Capstone Project – The Battle of Neighborhoods | Finding a Conducive Place in Lagos

1.0 Introduction

The purpose of this Capstone Project is to help people in traversing better facilities around their neighbourhood. It will help people executing a smart and efficient decision on choosing a great neighbourhood out of numbers of other neighbourhoods in Lagos, Nigeria.

Lots of people are migrating to various states in Nigeria and needed lots of research for a conducive place for their family. This project is for those people who are looking for better neighbourhoods. For ease of accessing to Cafe, School, Supermarket, medical shops, grocery shops, mall, theatre, hospital, like-minded people, etc.

This Capstone Project aims to create an analysis of features for a people migrating to Lagos to search the best neighborhood as a comparative analysis between neighbourhoods. It will help people to get the awareness of the area and neighbourhood before moving to a new city, state, country or place for their work or to start a new fresh life.

2.0 Data Section

2.1 Data Link:

https://en.wikipedia.org/wiki/Category:Populated_places_in_Lagos_State

Will use Lagos dataset which we scraped from Wikipedia. The dataset consists of latitude and longitude, neighbourhoods.

2.2 Foursquare API Data:

We will need data about different venues in different neighbourhoods of that specific borough.

"Foursquare" locational information was used to gain that information. Foursquare is a location data provider with information about all manner of venues and events within an area of interest. Such information includes venue names, locations, menus and even photos. As such, the foursquare location platform will be used as the sole data source since all the stated required information can be obtained through the API.

After finding the list of neighbourhoods, we then connect to the Foursquare API to gather information about venues inside each and every neighbourhood. For each neighbourhood, we have chosen the radius to be 100 meter.

The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes. The information obtained per venue as follows:

1. Neighborhood

2. Neighborhood Latitude

3. Neighborhood Longitude

4. Venue

5. Name of the venue e.g. the name of a store or restaurant

6. Venue Latitude

7. Venue Longitude

8. Venue Category

2.3 Map of North York
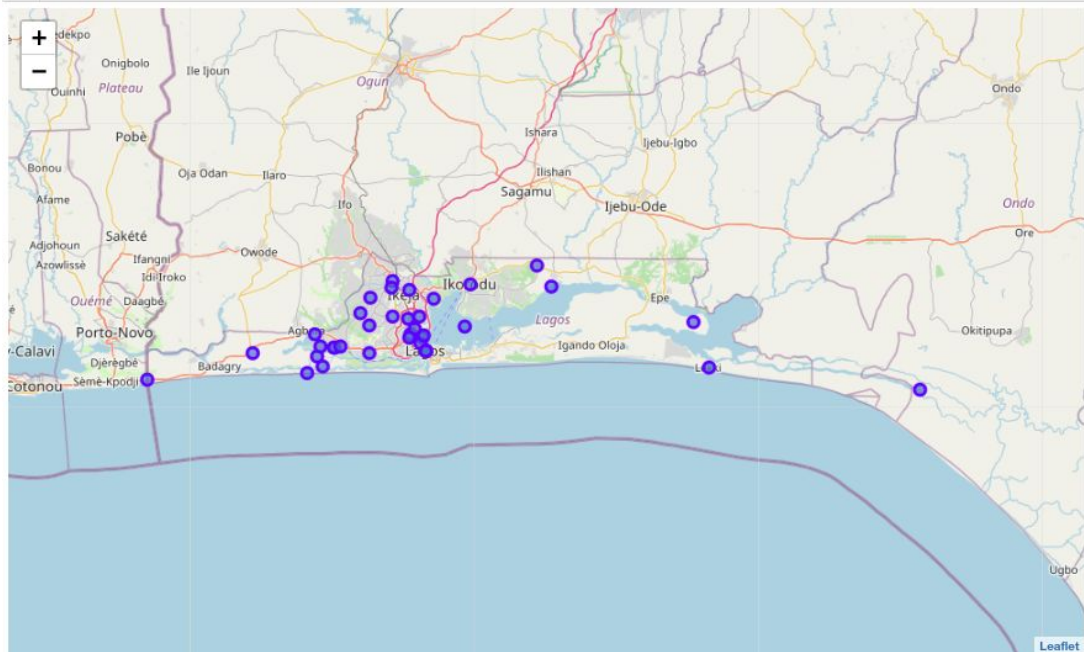
3.0 Methodology

Clustering Approach:

To compare the similarities of two cities, we decided to explore neighbourhoods, segment them, and group them into clusters to find similar neighbourhoods in a big city like New York and Toronto. To be able to do that, we need to cluster data which is a form of unsupervised machine learning: k-means clustering algorithm.

