Verifying Berlin neighborhoods

Interests

- The Berlin government has a budget to financially support a boroughs social development

- To gain the most benefits it is important to identify boroughs having the most demand

- Therefore a data analysis and modelling is done

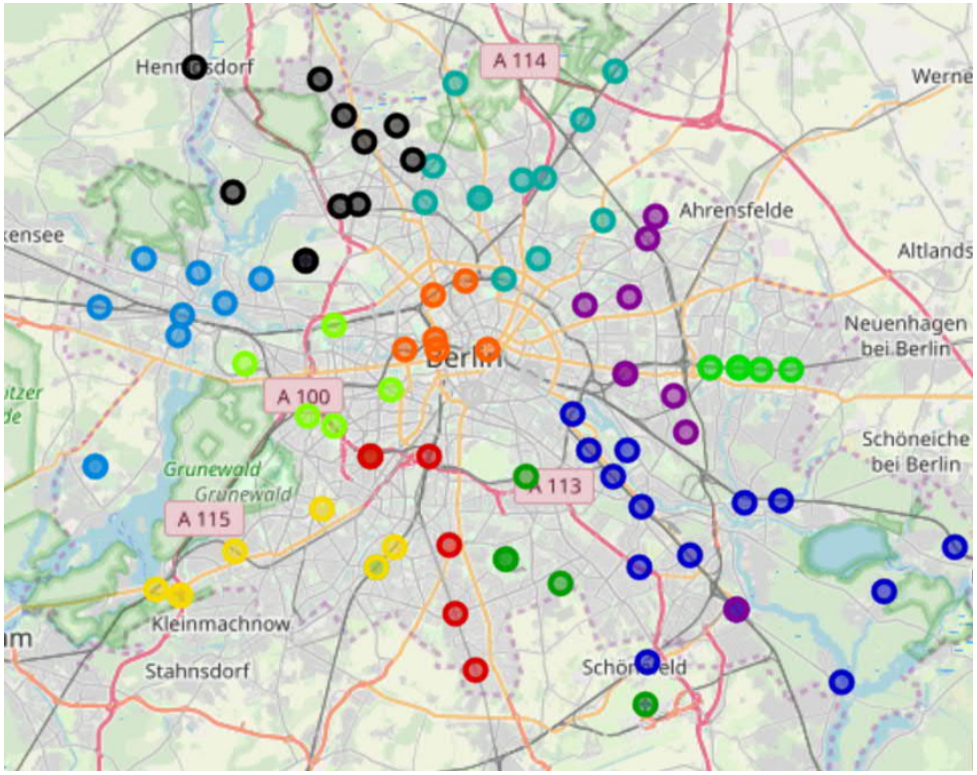
Data acquisition

- Wiki table including all boroughs and their neighborhoods:
https://de.wikipedia.org/wiki/Verwaltungsgliederung_Berlins

- Data set format
[number | neighborhood | borough | area km² | inhabitants | inhabitants per km

Nr.	Ortsteil	Bezirk	Fläche (km²)	Einwohner[2] (30. Juni 2019)	Einwohner pro km²
101	Mitte	Mitte	10,70	101.932	9526
102	Moabit	Mitte	7,72	79.512	10.299
103	Hansaviertel	Mitte	0,53	5.894	11.121
104	Tiergarten	Mitte	5,17	14.753	2854

Data acquisition



- Map of Berlin after adding longitude and latitude of the centroids by geolocator package

