Data Tools

The data for this project has been retrieved and processed through multiple sources. It has to be noted that the sources of data are scarce for South Africa. Therefore, to do data acquisition I had to fill data manually from external sources:

Neighbourhoods

The internet seems to have less data either in a form of a list for listing all the Johannesburg suburbs, as result I visited a webpage: https://www.roomsforafrica.com/dest/south africa/gauteng/johannesburg.jsp?tab=3, to search for Johannesburg's suburbs name. The suburbs will be used an alternative word for Neighbourhoods.

For each suburbs the geographical location for each was needed. I used the following webpage to retrieve the geographical coordinates for each listed Johannesburg suburbs: https://www.gps-coordinates.net/.

- The latitude and longitude of the neighbourhoods are retrieved using **Google Maps Geocoding API**. The geometric location values are then stored into the initial data frame.
- The venue data for each suburb in a data frame will be retrieved using a **Fouraquare API** and creating another data frame to contain all the venues details regarding its number of visits as compared with other venues within certain radius, and so forth.

Data scraping using **BeautifulSoup** library is one of the pleasurable data acquisition methods that can ease some cumbersome task for data scraping, but lack thereof of data which forms as a vitality for data analysis enforced me to create a json file from scratch, using acquired data through Jupyter Notebooks.