Using K-Means Clustering Approach | Most Common Venue

```
In [209]:  1  # add clustering labels
           2  neighborhoods_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)
           3
           4  Lagos_merged = lagos
           5
           6  # merge manhattan_grouped with manhattan_data to add latitude/longitude for each neighborhood
           7  Lagos_merged = Lagos_merged.join(neighborhoods_venues_sorted.set_index('Neighborhood'), on='Neighborhood', how='
           8
           9  Lagos_merged.head() # check the last columns!
```

Out[209]:

	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Adeniran Ogunsanya Street, Lagos	6.494436	3.357379	2	Shopping Mall	Stadium	Clothing Store	Convenience Store	Multiplex
1	Agege	6.625256	3.311209	2	Business Service	Bus Station	Clothing Store	Food Court	Food & Drink Shop
2	Ajangbadi	6.463640	3.170214	0	Hotel	Food	Women's Store	Clothing Store	Food Court
3	Akesan, Lagos State	6.547469	3.232404	2	Convenience Store	Women's Store	Fried Chicken Joint	Food Court	Food & Drink Shop
4	Akowonjo	6.610479	3.310308	1	Airport Terminal	Bank	Fast Food Restaurant	Bus Station	Women's Store

Most Common Venues near Neighborhood | Using Clustering

Most Common venues near neighborhood

```python
num_top_venues = 5

indicators = ['st', 'nd', 'rd']

columns = ['Neighborhood']
for ind in np.arange(num_top_venues):
    try:
        columns.append('{}{} Most Common Venue'.format(ind+1, indicators[ind]))
    except:
        columns.append('{}th Most Common Venue'.format(ind+1))

neighborhoods_venues_sorted = pd.DataFrame(columns=columns)
neighborhoods_venues_sorted['Neighborhood'] = Lagos_grouped['Neighborhood']

for ind in np.arange(Lagos_grouped.shape[0]):
    neighborhoods_venues_sorted.iloc[ind, 1:] = return_most_common_venues(Lagos_grouped.iloc[ind, :], num_top_ve

neighborhoods_venues_sorted.head()
```

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Adeniran Ogunsanya Street, Lagos	Shopping Mall	Stadium	Clothing Store	Convenience Store	Multiplex
1	Agege	Business Service	Bus Station	Clothing Store	Food Court	Food & Drink Shop
2	Ajangbadi	Hotel	Food	Women's Store	Clothing Store	Food Court
3	Akesan, Lagos State	Convenience Store	Women's Store	Fried Chicken Joint	Food Court	Food & Drink Shop
4	Akowonjo	Airport Terminal	Bank	Fast Food Restaurant	Bus Station	Women's Store

Work Flow:

Using credentials of Foursquare API features of near-by places of the neighbourhoods would be mined. Due to HTTP request limitations, the number of places per neighborhood parameter would reasonably be set to 100 and the radius parameter would be set to 500.

would be set to 500.

