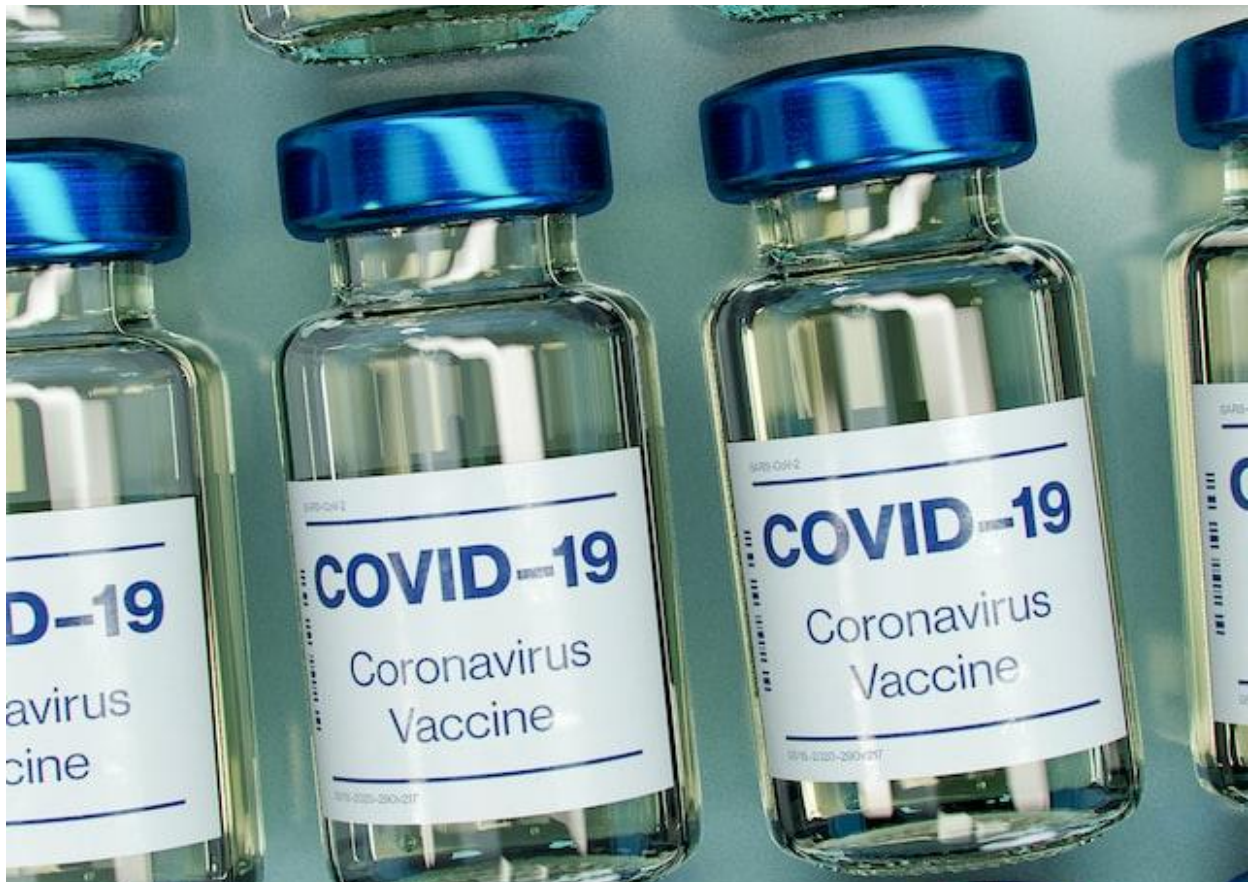# NIGERIAN COVID-19 ANALYSIS USING PYTHON

A USTACKY CAPSTONE PROJECT FOR THE DATA SCIENCE MICRODEGREE

Ogunlaiye Mosopefoluwa

INTRODUCTION

The first case of the CoronaVirus, a debilitating disease that caused a worldwide pandemic, was first discovered in China in late December of 2019 and spread around the world like a raging wildfire. By the 27th of February 2020, Nigeria recorded its first case of the disease.

Nigeria, a bustling country with diverse economic activities, was forced to declare a total lockdown of its 37 States after shutting down its major airports and borders. This severely affected its already struggling economy. Now that the pandemic is subsiding, there are efforts to revive the economy and gain back a sense of normalcy in her activities.

This project attempts to analyze the impacts of the pandemic on the country by creating insights from data science techniques, analytical skills and visualizations

The analysis shows the following

- The part of the country most affected by the virus
- The states with the most confirmed cases and casualties
- The relationship between confirmed cases and population in each of the States
- The daily infection, recovery and death cases in each state
- The impact on the Nigerian GDP
- The consequences of the change in the national budget that happened as a result of the pandemic.
- The relationship between the vulnerability index and the number of confirmed cases

DATA SOURCES AND METHODOLOGY

I would like to state here that this project was a little bit chaotic because the website of the Nigerian Centre for Disease Control, which is the official/sole source of up to date data regarding covid-19 in Nigeria has had issues for a number of months apparently. The data from that site couldnt be used due to this so I had to use the provided but not up to date data given in the project instructions. The faulty NCDC site also feeds the John Hopkins University site which gives data on the daily statistics so that data was also affected. So I had to use data up till 4th of August 2021 for that data set.

