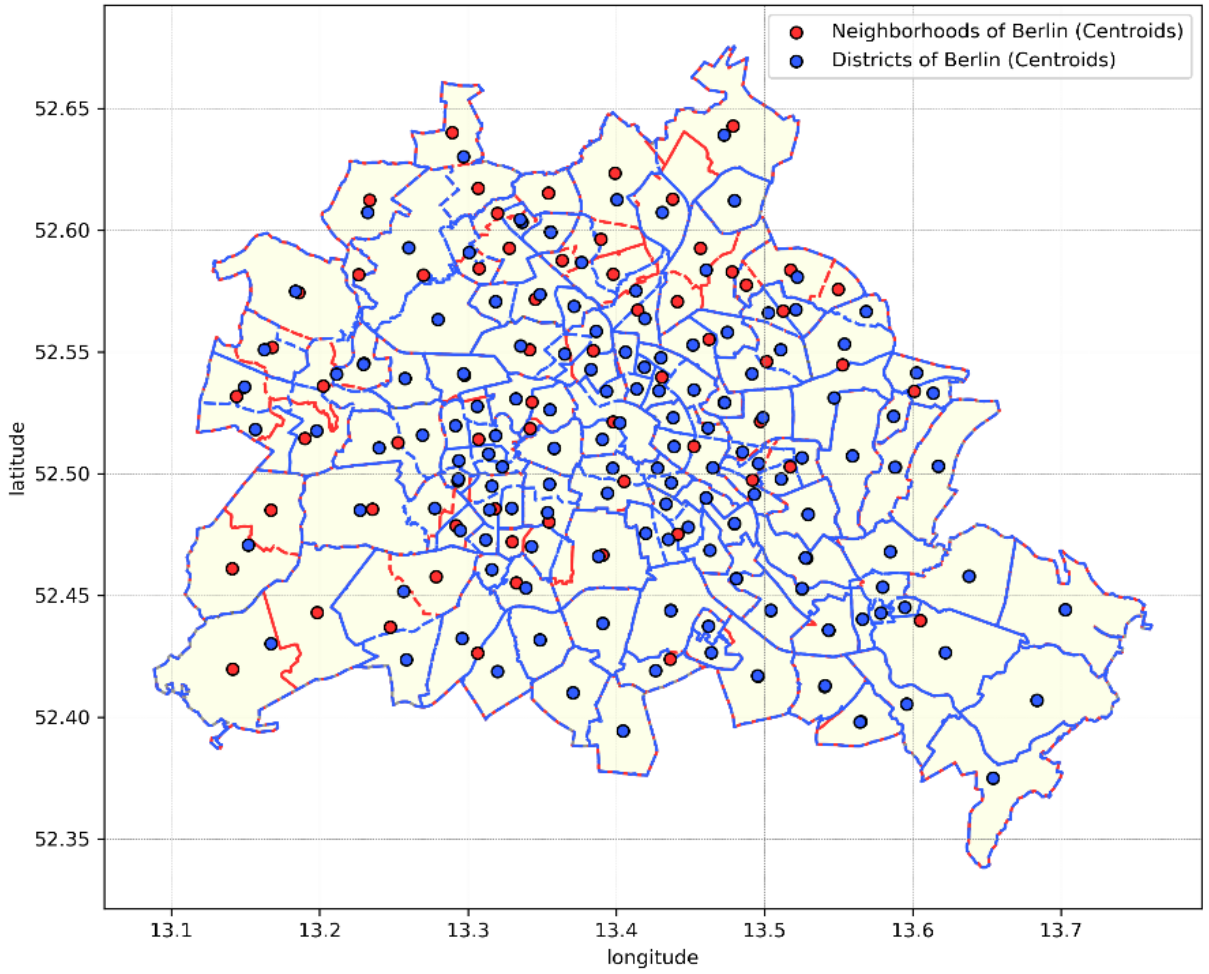
Berlin

- a city to live... but where?

