Explore and analyze rents and venues regarding relocating and house-hunting in Berlin, Germany

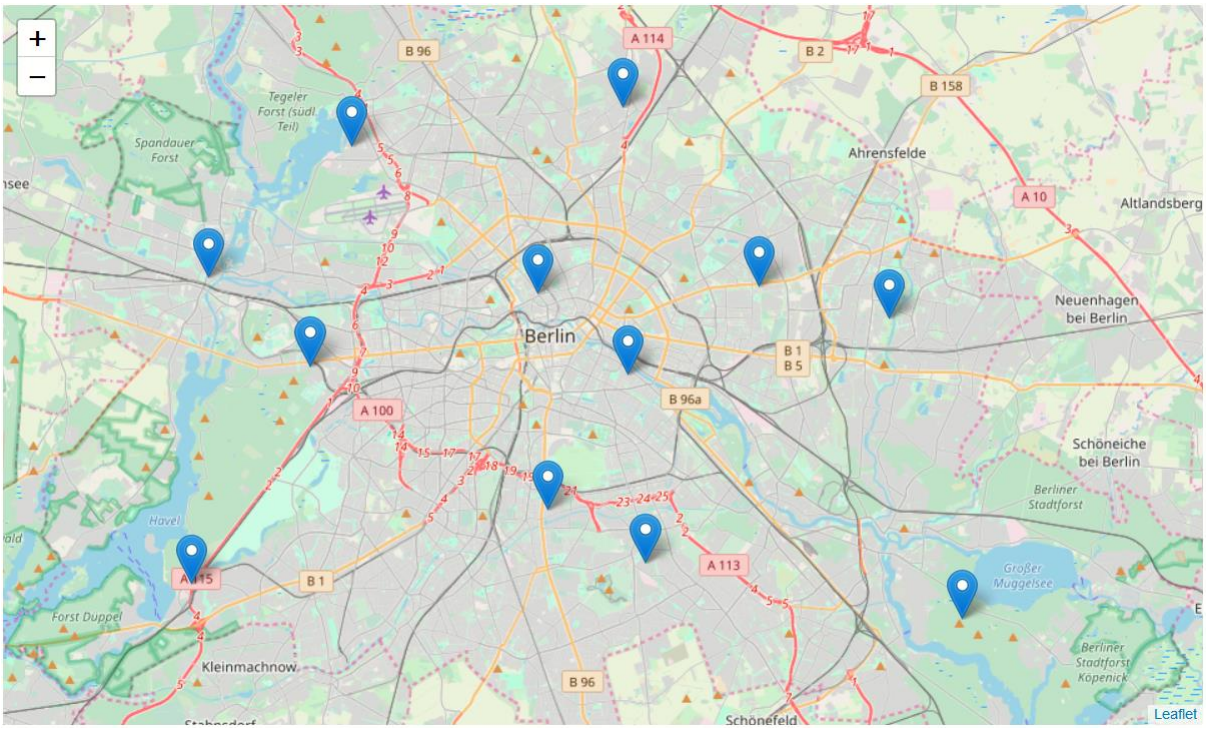
Srdan Suka - Capstone Project - The Battle of Neighborhoods (Week 2)

1. INTRODUCTIONAND BUSINESS PROBLEM DESCRIPTION

a. Although not an easy decision, nowadays, relocation has never been easier due to help of Internet, various services available online that minimize time for researching and pin pointing perfect location and as per budget affordable place to settle. With data science and available frameworks, libraries, models for fast analysis of city areas, narrowing search as per requirements and desires, people can shorten and prioritize the list of neighbourhoods of interest and find place to live within desired budget.

b. To illustrate the problem, I have been living in Zagreb for 20 years of my life, currently working remotely for Berlin based company. To be able to advance in my career, for me it would be necessary to relocate to Berlin, Germany. The question I need to ask myself is where exactly I would like to live in that big city.

