

## Data

We need to use the following Data:

- ⇒ List of neighborhoods in Hyderabad.
- ⇒ To plot the map latitude and longitude coordinates are required.
- ⇒ To perform clustering, we need venue data related to gaming cafes in Hyderabad.

## Sources and Methods used to extract data

Wikipedia pages related to the city will be the main source of data. We will use web scraping techniques to extract the data from these pages, with the help of python requests and beautifulsoup packages. Then we will get the geographical coordinates of the neighborhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighborhoods.

After that, we will use Foursquare API to get the venue data for those neighborhoods. Foursquare has one of the largest database of 105+ million places and is used by over 125,000 developers. Foursquare API will provide many categories of the venue data, we are particularly interested in the **gaming cafe** category in order to help us to solve the business problem put forward. This is a project that will make use of many data science skills, from web scraping, working with API, data cleaning and wrangling, using machine learning techniques and map visualization. In the next section, we will discuss about methodology and steps taken to analyze the data.