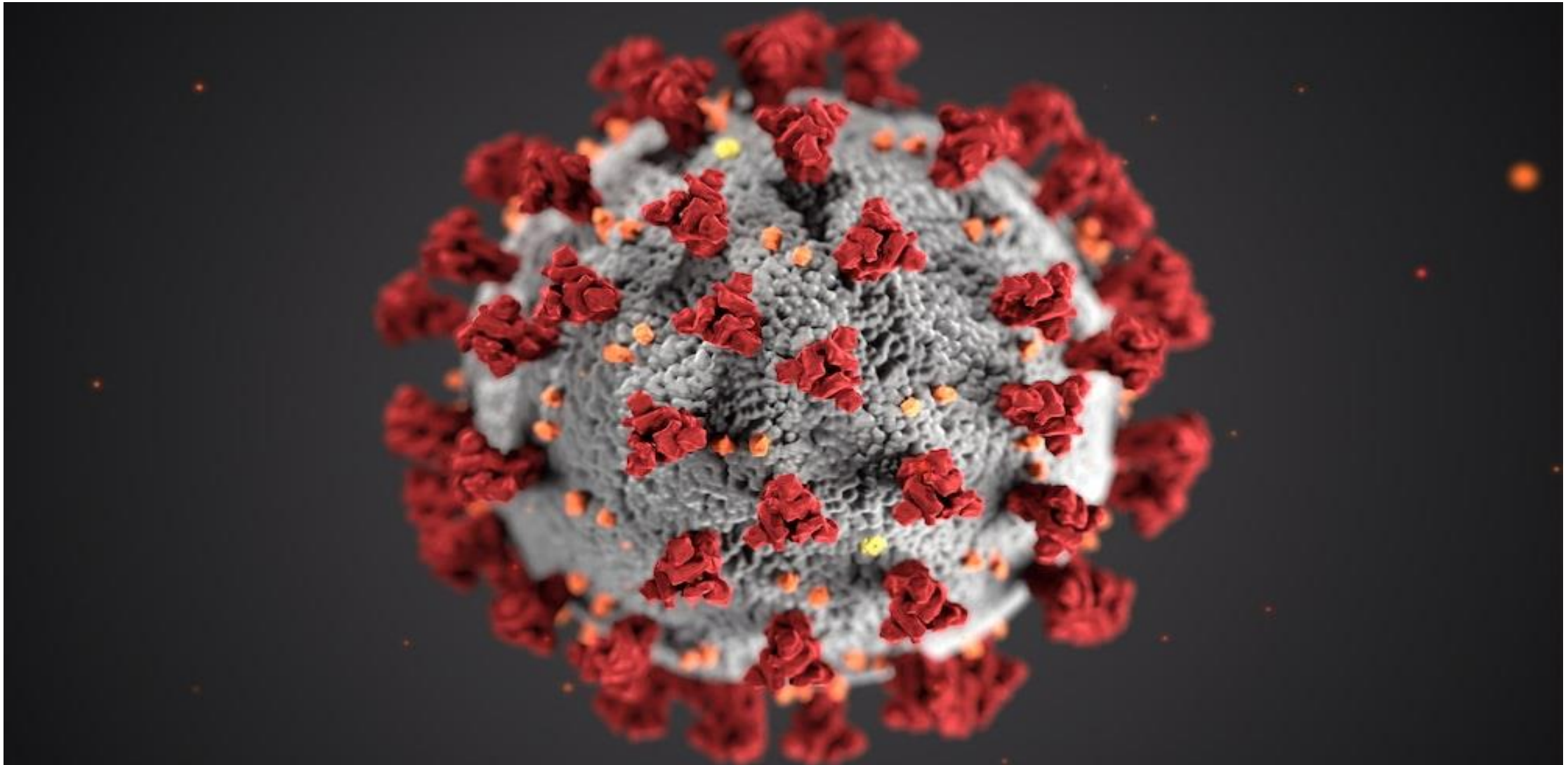
Nigeria COVID-19 Data Analysis using Python

Nigeria COVID-19 Data Analysis using Python



Project Overview

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.

Project 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.
- Report writing

Data Collection

The data source is divided into different parts, and combine all the different parts of the data to perform analysis and provide insights.

- The Nigeria Centre for Diseases Control (NCDC) monitors the country's COVID-19 situation, and releases data on metrics across all the 37 states in the country. Data can be obtained from NCDC COVID-19 official website by performing a web extraction or web scraping. The NCDC COVID-19 contains the affected states the number of confirmed cases, number of discharged cases, number of admission cases, and number of death.
- The Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE) publishes daily data on confirmed, death and recovered cases across different countries but we are to extract the data for Nigeria only.

- Nigeria Community Vulnerability Index data →The vulnerability index 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
  The data is used to determine the impact of COVID-19 on the economy.
- State Budget Data: This data shows the effect of COVID-19 on the economy for all the States in Nigeria.

Data cleaning and Collection of Covid-19 Nigeria Dataset using Webscraping(pandas)

Data cleaning and Collection of Covid-19 Nigeria Dataset Using web scraping without cleaning.

States Affected	No. of Cases (Lab Confirmed)	No. of Cases (on admission)	No. Discharged	No. of Deaths	
0	Lagos	104,286	1,143	102,372	771
1	FCT	29,535	9	29,277	249
2	Rivers	18,112	-3	17,960	155
3	Kaduna	11,672	2	11,581	89
4	Plateau	10,365	4	10,286	75

Data cleaning and Collection of Covid-19 Nigeria Dataset Using web scraping after cleaning.

	States Affected	No. of Cases (Lab Confirmed)	No. of Cases (on admission)	No. Discharged	No. of Deaths
0	Lagos	104286	1143	102372	771
1	FCT	29535	9	29277	249
2	Rivers	18112	-3	17960	155
3	Kaduna	11672	2	11581	89
4	Plateau	10365	4	10286	75

```
df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 37 entries, 0 to 36
Data columns (total 5 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   States Affected           37 non-null     object
 1   No. of Cases (Lab Confirmed)  37 non-null     object
 2   No. of Cases (on admission)   37 non-null     object
 3   No. Discharged            37 non-null     object
 4   No. of Deaths             37 non-null     int64
dtypes: int64(1), object(4)
memory usage: 1.6+ KB
```

Extract the first five affected states

States Affected	No. of Cases (Lab Confirmed)	No. of Cases (on admission)	No. Discharged	No. of Deaths	
0	Lagos	104286	1143	102372	771
1	FCT	29535	9	29277	249
2	Rivers	18112	-3	17960	155
3	Kaduna	11672	2	11581	89
4	Plateau	10365	4	10286	75

