Nigeria COVID-19 Data Analysis using Python

MUTIAT OLANREWAJU

SUPERVISED BY

MR. BLESSING ADESIJI

DATA SCIENCE MICRODEGREE
USTACKY

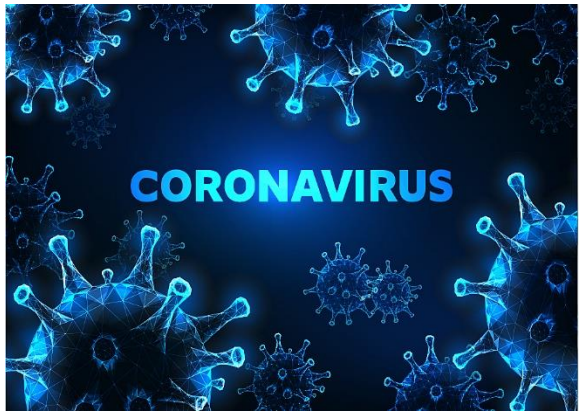
March, 2021

PRESENTATION OUTLINE

- Introduction
- Objectives
- Data Overview
- Methods
- Analysis
- Results
- Conclusion
- Suggestions on future work

INTRODUCTION

- Coronavirus disease (COVID-19) is an infectious disease caused by a newly discovered coronavirus, and it has affected major parts of the world.



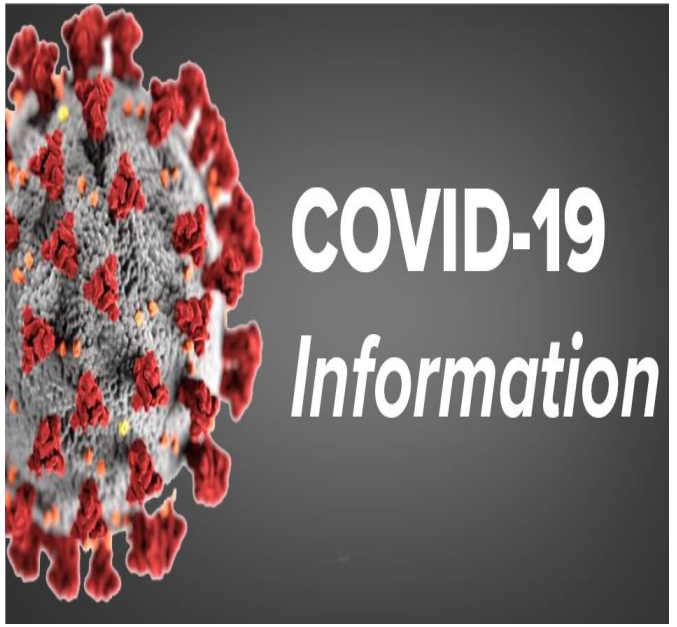
- Nigeria, a West-African country, has also been affected by the COVID-19 pandemic after recording its first case on 27th February 2020.

INTRODUCTION CONT.

- 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, python with data science & analytics skills were employed to collect data, explore the data, perform analysis, create visualizations, and generate insights.



OBJECTIVES

- Understand data collection process, in this case web scraping, and importing from data sources
- Understand the data cleaning and manipulation process.
- Develop data wrangling skills & data intuition.
- Know how to ask the right questions & find ways to provide answers.
- Develop visualization skills through the use of open-source libraries in Python.
- Generate insights from analysis.

DATA OVERVIEW

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

❑ DATA COLLECTION

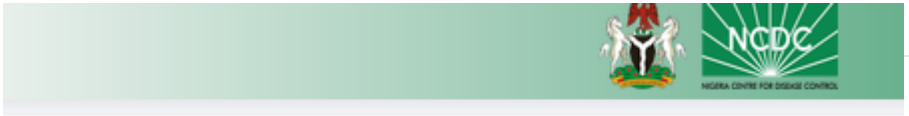
- Nigeria Centre for Diseases Control (NCDC) COVID-19 web scraping.
- Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE) data repository on daily data for Nigeria on confirmed, death and recovered cases.
- External Data:
  - Nigeria Community Vulnerability Index data
  - Real Domestic Gross Product Data
  - State Budget Data for Nigeria
  - Our World In Data (OWID) dataset

❑ DATA CLEANING AND PREPARATION

- Converts data to appropriate data type.
- Rename the columns of the scraped data.
- Remove comma(,) and white space in numerical data
- Extract daily data for Nigeria from the Global daily cases data

METHODS

- **NCDC WEB SCRAPPING**



Inspect NCDC website for data in HTML frontend

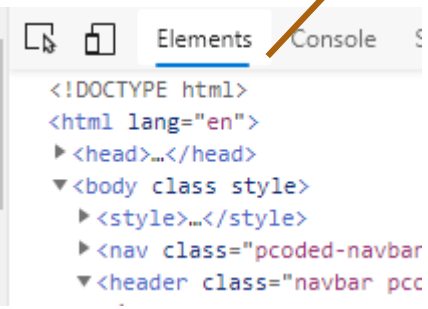
Import Necessary Libraries

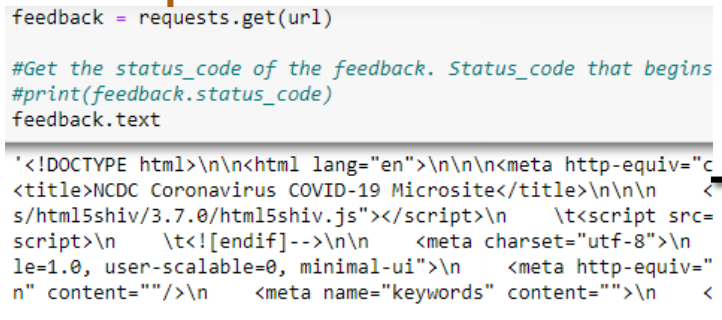
Using Request library in python to access the NCDC URL and download the HTML data from the website