Predicting valuable regions in Berlin for living

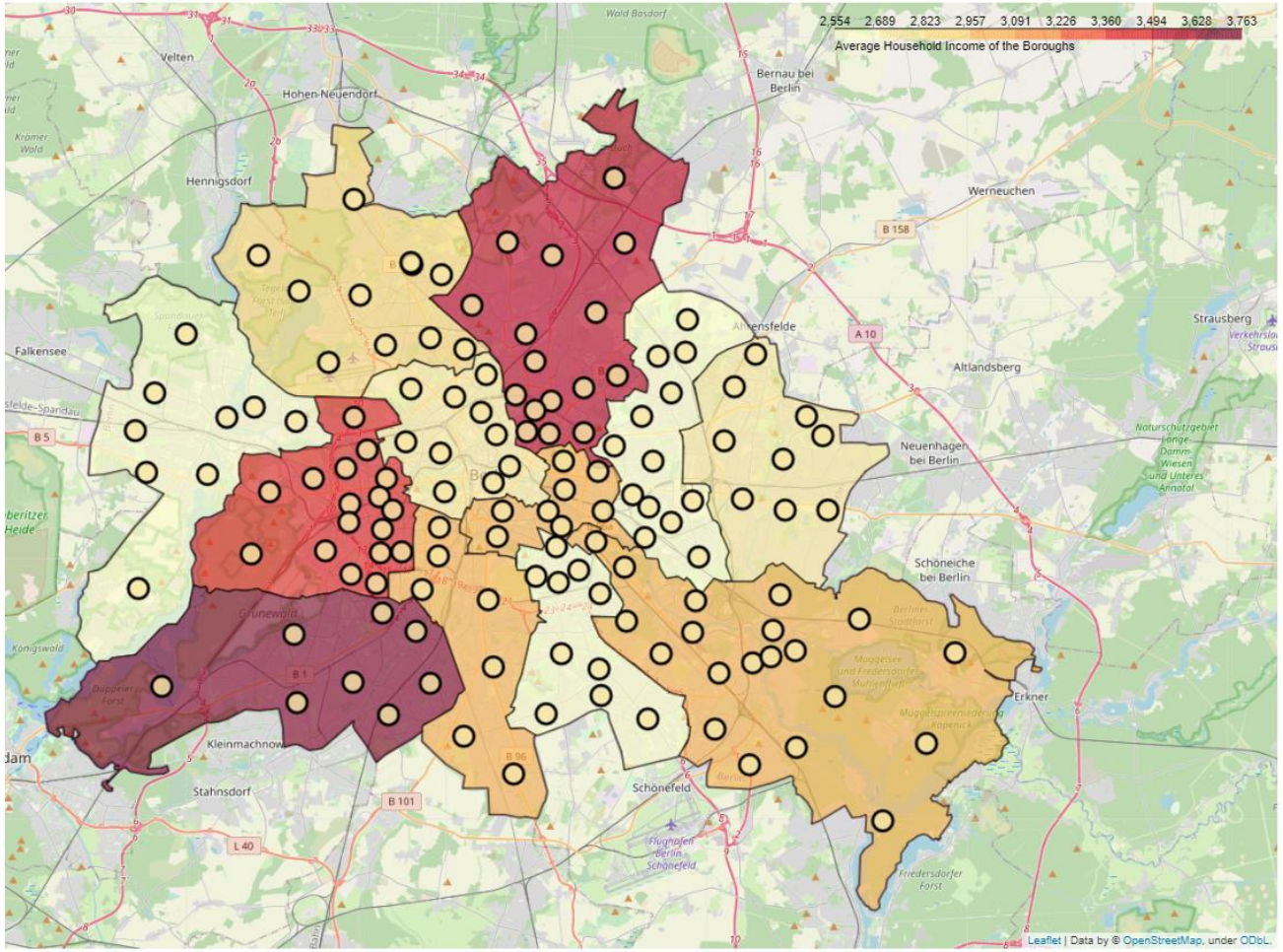
- Berlin as a city:
  - 3.7 million inhabitants
  - area of 892 km^2
- Capital city of Germany and the largest one
  - 12 boroughs
  - 96 neighborhoods
  - 138 district areas



Average Net-Household Income (Boroughs)

