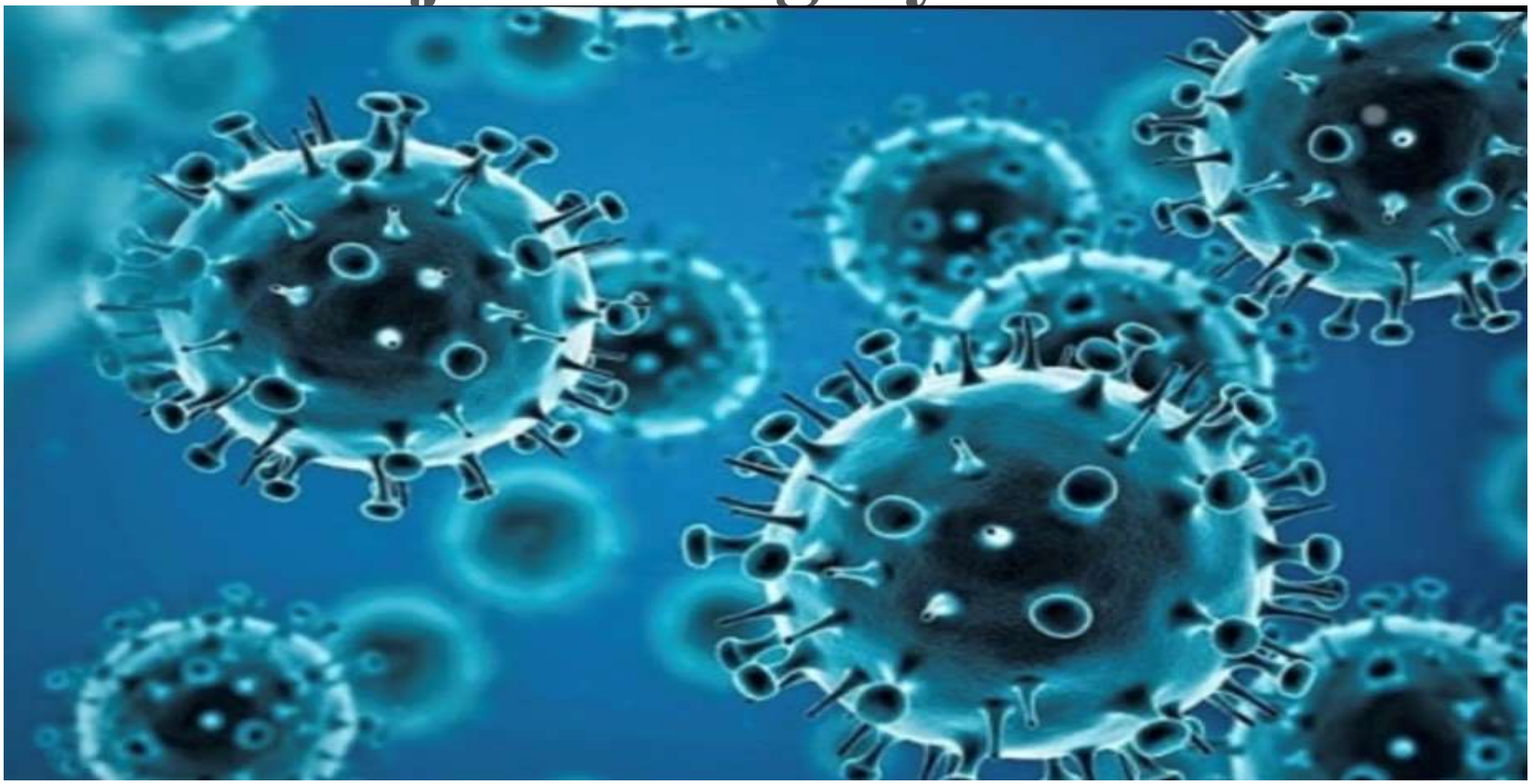
Nigeria Covid-19 Data Analytics Project Using Python



PROJECT OVERVIEW

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.

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.

PROJECT APPROACHES

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

- Web scraping, and importing from data sources.
- Good use of Pandas and visualization tools to communicate insights.
- Summary report documentation to derive more insights.

DATA INFORMATION

The data source is divided into different parts, and combined to perform analysis and provide insights.

- The Nigeria Centre for Diseases Control (NCDC) monitors the country's COVID-19 situation, and releases data on 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. The data for Nigeria will only be extracted from there.
- 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.
- The Real Domestic Gross Product(GDP) data for Nigeria will help to determine the impact of COVID-19 on the economy.
- State Budget Data will also be used to determine the impact of COVID-19 on the economy across the states in the country.

ANALYSIS QUESTION

- ✓ Some questions that my exploration seek to answer upon analysis of given data and the data collected from diverse sources are:
- ✓ Explore the distribution of the virus spread across each states, the numeric effect on each state through the distribution of the confirmed, discharged and the death cases in Nigeria states.
- ✓ 2.From John Hopkin's data, get the active cases in all countries and extracting the data for Nigeria out for analysis to help investigate the relationship between the aforementioned features and to draw conclusions based on their relation.
- ✓ Examine the connection between the number of confirmed cases, recovery cases, and deaths in order to calculate the infection rate throughout the country. Investigate and draw conclusions from the distribution of the infection rate over the date provided in the dataset.
- ✓ An exploration and evaluation of specific indicators within external data are being conducted. These indicators comprise of Fragility, vulnerability, epideiology, prevalence, health care, population density, transportation and socioeconomic factors.
- ✓ The data will be explored using these indices in order to create graphs and charts that are more effective for conveying information and drawing conclusions.
- ✓ The external GDP data will be analyzed to obtain a better understanding of the impact of covid-19 pandemic on the economy of the country.
- ✓ Deeper understanding of the impact of the virus on Nigeria's economy, and additional insights will be obtained from state and budgetary considerations.

ANALYSIS METHODOLOGY

❖ Web scraping, and importing from data sources.

❖ Good use of Pandas and visualization tools to communicate insights.

❖ Summary report documentation to derive more insights.

ANALYSIS EXPLORATION AND INFERENCES (RESULT)

Here, the use of diverse charts will be utilized to explain the result of my analysis and inferences made from the data in accordance to the analysis question.

