



# NIGERIA COVID-19

## DATA ANALYSIS USING PYTHON



## **PROJECT OVERVIEW**

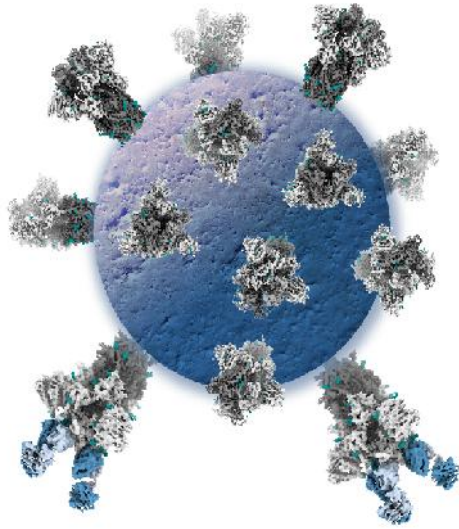
Nigeria is a country with 37 states - Federal Capital Territory included- and a fast-growing economic environment with about 200 million citizens. COVID-19 has affected several country activities as the country steadily progressed from its first case to shutting down major airports, state-wide lockdown, curfews, and reviving its economy. Coronavirus disease (COVID-19) is an infectious disease caused by a newly discovered coronavirus, and it has affected major parts of the world. Nigeria, a West-African country, has also been affected by the COVID-19 pandemic after recording its first case on 27th February 2020.

## **PROJECT OBJECTIVE**

In this project, data science and analytics skills are employed to collect data, explore the data, perform analysis, create visualizations, and generate insights.

- Use of web Scraping or extraction in collecting data from diverse sources
- Cleaning and preparation of data to standard format in order to be ready for analysis and insight getting from the data.
- Analysis and Visualization of data to find useful insight and trend from the data.
- Summary report or documentation to communicate the inferences





Efforts to combat the virus have also included widespread testing, contact tracing, quarantine and isolation measures, travel restrictions, and the development and distribution of vaccines. Several vaccines have been authorized for emergency use and are being administered to individuals to provide protection against COVID-19.

## REVIEW OF COVID-19

COVID-19, also known as Coronavirus Disease 2019, is a highly contagious respiratory illness caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). It was first identified in December 2019 in the city of Wuhan, Hubei province, China. Since then, it has spread globally, leading to a pandemic.

COVID-19 primarily spreads through respiratory droplets when an infected person coughs, sneezes, talks, or breathes. It can also spread by touching contaminated surfaces and then touching the face. The virus can cause a range of symptoms, from mild to severe, including fever, cough, fatigue, shortness of breath, loss of taste or smell, sore throat, and body aches. In severe cases, it can lead to pneumonia, organ failure, and even death.

Certain individuals, such as older adults and those with underlying health conditions, are more vulnerable to severe illness. However, people of all ages can contract the virus and transmit it to others, even if they are asymptomatic or have mild symptoms.

To mitigate the spread of COVID-19, public health measures have been implemented worldwide. These measures include wearing face masks, practicing physical distancing, frequently washing hands with soap and water, using hand sanitizers, avoiding large gatherings, and following local guidelines and restrictions.

## DATA INFORMATION

Data was extracted from diverse sources using different techniques which includes but not limited to web scrapping or extraction, forking and cloning of GitHub and other sources. Some brief information about the data is given below.

- ✓ The Nigeria Centre for Diseases Control (NCDC) which contains the states affected by the virus, the number of cases on confirmation and admission, number of discharged, and number of death.
- ✓ The Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE) publishes daily data on confirmed, death and recovered cases across different countries. We are to extract data for Nigeria only from these data source.
- ✓ Nigeria Community Vulnerability Index data. The vulnerability index was computed by considering several factors such as socio-economic status, population density, housing type, transportation, epidemiological, health system etc., these factors are known as themes.
- ✓ Real Domestic Gross Product Data for Nigeria. This data is needed to get the impact of COVID-19 on the Nigeria Economy.
- ✓ State Budget Data is provided also to support the analysis and infer on how the COVID-19 pandemics affected the economy of Nigeria and the state within it.



## ANALYSIS QUESTION

These are some question that my exploration seek to answer upon analysis of given data and data collected from diverse sources.

- Exploring the distribution of the virus across each states and the numeric effect on each state. Simply put, the distribution of the confirmed, discharged and the death cases in Nigeria state.
- Explore to get the correlation/relatedness of the features listed above and making inference based on the correlation.
- From John Hopkin's data, getting the active cases in all countries and extracting the out Nigeria out for analysis
- Analyze the relation between the confirmed cases, recovery cases and death instance and to use this to compute the infection rate across the country.
- Exploring and making inferences from the distribution of the infection rate based on the dates given in the date dataset.
- Investigating and analysis of the some certain index in the external data. These index includes fragility, vulnerability, epidemiology, prevalence, health care, population density, transportation and socio-economic index
- Exploration of data based on this indexes to get graphs and chart that is better for communication and inferences
- Analysis of the external GDP data to gain insight on the effect of COVID-19 pandemic on the country's' economy
- Further insight from respect state budget and revision to get more insight on the effect of the virus on Nigeria's economy.

## ANALYSIS METHODOLOGY

Based on the analysis questions the following methodologies are used and leverage on to gain sustainable insight into the data.

- Use of web Scraping or extraction in collecting data from diverse sources
- Cleaning and preparation of data to standard format in order to be ready for analysis and insight getting from the data.
- Analysis and Visualization of data to find useful insight and trend from the data.
- Summary report or documentation to communicate the inferences

## ANALYSIS EXPLORATION AND INFERENCES (RESULT)

In this session, I will be using diverse charts to explain the result of my analysis and inferences made from the data in accordance to the analysis question.

### ❖ Understanding the distribution of features in the NCDC DATA across each state

From the graph shown below - Figure 1.

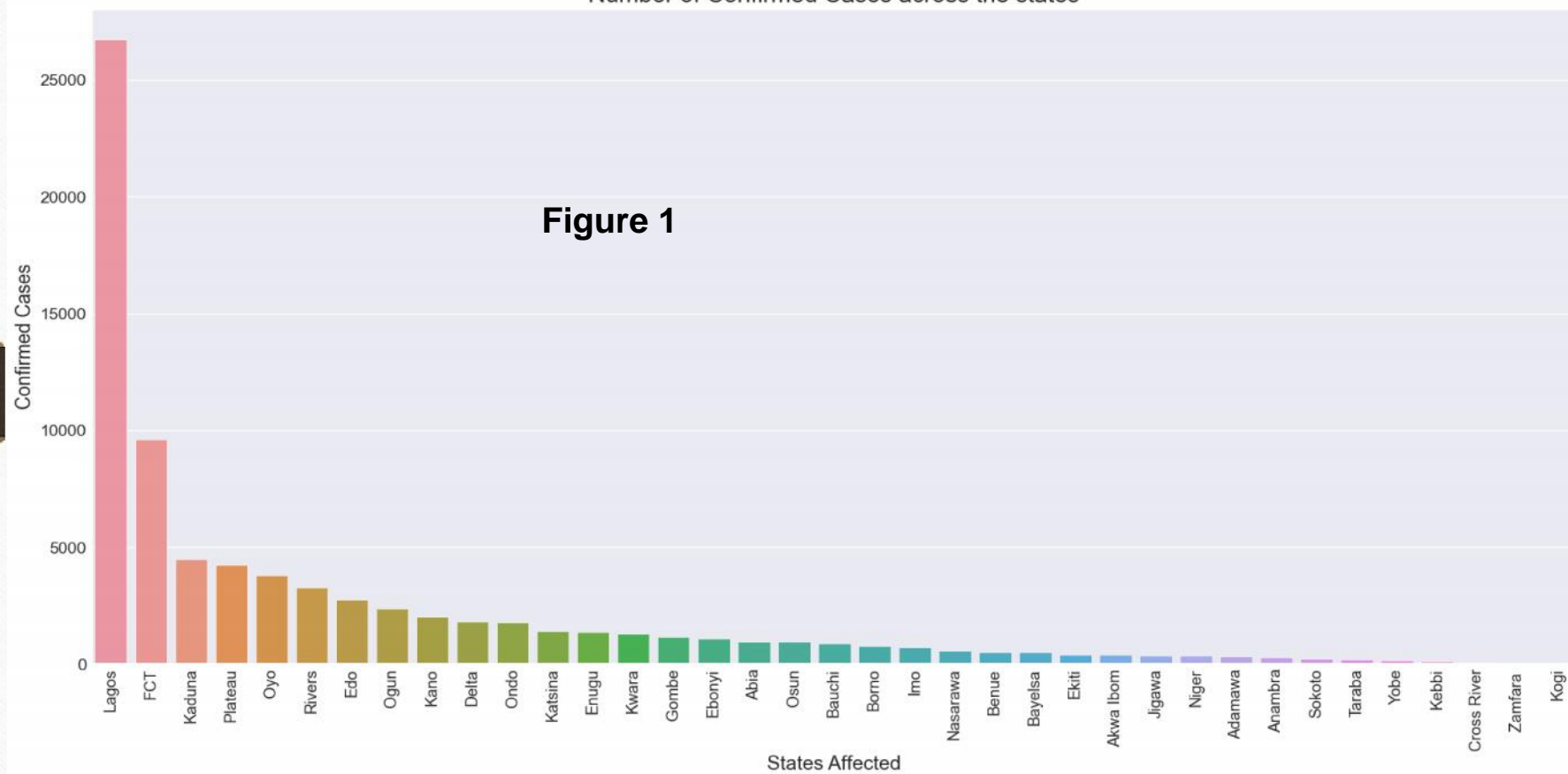
It shows from the inferences; on the bar plot graph, the growing of confirmed cases of COVID-19 across the states of Nigeria.