Renaming each columns

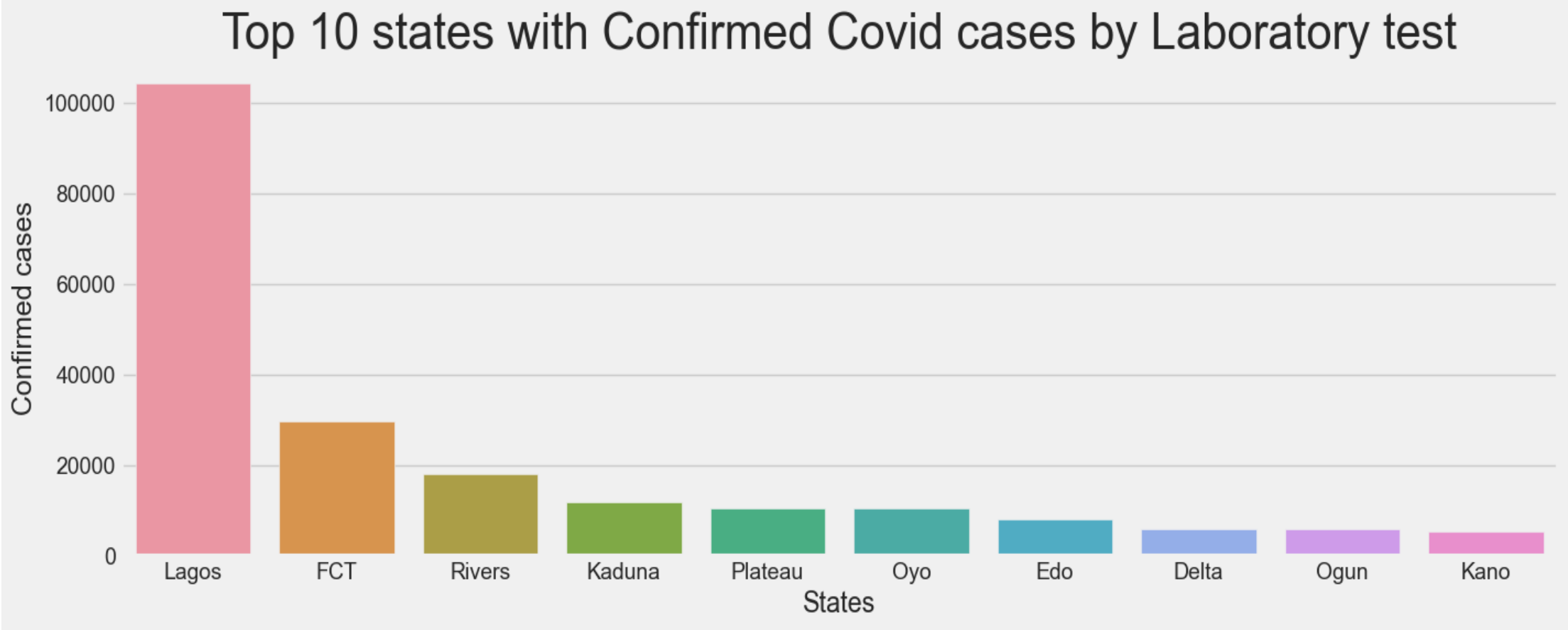
States	Confirmed cases	Admitted cases	Recovered cases	Death	
0	Lagos	104286	1143	102372	771
1	FCT	29535	9	29277	249
2	Rivers	18112	-3	17960	155
3	Kaduna	11672	2	11581	89
4	Plateau	10365	4	10286	75

Generate the statistical summary of the dataframe

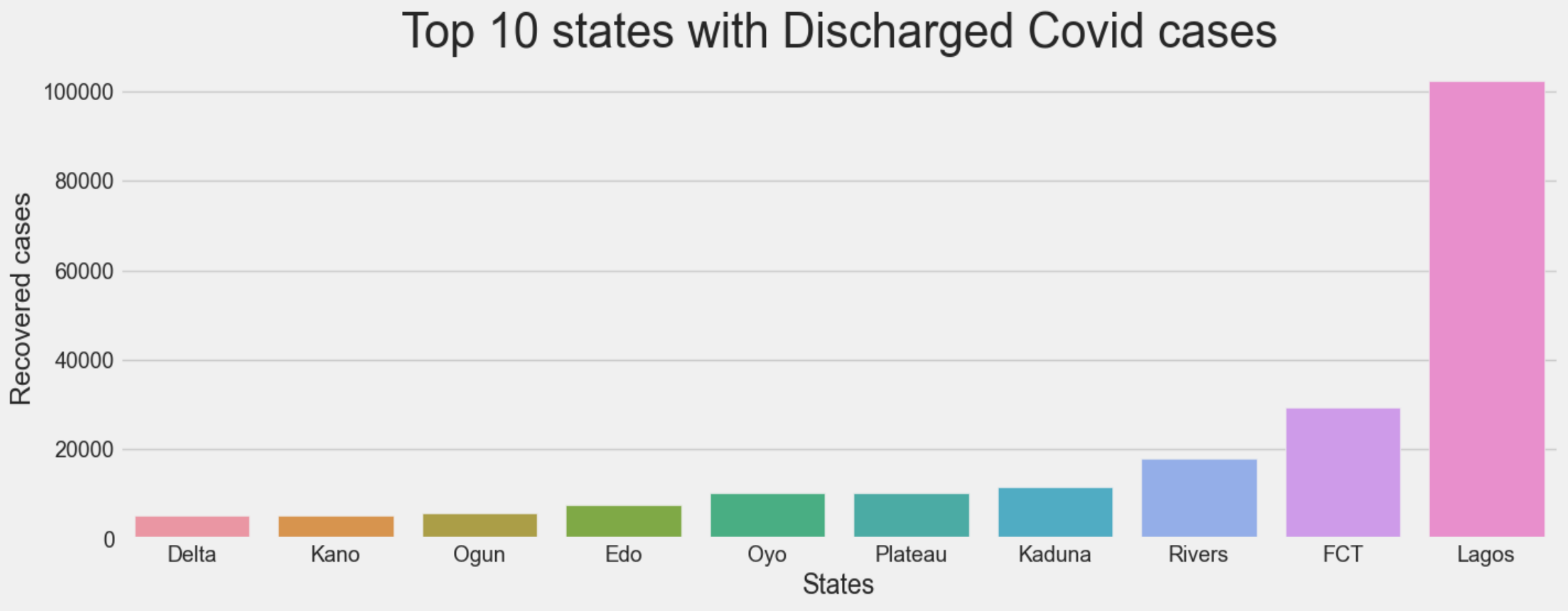
	Confirmed cases	Admitted cases	Recovered cases	Death	
count	37.000000	37.000000	37.000000	37.000000	
mean	7207.162162	96.189189	7025.702703	85.270270	
std	17349.875642	227.994498	17058.840709	134.897008	
min	5.000000	-3.000000	3.000000	2.000000	
25%	1373.000000	0.000000	1343.000000	25.000000	
50%	2691.000000	6.000000	2438.000000	38.000000	
75%	5429.000000	32.000000	5170.000000	89.000000	
max	104286.000000	1143.000000	102372.000000	771.000000	

From the above result

- The least death recorded is 2 while the highest death recorded is 771 and the average is approximatly 85.
- The highest Confirmed cases is 104285 and the least is 5 while the average corfirmed case is approximately 2120.
- The highest admitted cases recorded is 1142 and the average recorded is 96 approximately.
- Admitted cases took place everyday
- Highest number of patients recover is 102372
- Average number of covid19 deaths across all states in Nigeria is 37
- Highest number of deaths recorded in a state is 771 and minimum is 2.
- 75% of death cases is 89 deaths while 25% of death cases result 25.

