The Battle of Neighborhoods in Bonn

Analysing a City of choice: Bonn, Germany

Table of Content

Acknowledgement

This report is based on a Juypter notebook (shared on GitHub) which is based on a notebook created by Alex Aklson and Polong Lin. Copyright © 2018 Cognitive Class. The original notebook and its source code are released under the terms of the MIT License.

Objectives

"Write a full report consisting of all of the following components:

- **Introduction** where you discuss the business problem and who would be interested in this project.
- **Data** where you describe the data that will be used to solve the problem and the source of the data.
- **Methodology** section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.
- **Results** section where you discuss the results.
- **Discussion** section where you discuss any observations you noted and any recommendations you can make based on the results.
- **Conclusion** section where you conclude the report."

1. Introduction

The skills and tools learned in the previous sessions, specifically to use location data to explore a geographical location, will be applied by anlaysing the neighborhoods of the City of Bonn in Germany. For the "story telling" I will follow suggestion 2 of the project description and try to find answers for questions such as:

> *"If someone is looking to open a restaurant, where would you recommend that they open it? If a contractor is trying to start their own business, where would you recommend that they setup their office? If you would move for a job which neighborhood would you choose for housing?"*

As this report will be **peer-reviewed**, I assume that the reader knows the basics of Python Programming, API calls, the Folium Library, Choropleth maps, and k-clustering. As the reader might be unfamiliar with my City of Choice, i.e. the City of Bonn, I recommend that the interested reader consults the wikipedia page for additional information [1].

The population of Bonn is with around 300 000 people much smaller than the previously studied Cities of New York and Toronto. Nevertheless, the city of Bonn was from 1949 to 1990, the capital of West Germany. Still roughly a third of all ministerial jobs in Germany are located in Bonn and the headquarters of Deutsche Post DHL and Deutsche Telekom, both DAX-listed corporations, are in Bonn [1].

The Neighborhood Analysis exercised for New York and Toronto has clearly shown that location data retrieved from Foursquare severs provide valuable information about the local distribution of leisure venues. We have analysed the distribution of leisure venues in the context of **tourism** and of **moving from one neighborhood to another for a job offer**.

Additional to providing such information, I will demonstrate how a visualization of the spatial distributions of venues such as cafés and restaurants helps to **understand the field of competitors**. This information is not only valuable for the person who wants to start such a buisness, but as well for investors who have to decide if such a buisness might be sucessful.

2. A description of the data

As I was quite impressed by the provided examples using **choropleth maps** to visualize crime rates in San Francisco or Migration to Canada, I will use a **open data GeoJSON file** [2] for allowing choropleth map visualization in the context of analysing the City of Bonn.

Additionally, I will use publicly availiable data about the **population distribution** per **municipal district** from wikipedia [1].

The **geocoder package** allowing to retrieve **arcgis data** by API calls will be used to retrieve longitude and latitude values.

As a correction of geo data was necessary, additional geo data will be used by copy-pasting wikipedia data. A parser approach as for the Toronto neighborhood analysis did not prove to be efficient as not a single page, but several pages needed to be searched for information.

The core data for this analysis will be retrieved by API calls from **Foursquare** servers as for the Toronto and New York analysis.

3. Methodology section

The **usage of the data to solve the problem**, i.e. providing information and finding answers relevant for tourists, people who want to change neighborhood, business founders and investors will be **similar** to the approach applied in the Toronto and New York analysis.

The shared notebook on GitHub[3] includes code and markdown cells and therefore provides a full picture of the used methodology. For the **exploratory data analyis** the unique and the mean values for retrieved venues per neighborhood were calculated. For the final **clustering approach**, the unsupervised machine learning algorithm k-clustering was used. **Interferential statistical testing** was not systematically applied for this project, although a thorough analysis in particular for different k-values could have been valuable.

4. Results

The first result and starting point of the subsequent data visualization and clustering of data retrieved from Foursquare servers is shown in Figure 1. The four districts appear as differently colored areas. The color scale is such that increasing population is represented by an increase of redish color parts. The blue dots stand for the 51 neighborhoods of Bonn.