As hinted above the data for this project is gotten from different sources and combined to do the analysis from which we infer the insights. Here are some sources of the data:

1. The Nigeria Center for Diseases Control (NCDC) monitors the country's COVID-19 situation, and releases data on metrics across all the 37 states in the country. Data from NCDC COVID-29 official website (www.covid19.ncdc.gov.ng) was obtained by performing a web

extraction/ web scraping. This was done by querying the website API and passing the data into a pandas DataFrame using pd.read_html(). However, like I earlier stated, this source is unavailable for now because the site is down

2. 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 etcThese factors are known as themes. Each theme was broken into subthemes, and data was gathered from them to compute the overall vulnerability index score by weighing each theme equally. The Overall Vulnerability Index (CCVI Index) refers to the measure of the impact of the virus on a community after the virus arrives. It ranks from Very Low (0) to Very High (+1). Resource link – here

3. The Johns Hopkins University Data: 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 was accessed from the data repository, analyzed and insight was created from the data. Resource link- here

4. Real Domestic Gross Product Data: The data on Domestic Gross Product (GDP) for Nigeria was provided with the project details. This data was used to determine the impact of COVID-19 on the economy by comparing the Real GDP (Pre COVID-19) with Real GDP (During COVID-19). The analysis was done by plotting the GDP of each year (by Quarter) to show the difference and determine the impact of the negative virus on the economy Resource Link - here

5. State Budget Data:
   States across Nigeria reduced their initial budget due to the impact of COVID-19 on the economy. The budget data was provided with the project details. This data was analyzed to determine the negative impact of the virus on the economy. This analysis was done by calculating the difference between the initial and new budget of each state and calculating the percentage difference. The map of the percentage difference is then plotted to show the change in budget and comparison between the states.

RESULTS AND ANALYSIS

1. **Figures A - C** shows the graphs of the top 10 States with laboratory tested confirmed Covid Cases, recovering cases and the death cases
   - **Fig. A** shows Lagos State has the highest number of laboratory test confirmed Covid-19 cases with 26,708 infected. The FCT is the second most infected with 9,627 laboratory confirmed cases which is a significant gap from Lagos' statistics. This could be as a result of the differing population size between Lagos and the FCT/ Figure A:

Top 10 states in terms of Confirmed Covid cases by Laboratory test
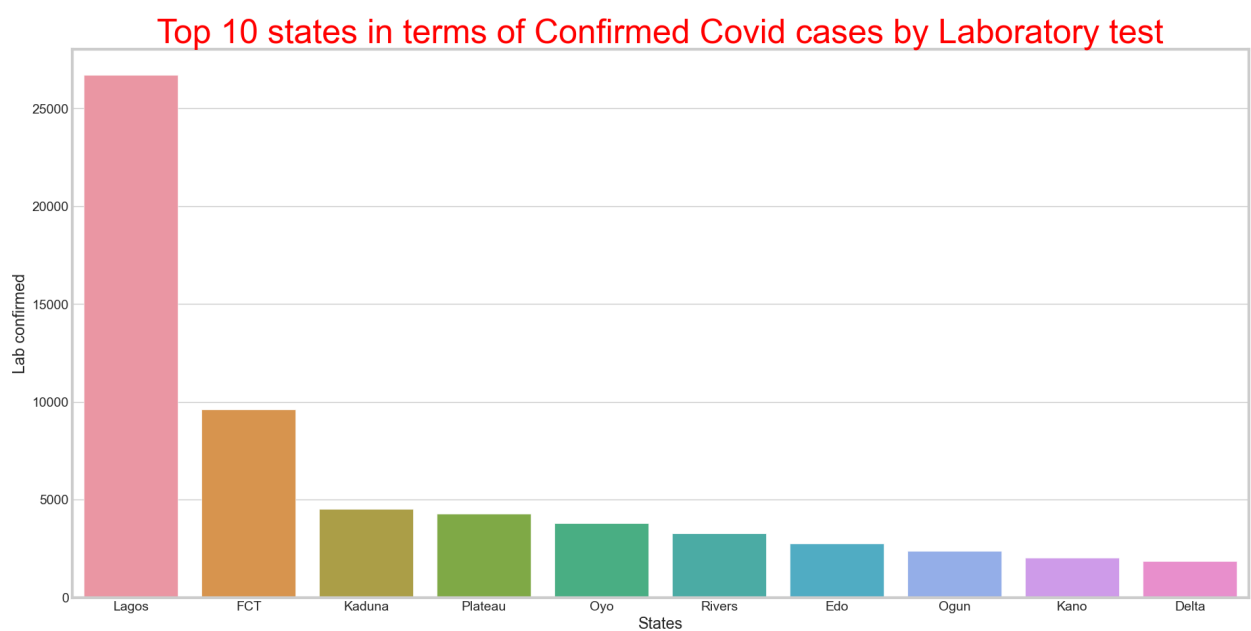
- **Fig. B** shows the top ten states with the highest discharge/recovery rates. Lagos has the highest statistics with 24,037 discharged and the FCT follows with 6,694 discharged. We can observe that the number of confirmed cases is slightly more than the number of recovery cases which indicates that even though the virus was still spreading as of the time this data was generated, a lot of people are recovering from the virus . Figure B:
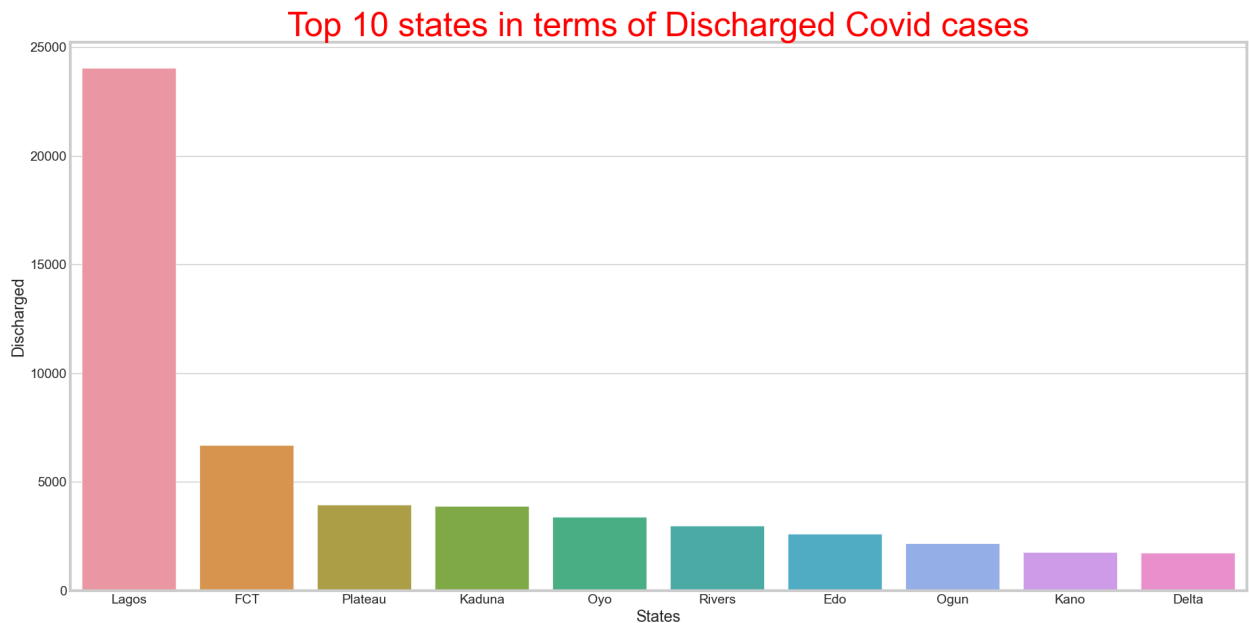
● **Fig. C** shows the number of death cases/fatalities. The number of deaths in Lagos is the highest with 236 recorded cases. Edo state follows with 113 deaths. One would expect that the FCT would record the second highest number of deaths since it has the second highest number of infections but this is not so. This could be as a result of a number of things like poor compliance to covid regulations or a lack of PPE in the hospitals in Edo state and a very high observance level in the FCT. However, generally speaking compared to the rate of infections and recovery, the death rate is indicative of general compliance to covid-19 protocols. Figure C:


