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

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.

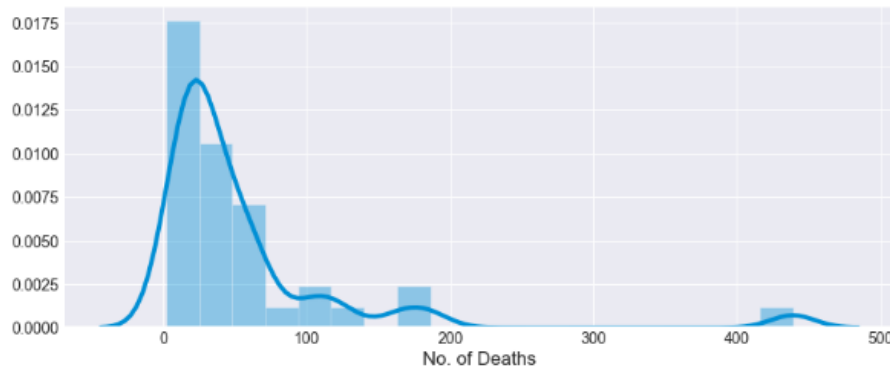❖ **Understanding the distribution of features in the NCDC DATA across each state**
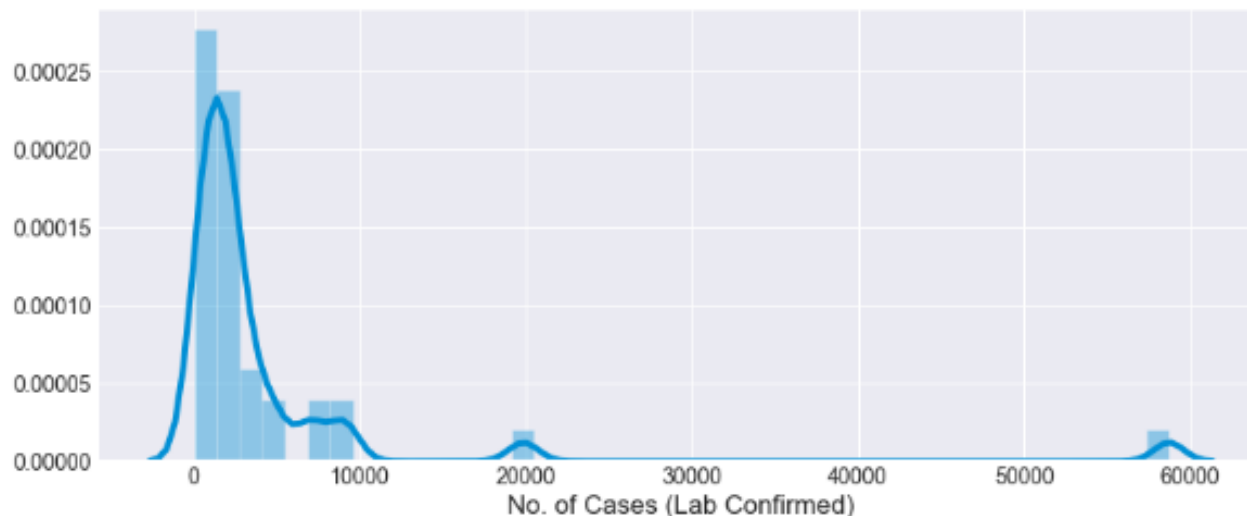
Figure 1.1
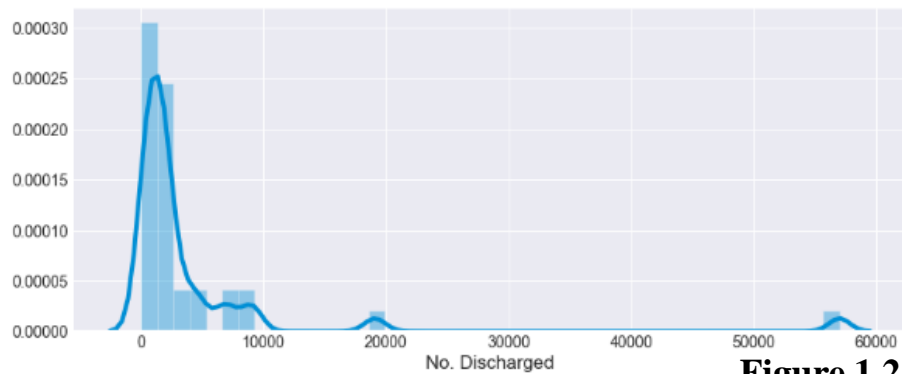
Figure 1.2

Figure 1.3

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, confirmation and death 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.
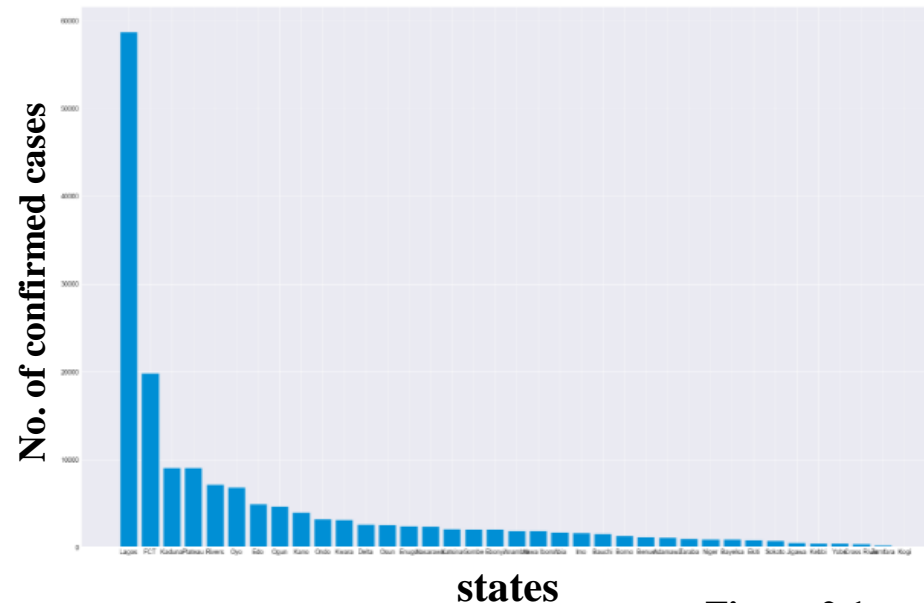
Inference:

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

The top 10 states with the highest number of confirmed cases are Lagos with 58713 cases, FCT with 19841 cases, Kaduna 9068 cases, Plateau 9060 cases, Rivers 7169 cases, Oyo 6855 cases, Edo 4907 cases, Ogun 4680 cases, Kano 3967 cases, and Ondo 3248 cases.