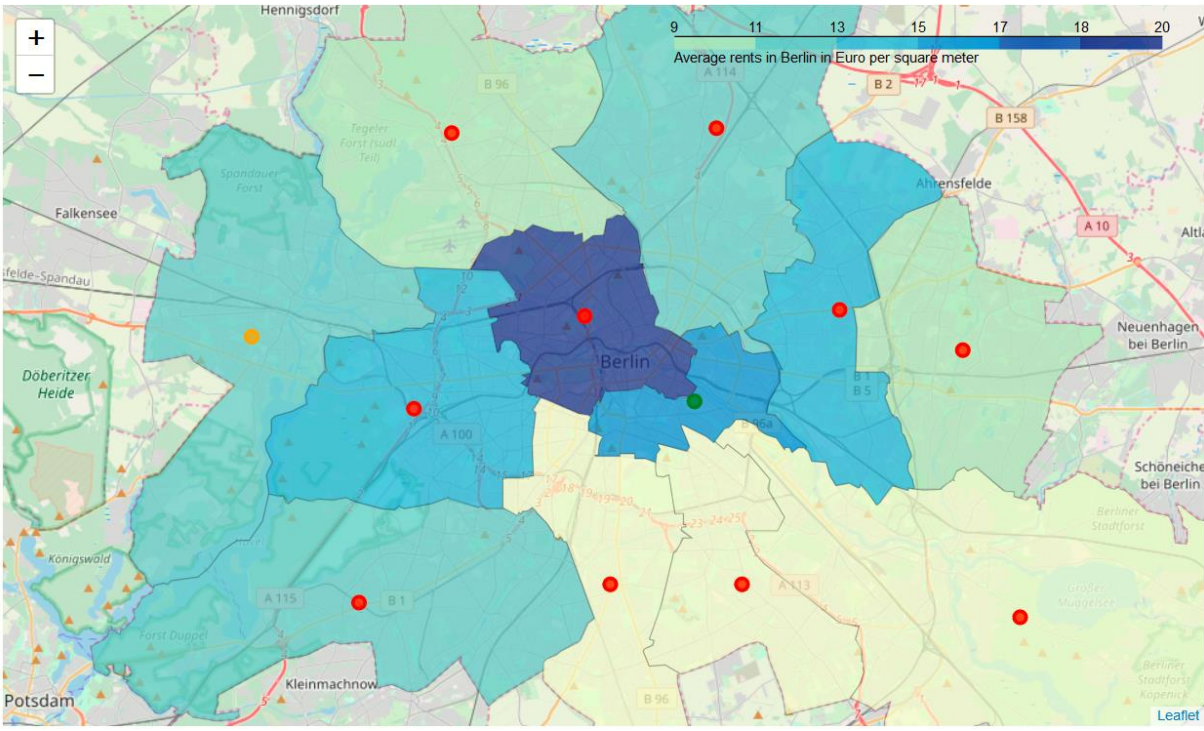
c. Berlin is the capital and largest city of Germany by both area and population. More than 3.5 million people are spread across 12 boroughs. Compared to the rest of Germany, the rents for flats are extremely high with trending upwards. Since Berlin is also the largest city by area we will focus to place near to center with higher social places density.



	Borough	Latitude	Longitude
0	Mitte	52.530880	13.381918
1	Friedrichshain-Kreuzberg	52.503703	13.431762
2	Pankow	52.592940	13.428462
3	Charlottenburg-Wilmersdorf	52.506477	13.257127
4	Spandau	52.536155	13.201278
5	Steglitz-Zehlendorf	52.434014	13.191542
6	Tempelhof-Schöneberg	52.458359	13.387560
7	Neukölln	52.441068	13.441185
8	Treptow-Köpenick	52.421899	13.614815
9	Marzahn-Hellersdorf	52.522653	13.575299
10	Lichtenberg	52.533323	13.503729
11	Reinickendorf	52.579943	13.279625

d. Since I like social places, restaurants, caffees, cinemas, theaters to enjoy new city I would like to find a place, my new neighbourhood, a home in my new city to live in. In order to determine the similarity, I will need to somehow each neighbourhood as a numerical vector then apply some machine learning technique, e.g., DBSCAN, KMeans, to cluster them into different groups.

