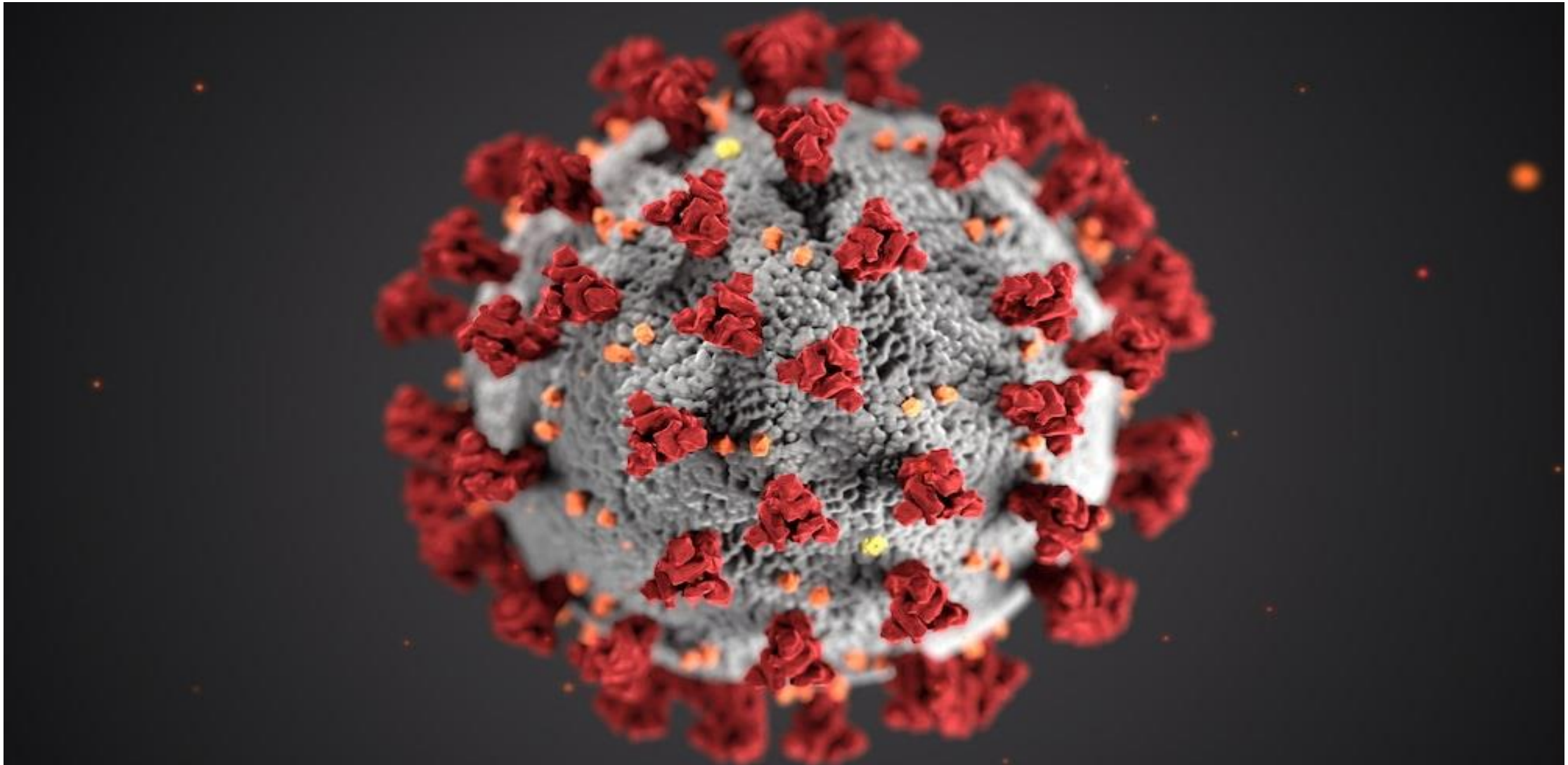
Nigeria COVID-19 Data Analysis using Python



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 Webscrapping(pandas)

```
df = pd.read_csv('covidnig.csv')
df.head()
```

	States Affected	No. of Cases (Lab Confirmed)	No. of Cases (on admission)	No. Discharged	No. of Deaths	
0	Lagos	26,708	2,435	24,037	236	
1	FCT	9,627	2,840	6,694	93	
2	Kaduna	4,504	579	3,877	48	
3	Plateau	4,262	280	3,948	34	
4	Oyo	3,788	368	3,374	46	

```
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
```

```
# Remove the comma in No. of Cases (Lab Confirmed) column
df['No. of Cases (Lab Confirmed)'] = df['No. of Cases (Lab Confirmed)'].str.replace(',',‘’)
df.head()
```

	States Affected	No. of Cases (Lab Confirmed)	No. of Cases (on admission)	No. Discharged	No. of Deaths	
0	Lagos	26708	2,435	24,037	236	
1	FCT	9627	2,840	6,694	93	
2	Kaduna	4504	579	3,877	48	
3	Plateau	4262	280	3,948	34	
4	Oyo	3788	368	3,374	46	

Remove the comma in No. of Cases (on admission) column

df['No. of Cases (on admission)']=df['No. of Cases (on admission)'].str.replace(',',‘’)

df.head()

	States Affected	No. of Cases (Lab Confirmed)	No. of Cases (on admission)	No. Discharged	No. of Deaths	
0	Lagos	26708	2435	24,037	236	
1	FCT	9627	2840	6,694	93	
2	Kaduna	4504	579	3,877	48	
3	Plateau	4262	280	3,948	34	
4	Oyo	3788	368	3,374	46	