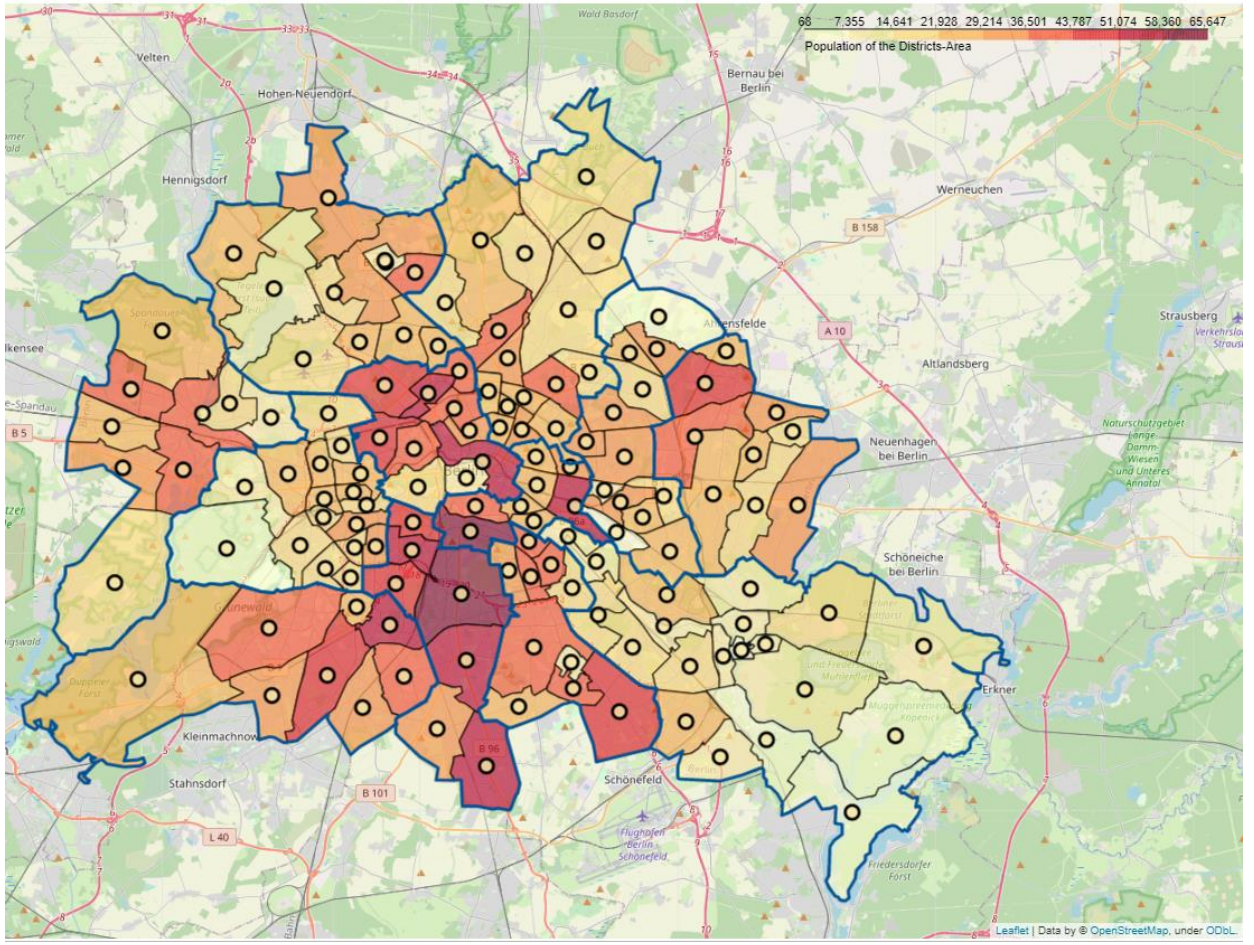
- Income-range of 2554 – 3763 €

- Top 3 boroughs
  - Steglitz-Zehlendorf (3763 €)
  - Pankow (3520 €)
  - Charlottenburg-Wilmersdorf (3484 €)
- Bottom 3 boroughs
  - Spandau (2687 €)
  - Lichtenberg (2640 €)
  - Neukölln (2554 €).



Population / Inhabitants of the Districts

- Boroughs with high amount
  - Tempelhof-Schöneberg (south)
  - Steglitz Zehlendorf (south-west)
  - Neukölln (south-east)
  - Pankow (north)
- Districts with high amount
  - Tempelhofer Vorstadt (65647)
  - Tempelhof (65094)
  - Alexanderplatz (57883)