With Lagos State having highest number confirmed cases compared to order states, across the country.

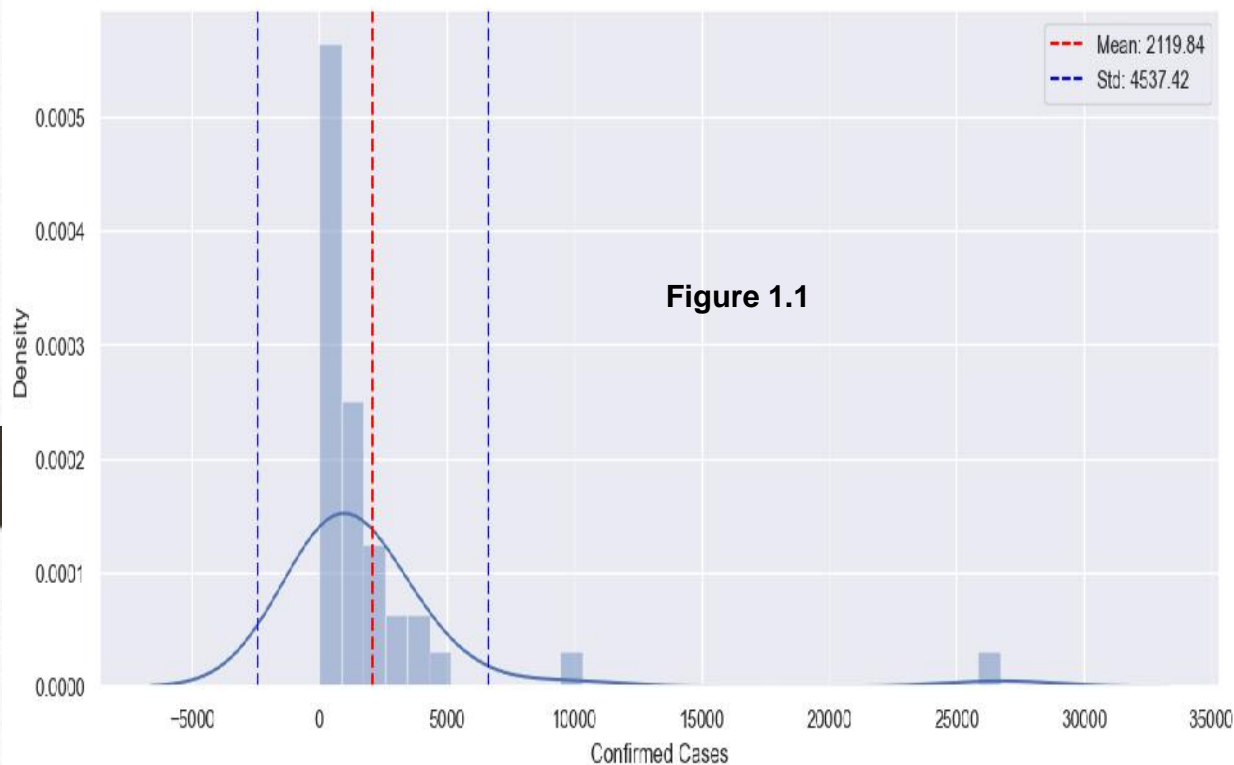
**Lagos State : 26,708 , FCT: 9,627, Kaduna: 4,504, Plateau: 4,262, Oyo: 3,788.**

**Rivers : 3,279, Edo : 2,768, Ogun : 2,382, Kano : 2,032, Delta : 1,843. And E.T.C**

Number of Confirmed Cases across the states







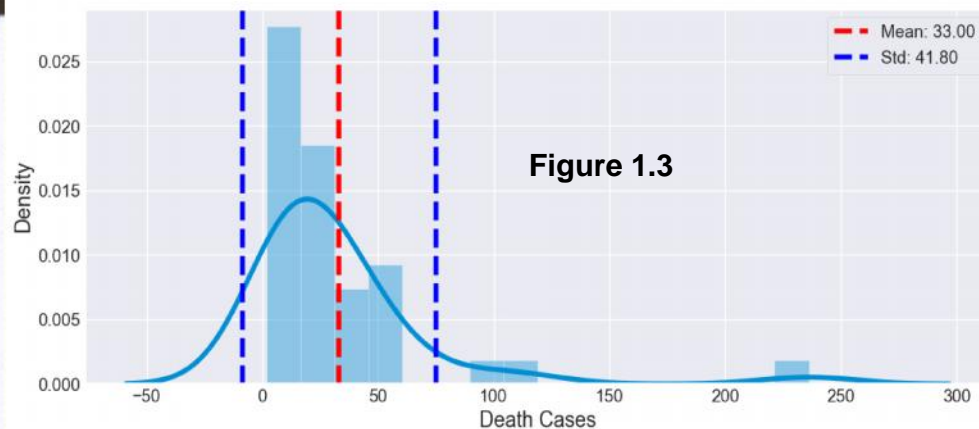
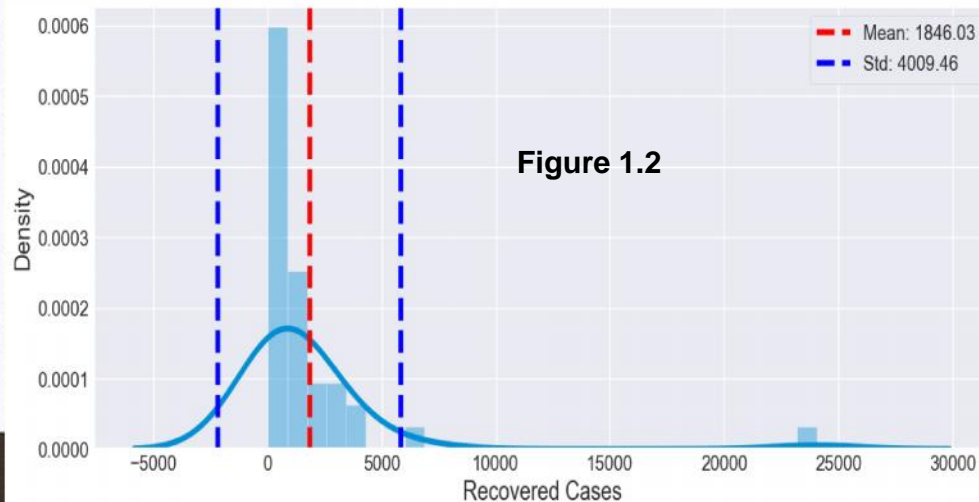
**STD:4537.42    MEAN: 2119.84    STD : 4537.42**

The resulting plot and annotations give an insight into the distribution of the confirmed COVID-19 cases, providing a summary of the data and highlighting the mean and standard deviation as key statistical measures.

Fig1.1 shows: The distribution plot visualizes the distribution of the 'Confirmed Cases', showing how the data is spread out across different values. It gives an overview of the frequency or count of different values, it shows increase in population, may lead to more out break of the COVID-19 virus . The mean value, annotated as 'Mean: <mean\_value>', represents the average value of the 'Confirmed Cases' . It indicates the central tendency of the data distribution, giving an idea of the typical value.

The standard deviation value, annotated as 'Std: <std\_value>', measures the spread or dispersion of the data around the mean. It indicates the variability or scatter of the 'Confirmed Cases' values. A larger standard deviation suggests a wider spread of the data points. From the results shows the pick of confirmed cases of the virus from the mean value, that is, has the density decreases; decrease in virus been spread across the states.



