EXECUTIVE SUMMARY ON NIGERIA'S COVID-19 DATA ANALYSIS

PURPOSE OF REPORT

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.

In this project, you will employ data science & analytics skills to collect data, explore the data, perform analysis, create visualizations, and generate insights.

The purposes of this report are:

- Examination of covid-19 impact on the country's indexes such as CCVI,
- Trend analysis of all cases (confirmed, discharged, and death)
- Regional index reaction to covid 19
- Quarterly GDP analysis from 2014-2020
- Impact of covid-19 on the budget
- Impact on Oil prices

METHODOLOGY

Data Overview

The data used in the covid-19 analysis consist of a data frame extracted from three different sources. They include the following:

1). NCDC DATASET

The data was extracted using web scraping from https://covid19.ncdc.gov.ng/ consisting of the number of confirmed cases, number of discharged cases, and death cases for all 36 states.

2). JOHN HOPKINS DATASET

This dataset is obtained from scraping the github repository that encompasses all cases(confirmed, discharged, and death) of several affected countries around the globe from the John Hopkins repository. But the cause of this study is limited to just Nigeria's, hence it is extracted.

3). COVID EXTERNAL DATASET

This data is further subdivided into 3 categories; The quarterly GDP (Gross Domestic Data) data, Covid-19 Indexes Impact data and budget data, and additional case statistics data.

4). OIL PRICE DATA

Just for further analysis oil price data was downloaded and extracted to see the overall trend pre-covid, covid, and post covid.

METHODS

Data Extraction

- The NCDC data was extracted from the website using web scraping and the function pd.read_html() was used to extract the data frame from the web page
- The John Hopkins dataset was extracted from the John Hopkins repository https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data which the raw data of the global confirmed, recovered and death cases were scrapped using pd.read_html().
  BeautifulSoup approach failed to capture all of the raw data for the confirmed and recovered cases leaving out the last date data entry, hence the imposition of the

pd.read_html() functions coupled with a lot of python editing to extract the data effectively.

● Covid external data was downloaded from their respective repositories and pd.read_csv was used to extract it from the device's storage.

● The U.S. Landed Costs of Nigeria Crude Oil (Dollars per Barrel) was downloaded from https://www.cbn.gov.ng/rates/DailyCrude.asp site and extracted from the device storage using pd.read_csv().

Analysis

After the data extraction the next step was to get the standard summary from the data's using the following attribute:

❖ .head()

❖ .describe()

❖ .info()

Now lets look at the analysis performed on each of these datasets

1). NCDC DATASET

● Basic insights derivation operations such as .info(), .describe() and viewing the complete data

● Plotted a bar charts of all the states affected to number of confirmed cases, no of discharged cases, and number of deaths

● Performed correlation analysis using scatterplots

● Observed the distribution of confirmed cases

● Also, another bivariate analysis such as determining the degree of relationship between the all three cases instances using relplot was carried out and a regression line was fitted on them

2). JOHN HOPKINS DATASET

● There are three subdivisions of the data; the number of confirmed cases discharged cases, and death cases across various countries.

● The data frames were extracted

● Basic insights derivation form .info(), .describe() and data.head()

- Introduced the pd.melt() function to convert all the confirmed, discharged and death date features into one date feature mapping to the case value
- Merged the three frames new dataframes created from using the pd.melt function
- Converted the date object to datetime using pd.to_datetime()
- Created and added Active cases feature by subtracting the deaths and discharged features from the confirmed features
- Extracted Nigeria's data out of all the various countries data
- Plotted a line graph of date to all three cases (i.e, confirmed,discharged and deaths) independently.
- Created and added infection rate feature using the diff() function on the confirmed cases
- Plotted date to infection rate for trend identification and to identify area of spikes and drops
- Used pandas masking operation to get date with maximum infection rate

3). EXTERNAL DATA

The external data helps us perform analysis to the end that insights as regarding the impact of covid-19 on various aspects of the country such as the GDP, Indexes, Budget data.

Index Analysis
- Basic insights derivation form .info(), .describe() were performed
- The index data was grouped by region to estimate the average index across different geopolitical zones in Nigeria. The **groupby()** function was used to perform this operation
- In other to arrive have larger data block to extract insight pd.merge() was used to combine the **NCDC** dataframe and the **Index** dataframe
- Insights on different indexes and how they affected each zone were determined
- Correlation analysis using scatterplot was performed
- Used nlargest function to get top 10 confirmed, discharged and deaths cases
- Plotted cases of top 10 CCVI index and population density
- Scatter plot of deaths and epidemiology index and fit a regression line on it

GDP Analysis
- The average GDP for each quarter was determined
- Converted the date column into **object** because if it were converted to datetime format the entries would be placed at seconds rather than years
- Used the melt function to get the gdp for all the years from 2014-2020 and splitted them into quarters mapped to their respective GDP's

● Plotted a bar chart to show GDP for all the years into the quarter subplots

Budget Data Analysis

● Created a new entry feature known as the **Percentage Budget Decrease** by subtracting the revised budget from the initial budget and divided by the initial budget
● Used **sns.jointplot()** to visualize the relationship between the Initial and Revised Budgets
● Used seaborn barplot to plot a bar chart for the states and their initial and revised budget
● Computed **average percentage decrease(APD)** in the budget across all states in Nigeria
● Extracted state with the minimum percent change
● Extracted state with the maximum percent change

OIl Price Data Analysis

● Extracted date with the maximum landed cost of crude oil in Nigeria from 2020 - 2021
● Extracted date with the minimum landed cost of crude oil in Nigeria from 2020 - 2021

FINDINGS/INSIGHTS

INDEX FINDINGS

❏ South West
● Some of the major factors that contributed to the increased cases in the south west are; epidemiology index which indicates disease vulnerability was high, and high population density
● Other factors that influenced the high recovery rate were the transportation index, socioeconomic index and good health care system/response compared to other states
❏ South East
● The major causes of the increased cases and death are: High occupancy by aged people, high population density, and high degree of health vulnerability due to exposure to other infections prior to covid's inception.
● The functionary determinants of the recovered cases are: good transportation system, and healthcare system.
❏ South South
● Confirmed cases arise from densely populated by aged people and high epidemiology index.
● It was curbed due to reduced socio-economic index impact and good transportation.
❏ North Central
● Major factors of the deaths and confirmed cases are: poor health facility, poor transportation system and high fragility.

- The recovered cases arise from the fact that the population density is small and moderate epidemiology effect.
  - ❏ North East
- The North East is operating negatively in all factors. Very fragile, poorest health care system and high socio-economic disadvantage.
  - ❏ North West
- Similar results with the Northeast just the slight difference in the fragility status as the North West's isn't as high as the North East.

- ❖ **Correlation Analysis**
- Areas with a high population density index(increased vulnerability as a result of the huge population) are reported to have a high number of confirmed cases
- Areas that are largely dominated by the aged who are susceptible to the virus have records for the high number of confirmed cases while areas with a low age index have a low number of confirmed cases

CASE STATISTICS
- 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, Kan 3967 cases,and Ondo 3248 cases.
- 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
- 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
- The maximum infection rate is 2464 we look further to discover when this occurred
- The date with the maximum infection rate was on the 23rd of January 2021 with an infection rate of 2464.
- Areas with low CCVI have a relatively high number of confirmed cases

GDP AND BUDGET INSIGHTS
- The average cumulative GDP is highest at the third quarter
- The year with the highest total GDP was the year 2019 and the lowest being 2020 most likely because of the pandemic (COVID-19)
- The analysis from the initial and revised budget data shows that the average percentage budget drop in Nigeria across all states is approximately 30%
- Katsina had the lowest percentage change of about 13% from 344Bn to 213Bn
- Cross River had the highest percentage change of about 87% from 1.1trn to 147bn

- Lagos has the highest budget prior to the pandemic with 1.7trn and after the revised budget which was approximately a 45% decrease from the year before was 920bn

OIL STATISTICS

- In recent data that is from 2020-date the date with the highest US landed cost of Nigeria crude oil was found to be on 15th Of February 2020 with an amount of 548$/bbl
- All the months from January 2020 to February 2021 excluding February 2020 and December 2020 have their values at the minimum for the range of 2020 to 2021 with 38.7$/bbl
- This trend plot shows that the US landed cost of Nigeria crude oil has been in a downtrend from the first quarter of 2020 and is still yet to brake the downtrend structure

RECOMMENDATIONS

The federal government should take note of the following of which if followed would prepare the country against the devastating impact on the lives of people and the economy as a whole

- The health system standards should be raised in order to effectively and efficiently attend to patients.
- Areas with high population density should have fast and reliable access to transportation
- The International Health Regulations(IHR) should be heavily placed on areas with aged people so they can get quick access to treatment outside of their borders
- The security conditions of the North East should be looked into to aid the permeability of health workers into the states and societies
- For further covid-19 studies impact should extend to other economic factors like education, oil market, stock market, agricultural sector and climatic evaluation.

REFERENCES

https://seaborn.pydata.org/

https://pandas.pydata.org/docs/reference/api/pandas.DataFrame

https://stackoverflow.com/

https://www.geeksforgeeks.org/

https://www.cbn.gov.ng/rates/DailyCrude.asp/

https://covid-static-assets.s3.amazonaws.com/US-CCVI/COVID-19+Community+Vulnerability+Index+(CCVI)+Methodology.pdf