 145.66   | 51.483   | 7.217     | 6809.186297  |
| <b>Wuppertal</b>            | 350046.0  | 168.39   | 51.267   | 7.183     | 7321.215864  |
| <b>Bielefeld</b>            | 333090.0  | 258.82   | 52.017   | 8.533     | 9076.616371  |
| <b>Bonn</b>                 | 318809.0  | 141.06   | 50.733   | 7.100     | 6700.805365  |
| <b>Münster</b>              | 310039.0  | 303.28   | 51.967   | 7.633     | 9825.325556  |

*Table 1 Largest cities of Germany with their population size, the city area, its location and its radius, assuming a circular size.*

|                             | 2015 pop  | Venue Number | Median Distance | Diversity score | Final Score |
|-----------------------------|-----------|--------------|-----------------|-----------------|-------------|
| <b>Munich (München)</b>     | 1450381.0 | 177          | 1342.0          | 9               | 5.586356    |
| <b>Berlin</b>               | 3520031.0 | 350          | 3679.5          | 9               | 5.275385    |
| <b>Frankfurt am Main</b>    | 732688.0  | 95           | 1380.0          | 9               | 4.714680    |
| <b>Hamburg</b>              | 1787408.0 | 185          | 2565.0          | 9               | 4.580506    |
| <b>Nuremberg (Nürnberg)</b> | 509975.0  | 55           | 1131.0          | 8               | 3.844996    |
| <b>Essen</b>                | 582624.0  | 40           | 1135.0          | 8               | 3.688168    |
| <b>Cologne (Köln)</b>       | 1060582.0 | 118          | 1358.0          | 7               | 3.604478    |
| <b>Düsseldorf</b>           | 612178.0  | 87           | 1427.0          | 7               | 3.226356    |
| <b>Bremen</b>               | 557464.0  | 43           | 1160.0          | 7               | 3.014199    |
| <b>Leipzig</b>              | 560472.0  | 51           | 1296.0          | 7               | 2.975001    |
| <b>Dortmund</b>             | 586181.0  | 32           | 1261.0          | 7               | 2.811982    |
| <b>Stuttgart</b>            | 623738.0  | 87           | 1197.0          | 6               | 2.748500    |
| <b>Bonn</b>                 | 318809.0  | 23           | 571.0           | 6               | 2.651234    |
| <b>Hanover (Hannover)</b>   | 532163.0  | 51           | 1942.0          | 6               | 1.718132    |
| <b>Münster</b>              | 310039.0  | 28           | 1736.0          | 6               | 1.666308    |
| <b>Bochum</b>               | 364742.0  | 22           | 1086.0          | 4               | 0.818252    |
| <b>Bielefeld</b>            | 333090.0  | 21           | 1428.0          | 4               | 0.503899    |
| <b>Wuppertal</b>            | 350046.0  | 18           | 3087.5          | 6               | 0.362260    |
| <b>Duisburg</b>             | 491231.0  | 28           | 4970.0          | 8               | 0.155145    |
| <b>Dresden</b>              | 543825.0  | 55           | 3920.0          | 6               | 0.000000    |

*Table 2 Summary features of each city. The city population (2015 estimate), the number of venues of interest, the median distance of the venues to the center, the diversity score, and the final score based on the used metric.*