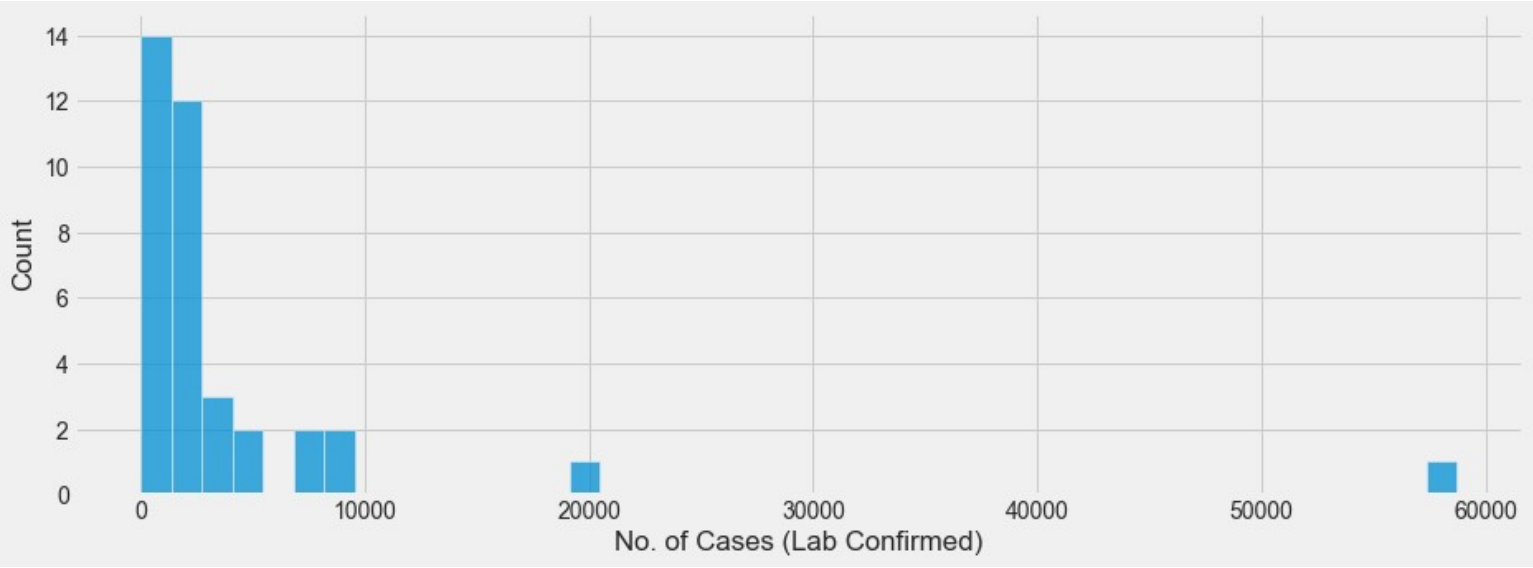


Fig 1.1

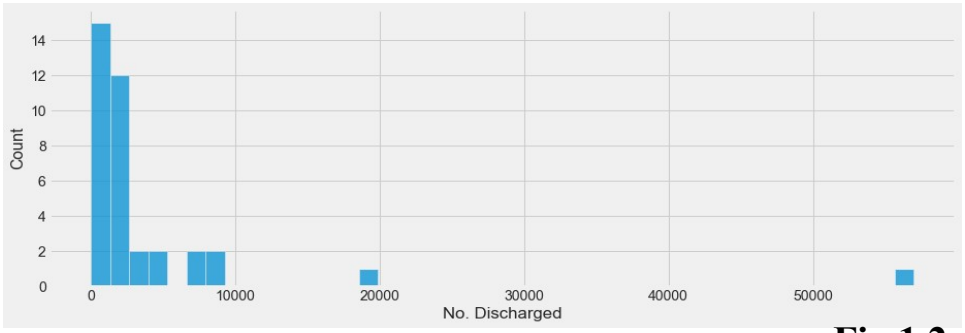


Fig 1.2

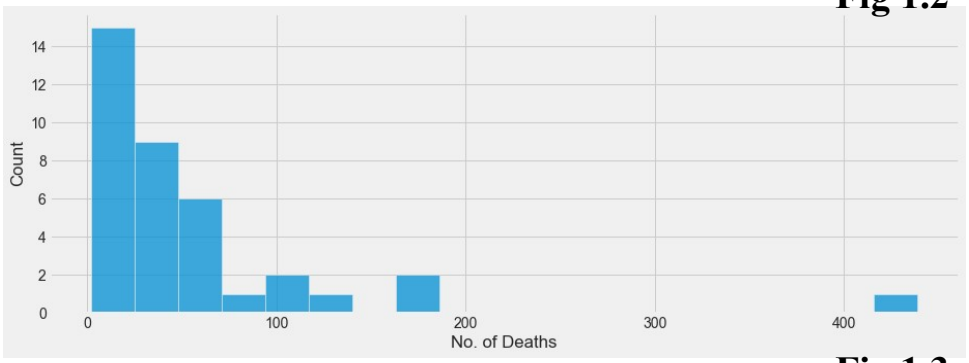


Fig 1.3

Conclusion:

Based on the virus distribution in Nigeria during the period under consideration, we observed that Lagos has a significantly higher number of confirmed cases compared to other states.

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

Upon examining the distribution of the three features in the NCDC data, we have observed a certain bias in the plot. This suggests that there are significantly higher numbers of discharges, confirmations, and deaths in a few states as compared to other states in the country. To gain a clearer insight into figures 1.1 to figures 1.3, we can from the bar plot below