e. Price index of rentals will have a major role in my decision and with desire to go as closer to center as possible we can see on the map price will go higher



f. For me, to be happy new "Berliner", I will need to tackle both problems, the rents and social places density for the city of Berlin. I will create a map where flat rent index is placed on Berlin and each neighborhood is clustered according to the venue density. Since square meter prices depend on at sizes, I will examine medium (60m2) and big size (100m2) flats. The map will be helpful for me making decisions in the future.

g. **Main goal of this project is to analyse all boroughs for social events and venues and compare the best ones fitting to my requirements vs rental prices.**

2. DataSETs and DATA Scraping

a. There are 12 neighborhoods in Berlin. We'll use label `Boroughs` to identify them.

b. In my initial dataframe we'll download geojson file with Polygon, latitude and longitude coordinates data in geojson format from GitHub repo https://data.technologiestiftung-berlin.de/data/bezirksgrenzen/bezirksgrenzen.geojson shape file.

c. We will also use borough names from the same file.

d. Data is simplified version of `https://data.technologiestiftung-berlin.de/data/bezirksgrenzen/bezirksgrenzen.geojson` and is uploaded to my GitHub repo https://raw.githubusercontent.com/srdansuka/BerlinBoroughs/master/berlin-bezirksgrenzen.geojson

e. In order address the mentioned problems, I will use the datasets/ sources listed below. A km2 shape link https://www.statistik-berlin-

brandenburg.de/regionales/kleinraumdaten/geometrienOD.asp?Kat=6301 containing a number of administrative areas in Berlin is provided by the statistical office of Berlin and Brandenburg. However, instead of transforming the shape file into GeoJSON format and extract-ing boroughs and boundaries, I will use the my GitHub repository https://raw.githubusercontent.com/srdansuka/BerlinBoroughs/master/berlin-bezirksgrenzen.geojson which has a GeoJSON file containing names and boundaries of Berlin's 12 boroughs. Once borough details are extracted, I will use centroids to obtain center coordinates (latitude / longitude) of each borough. Then, to obtain the most common venues in each bough, I will use Forsquare API. Fortunately, **FourSquare** offers free APIs for developers to access their database of venues. Each venue in their dataset is usually categorized into a venue category, which is described in their Developers Docs.

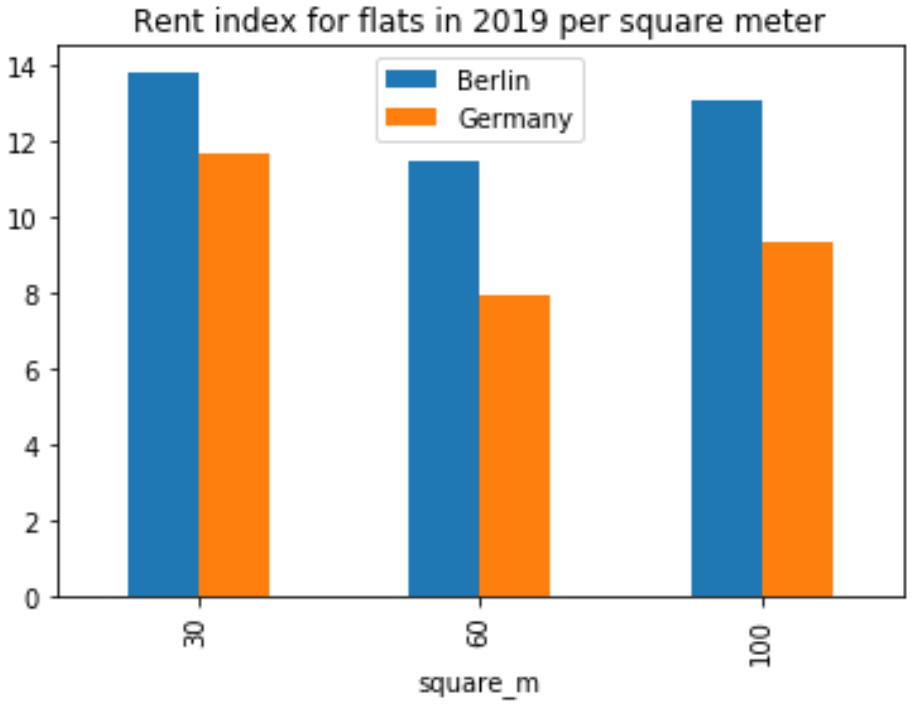
f. To obtain the details on flat rentals to get prices for July '19 listed and compared with historical data, I will scrape a platform for renting and buying flats named http://wohnungsboerse.net . Such data will enable me to create detailed map that:

g. displays neighbourhoods of Berlin in form of markers and clustered according to venue density

