ACCESSING THE POPULATION SIZE, DENSITY AND VENUES INFORMATION OF LAGOS NEIGHBOURHOODS

OBAFEMI ADEPOJU

MARCH 3, 2020

1.   INTRODUCTION

1.1   Background

Lagos is a state located in the south-western part of Nigeria. It is the country's most important commercial centre and the focal point of economic activities. Due to the economic prospects of Lagos and other factors, the state is very well populated.

Lagos is the most populous city in Nigeria despite it being the smallest state in the country. Its population accounts for about 27.4% of the national estimate. According to 2016 statistics, Lagos has a population size of 21 million and it is continually growing. It has been estimated that by 2050, its population will have doubled.

The large population of the state compared to its small size has created a problem that some stakeholders have termed as overpopulation. This problem has given rise to another problem - high traffic congestion in the state. While Lagos has population problem, not all parts of it are densely populated. There are 20 Local Government Areas in Lagos. These can be viewed as neighbourhoods for the administration of the state and they have districts under them.

1.2   Problem

It has been known that not all neighbourhoods of the state are overpopulated, however, it is difficult for most indigenes and travelers or migrants to know which part of the state is less dense in population so that they can plan ahead armed with vital information about where they wish to stay in the city.

Also, Lagos is a fast growing centre with many important venues for different activities. However, there is no easy way to access this vital information by anyone who needs it.

There is therefore, a need to have easy and intuitive means of accessing information about the population density of the different neighbourhoods (Local Government Areas) as well as postal codes and venues information. This project aims to create a visual access to the population distribution of Lagos neighbourhoods including a map that will incorporate the vital information with segmentation and clustering analysis.

1.3 Interest

Migrants, travelers, business investors and government will benefit from the ease of access to vital information that this project will provide.

2. DATA DESCRIPTION

The data for this project was obtained in unstructured format. A table was produced manually from information sourced from official documents and from the internet. The data requirement was divided into four parts based on how they were obtained.

1. Postal codes of the different neighbourhoods (Local Government Areas);
2. Lagos State official population data listing the Local Government Areas (neighbourhoods) with their population densities;
3. Geo-coordinates of all neighbourhoods of Lagos;
4. Venues information of the different neighbourhoods.

Postal codes of the neighbourhoods were obtained from different internet sources but mostly from nigeriapostcodes.com The codes were important as they distinguish one neighbourhood from the other and help in generating coordinates.

A crucial data for this project is the population data which gives all statistics about population size and density (per square kilometer, sq.km) of the 20 neighbourhoods. This data was obtained from the Lagos State Bureau of Statistics 2017 official document. The 2017 version of the document was the most recent version available for download at the time of working on this project. This data will be used on the map where it will be placed superimposed for all neighbourhoods on the map.

Geo-coordinates of the 20 neighbourhoods were all obtained online from different sites. There was no single official source for this data, hence, they had to be scraped from different sites one by one. This data will be used to locate the neighbourhoods on the map of Lagos State.

Lastly, venues information about the neighbourhoods were obtained using Foursquare API. The information obtained include name of the venues and the categories. The data will be cleaned, processed and passed through unsupervised machine learning for segmentation and clustering analysis.

All datasets obtained were entered into a table to generate a .csv file which was used to create a pandas dataframe for descriptive analysis and machine learning modeling.

3. METHODOLOGY

All analyses and visualizations were carried out using python programming language.

3.1   Data Setup

Github was used as the repository to host the prepared dataset before commencing analysis. The dataset had 6 columns namely: Postal Code, Neighbourhoods, Population, Population density, Latitude and Longitude. The columns contained information specific for the Neighbourhoods organized in each row.

	Postal Code	Neighbourhoods	Population	Population density	Latitude	Longitude
0	100283	Agege	1,415,547	60,768.47	6.619830	3.322165
1	102103	Ajeromi Ifelodun	1,966,700	103,258.63	6.455480	3.333860
2	100275	Alimosho	2,804,919	14,855.05	6.610556	3.295830
3	102102	Amuwo Odofin	719,337	2,931.16	6.450000	3.266667
4	102241	Apapa	715,792	13,568.42	6.449998	3.366665

Fig 1. A view of the first 5 rows of the dataset in a pandas dataframe

3.2   Exploring Population Density through Visualization

To have a visual understanding of the Population density of Lagos, bar charts and histograms were plotted using the Neighbourhoods and Population density data columns. The graphs were plotted with matplotlib using the scripting layer. For the histogram, in order to have well formatted x-axis ticks, numpy library was used to create the bin_edges and tick counts.