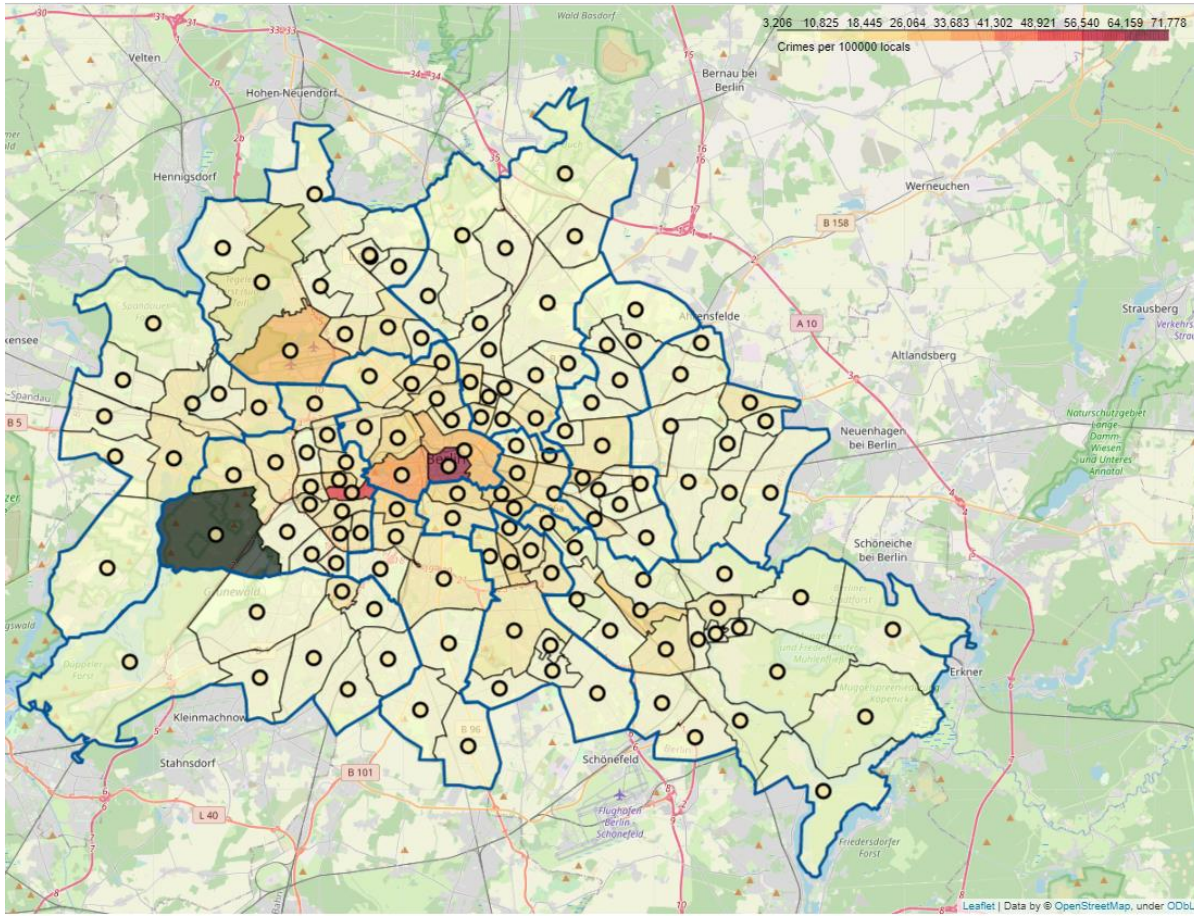


Crime Index of the Districts

$$HZ = \frac{Crimes}{Population} * 10^5$$

(Crimes per 100000 inhabitants)

- The highest crime index is reached at the districts
  - Regierungsviertel (HZ: 71778)
  - Kurfürstendamm (HZ: 52312)
  - Tiergarten Süd (HZ: 34507)
  - West 1 – Tegel-Süd (HZ: 31293) (airport)



1) First Selection of possible regions to live

- Method: Normalisation of the crime index and population
  - Min-Max Normalisation in the 0 to 1 range
- Define crime and population factors based on the first assumption
- Take all the data which fits both factors
  - All data with a crime factor < 0.4 (take the lowest crime regions)
  - All data with a population > 0.6 (take the highest population regions)

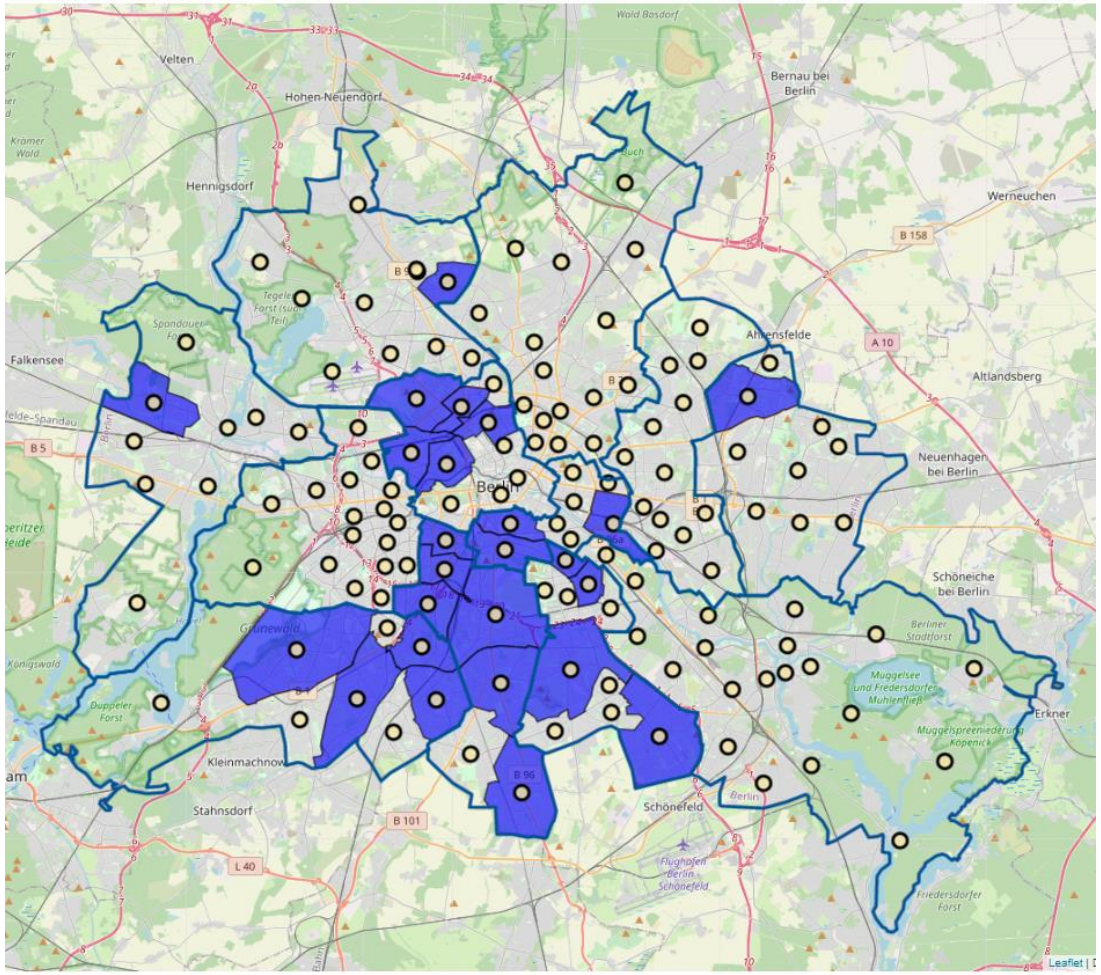
```
crime_pop_reduced = gdf_District_crime_pop_District_Area
crime_factor=0.4
population_factor=0.6
crime_pop_reduced = crime_pop_reduced[(crime_pop_reduced['normCRIME_per_100000_POP']<=crime_factor) & (crime_pop_reduced['normPOPULATION']>=population_factor)]
print("{} Districts are matching a lower crime factor of {} and a higher population factor of {} ".format(crime_pop_reduced.shape[0],crime_factor,population_factor))
crime_pop_reduced.sort_values(by="CRIME_per_100000_POP").head(3)
```

```
25 Districts are matching a lower crime factor of 0.4 and a higher population factor of 0.6
```