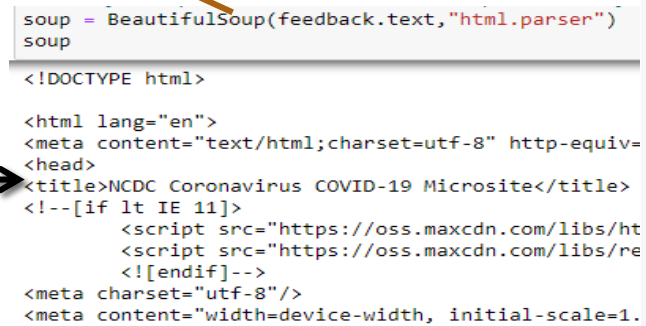
Using Beautiful Soup library to format the downloaded content into readable text

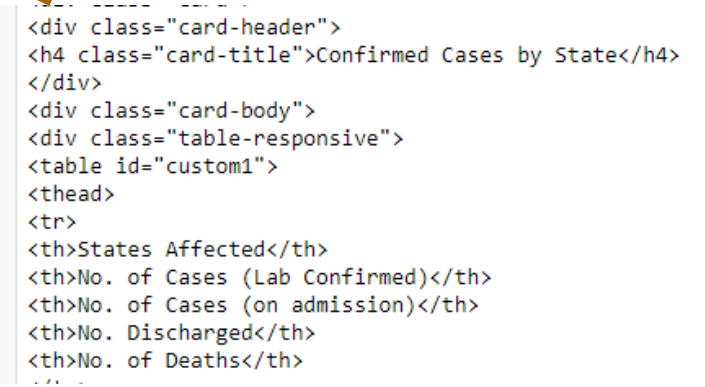
Extract the required table using the table id

Create a Csv file and open in append mode to store the extracted data



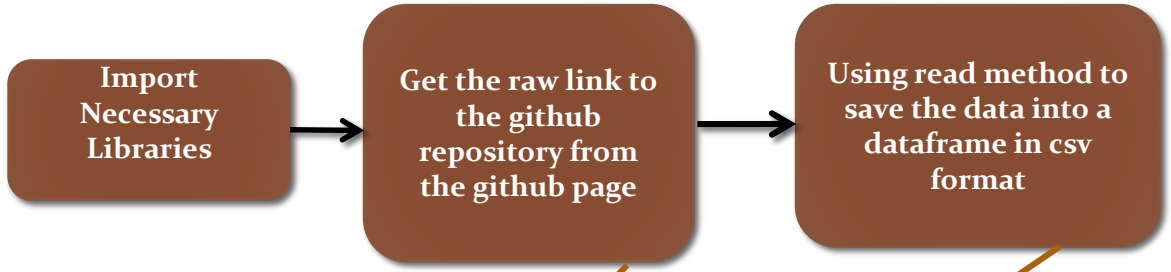






METHODS CONT.

John Hopkins Data Repository

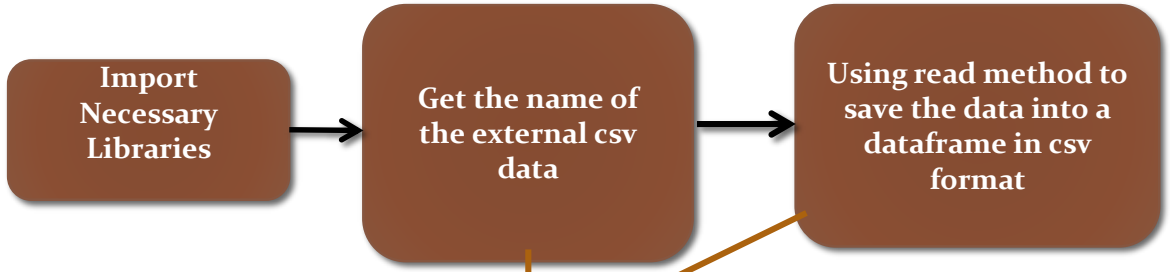



```
#Save the data from John Hopkins repository raw link
confirmed_global = 'https://raw.githubusercontent.com/
df_confirmed_global = pd.read_csv(confirmed_global)


recovered_global = 'https://raw.githubusercontent.com/
df_recovered_global = pd.read_csv(recovered_global)


deaths_global = 'https://raw.githubusercontent.com/CSS
df_deaths_global = pd.read_csv(deaths_global)
```

External Data




```
#Save the external data to a DataFrame using pd.read_csv() method
df_external = pd.read_csv('covid_external.csv')
```

```
#Save the budget data to a DataFrame using pd.read_csv() method
df_budget = pd.read_csv('Budget data.csv')
```

```
#Save the readgdp data to a DataFrame using pd.read_csv() method
df_readgdp = pd.read_csv('RealGDP.csv')
```

METHODS CONT.

