Capstone Project - The Battle of Neighbourhoods (Lagos, Nigeria)

Applied Data Science Capstone by IBM on Coursera

1. Introduction: Problem Background

This project deals with discussing the neighbourhoods of **Lagos city, The commercial capital of Nigeria** (The Most populous black nation) and the commercial hub of West Africa. It is the 4th wealthiest city in Africa, fourth place behind Johannesburg, Cairo and Cape Town.

This project would help people planning to start-up business in the city find the most ideal location.

The Foursquare API is used to access the venues in the neighbourhoods. Since, it returns less venues in the neighbourhoods, we would be analysing areas for which countable number of venues are obtained. Then they are clustered based on their venues using Data Science Techniques. Here the k-means clustering algorithm is used to achieve the task. Folium visualization library can be used to visualize the clusters superimposed on the map of the city. These clusters can be analysed to help small scale business owners select a suitable location for their need such as Hotels, Restaurants or Pizza shops.

2. Data Requirements

The Dataset is the Wikipedia page of Lagos state (https://en.wikipedia.org/wiki/List_of_Lagos_State_local_government_areas_by_population)

We shall explore Lagos city through its respective Local Government Areas (LGA) or Boroughs. The above link is a web page that shows the respective boroughs in Lagos State and each population figure. It's a public Wikipedia data page.

First, we'd save the weblink for The Lagos data, then get the source code HTML data from the Website and use The Beautifulsoup module to parse it. The next step is to get the table that contains the data we want to scrape from. Then finally, we'd iterate through each link of the table and append the text parts.

We than creating the data frame, we create a dictionary and append the LGA and corresponding Population data to it. Then we would attach a column of longitude and latitude by using the geopy.nominatim library to get each LGA latitude and longitude data.

The following steps are taken to analysed the data:

- We scrape the web page using the beautiful soup library

- Follow by using Foursquare API calls to retrieve geolocation data and then fetch the text data using the requests library

- We than convert it from JSON to Pandas data frame using the json_normalize module

- We shall use the folium library to render the maps and plot these via The Matplotlib library. Then we shall explore respective boroughs and analyse each area as a location for Start-up based on the aforementioned parameters.

- After this, we shall select our top location.