Remove the comma in No. of Cases (on admission) column

df['No. Discharged']=df['No. Discharged'].str.replace(',',‘’)

df.head()

	States Affected	No. of Cases (Lab Confirmed)	No. of Cases (on admission)	No. Discharged	No. of Deaths	
0	Lagos	26708	2435	24037	236	
1	FCT	9627	2840	6694	93	
2	Kaduna	4504	579	3877	48	
3	Plateau	4262	280	3948	34	
4	Oyo	3788	368	3374	46	

Convert No. of Cases (on admission) object to integer

```
df['No. of Cases (on admission)'] = df['No. of Cases (on admission)'].astype(str).astype(int)
```

Convert No. Discharged object to integer

```
df['No. Discharged'] = df['No. Discharged'].astype(str).astype(int)
df.info()
```

```
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     int32
 2   No. of Cases (on admission)   37 non-null     int32
 3   No. Discharged            37 non-null     int32
 4   No. of Deaths             37 non-null     int64
dtypes: int32(3), int64(1), object(1)
memory usage: 1.1+ KB
```

df.head(7)

	States Affected	No. of Cases (Lab Confirmed)	No. of Cases (on admission)	No. Discharged	No. of Deaths	
0	Lagos	26708	2435	24037	236	
1	FCT	9627	2840	6694	93	
2	Kaduna	4504	579	3877	48	
3	Plateau	4262	280	3948	34	
4	Oyo	3788	368	3374	46	
5	Rivers	3279	232	2987	60	
6	Edo	2768	52	2603	113	

Check the number of rows and columns present in the data using the shape attribute.

df.shape

(37, 5)

There are 37 rows and 5 columns.

```
#Rename columns
df.rename(columns={'States Affected': 'States',
                   'No. of Cases (Lab Confirmed)': 'Confirmed cases',
                   'No. of Cases (on admission)': 'Admitted cases',
                   'No. Discharged':'Recovered cases',
                   'No. of Deaths':'Death'}, inplace= True)
df.head()
```

	States	Confirmed cases	Admitted cases	Recovered cases	Death
0	Lagos	26708	2435	24037	236
1	FCT	9627	2840	6694	93
2	Kaduna	4504	579	3877	48
3	Plateau	4262	280	3948	34
4	Oyo	3788	368	3374	46

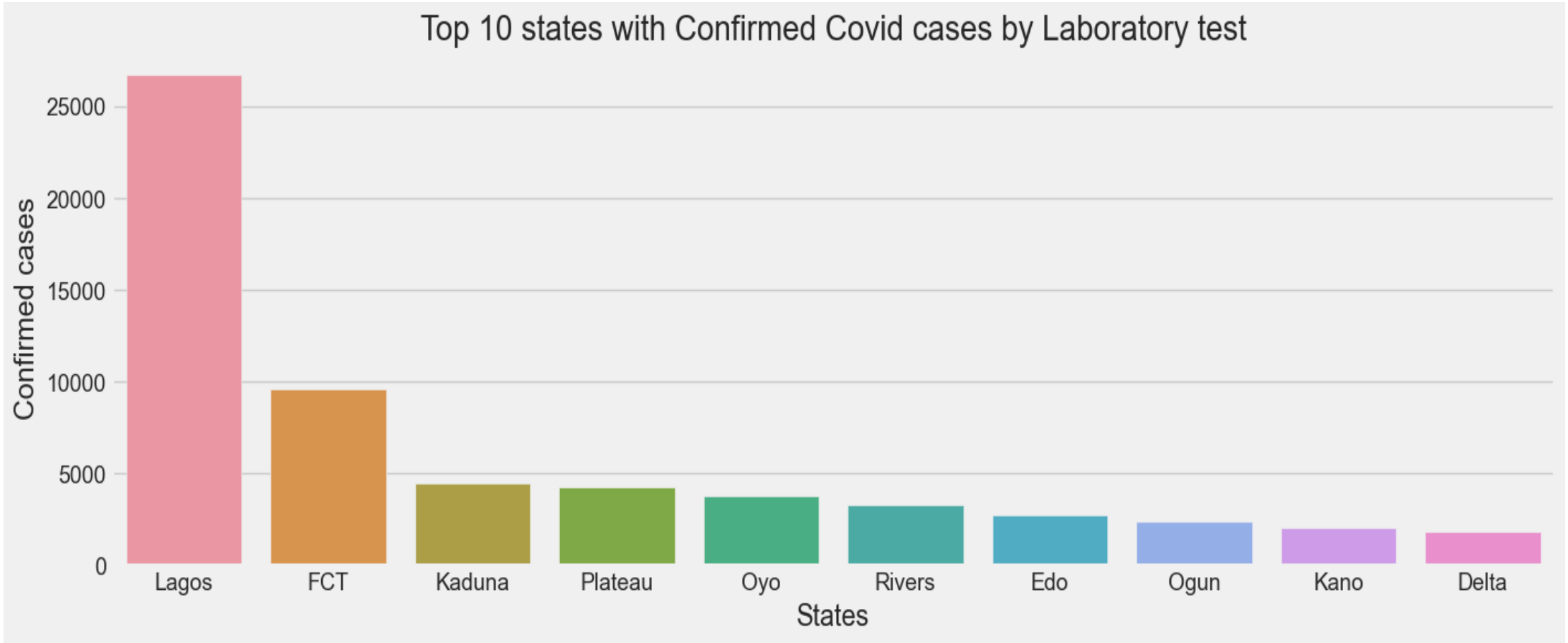
#Use describe function to generate the statistical summary of the dataframe

df.describe()