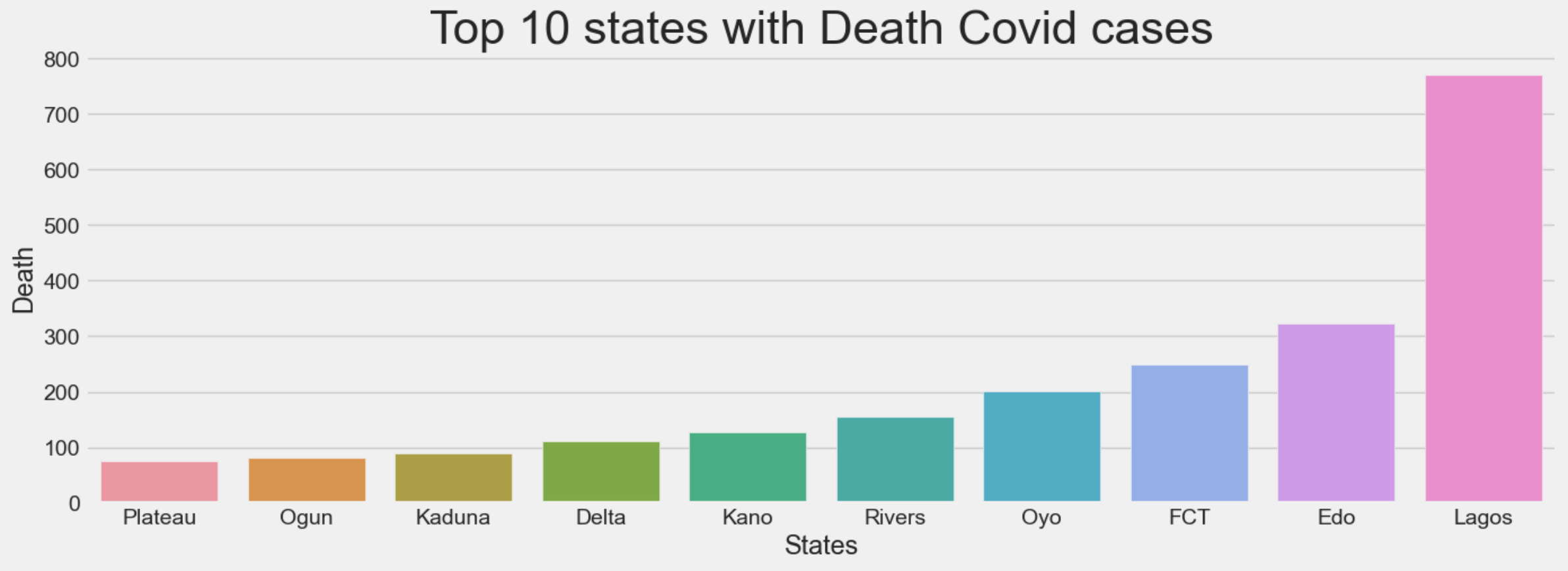


From the above result, Lagos State recorded highest number of confirmed cases with over 100000 while Kano, Delta and Ogun recorded least with less than 1000 patients.



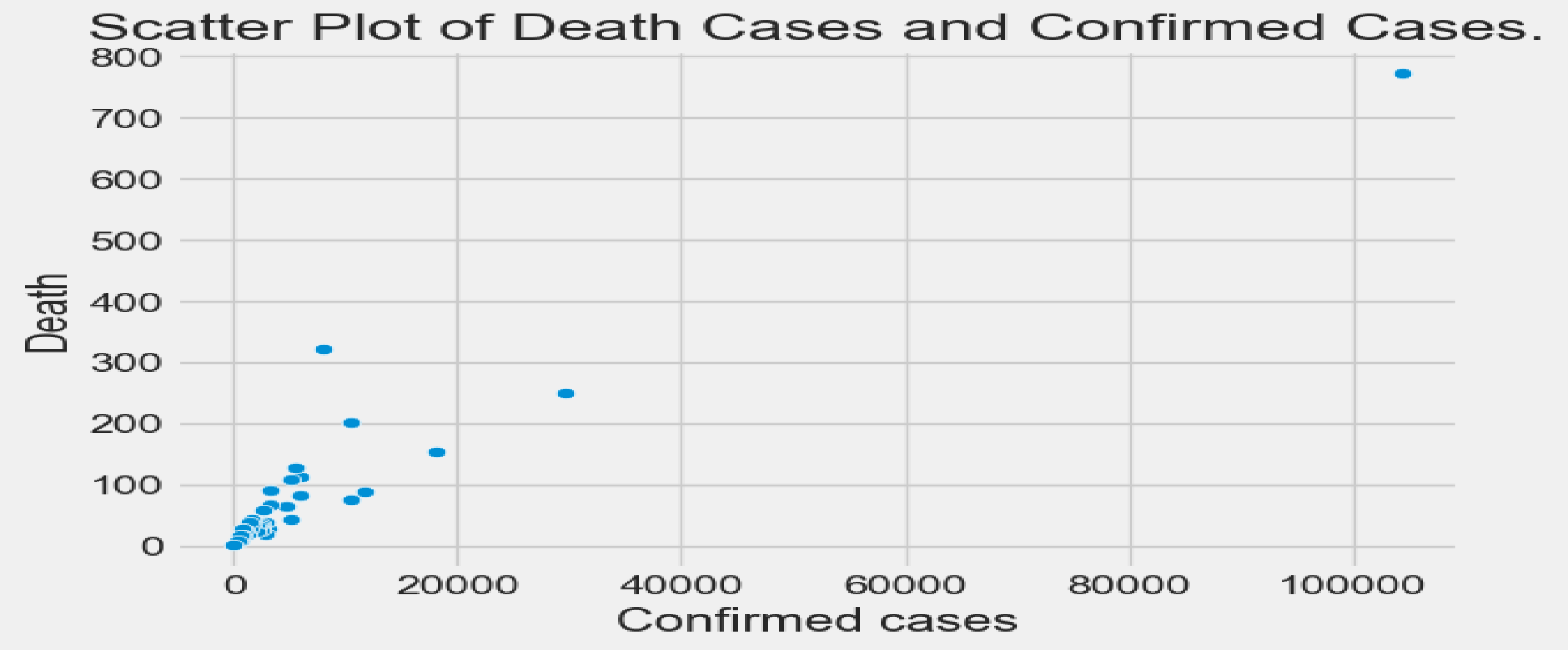
From the above result,

- Lagos State recorded highest number of discharged patients with over 100000 while Delta, Kano, Edo and Ogun recorded the least with less than 1000 patients.



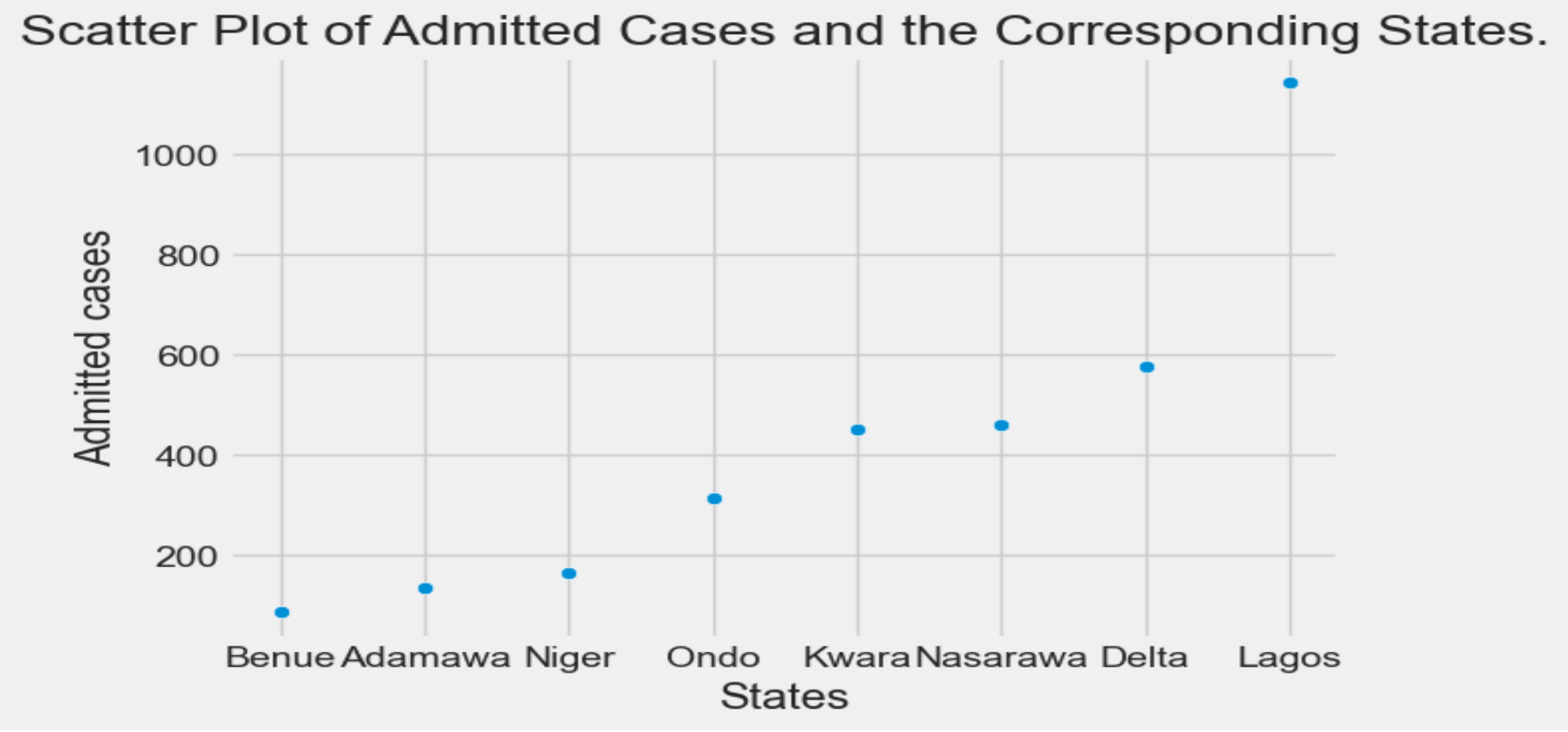
From the bar chart result above shows that:

- Lagos has the highest number of death patient with over 700.
- Edo is the second highest number of death rate follow by FCT with more than 300 and more than 200 respectively.
- Plateau, Ogun and Kaduna States recorded less than 100 death cases which is the least from the above result..