- Neighborhoods colored corresponding to their borough

Data acquisition

- The data retrieved by the wiki page can already be used as it is

- Only the parameter ‘Area (km^2)’ needs to be changed to float

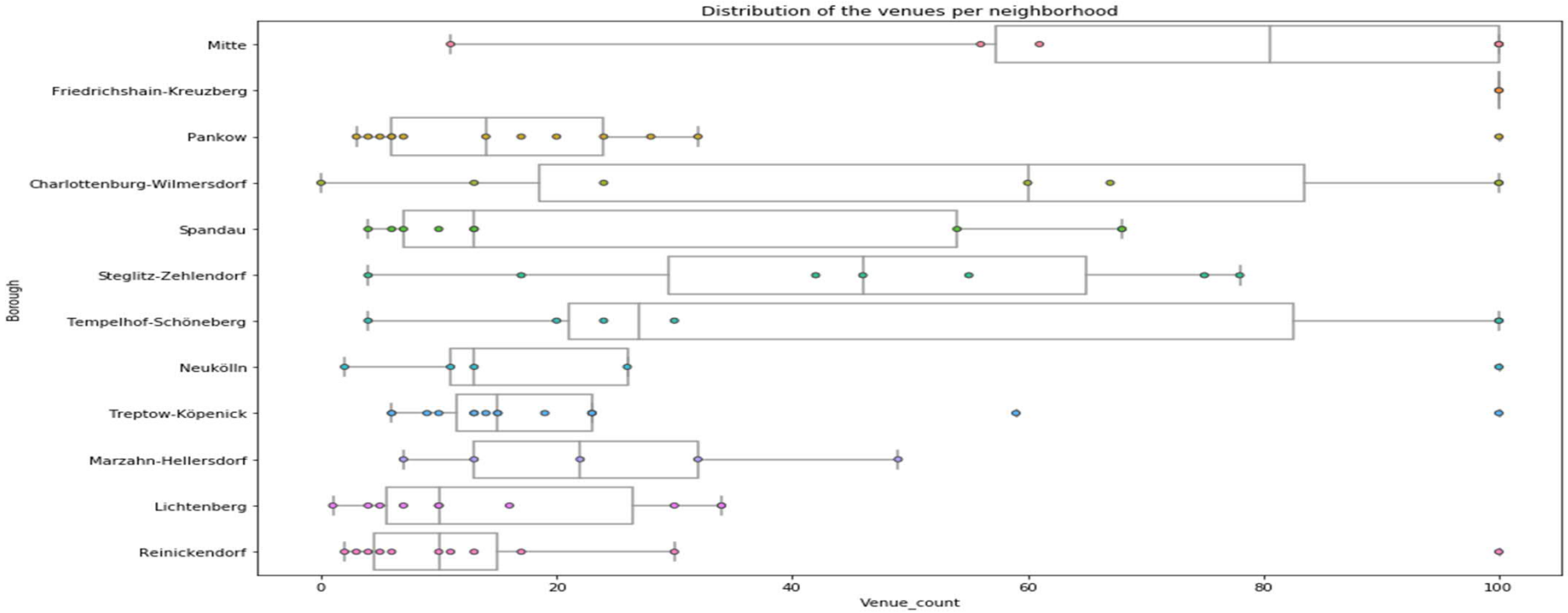
- A dynamic radius is determined depending on the area size of the neighborhood

Formula: Radius = MinMaxNormalization[‘Area (km^2)’] * mean([‘Area (km^2)’]) * 1000 / 2

- With this radius the Foursquare venues are called

Data exploration and analysis

- Reinickendorfs upper quartile is within 20 venues -> most neighborhoods have <20 venues



Data exploration and analysis

- 81.8% of Reinickendorfs neighborhoods have less than 20 venues

	Borough	Num_of_neighborhoods	Num_of_less_20_venues	Ratio: tot NH / <20 NH
7	Reinickendorf	11	9.0	0.818182
2	Lichtenberg	10	7.0	0.700000
11	Treptow-Köpenick	15	10.0	0.666667
8	Spandau	9	6.0	0.666667
6	Pankow	13	8.0	0.615385
5	Neukölln	5	3.0	0.600000
3	Marzahn-Hellersdorf	5	2.0	0.400000
0	Charlottenburg-Wilmersdorf	7	2.0	0.285714
9	Steglitz-Zehlendorf	7	2.0	0.285714
4	Mitte	6	1.0	0.166667
10	Tempelhof-Schöneberg	6	1.0	0.166667
1	Friedrichshain-Kreuzberg	2	NaN	NaN