Top 10 Death Cases

4
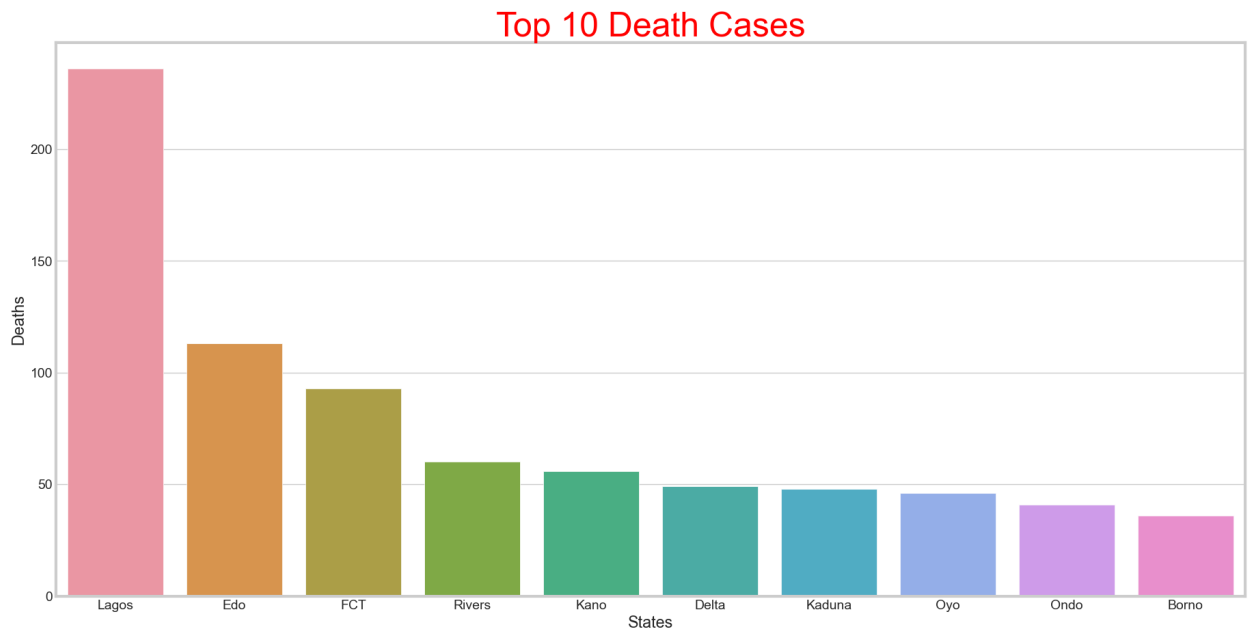
2. Fig. D shows the daily infection rate of the virus from 28th February, 2020 till date. The chart shows two waves of the virus, first being early 2020 and the second bigger wave towards the end of the year. Figure D:

.