	Confirmed cases	Admitted cases	Recovered cases	Death	
count	37.000000	37.000000	37.000000	37.000000	
mean	2119.837838	240.810811	1846.027027	33.000000	
std	4537.417740	595.255773	4009.464785	41.797794	
min	5.000000	0.000000	3.000000	2.000000	
25%	381.000000	25.000000	300.000000	11.000000	
50%	897.000000	57.000000	775.000000	21.000000	
75%	1843.000000	183.000000	1737.000000	36.000000	

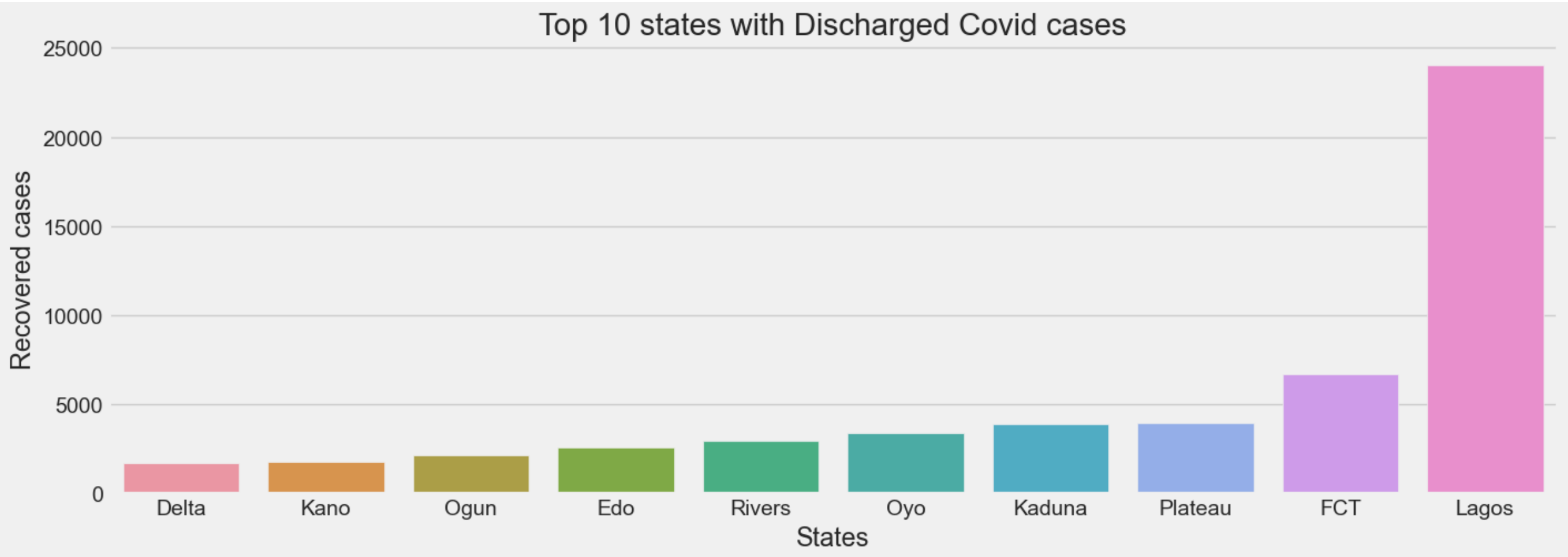
From the above result:

- The least death recorded is 2 while the highest death recorded is 236 and the average is 33
- The highest Confirmed cases is 26708 and the least is 5 while the average corfirmed case is approximately 2120
- The highest admitted cases recorded is 2840 and the average recorded is 241 approximately
- Admitted cases took place everyday
- Higheset number of patients recover is 24037
- Average number of covid19 deaths across all states in Nigeria is 33
- Highest number of deaths recorded in a state is 236 and minimum is 4.
- 75% of death cases is 36 deaths while 25% of death cases result 11.


Top 10 states with Confirmed Covid cases by Laboratory test
Confirmed cases
25000
20000
15000
10000
5000
0
Lagos
FCT
Kaduna
Plateau
Oyo
Rivers
Edo
Ogun
Kano
Delta
States