Figure 2.1

Inference:

The top 10 states with the highest discharged cases are Lagos with 56990, FCT 19104, Plateau 9002, Kaduna 9000, Rivers 7040, Oyo 6729, Edo 4715, Ogun 4627, Kano 3849, and Kwara 3067

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.

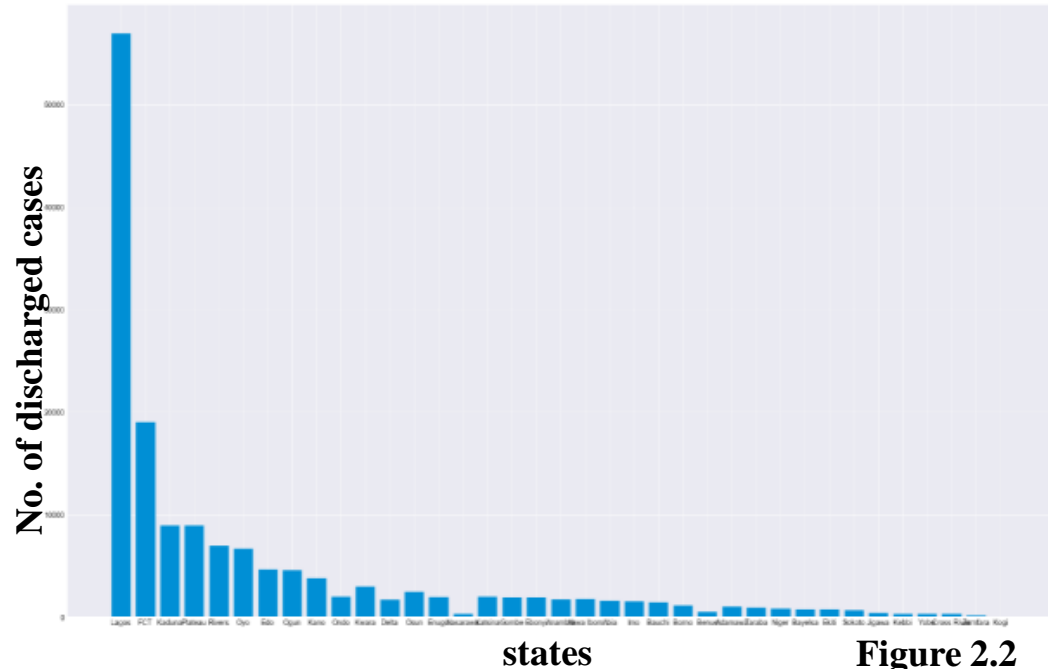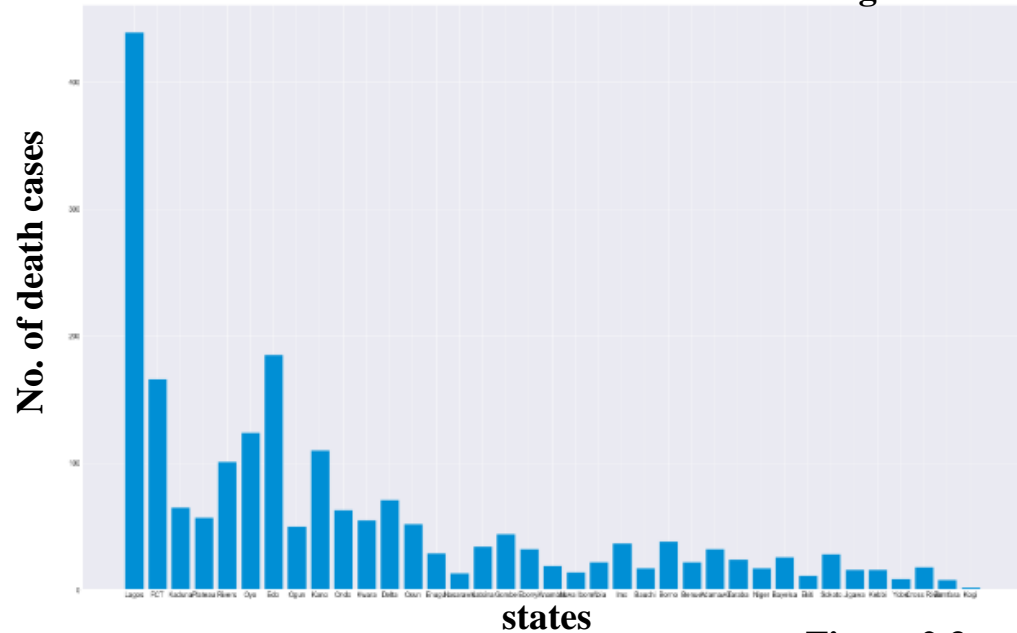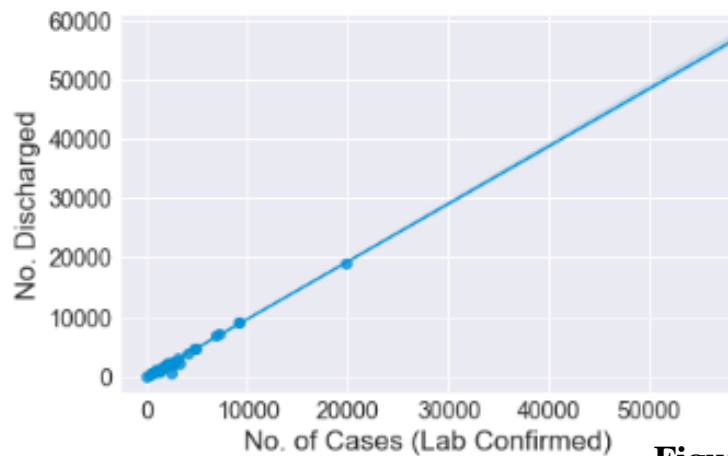The top 10 death cases are Lagos with 439, Edo 185, FCT 166, Oyo 124, Kano 110, Rivers 101, Delta 71, Kaduna 65, Ondo 6, Plateau 57.