1) First Selection of possible regions to live

- Moabit West
- Moabit Ost
- Brunnenstr. Nord
- Parkviertel
- Wedding Zentrum
- Südliche Friedrichstadt
- Tempelhofer Vorstadt
- Frankfurter Allee Süd FK
- Falkenhagener Feld
- Albrechtstrasse
- Lankwitz
- Drakestrasse

- Zehlendorf Nord
- Schöneberg Nord
- Schöneberg Süd
- Friedenau
- Tempelhof
- Mariendorf
- Lichtenrade
- Reuterstrasse
- Rixdorf
- Britz
- Rudow
- Marzahn Mitte
- MV 1 - Märkisches Viertel



2) Foursquare API for Venues

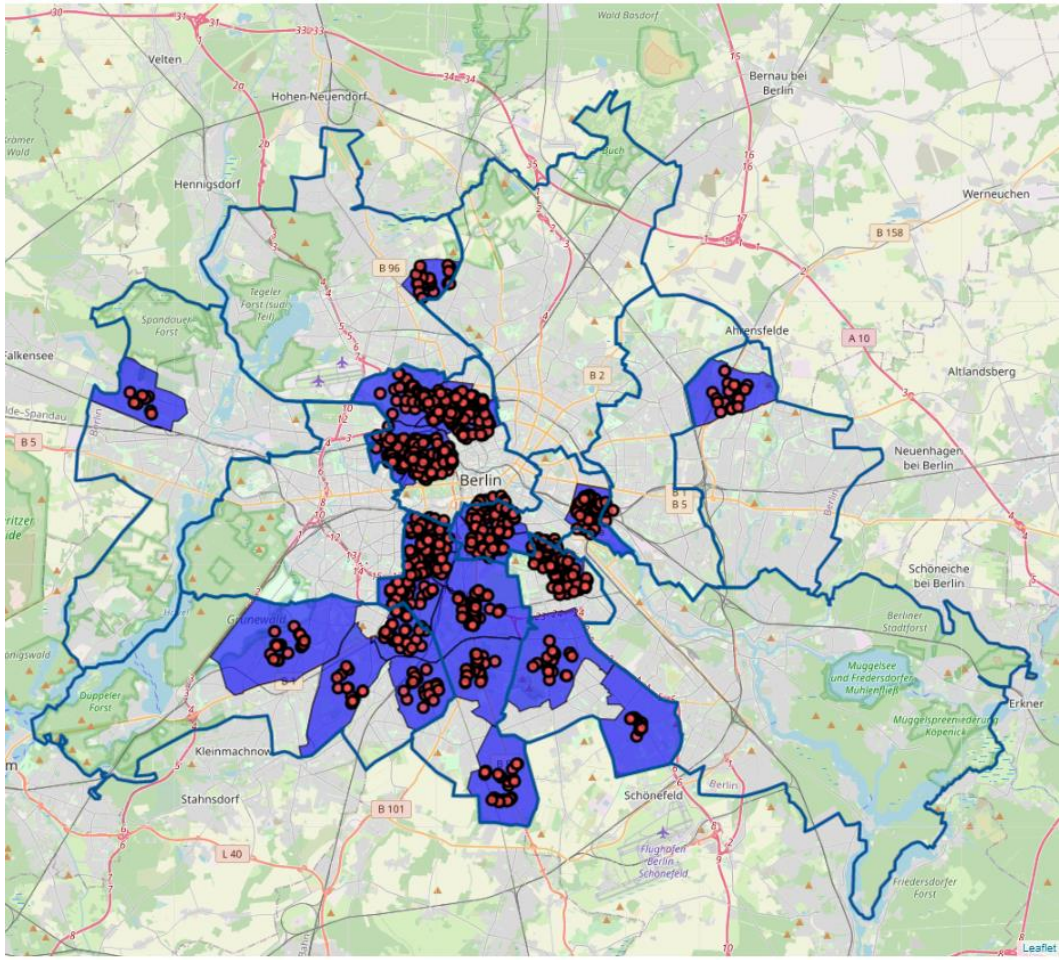
- Get venues via the Foursquare API
  - Radius 1000 m
  - Limit 100

```
Berlin_venues = getVenues(lat_lon["Name"],lat_lon["lat"],lat_lon["lon"],1000,100)
print('{} venues were returned by Foursquare.'.format(Berlin_venues.shape[0]))
Replacements = ["Venue"]
replace_characters(Berlin_venues,Replacements)
```

```
1497 venues were returned by Foursquare.
```

