SIMGE MULLAOGLU - 17.04.2021

The Best Place to Set Up Your Office in Berlin

Introduction

Berlin is the largest city in Germany and becoming more popular day by day. Berlin has more than 3.6 million population and 55 percent of the population is younger than 45 years of age, the average age was 42.7. [2] So we can say that Berlin is a young city.

Berlin has 12 boroughs and all of them have some characteristic places. Some of them are becoming more popular than others because of these characteristic places. [1] As we can see 80000 jobs were created by start-ups in Berlin in 2020. (Figure 1) Also, we can see that almost 35% of all German fintech start-ups locate in Berlin. (Figure 2)

Berlin is a living city and has so many different places like shopping centers, restaurants, coffee shops, offices, etc. All of these are some reasons for attracting people including expats and newbie startups.

Figure 1 - Jobs Created by Start-ups in Berlin [3]

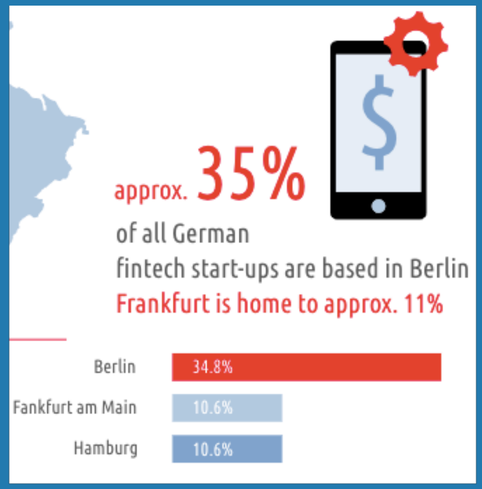


Figure 2 - Fintech start-ups with location percentage in Germany [3]

Business Problem

If you are looking for a place to set up your office, popularity is not enough to decide and you need to explore the real-world data. We always hear some complaints from people about the location of the office, or even in reviews on sites like Glassdoor, people consider the location of offices a minus or plus. Sometimes because of the distance or lack of transportation, they even quit their job. So this particular decision, with other plus features of your small company, of course, might attract more people to work with.

In this project, we will be answering the below questions:
1 - Which neighborhoods have more social places like restaurants, Coffee shops, and bars?
2 - In which neighborhoods the offices/workplaces are more common?
3 - Which neighborhoods have more common transportation centers?
4 - Which places are more attractive for people who are working in offices?
By answering these questions and combining them, we will be finding our final question:

Where should we open office of our startup to attract new workers?

Data Description

In this project, the data will be related to the Berlin location. I have three different data sets. I will clean the data, make some preparations and then combine all of them to get the final data set.

- The first data set was extracted online from geonames.org and in the form of a CSV file. [4] It is composed of the information of Berlin Postal Codes, and Location (Latitude/Longitude). The location column will be dividing two different columns in the data preparation phase since the column contains two different data and we will be using them separately in the project.

- The second data set was also extracted online from geonames.org. [4] It is in CSV file format and contains Postal Codes and Boroughs of Berlin. In data each Postal Code corresponds to a borough, so we are expecting that some postal codes are in the same borough.

- The third data set is extracting from foursquare by using Foursquare API for getting the most common venues in Berlin. The API is returning us the most common venues in Berlin, neighborhoods, latitude/longitude information, venue names, and venue categories.

Methodology

Data sets are downloaded and scraped from multiple sources and combined into one final data set. Null or meaningless values are checked and luckily there are not so many null or meaningless values. I decided to drop them from the dataset. Also, some transformations are made to specific columns like the Location column.

The third data set scraped by using Foursquare API - Venues Explore endpoint. After getting the venues in Berlin from the API call, put the data into a data frame. 7 columns and 3189 rows are obtained. Then we perform one hot encoding by calling the get_dummies function. This helps us to work with integers (0,1) rather than strings.