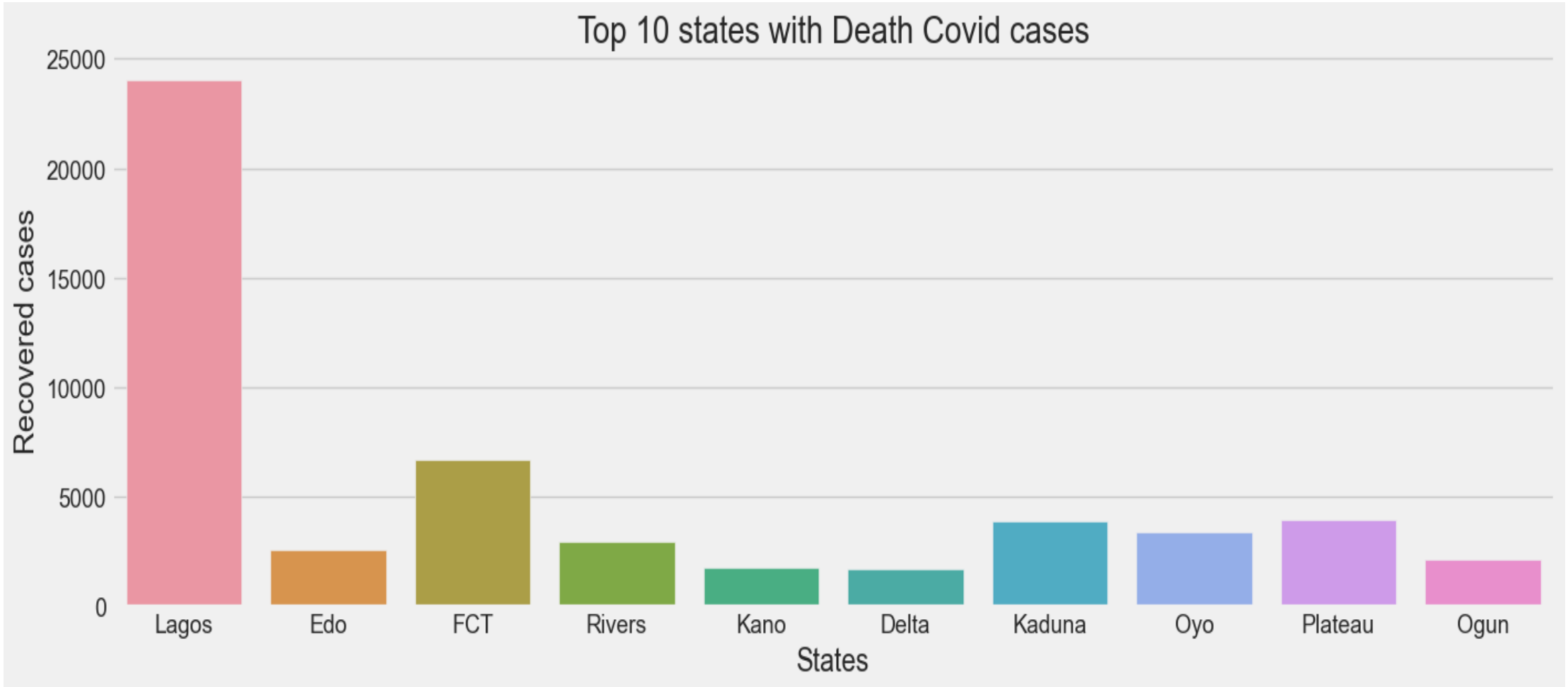
From the above result,

- Lagos State recorded highest number of confirmed cases with over 25000 while Kano, Delta and Ogun recorded least with less than 3000 patients.


Top 10 states with Discharged Covid cases
25000
20000
15000
10000
5000
0
Recovered cases
Delta
Kano
Ogun
Edo
Rivers
Oyo
Kaduna
Plateau
FCT
Lagos
States

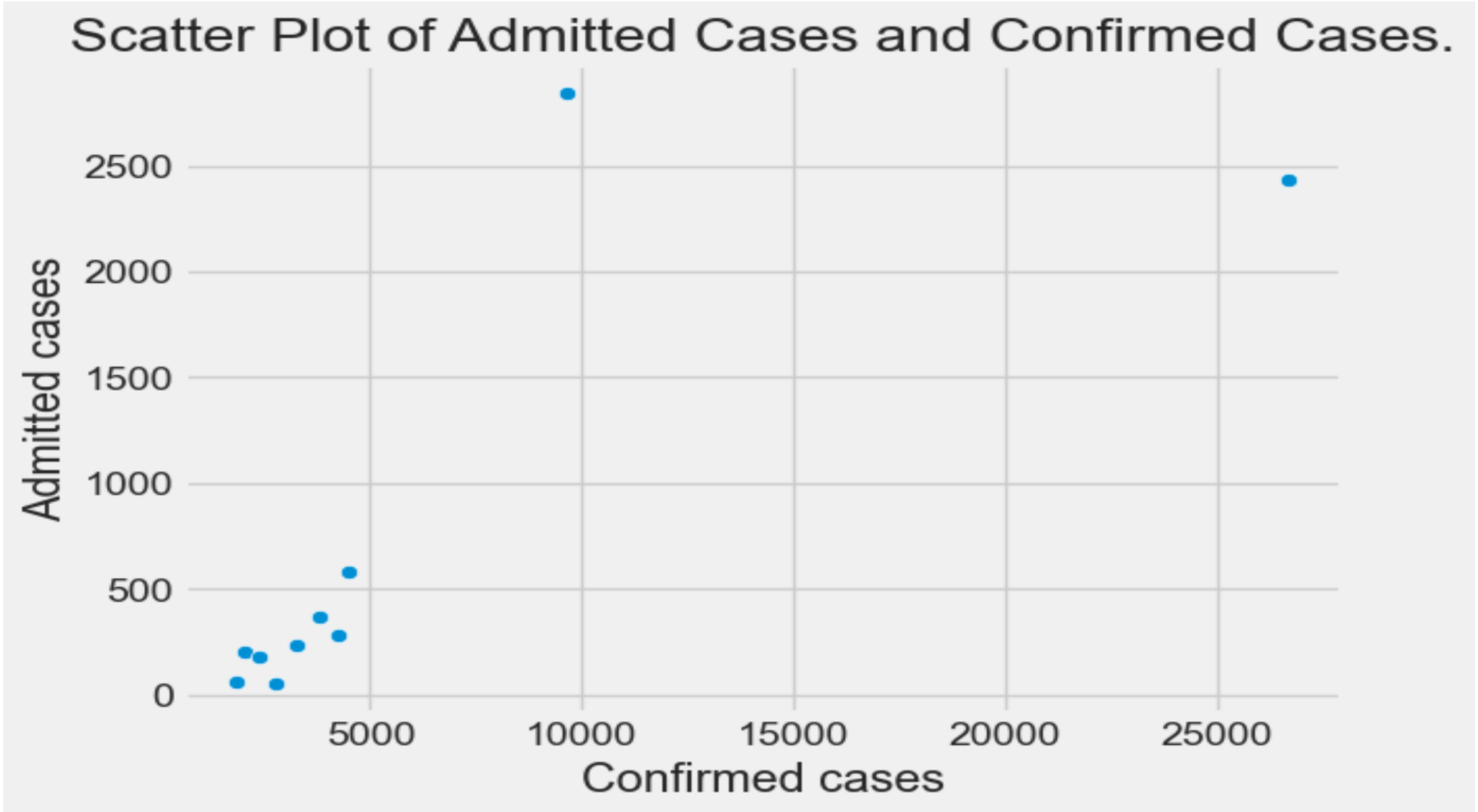
From the above result,

Lagos State recorded highest number of discharged patients with over 20000 while Delta, Kano and Ogun recorded the least with less than 3000 patients.


Top 10 states with Death Covid cases
25000
20000
15000
10000
5000
0
Recovered cases
Lagos
Edo
FCT
Rivers
Kano
Delta
Kaduna
Oyo
Plateau
Ogun
States

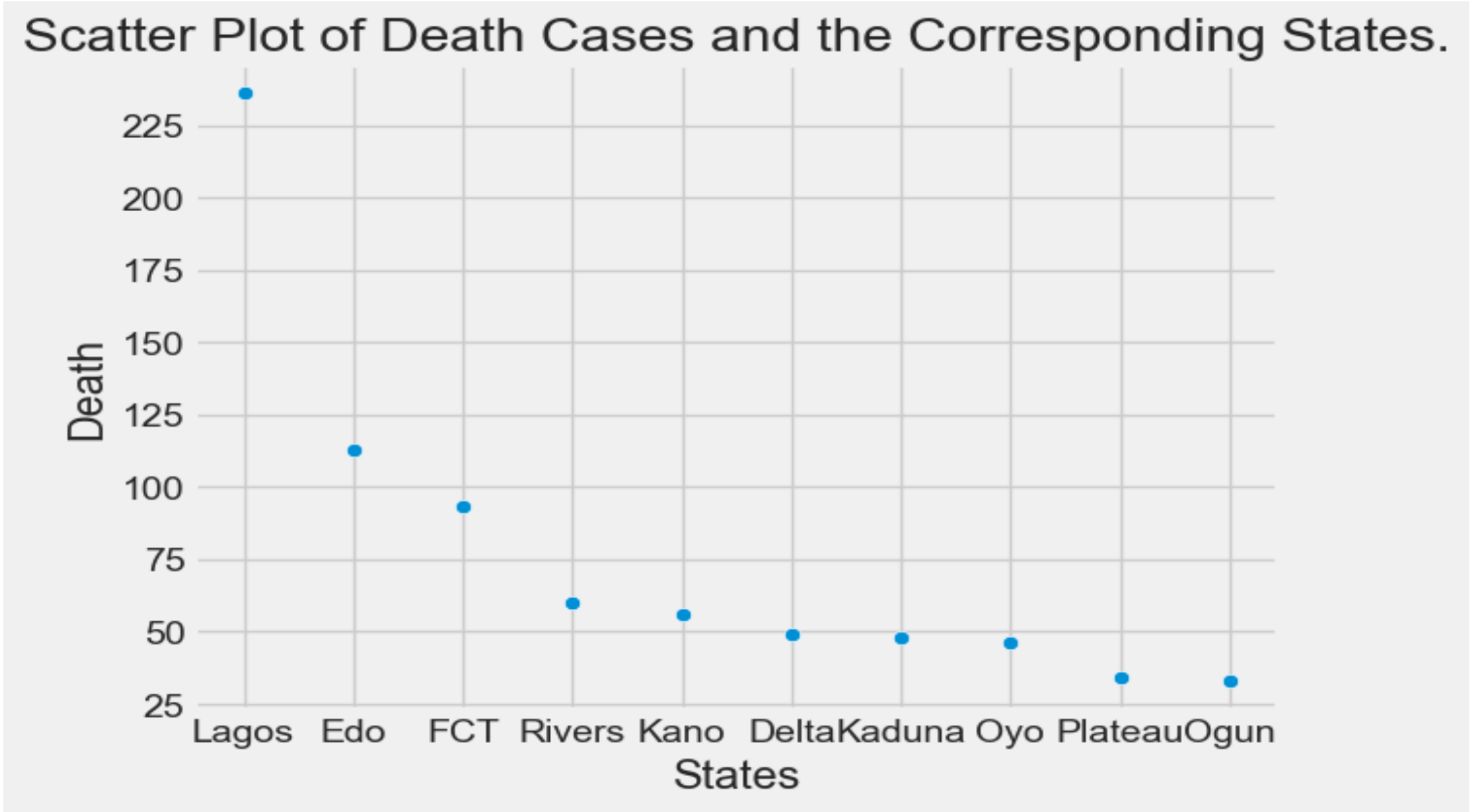
From the bar chart result above shows that:

- Lagos has the highest number of discharged patients with close to 25000 while the Delta and Kano recorded the least discharged patients with less than 3000.
- Abuja is the second highest number of discharged patients with more than 5000 but less than 10000.