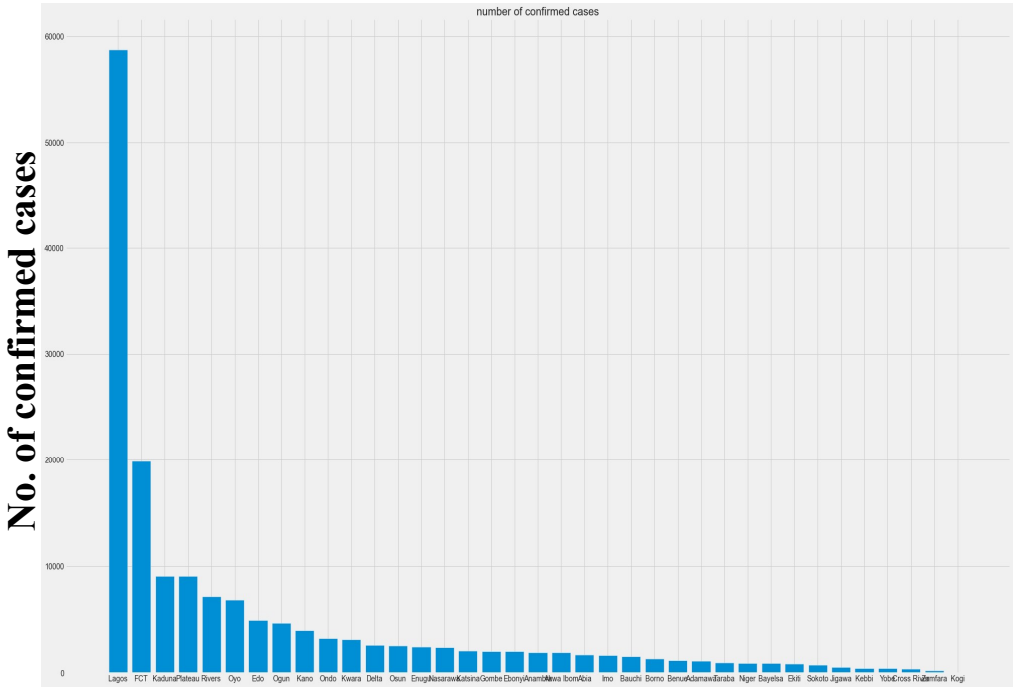
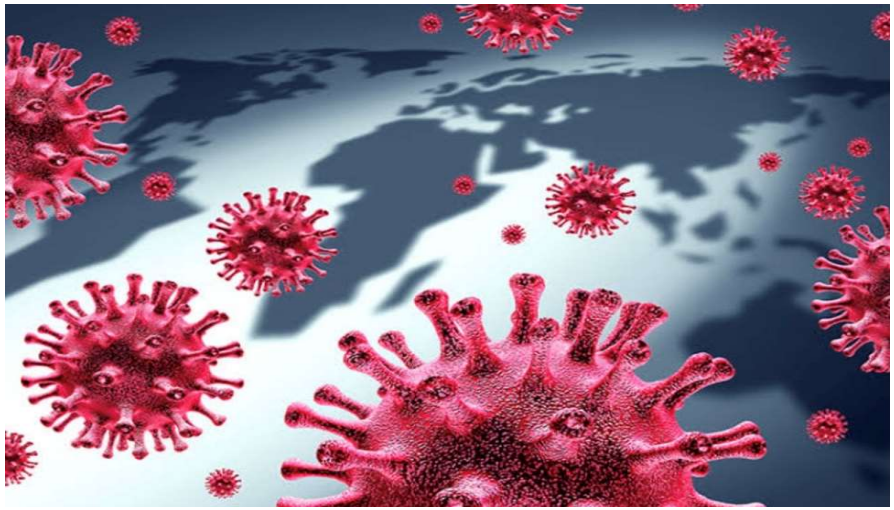


Figure 2.1



Conclusion:

The top 10 states with the highest number of 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.

Using this information, we can make a comparison with our prior findings and deduce that a large portion of the states with higher numbers of confirmed cases also exhibit elevated rates of discharge.

The top 10 states with the highest number of death cases are Lagos with 439, Edo 185, FCT 166, Oyo 124, Kano 110, Rivers 101, Delta 71, Kaduna 65, Ondo 63, and Plateau with 57.

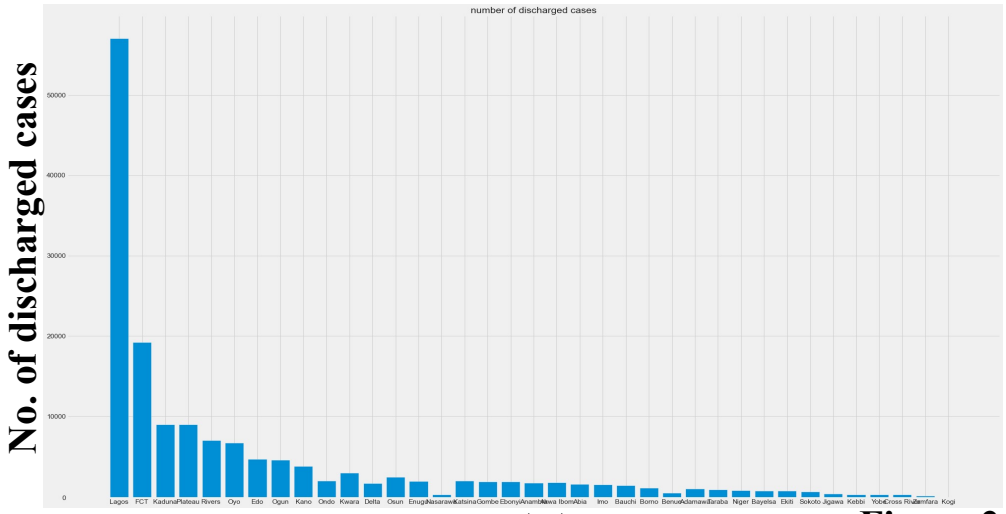


Figure 2.2

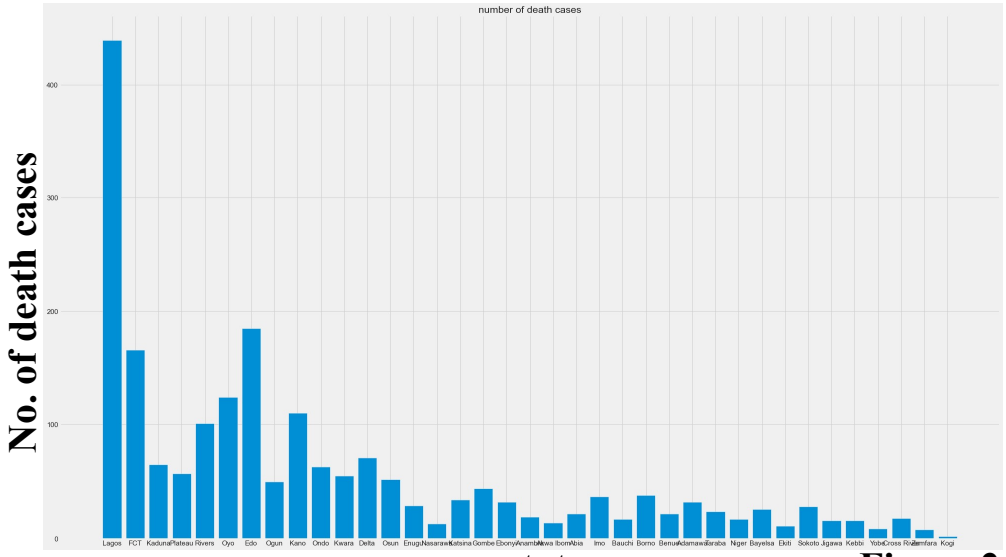
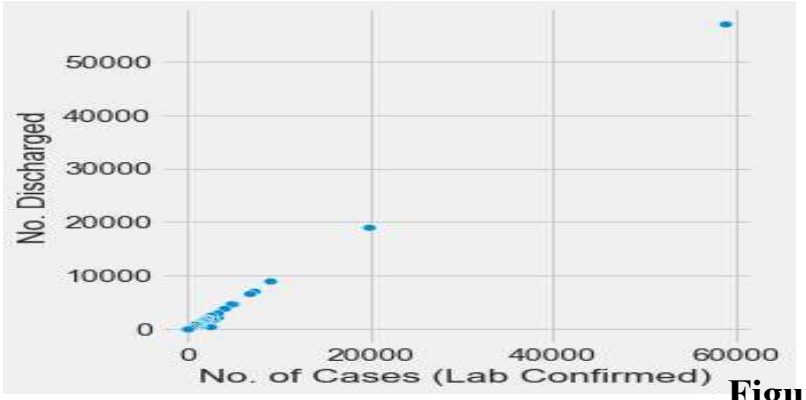


Figure 2.3



The above analysis relies on univariate analysis, but we can extend it by demonstrating that the correlation between the count of confirmed cases and the count of discharges is strong when considering the affected states. This relationship is represented by the slope of the bivariate plot, which indicates the discharge rate.

Figure 3.1

❖ **Correlation between the features of NCDC data**

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

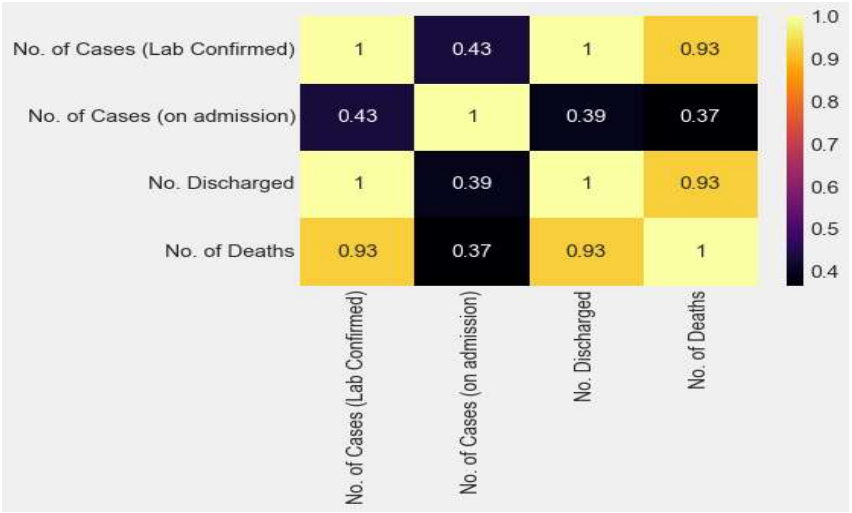


Figure 3.2 provides evidence in favor of the connection between the number of discharges and the number of cases. It demonstrates a strong correlation among confirmed cases, discharged cases, and death cases.

Figure 3.2

❖ The Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE) provides daily updates on confirmed, death, and recovered cases in various countries. We extracted data for Nigeria from this source, performed data transformation on the three datasets, and extracted relevant features. The plot below depicts the correlation between case confirmations, recovery rates, and death rates.