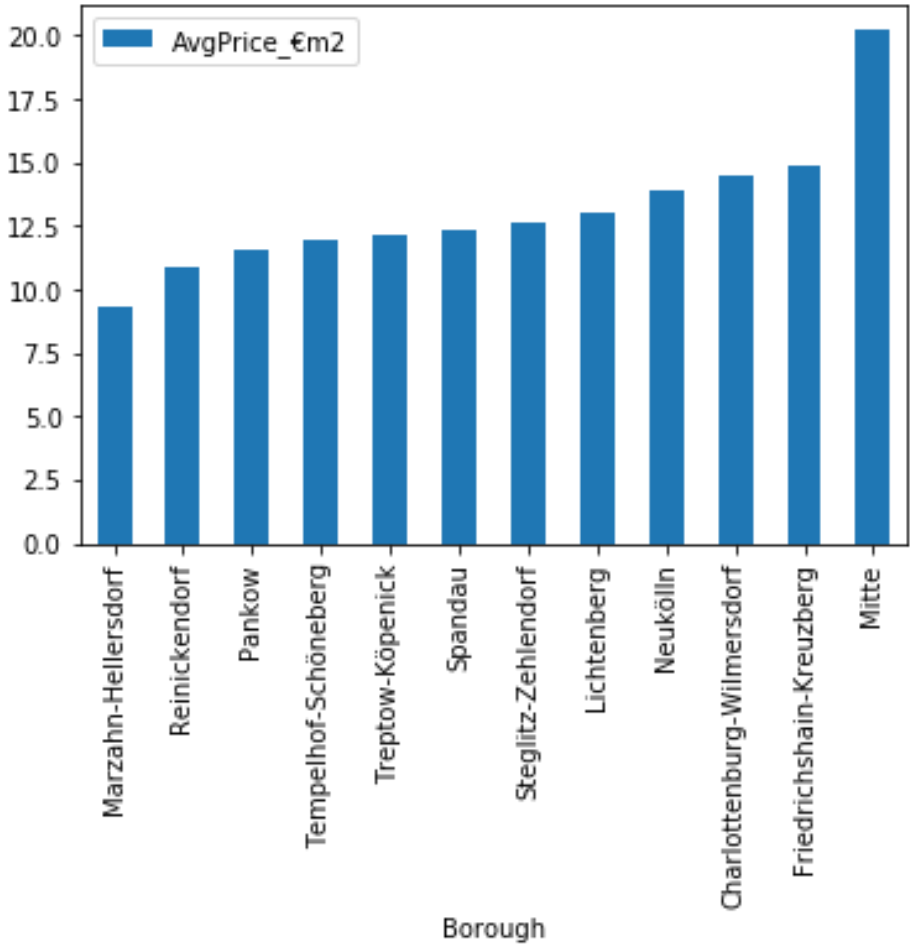
h. visualizes differences in rents

3. Renting FLAT IN BERLIN

a. Berlin, compared to the rest