

Data Description

In order to solve the proposed problem, two kinds of data are needed including the neighborhoods of Berlin and the corresponding venue data of these neighborhoods.

The data of neighborhoods of Berlin can be scraped from the Wikipedia(https://en.wikipedia.org/wiki/Category:Localities_of_Berlin). For this, we will use the method of requests in Python and the BeautifulSoup packages to obtain the data by scraping the above website. A list of neighborhoods can be obtained. Then, the Python Geocoder package will be applied to get the geographical coordinates of the neighborhoods so that a further step could be proceeded.

With the geographical coordinates, the Foursquare API will be used to further get the venue data for these neighborhoods. Foursquare has been widely used by over 125,000 developers with a large dataset of over 150 million places. Foursquare API provide a variety of categories of venue data, among which we are interested in the shopping mall category so that we can do our analysis. Therefore we will use the Foursquare API to get the venue data of shopping mall in Berlin.