Omdena
OPEN EDUCATION BY LOCAL CHAPTERS

Omdena is a global platform bridging mission-driven organizations with AI engineers, data scientists, and domain experts from diverse backgrounds to solve real-world problems. Omdena provides a streamlined collaborative environment to build innovative, ethical, and efficient AI and data science solutions.

Date: 26/8/2021

Omdena Lagos, Nigeria chapter is a part of Omdena that focuses on running open-source AI projects to solve challenges facing our local communities.

Date: 26/8/2021

Active collaborators

Opeyemi Adeniran: https://www.linkedin.com/in/opeyemi-taiwo-adeniran/
Edidong Esu : https://www.linkedin.com/in/edidiong-esu-1942a2129/
Ade-Onojobi Oluwadewalade: https://www.linkedin.com/in/dewalade/
Ismaila Lukman: https://www.linkedin.com/in/ismailukmanenegi/
Umar Fauwzziyyah: https://www.linkedin.com/in/fauzziyyah-ozohu-umar
Ajeyomi Adedoyin: https://www.linkedin.com/in/adedoyin-samuel-ajeyomi-83b80914b
Ayanlowo Babatunde: https://www.linkedin.com/in/ayanlola2002/
Egbigbe Osagie: https://www.linkedin.com/in/osagie-eboigbe-52b84794/
Ozeigbe Ojeifo: https://linkedin.com/in/ojeifo-oziegbe-516560146
Ayo Samuel: https://www.linkedin.com/in/sam-ayo
Nnaemeka Emmanuel: https://www.linkedin.com/in/emmanuel-nnaemeka-2b8378199/
Obembe David: https://www.linkedin.com/in/david-obembe-7b1636174
Kola-Adepoyigi Aderinsola:
https://www.linkedin.com/in/aderinsola-pj-kola-adepoyigi-733510109/
Kenneth Imade : https://www.linkedin.com/in/kenneth-imade/
Areago Adedayo: https://www.linkedin.com/in/adedayo-areago-2a6a94183/
Emmanuel Salako: https://www.linkedin.com/in/adekunle-salako-59b6801b1/