of Germany, is very expensive



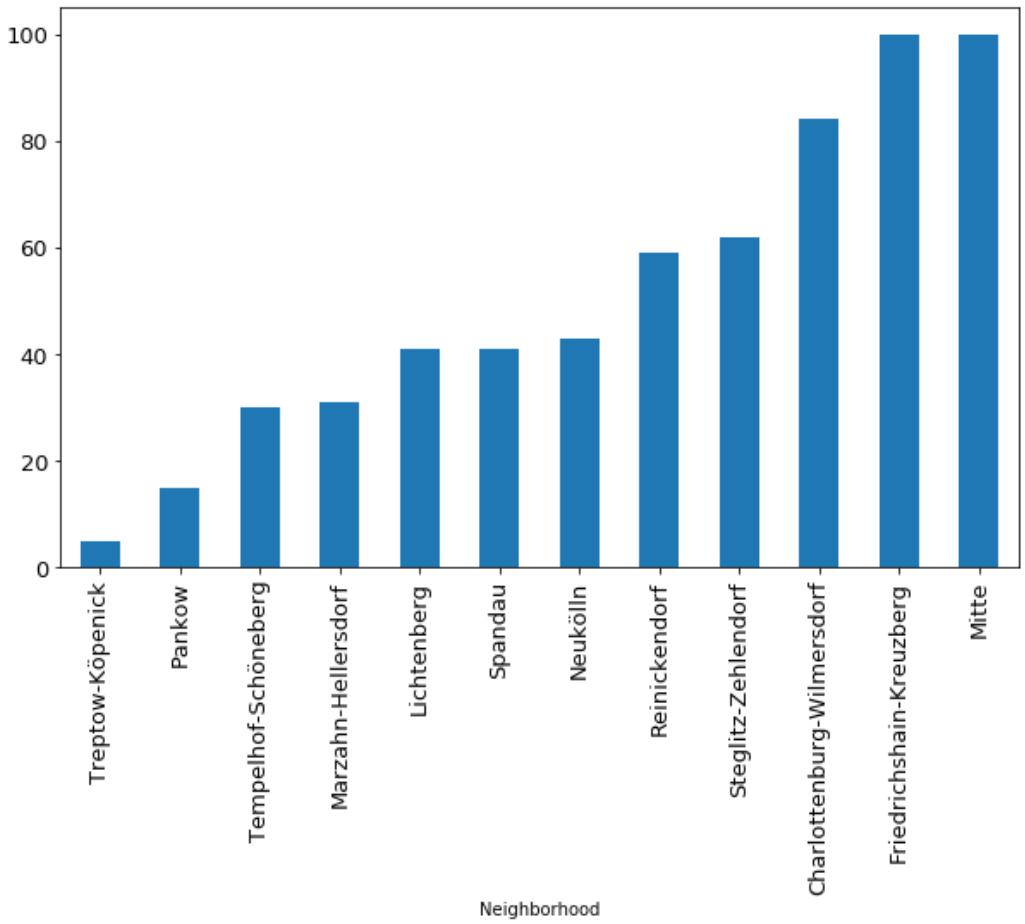
b. But with 12 boroughs, price goes from < €10 to > €20