3.3   Getting Coordinates

To create the first dataset, the Latitude and Longitude data of the Neighbourhoods were obtained from online GIS sites such as latitude.to, latitude.net and wikipedia. However, for the coordinates of Lagos State, geocoder library was used in a script and the values returned were assigned to a map variable for visualization.

3.4   Map creation - Map 1

The map of Lagos was created using folium library of python. The Neighbourhoods were placed superimposed on the map with their vital information accessible on click. The information such as Neighbourhood name, Postal Code, Population size and Population density were passed as parameters which were plotted on the map.

3.5   Venues data gathering

Foursquare API was used for venues data harvesting. The 20 Neighbourhoods were explored with Foursquare for the listing of their different venues. The API required passing in valid credentials along with parameters such as limit and radius. For this project, the limit set was 100, but it was not even approached in the results generated from the API. Also, the radius of exploration was set to 2000m because at a low value of 500m, less than 50 venues were returned for all the neighbourhoods combined. The venues were organized in categories so that each unique category of the results were identified.

3.6   Venues data analysis

Exploratory analysis was carried out on the venues data. The venues were grouped by their Neighbourhoods with their respective counts using the pandas *groupby* method. A mean of each venue category was determined per Neighbourhood to know their frequency of occurrence. With the frequency determined, a listing of top venues per Neighbourhood was done by sorting the

frequency in descending order. Then the Neighbourhoods were listed in a table with their most common venues arranged in descending order row by row.

3.7   Clustering analysis

The processed venues data was passed through kmeans clustering unsupervised machine learning algorithm in order to partition the Neighbourhoods into groups with similar characteristics. For this project, 5 clusters were used to partition the Neighbourhoods. The cluster labels were generated from the model and then assigned to the corresponding Neighbourhoods. The members of each of the 5 clusters were identified and grouped separately. Based on the unique characteristics of members of the clusters, a suitable name was assigned.

3.8   Map creation - Map 2

After clustering was completed, another map was generated showing all 20 Neighbourhoods. To easily visualize the cluster that each Neighbourhood belonged to, colour coding was done such that 5 unique colours represented each of the 5 clusters. So all members belonging to a particular cluster all had the same colours while those of other clusters had their respective colours.

4.   RESULTS

4.1   Population density across Neighbourhoods

A bar chart of the population densities showed the neighbourhoods that were experiencing overpopulation and those that were less populated.
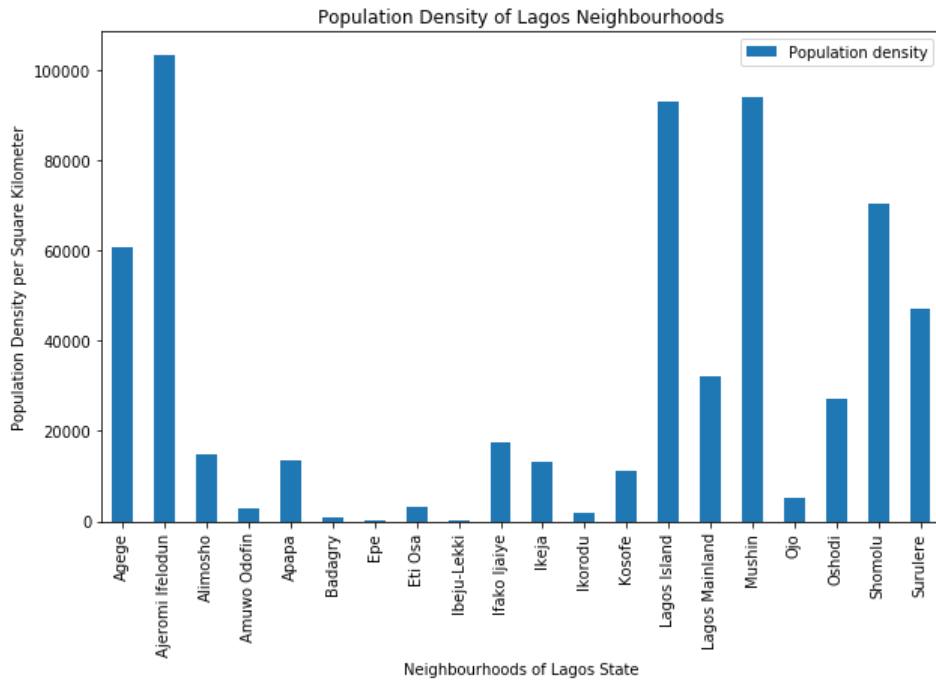
Fig. 2.　　Bar chart of Population density vs Lagos Neighbourhoods

