CAPSTONE PROJECT- BATTLE OF NEIGHBORHOODS

DATA

To solve the problem, we will need the following data:

• List of **neighborhoods** in **Bangalore**. This defines the scope of this project which is confined to the city of Bangalore, the capital city of the **Karnataka** state of India.

• **Latitude** and **longitude** coordinates of those **neighborhoods**. This is required in order to plot the map and also to get the venue data.

• Venue data, particularly data related to **Pizza places**. We will use this data to perform **clustering** on the neighborhoods.

SOURCES OF DATA AND METHODS TO EXTRACT THEM

- The data is stored in a **Wikipedia page** having the link as below: https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Bangalore
- This contains a list of neighborhoods in Bangalore with a total of 8 **boroughs** and 88 **neighborhoods**. The **boroughs** are central, northern, southern, eastern, western and many more.
- We will use web scraping techniques to extract the data from the Wikipedia page, 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 databases 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 **Pizza places** 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** such as **web scraping** (Wikipedia), **working with API** (Foursquare), **data cleaning**, **data wrangling**, **machine learning** (K-means clustering) and **map visualization** (Folium).