Daily infection rate
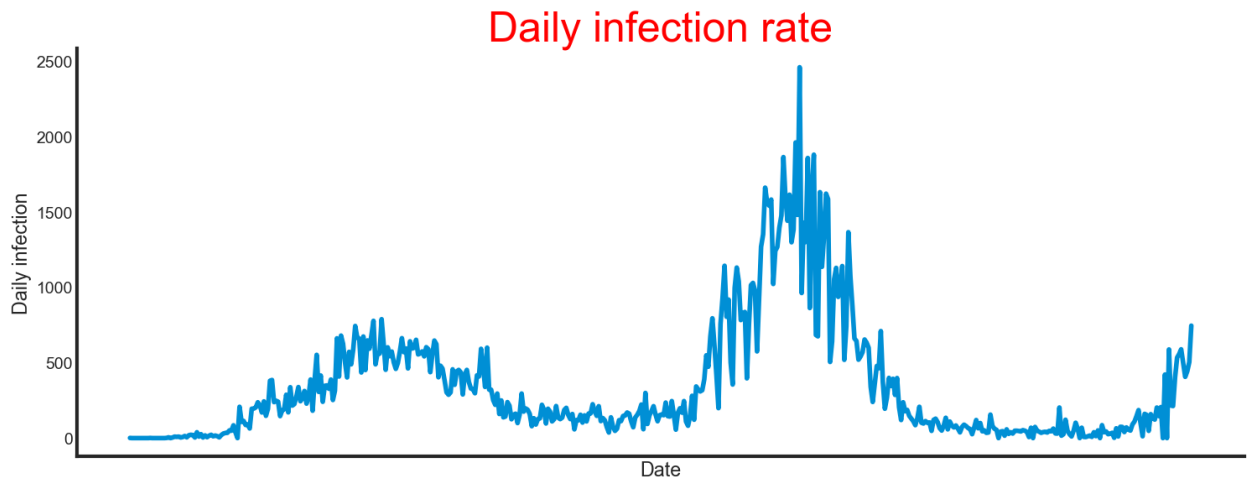
3. Fig. E Shows a combination of COVID-19 cases of Confirmed, recovered and death cases. The major spike in the recovery cases shows an effective approach towards the treatment of the virus. The death rate has no significant exponential growth. This indicates an effective approach to the treatment/preventive measures of transmitting the virus. Figure E:

Daily covid cases
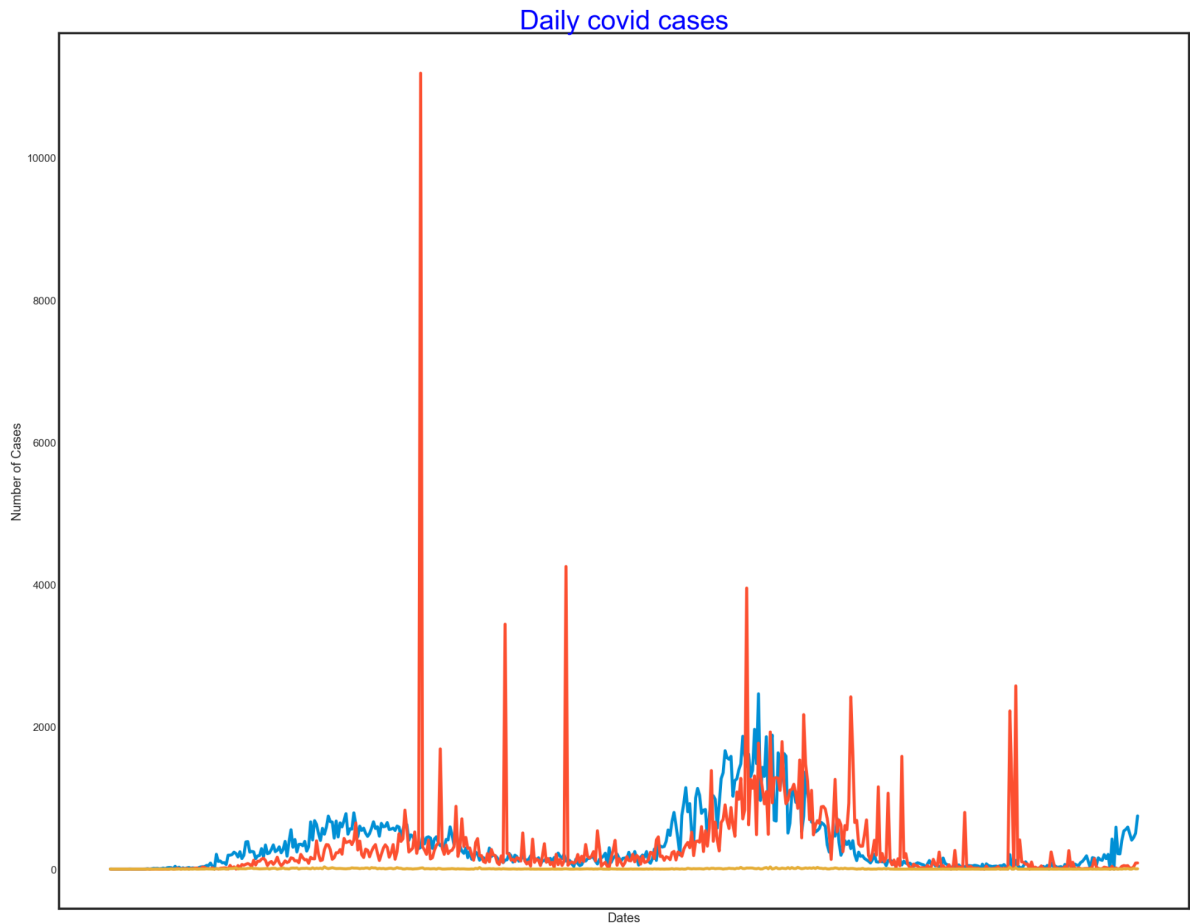
4. Fig. F Shows the relationship between the overall CCVI Index (Vulnerability Index) and the number of confirmed cases of COVID-19 in the top 10 infected states in Nigeria. The vulnerability index is a measure of how susceptible a state is to the virus (see Data Overview). The maps show no direct relationship between CCVI Index and the number of confirmed cases in each state. Lagos States, which has the highest number of confirmed cases (58,393) of the virus has a CCVI Index of 0.0. Kaduna and Kano state with a CCVI Index of 0.7 and 0.6 respectively, which should be more susceptible to the virus and hence, have more cases, both have less than 10,000 cases each.