4. BERLIN SOCIAL EVENTS PER BOROUGS

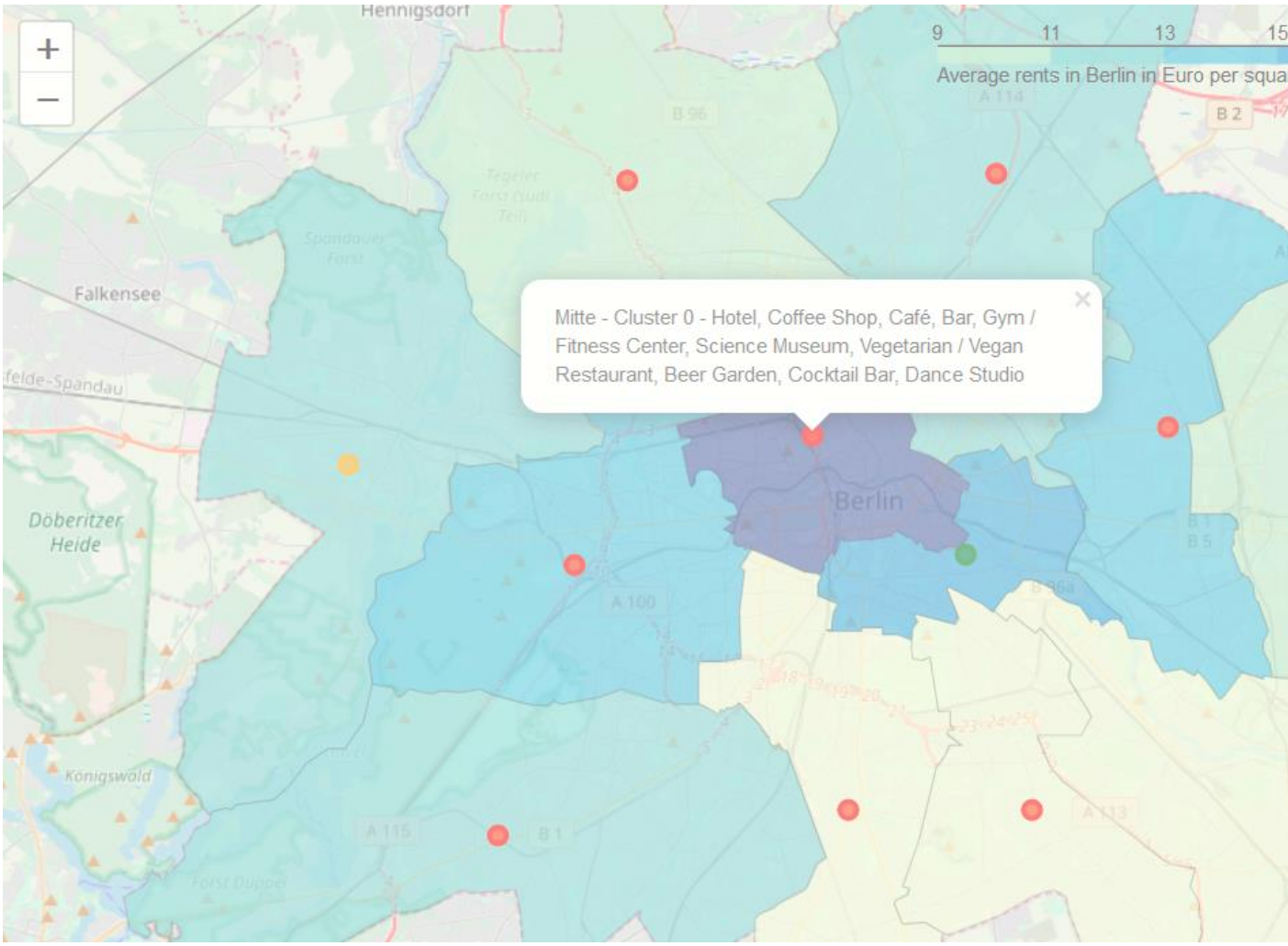
a. Boroughs closer to Mitte (Center) have more social events but are with higher rental prices

b. Borourgs on the edge are with less venues though but are with cheaper rentals



c. We have clustered boroughs in 3 Clusters
d. Forsquare API is used to scrape venues in Berlin boroughs
e. Data is displayed on map with 3 clusters described upon popup
    i. **Cluster 0** Hotel, Coffee Shop, Café, Bar, Gym / Fitness Center, Science Museum, Vegetarian / Vegan Restaurant, Beer Garden, Cocktail Bar, Dance Studio
    ii. **Cluster 1** German Restaurant, Café, Bar, Coffee Shop, Italian Restaurant, Cocktail Bar, Bakery, Wine Bar, Beer Bar, Ice Cream Shop
    iii. **Cluster 2** Supermarket, Pizza Place, Bus Stop, Restaurant, Hardware Store, Drugstore, Park, Clothing Store, Chinese Restaurant, Café