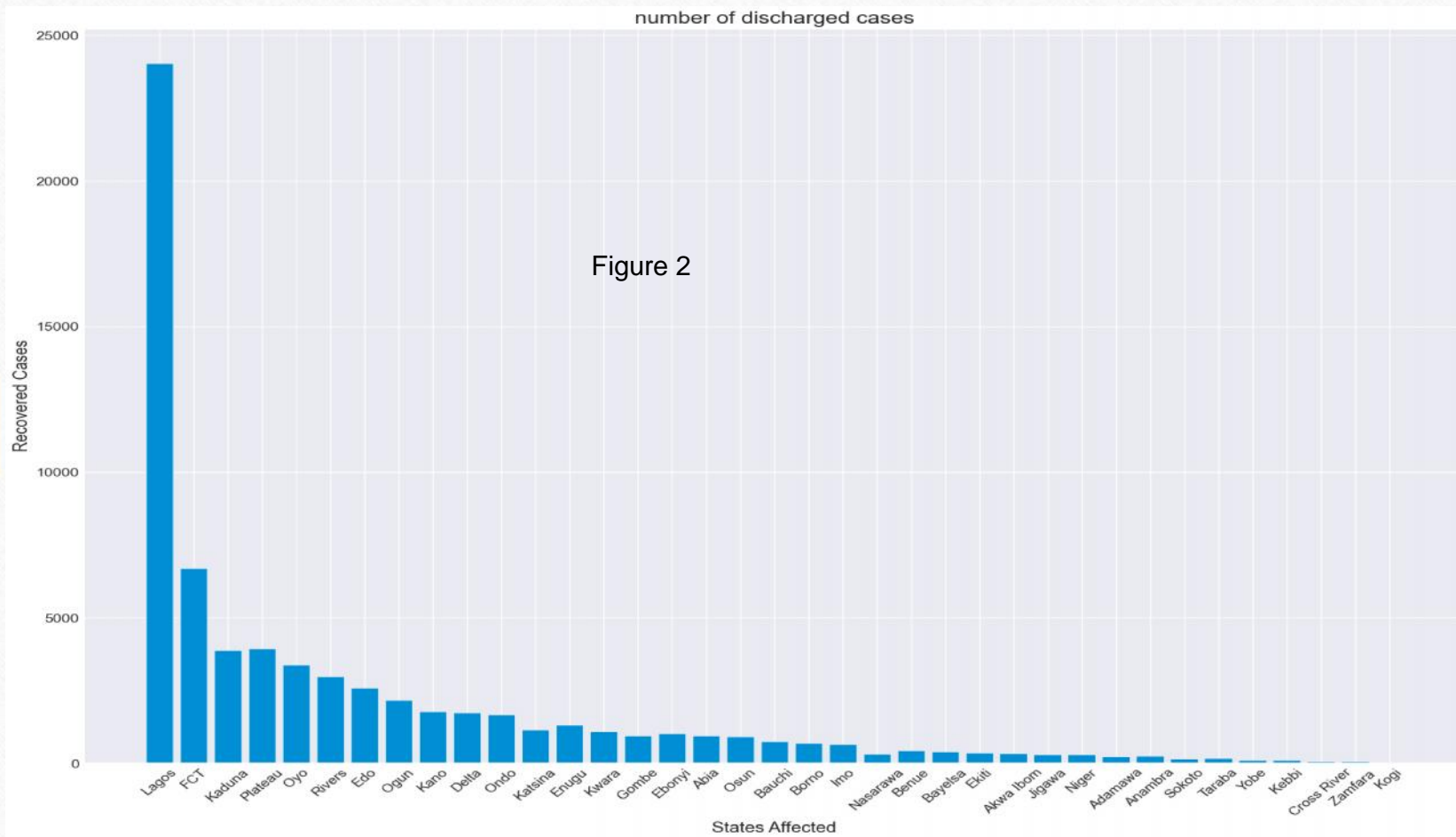


**Notice**, from the distribution of the three (3) features in the NCDC data, we notice a some sort of bias in the chart/plot. This is indicative that there is larger number of discharged, cases, confirmation cases and death cases in few states compared to other states across the nation. A more clearer insight on the figure 1.1 – 1.3 can be deduced from the bar plot of the same metrics. Shown from the distribution of the dataset, in the mean value of each.

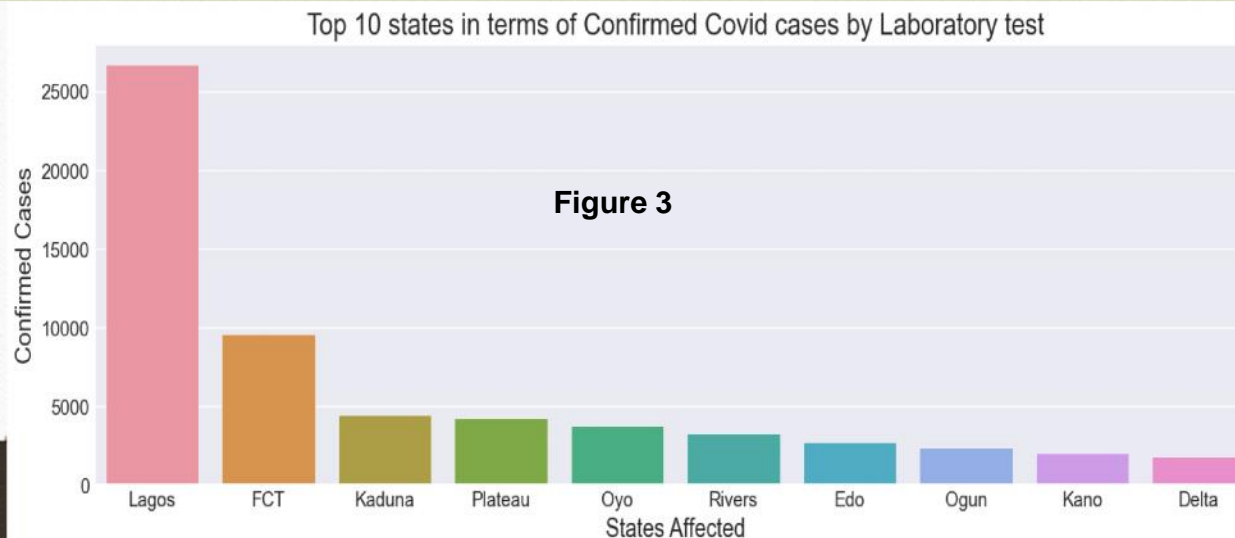
#### **Inference:**

We noticed from the distribution of the virus across Nigeria within the period being considered. Lagos overall Discharged cases(Recovered Cases) is very high compared to other state.

The top 10 states with the highest number of discharged cases are: Lagos with 24037 cases, FCT with 6694 cases, Plateau 3948 cases, Kaduna 3877cases, Oyo 3374 cases, Rivers 2987 cases, Edo 2603 cases, Ogun 2175 cases, Kano 1778 cases, and Delta 1737 cases.



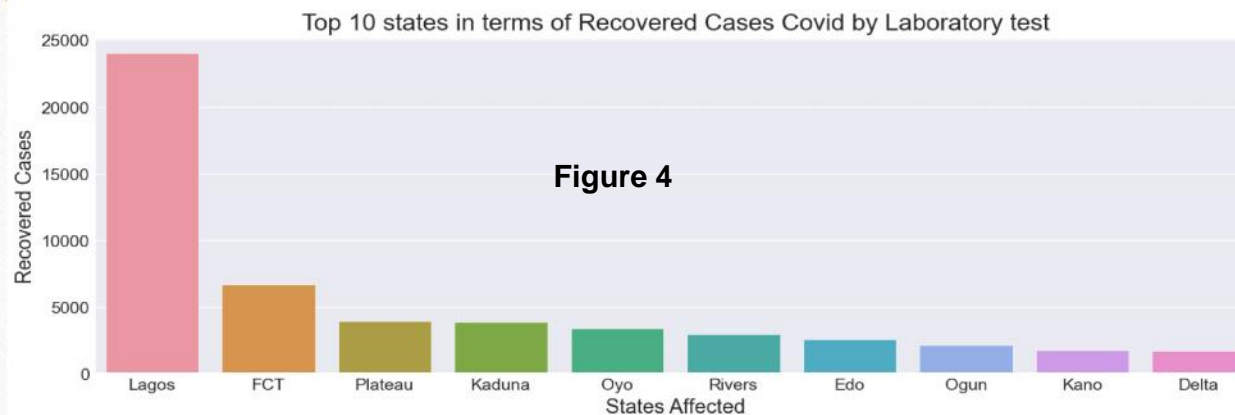




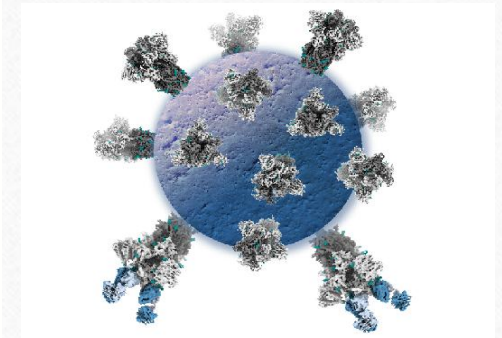
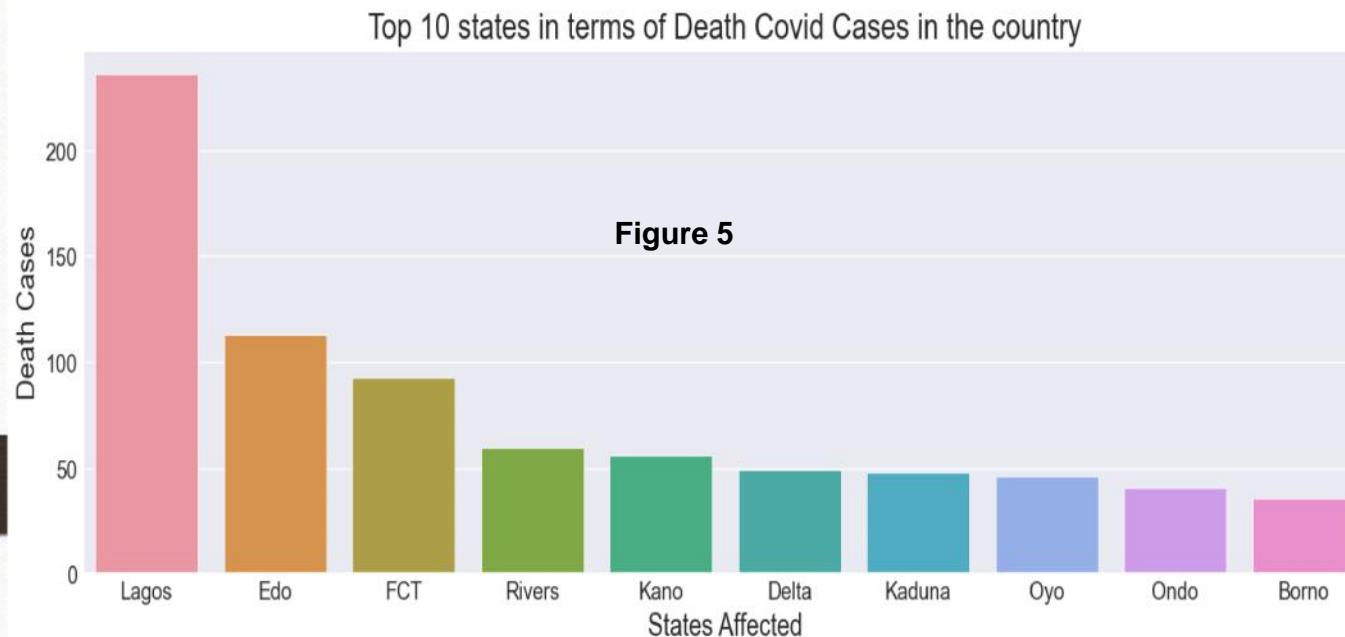
**Inference:**

From fig 3: The top 10 states with the highest discharged cases are Lagos with 26708, FCT 9687, Kaduna 4504, Plateau 4262, Oyo 3788, Rivers 3279, Edo 2382, Ogun 2382, Kano 2032 and Delta 1843.