- In Figure 1, it can be seen that **districts with higher population** have **more neighborhoods** than districts with lower population.

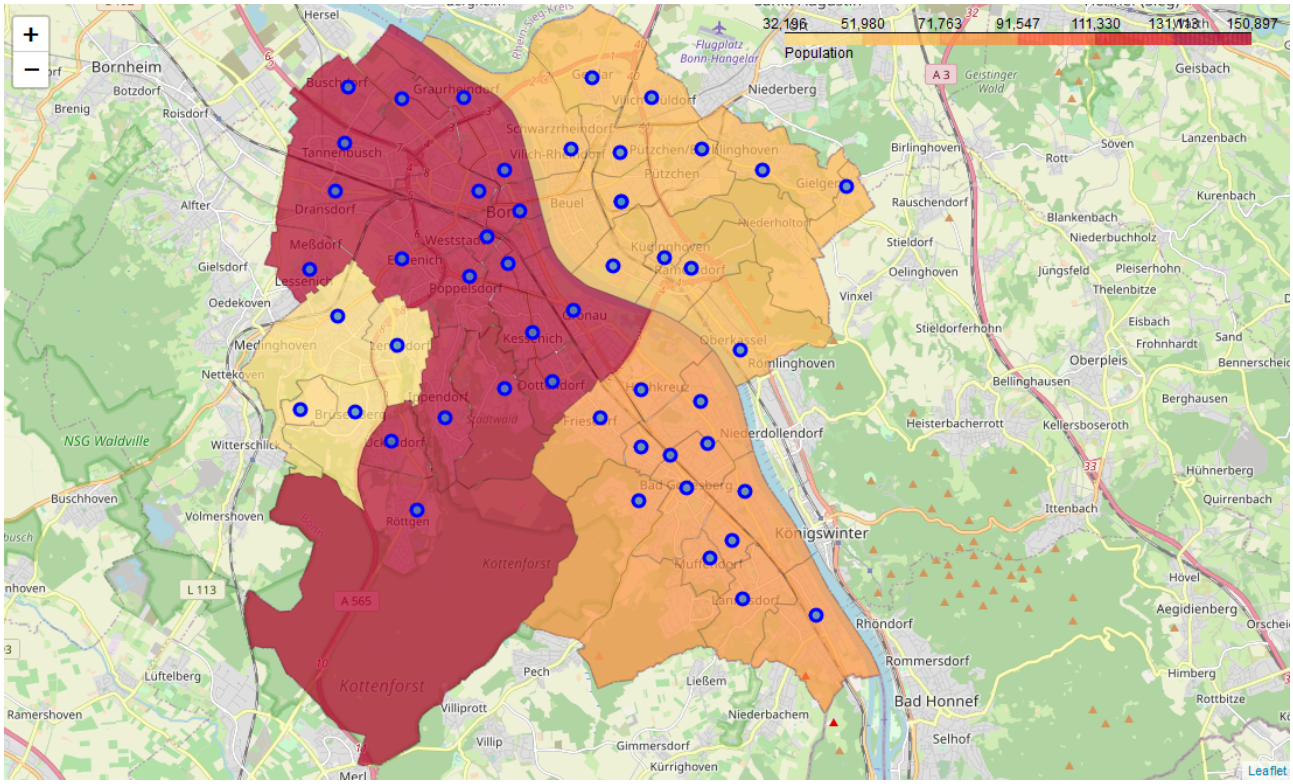

Figure 1: The choropleth map shows the population for all neighborhoods in a municipal district in colors ranging form yellow to red. The district of Bonn (in red) has the highest population. The district with the second highest population is Bad Godesberg and is situated south-east of Bonn. The district Beuel, shown north east of the district Bonn, is on the third place and the district Hardtberg which is the most western district has the lowest population. The blue circles are markers and the stand for the neighborhoods to be analysed.