❑ Approach to Data Analysis

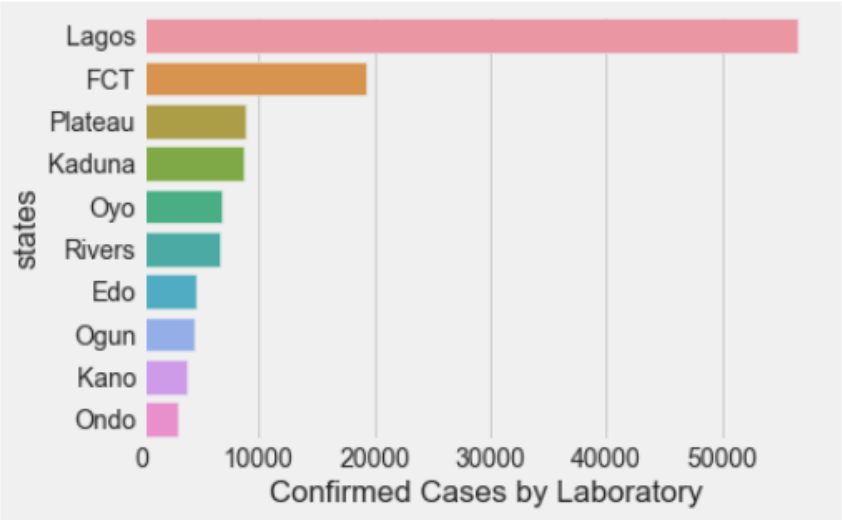
- **Data Collection:**
  - Appropriate method of web scraping was used to obtain DataFrame.
  - Dataset was imported from John Hopkins Github.
  - External Data was imported

- **Data Cleaning and Preparation**: Data cleaning was performed on the scrapped dataset in order to fix the data format. This includes
  - Converts data to appropriate data type.
  - Rename the columns of the scraped data.
  - Remove comma(,) and white space in numerical data
  - Extract daily data for Nigeria from the Global daily cases data.
  - Pandas dataframe was gotten using a groupedby method for Daily Confirmed Cases, Recovered Cases and Daily Death Cases in Nigeria using Date and Cases as columns.
  - Head, tail and info methods were used to display the overview of the dataset respectively.

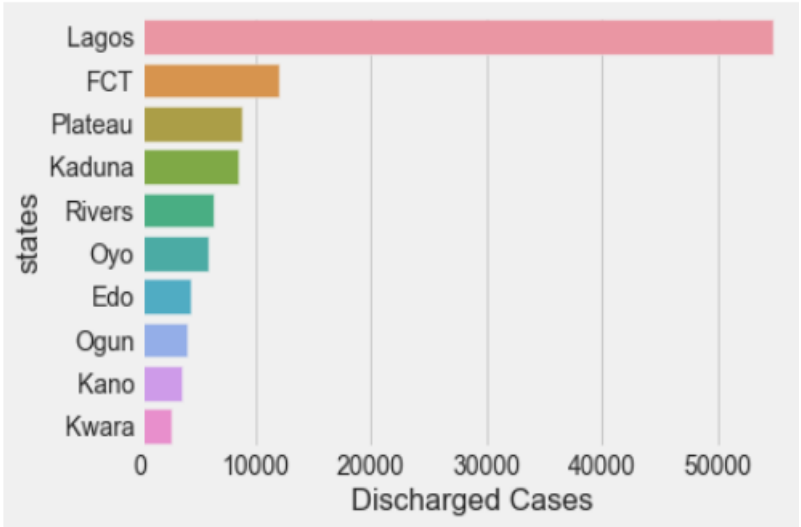
ANALYSIS

- Seaborn barplots that show the top 10 states in terms of confirmed, discharged and active Covid cases respectively by laboratory test were generated using the nlargest and sort method.

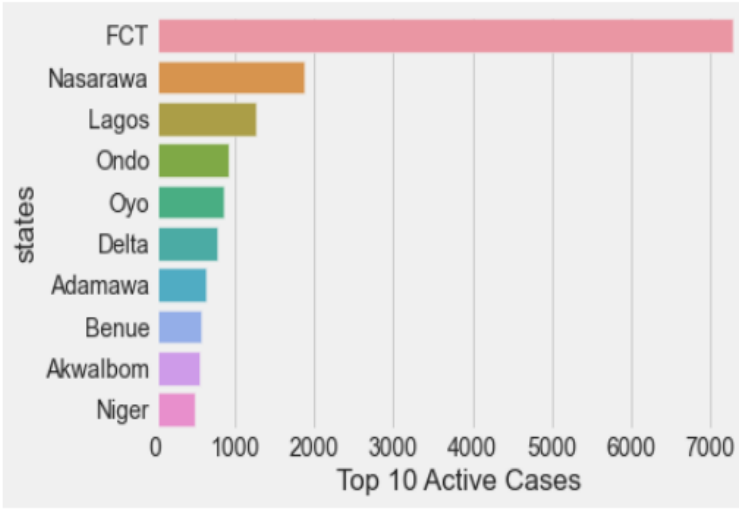
Text(0.5, 0, 'Confirmed Cases by Laboratory')



Text(0.5, 0, 'Discharged Cases')