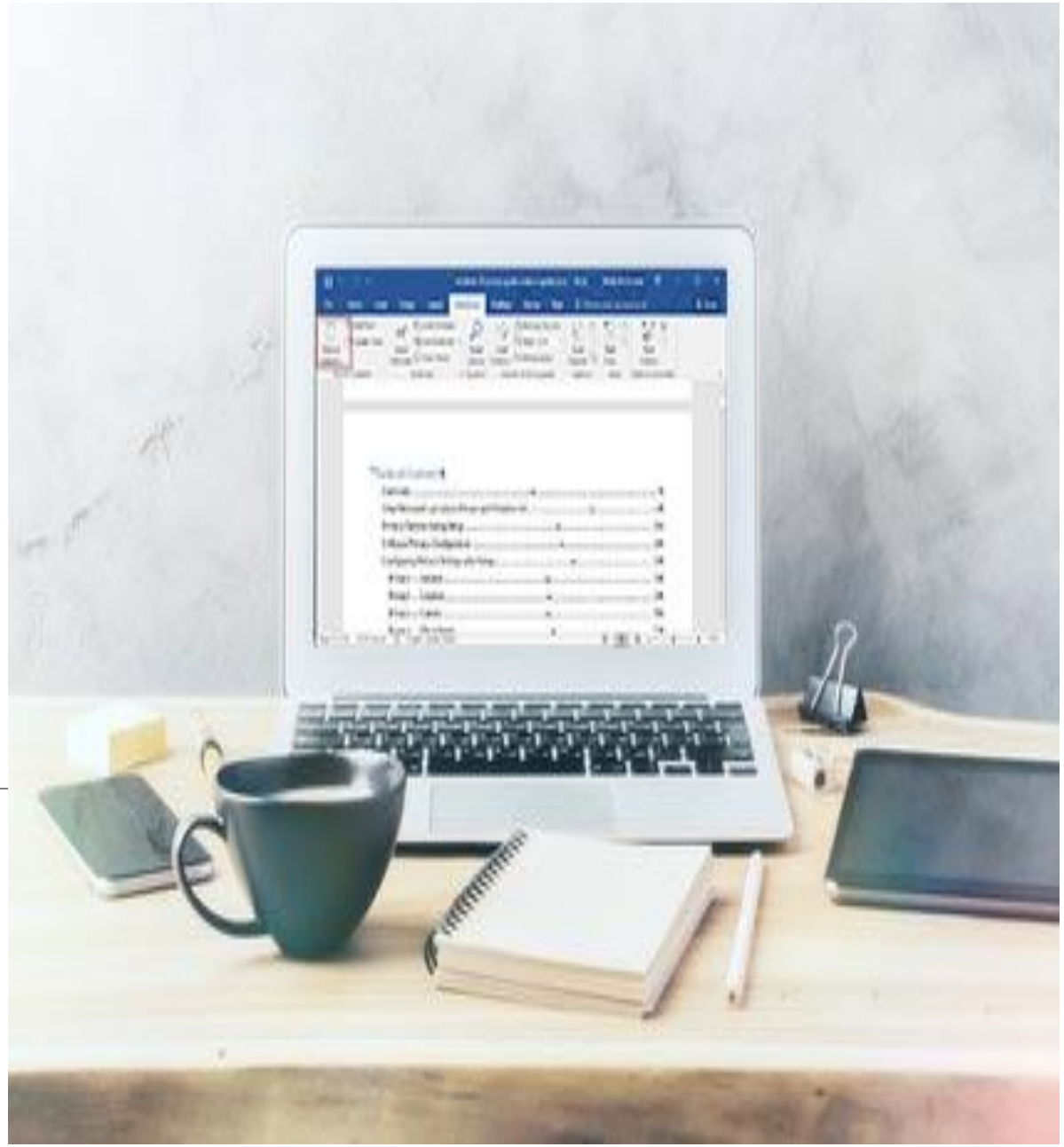
Omdena Lagos, Nigeria Chapter

Presents

A.i for Renewable Energy

Date: 26/8/2021

Table of Contents



1.0 Introduction

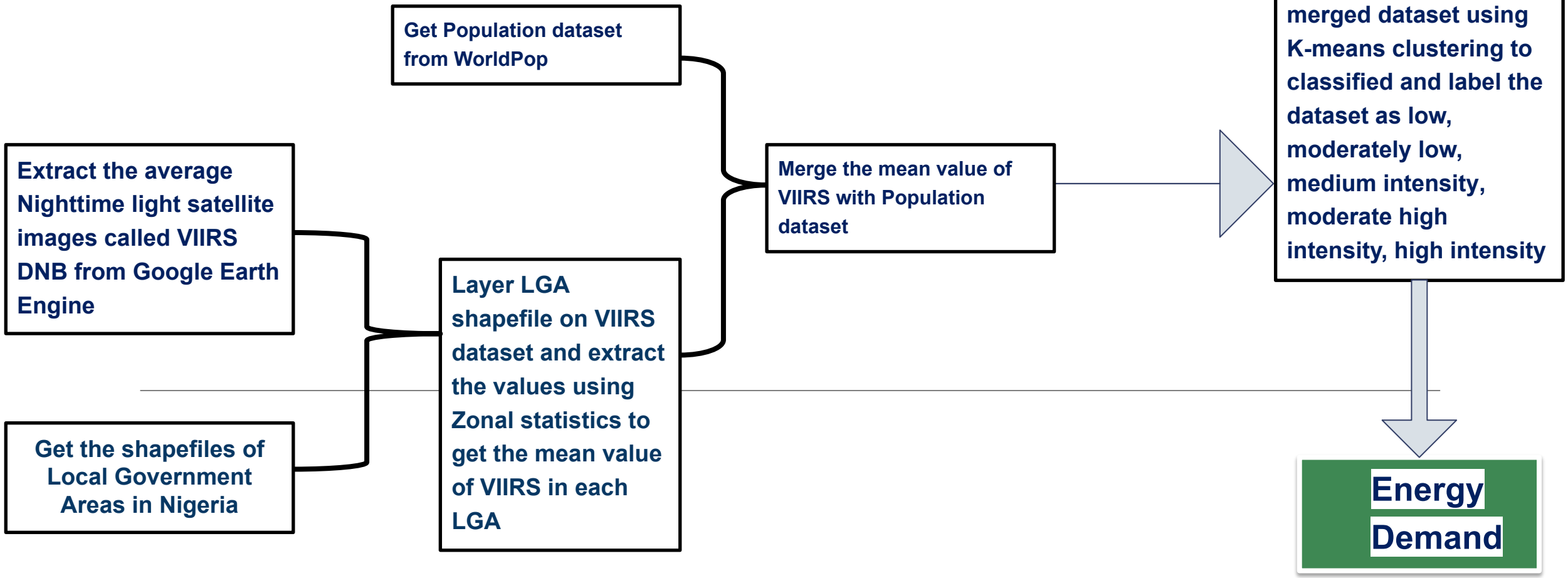
Despite the fact that Nigeria is one of the largest africa oil producer and one of the best producing oil countries in the world. Currently, only 45% of Nigeria's population is connected to the energy grid whilst power supply difficulties are experienced around 85% of the time and almost nonexistent in certain regions.