For the 51 neighborhoods, in particular by their latidude and longitude values, is is possible to use the Foursquare "explore" functionality to retrieve venues in a radius of 500 m around the input latidude and longitude values. It has to be noted that for some neighborhoods up to 25 venues (set as upper limit) could be retrieved and for other neighborhoods only 1 or 2 venues in a radius of 500 m around the input latidude and longitude values could be retrieved. It seems that for some neighborhoods only few entries exist in the Foursquare database. It could not be determined if the dataset is incomplete or if in these neighborhoods only few venues exist.

In Figure 2 the number of retrieved values per neighborhood are shown. The same color code as before was used and the increasing amount of a quantitiy, here the number of venues, is color coded by an increase of the red color component. It can be seen that only for few neighborhoods the chosen maximum value of 25 could be retrieved. Only one neighborhood in the district of Bad Godesberg (in the South) and a cluster of neighborhoods in the district of Bonn (north west) are venue rich.

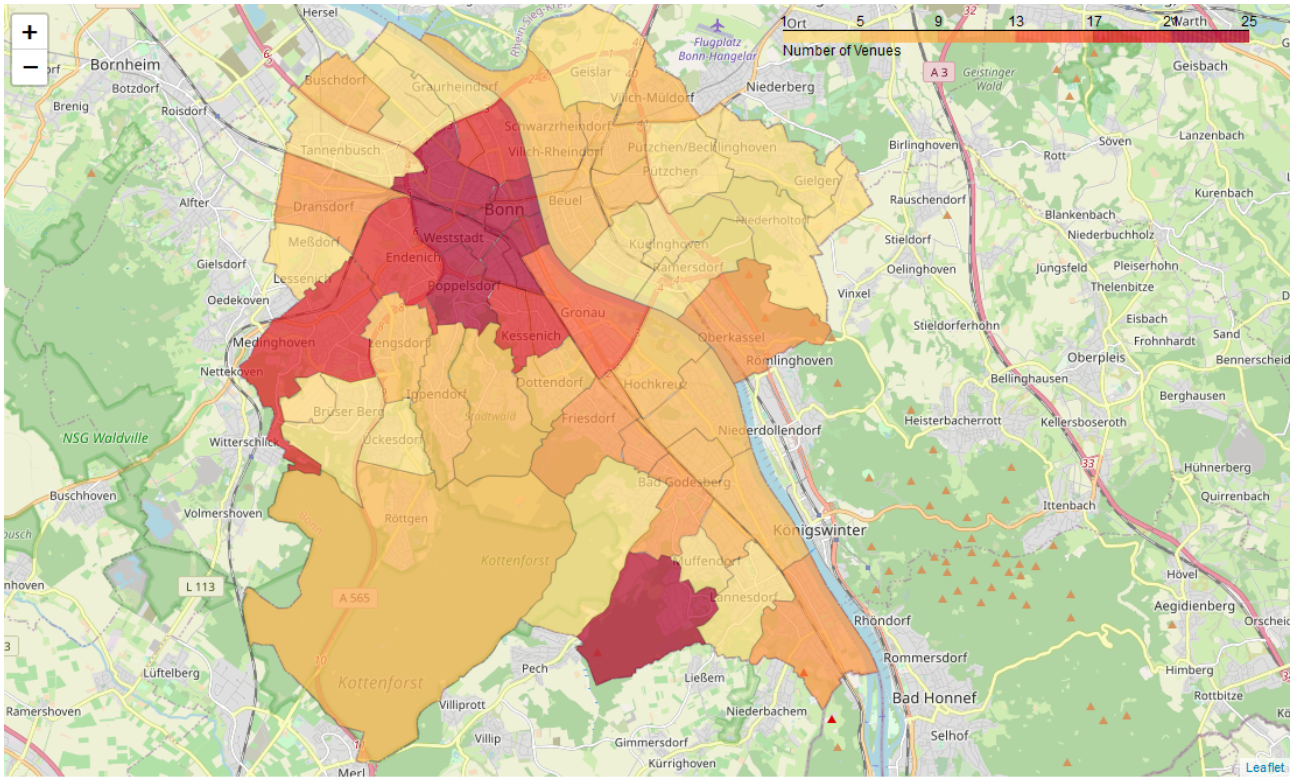

Figure 2: Number of venues per neighborhood as retrieved from Foursquare servers with maxium venue value of 25 and radius of 500 m around input longitude and lattidue value. Increasing amount of venues are illustrated by increasing red color component.