After grouping by Neighborhood and taking mean of them, we got values, categories, and neighborhoods. By sorting the values in our data frame and putting them in the newly created data frame that contains new columns like 1st, 2nd Most Common Venue, we got our final data frame.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Adlershof	Supermarket	Greek Restaurant	Convenience Store	Bus Stop	Soccer Field	Organic Grocery	Tram Station	Trattoria/Osteria	Taverna	Drugstore
1	Berlin-Mitte	Hotel	Coffee Shop	Italian Restaurant	Café	Bakery	Nightclub	Restaurant	Ice Cream Shop	Vietnamese Restaurant	German Restaurant
2	Charlottenburg	Italian Restaurant	Café	Hotel	Bakery	Coffee Shop	Supermarket	Vietnamese Restaurant	Asian Restaurant	Gourmet Shop	German Restaurant
3	Friedrichshain	Bar	Hotel	Bakery	Nightclub	Café	Pizza Place	Supermarket	Vietnamese Restaurant	Italian Restaurant	Drugstore
4	Kreuzberg	Café	Italian Restaurant	Coffee Shop	Bar	Ice Cream Shop	Pizza Place	Bakery	Turkish Restaurant	Vietnamese Restaurant	Cocktail Bar

Figure 3

At this point, I used the K-means clustering method from the scikit-learn library. Based on the data and after trying different cluster numbers, I choose 5 as a cluster number. I run k-means clustering by the number of 5 clusters and used the fit function. Also, the cluster labels are added. Then I merged this data set with the previous data set that contains postal code, neighborhood, latitude, and longitude by taking the neighborhood column as a reference. By using “geolocator.geocode(address)" I got latitude and longitude information of Berlin. I used the folium library for map rendering and visualized my clusters on the map in the Berlin location. (Figure 4)