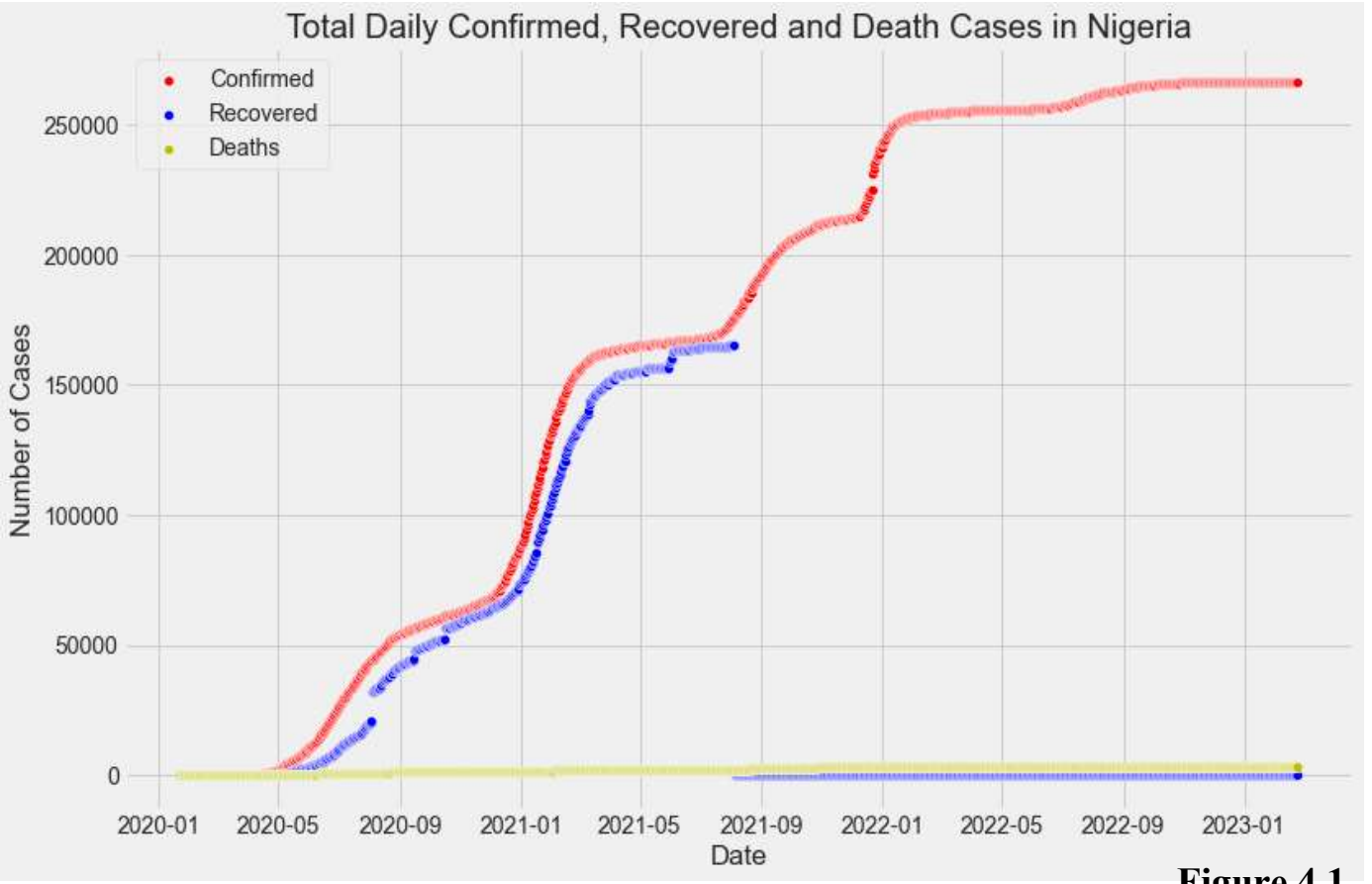


Figure 4.1

Conclusion

It can be deduced that some percentage of individuals who tested positive are recuperating.

