

## Data

Since my approach is rather straightforward, I do not rely on extensive data sources. I will only use the Foursquare API to acquire the coordinates from all public transport stations and medical facilities within a certain area. How will I use the data to solve the problem? First of all, I will extract, transform and load the respective datasets. Secondly, plotting the data within the folium-map based on real coordinates is important to get an impression of the distribution and whether it makes sense to analyze for possible clustering. However, the coordinates have to be translated into radians for mathematical calculations in order to calculate distances based on the Haversine Formula. Afterwards, results for each city will be concatenated into one data frame for statistical analysis.