4.0 Result Section

4.1 Map of Clusters in North York

Out[230]:
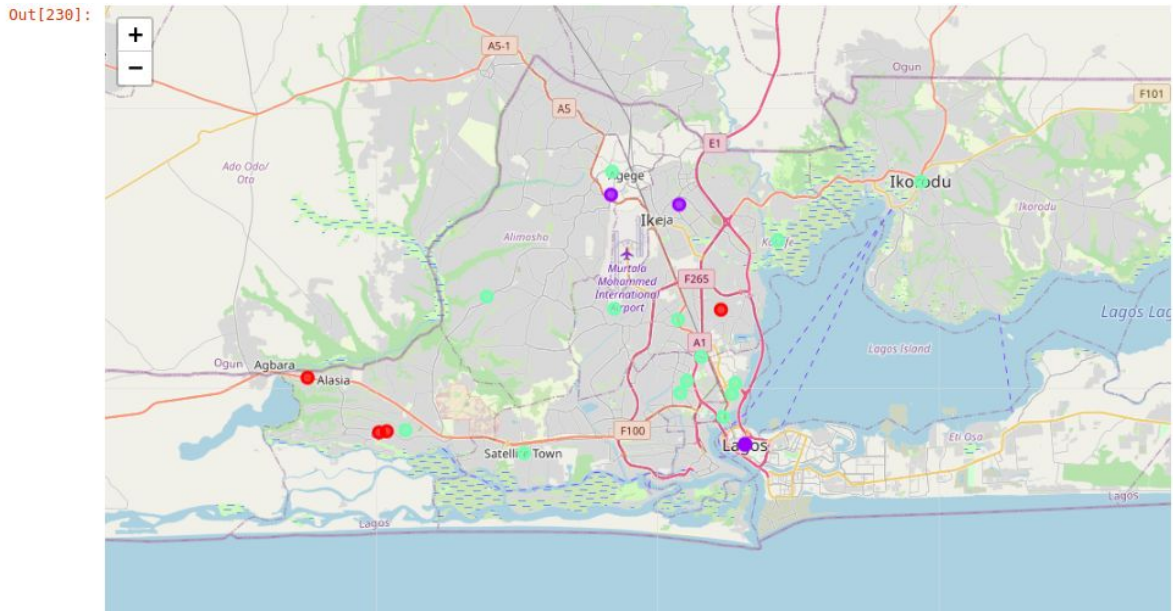
4.2 The Location:

Lagos is the most populous city in Nigeria and the African continent. Lagos is a major financial centre for all of Africa and is the economic hub of Lagos State. The megacity has the fourth-highest GDP in Africa and houses one of the largest and busiest seaports on the continent. It is one of the fastest-growing cities in the world.

4.2 Foursquare API:

This Capstone project has used Four-square API as its prime data gathering source as it has a database of millions of places, especially their places API which provides the ability to perform location search, location sharing and details about a business.

5.0 Discussion Section

5.1 Problem Which Tried to Solve:

The major purpose of this project is to suggest a better neighbourhood in a new city for the person who is shifting there. Social presence in society in terms of like-minded people. Connectivity to the airport, bus stand, city centre, markets and other daily needs things nearby.

- Sorted list of the house in terms of the possible neighbourhood with a good standard of living

6.0 Conclusion

In this Capstone project, using k-means cluster algorithm I separated the neighborhood into 10(Ten) different clusters and for 103 different latitudes and longitude from the dataset, which have very-similar neighborhoods around them. Using the charts above results presented to a particular neighbourhood based on average house prices and school rating have been made.

I feel rewarded with the efforts and believe this course with all the topics covered is well worthy of appreciation.

This project has shown me a practical application to resolve a real situation that has impacting personal and financial impact using Data Science tools.

The mapping with Folium is a very powerful technique to consolidate information and make the analysis and decision better with confidence.

6.1 Future Works:

This Capstone project can be continued for making it more precise in terms to find the best house in Lagos. Best means on the basis of all required things around and also in terms of cost-effectiveness.

Libraries Which are Used to Develop the Project:

- *Pandas: For creating and manipulating data frames.*
- *Folium: Python visualization library would be used to visualize the neighborhoods cluster distribution of using interactive leaflet map.*
- *Scikit Learn: For importing k-means clustering.*
- *JSON: Library to handle JSON files.*
- *XML: To separate data from presentation and XML stores data in plain text format.*
- *Geocoder: To retrieve Location Data.*
- *Beautiful Soup and Requests: To scrap and library to handle http requests.*
- *Matplotlib: Python Plotting Module.*