Figure 2.2

Figure 2.3

The analysis above is based on the univariate analysis, 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.

Figure 3.1

❖ Correlation between the features of NCDC data

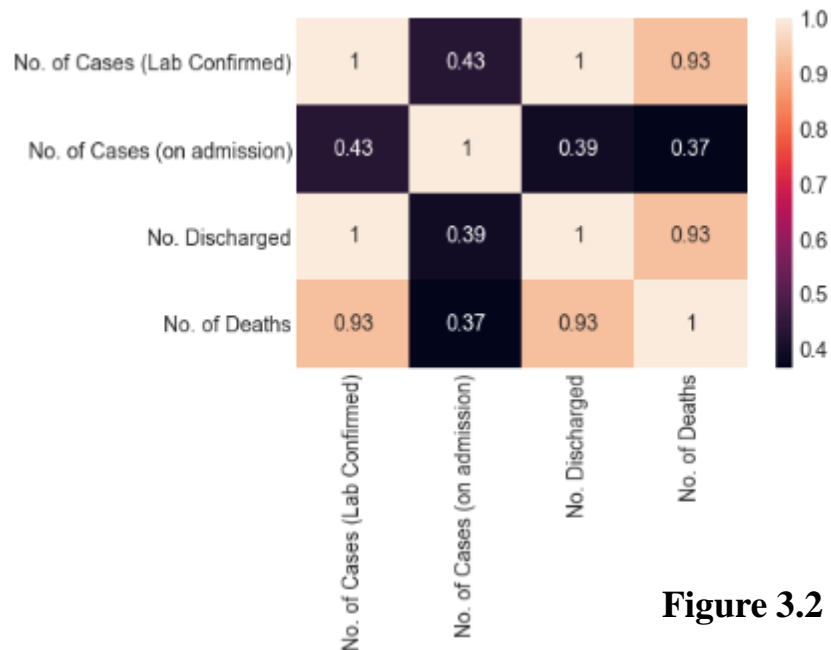Using Pearson correlation coefficient which shows the strength of relatedness between features

Inference:

In support of the relatedness between the No. of discharge and No. of cases. **Figure 3.2** shows that there is high correlation between the Confirmed cases, Discharged cases and the Death cases