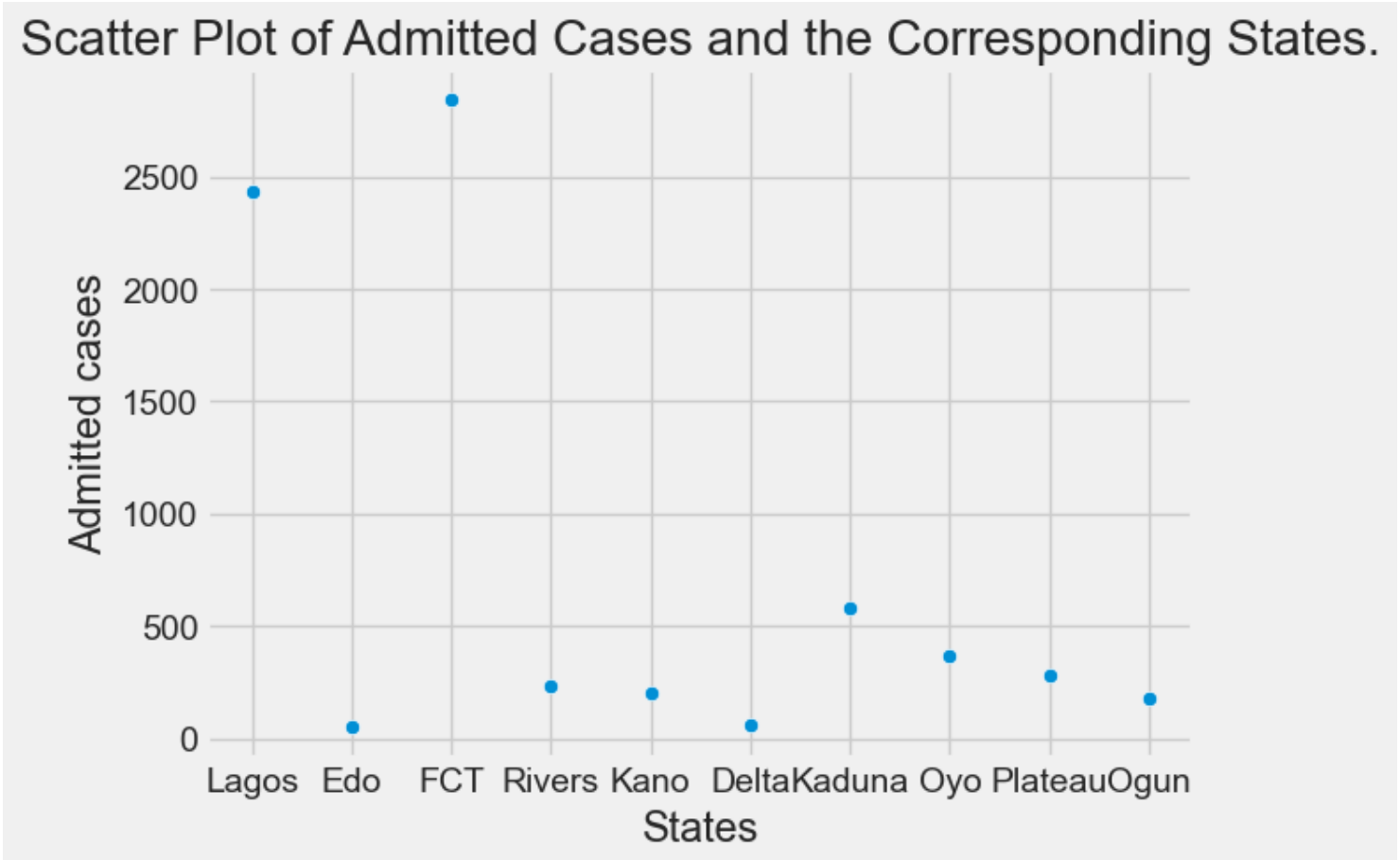

Scatter Plot of Admitted Cases and Confirmed Cases.
Admitted cases
2500
2000
1500
1000
500
0
5000
10000
15000
20000
25000
Confirmed cases

From the result above, it can be concluded that:

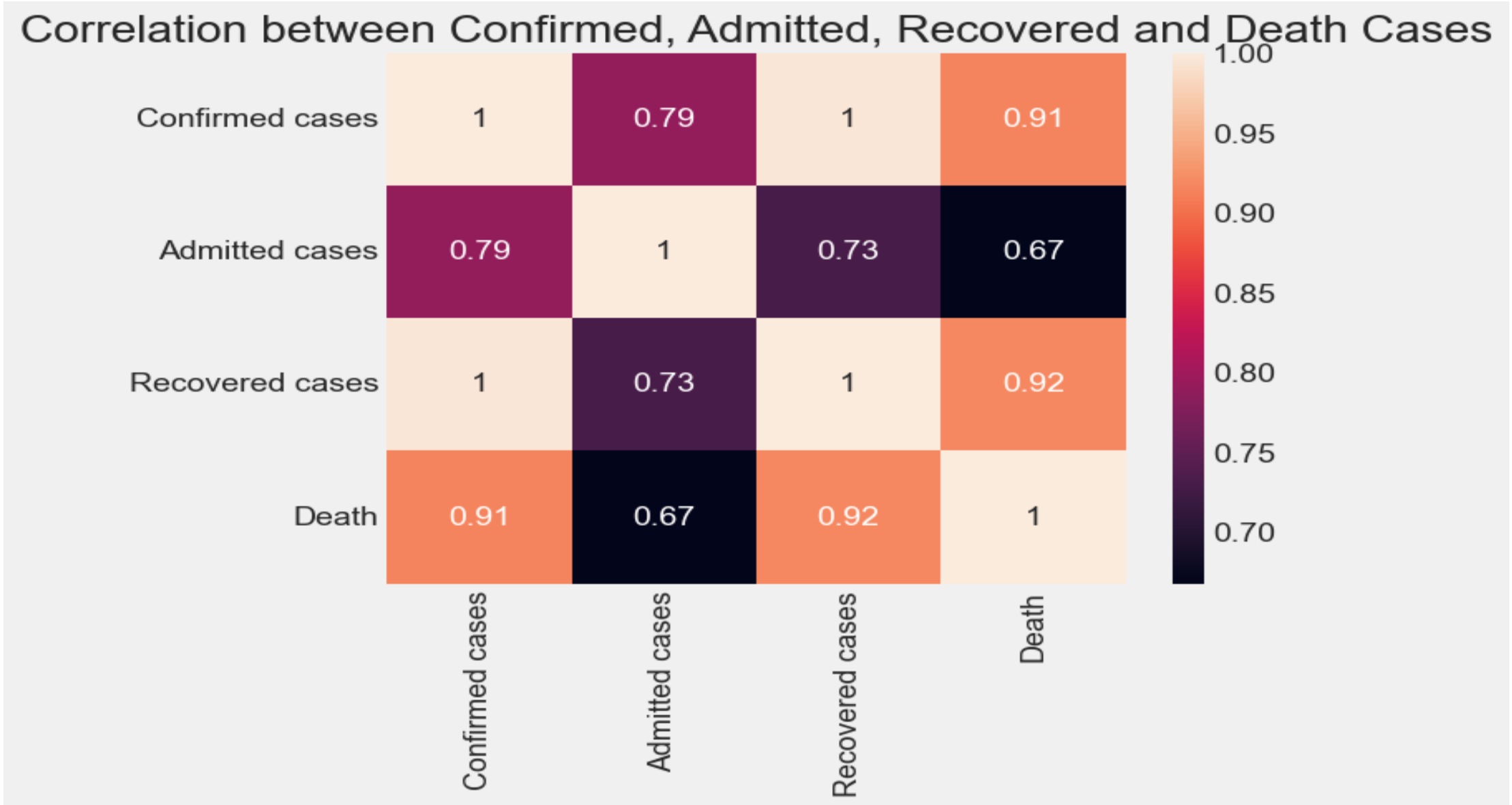
The number of admitted cases raise to over 2500 while Number of Confirmed Cases raise to over 25,000.


Scatter Plot of Death Cases and the Corresponding States.
Death
225
200
175
150
125
100
75
50
25
Lagos
Edo
FCT
Rivers
Kano
Delta
Kaduna
Oyo
Plateau
Ogun
States

From the graph above result, it is shown that Lagos highest number of death record of more 225 follow by Edo with more than 100 but less than 125 while both Plateau and Ogun recorded less record of death below 50.

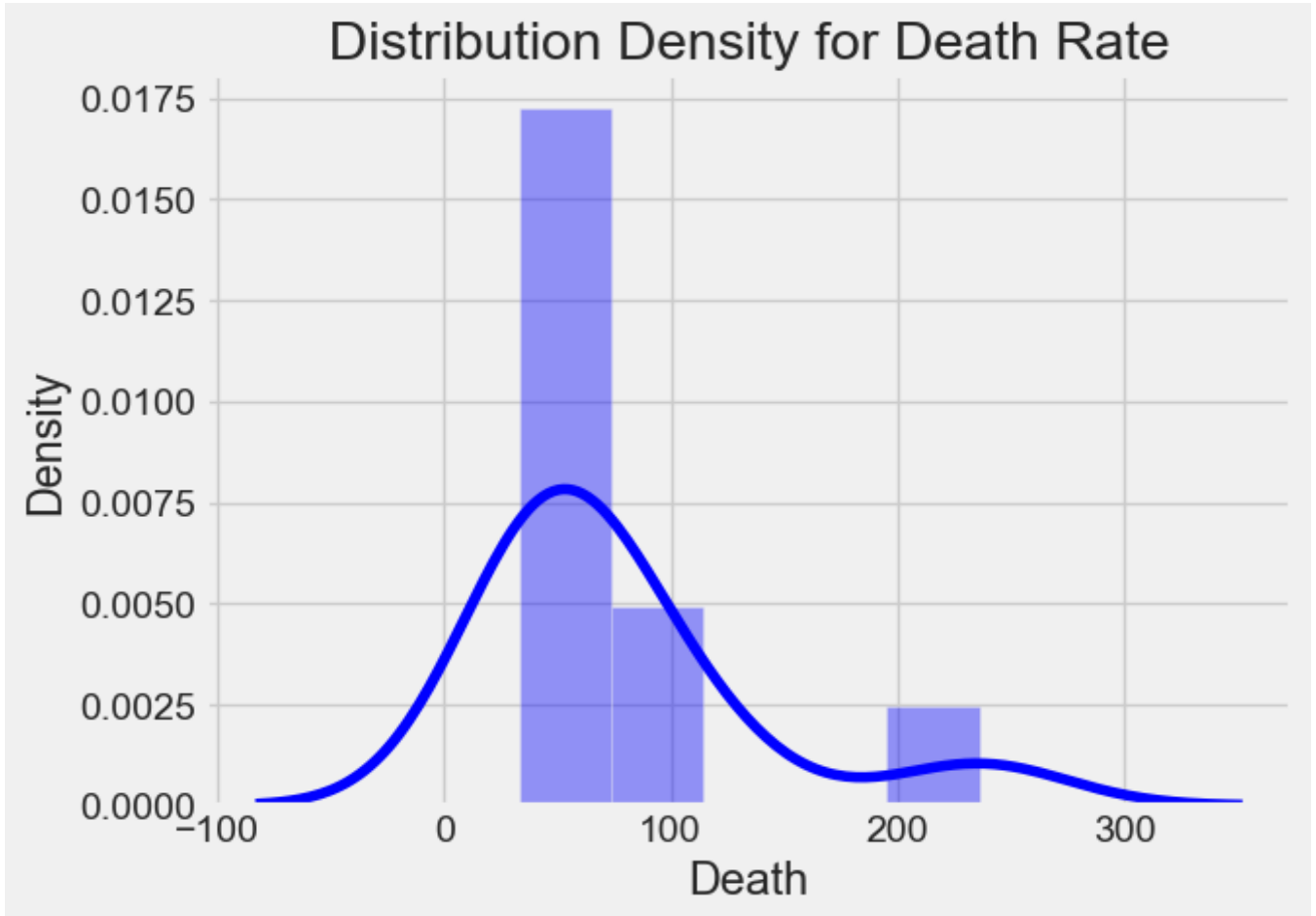

Scatter Plot of Admitted Cases and the Corresponding States.
Admitted cases
2500
2000
1500
1000
500
0
Lagos
Edo
FCT
Rivers
Kano
Delta
Kaduna
Oyo
Plateau
Ogun
States

The Scatter graph above shows that FCT recorded the highest number of admission case of more than 2500 patients while Edo and Delta had below 250 patients recorded