Data exploration and analysis

- Reinickendorf has most neighborhoods compared to the total area size

	Borough	Num_of_neighborhoods	Num_of_less_20_venues	Ratio: tot NH / <20 NH	Area (km^2)	Ratio * <20 NH
0	Reinickendorf	11	9.0	0.818182	89.40	7.363636
2	Treptow-Köpenick	15	10.0	0.666667	165.70	6.666667
4	Pankow	13	8.0	0.615385	103.26	4.923077
1	Lichtenberg	10	7.0	0.700000	52.02	4.900000
3	Spandau	9	6.0	0.666667	91.90	4.000000
5	Neukölln	5	3.0	0.600000	44.91	1.800000
6	Marzahn-Hellersdorf	5	2.0	0.400000	61.71	0.800000
7	Charlottenburg-Wilmersdorf	7	2.0	0.285714	64.62	0.571429
8	Steglitz-Zehlendorf	7	2.0	0.285714	102.47	0.571429
9	Mitte	6	1.0	0.166667	39.48	0.166667
10	Tempelhof-Schöneberg	6	1.0	0.166667	53.08	0.166667
11	Friedrichshain-Kreuzberg	2	NaN	NaN	20.18	NaN

Cluster models

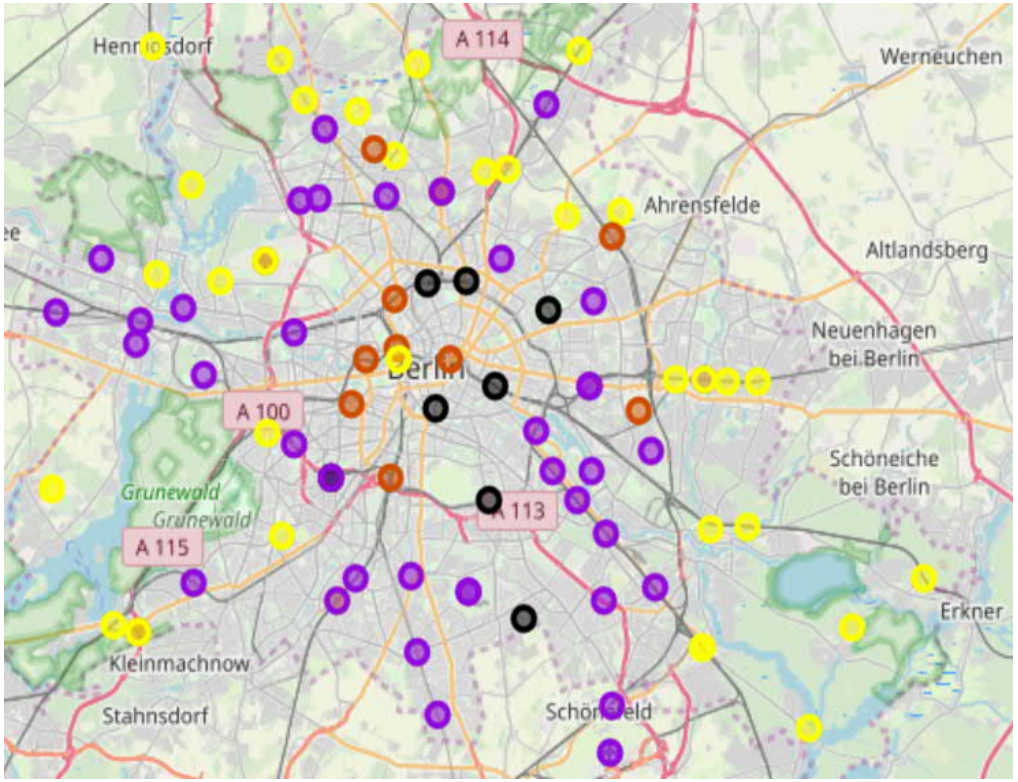
- Data preparation by one hot encoding of the venue categories

- Adding the inhabitants per area (km^2)

- Calculating the means

- The resulting data is used for kmeans modelling

Cluster models



- 4 different kmeans clusters are created:
Cluster 2 = 38 members
Cluster 0 = 36 members
Cluster 3 = 12 members
Cluster 1 = 9 members