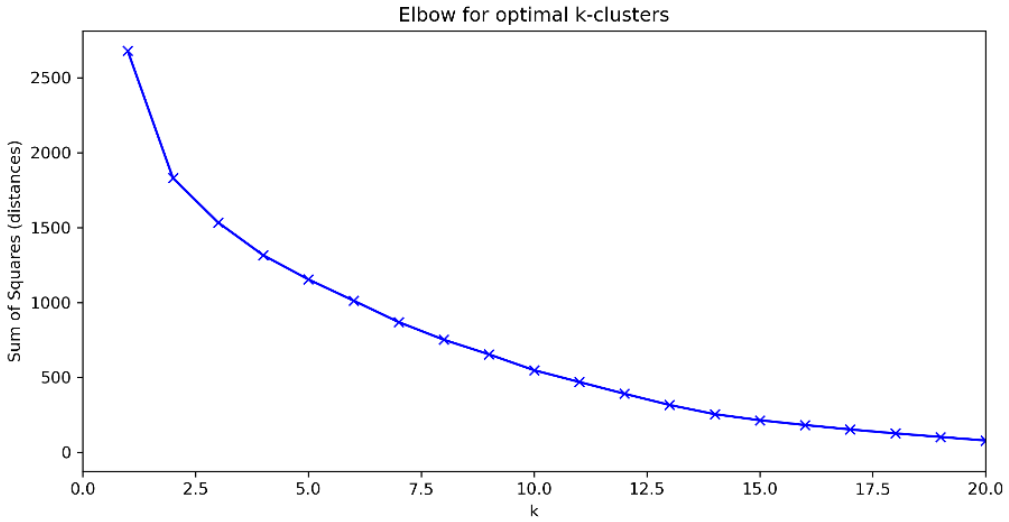
```
Berlin_venues.head(5)
```

	DISTRICT_AREA	DISTRICT_Latitude	DISTRICT_Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Moabit West	52.530806	13.332154	Einer dieser Tage	52.532388	13.331733	Café
1	Moabit West	52.530806	13.332154	Garcia Kaffeebar	52.527668	13.330624	Coffee Shop
2	Moabit West	52.530806	13.332154	Kallasch &	52.532869	13.336736	Bar
3	Moabit West	52.530806	13.332154	Zentrum fuer Kunst und Urbanistik (ZK-U)	52.533659	13.336372	Art Gallery
4	Moabit West	52.530806	13.332154	Hamberger	52.533230	13.332539	Market