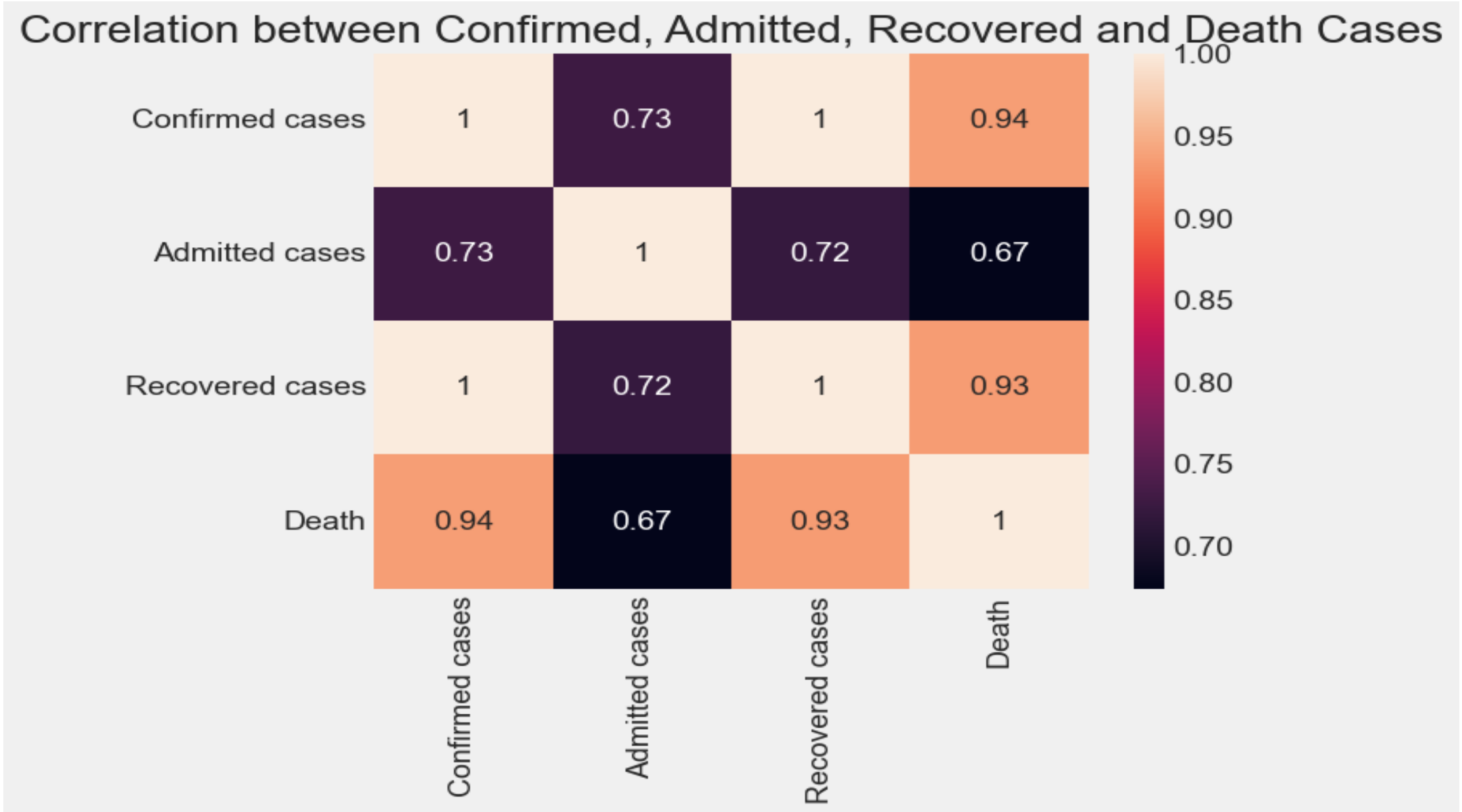


From the result above, it can be concluded that:

The number of confirmed cases raise to over 800 while Number of Confirmed Cases raise to over 100000

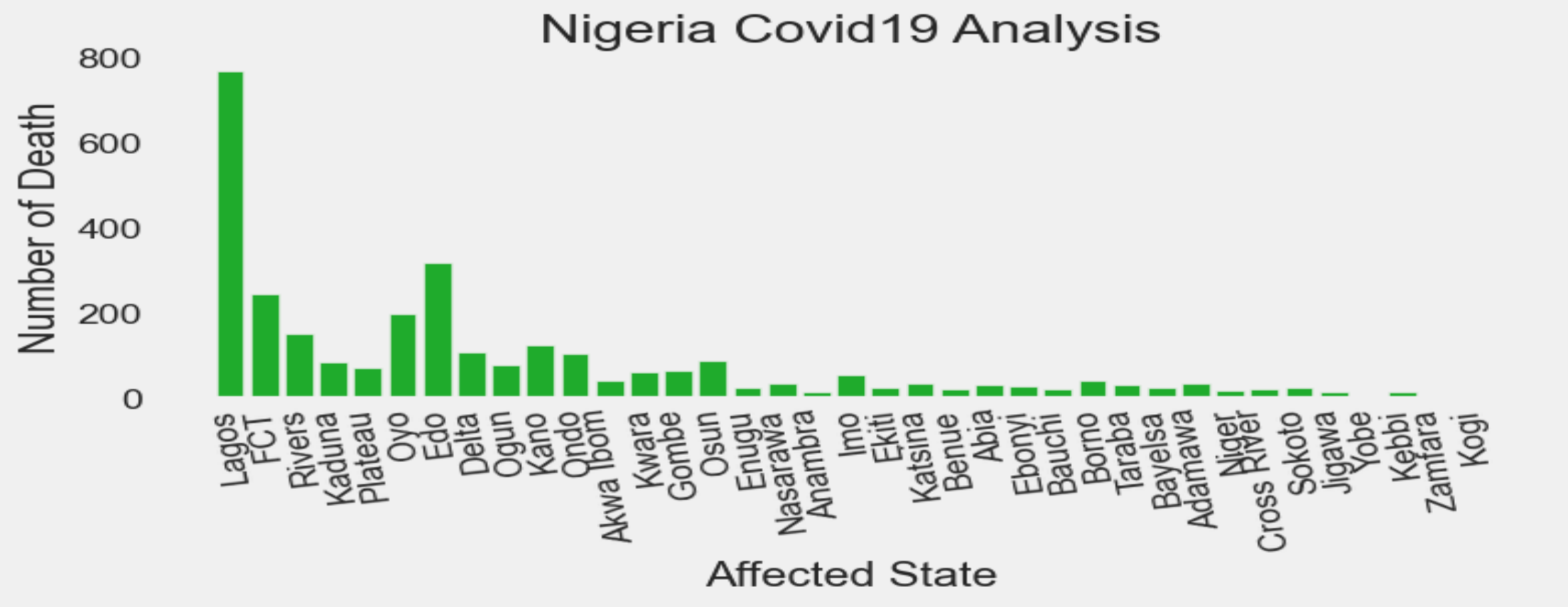


From the graph above result, it is shown Lagos recorded the highest number of admission case with more than 1500 patients follow by Delta close to 600