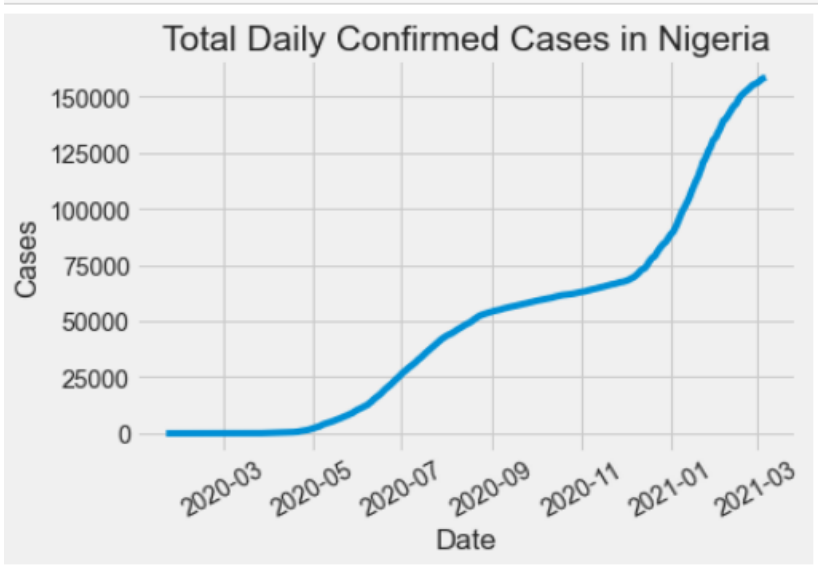
Text(0.5, 0, 'Top 10 Active Cases')

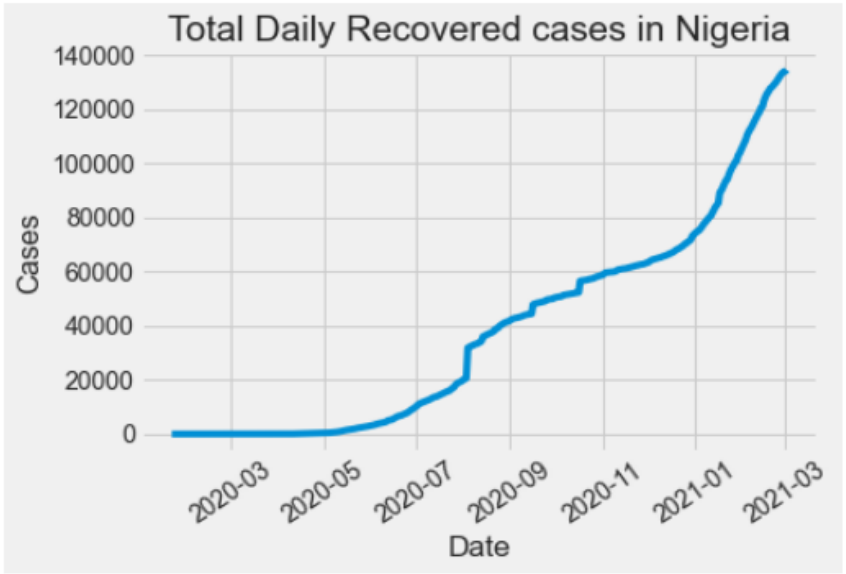


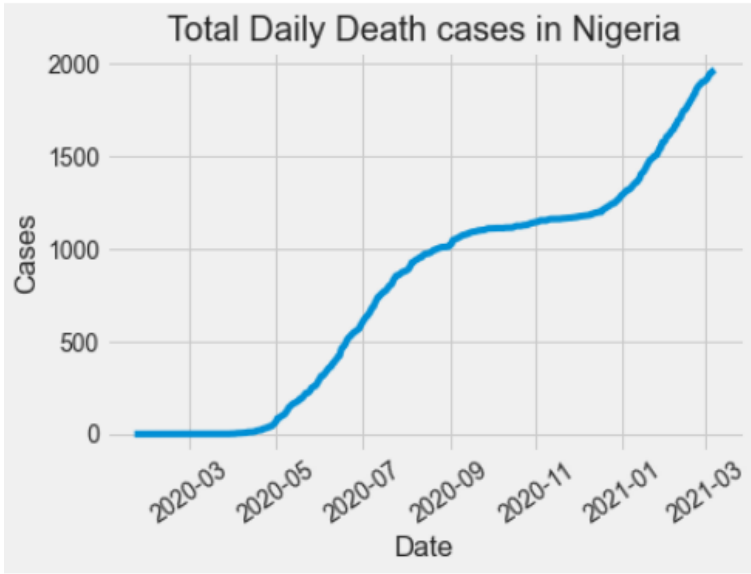
- **Insights**
  - Top 3 Confirmed and Discharge States are Lagos, Abuja and Pleteau.
  - It was shown from the chart above that FCT(Abuja) has the highest number of active cases, even though Lagos has the highest number of confirmed and discharged cases.

ANALYSIS

- Line plots for the total daily confirmed, recovered and death cases in Nigeria respectively were generated.



