# USTACKY MICRO DEGREE COURSE

## DATA SCIENTIST MICRODEGREE CAPSTONE PROJECT

---

**Executive Summary - Nigeria-COVID-19-Data-Analysis-Using-Python**

---

**NAME:** **ADEBANJO ESTHER CHIZOBA**

(Student)

**13th, July, 2023**

# Contents

# 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.

# Description

In this project, data science & analytics skills will be employed to collect data, explore the data, perform analysis, create visualizations, and generate insights.
The data source is divided into different parts, and you will combine the data to perform analysis and provide insights.

# Data collection

All the required libraries were imported and data was imported from the repository since the NCDC COVID-19 official website could not be accessed.

# Data cleaning and preparation

- Data were cleaned by dropping columns with a null value and zero value, removing commas from numerical values
- Data were also converted to appropriate datatypes
- Nigeria data was extracted from the world COVID-19 record for analysis
- The information was transposed to give it a better presentation
- Pandas DataFrame for Daily Confirmed, Recovered and deaths Cases were gotten

-

# Data analysis

The following analysis were carried out:

- Top 10 states in terms of Confirmed Covid cases by Laboratory test and Discharged Covid cases were generated
- Plot the top 10 Death cases
- A line plot for the total daily confirmed, recovered, and death cases in Nigeria was generated.
- The daily infection rate was determined to show the derivate of the total cases on a line plot
- The maximum infection rate for a day was calculated signifying the number of new cases and the date found.
- The relationship between the external dataset and the NCDC COVID-19 dataset was determined and a line plot of the top 10 confirmed cases and the overall community vulnerability index on the same axis was generated.
- The two datasets above were combined on a common column and the relationship between them was generated.
- A regression plot between two variables (Confirmed Cases and Population Density) was generated to visualize the linear relationships.

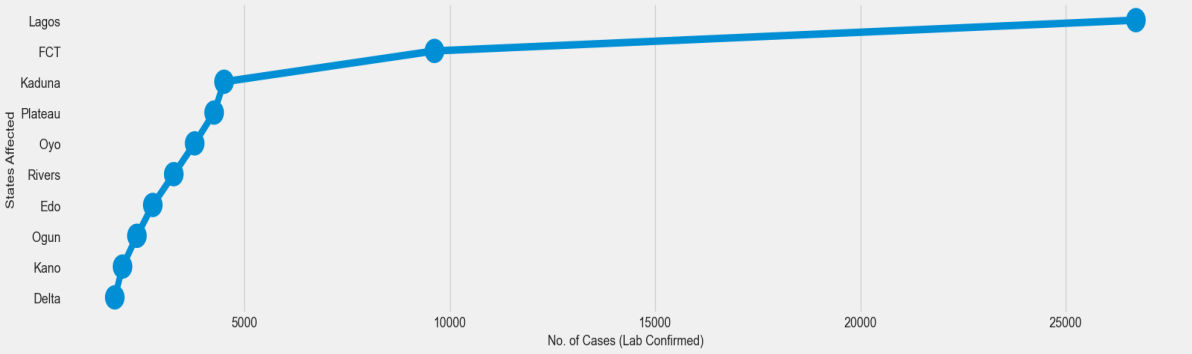
# Insight

- The global covid_19 dataset consists of 289 rows and 1147 columns while Nigeria data consist of 1 row and 1107 valid columns
- Top 10 states in terms of Confirmed Covid cases by Laboratory test and Discharged Covid cases and death cases are shown below:

## Top 10 Confirmed Covid cases

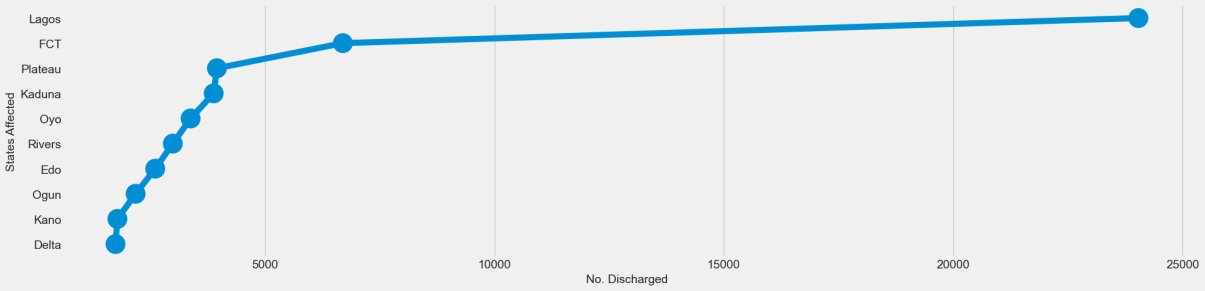
:

| | States Affected | No. of Cases (Lab Confirmed) |
|---|---|---|
| 0 | Lagos | 26708 |
| 1 | FCT | 9627 |
| 2 | Kaduna | 4504 |
| 3 | Plateau | 4262 |
| 4 | Oyo | 3788 |
| 5 | Rivers | 3279 |
| 6 | Edo | 2768 |
| 7 | Ogun | 2382 |
| 8 | Kano | 2032 |
| 9 | Delta | 1843 |



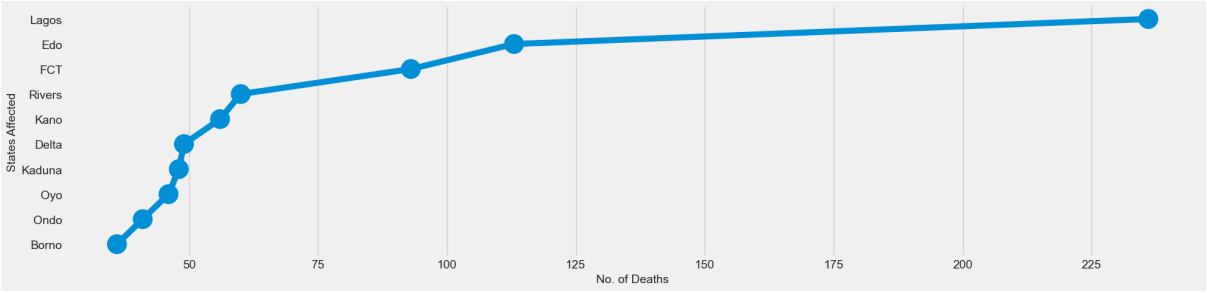

## Top 10 Discharged cases

| | States Affected | No. Discharged |
|---|---|---|
| 0 | Lagos | 24037 |
| 1 | FCT | 6694 |
| 2 | Plateau | 3948 |
| 3 | Kaduna | 3877 |
| 4 | Oyo | 3374 |
| 5 | Rivers | 2987 |
| 6 | Edo | 2603 |
| 7 | Ogun | 2175 |
| 8 | Kano | 1778 |
| 9 | Delta | 1737 |




## The top 10 Death cases

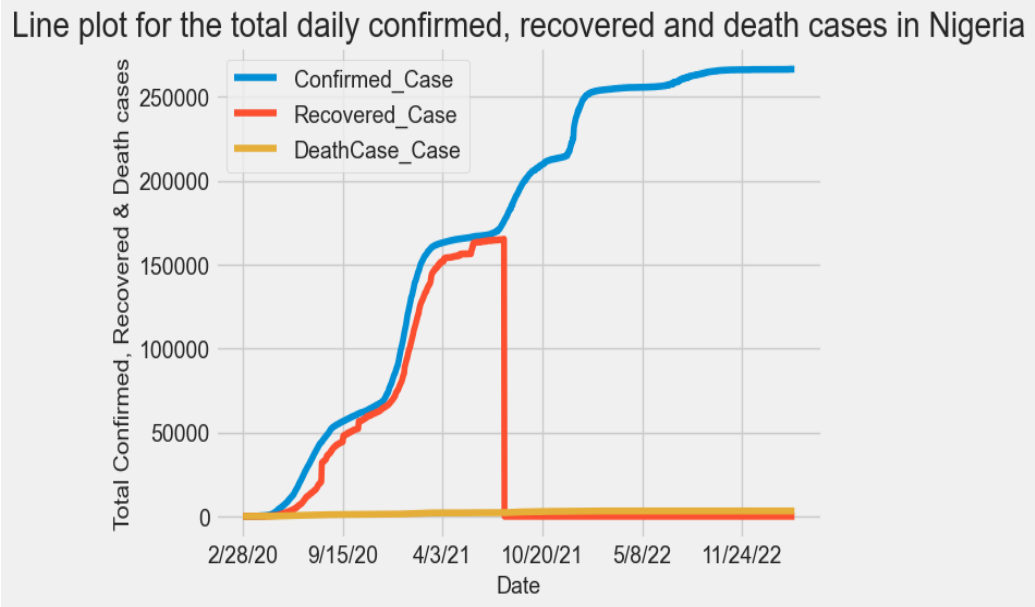
| | States Affected | No. of Deaths |
|---|---|---|
| 0 | Lagos | 236 |
| 1 | Edo | 113 |
| 2 | FCT | 93 |
| 3 | Rivers | 60 |
| 4 | Kano | 56 |
| 5 | Delta | 49 |
| 6 | Kaduna | 48 |
| 7 | Oyo | 46 |
| 8 | Ondo | 41 |
| 9 | Borno | 36 |