Correlation between Confirmed, Admitted, Recovered and Death Cases
Confirmed cases
Admitted cases
Recovered cases
Death
1
0.73
1
0.94
0.73
1
0.72
0.67
1
0.72
1
0.93
0.94
0.67
0.93
1
Confirmed cases
Admitted cases
Recovered cases
Death
1.00
0.95
0.90
0.85
0.80
0.75
0.70

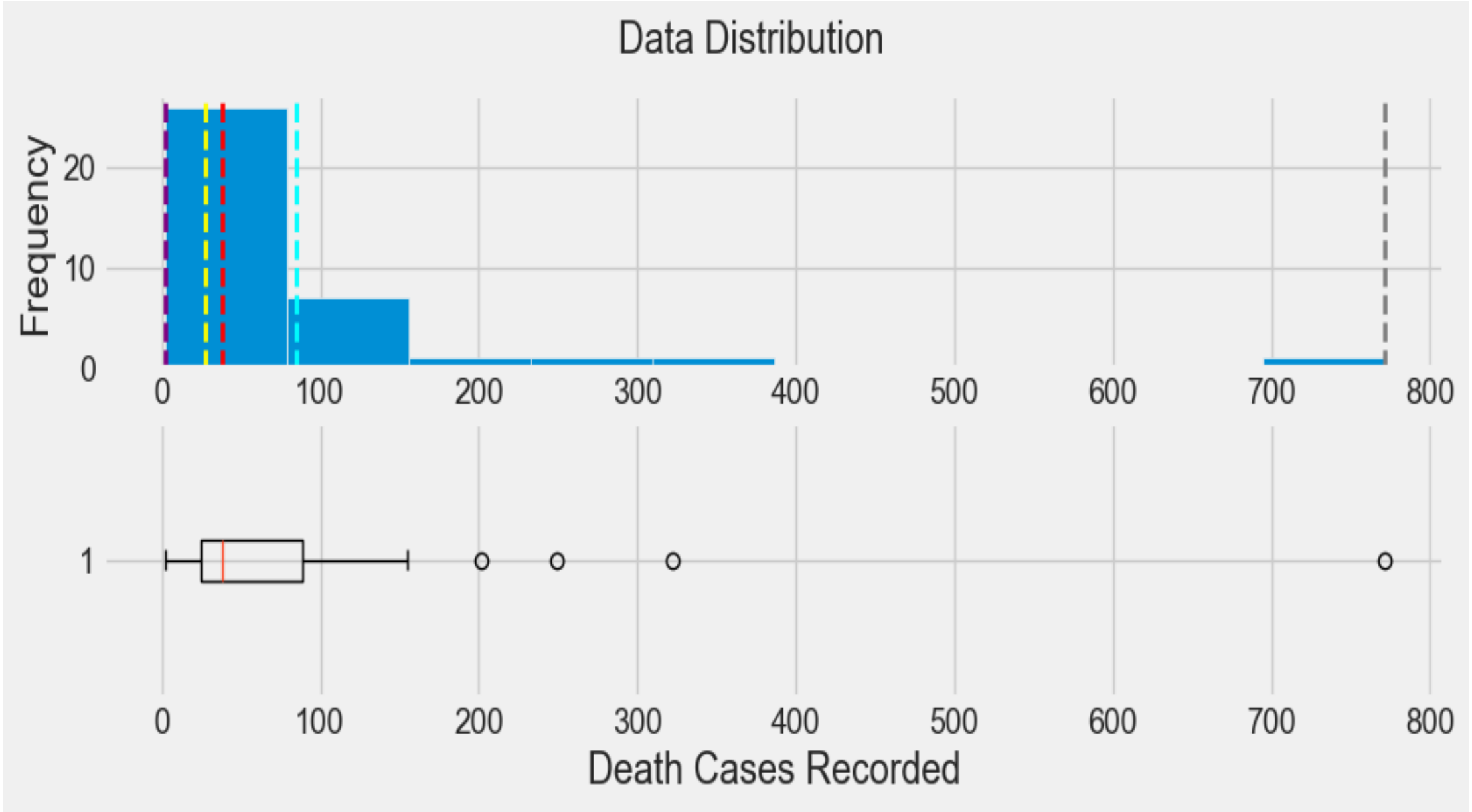
There is a high correlation coefficient between:

- Death and Confirmed cases, correlation coefficient is 0.94
- Recovered cases and Confirmed cases, correlation coefficient is 1
- Recovered cases and Death cases, correlation coefficient is 0.93
- Admitted cases and Confirmed cases, correlation coefficient is 0.73
- Confirmed cases to Confirmed cases, correlation coefficient is 1
- Death cases to Death cases, correlation coefficient is 1
- Recovered cases to Recovered cases, correlation coefficient is 1
- Admitted cases and Admitted cases, correlation coefficient is 1
- Low correlation coefficient between Admitted cases and Death cases, correlation



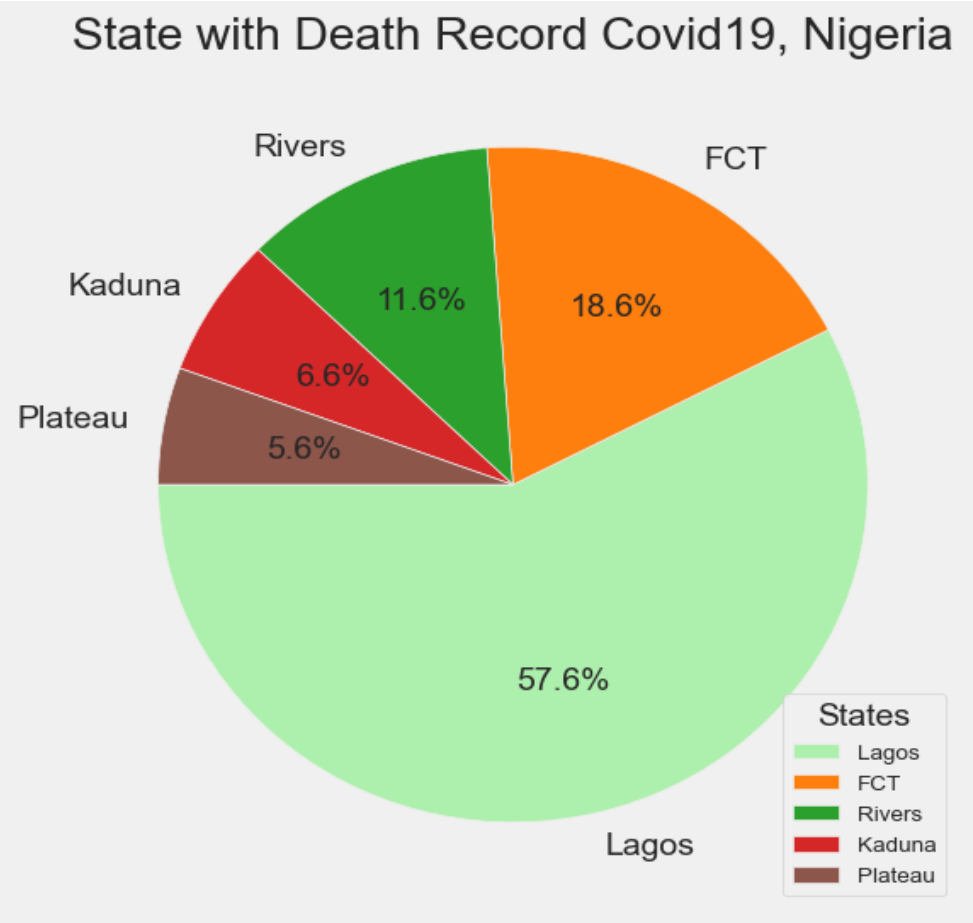
- From the bar chart above, it is shown that:
  - Lagos State has the highest death rate while Ogun and Plateau States recorded the least death rate in Nigeria.

.