The governments can dramatically reduce their carbon footprint by purchasing or directly generating electricity from clean, renewable sources and installing it to regions without power supply or not connected to the national electricity grid . The most commonly used renewable energy source is solar which will be our focus for this project

2.0 Problem Statement

The Idea is to get the lists of top Nigerian regions with high demand for electricity and also to find the best spot for solar panel Installation.

3.0 Methodology



3.1 Data Extraction Implementation

The night time task team extracted night time averages for the administrative regions in Nigeria. For the purpose of this project, we chose our regions to be the Local government areas in Nigeria (774 in total).

The dataset used for the night time is the **VIIRS dataset** (VIIRS Nighttime Day/Night Band Composites Version 1 is the name on Google Earth Engine). This contained the year range used in the project. This dataset is a composition of night time image collections covering the every year from 2012 -2021.

A **shapefile** is a folder that contains administrative segmentations with its descriptions. For this task, the LGA shapefile was layered on the VIIRS data set and the values extracted using Zonal statistics.

A Zonal statistics operation is one that calculates statistics on cell values of a raster (a value raster) within the zones defined by another dataset (in this case LGAs shapefile). After layering the LGA shapefile on the VIIRS dataset, the **mean value** of each zone in the raster was calculated. The Geemap package has a class that helped hasten the process.

The population data of each LGA was extracted from World Pop Open Repository and merged with the average night time mean data. Data extracted was handed over to the data gathering team where they added more features to the data, preparing it to be modelled and Visualized.

3.2 Exploratory Data Analysis

- There were 774 LGAs in the country with 36 states the states having a minimum of 6 LGAs and a maximum of 44 LGAs
- The LGA has an average land area of 1173. 07 square kilometer with 75% of the LGAs having a land area less than 1519.81 square kilometer.
- The minimum and maximum land area coverage of the LGAs are 11.65 and 10357.67 square kilometer respectively.
- Kano appears to be the most populated state with 13,724,250 residents, while Bayelsa is the least populated state having 1,099,952 residents.
- Lagos state has the highest Average Nighttime light mean of 159.433944 as was expected while Bauchi state has the least Average Nighttime light mean of 0.052742.