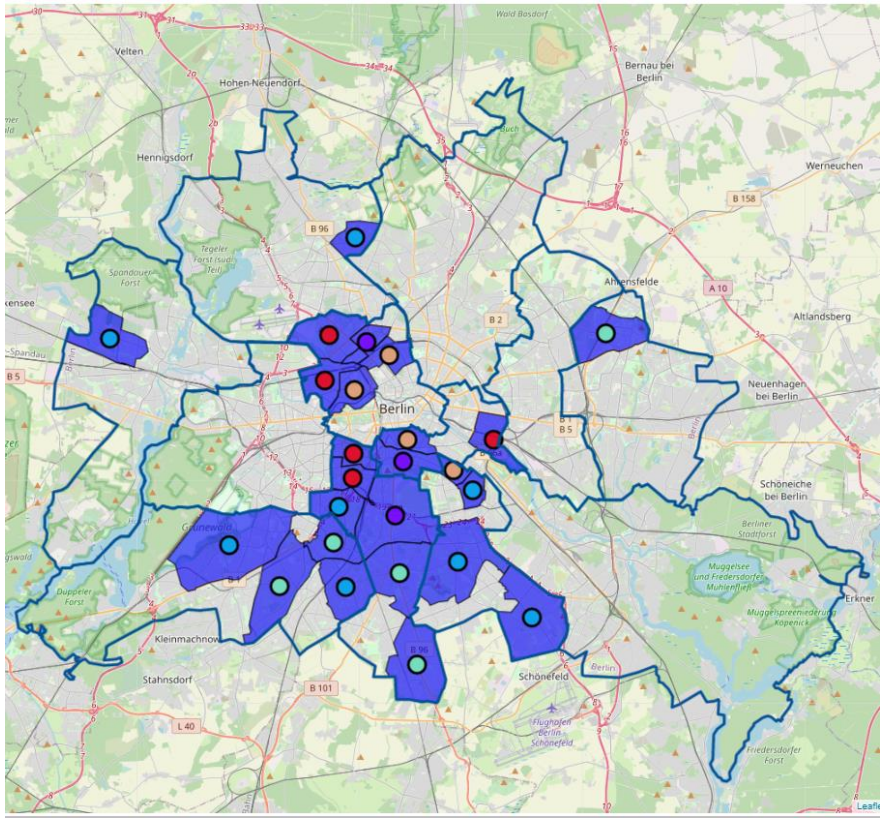


3) k-means Clustering

- K-means clustering of the venue data
  - *one-hot-encoding* of the data first
  - Elbow plot for optimal k-value
  - Index clusters to districts

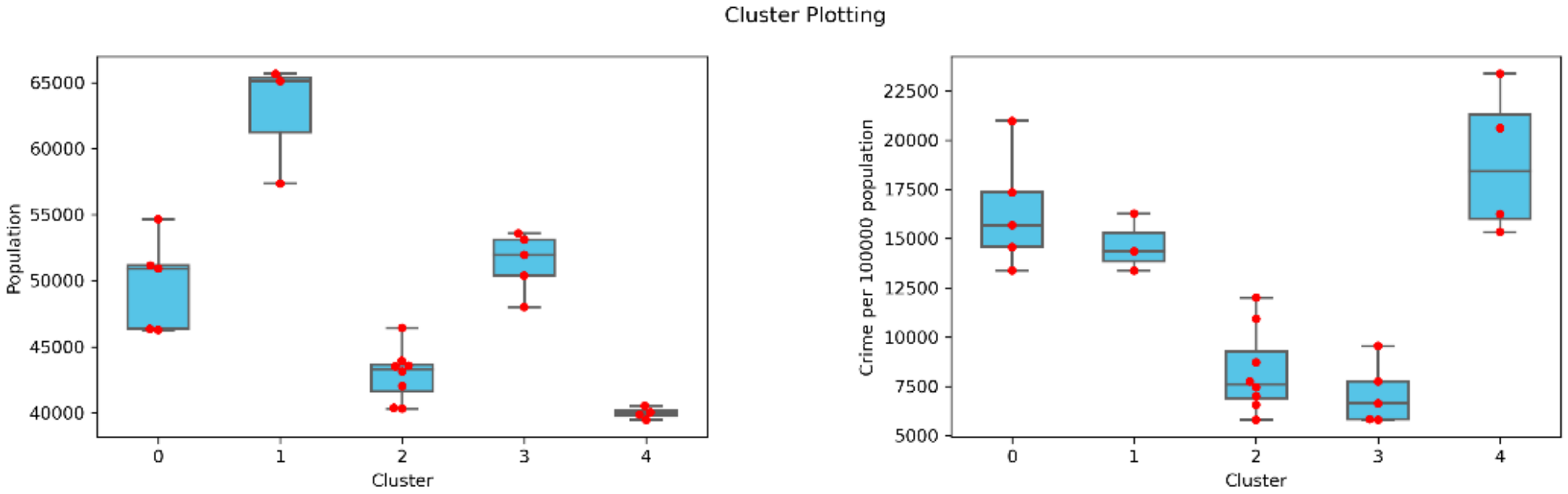


	Cluster_Labels	DISTRICT_AREA	POPULATION	CRIME_per_100000_POP	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	3	Albrechtstrasse	53582	5792.990183	Supermarket	Café	Chinese Restaurant
1	2	Britz	43514	10922.921359	Supermarket	Bakery	Liquor Store
2	4	Brunnenstr. Nord	39469	16248.194786	Bakery	Supermarket	Vietnamese Restaurant
3	3	Drakestrasse	50385	6628.957031	Supermarket	Bakery	Bus Stop
4	2	Falkenhagener Feld	42031	6552.306631	Liquor Store	Lake	Bus Stop



