Berlin Apartments and Venues

Georgios Papadopoulos

20.04.2020

1. The Idea

It has been three years since I moved to Berlin and I find it a fascinating and full of opportunities city. As David Bowie has said in the past Berlin is a cultural extravaganza.

A new startup is born every 20 minutes in Berlin, according to German business advisory firm, Gruenden. Berlin’s low rents, cool image and open attitude make the city an attractive hub for innovative minds.

Along with the limitless possibilities for people of art, there is also a huge tech scene which attracts investors, startups and people from all industries in order to settle and work. The difficulty appears when someone tries to find a place to live. It usually comes to a struggle to find an apartment and that can be a problem for people who are looking for a fast accommodation.

That being said an easy way to decide for place to live is the neighborhood and what is nearby.

The idea of this project is to sum up results from Immobilien Scout up to date and combine them with the Foursquare API venues requests in order for someone to better understand the neighborhood of each apartment.

Second step of this research is to cluster the neighborhoods of Berlin based on the venues and sort them based on the price per square meter. This is a value that is constantly changing.

This idea is addressing all the people that trying to find a place in this upcoming tech hub of Europe.

2. The Data

The data comes from two sources.

- **Web Scraping** from *Immobilien Scout* the most popular accommodation search site in Germany.

- **Foursquare API** requests for the venues near the specified location.

3. Methodology

Part 1

For the web scraping python library Beautiful Soup was used.

For the coordinates of the apartments python library geopy was used.

The plan was to combine the coordinates of each apartment with the Foursquare API in order to get what kind of venues are nearby and analyze the apartment surrounding as well as the respective neighborhoods.

- Create a function to scrape up to date data from Immobilien Scout

def immobilien_scraper(n, r1, r2)

User input:
n: number of pages to scrape
r1: number of rooms starting
r2: number of rooms ending

- Create a function to get the venues near the apartments queried from the site.

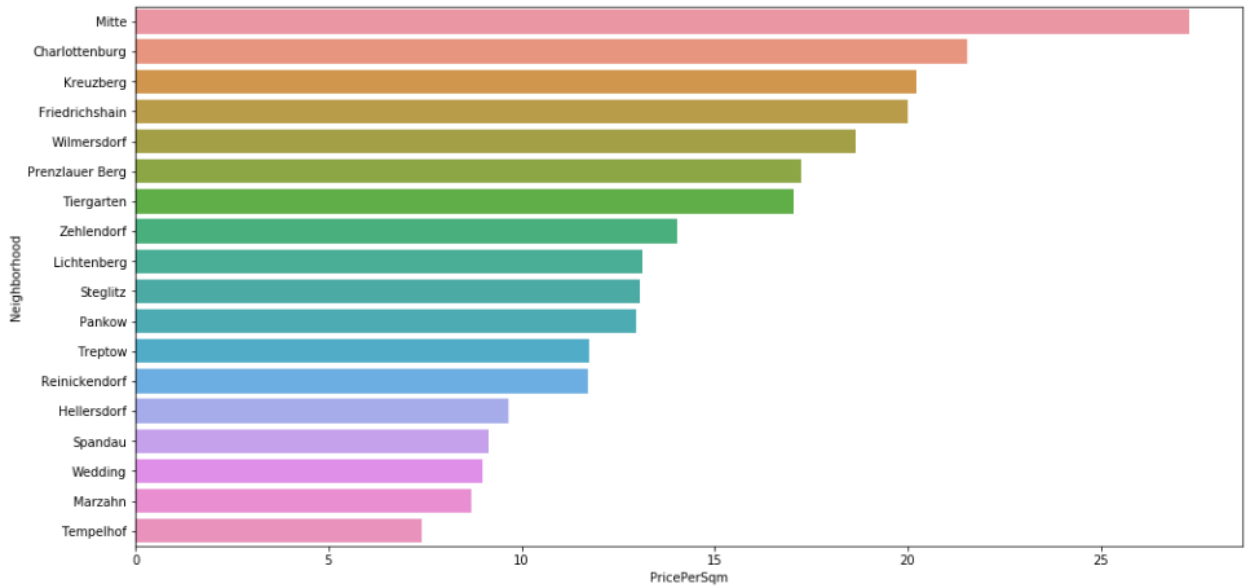
def getNearbyVenues(names, latitudes, longitudes, radius=500)

User input:
names: the addresses of the returned apartments
latitudes: the latitudes of the returned apartments
longitudes: the longitudes of the returned apartments

