Where to open a Bar in Berlin?

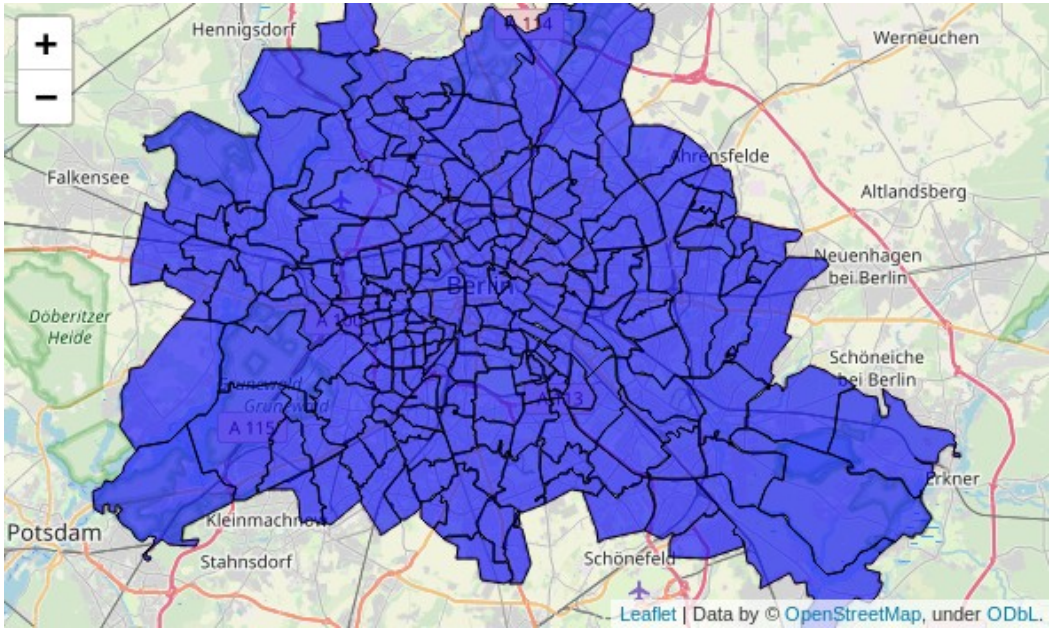
Karsten Poddig, 21.08.2019

The right location is essential for the success of a bar

- Berlin is probably the most important night life city in Europe
- Therefore there are already a lot of places for going out
- So it’s important to find a region
  - Which has not already to many bars on the one hand
  - And is still attractive in terms of venues on the other hand
- The aim of this analysis is to **investigate the relationship of bars and venues of other categories** with a Machine Learning model and to **use this model to find attractive regions** in Berlin for opening a bar.

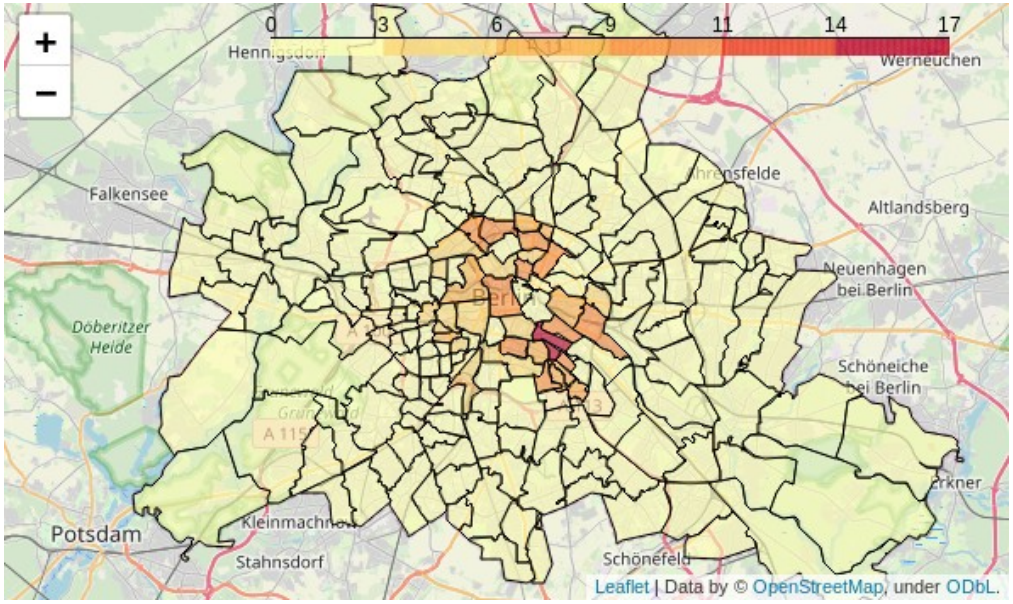
Collecting and Cleaning Data

- The Foursquare API is used to find venues in each region of Berlin
- The regions of Berlin come from a geojson-file from https://data.technologiestiftung-berlin.de/data/plz/plz.geojson
- The venues are grouped into categories and the most frequent categories are selected as independent variables for the Machine Learning algorithm
- Postal code regions of Berlin:

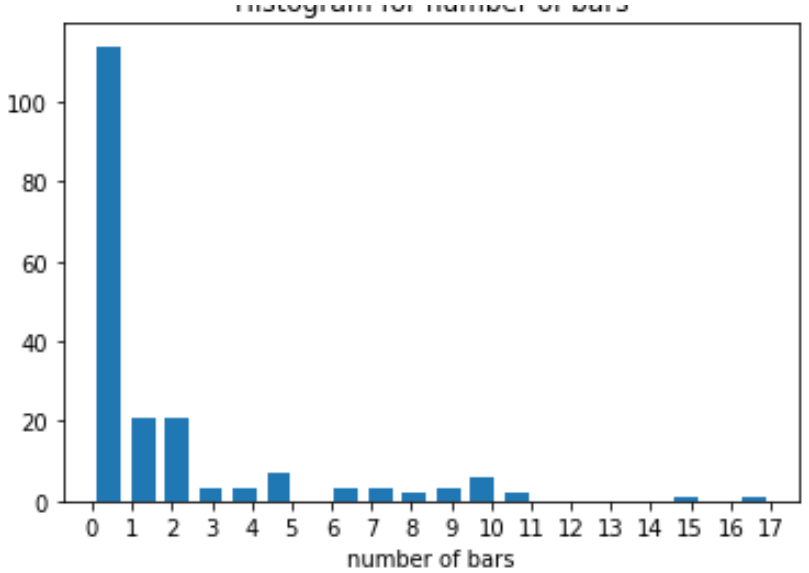


Exploratory Analysis

- How many bars are there in each region?



- Number of bars in each region



- Histogram of number of bars per region

Regression Model

Category	
Supermarket	322
Café	291
Italian Restaurant	226
Hotel	162
Bakery	159
Bar	154
Coffee Shop	124
Ice Cream Shop	117
German Restaurant	115
Bus Stop	113
Park	106
Vietnamese Restaurant	103
Drugstore	89
Plaza	79
Restaurant	78
Pizza Place	71
Gym / Fitness Center	64
Doner Restaurant	60
Asian Restaurant	57
Cocktail Bar	54
Organic Grocery	51
Tram Station	49
Greek Restaurant	48

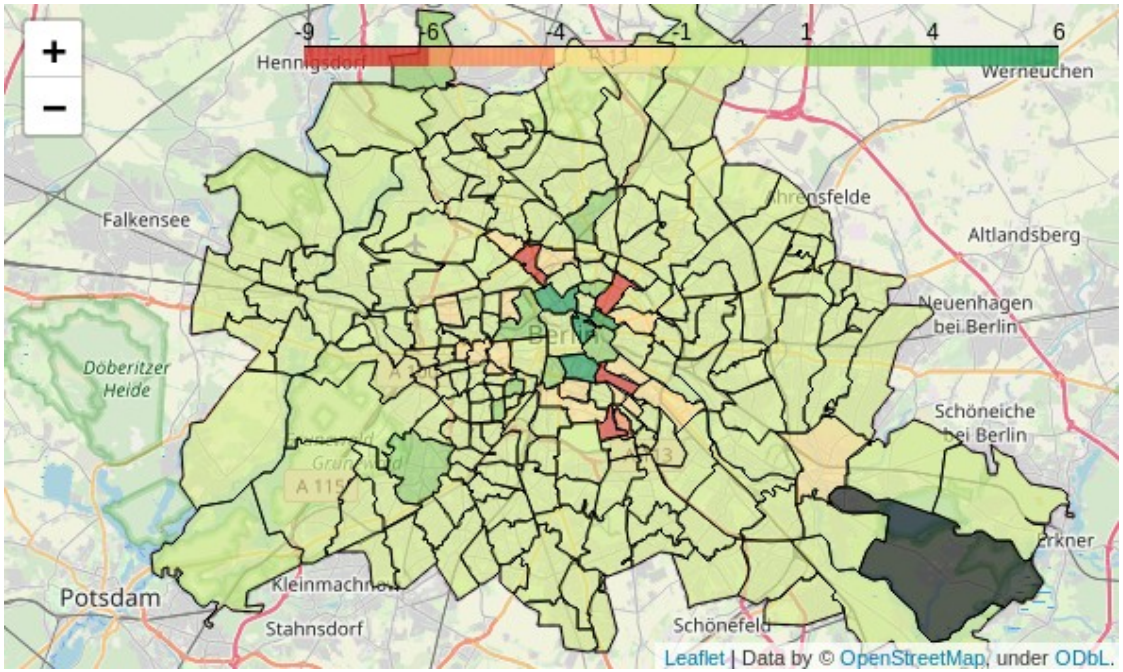
- independent variables: How many venues of other categories are there?
- The most frequent categories are used as independent variables!

Regression Model

- We apply a 1-Layer Neural Network with a rectified linear unit as activation function (relu) to predict the number of bars in each region
- This is simply a linear model with maximizing the output to zero
- The model yields a mean absolute error (MAE) of 0.74

Results

- The plot shows the difference of predicted and real number of bars:



- In regions with a high difference the predicted number of bars is much higher than the real number. **These regions (green) seem to have a lack of bars and could be attractive regions.**

Results

- Let’s take a look at the venues of the top five regions (right):
- Indeed in these regions there are not many bars compared to the number of venues of other categories.

postal_code	10115	10178	10969	10179	10963
Asian Restaurant	0	1	0	0	1
Bakery	1	0	2	1	1
Bar	5	0	3	0	1
Bus Stop	0	0	0	0	0
Café	3	2	3	3	3
Coffee Shop	7	4	3	0	1
Doner Restaurant	0	0	0	0	0
Drugstore	1	0	0	1	0
German Restaurant	1	1	1	1	0
Gym / Fitness Center	0	1	1	1	1
Hotel	4	3	3	10	4
Ice Cream Shop	2	2	0	0	0
Italian Restaurant	3	0	2	0	0
Park	0	0	2	0	2
Pizza Place	0	0	1	0	0
Plaza	0	1	0	1	0
Restaurant	4	2	0	0	1
Supermarket	1	0	0	0	0
Vietnamese Restaurant	2	3	1	1	0

Conclusion

- This analysis is based on a small amount of data (since some venues are not found by the Foursquare API).
- There is room for improvement if one could collect more venues.
- Nevertheless five regions were found with a relatively small number of bars compared to the number of venues from other categories.