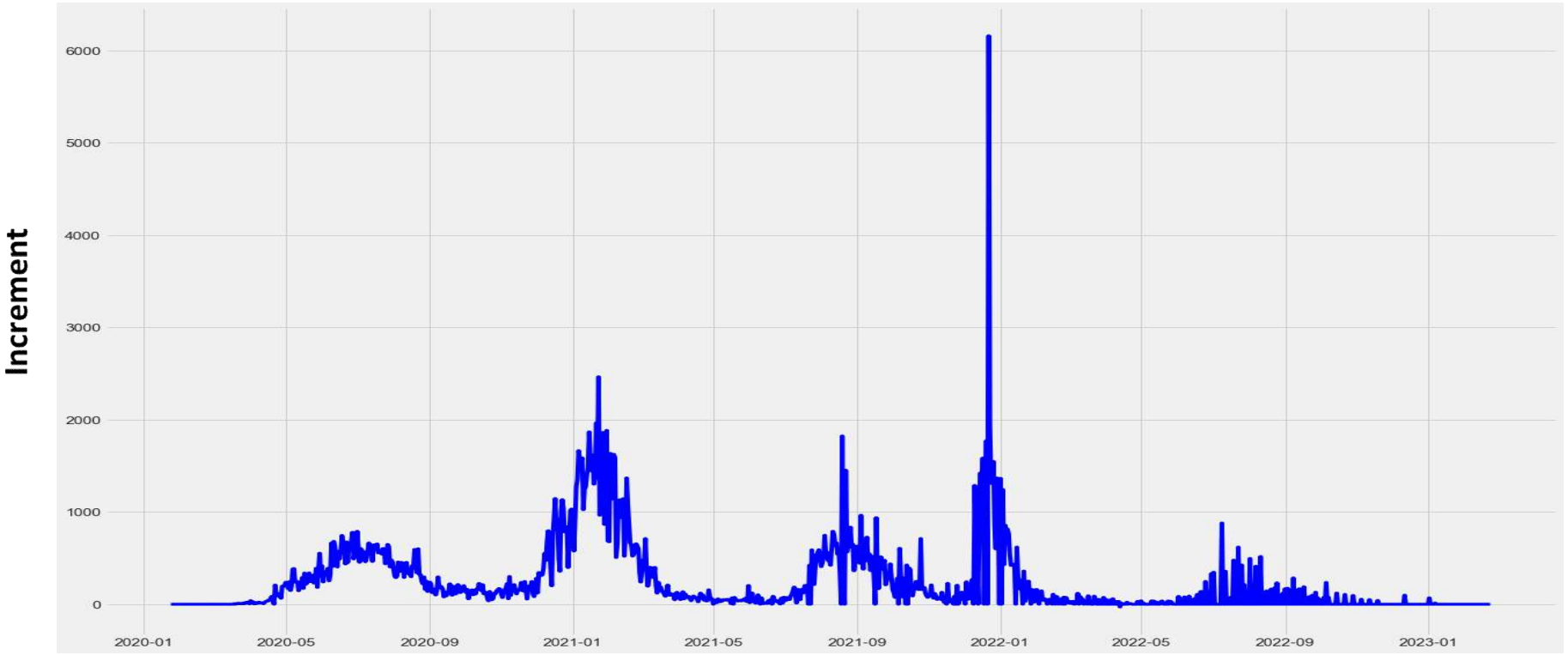


Figure 4.2

Conclusion

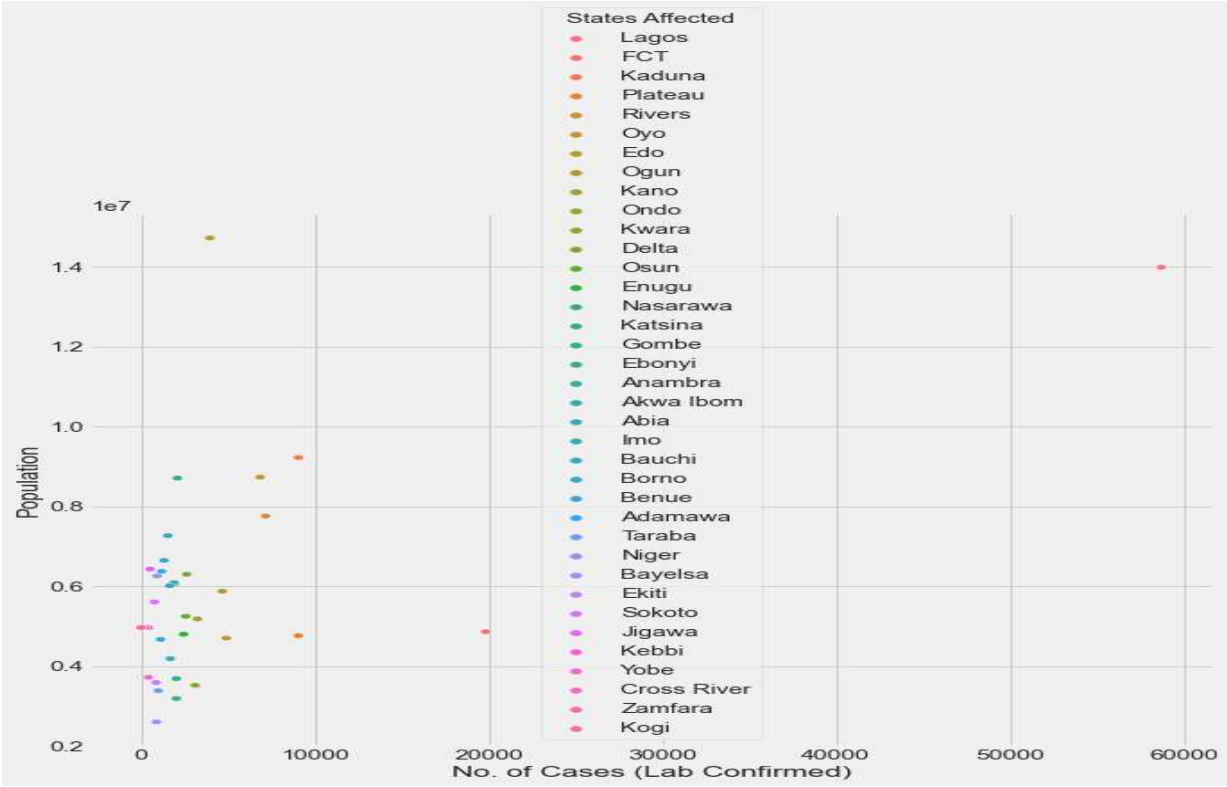
After calculating the increase in confirmed cases on a daily basis, it becomes evident that the infection rate experienced a significant surge between September 2021 and early January 2022. The highest recorded increment was 6158 cases. The date that had the highest infection rate was January 23, 2021, with a total of 6158 cases.

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

I conducted an analysis of specific indices in the external data, such as fragility, vulnerability, epidemiology, prevalence, healthcare, population density, transportation, and socio-economic indicators. By exploring the data based on these indices, I created graphs and charts that effectively communicate the statistical inferences. The charts presented below illustrate the relationships found.

Figure 5.1

In **Figure 5.1**, the correlation of the virus is displayed in relation to the affected population in each state.



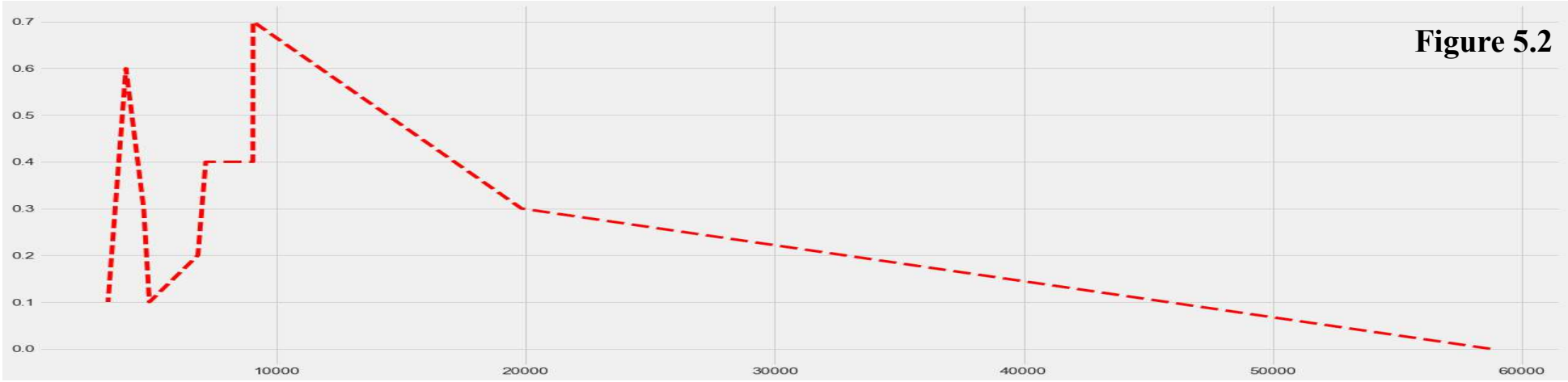


Figure 5.2

The Graph above shows the areas with low CCVI have relatively high number of confirmed cases.

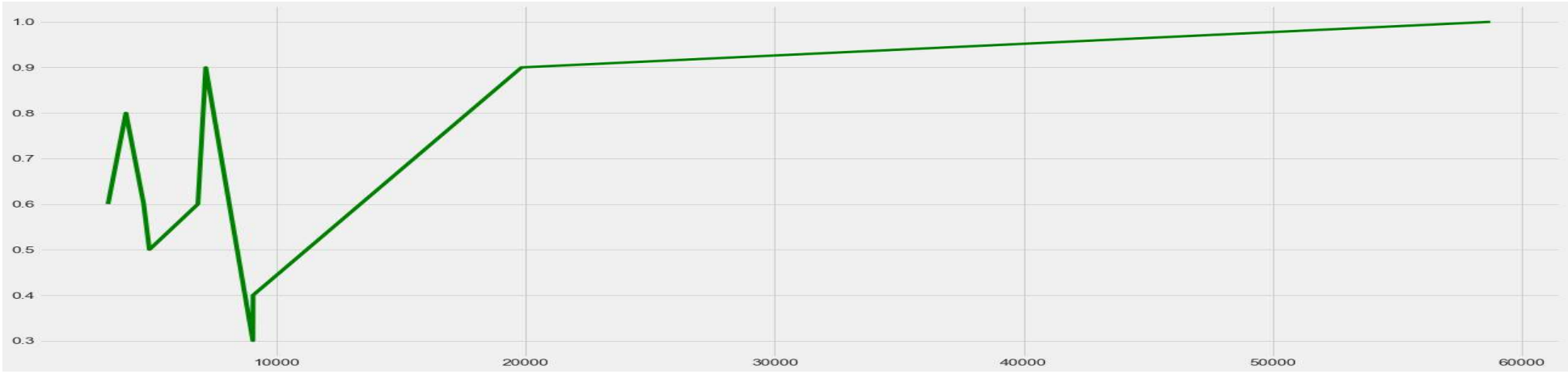


Figure 5.3

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

Quarterly GDP data analysis.