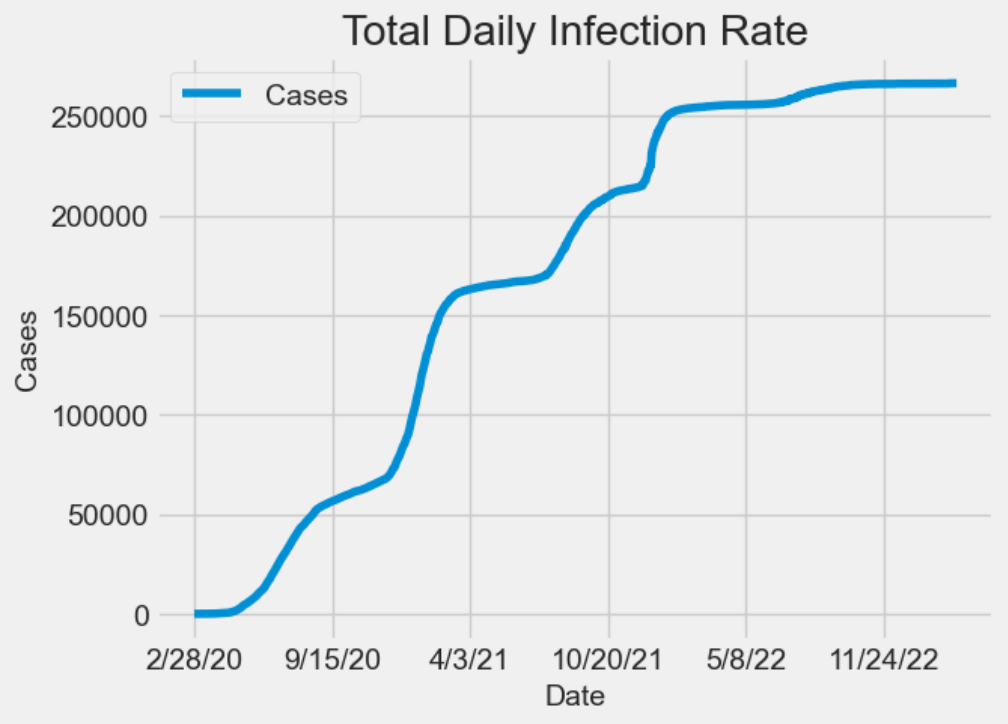
**Total daily confirmed, recovered, and death cases in Nigeria**

:

| Date | Confirmed_Case | Recovered_Case | DeathCase_Case |
|---|---|---|---|
| 2/28/20 | 1 | 0 | 0 |
| 2/29/20 | 1 | 0 | 0 |
| 3/1/20 | 1 | 0 | 0 |
| 3/2/20 | 1 | 0 | 0 |
| 3/3/20 | 1 | 0 | 0 |
| ... | ... | ... | ... |
| 3/5/23 | 266598 | 0 | 3155 |
| 3/6/23 | 266598 | 0 | 3155 |
| 3/7/23 | 266598 | 0 | 3155 |
| 3/8/23 | 266598 | 0 | 3155 |
| 3/9/23 | 266598 | 0 | 3155 |

