

# Berlin Venues Project

Simge Mullaoğlu

## 1- Introduction/Business Problem

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 created by start-ups in Berlin in 2020. (Figure 1) Also we can see that almost 35% of all German finch 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.

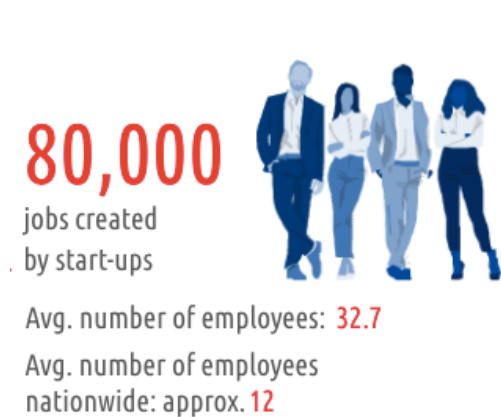


Figure 1 - Jobs Created by Start-ups in Berlin [3]

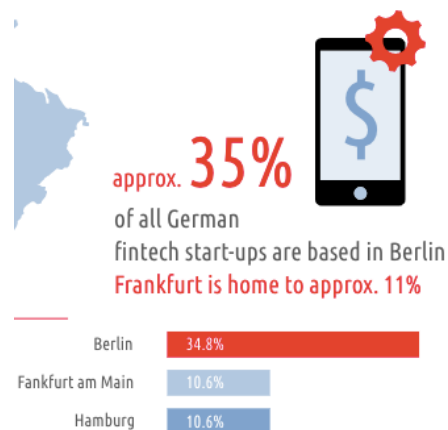


Figure 2 - Fintech start-ups with location percentage in Germany [3]

As I mentioned, there are a lot of demands for living in Berlin. It has so many places to attract young people, job seekers, startups.

**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 below questions:

1 - Which neighborhoods have more social places like restaurants, Coffee shops and bars?

2 - In which neighborhoods the offices/work places are more popular?

3 - Which neighborhoods have more popular 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 our office of tech startup to attract new workers?**

We will be exploring boroughs and venues of Berlin to find out where is the best place to locate your new office. At the end of the project, we will be giving suggestion of locations to CEOs or Managers of startups for setting up their new offices.

## 2- Data Description

In this project, the data will be related with 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.

First data set extracted online and form of csv file. It is composed of the information of Berlin Postal Codes, and Location (Latitude/Longitude). In this data, each postal code has their unique coordinate data in Location column. Location column will be dividing two different columns in data preparation phase, since the column contains two different data and we will be using them separately in the project. Postal Codes are numeric data with five figures(e.g. 10115), and Location column is Varchar data type with latitude and longitude information (e.g. '52.532/13.385').

Second data set also extracted online with 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. Postal codes are numeric (integer) with five figures like 10115 and Borough column is varchar data type (e.g. 'Berlin-Mitte'). This data set will be combining with others later based on Postal Code column.

Third data set is extracting from foursquare by using **Foursquare API** for getting the most popular venues in Berlin. The API is returning us most popular venues in Berlin, neighborhoods, latitude/longitude information, venue names and venue categories. After the getting and combining data sets, I will be using Folium library for map rendering. I will give details in methodology section.

By combining these three data sets I will be exploring, clustering, and make data analysis based on combined data.

## 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/>