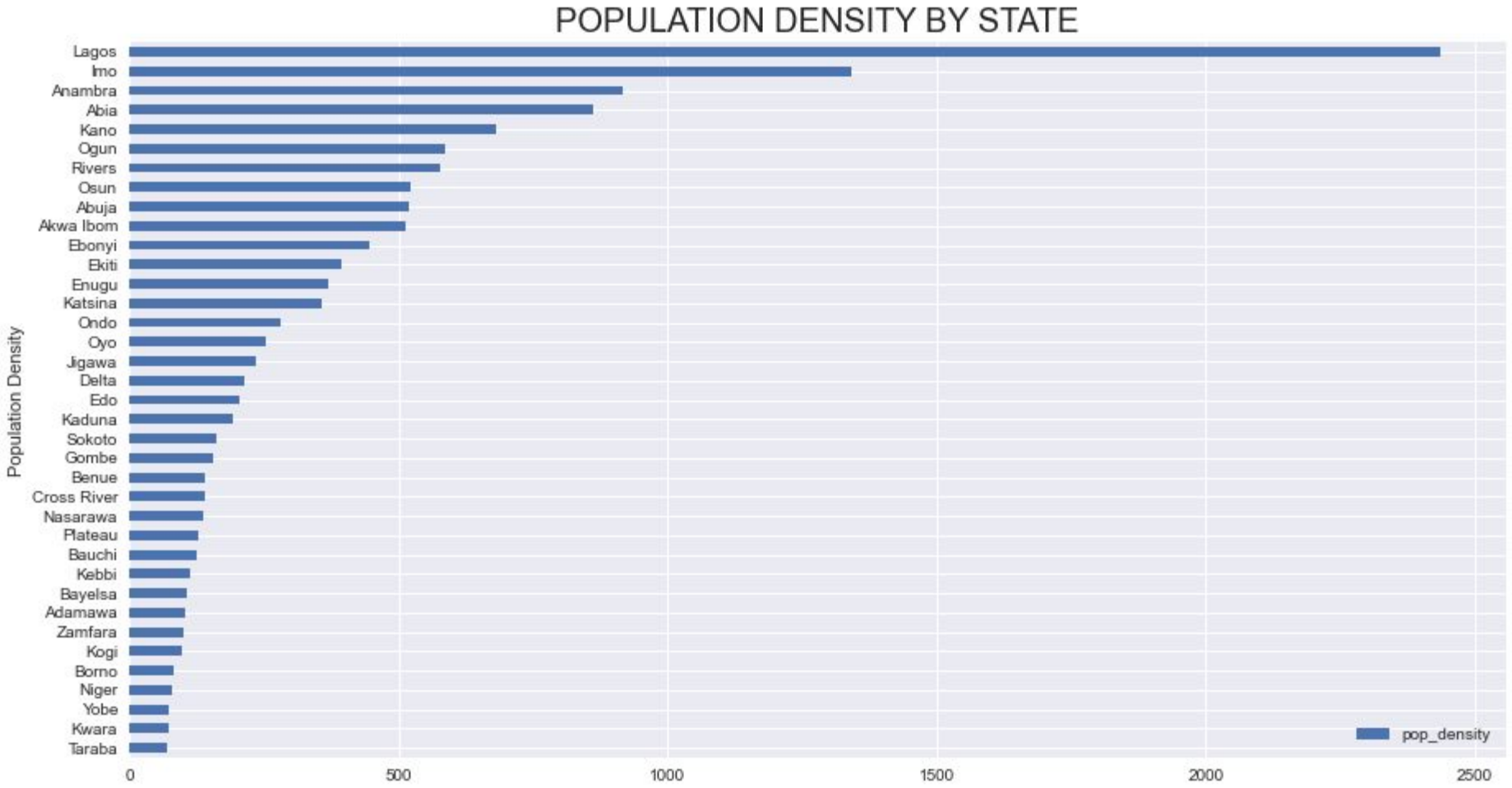
3.2.1 Feature Engineering

Feature Engineering is the process of using domain knowledge to extract features from raw data. From our data, we derived a new feature, population density.

Population density is the number of people living in a unit area per square kilometer. Mathematically, population density= Population/Area (km^2)

Reason

It is the most effective way to look at each LGA, its total occupants and land area. It eliminates any bias towards size of population or the land area of an LGA.



Date: 26/8/2021

3.2.2 Feature Selection and Scaling

The two most significant feature used factored in the model are Population density and Average Nighttime mean. Using all features would require we perform feature encoding, this would increase the number of features in the data to be learnt by the model and would introduce complexities or clusters overlapping. Simplicity is better than Complexity.

Feature Scaling: This is a method used to normalize the range of independent variables or features of data.

The two features used in our model has a clear different range of values. This can cause the model (Kmeans) to be biased towards features having larger values (population density).