- In Figure 2, it can be seen that some neighborhoods are **venue rich** (25 venues or more, color coded in red) and other neighboorhodds are **poor in venues** (color coded in light yellow).

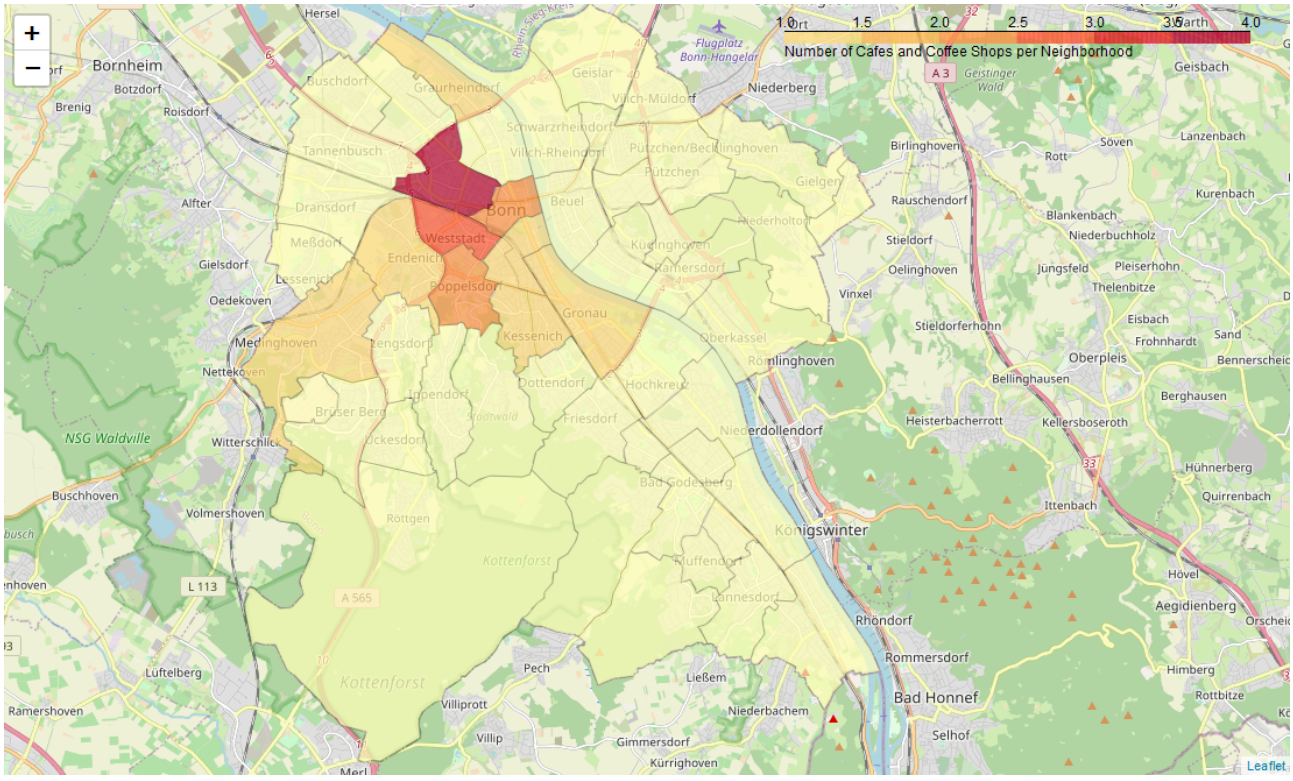

Figure 3: Number of Cafés and Coffee Shops. The retrieved data set with venues per neighborhood was filtered and only venues with venue category "Café" or "Coffee show" are shown.

