**Nigeria COVID-19 Data Analysis using Python**

By

**RAIMI AZEEZ BABATUNDE**

Ustacky Capstone Project

### INTRODUCTION

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.  
  
This project is aimed at employing data science & analytics skills to collect data, explore the data, perform analysis, create visualizations, and generate insights on the effect of the virus on the country.

This analysis shows:

* The South-West geo-political region of the country is most affected by the virus,
* The negative effect of the virus on the Gross Domestic Product (GDP) and budget change of the country,
* The relationship between the Vulnerability index (CCVI Index) and the number of confirmed cases,
* The top states with the most confirmed cases of the virus and casualty,
* The relationship between the confirmed cases in each state and the population,
* The daily infection, recovery and casualty rate of cases in the country.

### DATA OVERVIEW & METHOD

The data for this analysis was sourced from different location and combined to perform analysis and provide insights. The data sources include:

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)](http://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().
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 etc., these 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 equally each theme.

* 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 – https://covid-static-assets.s3.amazonaws.com/Africa+CCVI+methodology.pdf

1. 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 – www. https://github.com/CSSEGISandData/COVID-19/tree/master/csse\_covid\_19\_data/csse\_covid\_19\_time\_series

The data for Nigeria was extracted from the global data using the “for loop” and passed into a new data frame to show the daily confirmed, recovered and death cases of COVID-19 in Nigeria.

1. 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 deference and determine the impact of the negative virus on the economy

Resource Link - https://www.pwc.com/ng/en/assets/pdf/economic-alert-october-2020.pdf

1. 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.

### ANALYSIS & RESULTS

### Fig 1-3 - Maps of Top 10 States Confirmed, Discharged/Recovered cases COVID-19 cases in Nigeria.

### Fig. 1 shows Lagos State has the highest cases of COVID-19 confirmed, discharged and death cases, recording 59,047 confirmed cases, 56,990 discharged and 439 death cases as at 1st June, 2021. Other states have a significantly lower confirmed case with FCT being the second highest recording up to 19,868 confirmed cases, 19,110 discharged and 166 death cases, and Ondo been the 10th highest recording up to 3,311 confirmed cases, 3,219 discharged and 64 death cases of COVID-19.

### ../covidproject/output_43_0.png

### ../covidproject/output_47_0.png

### ../covidproject/output_51_0.png

### Fig 4 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. This shows a high spike in infection rate between December, 2020 and early March, 2021 indicating the massive second wave of the virus.

### ../covidproject%20v2/output_82_0.png

### Fig. 5 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.

### ../covidproject/output_64_0.png

### Fig. 6 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.

### ../covidproject/output_86_0.png

### Fig. 7 Shows the regression plot between the number of confirmed cases and population density of each states.

### ../covidproject/output_90_0.png

### Fig. 8i, 8ii, 8iii shows the daily confirmed, recovered and death cases of COVID-19 respectively, clearing showing the increasing rate if the virus from 28th February, 2020 till date. Fig 8iv shows the combined graph of the 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.

### ../covidproject%20v2/output_66_0.png

### ../covidproject%20v2/output_68_0.png

### ../covidproject%20v2/output_70_0.png

### ../covidproject%20v2/output_71_1.png

### Fig. 9 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 have 28.6% budget change.

### ../covidproject/output_104_0.png

### Fig. 10 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.

### ../covidproject/output_116_0.png

### Fig. 11 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.

### ../covidproject/output_118_0.png

### 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.

### ../covidproject/output_124_0.png

### 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:

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

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

### 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,

### There is no direct relationship between the CCVI Index and the number of lab confirmed 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,

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

### 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,

### 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.