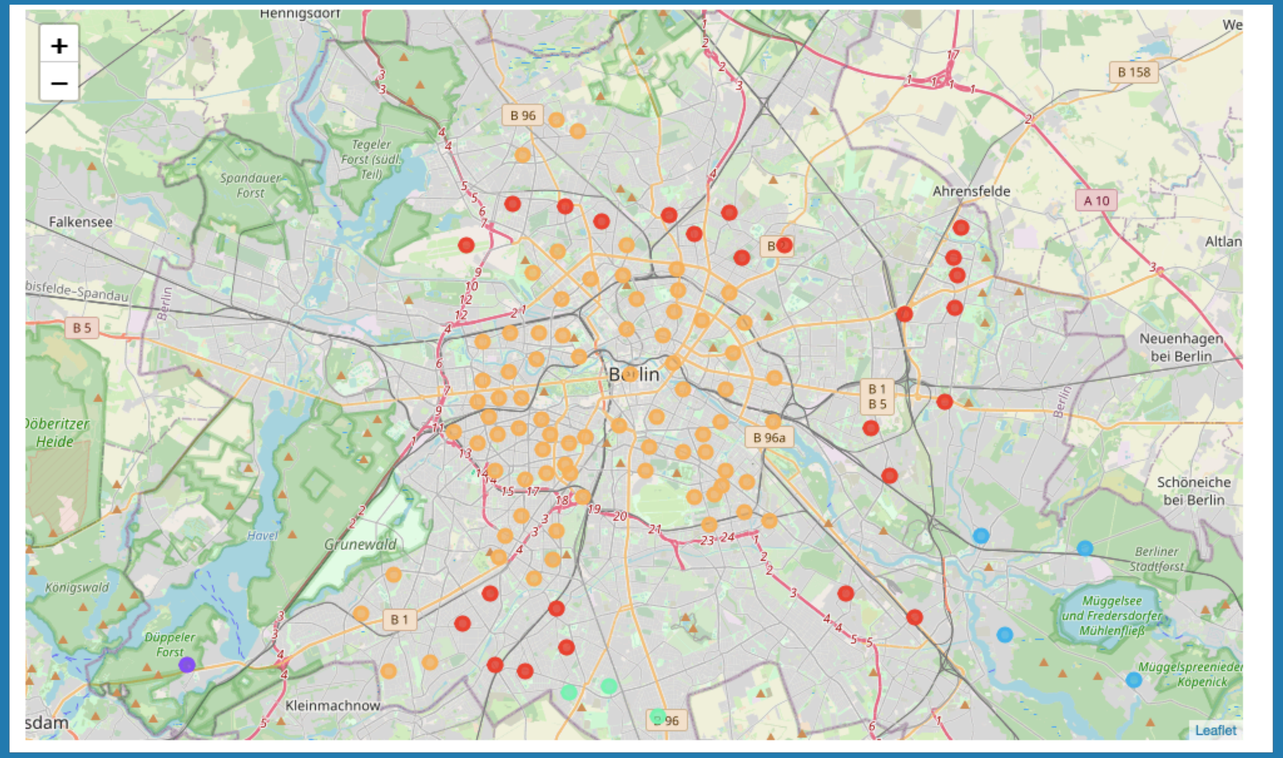


Figure 4 - Clusters in the Map

Analysis

After getting the data from Foursquare by using Foursquare API, I analyzed the data. By getting the count of the venues according to the neighborhoods, we can see that the most venues are located in Neukölln (421), Kreuzberg (363), Berlin-Mitte (342), and Charlottenburg (317). (Figure 5)

Also, I analyzed the data by grouping by Venue Category and filtering with Plaza to see which places have this category. We can see some of these places below. (Figure 6)

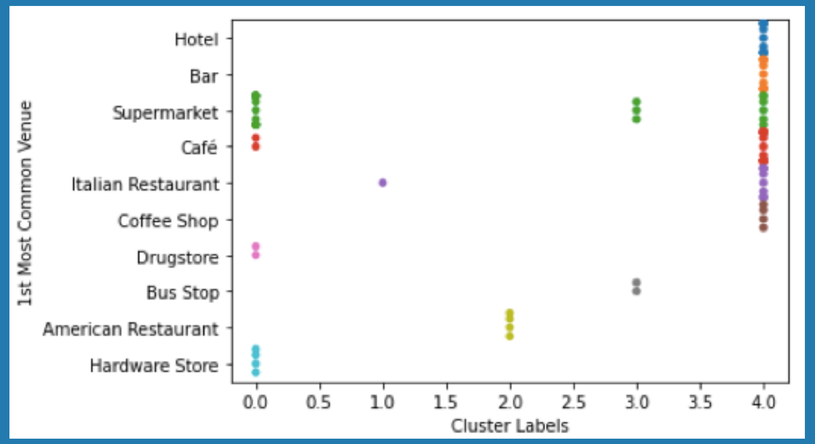
Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Adlershof	20	20	20	20	20	20
Berlin-Mitte	342	342	342	342	342	342
Charlottenburg	317	317	317	317	317	317
Friedrichshain	171	171	171	171	171	171
Kreuzberg	363	363	363	363	363	363
Köpenick	14	14	14	14	14	14
Lankwitz	18	18	18	18	18	18