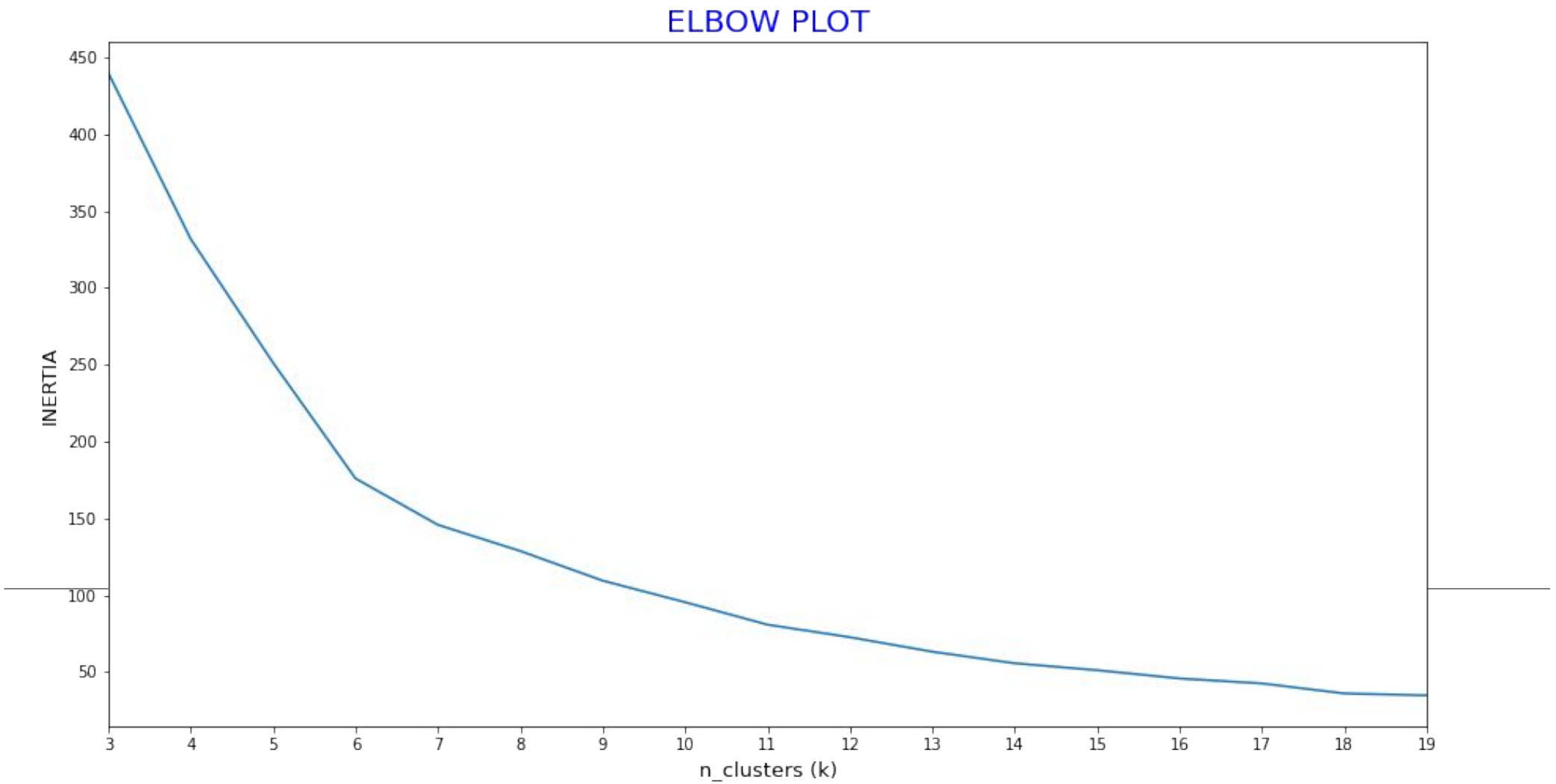
To avoid this, we had to scale the population density using StandardScaler estimator which transform each value as representation of the extent to which it deviates from the mean (0).
formula; $z = x - \mu / \sigma$

3.3 Model Building

The model used was Kmeans. It is a distance-based algorithm that tries to partition the dataset into K pre-defined distinct non-overlapping subgroups (clusters) where each data point belongs to only one group.

Prior to the instantiating the model and setting a pre-defined K clusters, an elbow plot was used to check for the optimal k clusters for the data the model was to learn from.

Although the optimal clusters for the data suggested by the elbow plot seems to fall around 11, we choose to settle for **5 clusters**, the reason behind this was to avoid too many cluster as having too much clusters would mean resources would be pulled for several sections. This would require more labor and different procedures for the installation the solar panels at strategic positions.


ELBOW PLOT
INERTIA
450
400
350
300
250
200
150
100
50
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
n_clusters (k)

