**CLASSIFICATION OF CITIES IN NIGERIA USING FOURSQUARE LOCATION DATA**

**INTRODUCTION**

**Background of the study**

Nigeria is the largest country in West Africa and the most populous country in Africa as a whole, with over 206 million people. Nigeria’s population makes up 2.64% of the total world population. The population density in Nigeria is 226 per km2, that is 586 people per mile2. The total land area is 910,770km2 (351,650sq miles). 52% of the Nigeria population is urban and the rest rural.

As a resident and citizen of Nigeria, I have decided to use Nigeria as a whole in my project.

Nigeria is a very beautiful country brimmed with many inspiring natural attractions such as mountains, beaches, forests, city skylines, deserts, local landscapes, to mention but a few. It is a country that is blessed with so many natural resources such as petroleum, natural gas, tin, iron ore, coal, limestone, lead, zinc and arable land. The oil and gas sector accounts for about 10% of gross domestic product, and petroleum exports revenue represents around 86% of total exports revenue ([www.opec.org](http://www.opec.org)).

Nigeria also has excellent cuisines which consists of food items and dishes from the hundreds of ethnic groups that comprise Nigeria.

**Business Problem**

For the above stated facts, Nigeria is a country that has attracted foreign investors and tourists but there is no readily available information to guide potential investors and tourists on the cities to invest in or visit. The target stakeholders for this project are investors and tourists.

To solve the above problem, we can create a map that shows major cities in Nigeria clustered according to the venues that surround these cities. This will enable investors know cities to invest in and enable tourists know which cities to visit according to their interests and preferences.

**DATA SECTION**

**Data Description**

It was almost impossible to get any data for locations in Nigeria with their geographical coordinates. However, I was able to come across data containing location of all the major cities in Nigeria and their geographical coordinates.

The data used in this project include:

1. List of major cities in Nigeria and their geographical coordinates scraped from <https://www.latlong.net/category/cities-160-15.html>
2. The foursquare API location data will be used to explore venues surrounding each city.

**METHODOLOGY**

**Data Preprocessing**

The Nigeria data containing cities with their respective geographical coordinates, was used to populate a pandas dataframe. With this dataframe, we could start manipulating the data.

A screenshot of a cell phone

Description automatically generated

Fig 1: Dataframe of cities.

The python folium library was used to visualize the geographical details of the cities in Nigeria. A map was created using the longitude and latitude values to get the map as shown below:

A picture containing text, map

Description automatically generated

Fig 2: Map of cities.

Next, we utilized the Foursquare API location data to explore the cities and segment them. My Foursquare credentials were defined. The Foursquare credentials are essential for exploring places.

The first city that was selected was Ikeja in Lagos, Nigeria. We got the top 1000 venues within a radius of 5000 meters. To do this, we had to create the GET request URL. This url is used to get the JSON data that contains all the venue data for a given or selected location.

In the next step, we used a get\_category\_type function which is a function that extracts the category of a particular venue. After this step, the JSON data is cleaned and structured into a new pandas dataframe as shown below:

A screenshot of a cell phone

Description automatically generated

Fig 3: Dataframe showing categories of venues.

We also show how many venues were returned by Foursquare.

**Exploring cities in Nigeria**

We get all the cities, and all their respective venues and populate them in a pandas dataframe. We also check how many venues were returned for each city.

A screenshot of a social media post

Description automatically generated

Fig 4: Number of venues returned for each city.

We also find how many unique categories can be curated from all the returned venues. We have a total of 39 unique categories.

**Analyzing each city**

Here, we apply one hot encoding to all venues. So now, the number of columns become 61. Next, we group rows of cities by taking the sum of the number of occurrences of each category.

Now, we create a new dataframe and display the top 10 most common venues for each city as shown below:

A screenshot of a cell phone

Description automatically generated

Fig 5: Top 10 most common venues.

**Clustering Cities**

For clustering, K-means clustering was used to cluster cities into 6 clusters. We created a new pandas dataframe that includes the clusters as well as the top 10 venues for each city.

**RESULTS**

Below is a pandas dataframe showing the top 10 venues for each city and the clusters they belong to.

A screenshot of a cell phone

Description automatically generated

Fig 6: Top 10 venues and clusters.

Now, let’s visualize the resulting clusters.

A picture containing text, map

Description automatically generated

Fig 7: Clustered cities.

After examining each cluster to determine the discriminating venue categories that distinguish each cluster, the results are shown below:

A screenshot of a cell phone

Description automatically generated

Fig 8: Cluster 0.

A screenshot of a cell phone

Description automatically generated

Fig 9: Cluster 1.

A screenshot of a cell phone

Description automatically generated

Fig 10: Cluster 2.

A screenshot of a cell phone

Description automatically generated

Fig 11: Cluster 3.

A screenshot of a cell phone

Description automatically generated

Fig 12: Cluster 4.

A screenshot of a cell phone

Description automatically generated

Fig 13: Cluster 5.

**DISCUSSIONS**

For the first cluster, which is cluster 0, you can see that the most common venues are convenient stores and banks. There are no nearby hotels in cluster 0. Therefore, potential investors who are interested in starting up a hotel business can start in cities that are in cluster 0. Cities in this cluster will not attract tourists because there are no hotels nearby for tourists to stay in.

Cities that make up cluster 1 have restaurants as the most common venues and a lot of tourist centers. These cities will attract a lot of tourists.

Cluster 2 has hotels, tourist centers and departmental stores as its most common venues. Cities that make up this cluster will very well attract a lot of tourists and investors.

Cluster 3 has a lot of venues that will attract tourists, but there are very few hotels in these cities that make up cluster 3. Potential investors who are interested in the hotel business can invest in these cities.

Cluster 4 encompasses only one city. The most common venue in this cluster are bars. The city in this cluster will attract a lot of potential investors.

Cluster 5 is made up of only one city just like cluster 4. Camping grounds are the most common venues in this cluster. This city will attract a lot of potential investors.

**CONCLUSION**

Even though the venues were limited to 1000 venues and within a radius of 5000 metres, enough venues were not generated by the Foursquare location data. This is as a result of the Foursquare API location data not working well with the Nigeria dataset.

However, with the few venues that were explored, we were able to give some information about major cities in Nigeria.

**REFERENCES**

1. https://www.opec.org/opec\_web/en/about\_us/167.htm

2. https://www.worldometers.info/world-population/nigeria-population/