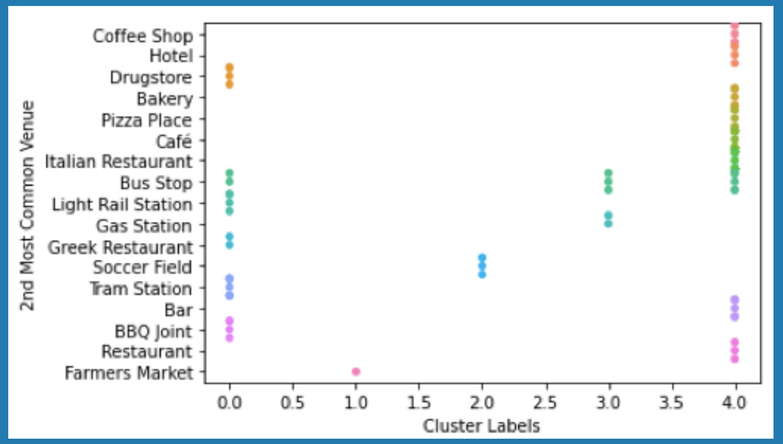
Figure 5

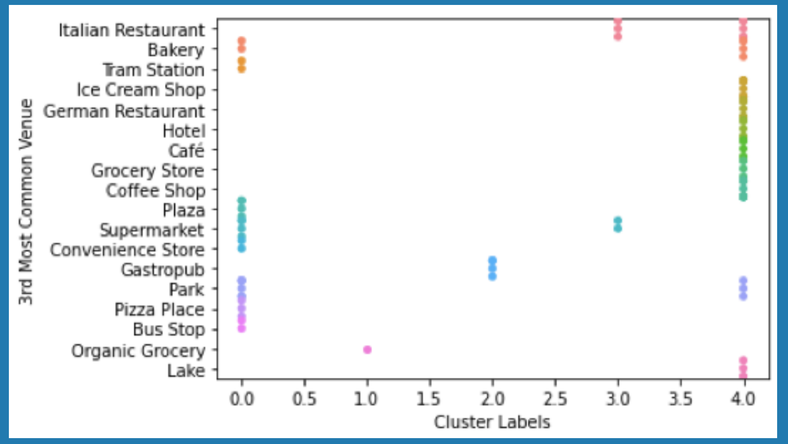
1302	Wilmersdorf	52.497	13.314	Olivaer Platz	52.499373	13.314636	Plaza
1342	Wilmersdorf	52.494	13.303	Hochmeisterplatz	52.495356	13.300835	Plaza
1354	Wilmersdorf	52.498	13.29	Henriettenplatz	52.496741	13.290995	Plaza
1366	Wilmersdorf	52.498	13.29	Rathenauplatz	52.495136	13.285539	Plaza
1389	Wilmersdorf	52.485	13.313	Heidelberger Platz	52.480962	13.312152	Plaza
1405	Wilmersdorf	52.482	13.329	Bundesplatz	52.479507	13.328540	Plaza
1424	Wilmersdorf	52.499	13.326	Ludwigkirchplatz	52.497055	13.323500	Plaza
1460	Tempelhof	52.497	13.343	Viktoria-Luise-Platz	52.495768	13.341866	Plaza
1498	Tempelhof	52.492	13.339	Viktoria-Luise-Platz	52.495768	13.341866	Plaza
1502	Tempelhof	52.492	13.339	Prager Platz	52.493245	13.332858	Plaza
1520	Tempelhof	52.492	13.339	Bayerischer Platz	52.489078	13.339784	Plaza
1549	Tempelhof	52.494	13.353	Winterfeldtplatz	52.497347	13.353790	Plaza

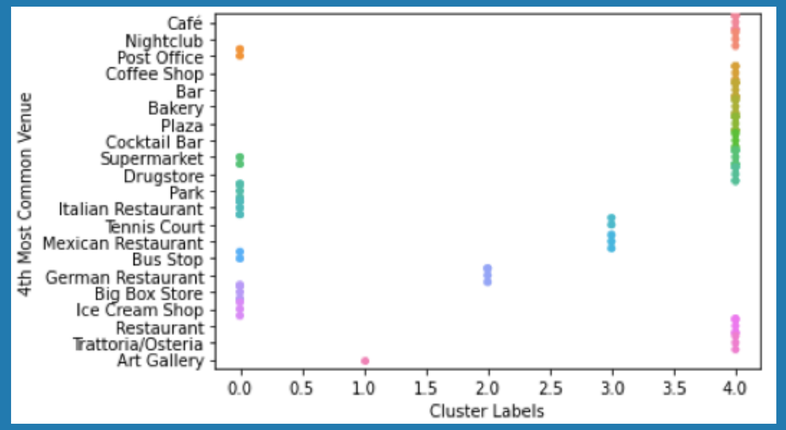
Figure 6

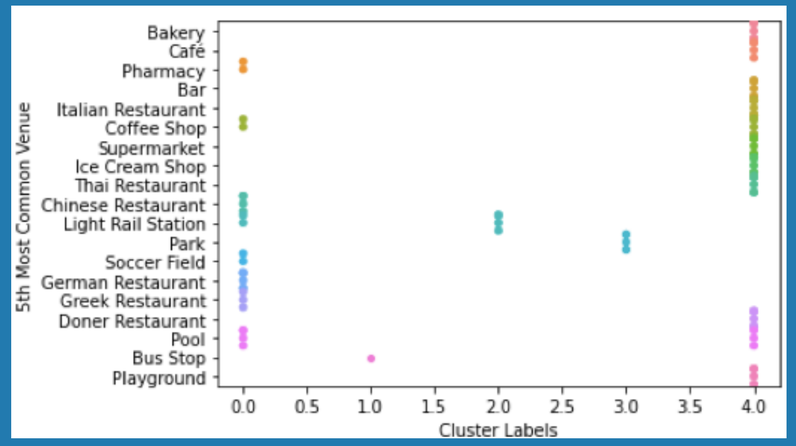
We can see details about the categories in the cluster

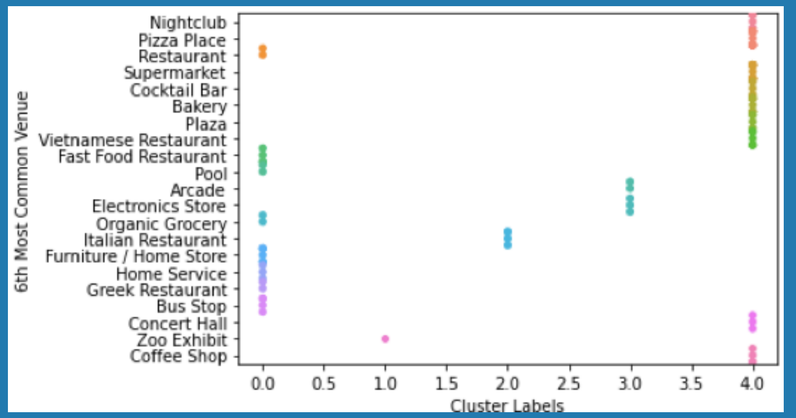












Results

As we can see from the analysis, the cluster divided based on some characteristics. For example, in cluster 1 Plaza, Supermarket, Restaurants, Rail Station, and Stores are common. Also in Cluster 5, Restaurants, Plaza and Coffee shops are common.

When some neighborhoods have some specific common places, other neighborhoods have others. For example, we can see that nightclubs or cocktail bars are more common in Cluster 5. Soccer Field is more common in Cluster 3. Moreover, transportation points are more common in Cluster 1. So we can say that if we are looking for a specific category, we can find the place where it is more common on the clusters.

Discussion

From the observations, we can say that Cluster 1 and Cluster 5 have Plaza as common venues. Plaza itself is not enough for locating your office in that neighborhood.

