

## 1. Introduction/Business Problem

Berlin, the capital of Germany and one of the most important cities in Europe, is attractive for tourists and investors for many reasons. Among them belong its vibrant culture, fabulous food, intense night-life and centuries-old history. Although the city is pretty big, there are some key areas which are of most interest for business owners and investors.

The classification and segmentation of Berlin central neighborhoods with respect to popular venues, as well as the analysis of population demographics and tourists overnight stays is aimed to support finding those areas and thus support the decision making of following stakeholders:

- Any contractor or entrepreneur who is aiming to open new business (e.g. hotel, restaurant, cafe, yoga studio), and while looking for the right location has to take into consideration factors such as target group or possible surrounding businesses and competitors. Tourist visit ratio, most common venues, and average age of residents of given area are important parameters and are subject of the analysis presented in next chapters.
- Investors in commercial real estate. Investing in the right property includes understanding of local demographics, having overview of the type of successful businesses in the area or of its attractiveness for tourism.
- City planners when developing the design and monitoring the development of public spaces, parks, transport, and community infrastructure in each neighborhood.

## 2. Data

To address and analyze the problem of interest presented above, data was drawn from three separate data sources:

- [Official Berlin Census Data](#) for up-to-date demographic and touristic data for the city of Berlin. The information is part of the [Creative Commons Namensnennung 3.0 Deutschland Lizenz](#) 2019.

The official population statistics of Berlin consist of various statistics, each of which has its own function. Our main focus is the number of residents in Berlin divided by postal code, district and age groups. This information is available in xsml-format under following link (last updated on June 30, 2019): [Link to file](#)

To illustrate domestic tourism, accommodation facilities with ten or more guest beds and all campsites for holiday camping with ten or more parking spaces are surveyed monthly by the official authorities in Germany. The results are then published and publicly available on the internet. The tourist data of interest for this project is the number of guest arrivals and overnight stays for each district (Bezirk) of Berlin for the

period January 2019 till September 2019. It can be downloaded in XLSX format from following link: [Link to file](#)

- [OpenDataSoft](#) to get the geo coordinates (longitude and latitude) of each postal code area. This data is available under the Open Database License. For more information and credits visit "[© OpenStreetMap contributors](#)"

I downloaded from the following link the CSV file containing postal codes of all neighborhoods within Berlin with the corresponding gps coordinates which are required for executing requests and retrieve information from Foursquare API:  
[Link to File](#)

- [Foursquare Developer](#) to get the most popular venues for each postal code area of interest using Foursquare technology and its user-generated content data.