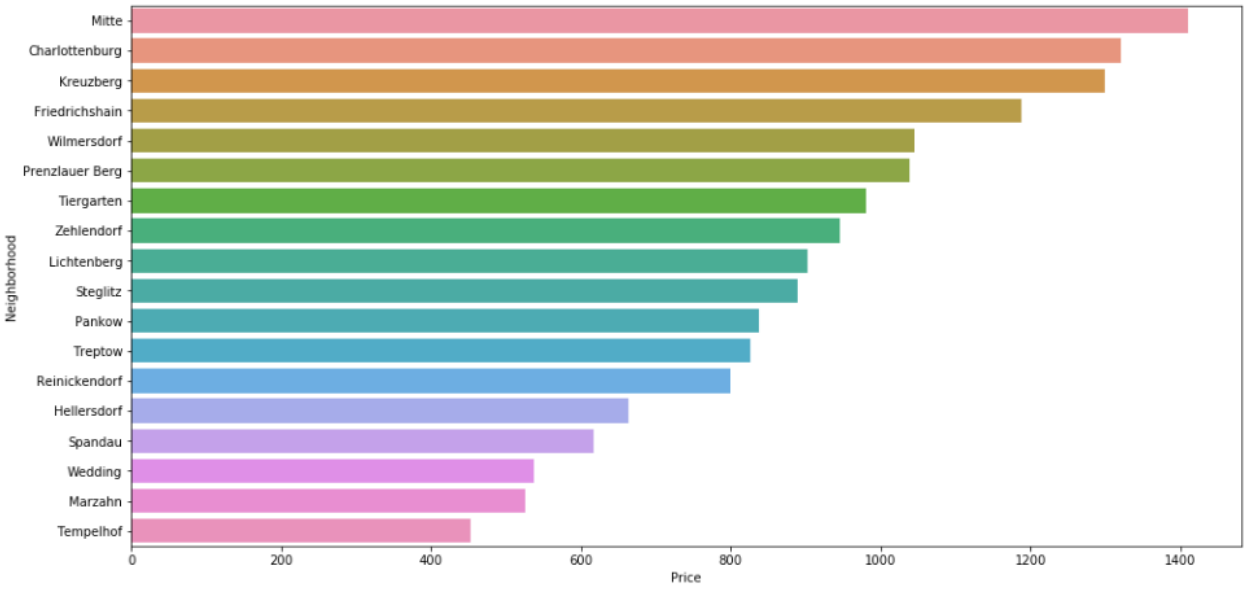
- Perform some exploratory data analysis on the results in order to get a better picture of the situation in Berlin.

Below are two graphs that were result of our exploratory data analysis.

Average price per square meter of each neighborhood



Average price per neighborhood



Part 2

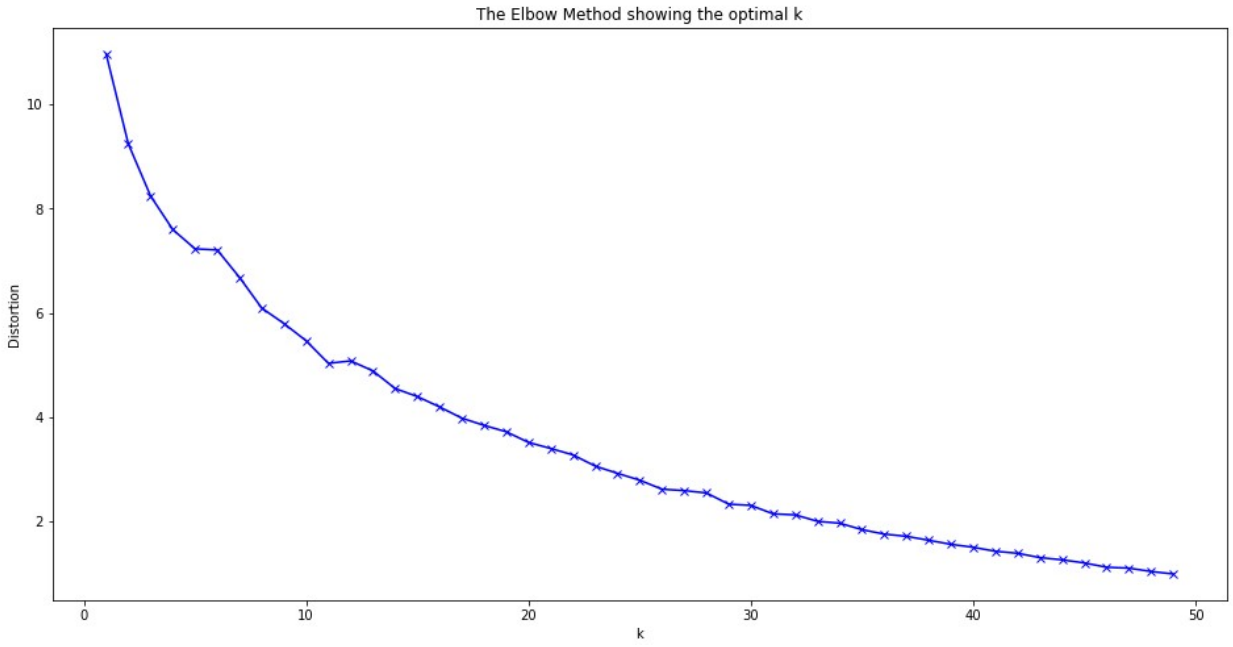
Clustering using K-Means algorithm. The K-Means algorithm is one of the most popular unsupervised machine learning algorithms. The goal of this technique is to find similarities to the data point and group similar points together.

In this situation the clustering will be performed using the similarities in the venues near each apartment scraped in Berlin.

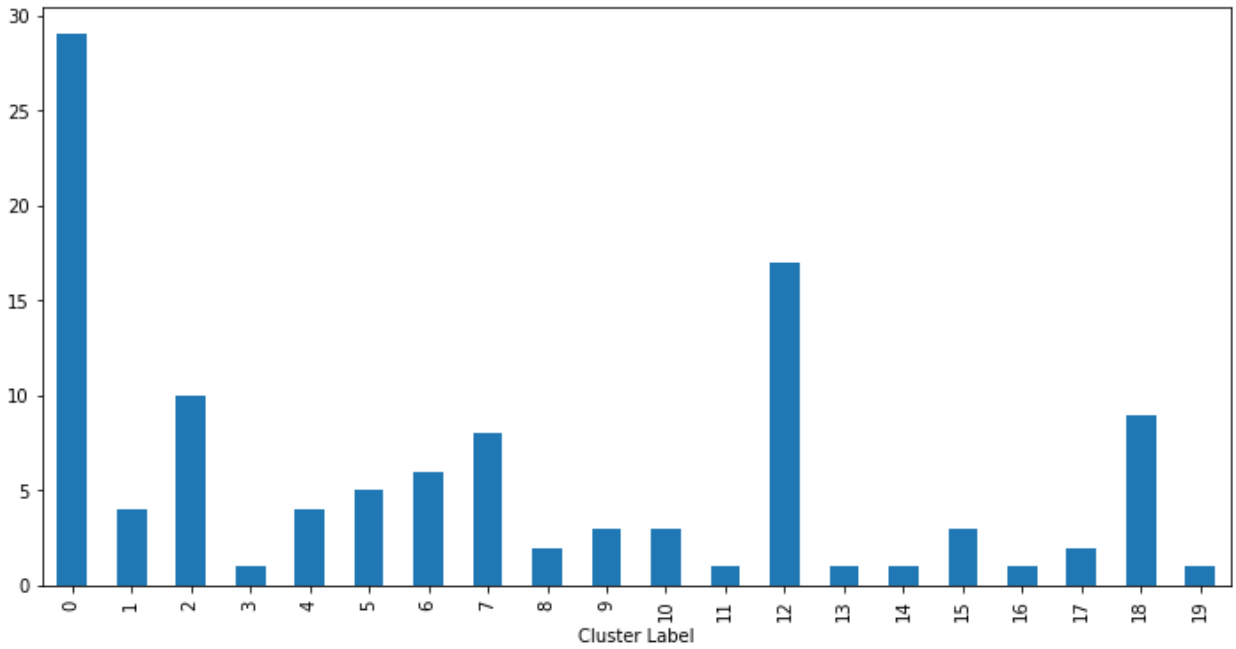
We will consider the top 20 most common venues in order to develop the clustering analysis.

Important thing here is to determine the optimal k for our K-Means algorithm. For this purpose the elbow method was used.

Unfortunately the elbow method did not produce clear results and choosing a high value of k would be out of the goal of this project.

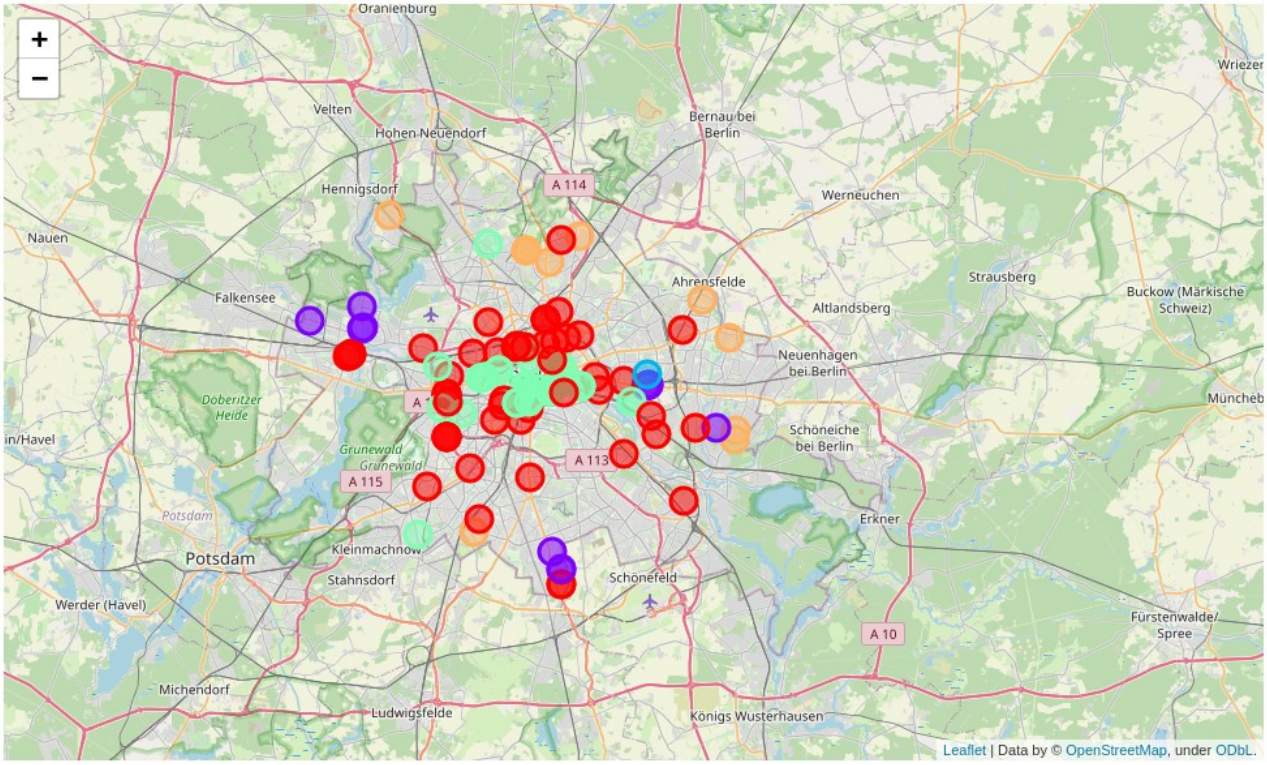


So after some tests the number of 20 was chosen because the number of apartments in each cluster seemed more equally distributed.



