

# Coursera Project

## Opening a Café in Berlin, Germany

### Introduction

Coffee houses have been an important social gathering point in Europe and their appearance encouraged several cultural and political transformations during the 17<sup>th</sup> and 18<sup>th</sup> centuries. They provide a forum for exchanging views and nurturing public opinion across the social spectrum. Nowadays the number of cafes in Berlin are increasing in the rapid rate and provide great revenue to the owners and property developers. Selecting the location of cafes plays an important role in the success of it due to the competition with other emerging cafes and famous coffee chains. Many entrepreneurs will be interested in opening a café in the capital city of Germany.

### Business Problem

The aim of this capstone is to analyse and select the best location in Berlin, Germany to open a new café. By using data science methodology and machine learning techniques like clustering, the project can provide solutions to answer the business question: If an entrepreneur wants to open a café in Berlin, Germany, where would you recommend him to open it?

### Data

The Data need are as follows:

- 1.The list of neighbourhoods in Berlin. Our project is confined to the city of Berlin, Germany , Europe.
- 2.The latitude and longitude of the neighbourhoods of Berlin.
- 3.Venues data, particularly data related to cafes. This data is used to perform clustering on the neighbourhoods.

### Methodology

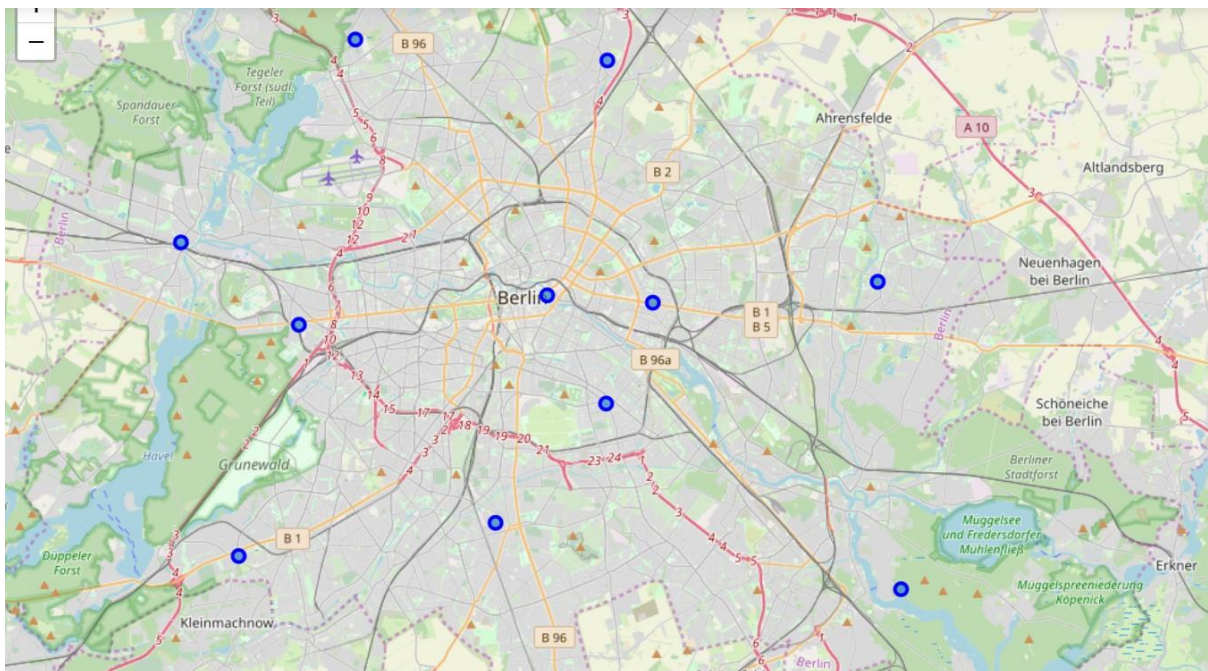
First, we will need to get the list of neighbourhoods in Berlin. We can get this list from the Wikipedia page ('[https://en.wikipedia.org/wiki/Neighborhoods\\_and\\_neighborhoods\\_of\\_Berlin](https://en.wikipedia.org/wiki/Neighborhoods_and_neighborhoods_of_Berlin)').We need to extract this list using web scrapping method. BeautifulSoup package and Pythons requests packages are used for extracting the data. After getting the list, we need to find the

geographical coordinates of the neighbourhoods of Berlin. To do so we can use the Geocoder package which will convert the address into its constituent latitude and longitude coordinates.