Based on this we can compare it with our previous result and infer that majority of the states with more confirmed cases also have high discharge rate.



Form fig4: The top 10 death cases are Lagos with 24037, FCT 6634, Plateau 3948, Kaduna 3877, Oyo 3374, Rivers 2987, Edo 2603, Ogun 2175, Kano 1778, and Delta 1843.,.

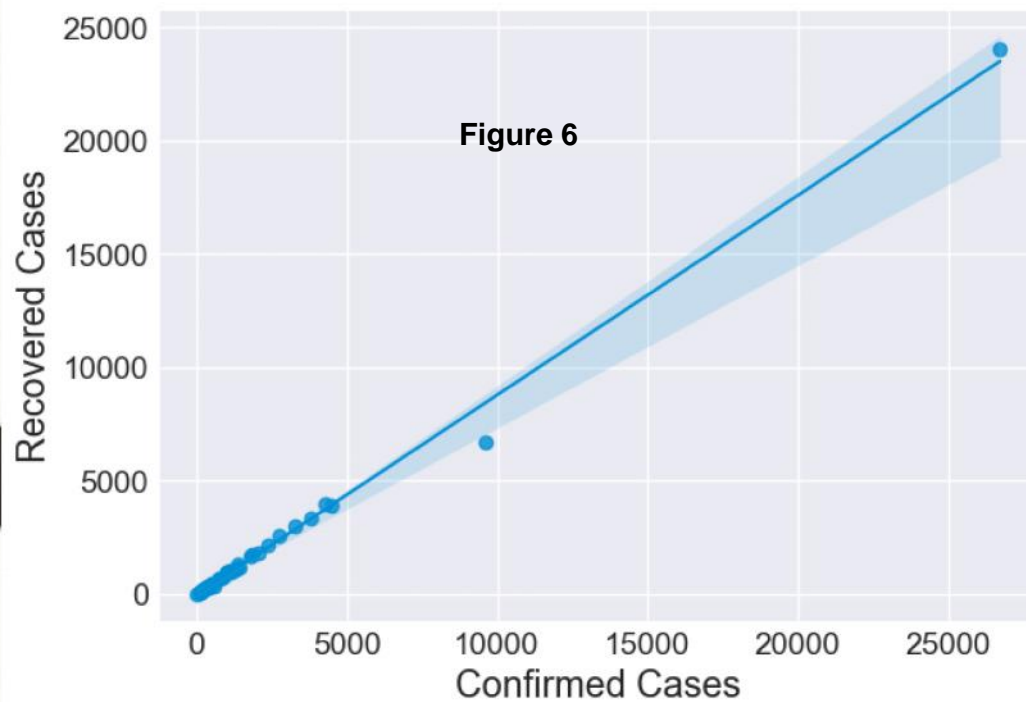


**Inference:**

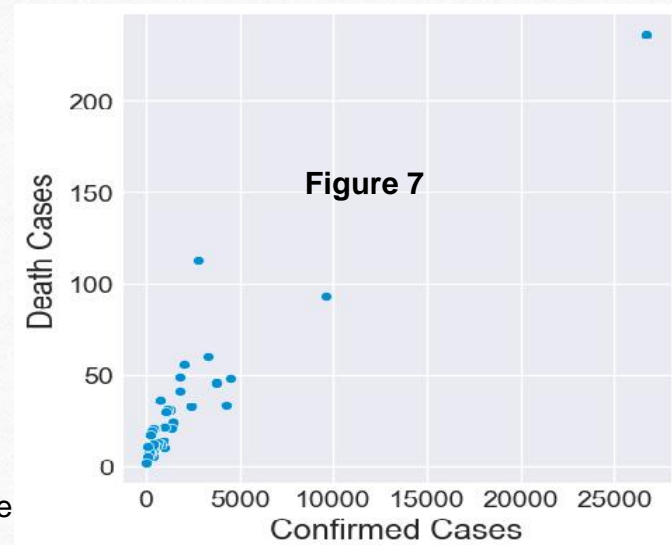
From Fig 1.3 and Fig 5: Shows the distribution of the dataset; of the death rate cases increase due to Increase in population in some states. Lagos having highest number death Cases follow by Edo, FCT, Rivers, Kano, Delta, respectively.

The top 10 states in terms of Death Cases of the virus (COVID-19) across the country are as follows: Lagos with 236, Edo 113 , FCT 93, Rivers 60, Kano 56, Delta 49, Kaduna 48, Oyo 46, Ondo 41 and Borno 36 while Kogi has no death cases.





The analysis from fig 6 is based on the univariate analysis. Overall, the regression plot generates the visualization and analysis between the relationship of the number of 'Confirmed Cases' and 'Recovered Cases' using the dataset; but we can further show the relatedness between the number of confirmed and number of discharge is high based on state affected. This account rate of discharge being the slope of the bivariate plot.



The analysis in fig 7 provides insights into the relationship between two variables, the relationship between the "Death Cases and Confirmed Cases". That due in some states the Confirmed Cases of the virus are high; the Death rate varies not high as the confirmed cases. Some state recorded low death rate While some no death.

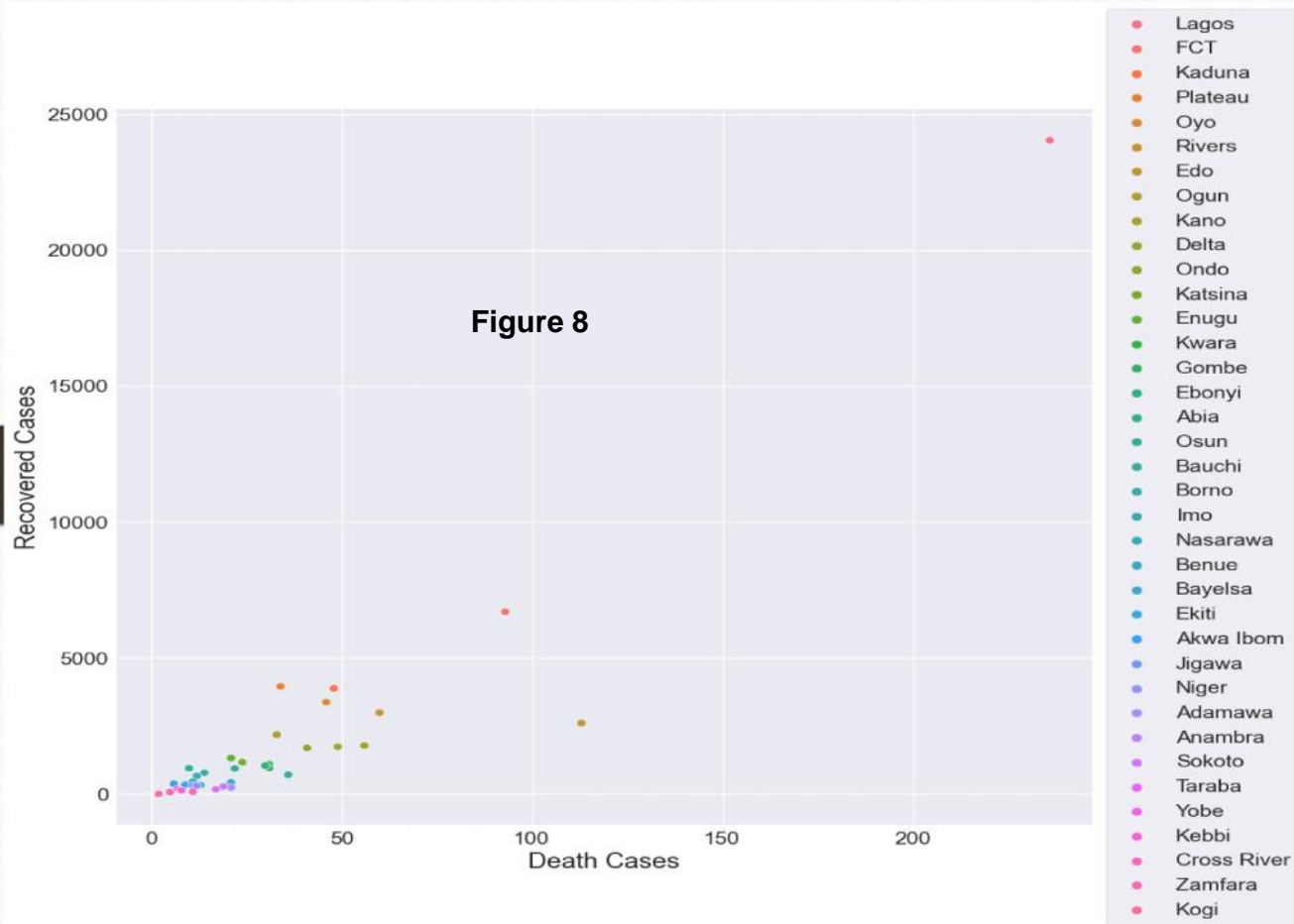


Figure 8: From the scatter plots between the “Recovered Cases and Death Cases”; the graphical representation displays the relationship between the two variables. The relationship shows that the Recovering rate is high compared to Death rate across the states. Lagos state with high recovering rate than Kogi state showing the lowest death rate and low population rate.