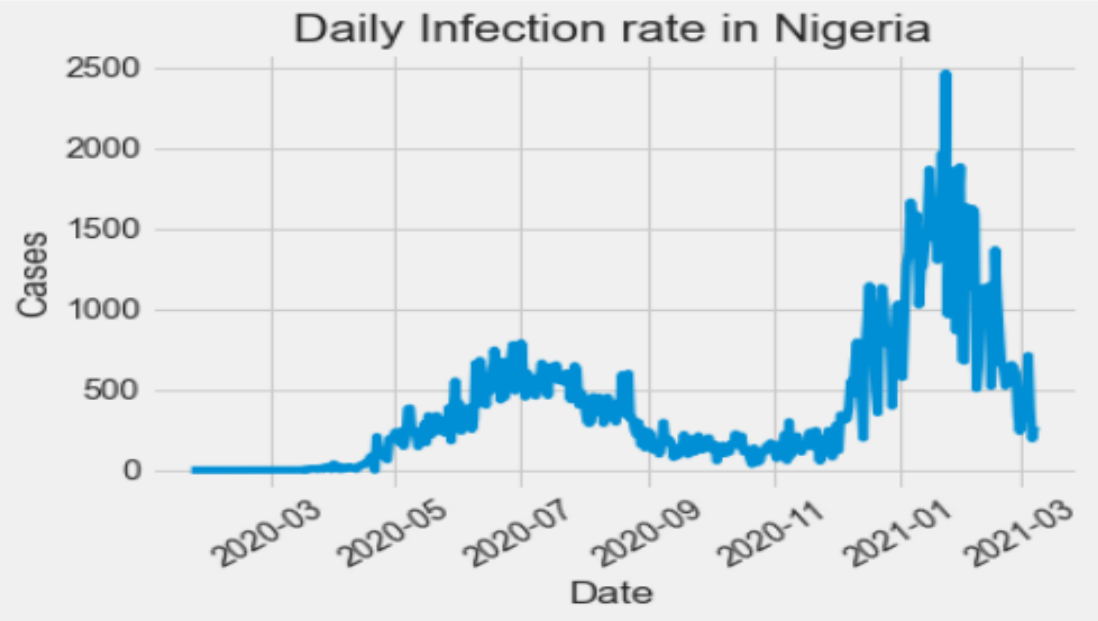




- **Insights**
  - It can be seen that as the confirmed cases are increasing so the recovered cases.
  - It was shown from the charts that since the confirmed cases, there is a fast increase in the death cases from May 2020 until September 2020 to January 2021.

ANALYSIS

- The daily infection rate was determined using the Pandas diff method to find the derivate of the total cases and a plot was generated from the result.



- **Insight**
  - The maximum figure can be observed to be around late January 2021. It is really important that people continue keep to the preventive measures during the festive period.

- Maximum infection rate for a day (Number of new cases) was calculated.
- The date of the result was extracted.

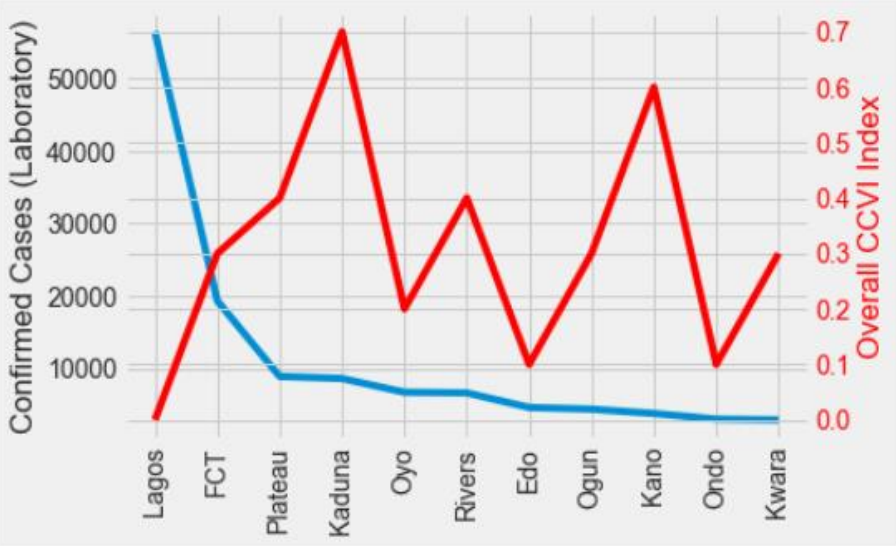
	Date	Cases	Max_Inf_Rate
367	2021-01-23	120602	2464.0

- **Summary**
  - The maximum infection rate was recorded on 23rd of January, 2021. Meaning there were 2464 new cases on that day.

ANALYSIS

- The relationship between the external dataset and the NCDC COVID-19 dataset was determined.

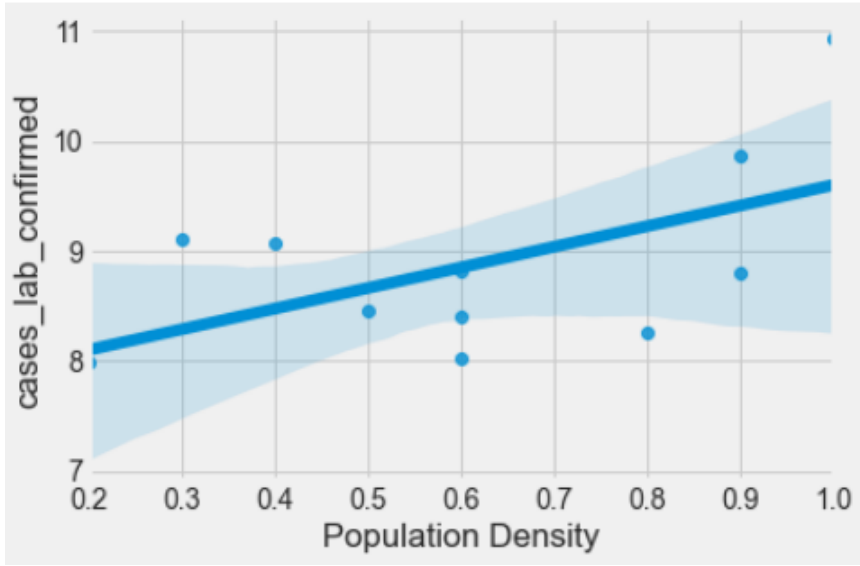
- Here, a line plot of top 10 confirmed cases and the overall community vulnerability index was generated on the same axis.