After gathering the data, we will place our data into a pandas DataFrame and visualize the neighbourhoods in a map using the folium package. This shows us the geographical location of the city of Berlin. Next, we will use the Foursquare API to get the top 100 venues that are within a radius of 2000. Before using the API we need to create a Foursquare Developers Account in order to obtain the Foursquare ID and the secret key. Now we make API calls to Foursquare by passing the coordinates of the neighbourhoods. Foursquare will return the venue data in JSON format and extract the venue name, category, latitude, longitude. With this data we can analyse the venues and their category in each neighbourhood.

We are finding the cafes in each neighbourhood so we will filter 'Café' as venue category for the neighbourhoods. This will give us the exact number of cafes in each neighbourhood. Lastly, we will be performing clustering on the filtered data using the K-means clustering technique. This technique will cluster the neighbourhood into 3 clusters and find the frequency of Cafes in each neighbourhood.

This will help us analyse and identify the neighbourhoods which will be applicable to open a new café with immediate success by identifying the neighbourhoods with less and more concentrations of cafes.



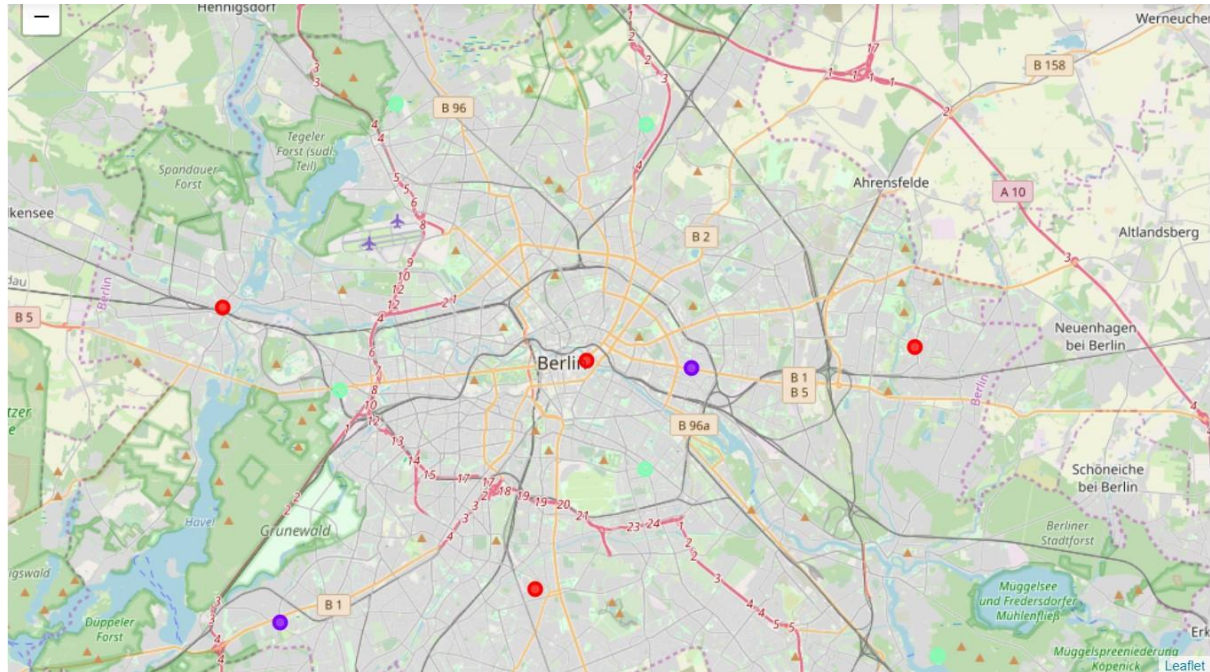
## Result

K-means Clustering shows the divides the neighbourhoods into 3 clusters based on the frequency of cafes in the neighbourhood.

Cluster 0: Neighbourhoods with less concentration of cafes. (red colour)

Cluster 1: Neighbourhoods with high concentration of cafes. (blue colour)

Cluster 2: Neighbourhoods with moderate concentration of cafes. (mint colour)



## Conclusion

1. 'Mitte' which is often referred to as Berlin's centre, due to many famous monuments, has a cluster 0, so there are fewer cafes in the neighbourhood. It will be a huge advantage to set up a café in city centre which attracts a huge number of tourists.
2. 'Charlottenburg-Wilmersdorf' has a moderate number of cafes. Setting up a café in this neighbourhood can be a great prospect as this neighbourhood is near to the Olympic Stadium which hosts many events including Bundesliga games for Hertha Berlin team.
3. Neighbourhoods with low concentration of cafes are the suburbs of Berlin except the 'Mitte'.
4. Neukölln is renowned for many famous museums and has a moderate number of cafes in the neighbourhood.

