

IBM Applied Data Science Capstone

# Moving to Berlin

A Neighbourhood Guide for Newcomers to Berlin



Joshua Gladwin

30 January 2020

## Introduction

Berlin is one of the fastest growing tech hubs in Europe. With Brexit uncertainty threatening London's place as the tech capital of Europe, many in the tech industry are looking to relocate to Berlin. The city is noted for its high quality of life and, in comparison to London, low cost of living. Berlin is an international city, with a growing international community. The city is famed for its open and friendly attitude towards women and LGBT people.

## Business Problem/Target Audience

Berlin's tech companies are often hiring from other cities, such as London and San Francisco, to make sure that they have the best employees with the right skills for their needs. This often requires a relocation process, to aid the new employees move to and settle into Berlin. The aim of this project is to provide newcomers to Berlin with a guide to the various neighbourhoods in Berlin. Many people may have visited Berlin at some point before relocation, but the experience of a tourist is much different to that of someone who lives in Berlin. It is important that a company helps its employees settle into their new city as quickly as possible. If there are problems finding the right neighbourhood to live in, this may have a negative impact on the employee's performance at work in the crucial early stages of the role. An employee may have to relocate once they have started in the role causing much unnecessary stress. In the worst-case scenario, new recruit may decide to return back to their previous location. It is therefore important that the company does its utmost to ensure that there is a seamless transition from one city to another.

## Data

There are three data sources that are required to solve this task:

- List of the neighbourhoods in Berlin. This forms the backbone of the analysis, providing the main data points to be analysed.
- Latitude and longitude coordinates of the neighbourhoods. This allows the neighbourhoods to be plotted on a map, to better visualise the neighbourhood clusters
- Venue data for each of the neighbourhoods. This provides the information needed to cluster the neighbourhood data.

Firstly, the neighbourhood data is gathered from Wikipedia. Here we can find a list of the 96 *Ortsteile*, the 96 official neighbourhoods within the city-state of Berlin, as designated by the Berlin State government.

Next, using the Geocoder Python package, we can obtain the latitude and longitude of each of the neighbourhoods. With the coordinates of the neighbourhoods, we can now plot the geographic location of each of the neighbourhoods using the mapping library Folium.

Finally, using the Foursquare Venues API, we can get a list of the venues found in each of the neighbourhoods. The venues data can then be used to compare the similarities between each of the neighbourhoods, using K-Means clustering, to group the neighbourhoods into distinct clusters.

## Bibliography

NomadList, 2019. *Cost of Living in Berlin*. [Online]

Available at: <https://nomadlist.com/berlin>

[Accessed 22 01 2020].

Wikipedia, 2019. *Boroughs and neighborhoods of Berlin - Wikipedia*. [Online]

Available at: [https://en.wikipedia.org/wiki/Boroughs\\_and\\_neighborhoods\\_of\\_Berlin](https://en.wikipedia.org/wiki/Boroughs_and_neighborhoods_of_Berlin)

[Accessed 22 01 2020].