The analysis presents the quarterly GDP data for various years between 2014 and 2020. It is observed that the average cumulative GDP reaches its peak in the third quarter. However, a decline in the quarterly GDP is noticed in the same column from 2019 to 2020, which coincides with the onset of the pandemic. The distribution for each year is more evident in the **Figure 6.1** analysis.

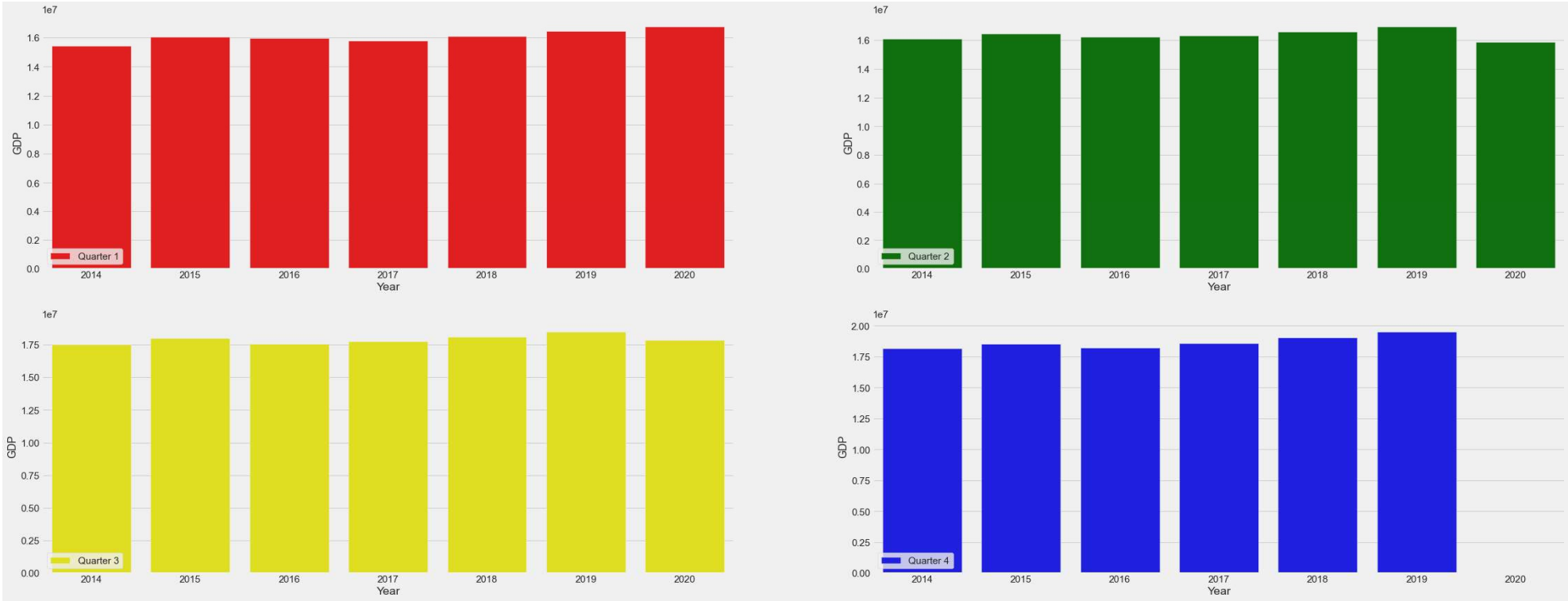


Figure 6.1

❖ **COVID-19 NIGERIA DATA ANALYSIS**

We can observe similar conclusions when comparing this data analysis to the previously analyzed NCDC data. One notable finding is that the distribution of the different features in the dataset, such as the Number of Lab Confirmed Cases, Number of Cases on Admission, Number of Discharged Cases, and Number of Deaths, exhibits a bias that closely resembles that of the NCDC data.

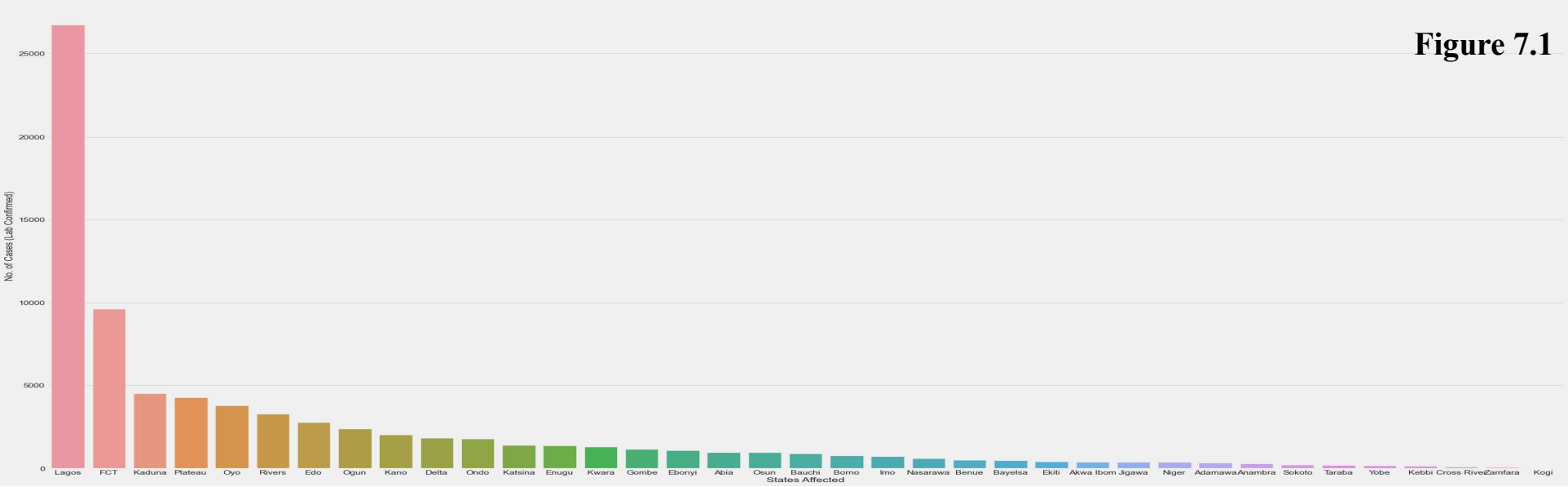


Figure 7.1

Conclusion:

Our analysis of the virus distribution in Nigeria during the specified period reveals that Lagos has a notably higher number of confirmed cases in comparison to other states.

