**NIGERIA COVID-19 DATA ANALYSIS**

**EXECUTIVE SUMMARY**

The data used for this project includes:

* Data from NCDC covid-19 official website, I was able to get this data by scraping it from NCDC’s website using python’s beautiful soup, features in this data includes: Number of cases confirmed in the laboratory and on admission, Number of discharged patients and Number of patients who died.
* Data from The Johns Hopkins University Center for Systems Science and Engineering repository, they publish daily data on confirmed, death and recovered cases across different countries in their repository.
* The Nigeria Community Vulnerability Index data, which 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.
* Real Domestic Gross Product Data of Nigeria, and the State Budget Data.

Coronavirus disease (COVID-19) is an infectious disease caused by a newly discovered virus, 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.

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

**PROJECT OVERVIEW**

**DATA OVERVIEW**

After performing exploratory analysis on the data, I was able to derive the following insights:

* The top ten states in terms of confirmed covid-19 cases by Laboratory test are: Lagos, FCT, Kaduna, Plateau, Rivers, Oyo, Edo, Ogun, Kano, Ondo.
* The top ten states with the highest number of discharged covid-19 patients are: Lagos, FCT, Kaduna, Plateau, Rivers, Oyo, Edo, Ogun, Kano, Ondo.
* The top ten states with the highest number of deaths are: Lagos, Edo, FCT, Oyo, Kano, Rivers, Delta, Kaduna, Ondo, Plateau.
* The highest number of confirmed cases recorded in Nigeria was on the 23rd of January, 2021. On this day, Nigeria recorded 2,464 covid-19 cases.
* The highest number of recovered cases recorded in Nigeria was on the 4th of October, 2020. On this day, 11,188 patients recovered from covid-19.
* Nigeria recorded the highest number of deaths on the 16th of June, 2020. 31 covid-19 patients lost their lives.
* After merging NCDC data with Nigeria Community Vulnerability Index data, I was able to detect a negative relationship between number of cases confirmed in Lab and overall CCVI Index that is the higher the number of confirmed cases the lower the overall community vulnerability index in the states except in FCT.
* There is a direct relationship between population density and number of cases confirmed in the lab, the higher the population density the higher the number of cases confirmed in the laboratory.
* There is also a negative correlation between Age Index Score and number of cases confirmed in Lab with Age Index Score and Number of Deaths that is when we have a low Age index score there are more confirmed cases and more number of deaths and vice versa.
* Enugu, Abia, Imo are the first 3 states with the highest number of people with Acute IHR that is people that got infected and are expected to require acute care, these 3 states are also seen to have a low health index score, but the number of deaths in these three states are low, but Lagos with a very high number of deaths has a low health index score and a high percentage of people who got infected and are expected to require acute care. This implies that despite the good health system in Lagos, more people still die.
* People with low age index score, are at lower risk of needing acute care when infected but people with high age index score are at higher risk of needing acute care when infected.
* The GDP of quarter 1 and quarter 2 of each year are always lesser than the GDP of quarter 3 and 4. 2020 had the lowest GDP for quarter 2 compared to quarter 2 of previous years due to the impact of covid 19 on the economy.

**ANALYSIS AND RESULT**