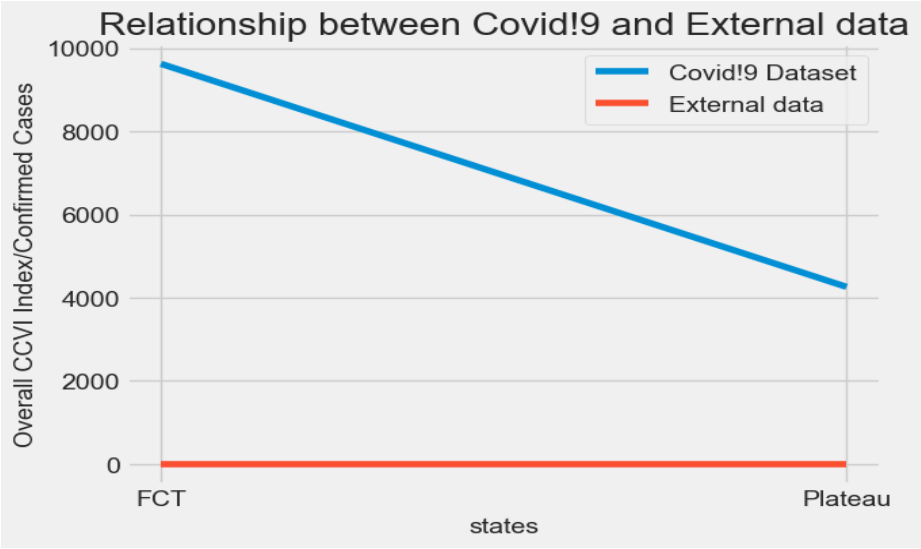



**Daily Infection rate**

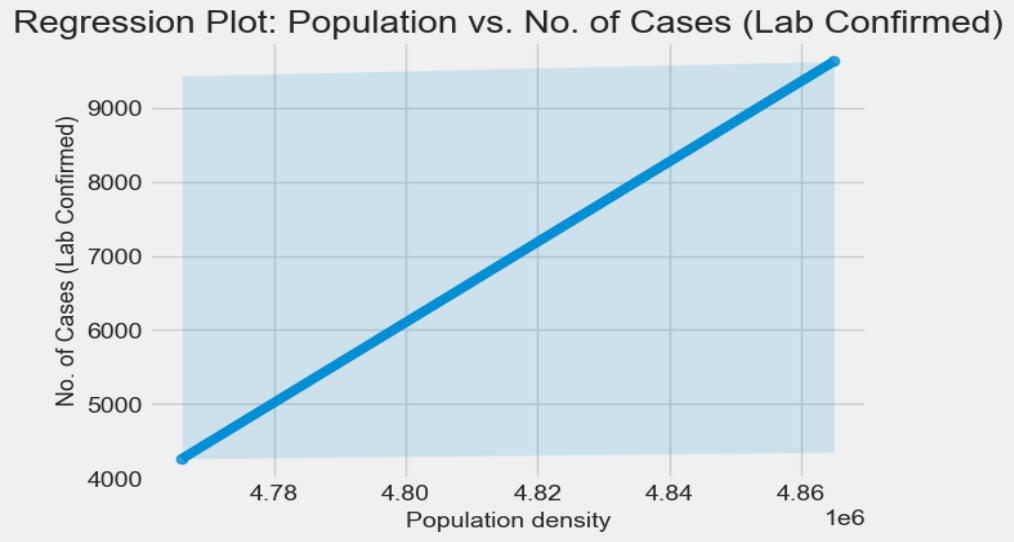



The maximum infection value for a day was found to be 103799897.0 and occurred on 3/9/23

A line plot of the top 10 confirmed cases and the overall community vulnerability index



A regression plot between two variables (Confirmed Cases and Population Density)




The effect of the pandemic on the economy is shown by the graph below: