

Exploring venues in Berlins neighborhoods

Introduction

Due to rising rental prices in the city center of Berlin, more and more people move to the suburbs. However, most popular venues are situated in the city center. Are there enough venues for the rising number of inhabitants in the suburbs of Berlin? This would be an interesting question especially for people planning to start a venue. I want to help them find out in which neighborhood there is still a shortage of venues based on the number of inhabitants in this neighborhood. Depending on which type of venue they want to open, I will show which neighborhood already has enough venues of this kind and which lack these types of venues.

The data analysis will also be interesting for people who want to move to Berlin or tourists who are on vacation and decide where they want to stay. Depending on what they want in their neighborhood - nature and sights or rather restaurants and nightlife or maybe child-friendly places like playgrounds - they will get an overview of which neighborhood has the most to offer. For people moving to Berlin it will also be interesting to see whether there is a district in which there are quite a few venues but only little number of inhabitants. The rental prices might be lower in this district as it doesn't seem so popular.

Therefore the data showing which neighborhood is famous for which venue will be relevant for quite a few different target groups.

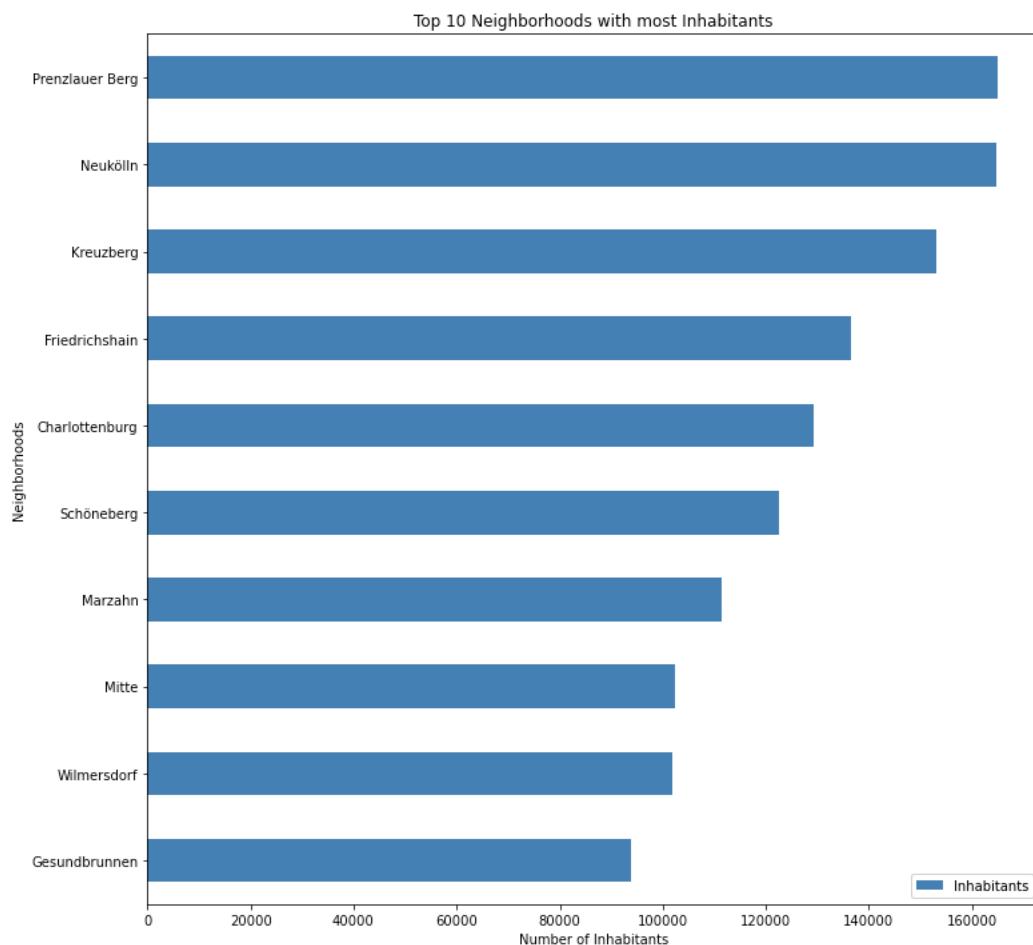
Data

The data I will be using is the 2nd table from https://de.wikipedia.org/wiki/Verwaltungsgliederung_Berlins as it contains the number of inhabitants per neighborhood from the end of last year. I didn't want to use the boroughs as these unite many neighborhoods into one borough. This would be too imprecise. As this is German data, numbers have a dot instead of a comma to indicate 3 places. To continue using this data in a correct way I transformed it into the American decimal system. In this table there was also the area of the neighborhood given in km². As I will work with Geojson data I deleted the area column. From the geocoder library coordinates for each neighborhood were read and added to the dataframe. These coordinates are then used to get the venues for each neighborhood from Foursquare. Based on the latitude and longitude values of each district we get all venues in a radius of 600 around these coordinates. I chose this radius as it gives the highest number of venues per district without too many duplicates. A higher radius would produce more duplicates whereas a smaller radius would miss out on certain venues. For each venue I pull its name as well as its coordinates and venue category. The coordinates are needed to visualize each venue on the map. The venue category facilitates the comparison between neighborhoods. To see which venue category there is per neighborhood we use one hot encoding so there is only one row per neighborhood. We still have 271 columns indicating that there are way too many venue categories. I group them into 11 main categories and delete the detailed categories. This helps to generate bar charts that answer the question on which kind of venues there are per neighborhood.

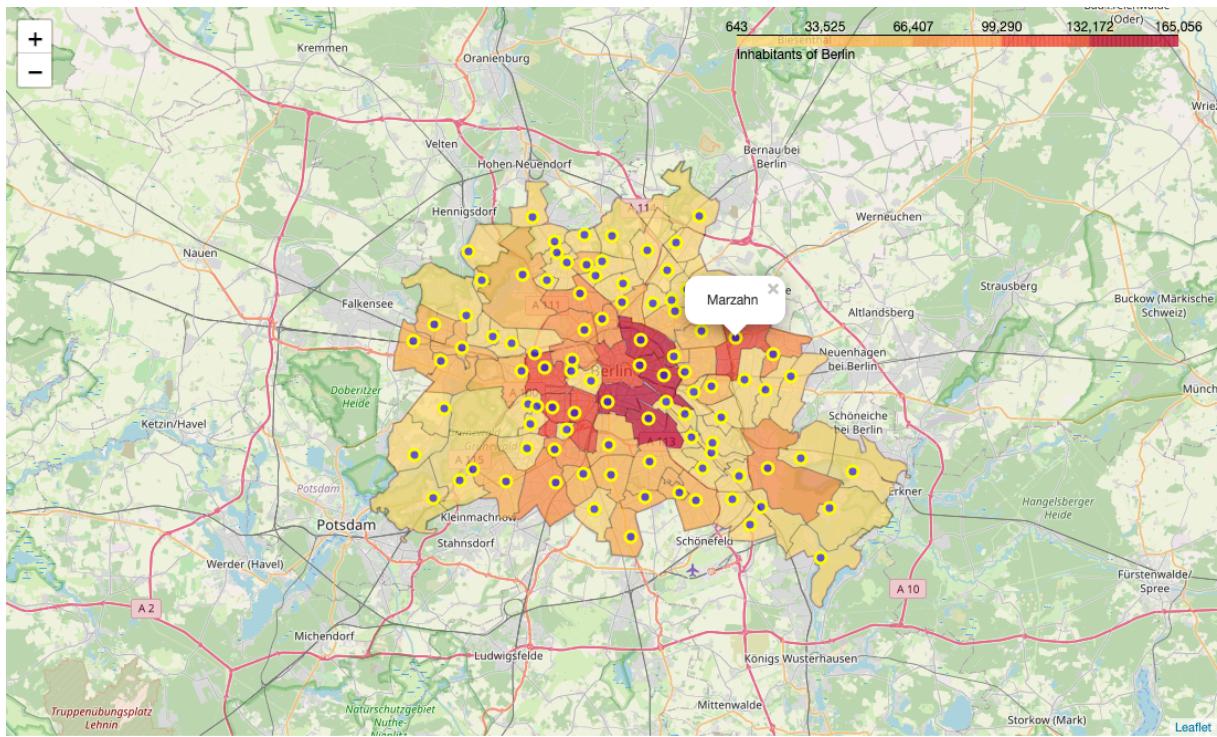
To mark the areas of Berlin in a map, I also downloaded the Geojson file of the neighborhoods in Berlin: <https://opendata-esri-de.opendata.arcgis.com/datasets/esri-de-content::ortsteile-berlin>. This allows to create Choropleth maps which make it easier to visualize two pieces of information on one map without appearing to cluttered. The goal is to visualize both inhabitant as well a venue data in one map to answer the initial question in which neighborhood one can find the highest number of venues compared to the number of inhabitants.

Methodology and Results

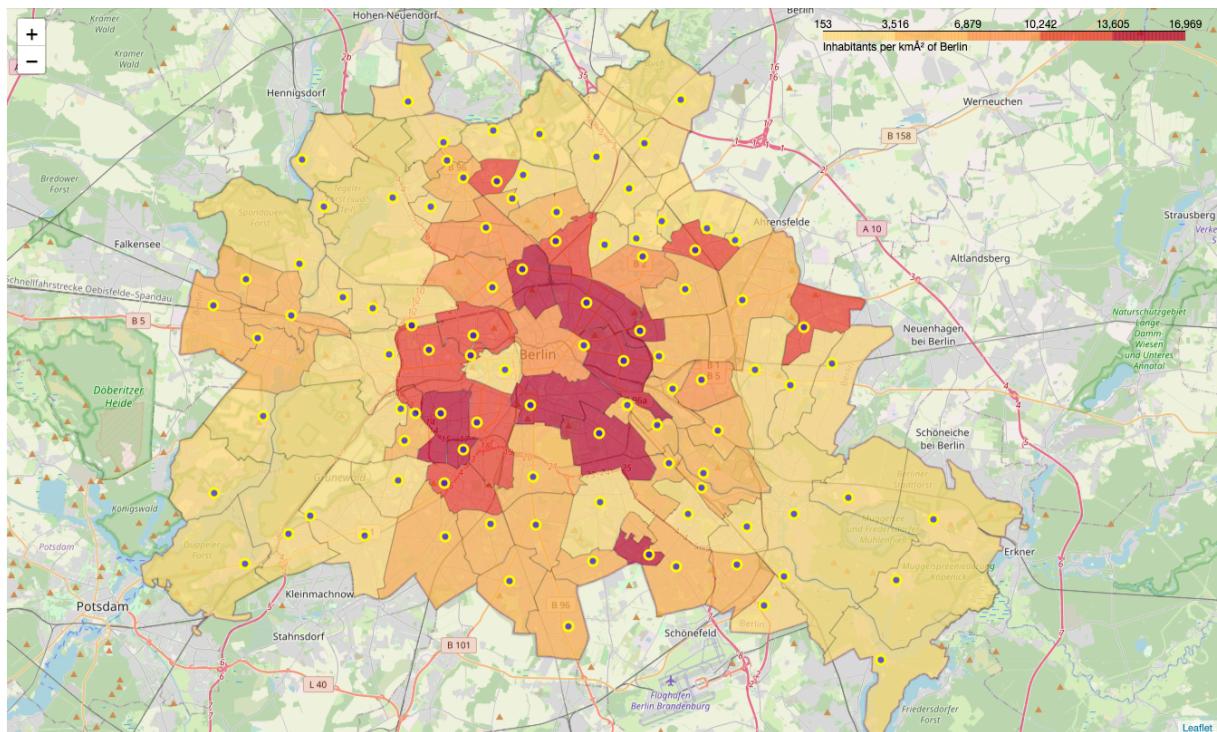
First of all I generated a bar chart of the 10 neighborhoods with the most inhabitants in Berlin to get a feeling of the data. One can see that districts in the eastern center of Berlin like Prenzlauer Berg, Neukölln, Kreuzberg and Friedrichshain are the 4 most populated neighborhoods of Berlin. The western center of Berlin is placed 5th and 6th with Charlottenburg and Schöneberg. One suburb is in the top 10 Neighborhoods ranked seven, Marzahn, followed by other city center districts.



To get a better overview of the distribution of the inhabitants, I visualize the data on a Choropleth map with markers of the district names. We can see that many inhabitants live in the Eastern center of Berlin but also in the outer district of Marzahn as we already found out in the bar chart.

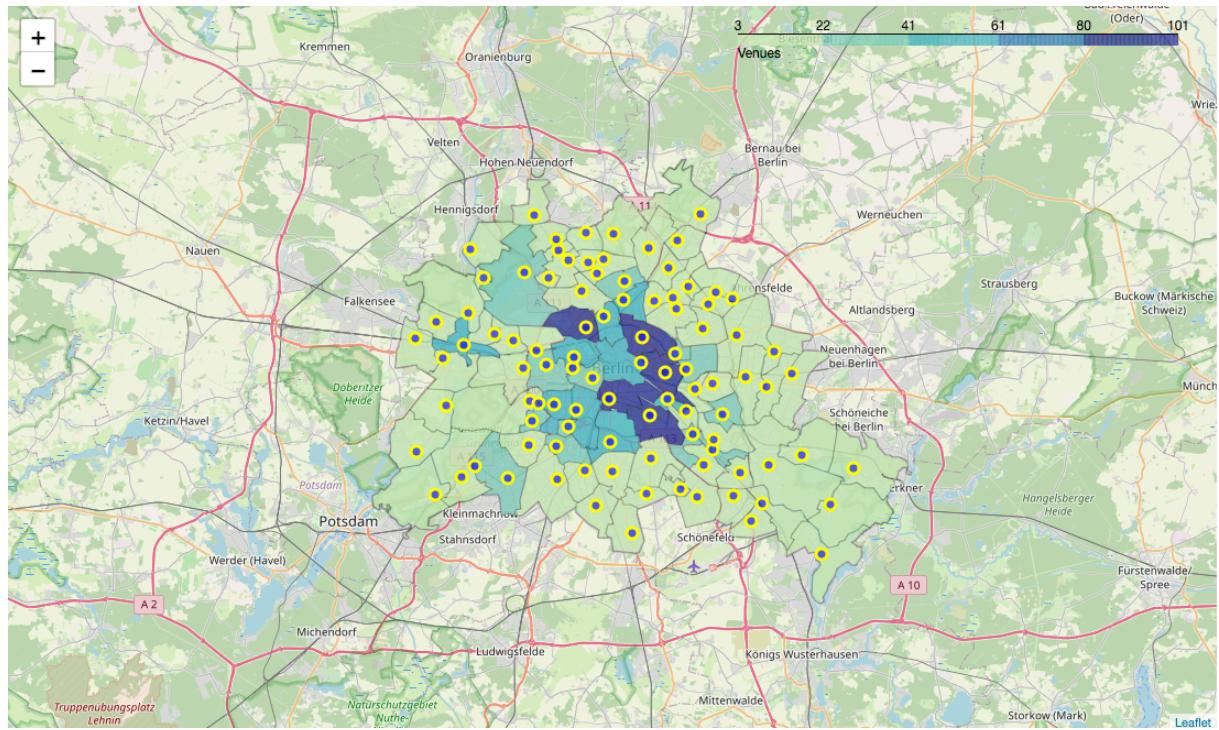


To see whether it gives me more information I also visualize the number of inhabitants per square kilometer. Two more districts in the outer eastern part of Berlin have quite a few inhabitants per km²: Hellersdorf and Neu-Hohenschönhausen. There is also one small area in the south called Gropiusstadt that has many inhabitants for its area, as well as another area in the north called Märkisches Viertel.

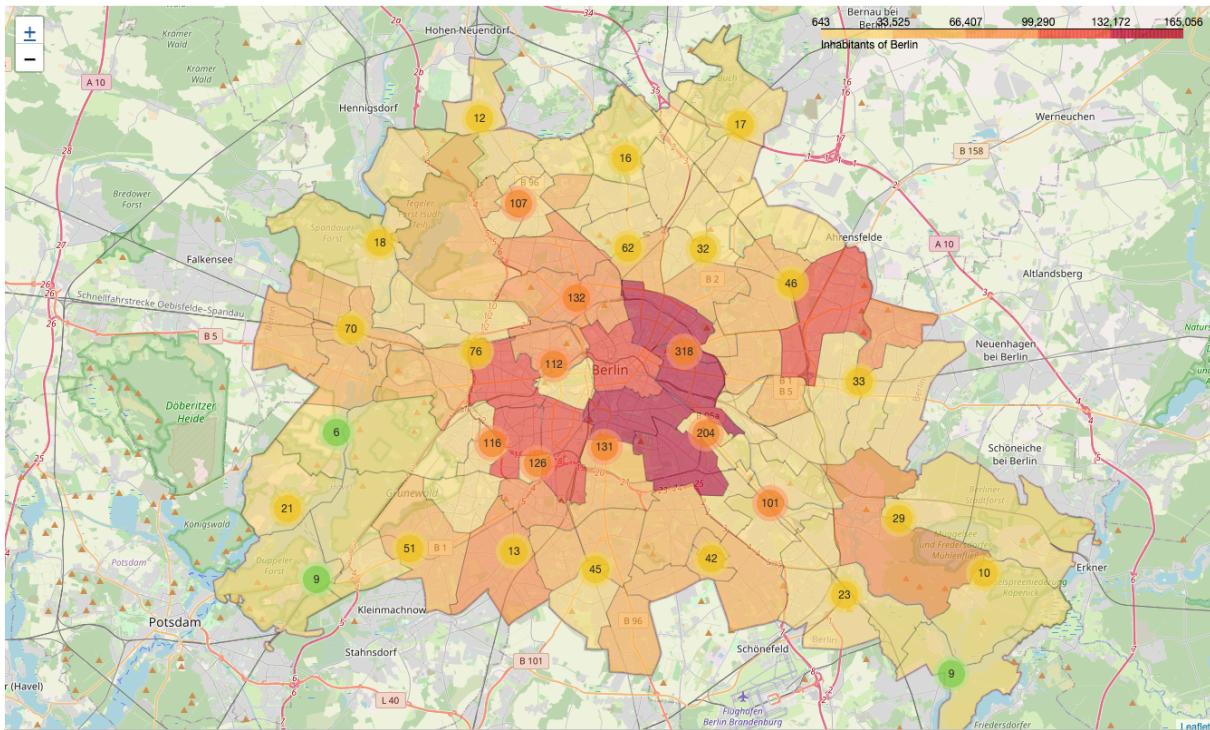


Looking at the foursquare data we can see which districts have the highest amount of venues: Neukölln, Kreuzberg, Friedrichshain, Prenzlauer Berg, as well as Wedding. There are also some suburbs like Spandau that have just as many venues as the center of Berlin (40-

60). Also the northern suburb Tegel and the southern suburb Zehlendorf have more venues than other suburbs of Berlin (22-40).



Now that we have both data, we combine it on a map. We can see that where there live fewer people, there are fewer venues. We can see that Wedding, Friedrichshain, Prenzlauer Berg, Kreuzberg and Neukölln are the Neighborhoods with the highest number of venues. However, districts like Marzahn, Köpenick or Lichtenfelde have quite a lot inhabitants but only few venues to go to. Tegel has quite a few venues but not so many inhabitants. As Tegel airport has been closed only recently, there is quite a lot of potential for it to increase its inhabitants.



Which venues are most common in which district?

For people who like breakfasts in Cafes the eastern districts like Prenzlauer Berg and Friedrichshain have a lot to offer, followed by the northern district Wedding.

After a nice breakfast one can go shopping again in Prenzlauer Berg, Wedding or also Kreuzberg which is not far off. A little longer but also good for shopping is Spandau which is a suburb in the western part of Berlin.

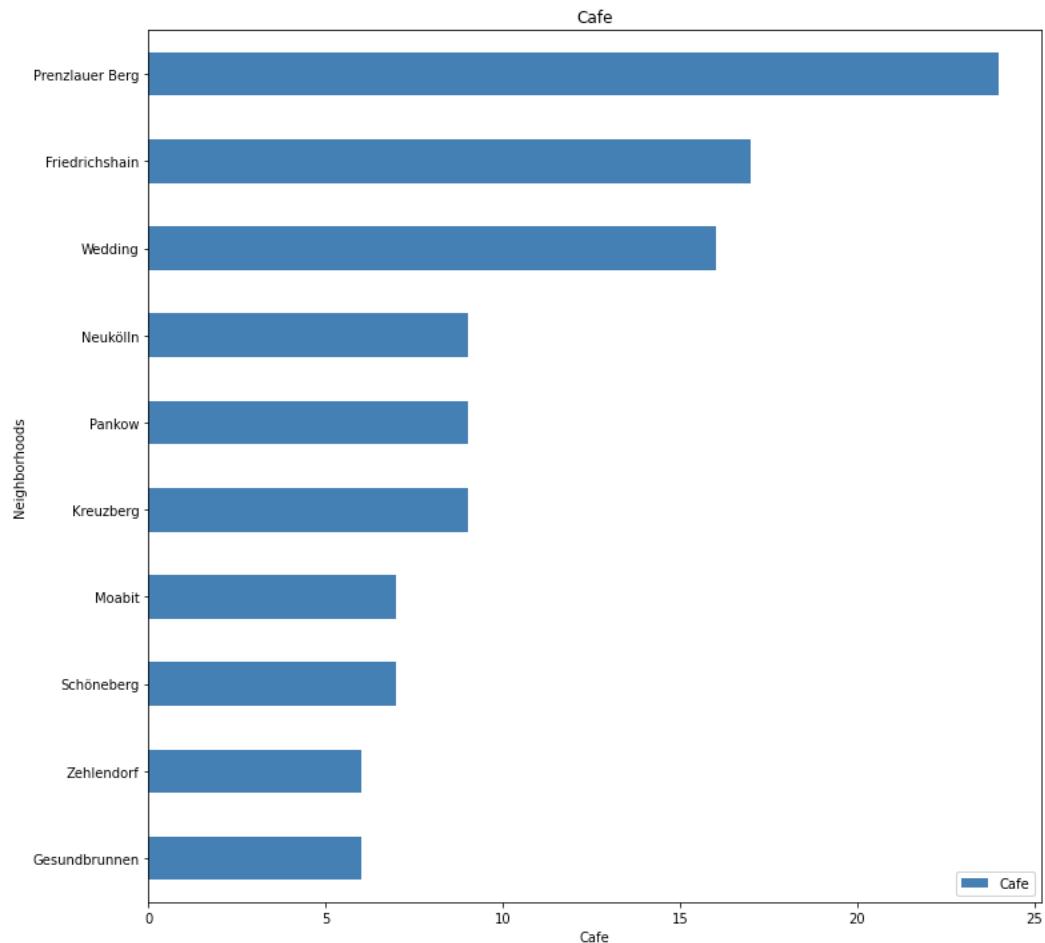
For entertainment venues like Laser Tag, Planetarium, Comedy Club , Go Kart Tracks or Football Stadium, Neukölln, Wedding and Kreuzberg are again the places to go to.

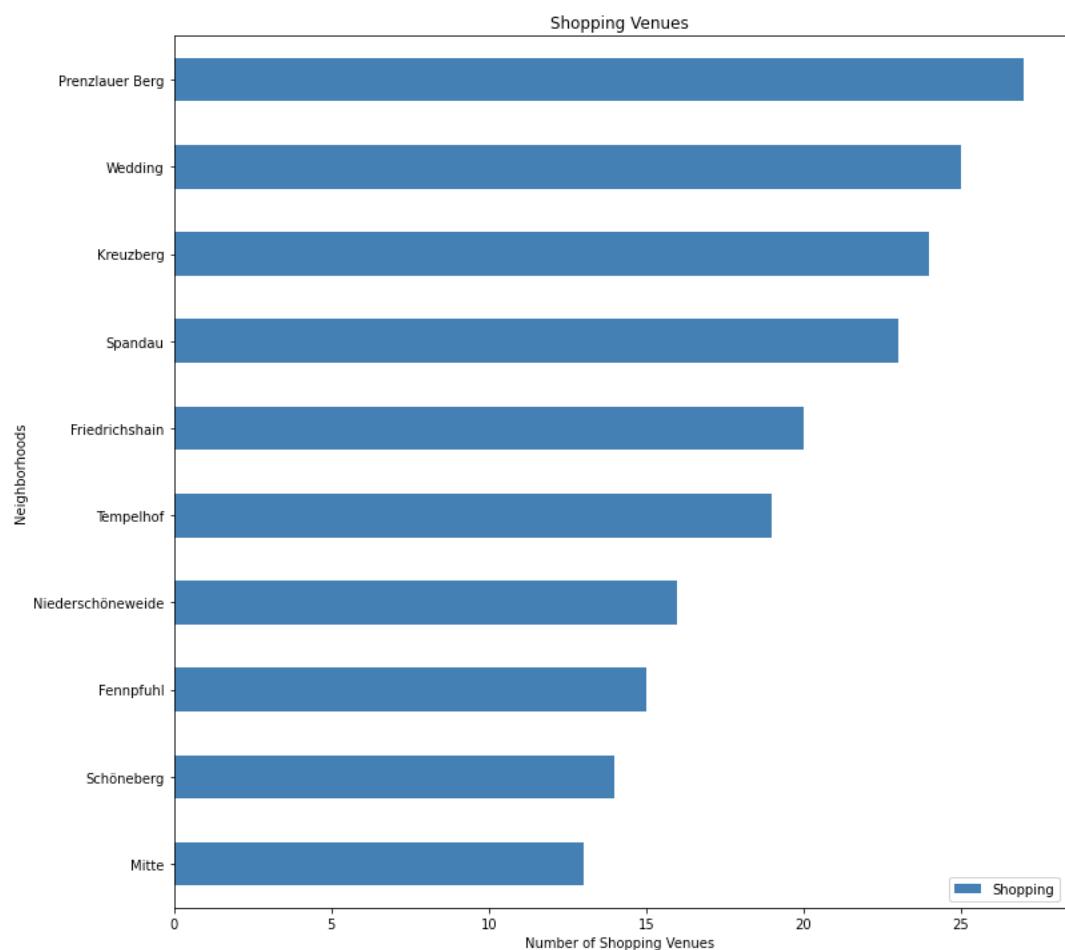
If one prefers sights and nature instead of shopping or entertainment, Mitte, Schöneberg and Grunewald are the neighborhoods to go to.

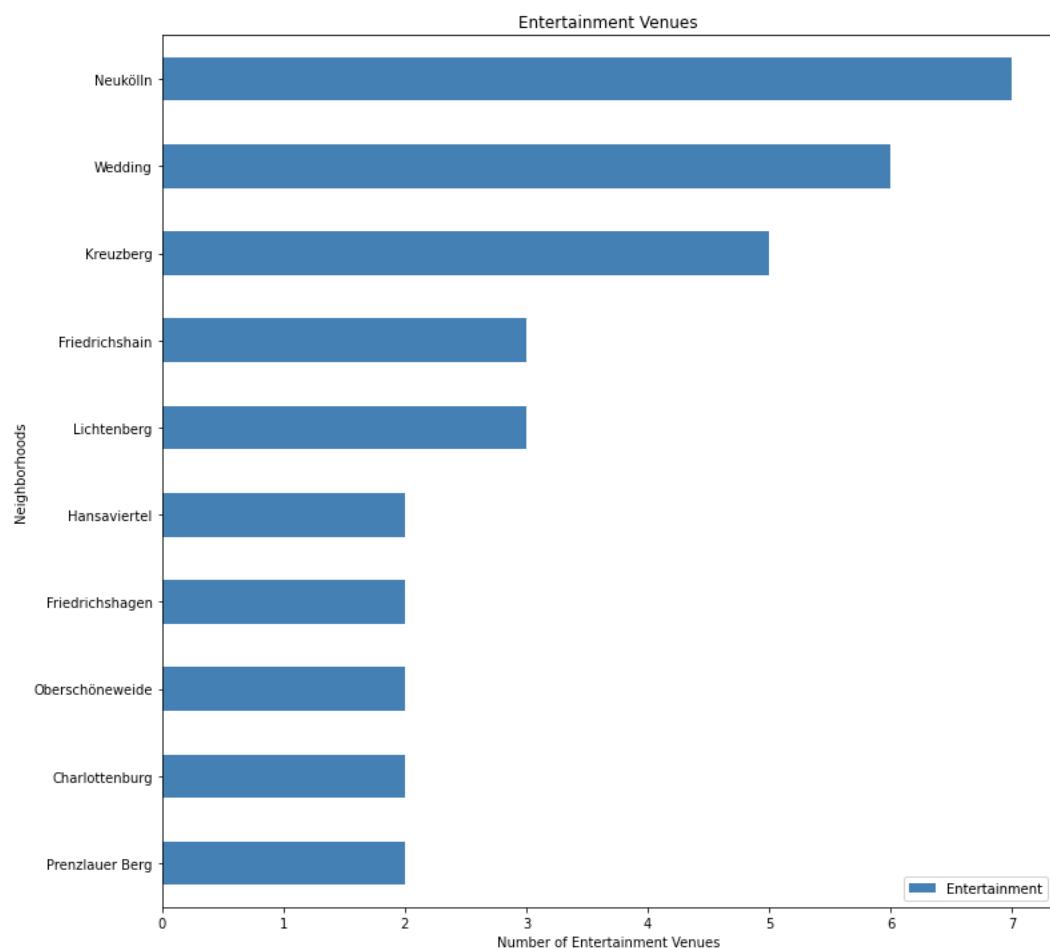
People who like to eat in restaurants best be advised to again visit Wedding, Neukölln, Friedrichshain or Kreuzberg.

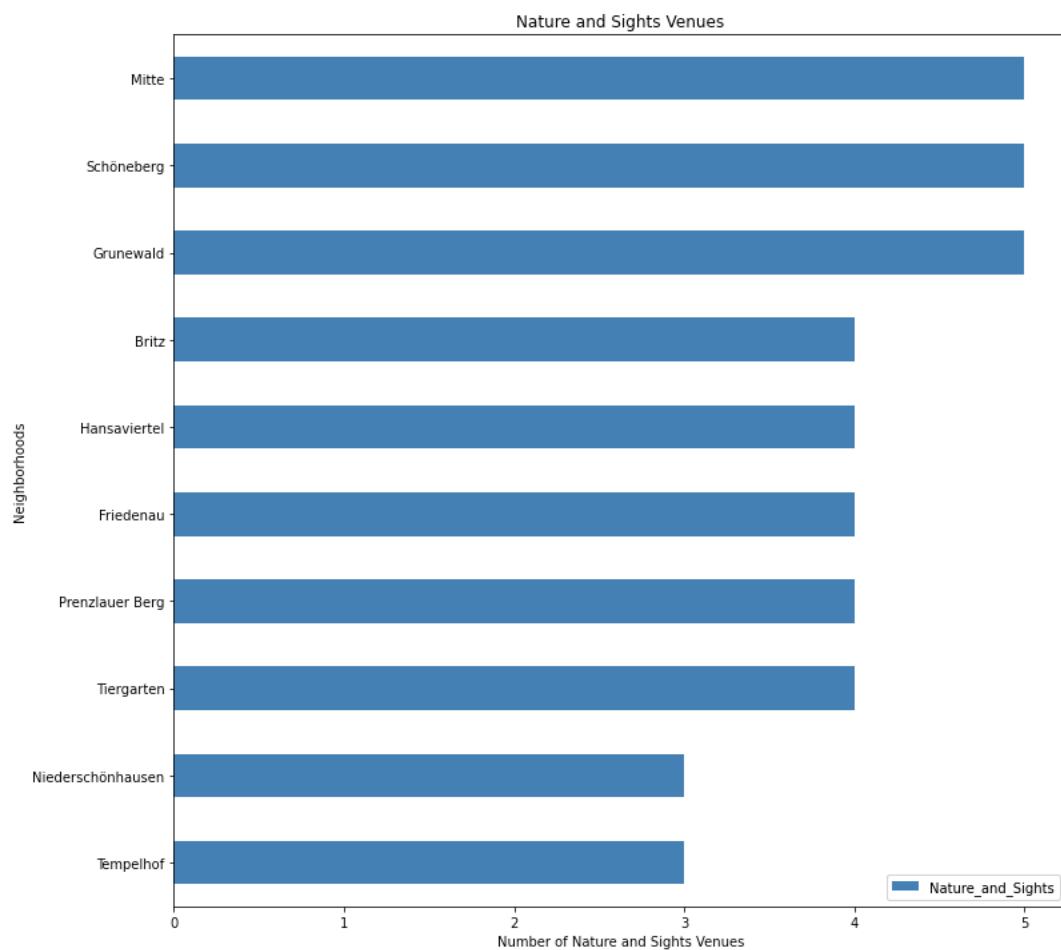
For nightlife like pubs and clubs Neukölln is by far the neighborhood with the most nightlife venues.

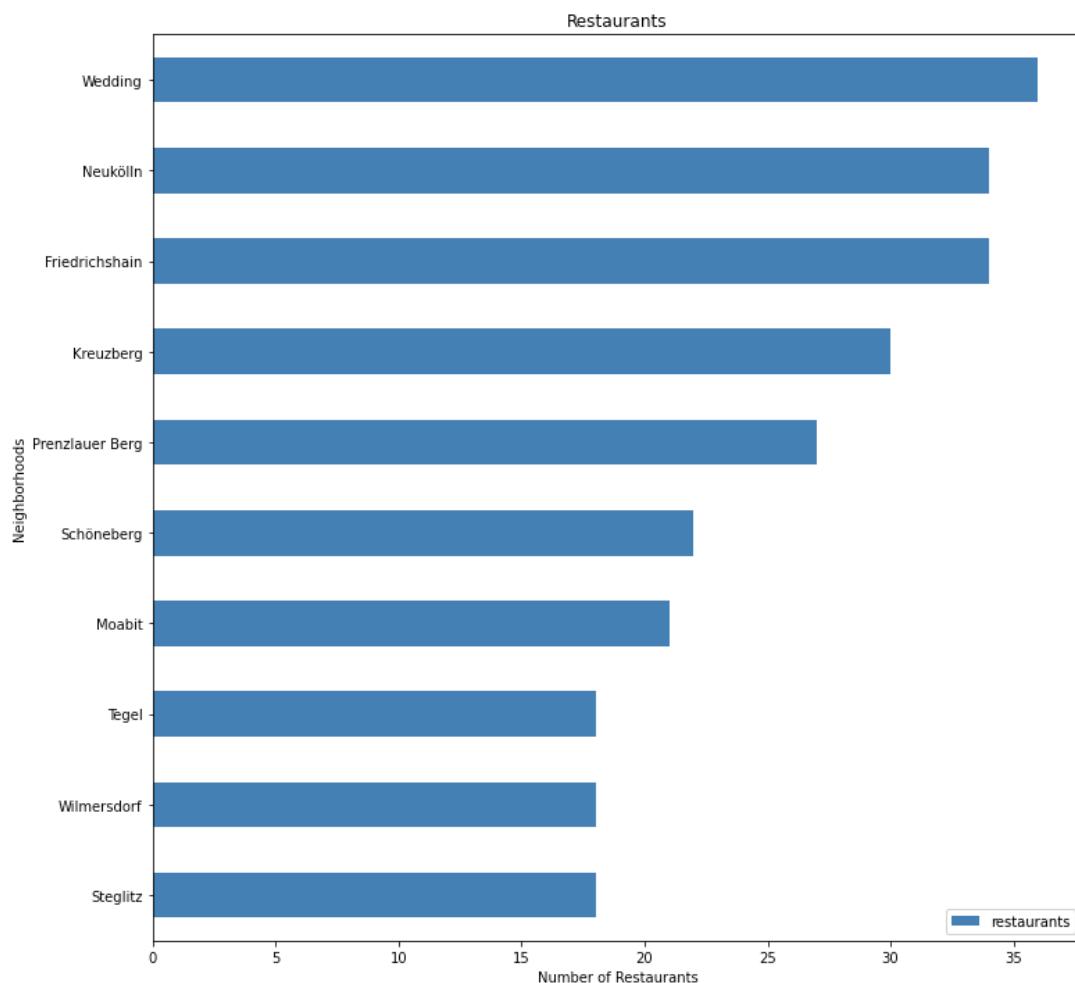
The other 5 categories simply didn't have enough data to give a good indication of which district has the most venues.

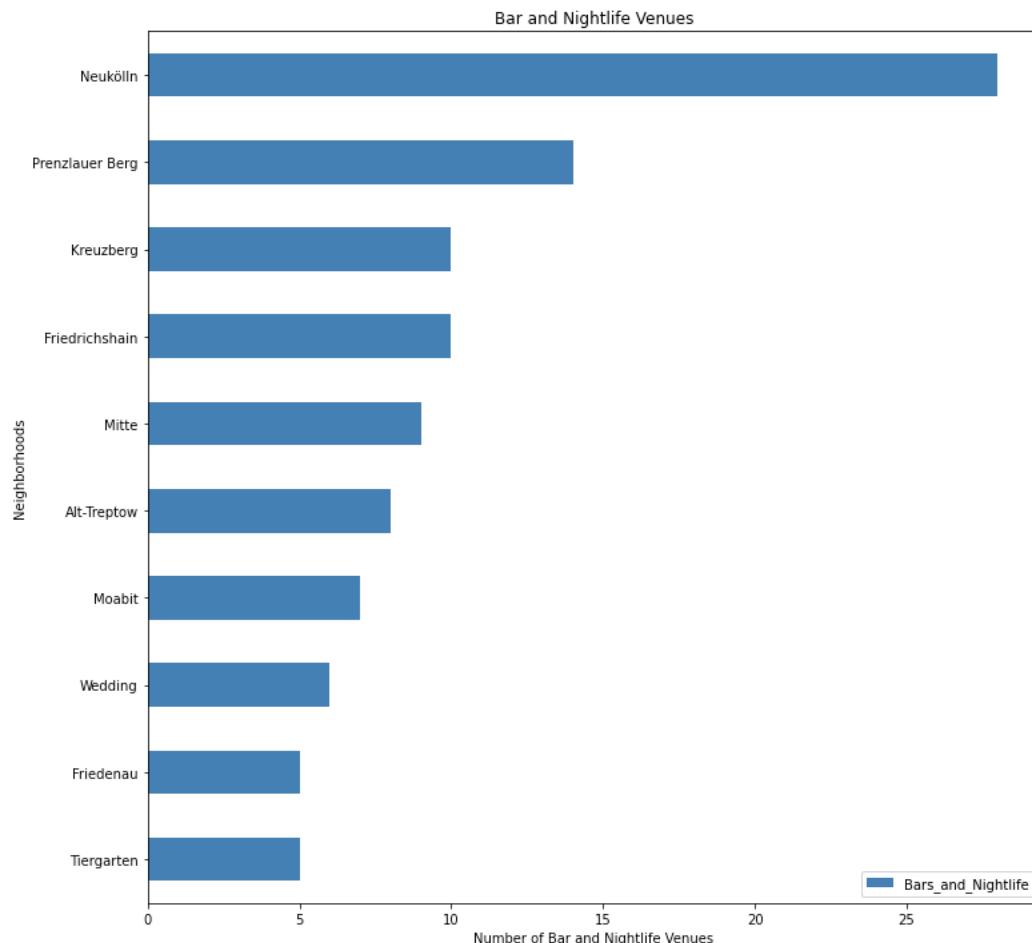








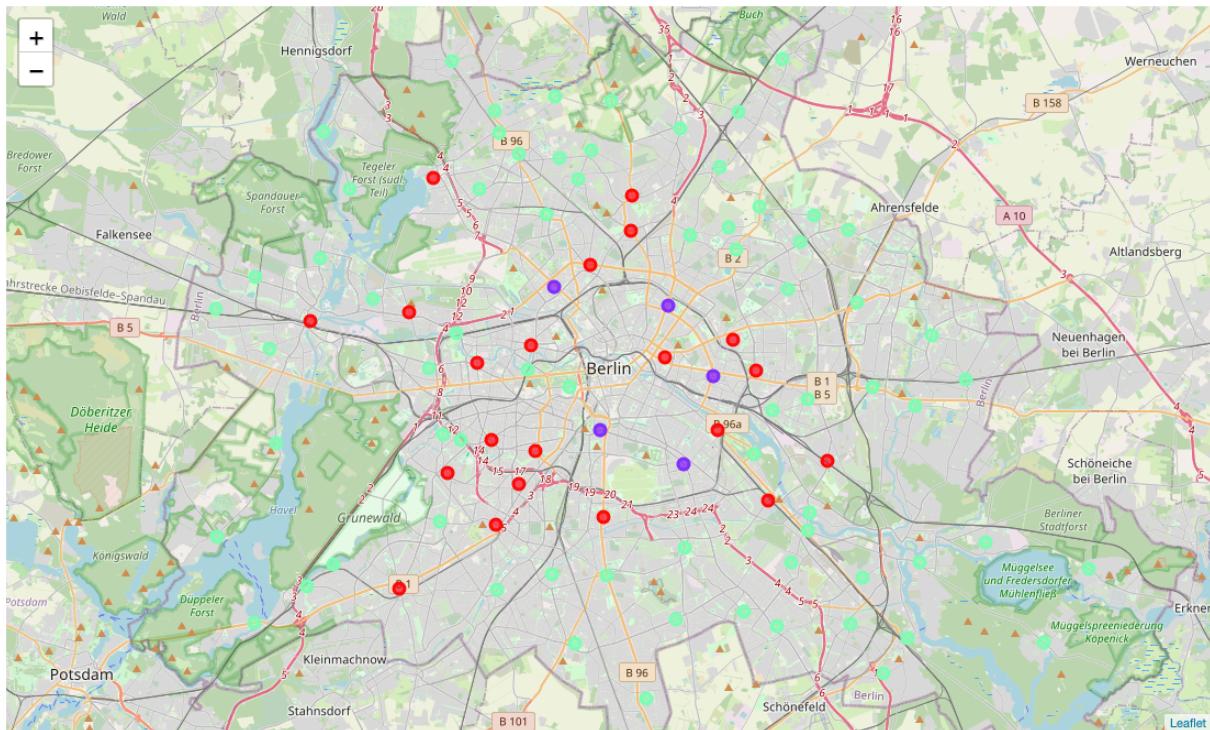




Looking at the cluster analysis of most common venues per district we can see that there are three clusters that are similar to the number of venues we find in these districts. The first cluster only includes the 4 neighborhoods that have the highest number of venues: Neukölln, Friedrichshain, Prenzlauer Berg, Kreuzberg and Wedding (purple). This area is most popular for its Cafes and Bars.

The second cluster represents the districts that have many restaurants, cafes and supermarkets (red). These are in the outer skirts but closer to the center than the third cluster neighborhoods, except for Tegel, Spandau and Zehlendorf.

The third cluster is more diverse and mainly includes suburbs of Berlin, only two city center districts are to be found in this cluster: Hansaviertel and Tiergarten (green). It has supermarkets as well as Playground, Shopping and Nature Reserves in it. These are also the districts with the fewest venues.



Discussion

The data presented in this report has to be interpreted with care. As Foursquare only allows a certain amount of calls per day, the venues presented here are not at all complete. Especially when looking at a certain kind of venue, there are only 3-5 entries per district, so that it is not possible to compare districts.

Not all neighborhoods are the same size, depending on the radius one sets for the neighborhood, small neighborhoods which have a border to neighborhoods with many venues suddenly get more venues assigned.

Grouping has been done manually, therefore it is all based on my personal judgement and some groups got quite big as shopping no matter if food or clothes, all appeared in one group, this should also be looked at separately.

Conclusion

Analyzing the foursquare data one can see that the eastern city center of Berlin has the highest number of venues, followed by the northern district Wedding, especially when it comes to Restaurants, Cafes, Nightlife, Shopping and Entertainment. If one prefers sights and nature instead of shopping or entertainment, Mitte, Schöneberg and Grunewald are the neighborhoods to go to.

All in all one can say that where there are fewer inhabitants there are fewer venues. However, districts like Marzahn, Köpenick or Lichtenfelde have quite a lot inhabitants but only few venues to go to. Tegel has quite a few venues but not so many inhabitants. As Tegel airport has been closed only recently, there is quite a lot of potential for it to increase its inhabitants.

Therefore people who move to Berlin and like the action might want to move to the eastern center of Berlin or Wedding. This also seems the right place for people visiting Berlin. However, as the rental prices are quite high in these districts it might be worth checking out the surrounding districts or Tegel which has quite a lot of venues but not as many inhabitants yet. If one prefers nature and quiet, one might want to look for flats in Schöneberg and Grunewald.

For people who want to open a venue, Marzahn, Köpenick and Lichtenfelde are the places where there are still some venues needed. Obviously more research needs to be done to figure out if people who live there are willing to pay for new venues and which venue type is still missing in these neighborhoods.