Data Distribution
Frequency
20
10
0
0
100
200
300
400
500
600
700
800
1
0
100
200
300
400
500
600
700
800
Death Cases Recorded

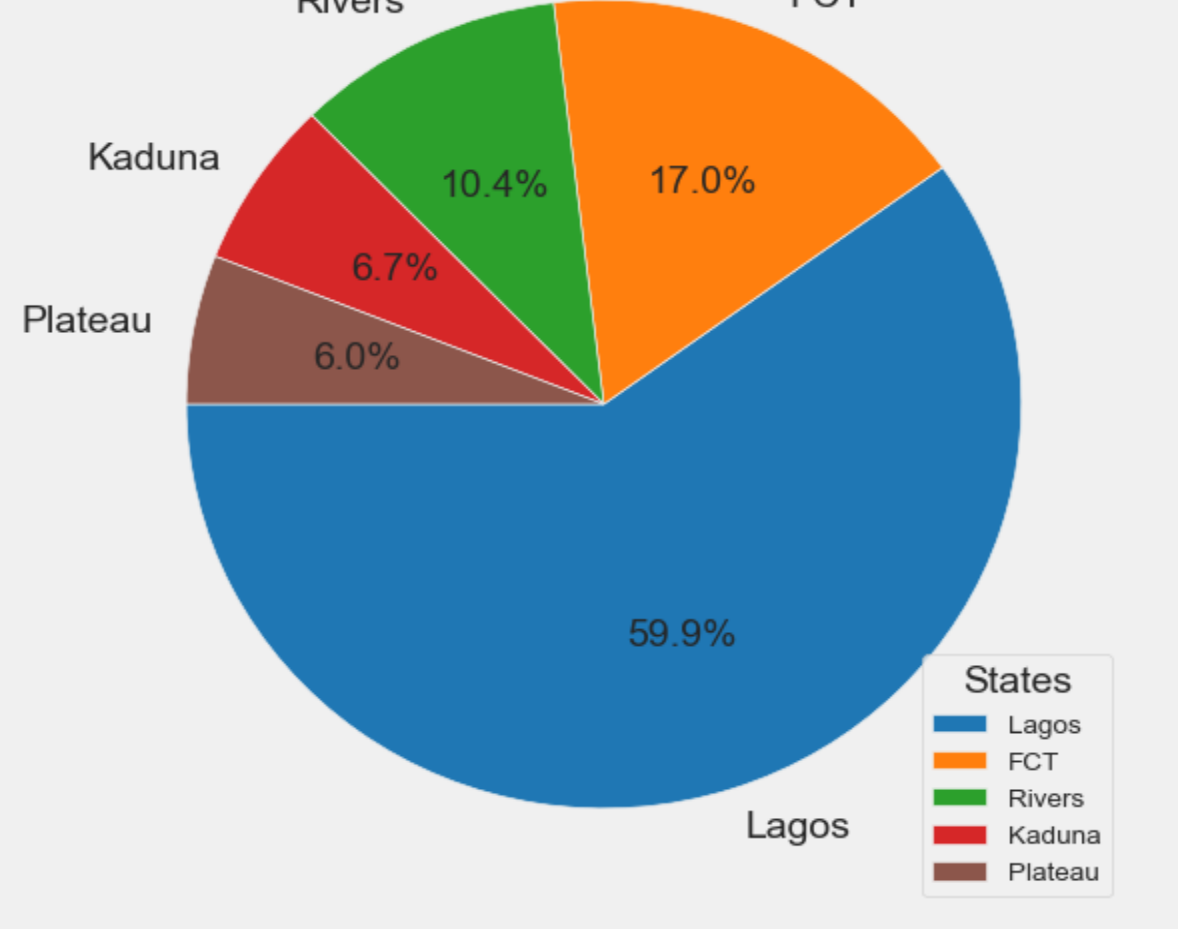
Comments:

- The yellow dash vertical line signifies mode(most occurrence) rate of death
- The dashed vertical red line signifies median(middle) rate of death
- The dashed vertical cyan line signifies mean rate of death
- The dashed vertical red line signifies median(middle) rate of death
- The dashed vertical purple line signifies least rate of death
- The dashed vertical gray line signifies highest rate of death



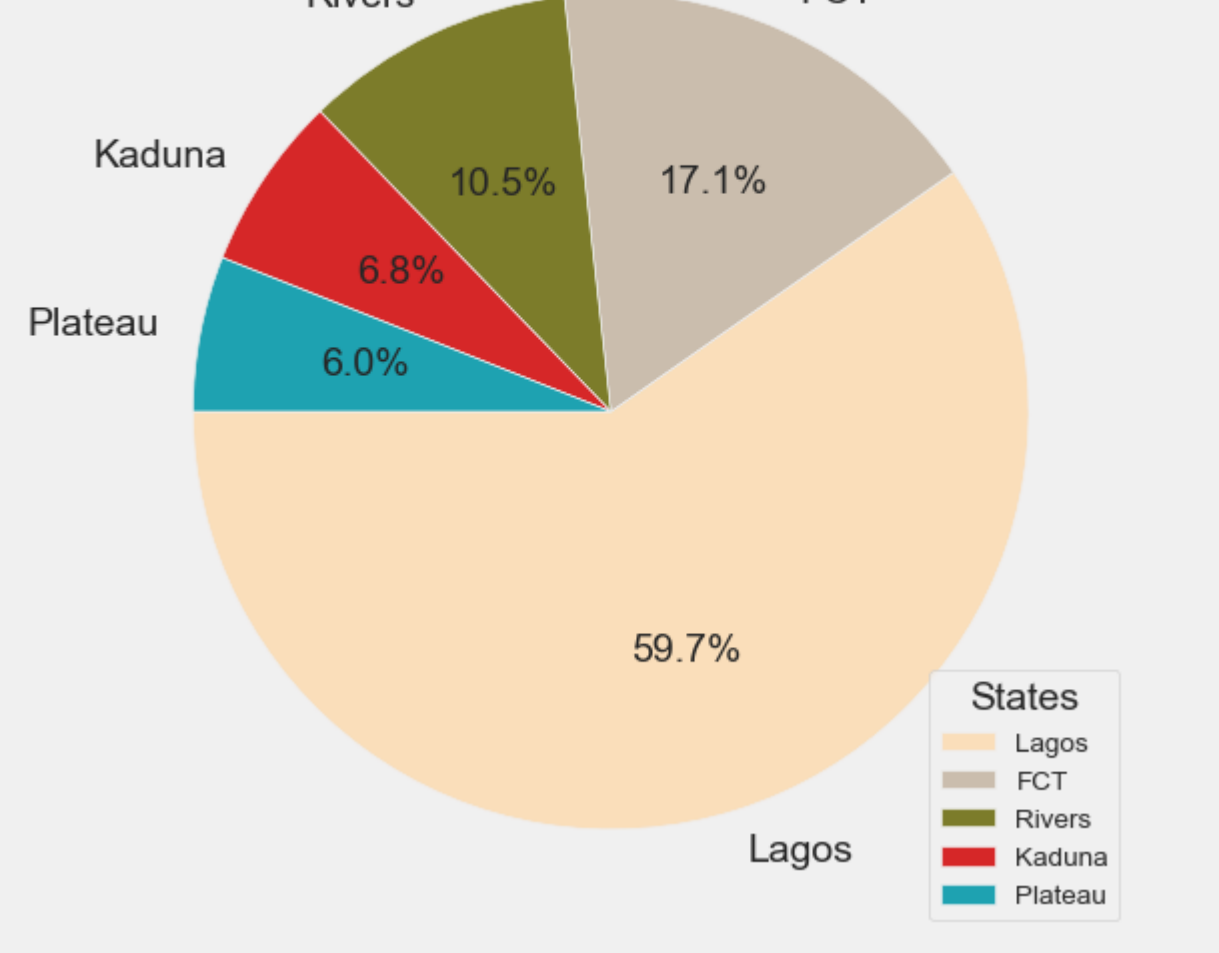
Comment:

Lagos state recorded the highest death of 57.6% follow by FCT with 18.6% while Plateau has the least record of 5.6% Therefore it is unsafe to live in Lagos.