- **Observation**

  You will notice that the states with the highest vulnerability index score in relation to the confirmed cases are Northern states(Kaduna and Kano) with weak health systems, more poverty etc. Lagos has a high number of confirmed cases but has less vulnerability index score because of good health system, more social distancing etc

- Here, a regression plot between two variables was generated to visualize the linear relationships - Confirmed Cases and Population Density.

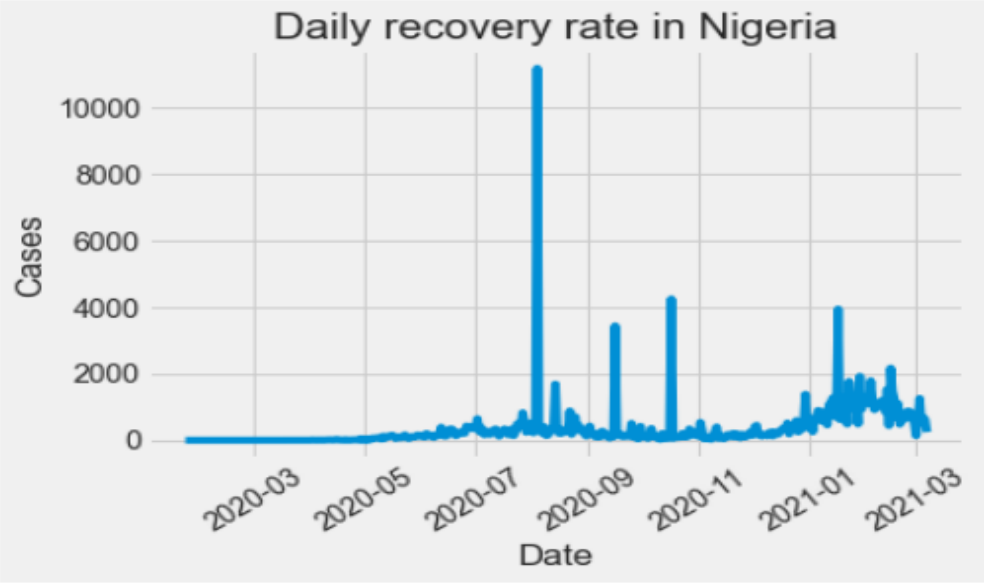


- **Observation**

  Confirmed cases and Population density have positive correlation. There are high cases in areas that are densely populated.

FURTHER ANALYSIS

- The daily recovery rate was determined using the Pandas diff method to find the derivate of the total recovery cases and a plot was generated from the result.



- **Insight**
  - The maximum figure can be observed to be around late August 2020.

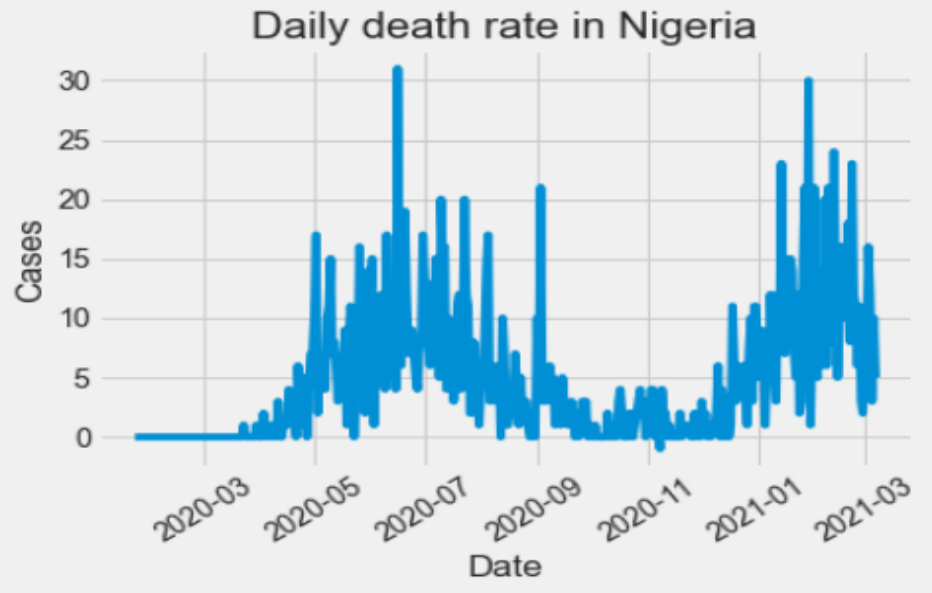
- Maximum recovery rate for a day (Number of recovery cases) was calculated.
- The date of the result was extracted.

	Date	Cases	Max Recovery Rate
195	2020-08-04	31851.0	11188.0

- **Summary**
  - The maximum recovery rate was recorded on 4th of August, 2020. Meaning there were 11188 new recovery cases on that day.

FURTHER ANALYSIS

- The daily death rate was determined using the Pandas diff method to find the derivate of the total death cases and a plot was generated from the result.