- map illustrating the clusters and neighborhoods

- Cluster 1 and 3 are mostly centered and considered in the following

Cluster models

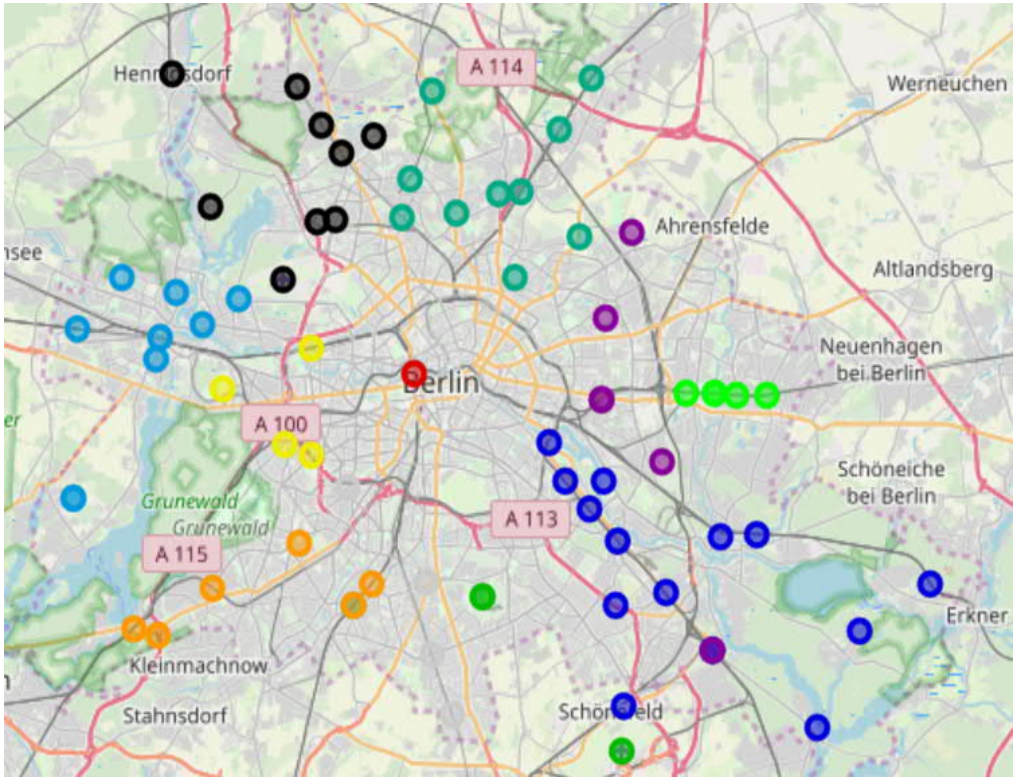
- Cluster 2 and 0 are defined as ‚support required‘

- Number of neighborhoods being in cluster 2 and 0

- All neighborhoods of Treptow – Köpenick, Marzahn – Hellersdorf and Spandau belong to cluster 2 and 0

	Borough	Cluster_Labels_Count	Num_of_neighborhoods	Ratio
0	Treptow-Köpenick	15	15	1.000000
3	Spandau	9	9	1.000000
6	Marzahn-Hellersdorf	5	5	1.000000
5	Steglitz-Zehlendorf	6	7	0.857143
1	Pankow	11	13	0.846154
2	Reinickendorf	9	11	0.818182
4	Lichtenberg	7	10	0.700000
8	Tempelhof-Schöneberg	4	6	0.666667
9	Neukölln	3	5	0.600000
7	Charlottenburg-Wilmersdorf	4	7	0.571429
10	Mitte	1	6	0.166667

Cluster models



- Map of neighborhoods being in cluster 2 and 0

- Treptow – Köpenick (dark blue) is mostly countryside area

- Spandau (light blue) is working area

- Marzahn – Hellersdorf (light green) is a living area and the neighborhoods are close together

Conclusion

- Data analysis shows that Reinickendorf needs support in their social development

- Cluster modeling shows that Treptow – Köpenick, Marzahn – Hellersdorf and Spandau have highest demand

- Marzahn – Hellersdorf will be in focus because people living their need to be satisfied instead of attracting new people to Treptow – Köpenick or Spandau

Conclusion – Final statement

- The neighborhoods of Marzahn – Hellersdorf are closed together

- So the financial support can probably be reached with lower financial resources

- Then also Reinickendorf can be supported additionally

The recommendation is to support Marzahn – Hellersdorf first and Reinickendorf further on