Overall CCVI Index & Lab Confirmed Cases
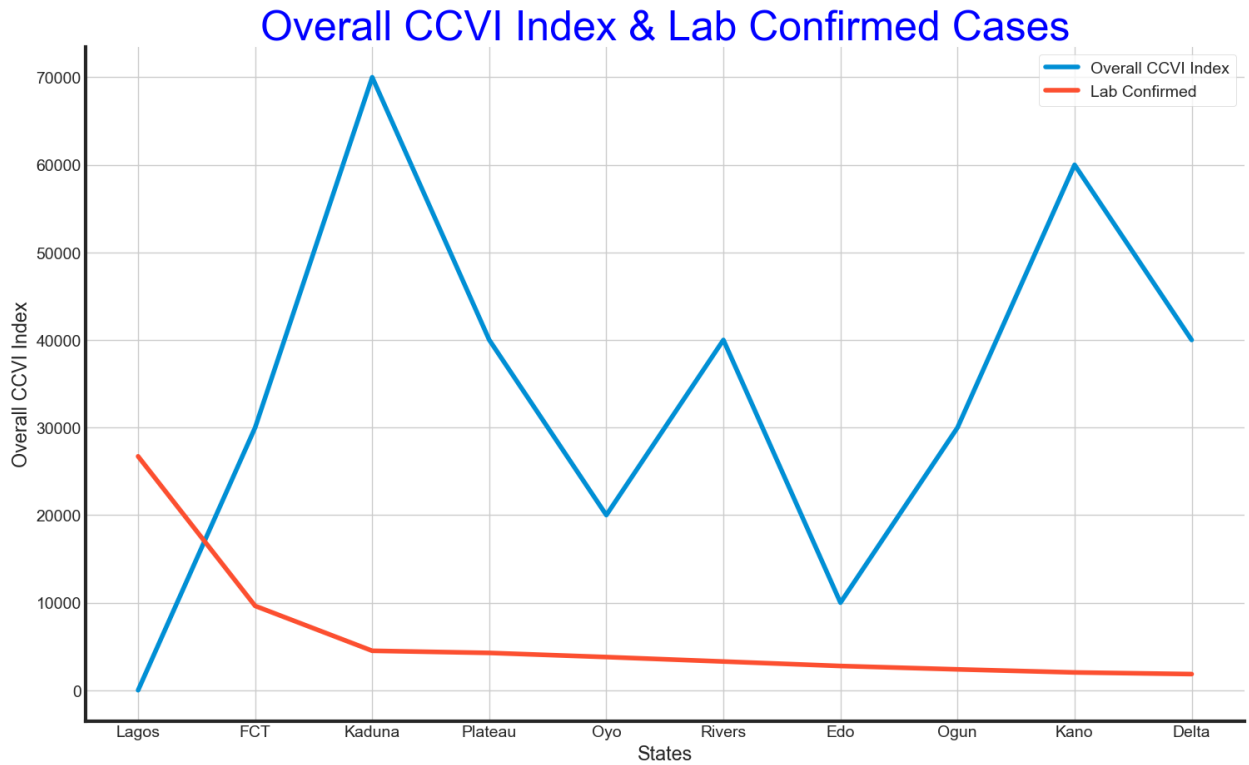
5. Fig. G Shows the regression plot between the number of confirmed cases and population density of each state.

Regression plot between confirmed cases and population density
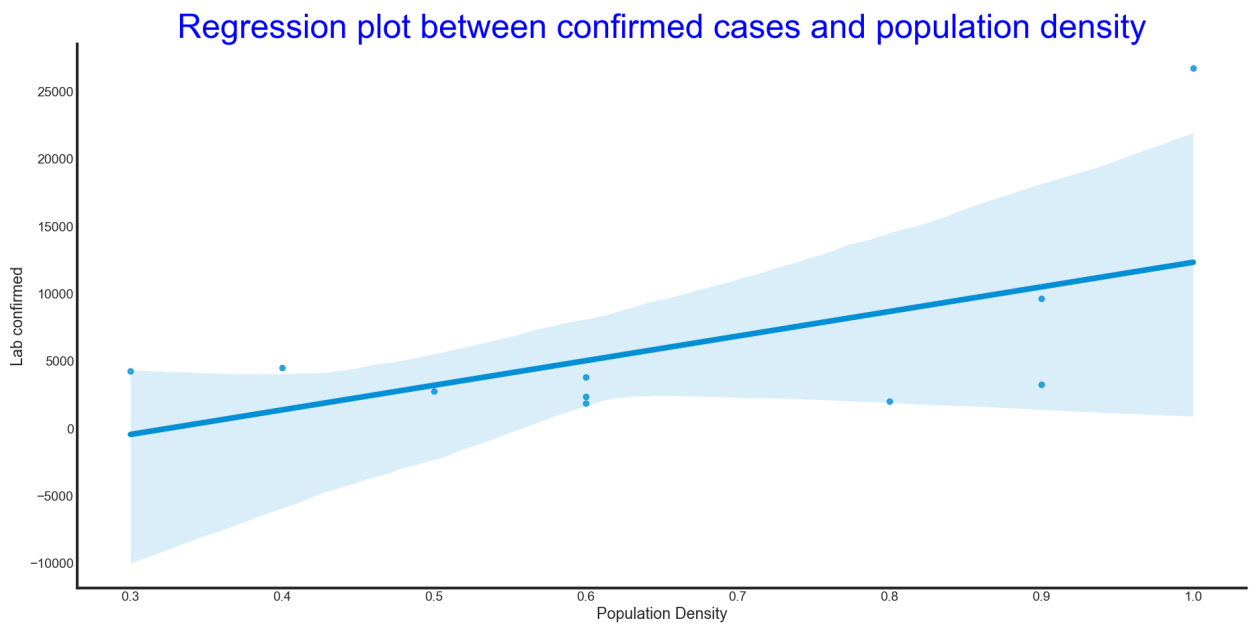
6. Fig. H is a map of the confirmed, recovered and death cases of COVID-19. The map shows a continued growth in the confirmed and recovered cases of the virus. This also indicated an effected response to the treatment and preventive methods of spreading the virus. The death rate also indicates an effective approach to curtailing the virus.