3.3.1 Model Evaluation

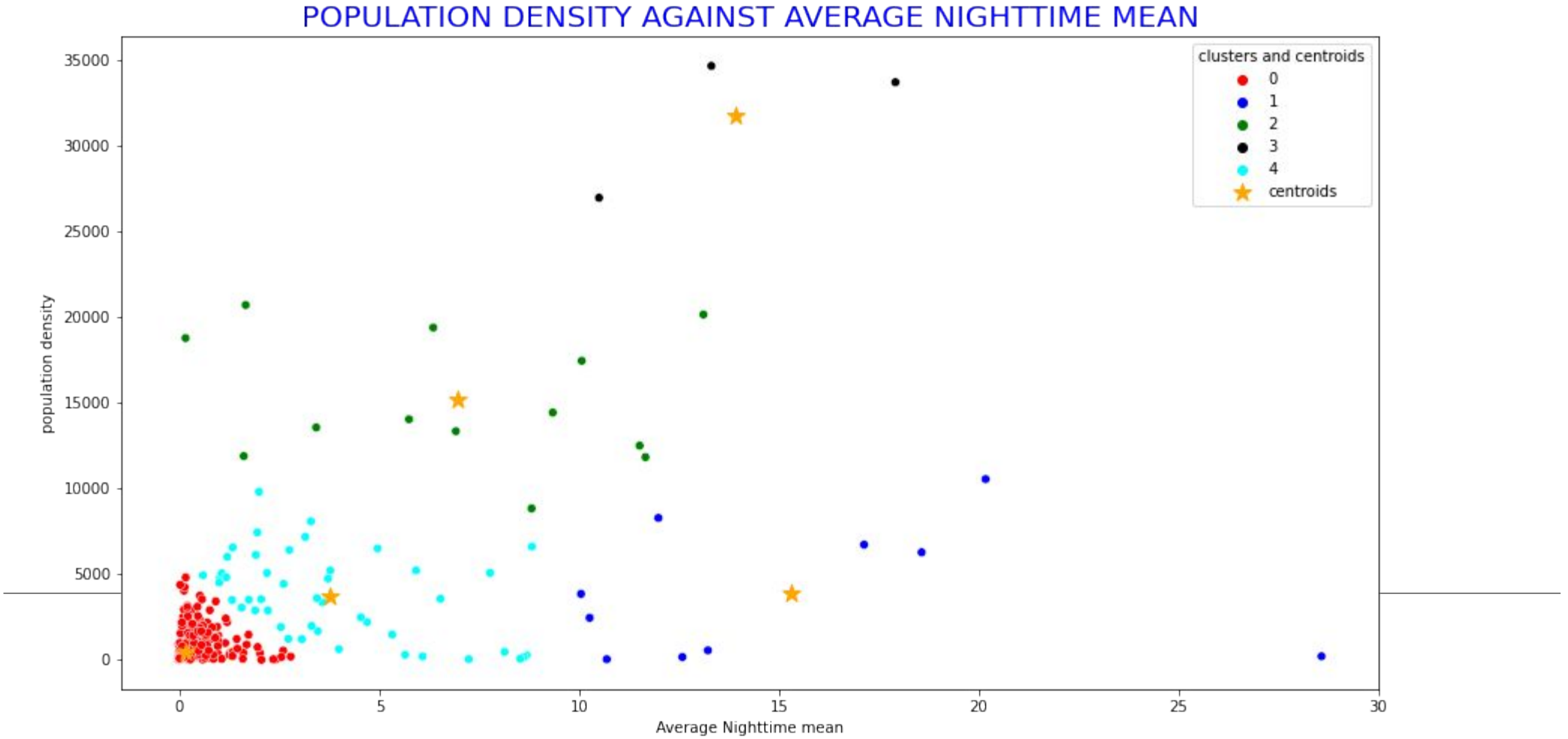
The evaluation metrics used was silhouette score. It is a metric used to calculate the goodness of a clustering technique and it ranges from -1 to 1.

When the value is 1, it means clusters are well apart from each other and clearly distinguished and when the value is 0, it means clusters are indifferent, or there is no significant difference between clusters. When it is -1, the clusters are wrongly assigned

.

$$\text{silhouette score} = (b - a)/\max(a, b)$$

The silhouette score of the model was 0.82 in 2d.p



Date: 26/8/2021

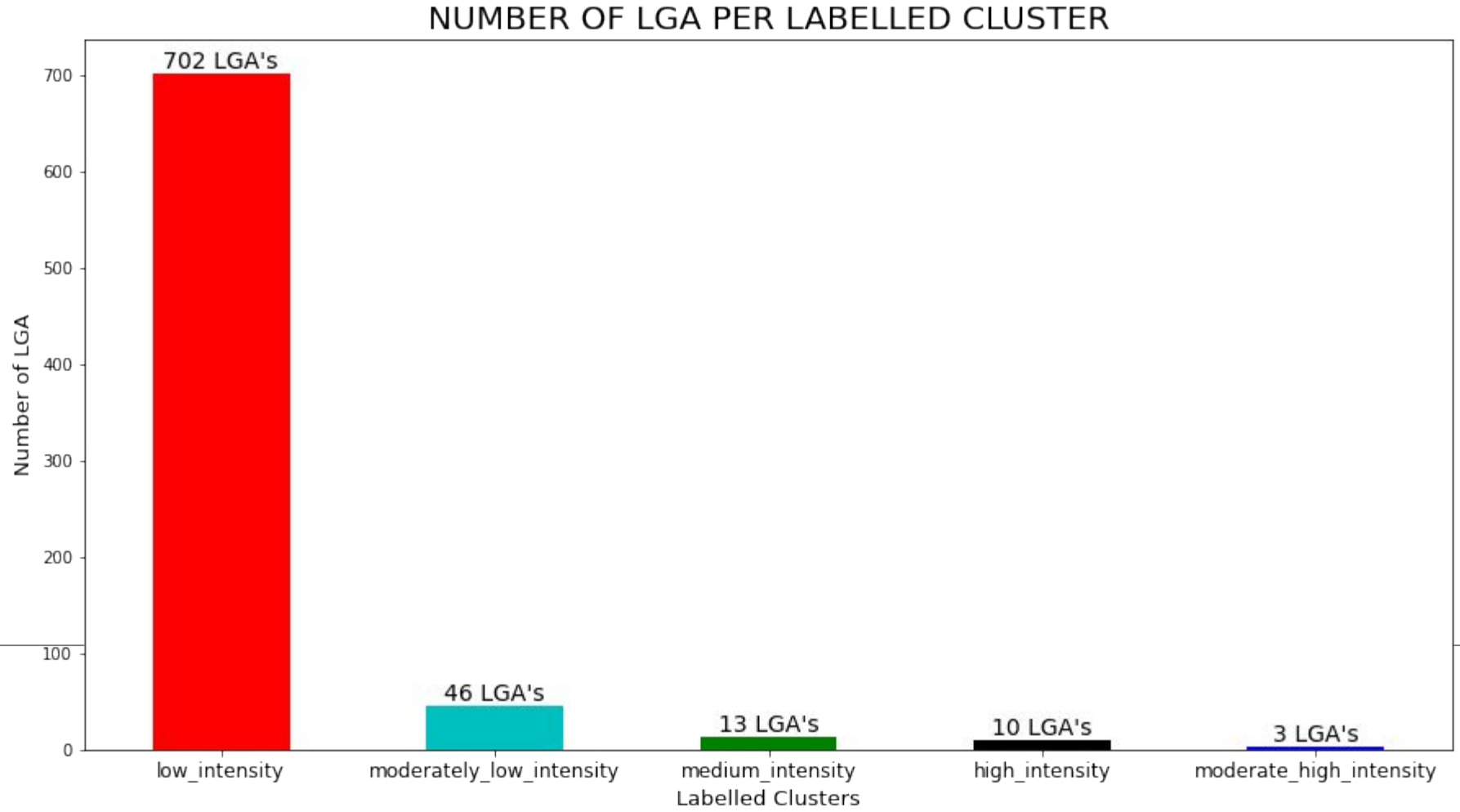
3.3.2 Clustering Labelling

The clusters generated by the model were renamed based on how much luminous intensity they emit.

From least luminous intensity to the most;

- low intensity, moderately low intensity, medium intensity, moderate high intensity, high intensity.

During solar panel installation, LGAs that fall within low intensity cluster are placed as topmost priority while LGAs within the high intensity cluster are placed last.


NUMBER OF LGA PER LABELLED CLUSTER
Number of LGA
0
100
200
300
400
500
600
700
702 LGA's
46 LGA's
13 LGA's
10 LGA's
3 LGA's
low_intensity
moderately_low_intensity
medium_intensity
high_intensity
moderate_high_intensity
Labelled Clusters

4.0 Result & Discussion

Kmeans clustering algorithm was used in modelling with a pre-defined clusters of 5.

Although, the elbow plot suggested an optimal clusters of 11, the chosen clusters was set so as to avoid too many clusters which would cause an increase in resources to be designated to those subgroups (clusters).

The model had a silhouette score of approximately 0.82 upon evaluation signifying clear and distinguished clusters.

The clusters were labelled from least luminous intensity to the most in the order; low intensity, moderately low intensity, medium intensity, moderate high intensity, high intensity. With low intensity being top on solar panel installation and high intensity coming last.

The energy demand was visualized based on the 5 clusters gotten from the model along with their local governments areas to determine the best spot for solar panel Installation