Figure 3.2

❖ 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.
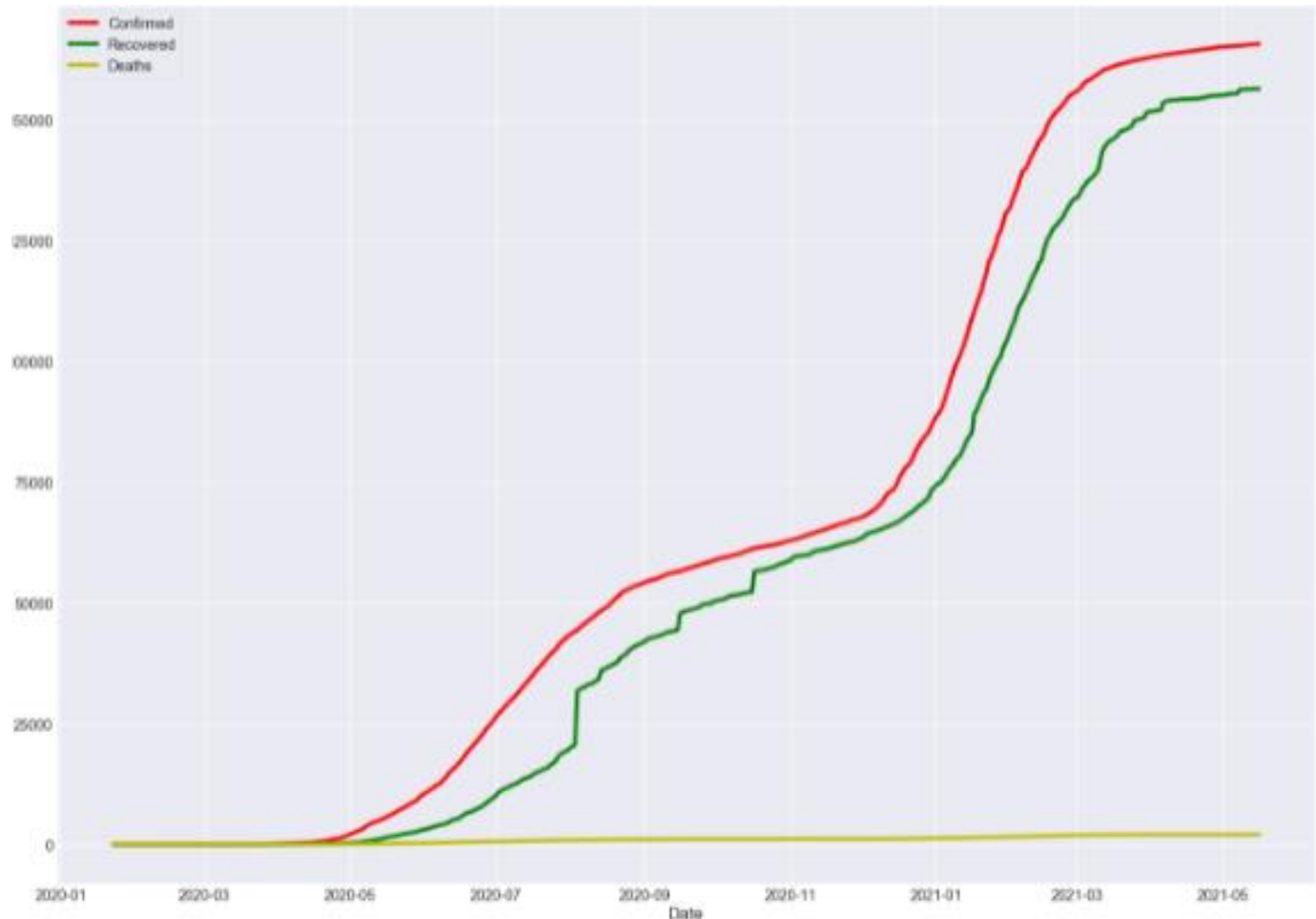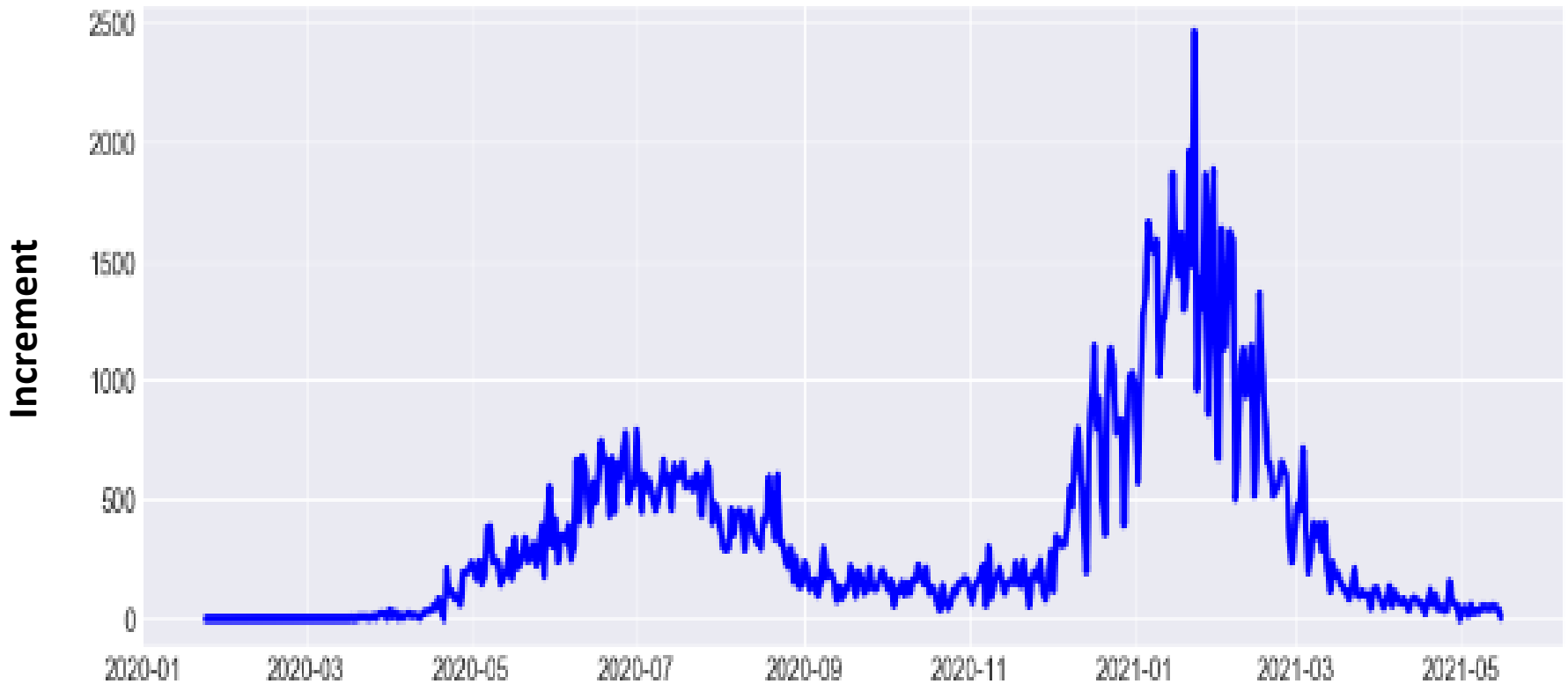

Figure 4.1

Figure 4.2

Inference:

Upon computation of the isncrement 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 November 2020 to early March 2021. Maximum infection increment is 2464. The date with the maximum infection rate was on the 23rd of January 2021 with an infection rate of 2464.

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

I 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. Below are some chart showing the relationships

Figure 5.1

Figure 5.1 shows the relationship or spread of the virus based on the population of respective state affected.
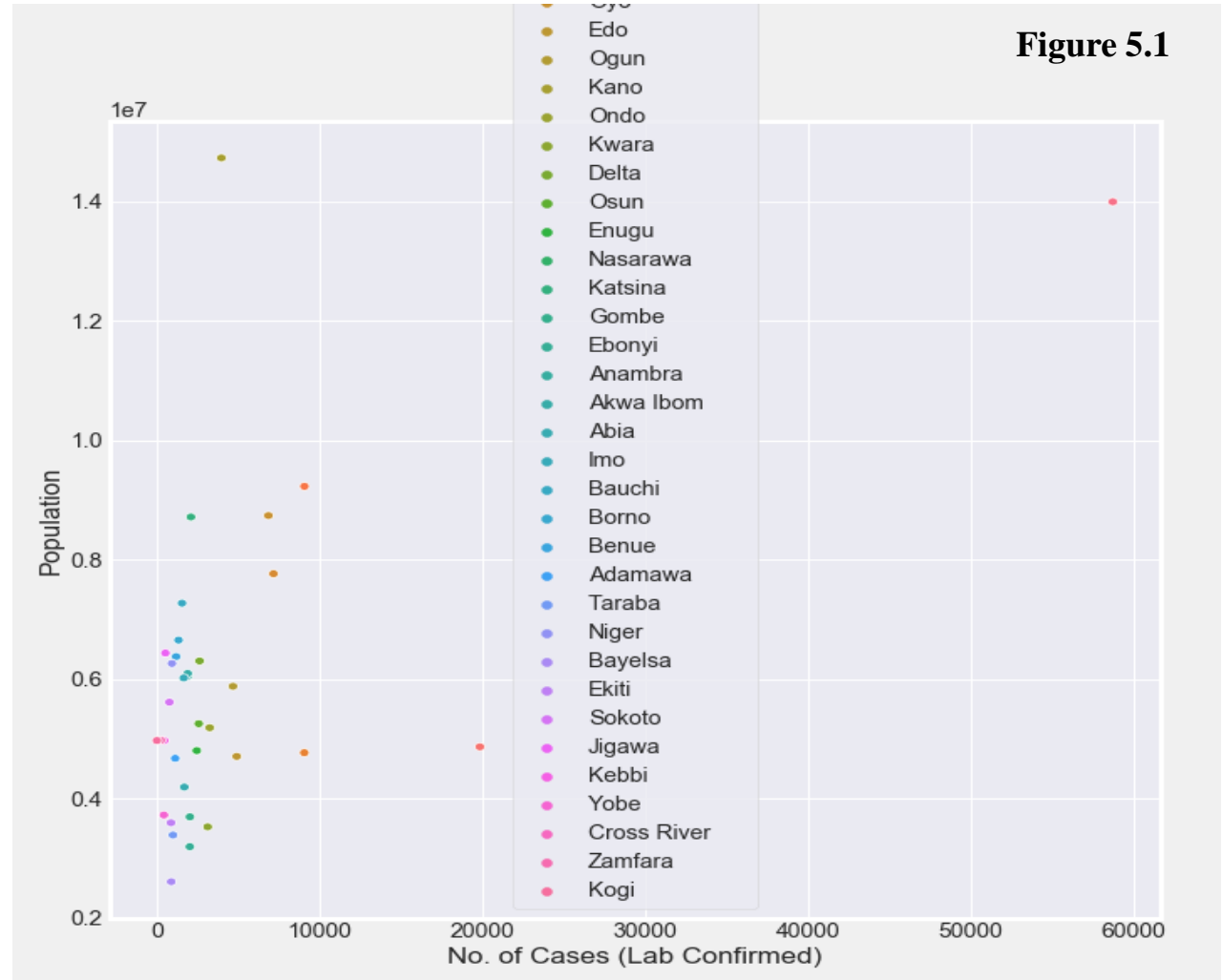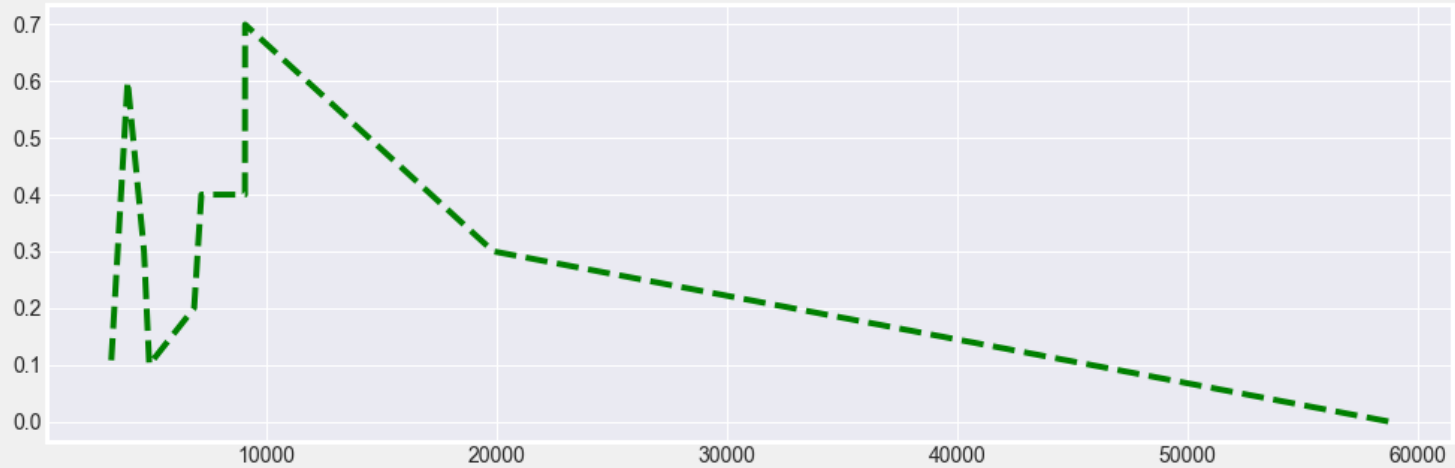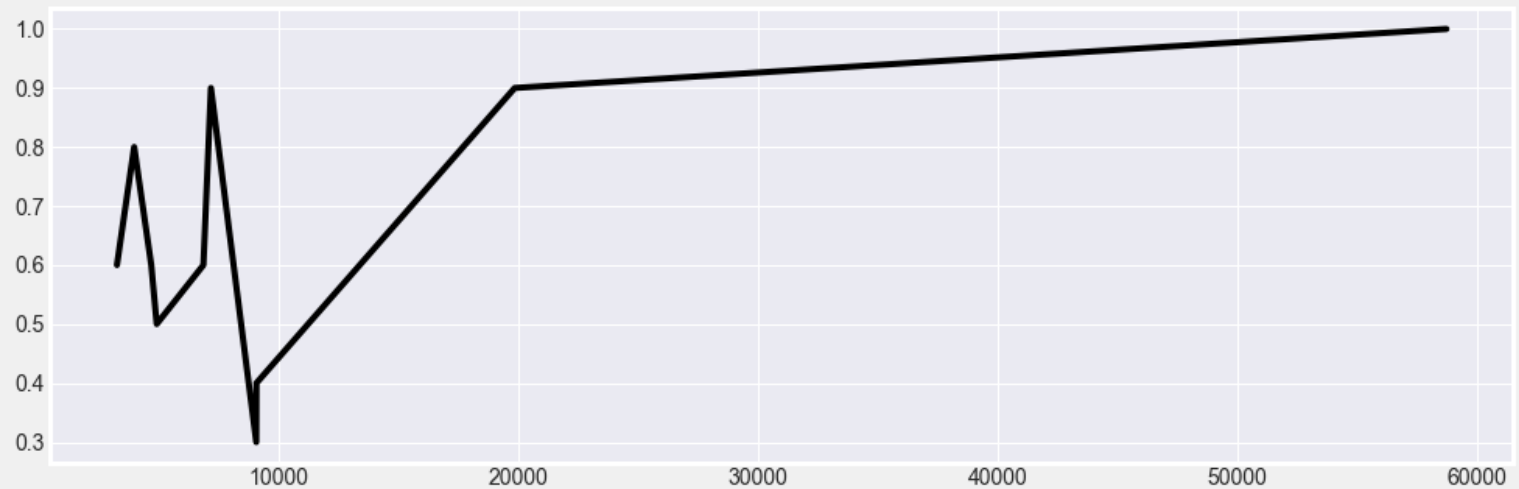
Figure 5.2