In order to know the number of overpopulated Neighbourhoods compared to those with low population, a histogram was plotted.
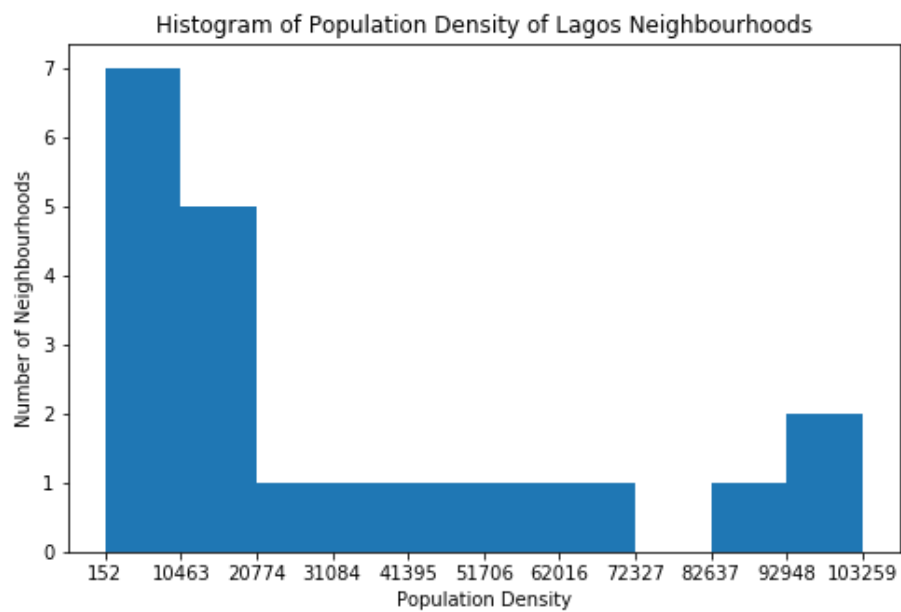

Fig. 3. Histogram of Lagos Neighbourhoods Population density

The histogram shows that out of 20 Neighbourhoods, **12** were actually of low population densities, **3** were overpopulated and the others were of medium densities.

4.2 MAP 1

A map of Lagos showing all 20 Neighbourhoods represented as blue points which when clicked displays a popup revealing vital information such as Neighbourhood name, postal code, population size and population density. The map shows that some Neighbourhoods are quite far from others such as Badagry and Epe while many others are close together.
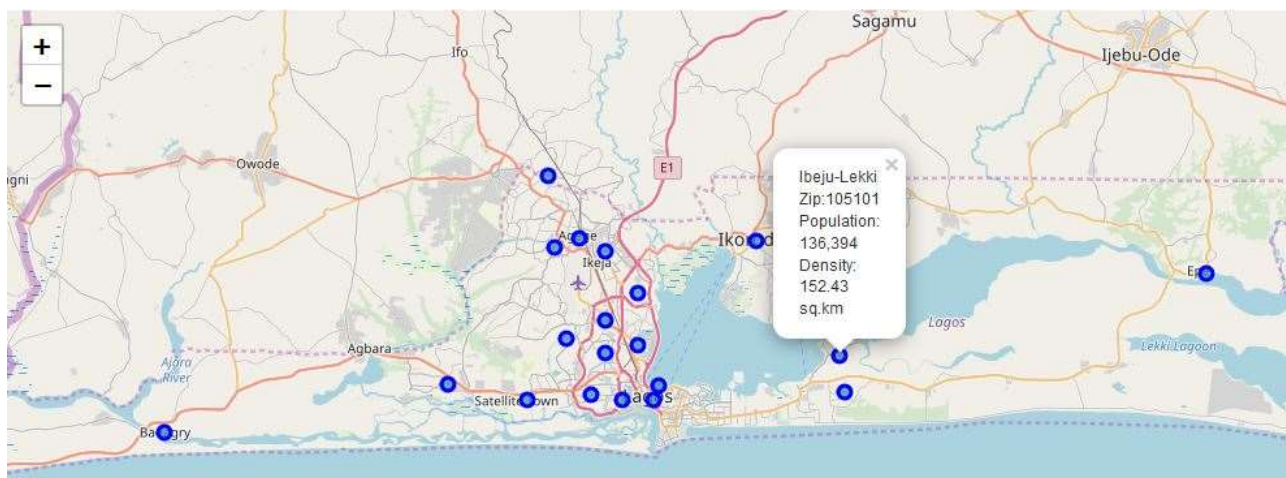