Map showing confirmed, recovered and death cases
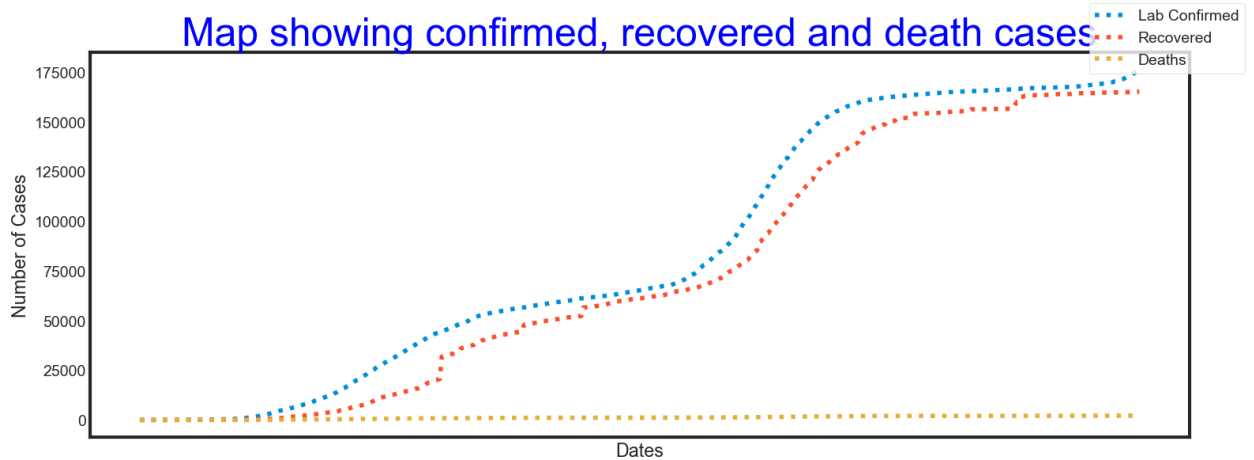
7. Fig. I shows the Percentage change in budget between the pre-covid period and during the covid period. This map shows Cross River state to have the highest change in budget of 86.6%, a state which has only 394 cases of COVID-19 (34th highest state of confirmed cases). Lagos State with the highest confirmed cases of the virus (58,393 cases) have 45.2% budget change and FCT, the state with the second highest confirmed cases, has 28.6% budget change.

Percentage Change in Budget
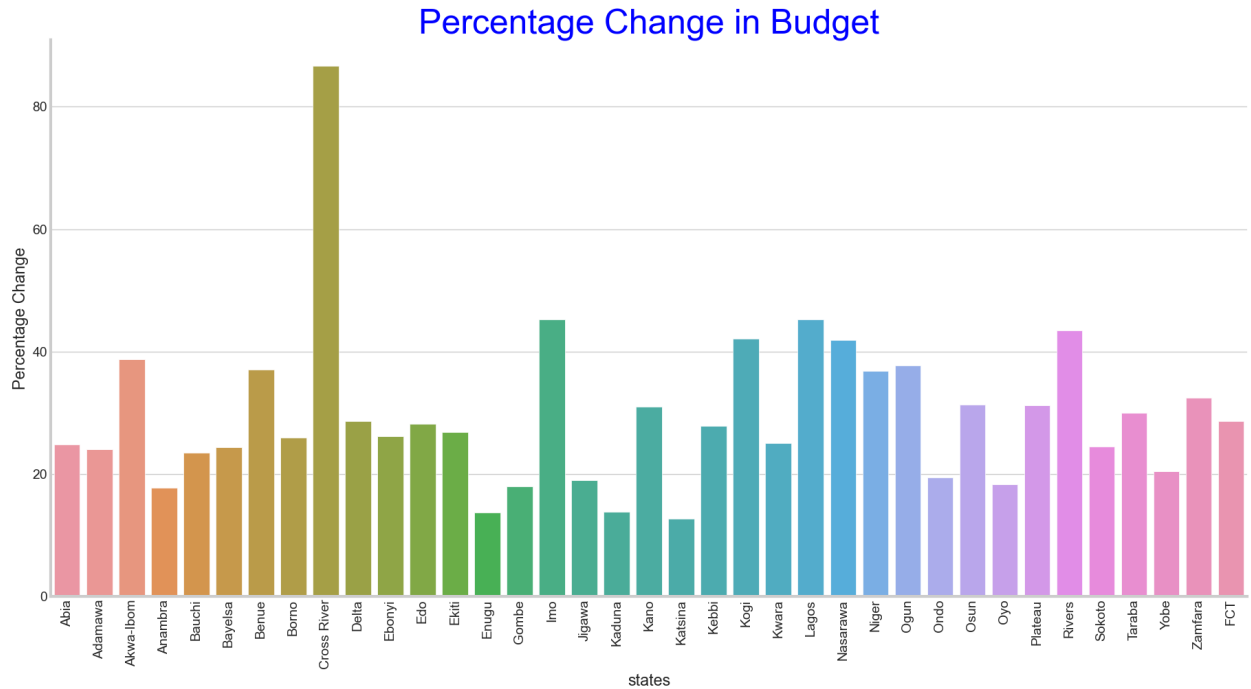
8. Fig. J Nigeria is divided into six geo-political regions. This shows the number of confirmed cases of COVID-19 by Region in Nigeria. South-West Region has the highest number of confirmed cases which includes states like; Lagos, Oyo, Ondo, Ogun, etc. followed by North Central which includes states like FCT, Kogi, Plateau, etc.

