USTACKY

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 lockdowns, curfews, and reviving its economy.

Description

In this project, data science & analytics skills (web scraping)were employed to collect data, explore the data, perform analysis, create visualizations, and generate insights.
The data source is divided into different parts, and some data were combined to create another dataset used to perform analysis and provide insights.

Data collection

After importing all required libraries, datasets were collected from The Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE) repository using the Urlib.request since the NCDC COVID-19 official website could not be accessed. These datasets includes:

- Global daily confirmed cases
- Global daily recovered cases
- Global daily death cases
- Covid_external data
- Budget
- Real GDP

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 dataframe was transposed to give it a better presentation
- Pandas DataFrame for Daily Confirmed, Recovered and deaths Cases were gotten

Data analysis

The following analysis was 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. 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.
- Cases and Population Density) was generated to visualize the linear relationships.
- Analysis of the real GDP data shows that the GDP fell from 413801074.72 Pre-Covid_19 to 15890000.0 during Q2 2020, a %change of 96%. A visualization was also presented.
- A bar plot showing the initial and revised budget for each state was presented.

Insight

- The global covid_19 dataset consists of 289 rows and 1144 valid columns while Nigeria's data consist of 1 row and 506 valid columns
- From the bar chart in Figure 1, Lagos has the most confirmed cases in Nigeria with 26708 cases while Delta state has the least cases with 1843 cases.
- From the bar chart in Figure 2, Lagos has the most discharged cases in Nigeria with 24037 cases while Delta state has the least cases with 1737 cases.
- From the bar chart in Figure 3, Lagos has the most death cases in Nigeria with 236 cases while Delta state has the least cases with 36 cases.
- Figure 4 shows that very few death cases in Nigeria
- The maximum infection rate for a day was 103799897 and this happened on 3/9/23.
- Analysis of the real GDP data shows that the GDP fell from 413801074.72 Pre-Covid_19 to 15890000.0 during Q2 2020, a %change of 96%.
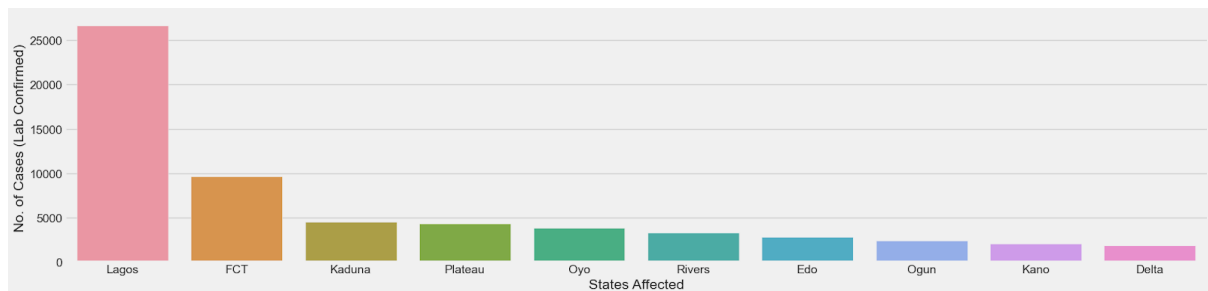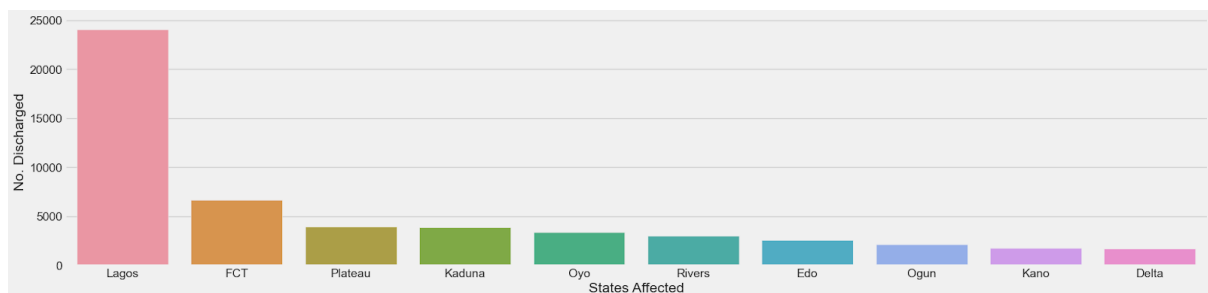

Figure 1: Top 10 Confirmed Covid cases
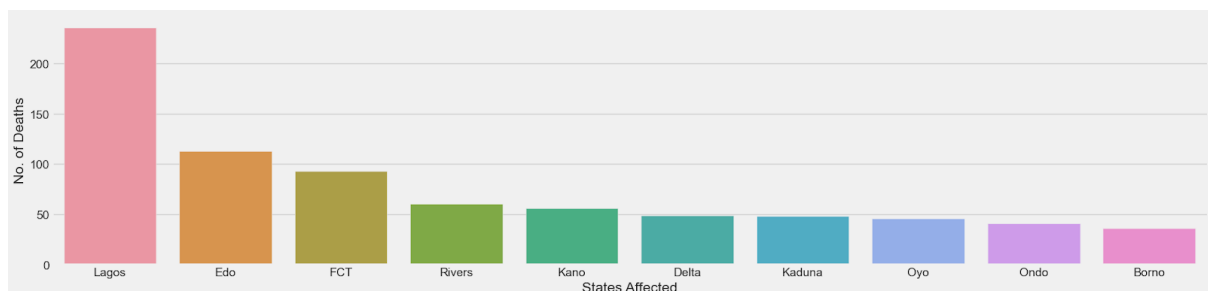

Figure 2: Top 10 Discharged cases

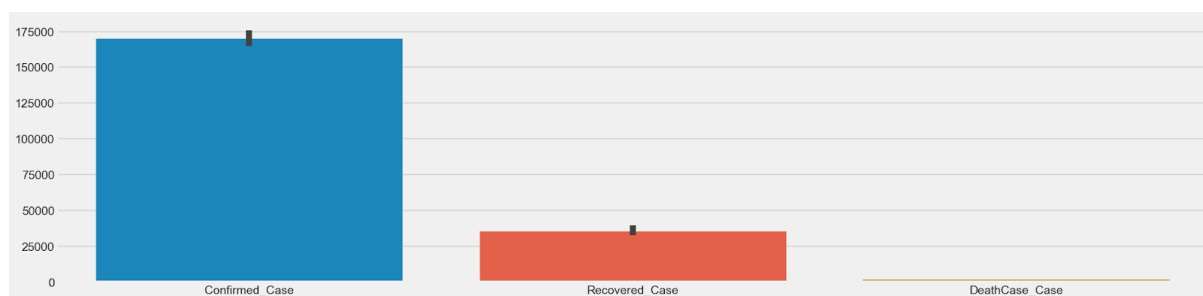*Figure 3: The top 10 Death cases*

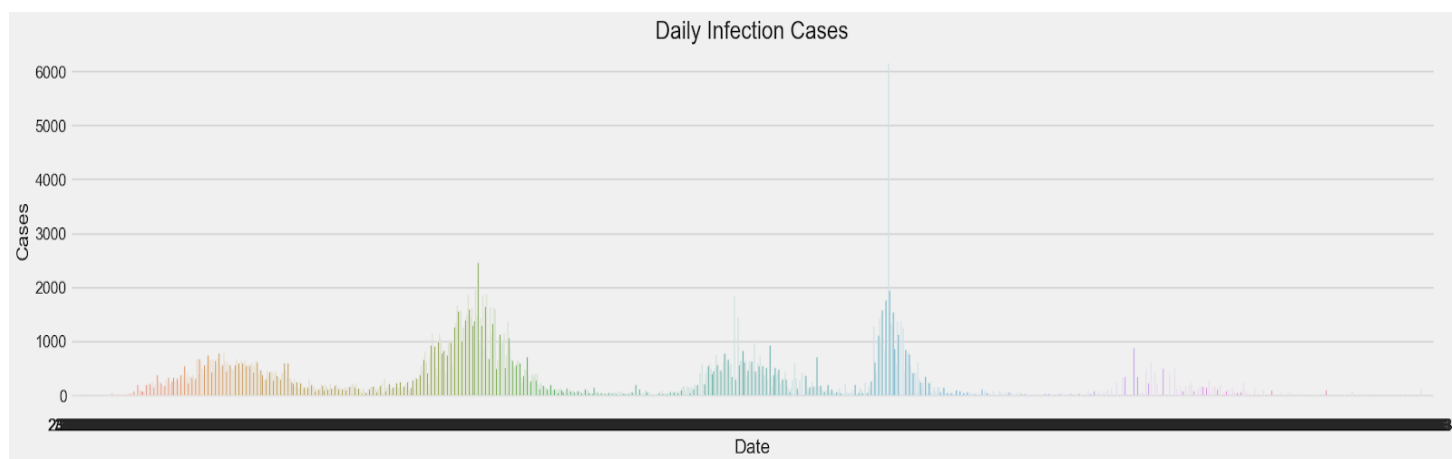*Figure 4: Combined Cases (Confirmed, Recovered & Death)*
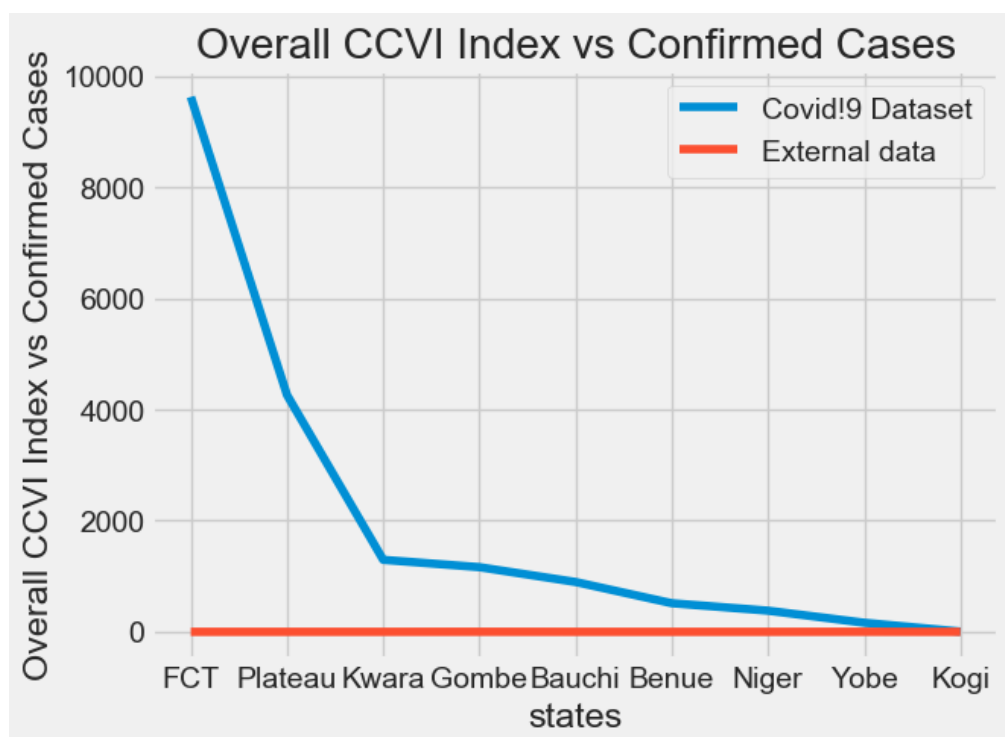

Figure 5: Daily infection rate

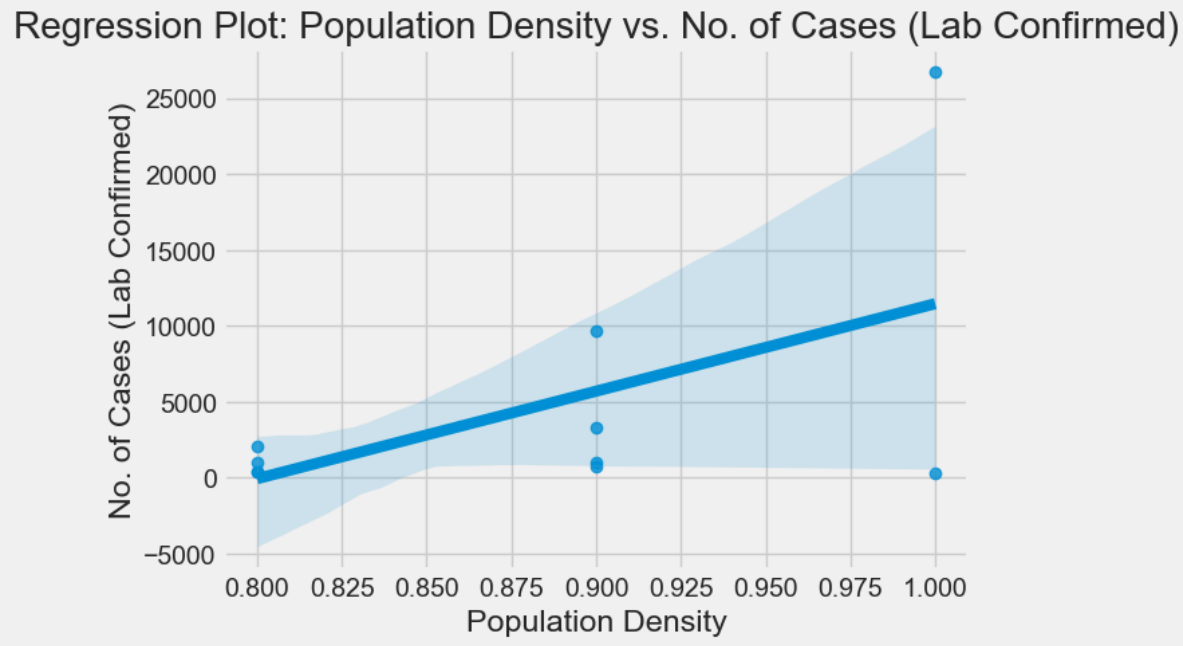*Figure 6: Top 10 confirmed cases and the overall community vulnerability index*

Figure 7: Regression plot between two variables (Confirmed Cases and Population Density)
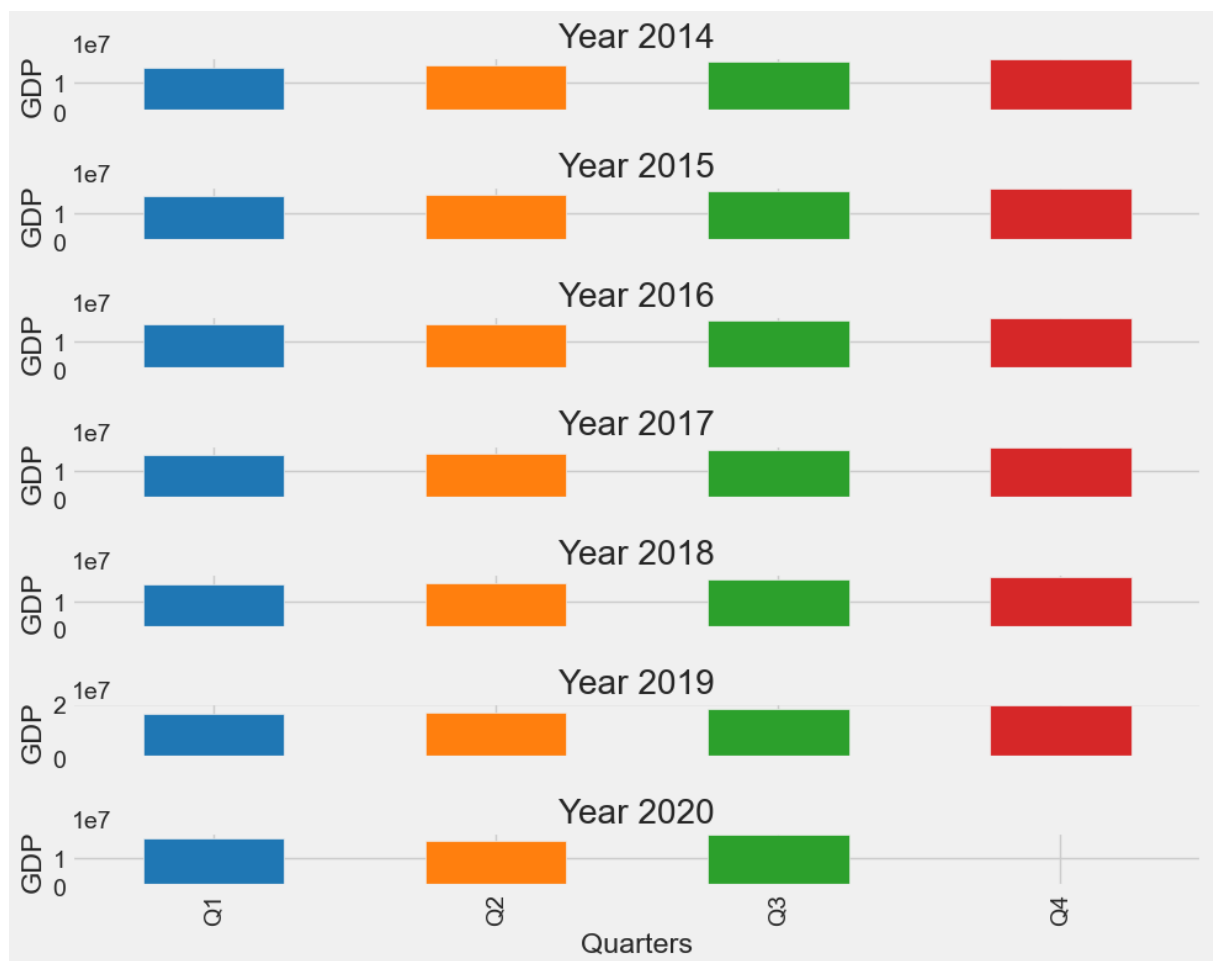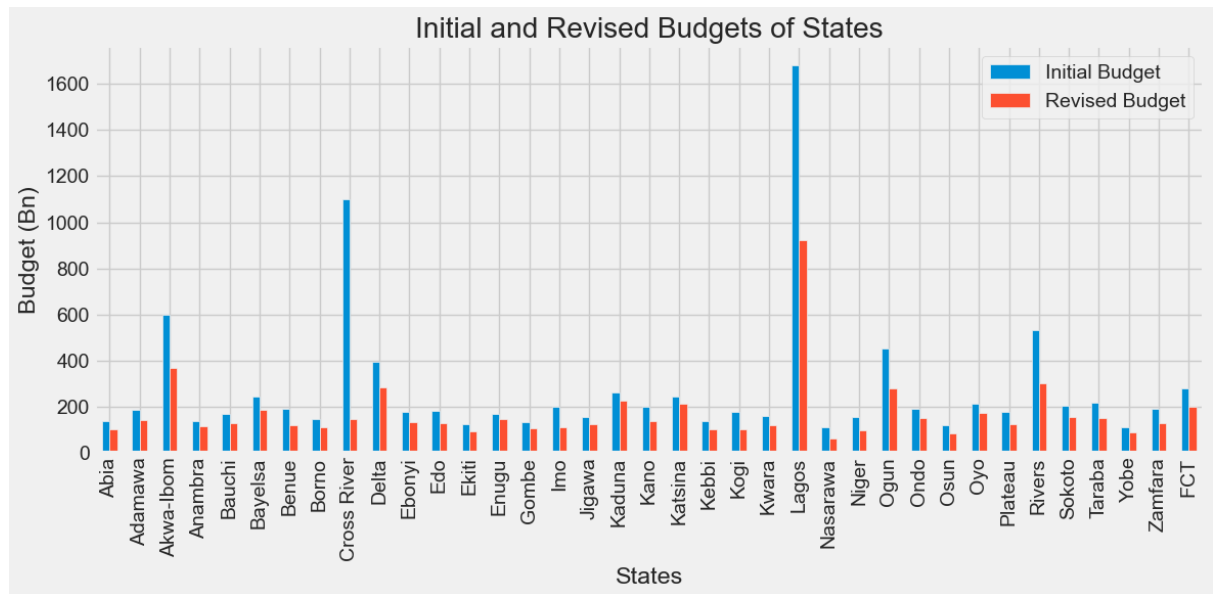

Figure 8: Real GDP

Figure 9: Initial and Revised Budget