Fig. 4. Map of Lagos 20 Neighbourhoods with vital information displayed in a popup

4.3 Venues Data

After querying the Foursquare API for venues of each Neighbourhood through passage of their respective coordinates to the API, just a few venues were returned. Even though the radius was set to a high value of 2000m contrary to that set for Neighbourhoods in other locations (which usually ranged from 500m - 750m), the returned results was still low.

A total of **234** venues and **77** categories were returned for the 20 Neighbourhoods queried. **Ikeja** Neighbourhood had the highest number of venues returned followed by **Lagos Island** & **Lagos Mainland** with values of 38, 28 & 28 respectively (the second and third Neighbourhoods sharing the same venue count).

Fig. 5. Venue count for Neighbourhoods with highest values

Epe Neighbourhood had the lowest number of returned venues with a value of 1.

Fig. 6. Venue count for Neighbourhood with least value

A view of top 5 venues returned for Neighbourhoods with high counts (Ikeja, Lagos Island and Lagos Mainland)

Fig. 7. Top 5 venues for Neighbourhoods with highest venue counts

4.4 Kmeans cluster analysis

The result of kmeans clustering of the processed venues data is shown in the cluster labels returned for each Neighbourhood.

	Postal Code	Neighbourhoods	Population	Population density	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
0	100283	Agege	1,415,547	60,768.47	6.619830	3.322165	0	Bus Station	Fast Food Restaurant	Market	Convenience Store	Shopping Mall	Wine Shop
1	102103	Ajeromi Ifelodun	1,966,700	103,258.63	6.455480	3.333860	0	Bus Station	Harbor / Marina	Rest Area	Chinese Restaurant	Coffee Shop	Convenience Store
2	100275	Alimosho	2,804,919	14,855.05	6.610556	3.295830	0	Bus Station	Fast Food Restaurant	Park	Boutique	Campground	Bus Stop
3	102102	Amuwo Odofin	719,337	2,931.16	6.450000	3.266667	2	Hotel	Harbor / Marina	Park	Electronics Store	African Restaurant	Spa
4	102241	Apapa	715,792	13,568.42	6.449998	3.366665	1	Shopping Mall	Indian Restaurant	Boat or Ferry	Pier	Pizza Place	Playground

