

# **Clustering Neighbourhoods in Berlin City (Germany) To Find Out Which Neighbourhoods Require More Public Parks, Based On Public Parks Available To Residents**

## **2. Data Collection**

Data will be collected from the following data sources:

### **1. Berlin Neighbourhood Data:**

Data will be scraped from the following Wikipedia page  
[https://en.wikipedia.org/wiki/Boroughs\\_and\\_neighborhoods\\_of\\_Berlin](https://en.wikipedia.org/wiki/Boroughs_and_neighborhoods_of_Berlin)

The data includes 96 localities in the Berlin city.

### **2. Location Coordinates for each Neighbourhood:**

The Geocoder Python package will be used to get latitude and longitude of each Neighbourhood.

### **3. Berlin Venue Recommendations from FourSquare API**

(FourSquare website: [www.foursquare.com](http://www.foursquare.com))

The FourSquare API will be used to explore neighborhoods in Berlin City. The Foursquare explore function will be used to get parks in each neighborhood, and then use this as one of feature to group the neighborhoods into clusters. The following information will be retrieved:

- Venue Name
- Coordinates : Latitude and Longitude
- Category Name