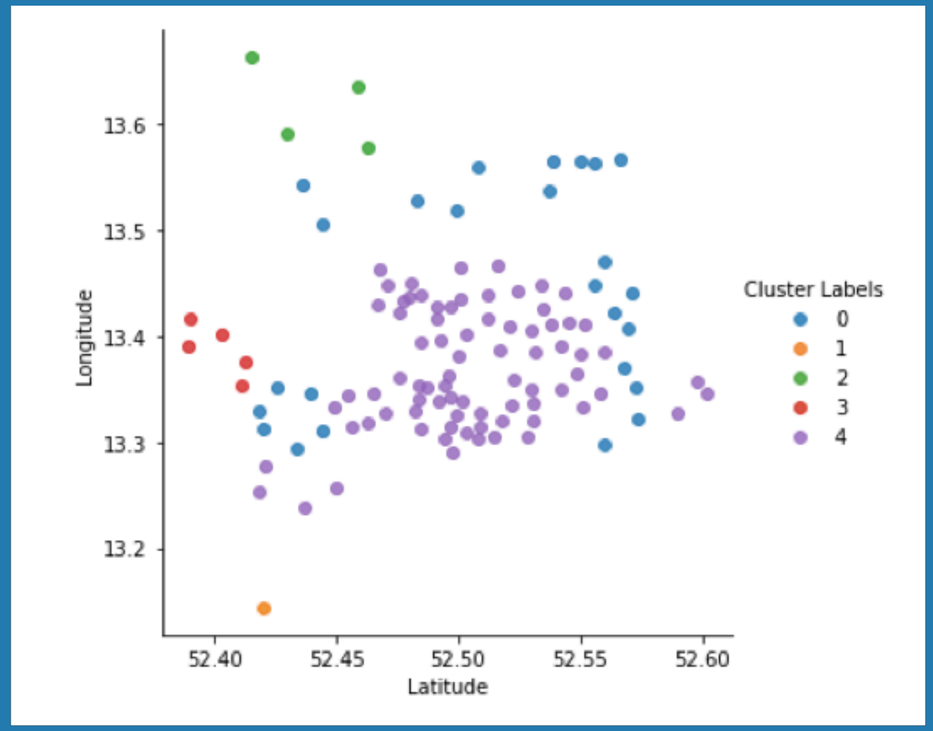


Figure 7 - Clusters based on Location (Latitude/ Logitude)

For example, locating your office in Köpenick might not be a good idea, since we can see that Soccer Field and American Restaurant are common places.

➡ Wilmersdorf might be a good option, since it has already Plaza as a common place, and it has Cafe, Restaurants, Hotel, Bakery, Supermarket, even Boutique.

➡ Tempelhof might be another good option since again it has already Plaza as a common venue, and it has common Restaurants, Cafe, Bar, and Supermarket.

➡ Zehlendorf and Schöneberg are also might be good options since, besides Plaza, it has Restaurants and Cafe. Also, Zehlendorf has Lake and Park where people can go for relaxing.

➡ Lichterfelde is another option for locating your office, since it has Restaurants, Supermarket, Rail Station and Bus Stop. Other neighborhoods in Berlin might not be a good option for locating your office.

Conclusion

In this project, I examined the neighborhood and venues in Berlin city. I was looking for an answer for people like CEOs or business owners, co-founders where they should locate their offices. After performing analysis for each cluster and neighborhood, I got my final results. Although all neighborhoods in Berlin have some other common venues, not every one of them is well suitable for a start-up or office.

If you are looking for a place to set up your office and looking for the best place for your employees, you need to meet up with their expectations. Having restaurants, coffee shops, public transportation, park, etc. make the place more attractive for people. Based on my analysis, I can say that locating your offices in Wilmersdorf, Tempelhof, Zehlendorf, Lichterfelde, and maybe in Schöneberg and Steglitz will attract your current or future employees.

References

[1] https://theculturetrip.com/europe/germany/berlin/articles/the-10-coolest-neighbourhoods-in-berlin/

[2] https://www.businesslocationcenter.de/en/business-location/berlin-at-a-glance/demographic-data/#:~:text=The%20population%20of%20Berlin&text=With%20its%20roughly%203.77%20million,the%20average%20age%20was%2042.7.

[3] https://www.businesslocationcenter.de/en/startup-capital-berlin/

[4] https://www.geonames.org/postalcode-search.html?q=14169&country=DE