4. Results

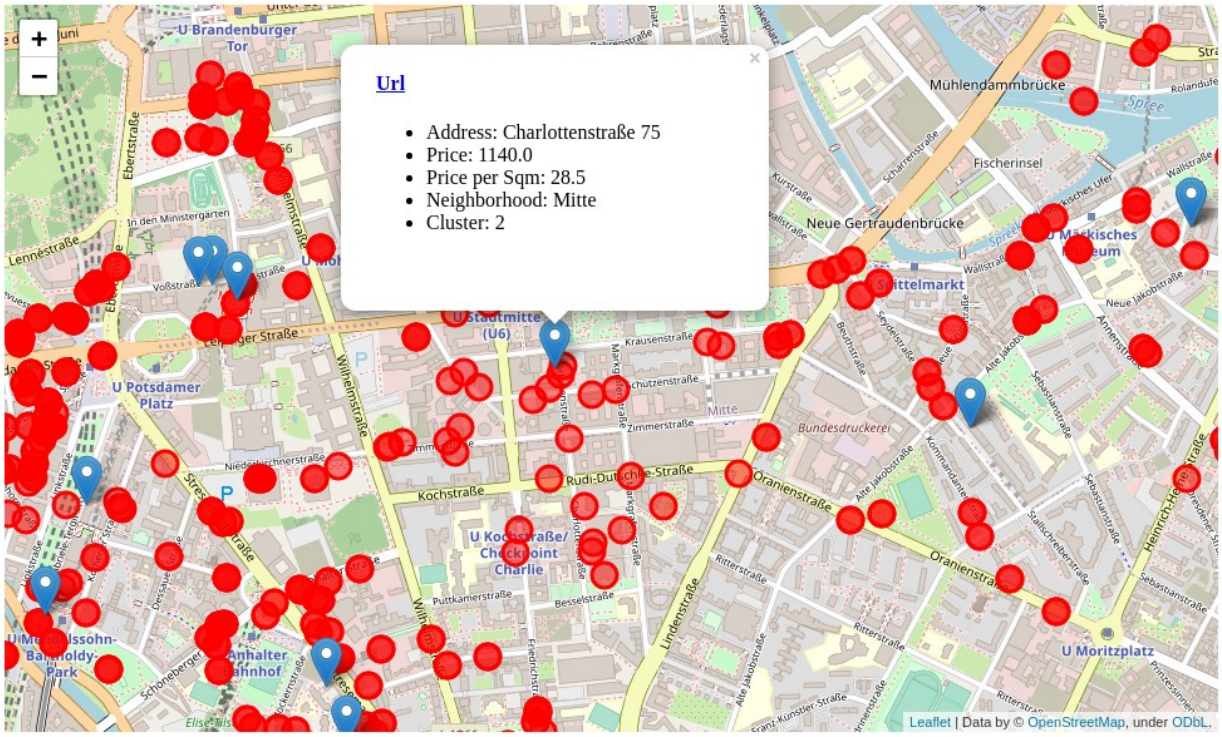
The clusters of the apartments were visualized based on the foursquare results .



There are 273 unique venue categories in our scrape results. The top 20 of them are visible below.

Hotel	166
Café	138
Supermarket	113
Bakery	91
Italian Restaurant	88
Coffee Shop	77
German Restaurant	54
Ice Cream Shop	51
Plaza	51
Bus Stop	49
Park	49
Gym / Fitness Center	43
Bar	42
Drugstore	38
Vietnamese Restaurant	36
Restaurant	36
Clothing Store	35
Asian Restaurant	31
Organic Grocery	30
Pub	30

The important result is that we managed to cluster real time and up to date apartments scraped from Immobilien Scout based on the venues nearby. User can explore each cluster individually and choose the one that suits him better.



One interesting feature in the map above is that the user can click on an apartment if it is interesting and then directly be prompted to Immobilien Scout via a hyperlink in order to find all the relevant information and the application process.

5. Discussion

Things to keep in mind:

- This research was conducted for two room apartments
- The user is free to query from the scraper the room that best suit his needs and then reproduce the analysis.
- It is functional so it can be reused
- Results might differ from the ones presented here as the data are real time

Potential issues:

- Sometimes geopy returns wrong coordinates probably because of mistake in its database or overlap with similar address.
- Be aware of Foursquare API limits.
- Potential overload of folium (always depending on user’s machine) when scraping too many pages of apartments

6. Conclusion

It seems that Berlin apartments were able to be clustered based on the venues nearby with the help of Foursquare API.

Sample of the resulting dataframe of clusters is visible below

	Address	Url	PricePerSqm	Cluster Label	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Ruppiner Chaussee 331	https://www.immobilienscout24.de/expose/72582744	10.015955	5	Supermarket	Health & Beauty Service	Gas Station	Flea Market	Event Space
1	Flämingstraße 70	https://www.immobilienscout24.de/expose/115916620	10.100000	5	Supermarket	Tram Station	Yoga Studio	Film Studio	Event Space
2	Havelländer Ring 33	https://www.immobilienscout24.de/expose/116931340	7.922824	5	Supermarket	Tram Station	Bus Stop	Yoga Studio	Flea Market
3	Streitstraße 28	https://www.immobilienscout24.de/expose/116460399	11.524971	5	Supermarket	Bus Stop	Pet Store	Drugstore	Turkish Restaurant
4	Nahariyastr. 43	https://www.immobilienscout24.de/expose/116452133	7.996977	5	Bus Stop	Supermarket	Fondue Restaurant	Farmers Market	Fast Food Restaurant

Future Research

I encourage all machine learning enthusiasts to engage themselves in further developing K-Means for better k optimizing an thus better clustering results. Moreover something that could have an impact to all people trying to determine the pricing in Berlin apartments would be to define if and how much could nearby venues define the price of a flat. Continuous data acquiring and statistical analysis would be crucial for this task.