- **Insight**
  - The maximum death figure can be observed to be around June 2020. Also, a close value was recorded around February 2021.

- Maximum death rate for a day (Number of death cases) was calculated.
- The date of the result was extracted.

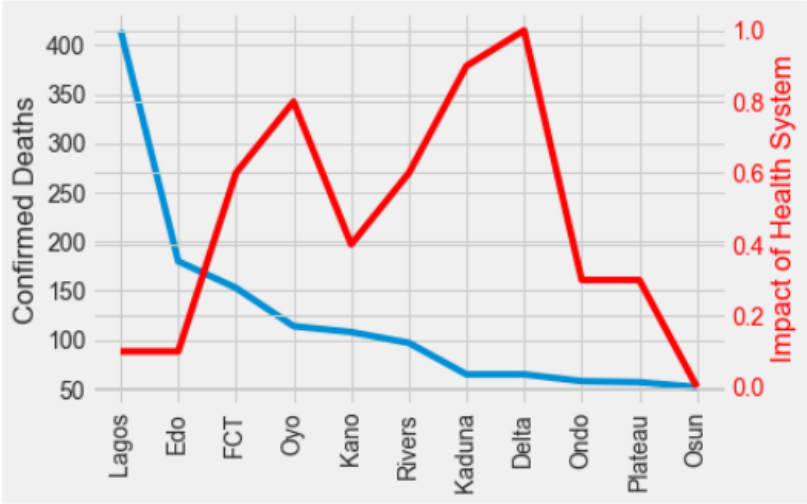
	Date	Cases	Max Death Rate
146	2020-06-16	455	31.0

- **Summary**
- The maximum death rate was recorded on 16th of June, 2020 with 31 new death, even though there were only 455 new cases that day.

FURTHER ANALYSIS

- The relationship between the external dataset and the NCDC COVID-19 dataset was determined.

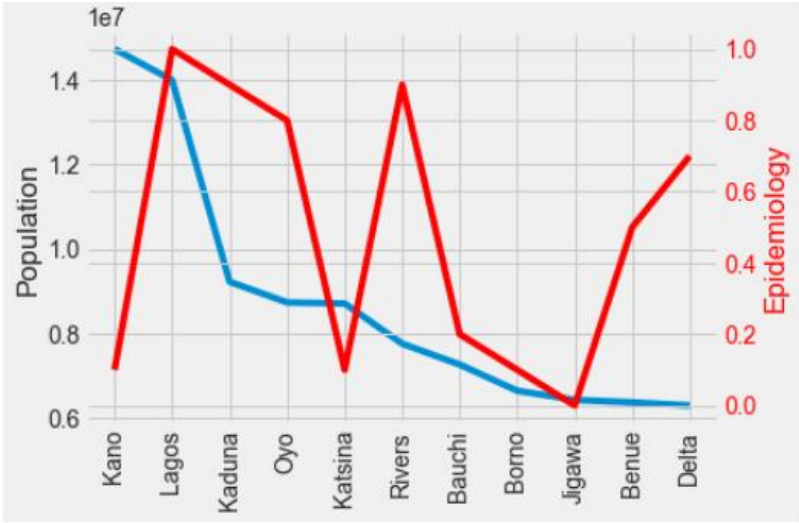
- Here, a line plot of top 10 death cases and the Impact of Health System was generated on the same axis.



- **Observation**

  You will notice that the states with the highest vulnerability index score(Heath System) in relation to the death cases are Delta, Kaduna and Oyo states with weak health systems, less social distancing etc. Lagos has a high number of death cases but has less vulnerability index score because of good health system, more social distancing etc.

- Here, a line plot of top populated states and their epidemiology was generated on the same axis.



- **Observation**

  Kano has the largest population but unfortunately the epidemiology in the state and others like Katsina and Jigawa are so low which means the health system that deals with the control of the disease is very small hence makes them more vulnerable.

RESULT

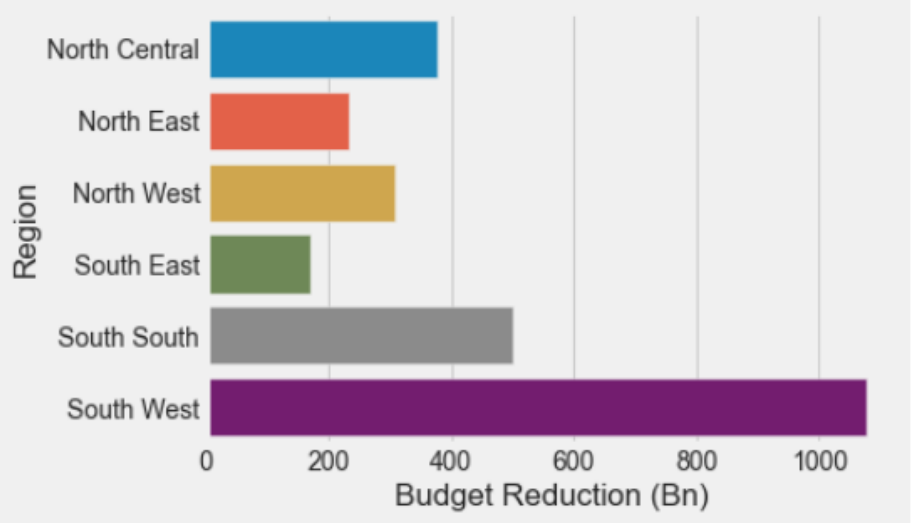
IMPACT OF COVID-19 ON THE NIGERIA ECONOMY

- To determine the effect of the pandemic, the initial and revised budget from the budget data were explored.