Total Daily Confirmed, Recovered and Death Cases in Nigeria

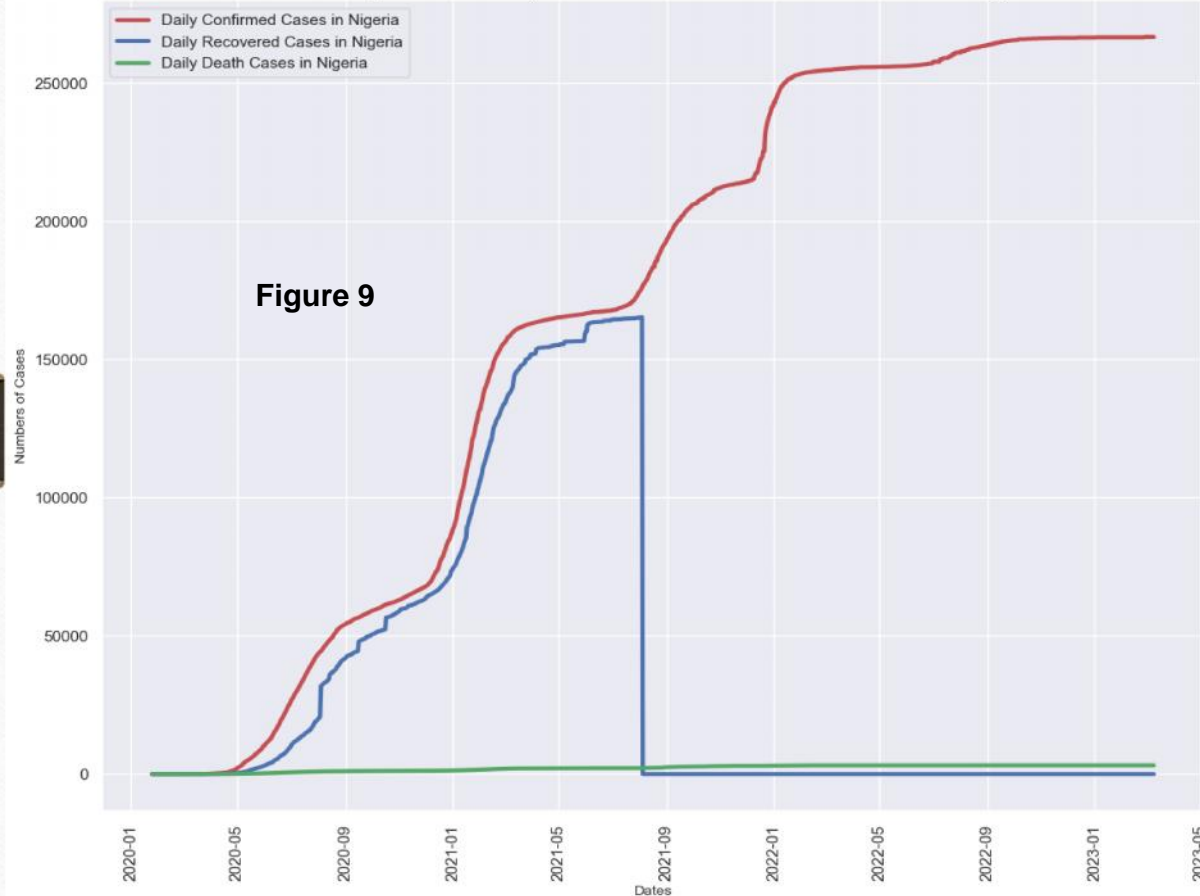
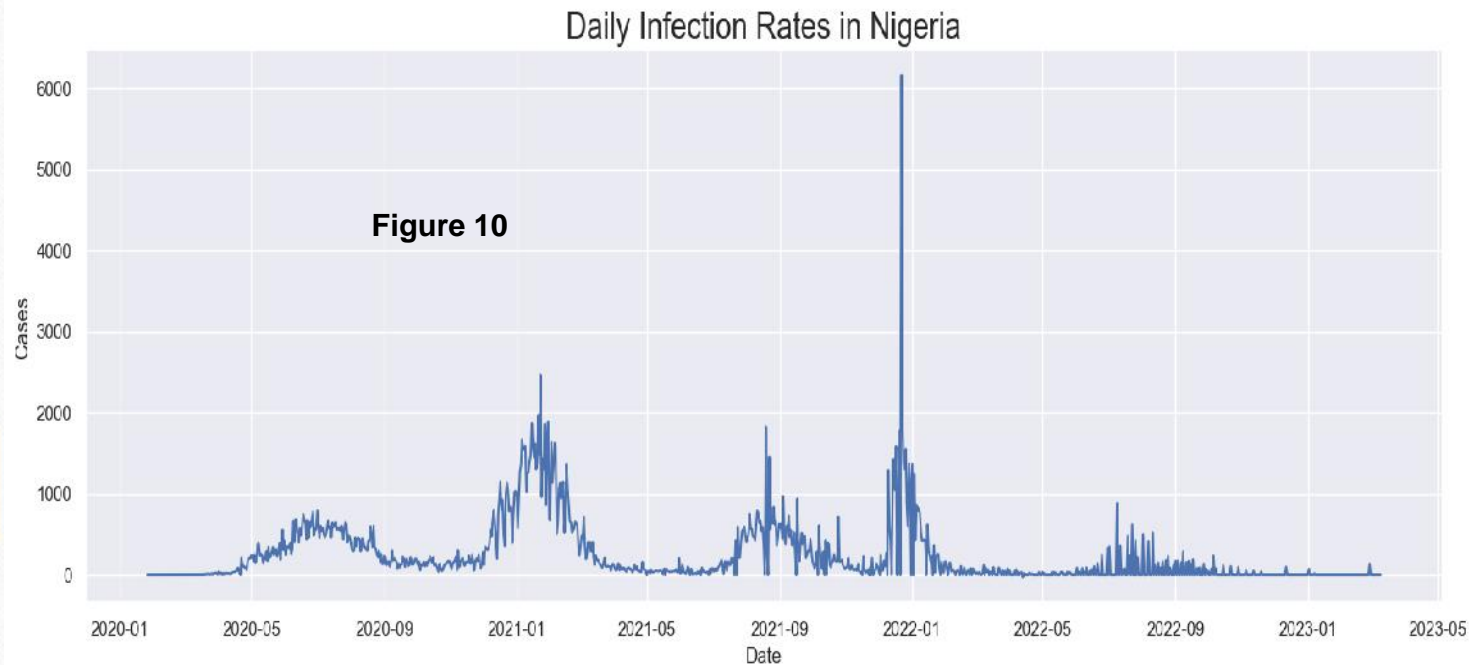


Figure 9

❖ The Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE) publishes daily data on confirmed, death and recovered cases across different countries. We are to extract data for Nigeria only from these data source. After the data transformation of the three (3) dataset and extraction of needed features. The plot below shows the relationship between case confirmation, recovery rate and death rate.

#### Inference:

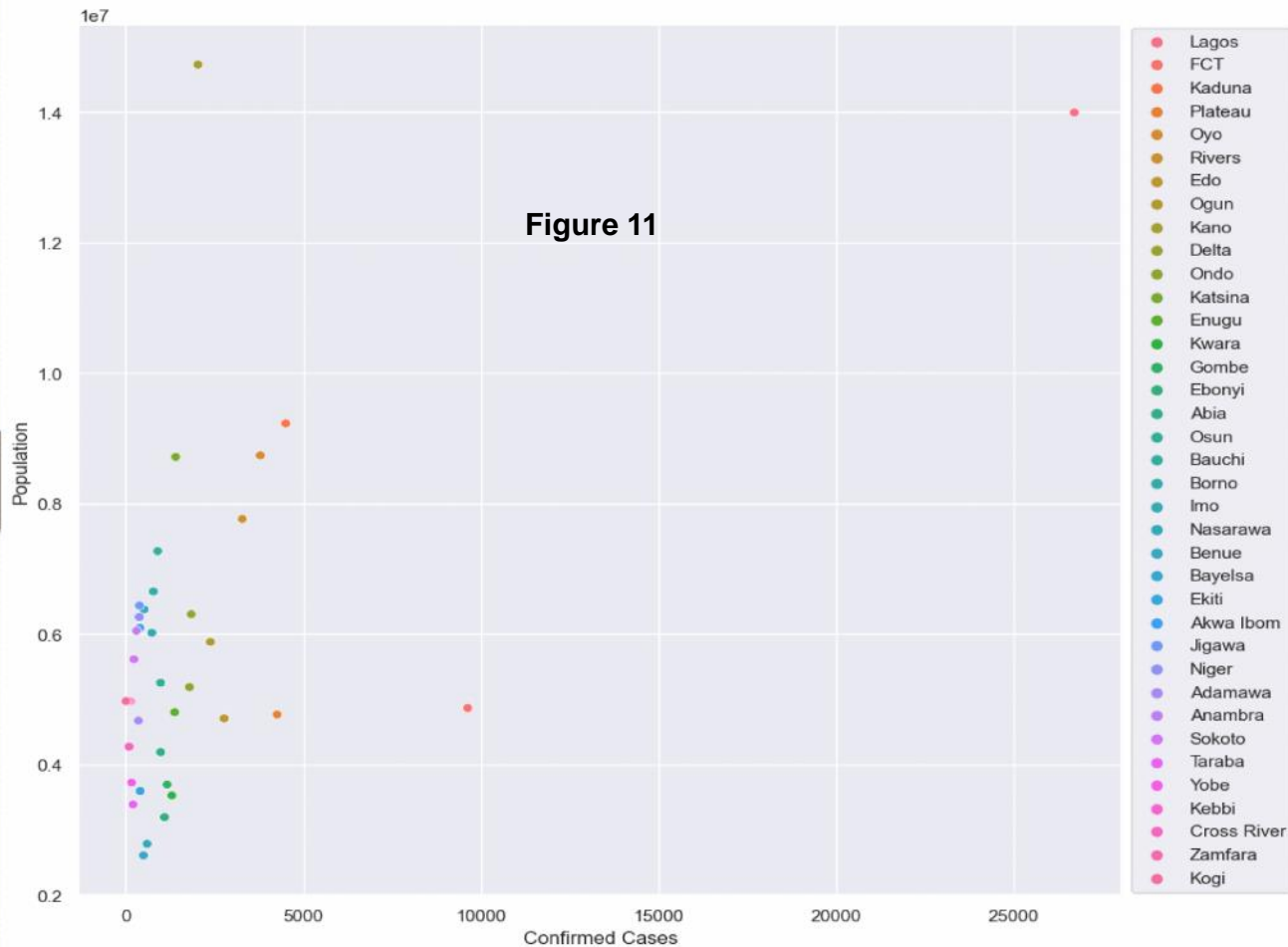
We can infer that a large proportion of those that were confirmed positive have recovered from the infection.



**Inference:**

Determining the daily infection rate, find the derivate of the total cases. Upon computation of the increment based on the change in confirmed cases on a daily basis. We can see that there was a huge spike in the infection rate from around January 2021 and January 2022. Maximum infection increment is 6158.0. The date with the maximum infection rate was on the 22nd of December 2021 with an infection rate of 6158.0.



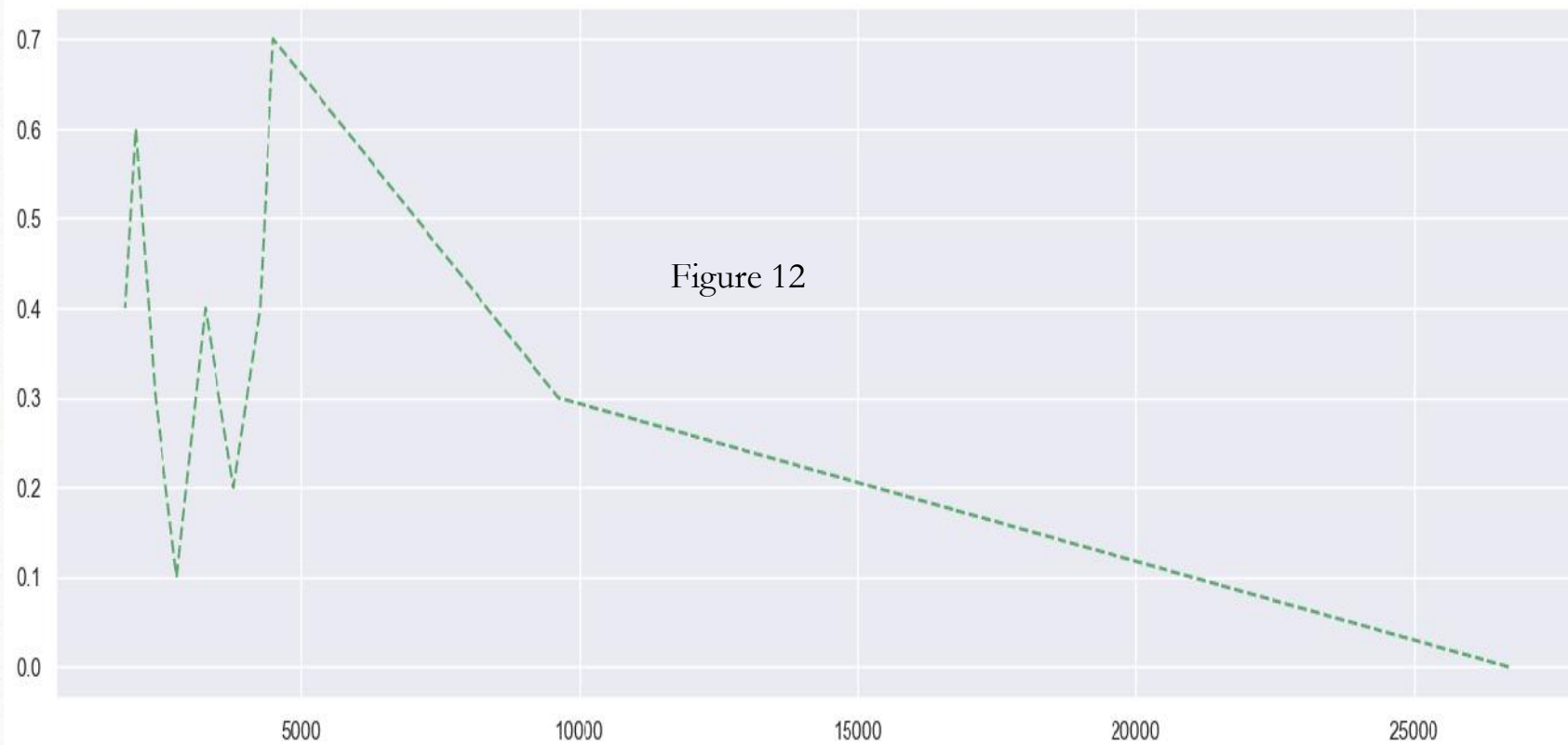


❖ **Analysis from statistical summary of COVID-19 data from external source.**

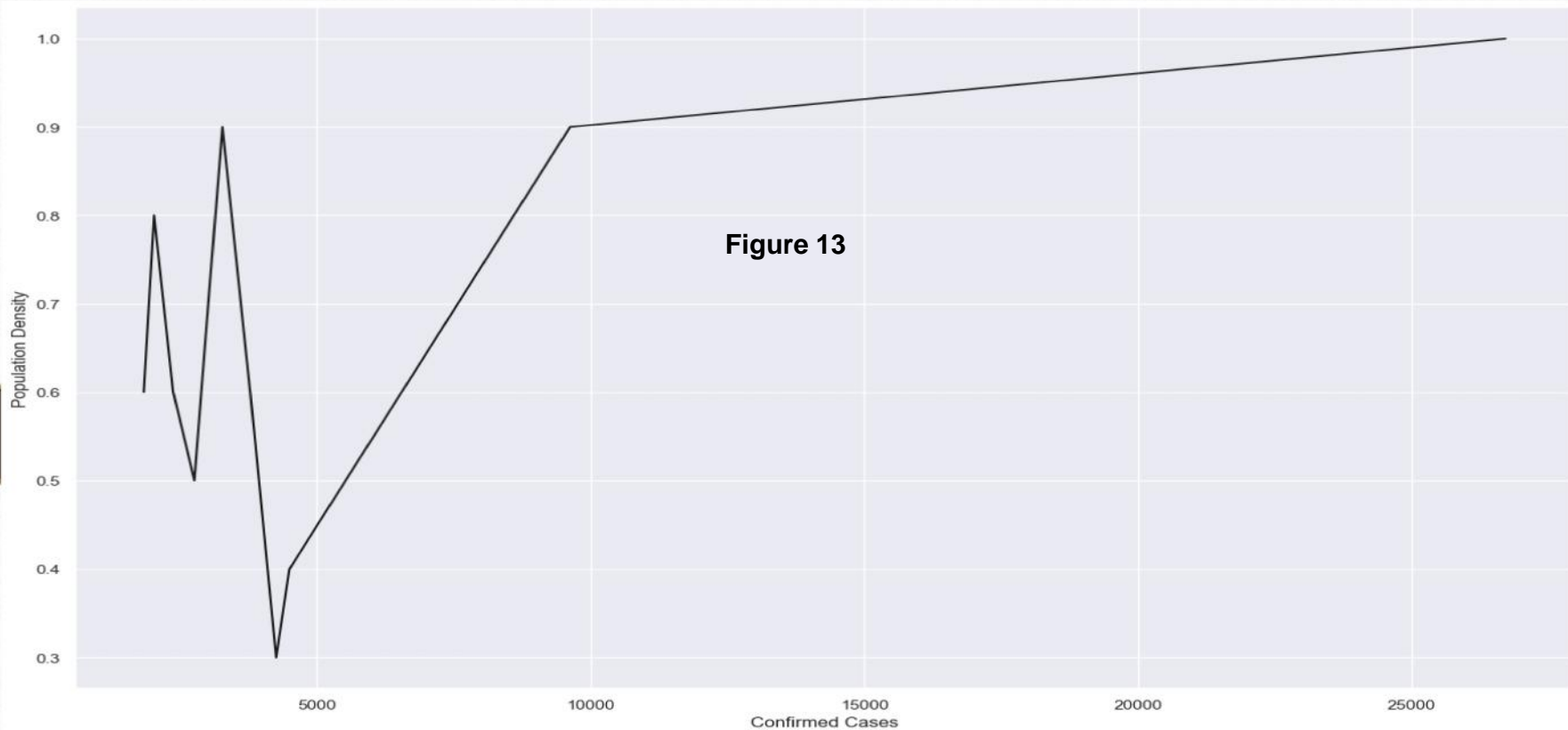
The Investigated and analysis of the some certain index in the external data.

These index includes fragility, vulnerability, epidemiology, prevalence, health care, population density, transportation and socio-economic index. Exploration of data based on this indexes to get graphs and chart that is better for communication of the statistical inferences. Fig.11 show the show the relationship how the virus spread across the states , based on the population of respective state affected.