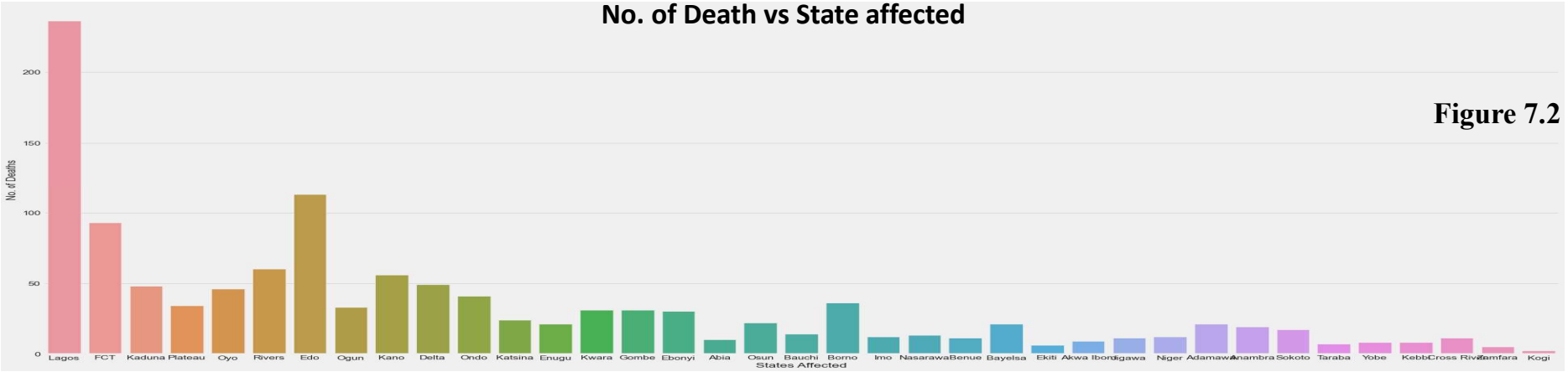


Figure 7.2

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.

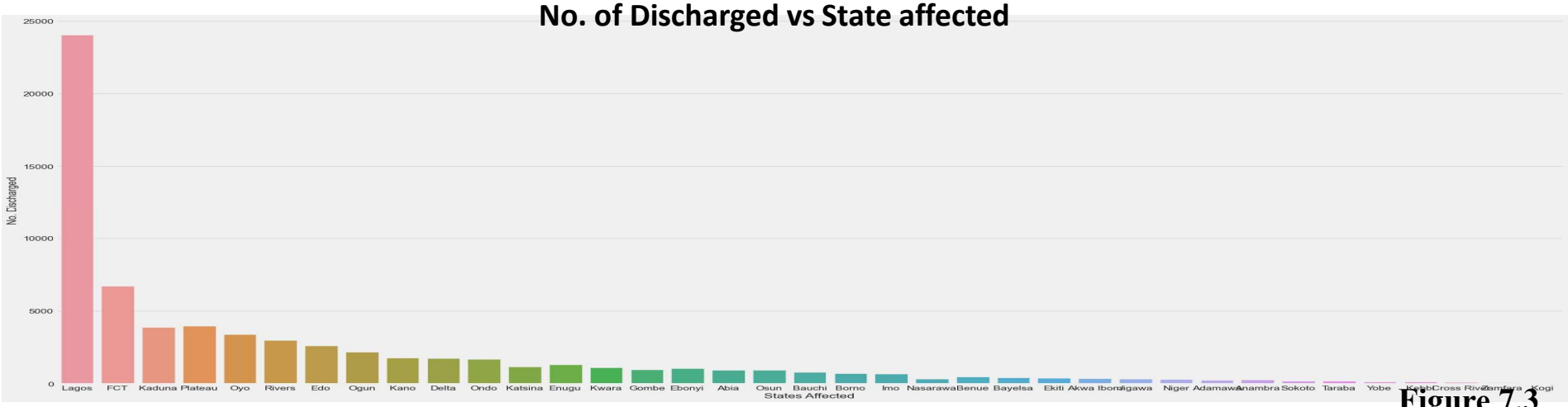
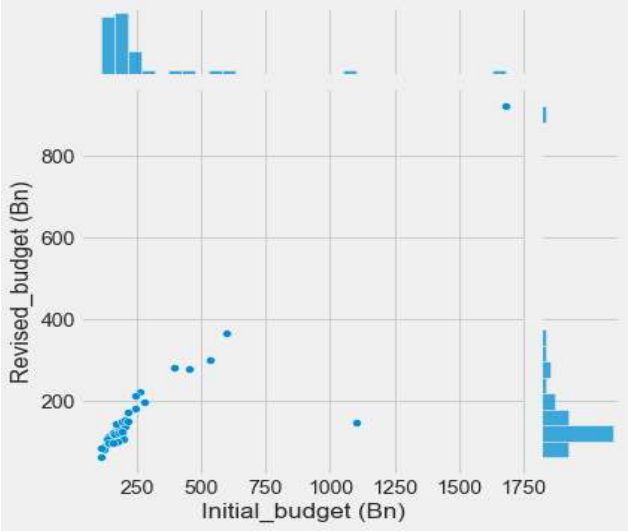


Figure 7.3

❖ BUDGET DATA ANALYSIS

As a result of the economic impact of COVID-19, states in Nigeria reportedly decreased their initial budgets. This dataset was collected to evaluate the extent of the pandemic's impact on the economy. Using the data, I calculated the percentage increase in the budget and plotted a graph to visualize the relationship between the initial and revised budgets:



Conclusion:

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:

After examining the data and drawing various conclusions, it can be inferred that the COVID-19 pandemic had a significant impact on both the economy and the lives of Nigerian citizens during the period when the virus was prevalent. It is crucial that the country's leaders take measures to combat the pandemic's effects, both in terms of healthcare and economic mitigation.