- The table below shows the changes in the initial and revised budget per region.

	Initial_budget (Bn)	Revised_budget (Bn)	cases_lab_confirmed	changes
region				
North Central	1136.08	760.00	33249.0	376.08
North East	951.13	720.01	6448.0	231.12
North West	1385.07	1078.50	16323.0	306.57
South East	640.86	470.20	6865.0	170.66
South South	1526.08	1026.45	16402.0	499.63
South West	2774.80	1699.20	73553.0	1075.60

- Here is a plot generated from the resulting table



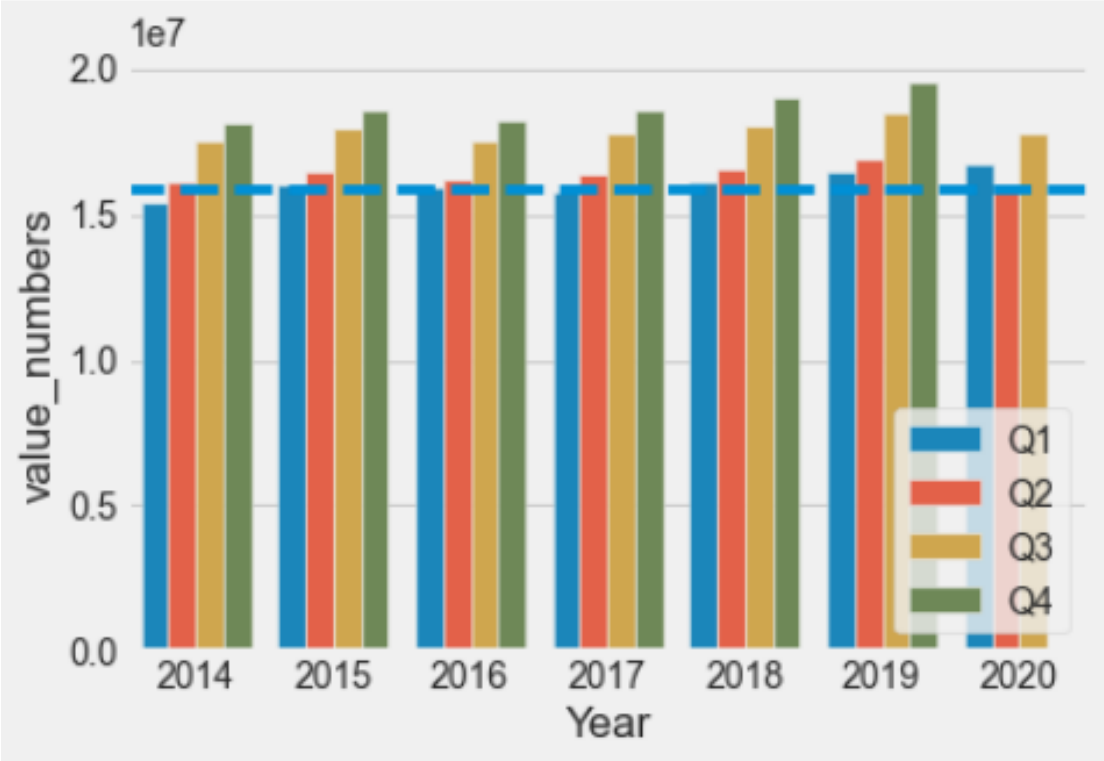
- **Observation**

From the table and plot, the south west region recorded the highest amount of confirmed cases, and also the highest amount of reduction in budget. This is because, the south west region experienced the longest duration of Lock-down, and as such was forced to shut down major economic activities. The effect is seen in budget reduction.

IMPACT OF COVID-19 ON THE NIGERIA ECONOMY

- The Real GDP value Pre-COVID-19 was compared with Real GDP in 2020 (COVID-19 Period, especially Q2 2020).

- From the Real GDP Data, a barplot was generated using the GDP values for each year & quarters.



- **Observation**

From the chart, it can be seen that since 2014, the GDP has been increasing gradually until 2020 when the economy experience a pandemic and contrary to the normal trend of increment, there has been reduction from the quarters of the year 2020. During the second quarter in 2020 (Q2) the economy size has been pulled down to the size of the first quarter in 2017 due to the several movement restriction on some products and services. This means the economy might not balance back from the pandemic to the pre-pandemic level in the next two years(2023).

CONCLUSION

- Using python with data science & analytics skills to collect data, explore the data, perform analysis, create visualizations, and generate insights, the inherent impact of the pandemic on the Nigeria economy was identified and proper planning should be done.

- It is important to note that from the analysis done, the economy might not balance back from the pandemic to the pre-pandemic level in the next two years(2023).

SUGGESTION ON FUTURE WORK

- Ability to get extract COVID 19 data from post and comments on social media, analyse and generate insights on the breakdown of the effect on:
  - frontline and essential workers
  - Primary, Secondary and Tertiary Education
  - Employment rate and effect on each job sectors

- Ability to do analysis on the rate of vaccine availability, vaccine collection per confirmed cases and non confirmed cases and effect on the confirmed cases.

- Ability to include Artificial Intelligence (AI) into the project to design a model that train and predict the effect and opportunities of COVID 19 in the next few years.

Thank You