**SHOMOLU LOCAL GOVERNMENT AREA: AN INDEPTH EXPLORATORY DATA ANALYSIS USING PYTHON**

**(Pandas and Seaborn)**

**TABLE OF CONTENTS**

1. THE STUDY AREA (SHOMOLU LGA)
2. METHODOLOGY
3. EXPLORATORY DATA ANALYSIS
4. CONCLUSION

* **Do you know that a large majority of the residents in Shomolu LGA are traders?**
* **Do you know that there are more females involved in Trading/Commerce in Shomolu LGA than males?**
* **Do you know that the average house rent paid per month is between 5000 and 10000 naira monthly?**
* **Do you know that majority of Shomolu’s residential population are youths within the age group of 30 – 40 years, closely followed by the 21 – 30 age bracket?**
* **Do you know that a large majority of residents in Shomolu LGA are tenants?**
* **Do you know that Most landlords or owners of dwelling places are married and are over 50 years of age?**
* **Do you know that most of the artisans in Shomolu earn between #20,000 and #40,000 per month?**

**THE STUDY AREA**

Shomolu, also spelled “Somolu” town, is a residential suburb located in the north of Lagos state in southwestern parts of Nigeria. The town is plagued by problems of overcrowding, poor housing and inadequate sanitation. Its local activities include work in leather, handicrafts and printing.

The population of the LGA according to 2006 census is 402,673 persons. However, Shomolu local government area was formerly known as Mushin East local government. The present Somolu Local Government comprises areas like Community Road, Akoka, areas east of Ikorodu road up to anthony Oke side interchange, including Somolu, Bashua, Bariga, some parts of Akoka, Igari, Obanikoro, Pedro village, Abule Okuta, Seriki village, Apelehin, Ilaje.

Shomolu Local Government is essentially inhabited by the Yorubas’, prominent among them are the Ijebus, Egbas, Awories and Ilajes. Other zoruba ethnic groups such as Oyo, Osun and Ekiti are represented in the area. However, other ethnic groups from the East and Northern parts of the country are equally large in number in the Local Government Area.

While the community in time past has been embroiled in violence and disorder (one which could equally be attributed to the failure of the society to provide viable alternatives), a new and different mentality is feeding a much more proactive and positive temperament. To the people of Shomolu, the canal which runs from Shomolu to Folagoro, Akoka, etc. is a symbolic divider. It is popularly known as “middle belt” by the indigenes as it serves as neutral ground and often a space of demarcation between the two most important factions of the community’s gangs.

Shomolu is the biggest printing hub in Nigeria and is complemented with other commercial activities such as trade of foodstuffs, utilities etc.

Residents of the study area practice Christianity, Islam and few percentages are traditional rulers.

The educational system in Shomolu cannot be termed poor asLagos state government operates state schools and the education system is the 6-3-3-4, which is practiced throughout the whole country. The levels are primary, junior secondary school, senior secondary school and university. List of some of the schools in Shomolu include; Binta Girls High School, CMS Grammar School, Igbobi College

There are not too many landmarks in Shomolu since its evolution, some of the few includes;

* **CMS Grammar School, Lagos**

The CMS Grammar school in the Bariga district of Lagos is the oldest secondary school in Nigeria, founded on the 6th of June 1859 by the Church Missionary Society. For decades it was the main source of African Clergymen and administrators in the Lagos colony.

* **Yaba College**

Yaba Higher College was founded in 1932 in Yaba, now a suburb of Lagos in Nigeria to provide tertiary education to Africans, mostly in vocational subjects and teaching. The College staff were transferred to start the University of Ibadan in 1948 and the college premises were used for the new Yaba College of Technology.

* **Bariga**

Bariga is a district and a suburb in Shomolu local government area of Lagos State. It is known to be the location of the oldest secondary school in Nigeria.

**2.0 METHODOLOGY**

The dataset used in this Exploratory analysis was primarily collected by Students of the department of Urban and Regional Planning set of 2015/2016 academic year, University of Lagos, while in the process of preparing a master plan for Shomolu LGA in 2018.

The dataset stems from a set of 181 questionnaires administered to the residents of Shomolu LGA in March of 2018. Evan Morris formula was used in calculating the sample size. Which was gotten to be approximately **181** out of a total of 631,857 persons as at March 2018. **Random sampling technique** was used to administer the questionnaires across the local government area.

The **python programming language** was the major tool used in the exploratory analysis of the dataset acquired. Other tools used includes:

**Pandas:** A python data analysis library which is used primarily for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical/categorical tables and time series.

**Jupyter notebook**: A jupyter notebook is an interactive python notebook that allows you to run your code in sequences as well as create and share documents that contain live code, equations, visualizations and narrative texts. You can decide and specify which cell runs first in the notebook environment. JUPYTER is actually an acronym for 3 languages. Julia, Python and R. it was built primarily for these three languages.

**Seaborn**: Python’s library for expert visualizations. This library handled virtually all of the visualizations in this exploratory analysis. Basically it provides high level interface for drawing attractive and informative statistical graphics.