Inference: Graph showing the areas with low CCVI have relatively high number of confirmed cases

Figure 5.3

Inference: Densely populated area also contributed to the source capsule of high number of confirmed cases

❖ Quarterly GDP data analysis.

This gives us the quarterly GDP in different years ranging from 2014 to 2020. It is noticed from the third quarter based on the calculation for the average cumulative GDP that it has the highest value but in that column, we notice a fall in the quarterly GDP from 2019 to 2020, which is the incept of the pandemic. This is more clear the distribution for respective years in **Figure 6.1**
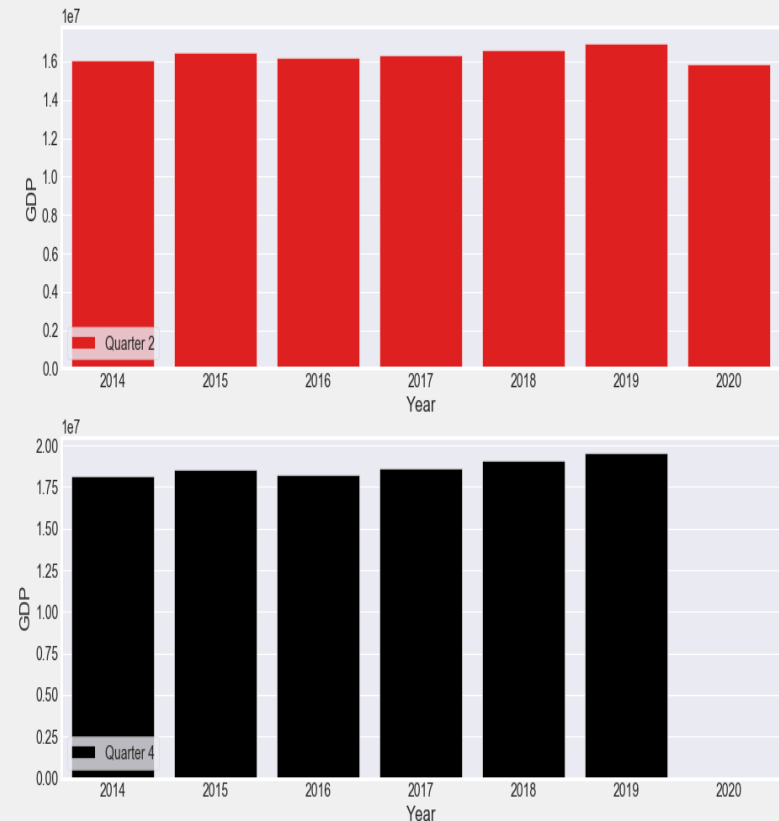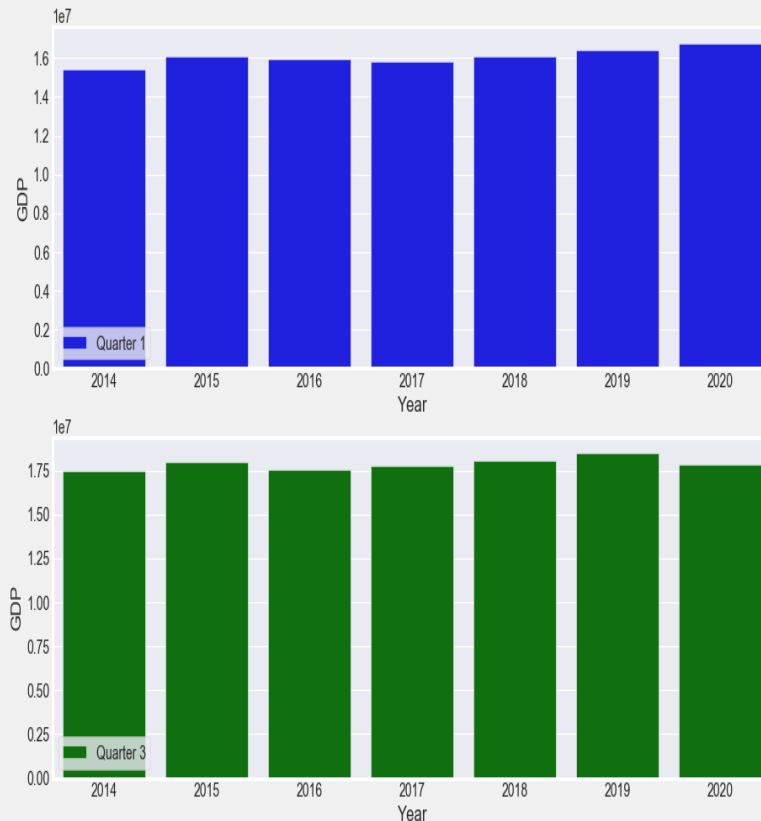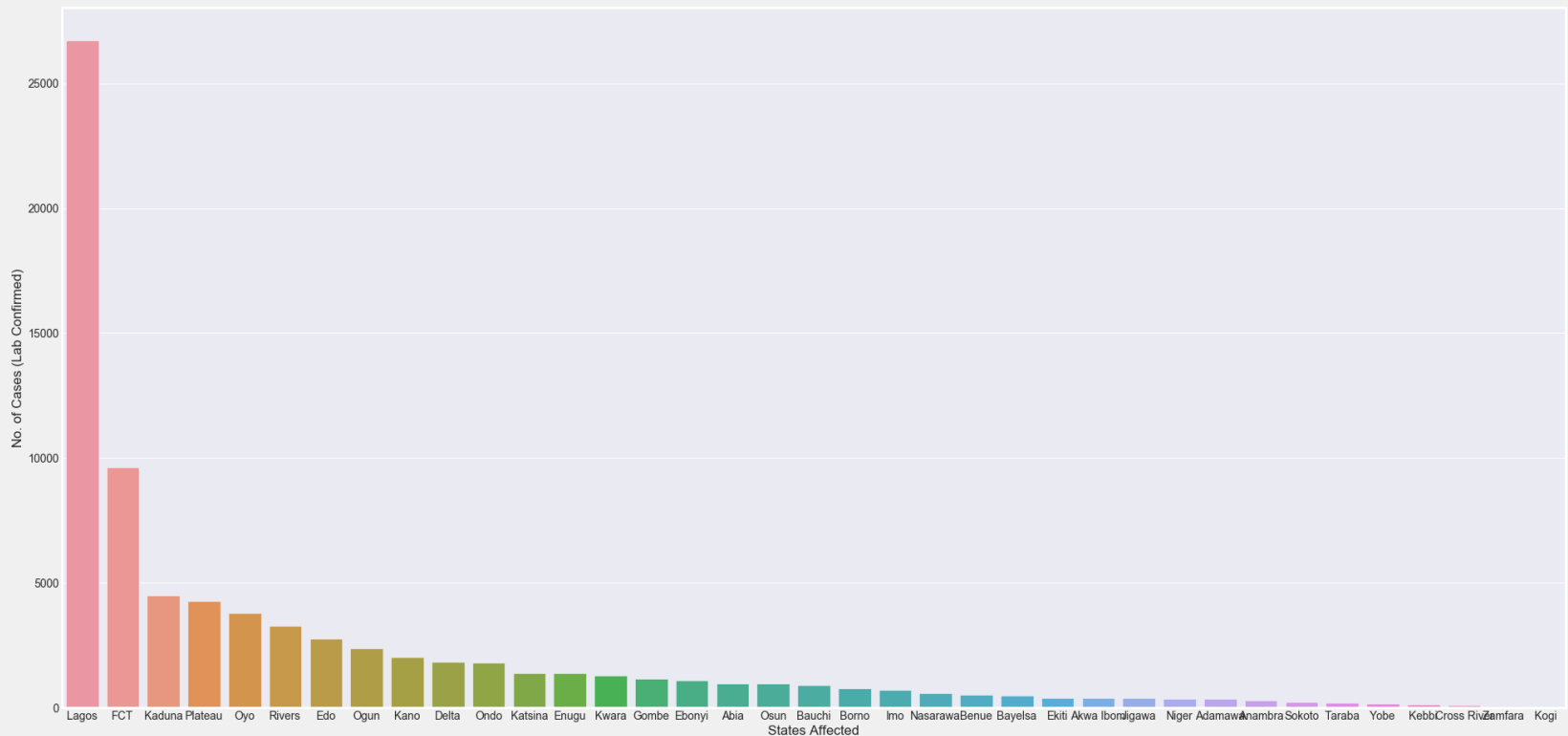

Figure 6.1

❖ COVID NIGERIA DATA ANALYSIS.

This gives similar inferences with the NCDC data analyzed above, we can noticed the bias in the distribution of the different features in the dataset which includes Number of Cases (Lab Confirmed), Number of Cases (on admission), Number of Discharged and Number of Deaths which is similar to that of NCDC data

Figure 7.1

Inference: We noticed from the distribution of the virus across Nigeria within the period being considered. Lagos overall confirmed cases is very high compared to other state.