How it works

- We used Streamlit app to visualize the model

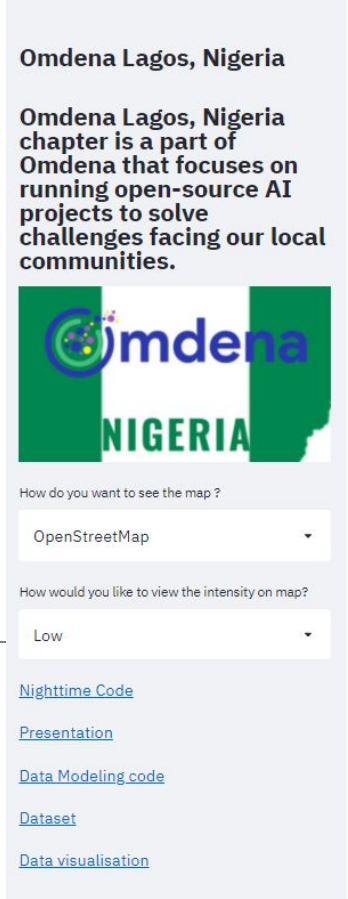
Steps to start using streamlit

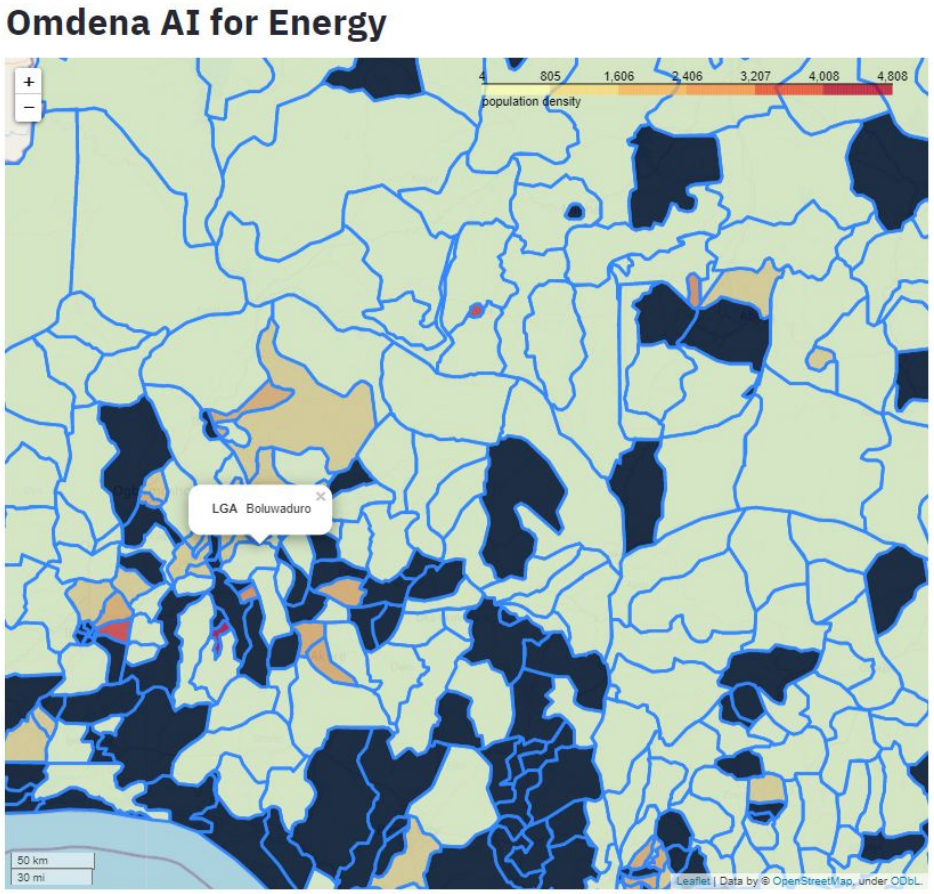
- Make sure that you have Python 3.6+ installed
- Install Streamlit using PIP and run the "hello world" app
- For further information on how to use this framework to build your own application please visit: https://docs.streamlit.io/en/stable/

Map Visualization of Population and Energy Intensity

Link to the streamlit App

https://share.streamlit.io/osabobo/omdena_app/app7.py





Date: 26/8/2021

5.0 Conclusion

The project over-all will help governments, NGOs and private sectors to survey and validate locations before installing solar panels to regions

The other aspect of the project may help government, NGOs and private sectors to dentify when to use renewable energy.

6.0 Recommendation

Experimenting on more models with better performance metrics to validate such models

6.0 Appendix

Raw data used
https://github.com/OmdenaAI/omdena-nigeria-energy/blob/main/src/final%20deliverables/Dataset_final/Nighttime_and_population_data.csv

Jupyter notebook for EDA
https://github.com/OmdenaAI/omdena-nigeria-energy/blob/main/src/final%20deliverables/task/Data_Modeling/Explainatory%20Data%20Analysis.ipynb

Jupyter notebook for data modelling
https://github.com/OmdenaAI/omdena-nigeria-energy/blob/main/src/final%20deliverables/task/Data_Modeling/Omdena%20cluster%20model.ipynb


Thank You!