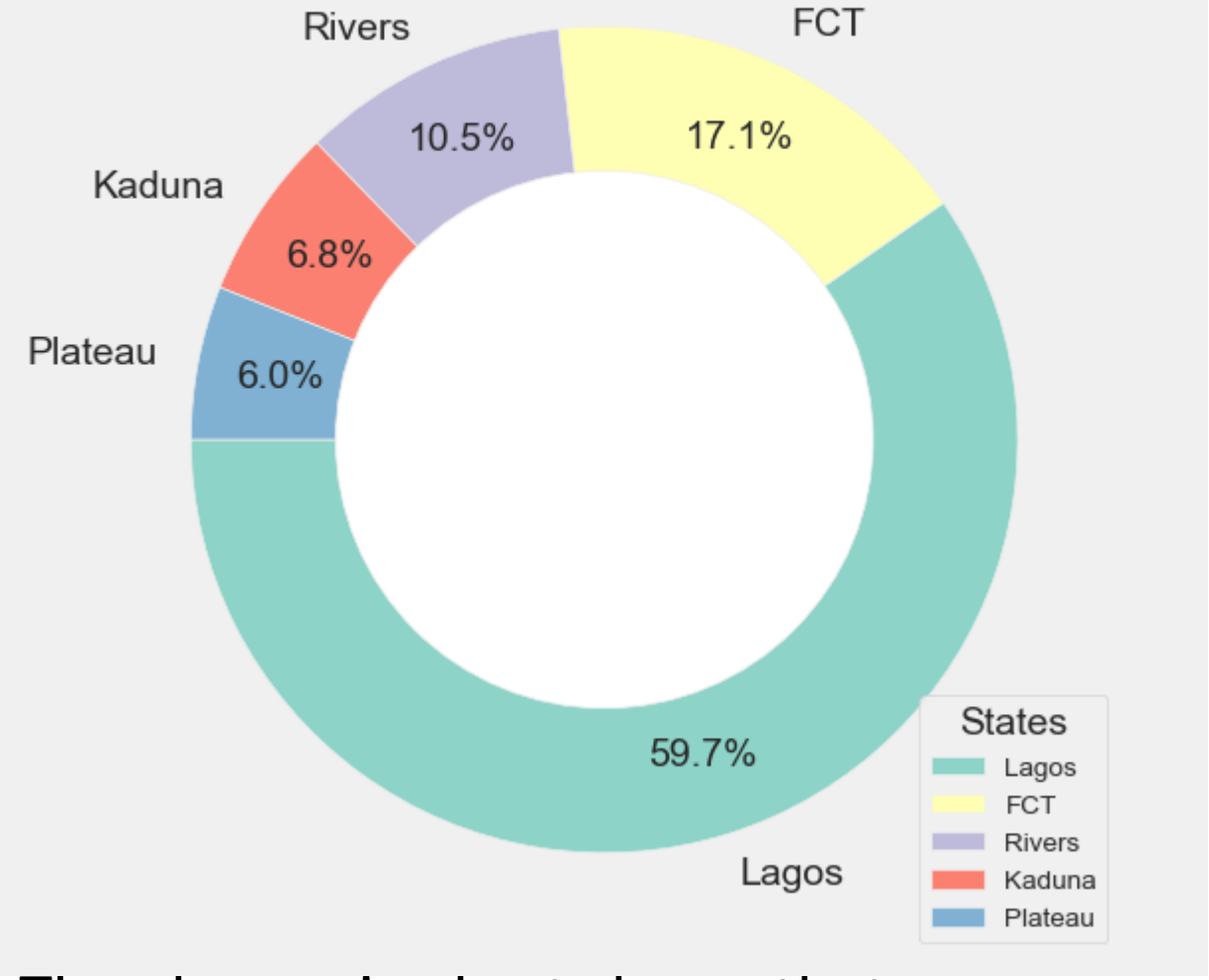
Base on the Pie Chart shown above

- Lagos state has the highest confirmed case record of 59.9% follow by FCT with 17.0% while Plateau and Kaduna states has the least confirmed case recorded with 6.0% and 6.7% respectively. Covid19 virus tends to spread more in Lagos state and FCT than any other states in Nigeria..



Comment:

- Lagos state is the state with the highest record of discharged with 59.7% follow by FCT with 17.1% while Plateau has the least record of discharged record with 6.0%..



The above pie chart shows that:

Lagos is the state with the highest record of discharged with 69.5% follow by FCT with 17.1% while Plateau has the least record of discharged.

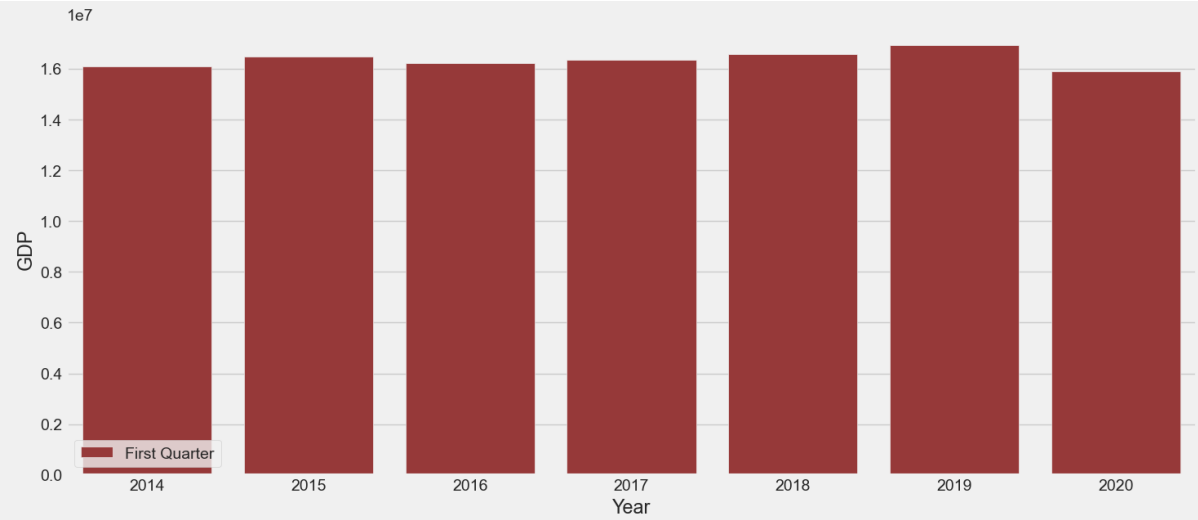
Collection of John Hopkins data

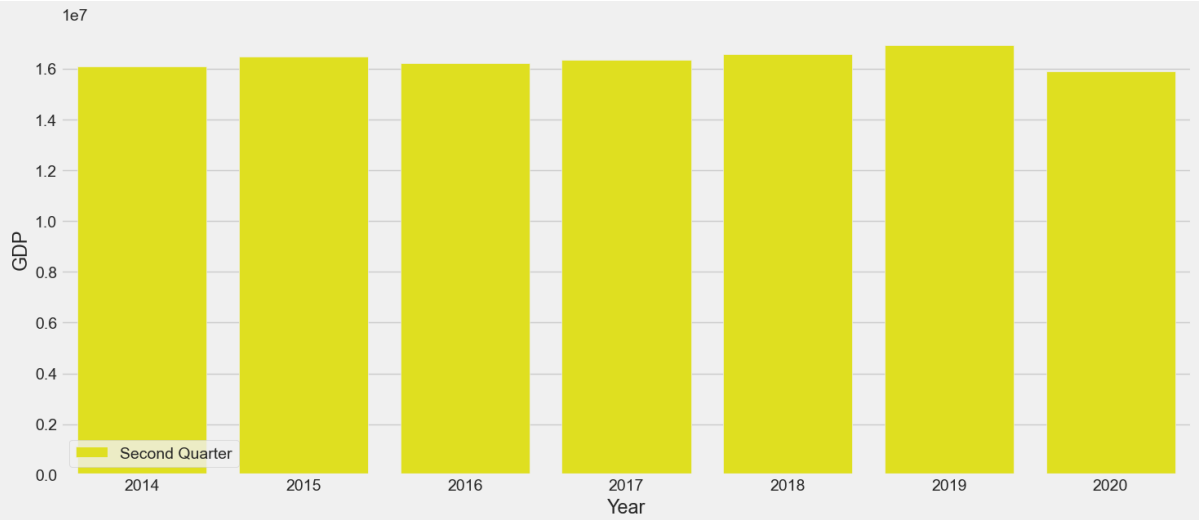
- Global Daily Confirmed Cases - Click https://github.com/CSSEGISandData/COVID-19/blob/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_confirmed_global.csv
- Global Daily Recovered Cases - Click https://github.com/CSSEGISandData/COVID-19/blob/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_recovered_global.csv
- Global Daily Death Cases - Click Here https://github.com/CSSEGISandData/COVID-19/blob/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_deaths_global.csv

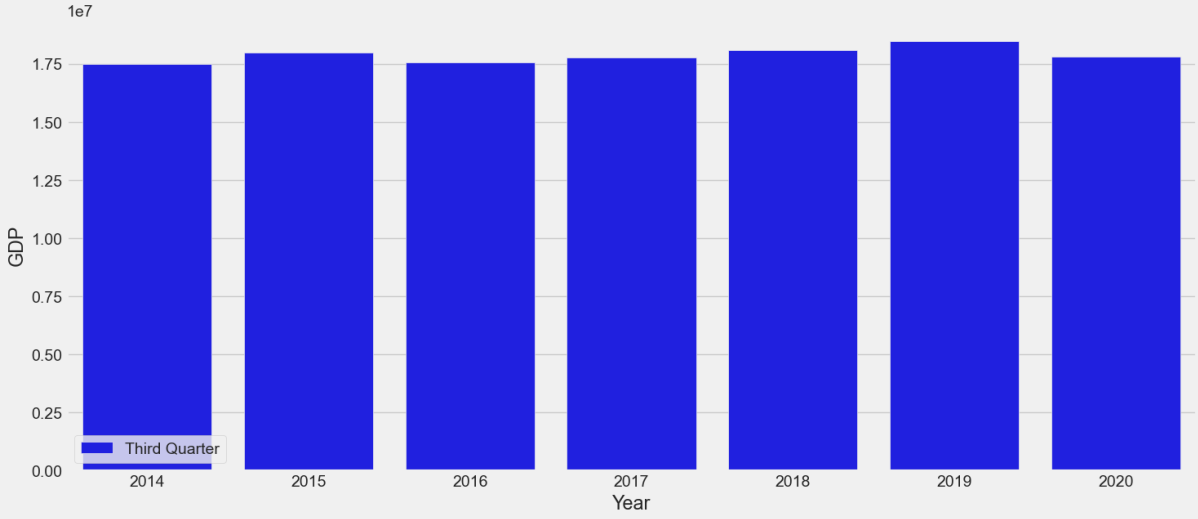
External Data

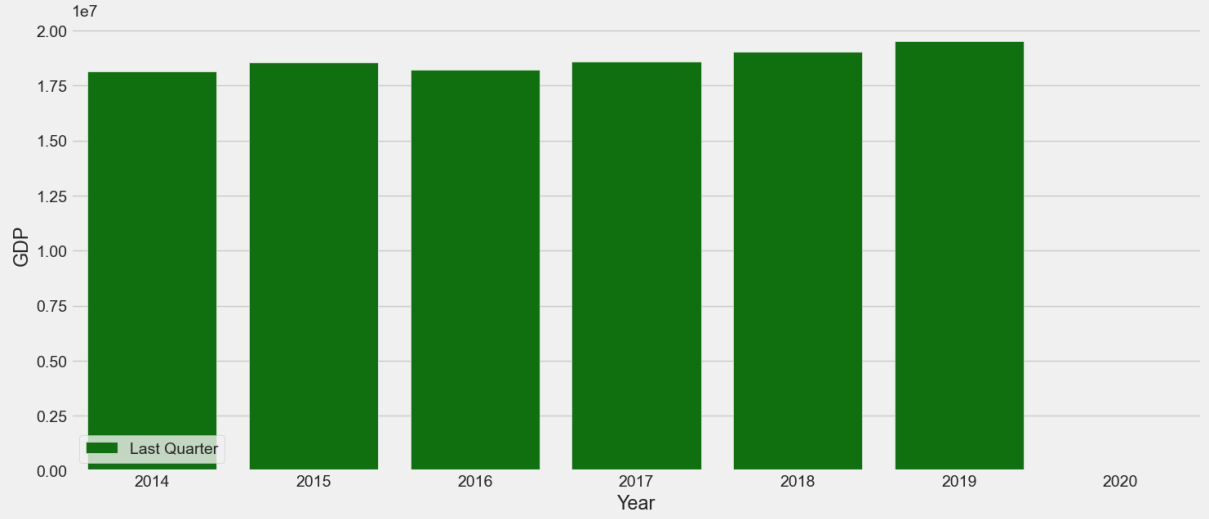
Year	Q1	Q2	Q3	Q4	
0	2014	15438679.50	16084622.31	17479127.58	18150356.45
1	2015	16050601.38	16463341.91	17976234.59	18533752.07
2	2016	15943714.54	16218542.41	17555441.69	18213537.29
3	2017	15797965.83	16334719.27	17760228.17	18598067.07
4	2018	16096654.19	16580508.07	18081342.10	19041437.59
5	2019	16434552.65	16931434.89	18494114.17	19530000.00
6	2020	16740000.00	15890000.00	17820000.00	0.00

Bar plot using the GDP values for each year and quarters









Conclusion

It is no longer a news that COVID-19 pandemic has left no part of the world untouched. The virus emerged from Wuhan, China, and spread to all other country for example Nigeria. Base on the analysis carried out so far, COVID-19 pandemic has an enormous effect on Nigeria health, education, transportation, finance, telecommunication just to mention few. It is high time for leadership, citizens and state institutions to come together for solutions to prevent the deadly virus from spreading.

The following were applied and achieved successfully during the course of the Nigeria covid-19 data analysis project:

- Data collection process from data sources.
- Data cleaning and manipulation of data.
- Develop visualization skills through the use of open-source libraries in Python.
- Generate insights from analysis

Recommendations

Based on the conclusions presented, the following recommended to anyone who work in the field of data:

- The source of data collection should be trusted and secured.
- Data collection and data cleaning should be more focused on.
- Data should be welled Visualized.
- Build a model that can be used to predict the future information etc.

Limitations

- A limited sample size or lack of reliable data such as missing data.
- Deficiencies in data measurements e.g. a questionnaire item not asked that could have been used to address a specific issue.
- Lack of reading multiple files at once without using iteration.

Reference

- https://github.com/Ustacky-dev/Nigeria-COVID-19-Data-Analysis-Using-Python/blob/main/covidnig.csv
- https://github.com/CSSEGISandData/COVID-19/blob/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_confirmed_global.csv
- https://github.com/CSSEGISandData/COVID-19/blob/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_recovered_global.csv
- https://github.com/CSSEGISandData/COVID-19/blob/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_deaths_global.csv
- https://github.com/Ustacky-dev/Nigeria-COVID-19-Data-Analysis-Using-Python/blob/main/Budget%20data.csv
- https://github.com/Ustacky-dev/Nigeria-COVID-19-Data-Analysis-Using-Python/blob/main/RealGDP.csv

Reference continues

- https://www.kaggle.com/learn/pandas
- https://pythoninstitute.org
- https://www.coursera.org/learn/python-for-applied-data-science-ai
- https://github.com/Ustacky-dev/Nigeria-COVID-19-Data-Analysis-Using-Python/blob/main/covid_external.csv
- https://www.w3schools.com/python
- Python how to program by Deitel Deitel