Other options that could be used for this exploratory analysis are the SPSS software or the Microsoft Excel package but were discarded, because it’s more strenuous to work with SPSS having to deal with encoding the values in a spreadsheet and decoding it for visualizations, and its visualizations are very poor, unlike the python’s data visualization libraries. While the Microsoft Excel Package is fair with its visualizations, it gets slow when loaded with huge datasets.

**3.0 EXPLORATORY DATA ANALYSIS**

The dataset stems from a set of 181 questionnaires administered to the residents of Shomolu LGA in March of 2018. Evan Morris formula was used in calculating the sample size, which was gotten to be **181.** Random sampling technique was used to administer the questionnaires.

Loading the dataset into a pandas DataFrame. Let’s explore!

Calling the pandas **info()** method to the dataset

The dataset set comprises of **181 rows** and **60 columns**. For simplicity, the dataset would be stratified into the following broad indicators and would be explored accordingly. These indicators include:

1. Socio-Economic Characteristics
2. Land Use Characteristics
3. Infrastructures and Facilities
4. Source of Attraction to the study area
5. Education
6. Health
7. Security
8. Social Vices
9. Tax
10. Housing condition

For an in-depth analysis, these indicators would be explored sequentially in a series of post. Starting from the Socio-economic characteristics which would discussed and visualized here, down to the average Housing conditions.

**Socio-economic characteristics**:

This looks at the socioeconomic situation of people living in Shomolu; **Education**, **employment** **status**, **household income per month**, **occupation**, **mode of** **transportation, age** and **sex distributions** among others will be duly explored. Running a quick value count on these indicators we notice the following:

**Occupation**: The occupation distribution of residents in Shomolu LGA includes **Traders**, **Artisans**, **civil servants**, **Industrial workers**, and the **retired**. With a large majority as **Traders** (about 60%) followed by Artisans.

**Rent per month and Tenancy type**: A large majority of the residents in Shomolu are **Tenants**, from the data collected. Very few are landlords or people living in inherited apartments. The **rents** paid per month ranges from #**5000** up until **#30000** and above. With a large portion of the residents paying between **#5000** and **#10000** as housing rent monthly.

**Age and Sex:** More males were sampled than females, very few of the respondents were above 50 years, the most common age group in this region is between 30 – 40 years. Followed closely by the 21 – 30 age group. It is sufficing to say that Shomolu has more youths in population than the older generation.

**Household Income Per Month:** The income generated by each household in the study area ranges from **below #20,000** to **above #100,000.** In this distribution, there are two peaks, which are the below **#20,000 group** and **the #20,000 - #40,000**. With that being said, it is also worthy to note from the analysis that the most frequent mode of transportation in Shomolu LGA is public transportation i.e buses, motorcycles, tricycles popularly called **maruwa**

**Now to some serious socio-economic exploratory analysis**.

Visualizing the relationship between **Marital Status** and **Age** in shomolu LGA, utilizing Seaborn’s insightful **countplot()** and **catplot()** methods, the following are made evident:

1. The marital status distribution ranges from married, single, divorced and Separated.
2. More people in Shomolu LGA are married than single
3. Most married people in Shomolu are within the age range **of 31 - 40 years** closely followed by **41 – 50 years’** age group.
4. Nobody Under 20 years of age in Shomolu LGA is married.
5. Very few people are either divorced or separated.

Really fascinating isn’t it?

Adding another indicator **Tenancy Type** to the plot and visualizing it, we also discover the following:

1. A very large majority of the people living in Shomolu LGA are **tenants**
2. Most landlords or owners of dwelling places are married and are over 50 years of age.
3. Less than 5% of married couples between the Ages of 31 – 40 years inherited the houses they live in.

Say we want to see the relationship between **Occupation** and **Household income per month**

Seaborn’s **countplot()** to the rescue.

1. Firstly, we notice very glaringly that the **Trading/commerce** occupation takes the highest count across all Household Income Level Categories in Shomolu LGA, From below **#20,000** up until above **#100,000**. That is in accordance to the earlier stated fact that **majority of the residents of Shomolu are traders or involved with commerce.**
2. Another interesting deduction one can make from the visualization is that in the highest household income category i.e #100,000 and above industrial workers and other occupations actually come to the limelight
3. And lastly, most of the Artisans in Shomolu earn between **#20,000** and **#40,000**.

There are still loads of information to be mined from this dataset, but I would leave that to you.

Finally checking the relationship between **Gender** and **Occupation** distribution in Shomolu LGA still using seaborn’s **countplot()**

From the beautiful visualization above, the following can be inferred:

1. There are more females involved in Trading/Commerce in Shomolu LGA than males.
2. There are more Males as civil servants than Females
3. More Males are industrial workers.
4. More Males are artisans than females in Shomolu LGA.

You could get so much more information from these plots though. but I would leave that to you.

**4.0 CONCLUSION**

.

And that’s all on the socioeconomic characteristics of residents in Shomolu LGA, in the next series of post we would be diving right into an in-depth exploratory analysis of the other indicators. The next post would particularly be on **Land use characteristics** in Shomolu LGA.

You can find the code to this article as well as the dataset on my github here,