A more detailed classification solely based on the quantity of retrieved venues (as shown in Figure 2) could be used for statistical analysis and classification. Instead of pursuing this approach, the focus was layed on a qualitative analyis and in Figure 3 the number of Coffee Shops and Cafés per neighborhood is shown. It can be seen that the highest amount of Cafés and Coffee shops are found in the center of Bonn. It seems questionable that the dataset on the Foursquare servers is complete and a low number of Coffee shops outside the center of Bonn might be an artefact. Nevertheless, assuming that the data set is complete, the presented visualisations allows to **understand the field of competitors** in case of dealing with the question of opening a café or coffee shop in a particular neighborhood.

- In Figure 3, it can be seen that the **center of Bonn** is **rich in cafés and coffee shops**.

Following the procedure presented in the analysis of neighborhoods in New York and Toronto, an unsupervised machine learning algorithm (k-clustering) was applied to check if such a clustering approach helps to better understand the venue data set retrieved from Foursquare servers. As explained in the methodology section, interferential statistical testing was not systematically applied. Instead it was found by trial and error that a value of k = 4 let to a classification which could be relatively easy interpreted and provided some additional value. The result of the k-clustering, i.e. the visualization of the four clusters is shown in Figure 4. The majority of the neighborhoods belong to two different clusters shown in cyan or red. Only two neighborhoods appear to the remaining two clusters (shown in purple and ochre).

- In Figure 4, it can be seen that for **k = 4**, most neighborhoods are classified as belonging to the c**luster represented by red markers** or as belonging to the **cluster represented by cyan markers**.