Confirmed Cases by Region
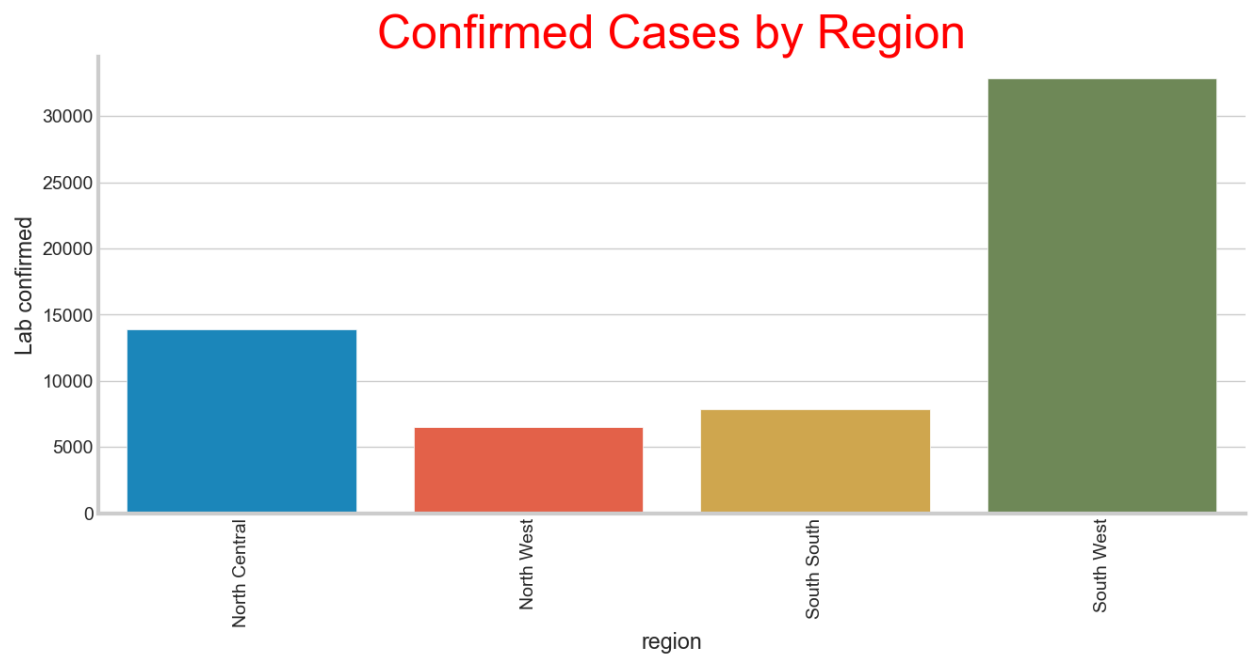
9. Fig. K Shows the relationship between the population of states and the number of cases in the states. The map shows a direct relationship between the population of regions in the country and the number of confirmed cases except the North-West region which has a very low number of confirmed cases compared to the population.

Relationship Between Population and Confirmed Cases
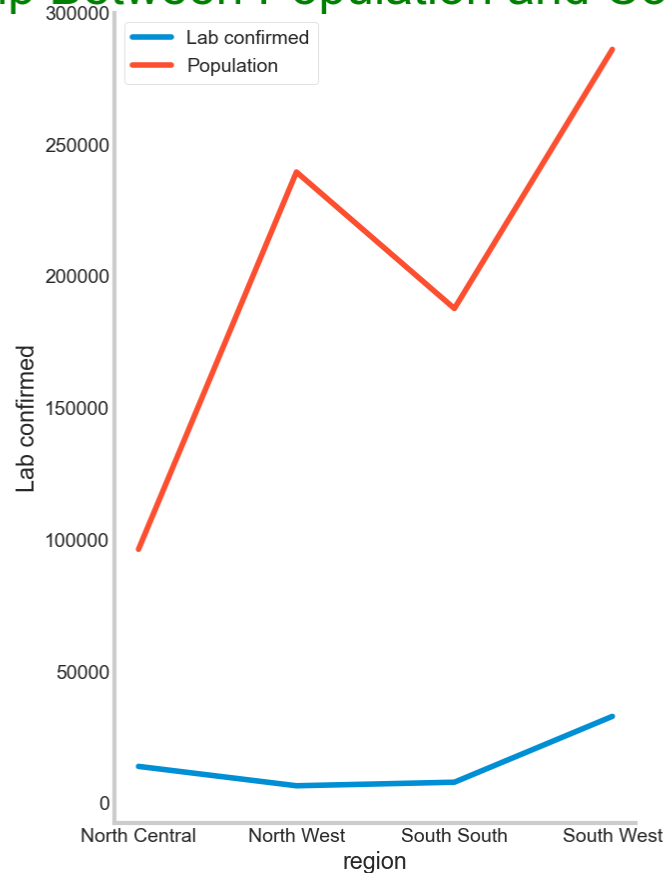
10. Fig. 12 Shows the Gross Domestic Product of the country between 2014 – 2020.   This shows the second and third quarter of 2020 (The covid period) has the lowest GDP since 2014.