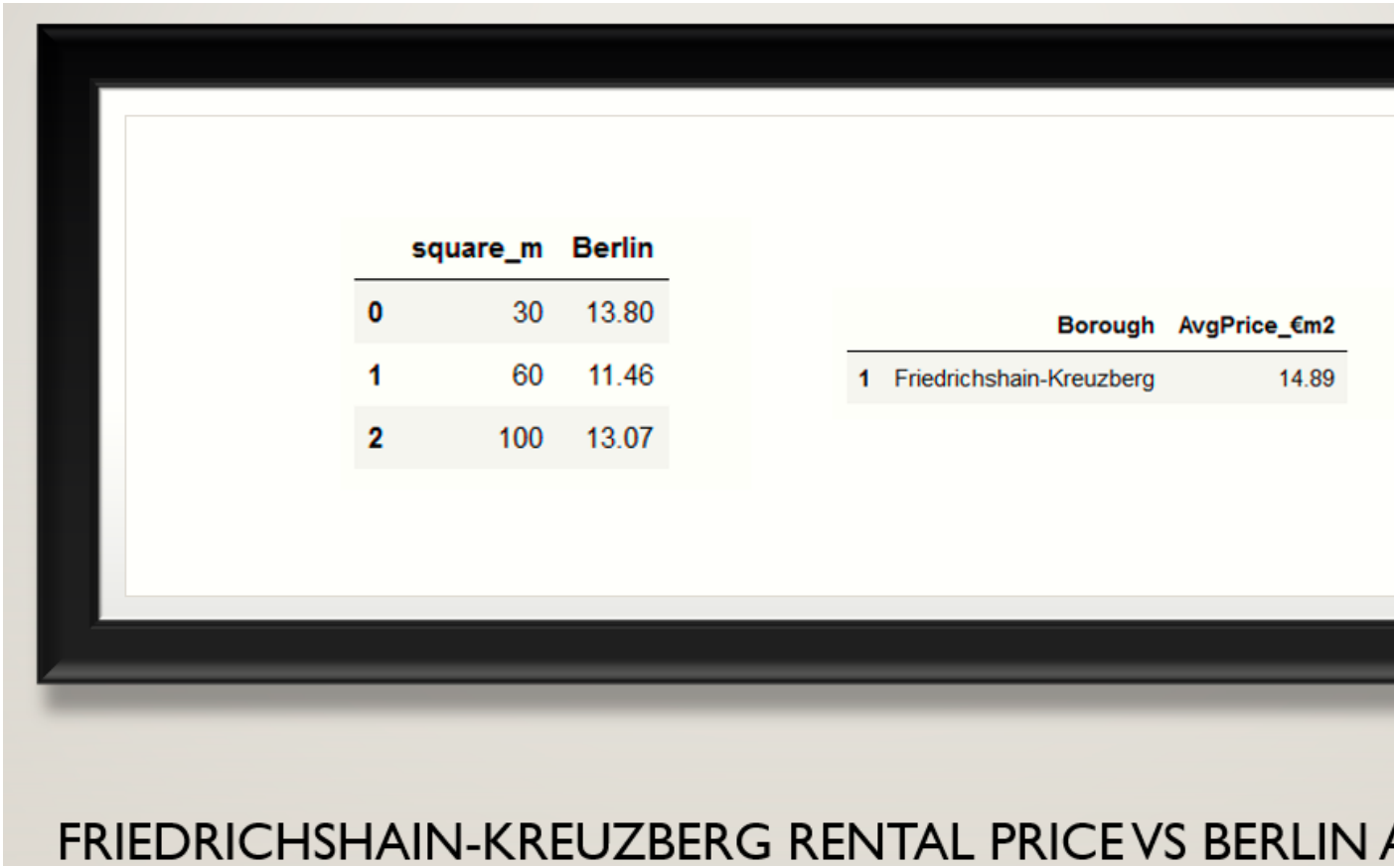
Clustered venues



5. CONCLUSION

- With prices close to Berlin average, close to center and with vibrant social events and venues for younger people, *Friedrichshain-Kreuzberg* seems to be the best candidate to rent a flat.

	ID	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Mos Commo Venu
1	02	Friedrichshain-Kreuzberg	52.503811	13.428234	1	German Restaurant	Café	Bar	Coffee Shop	Italia Restaura



	square_m	Berlin
0	30	13.80
1	60	11.46
2	100	13.07

	Borough	AvgPrice_€m2
1	Friedrichshain-Kreuzberg	14.89

6. FINAL WORD

- Berlin is very vibrant city with lot of opportunities for young people, either interested in culture or technology.
- Prices of rentals are rising so who ever is considering to relocate, beter now then later since prices are rising almost linear.
- Berlin is rich with venues, social events and places for everyone and it’s possible to find place to live close to city center for avg rental price.