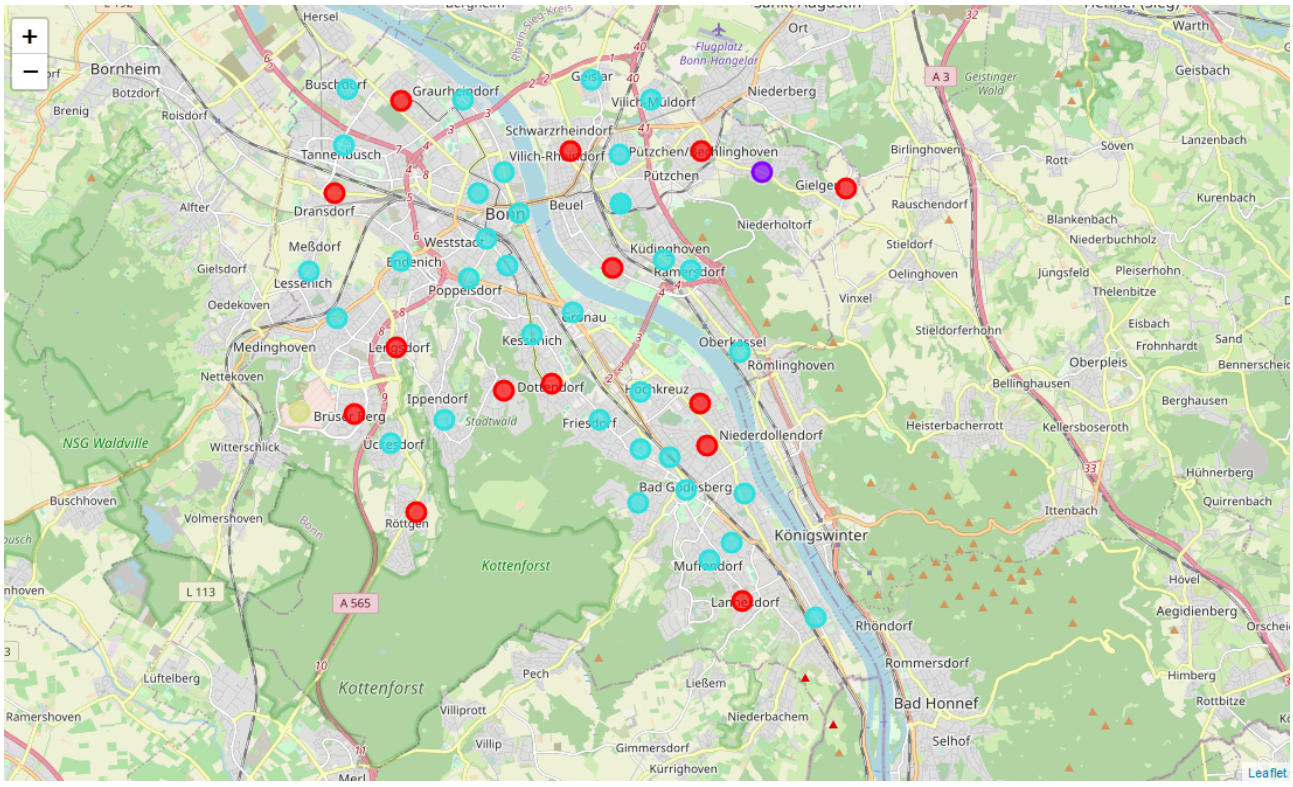

Figure 4: Results of k-clustering (with k = 4). The unsupervised machine learning algorithm was classifying venue categories per neighborhood into 4 clusters shown in red, cyan, purple and ochre.

The two clusters with only one value will be neglected in the subsequent analysis. Nevertheless, these two clusters were necessary for the k-clustering algorithm as they allowed to treat outliers and

niche venues. The analysis of the values classified as belonging to the cluster shown in red reveal that the most common venue categories for neighborhoods belonging to this cluster are supermarkets and tramstations. In Figure 5 the venue categories "supermarket" and "tram station" per neighborhood are shown. It can be seen that there is a certain overlap with the identified clusters, but as well some exceptions as neighborhoods with supermarkets and tram station were nevertheless classified as belonging to the second cluster with more diverse venues. The cluster were supermarket and tram stations are the dominant venue categories is therefore different from neighborhoods where supermarket and tramstations exist.

- In Figure 5, it can be seen that neighborhoods were venues belonging to the venue category **supermarket** or **tramstations** are h**ighlighted in red,** are **only partly overlapping** with neighborhoods belonging to the **cluster** were supermarket and tramstations are the **dominant venue categories**.

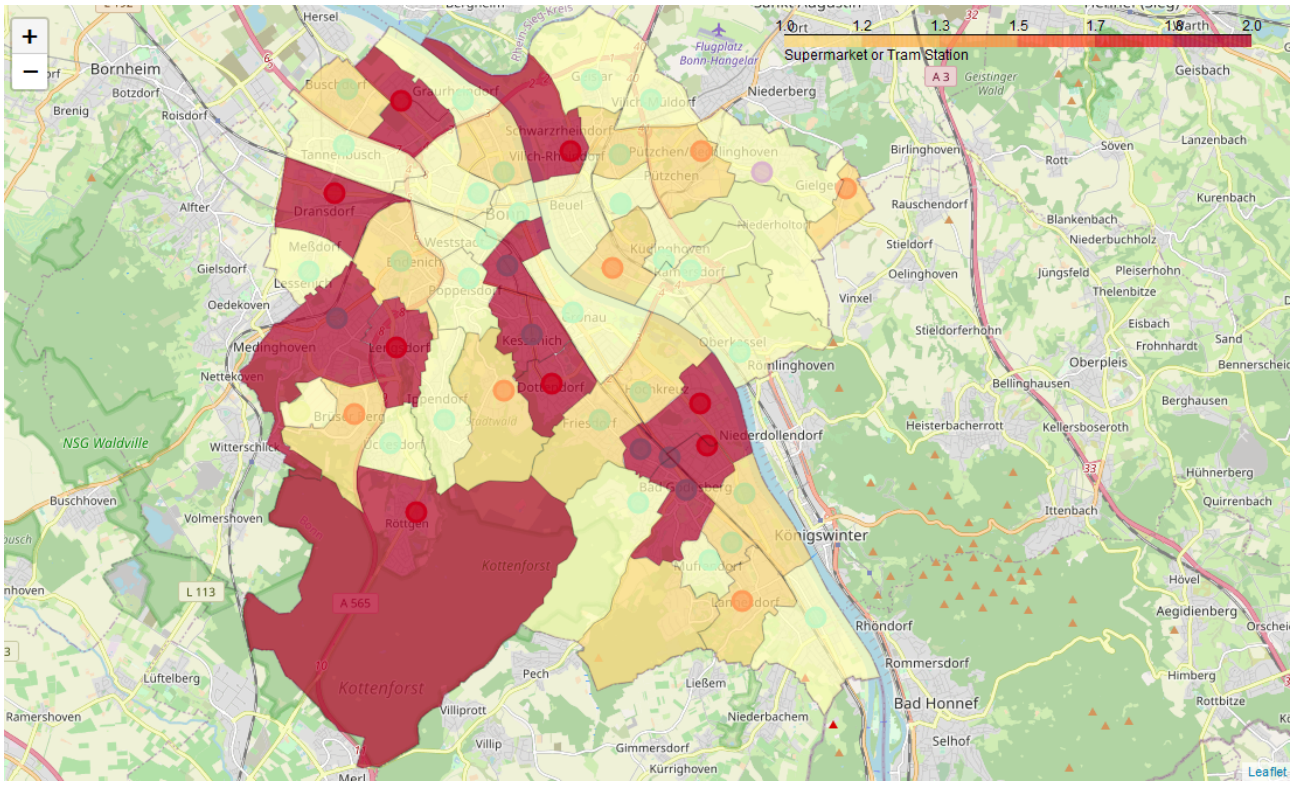

Figure 5: Visualization of Supermarkets and Tram Stations per Neighborhood. Neighborhoods with venues belonging to the venue category Supermarket or Tram station are only parly overlapping with the neighborhoods belonging to the cluster where Supermarket and Tram Station are the dominant venue categories.

5. Discussion

The data analysis and the visualization of the data in five Figures allow to have a closer look at the spatial distribution of population and venues in Bonn. Visualization of particular venue categories such as Coffee shop and Café allow to study the field of competitiors. In my opinion, the retrived data set might be incomplete as the Foursquare data (limit of retrieved venues = 25) shows no venues of this type for some neighborhoods. This seems unrealistic to me. Despite concerns if the dataset is complete or not, the visualization clearly shows that a coffee shop hot spot exists in the center of Bonn. The classification approach using k-clustering helps to understand the data. The two dominant clusters can be distinguished as diverse neighborhoods and rather monotone neighborhoods where supermarkets and tramstation are the most common venue categories. A

further analysis revealed that supermarket and tramstation themself are not an indicator if a neighborhood is monotone or diverse in terms of venue. It is their first and second place of most common venue category type which is a good indicator of a rather monotonous neighborhood.

6. Conclusion

The results shown in the report might help a future buisness founder or investor to understand the field of competitors in the coffee shop business. Nevertheless, the datasets seems to be incomplete and relying only on Foursquare data, despite being a good first indicator, might be insufficient to get the full picture. A classification of the different neighborhoods based on the most common venue category, led to the conlusion that neighborhoods, where tram stations and supermarkets are the most common venue category, are rather monotone. If you want to live in a diverse neighborhood in Bonn, and close to the coffee shop hot spot in the center, then this "Battle of Neighborhood" analysis might help you to find a neighborhood belonging to the cluster representing a high diversity of venues.

References

[1] “https://en.wikipedia.org/wiki/Bonn.”, Wikipedia.
[2] “https://opendata.bonn.de/dataset/fl%C3%A4chen-der-ortsteile.”, .
[3] Zarathustra, M., “https://github.com/meleagros-zarathustra/Coursera_Capstone/blob/master/Submission_The_Battle_of_Neighborhoods_in_Bonn.ipynb.”