Fig. 8. Neighbourhoods with their cluster labels after kmeans clustering

4.5 Map after clustering

The map of Lagos recreated with the new cluster labels generated shows the Neighbourhoods with similar characteristics. Agege, Ajeromi Ifelodun & Alimosho are some Neighbourhoods clustered together.
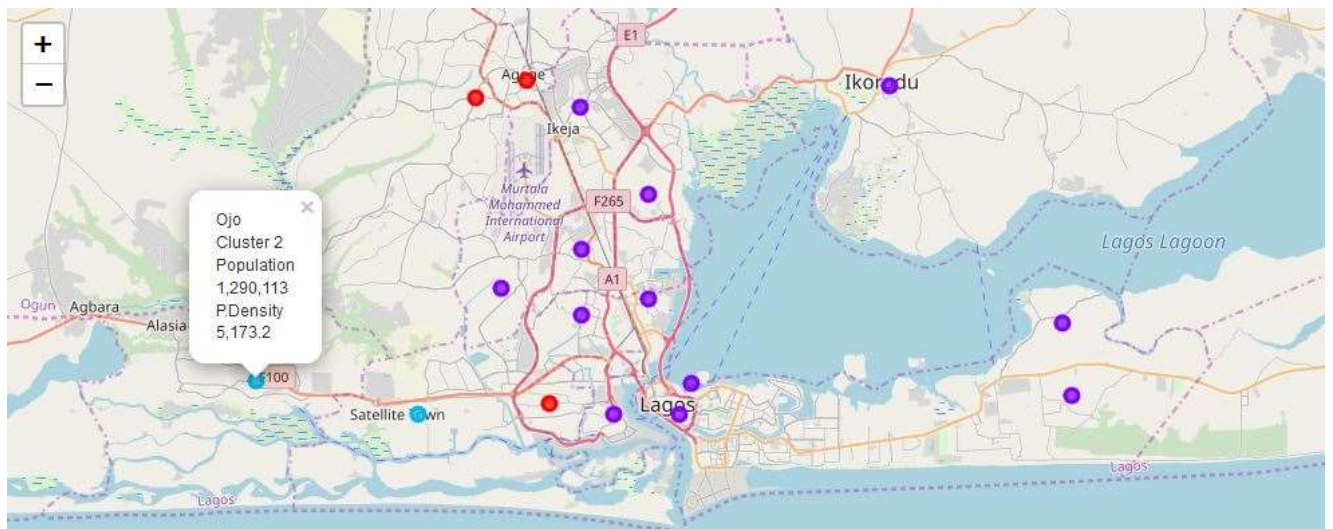

Fig. 9. Map showing Neighbourhoods with similar characteristics having the same colour

Following the Neighbourhood clustering, the five clusters created were characteristically named as follows:

Cluster 0 – Social and Business Hub

Cluster 1 – Hospitality, Trading and Events Hub

Cluster 2 – Hospitality and Business Hub

Cluster 3 – Business Hub and Eatries

Cluster 4 – Historical centre

5. DISCUSSION

Lagos is a highly populated city with 20 Local Government Areas referred to as Neighbourhoods. The Neighbourhoods of Lagos have varying population densities showing that the concentration of people is localized, not spread across the state. Therefore, the problem of overpopulation can be explored in terms of the particular Neighbourhoods that are experiencing it rather that saying the entire state is overpopulated.

From the bar chart and histogram, it was noted that there were more less-population-dense Neighbourhoods than more-dense ones. The factors causing density in the different Neighbourhoods can be tackled by Government stakeholders to allow for more even spread of the population.

The first map created showed all the vital information about Lagos Neighbourhoods. The information was accessible upon clicking of the dots on the map. This will prove very useful to anyone such as a traveler, migrant or even an investor looking for fast and easy access to information about the population distribution of Lagos Neighbourhoods.

The venues result for the Neighbourhoods was quite poor given the radius specified. The low result can be attributed to insufficient map data available in the Foursquare database for Lagos Neighbourhoods. The fact remains that Lagos is still a developing city in a developing country. Therefore, it will take some time before the map presence of all venues become explorable on maps. Also, another factor that might have impacted the result was the coordinates for each

Neighbourhood specified. It means that if the coordinates are adjusted slightly to the centre of the Neighbourhoods where venues are located, then higher venue counts might be returned.

Kmeans clustering helped to group similar Neighbourhoods together on the map. This segmentation would prove useful to anyone looking for insight into the type of venues that are trending in any Neighbourhood they may be interested in within Lagos.

6.   CONCLUSION

The use of a map to explore and visualize the population statistics of Lagos Neighbourhoods is a very effective way of providing fast and easy access to vital information for anyone that might be interested in it. It will help to foster geographical literacy and enable informed planning for migration, settling and business establishment purposes.

Lagos Neighbourhood clustering helped to provide good insight into the kind of venues common to Neighbourhoods and the viewing of Neighbourhoods of similar characteristics at a glance.