Figure 7.2
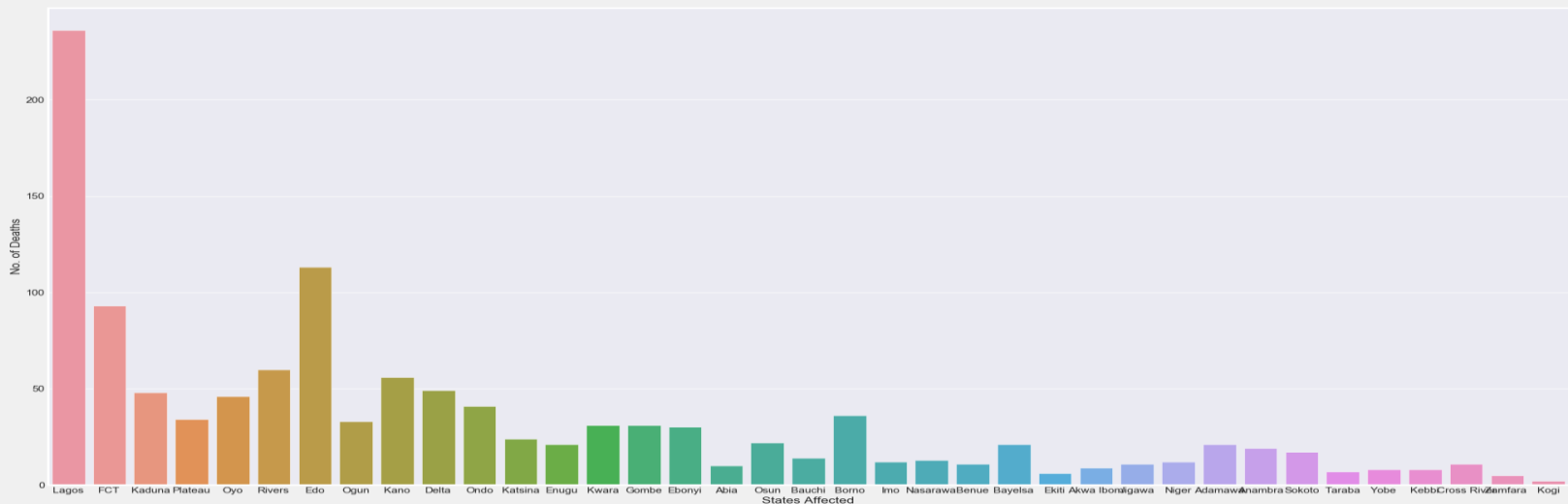

No. of Death vs State affected

Figure 7.2 & Figure 7.3: This is similar to Figure 2.2 and 2.3 in terms of the distribution. The plot shows distribution of number of death and number of discharged across each state with Lagos as lead in this metrics.
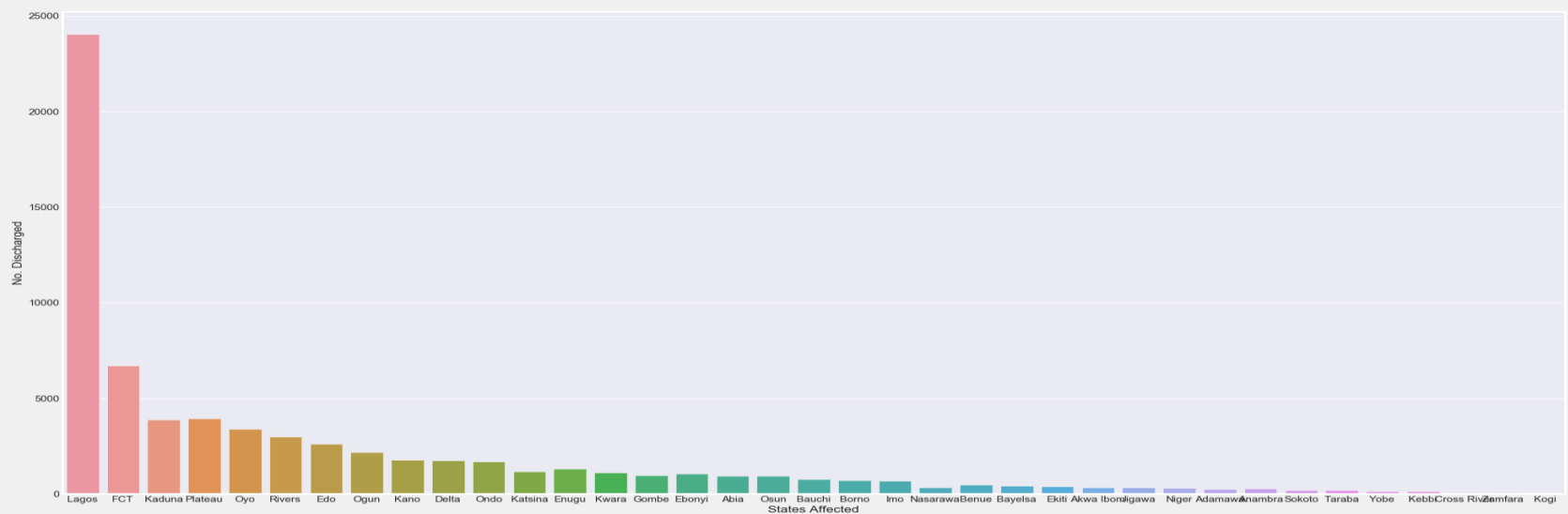

No. of Discharged vs State affected

Figure 7.3

❖ 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:
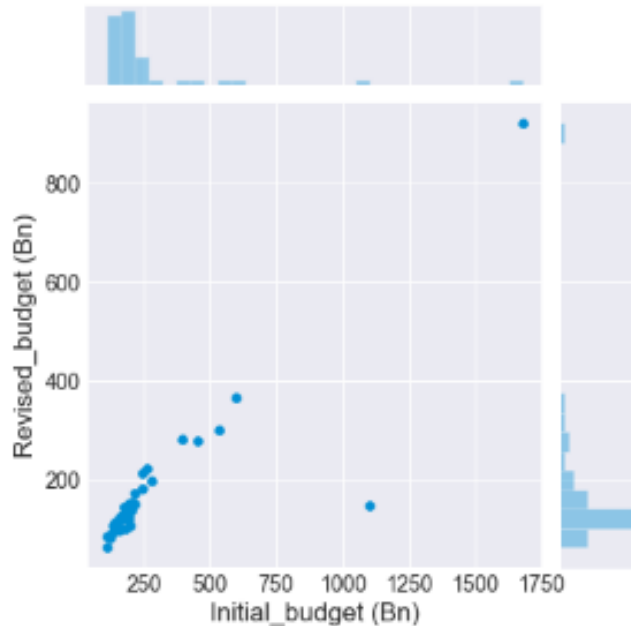
Inference:
The average percentage change in budget $(\%) = 29.7\%$

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.