Real GDP
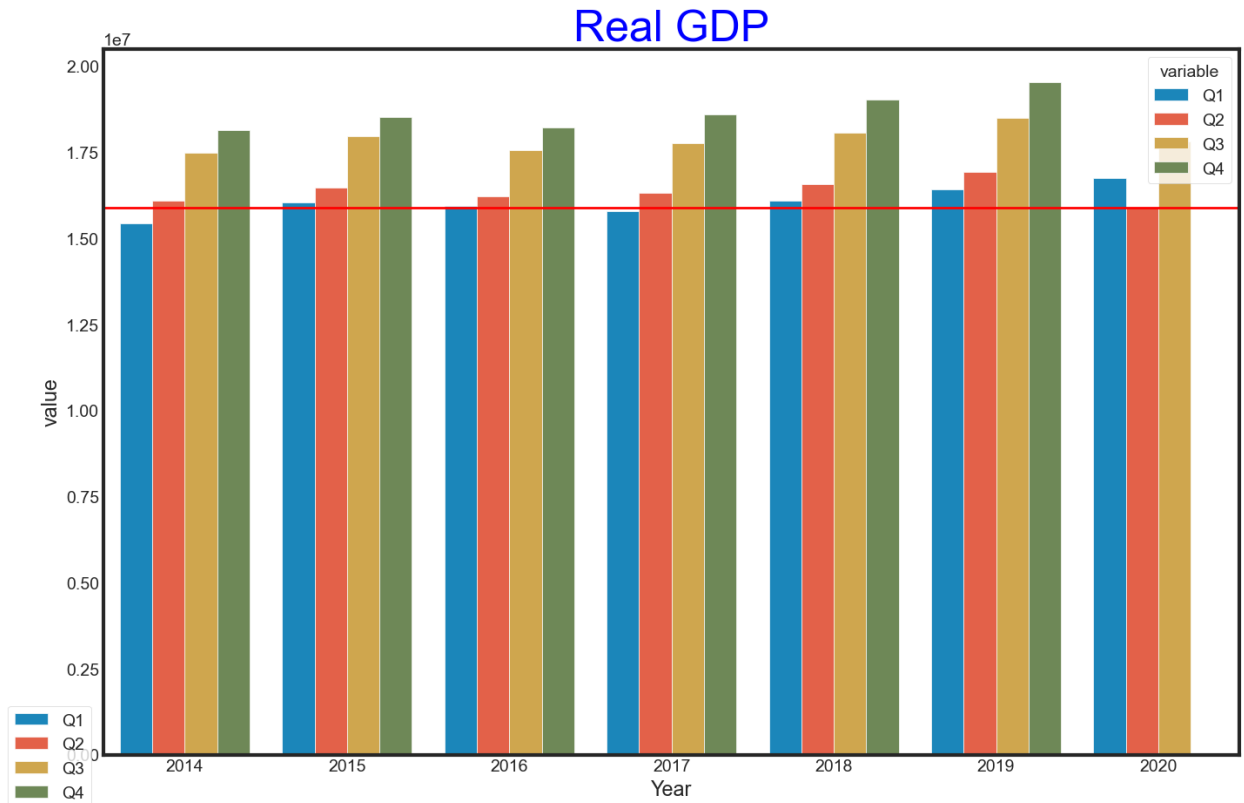
CONCLUSIONS

The purpose of the project is to analyze COVID-19 data and provide insight on the data. Analysis of the data used for this project shows different results which has led to conclusions that:

1. Lagos state and FCT are the top two states with the confirmed, recovered and death cases of COVID-19 virus in the country,

2. The South-West region of Nigeria has the highest number of confirmed cases in the Country,

3. The daily cases of COVID-19 indicates two waves of similar pattern of the virus and the rate of recovered and death cases shows an effective approach towards curtailing and treatment of the virus,

4. There is no direct relationship between the CCVI Index and the number of lab confirmed

11

cases, this can be as a result of not fully understanding and applying factors which makes the spread in the calculation of the CCVI Index or a lack of proper reporting of confirmed cases in some states/region of the country,

5. The budget change of the states in the country has no direct relationship with the number of cases from the state,

6. Either there is incorrect/insufficient report of confirmed cases in the North-West region of the country or the North-West region of the state took some preventive measure in the spread of the virus to the region which is really effective and should be studied,

7. Finally, the covid pandemic has negatively affected the country economy which is evident in the GDP of the country in 2020.

FUTURE WORK SUGGESTION

I would strongly suggest further study of the daily confirmed cases in the country and each region, in comparison with global cases to study the pattern of waves of the infection to prevent another wave and put preventive methods in place.

Also, the mode and method of recording confirmed cases in the country should be looked into to ensure accuracy especially in the Northern region of the country.