Figure 12



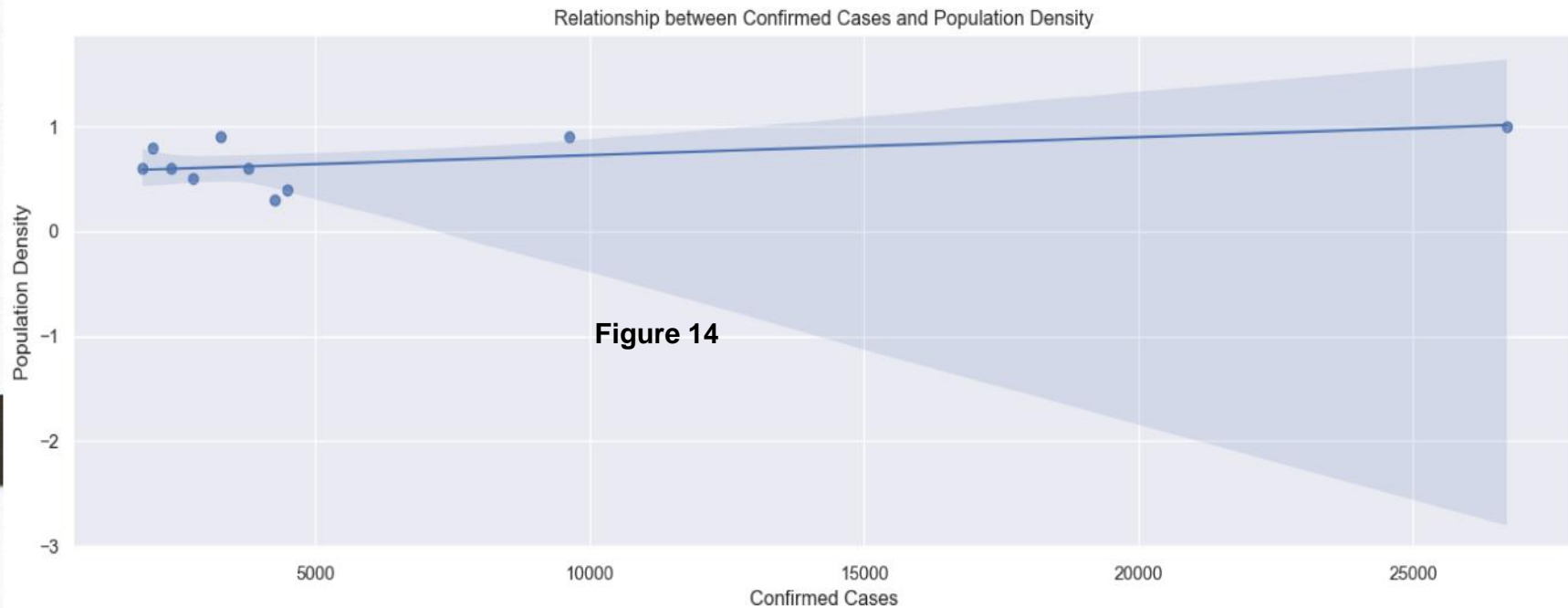




**Figure 13**

**Inference: From Fig.12:** Graph showing the areas with low CCVI have relatively high number of confirmed cases

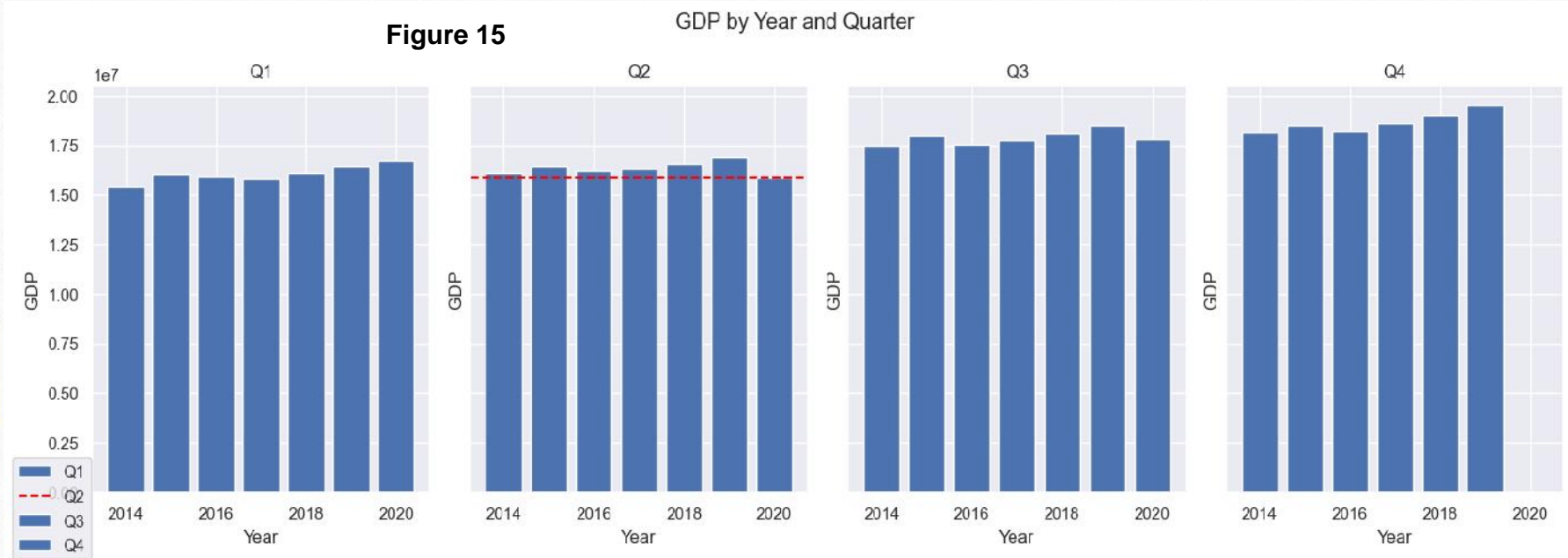
**Inference: From Fig.13:** Densely populated area also contributed to the source capsule of high number of confirmed cases



Once you generate the regression plot, observe the trend and relationship between the confirmed cases and population density:

- If the regression line has a positive slope, it indicates a positive correlation between confirmed cases and population density. As population density increases, the number of confirmed cases tends to increase.
- If the regression line has a negative slope, it indicates a negative correlation, suggesting that as population density increases, the number of confirmed cases tends to decrease.
- If the regression line is flat, it suggests no significant linear relationship between the two variables.

## ❖ QUARTERLY GDP DATA ANALYSIS



**GDP(Gross Domestic Product) data involves several aspects:**

- 1. Tracking Economic Growth:** GDP data allows analysts to monitor the overall economic performance of a country. By comparing GDP figures from one quarter to another, it is possible to assess the rate of economic growth or contraction.
- 2. Identifying Trends:** Examining quarterly GDP data helps identify short-term economic trends. It allows economists, policymakers, and investors to understand whether the economy is expanding or contracting, and whether any patterns or cycles are emerging.



3. **Sectoral Analysis:** GDP data can be further broken down by sectors such as agriculture, manufacturing, services, etc.

Analyzing sectoral contributions to GDP provides insights into which industries are driving economic growth or facing challenges.

4. **Business Cycle Analysis:** GDP data is often used to determine the phase of the business cycle. By examining fluctuations in quarterly GDP figures, economists can identify periods of expansion, recession, or recovery in the economy.

5. **Policy Formulation:** Governments and policymakers utilize quarterly GDP data to assess the effectiveness of economic policies and make informed decisions. It helps them identify areas of the economy that require attention or intervention.

6. **International Comparisons:** Comparing quarterly GDP data across different countries allows for cross-country analysis.

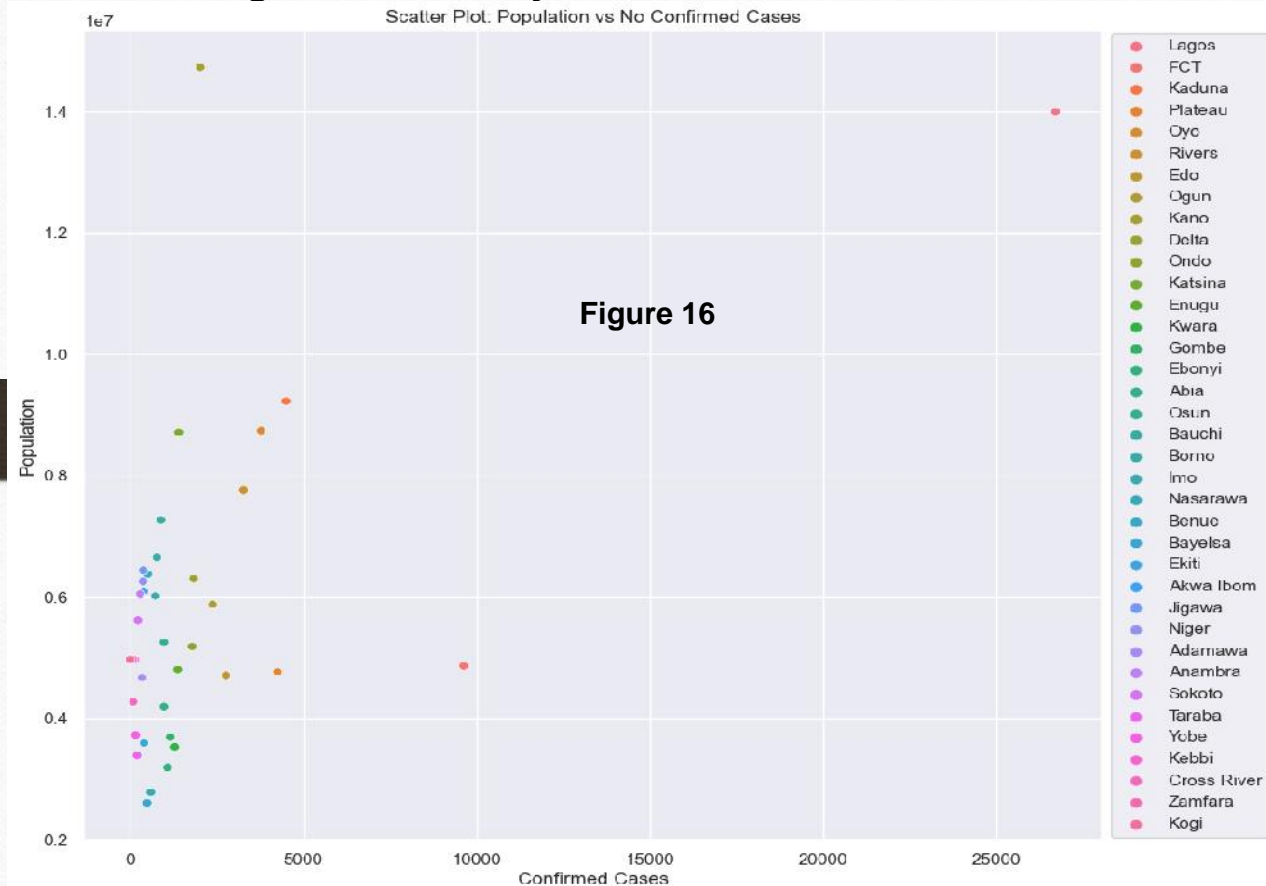
It helps understand relative economic performance, competitiveness, and identify potential areas for collaboration or learning.