4) Check Crime and Population (Cluster)

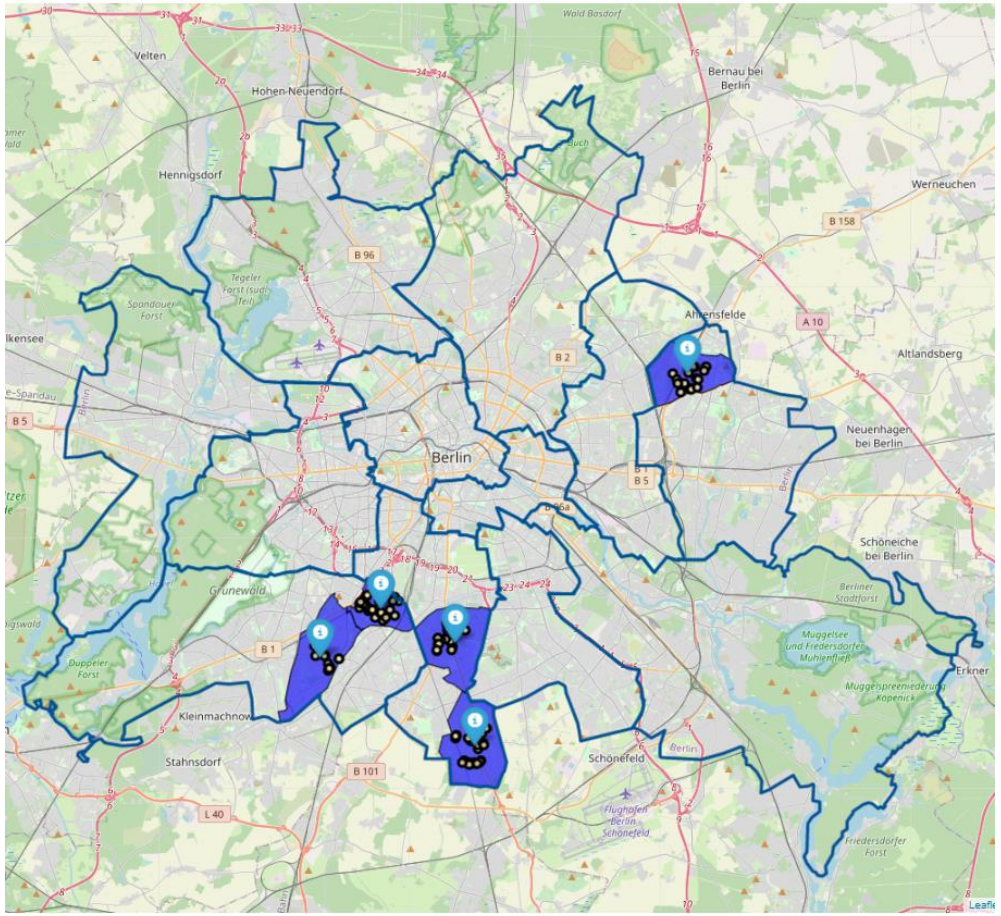
- Via boxplots (Crime and population against cluster assignment) the crime and population values can be assigned
- Check which cluster has the lowest crime index ➡ cluster 3



5) Cluster with lowest Crime Index

- Via the boxplots extract all districts with the lowest crime rate

	Cluster_Labels	DISTRICT_AREA	POPULATION	CRIME_per_100000_POP	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	3	Albrechtstrasse	53582	5792.990183	Supermarket	Café	Chinese Restaurant
3	3	Drakestrasse	50385	6628.957031	Supermarket	Bakery	Bus Stop
8	3	Lichtenrade	51955	5824.271004	Supermarket	Soccer Field	Bakery
10	3	Mariendorf	53106	7739.238504	Supermarket	Bank	German Restaurant
11	3	Marzahn Mitte	48008	9550.491585	Tram Station	Supermarket	Cultural Center



6) Sum of Venues

- In order to decide which region of the cluster is the most liveable one the venues are summed
  - Low crime index ➡ safe surrounding
  - High amount of venues ➡ most possibilities for action

```
The top three relevant district areas are Albrechtstrasse, Lichtenrade and Mariendorf
```

	Cluster_Labels	DISTRICT_AREA	POPULATION	CRIME_per_100000_POP	SUM_VENUES	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	3	Albrechtstrasse	53582	5792.990183	31	Supermarket	Café	Chinese Restaurant	Park	Trattoria/Osteria
1	3	Lichtenrade	51955	5824.271004	28	Supermarket	Soccer Field	Bakery	Doner Restaurant	Light Rail Station
2	3	Mariendorf	53106	7739.238504	28	Supermarket	Bank	German Restaurant	Pool	Bakery