7. **Investment Decisions:** Investors analyze quarterly GDP data to make informed investment decisions. It provides insights into the overall health of the economy, potential risks, and opportunities that may impact specific industries or sectors.

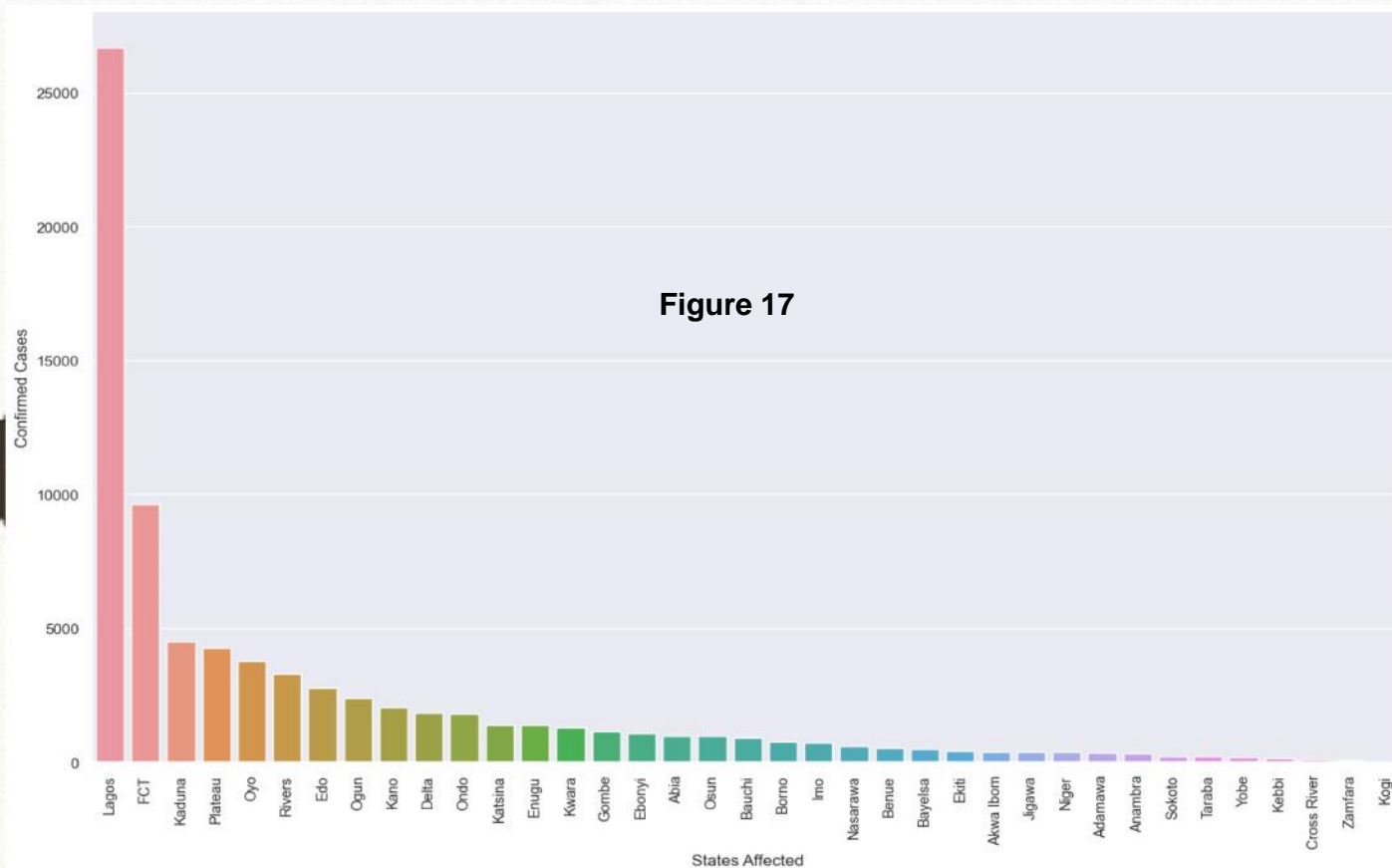
Overall, quarterly GDP data analysis provides valuable information about the current state of the economy, short-term trends, and sector-specific performance. It helps stakeholders make informed decisions, formulate policies, and respond to economic changes effectively.

From figure 15 above show data analysis of GDP of the country from 2014 to 2019. It is noticed from the average cumulative GDP; base the on the calculation the fourth quarter has the highest GDP due low pandemic occurrence . There is a drop in the GDP in 2014 and 2016, which is the incept of the pandemic. This is more clear the distribution for respective years in Figure 15

## COVID-19 Nigeria Data Analysis from the external dataset



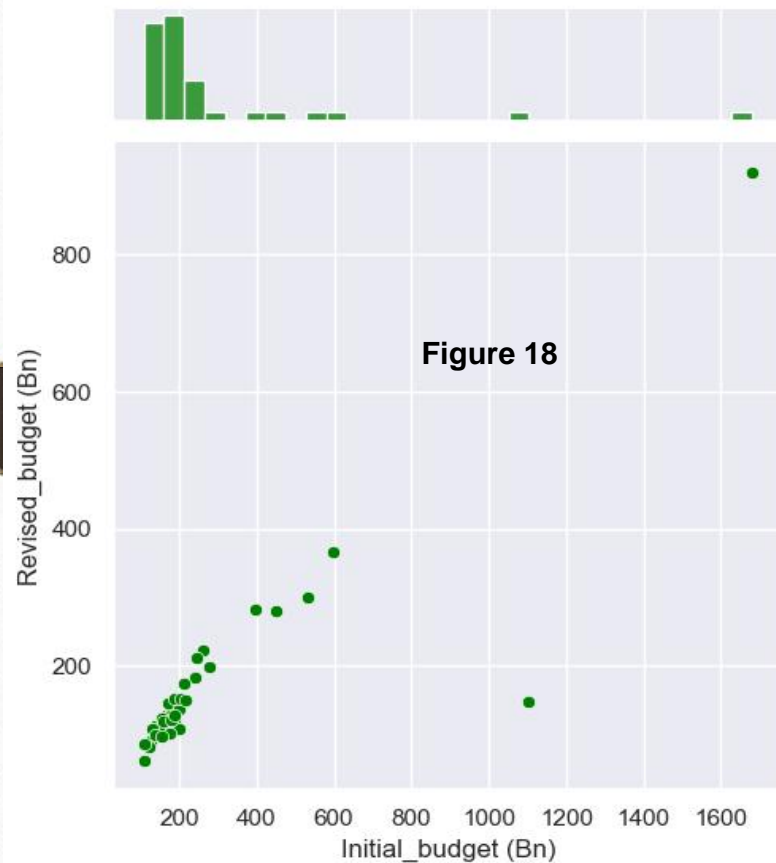
The scatter plot shows that there is a positive correlation between the number of confirmed COVID-19 cases and the population of a state. This means that states with larger populations tend to have more confirmed COVID-19 cases. However, there is also a lot of variation in the data, so it is important to not make too many inferences from this scatter plot. From Fig16. The scatter plot shows the relationship between the number of confirmed COVID-19 cases and the population of different states in Nigeria. The points that are clustered on the scattered plot. That is, low number confirmed COVID-19 Cases with low population. The point far from the clustered point; shows, high Confirmed cases due to high population.



From fig 17 shows the number of confirmed COVID-19 cases in different states in the Nigeria. The bar plot shows that Lagos, FCT, Kaduna, Plateau and Oyo have the most confirmed COVID-19 cases.



## ❖ Budget Data Analysis



It was stated that states across the country reduced their initial budget due to the impact of COVID-19 on the economy. The data is to be used to determine the impact of COVID-19 on the economy. From this data, I was able to compute the percentage increment in the budget. The plot below shows the relationship between the initial and revised budget:

### Inference:

The average percentage change in budget (%) = 29.7%

The minimum percentage change in budget (%) = 12.7% (Kastina State)

The maximum percentage change in budget(%) = 86.6% (Cross River State)

### CONCLUSION

Based on the analysis and series of inferences above from various data provided, we can infer that the COVID-19 pandemic has a great effect on the economy of Nigeria and lives of many citizens of the country during the prevalence period of the virus. It is noteworthy the leaders of the country must try as much as possible to combat this pandemic cause in terms of health and economic mitigations.