Correlation between Confirmed, Admitted, Recovered and Death Cases
Confirmed cases
Admitted cases
Recovered cases
Death
1
0.79
1
0.91
0.79
1
0.73
0.67
1
0.73
1
0.92
0.91
0.67
0.92
1
Confirmed cases
Admitted cases
Recovered cases
Death
1.00
0.95
0.90
0.85
0.80
0.75
0.70

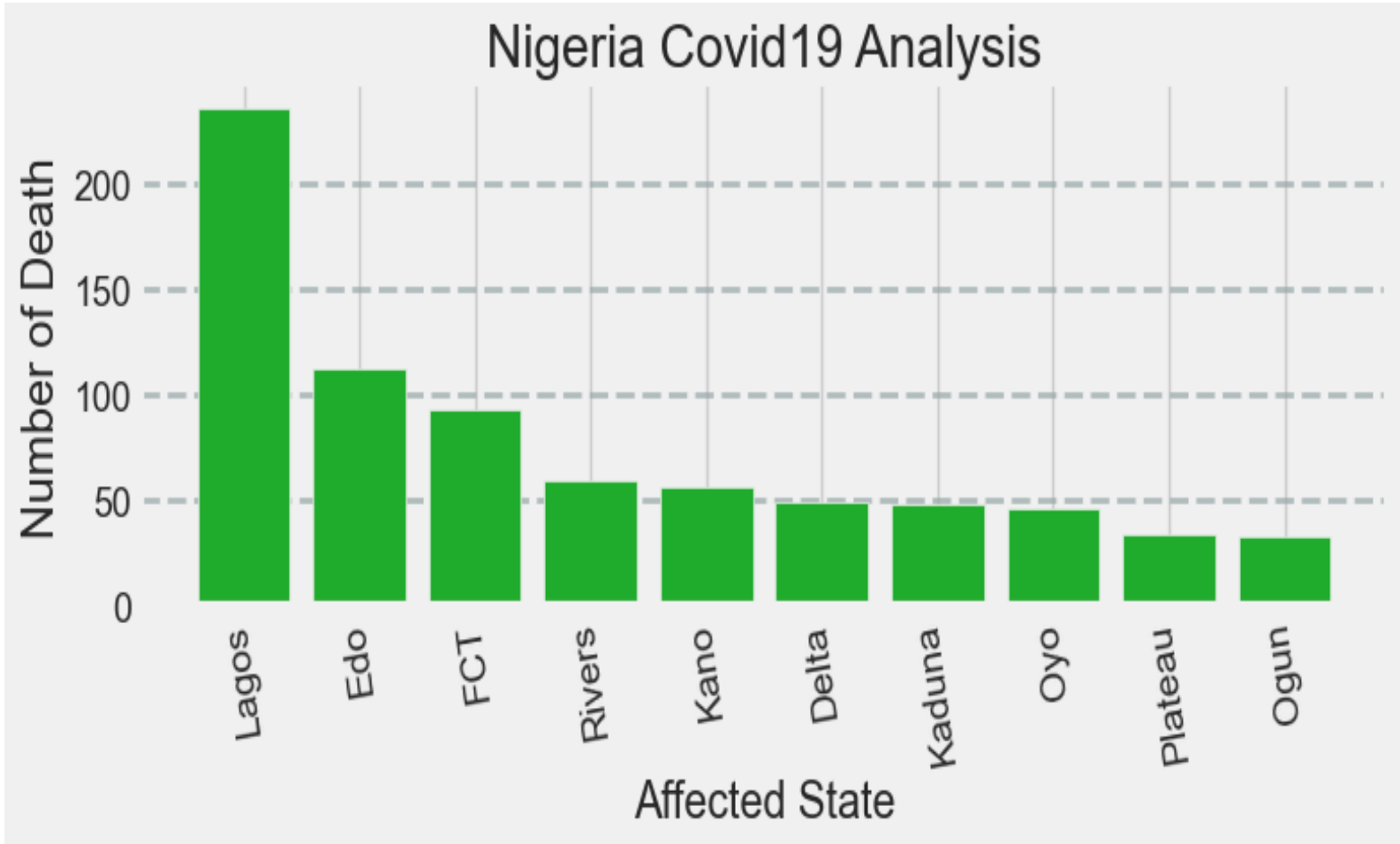
From the above result,

- There is a high correlation coefficient between:
- Death and Confirmed cases, correlation coefficient is 0.91
- Recovered cases and Confirmed cases, correlation coefficient is 1
- Recovered cases and Death cases, correlation coefficient is 0.92
- Admitted cases and Confirmed cases, correlation coefficient is 0.79
- 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 coefficient is 0.67

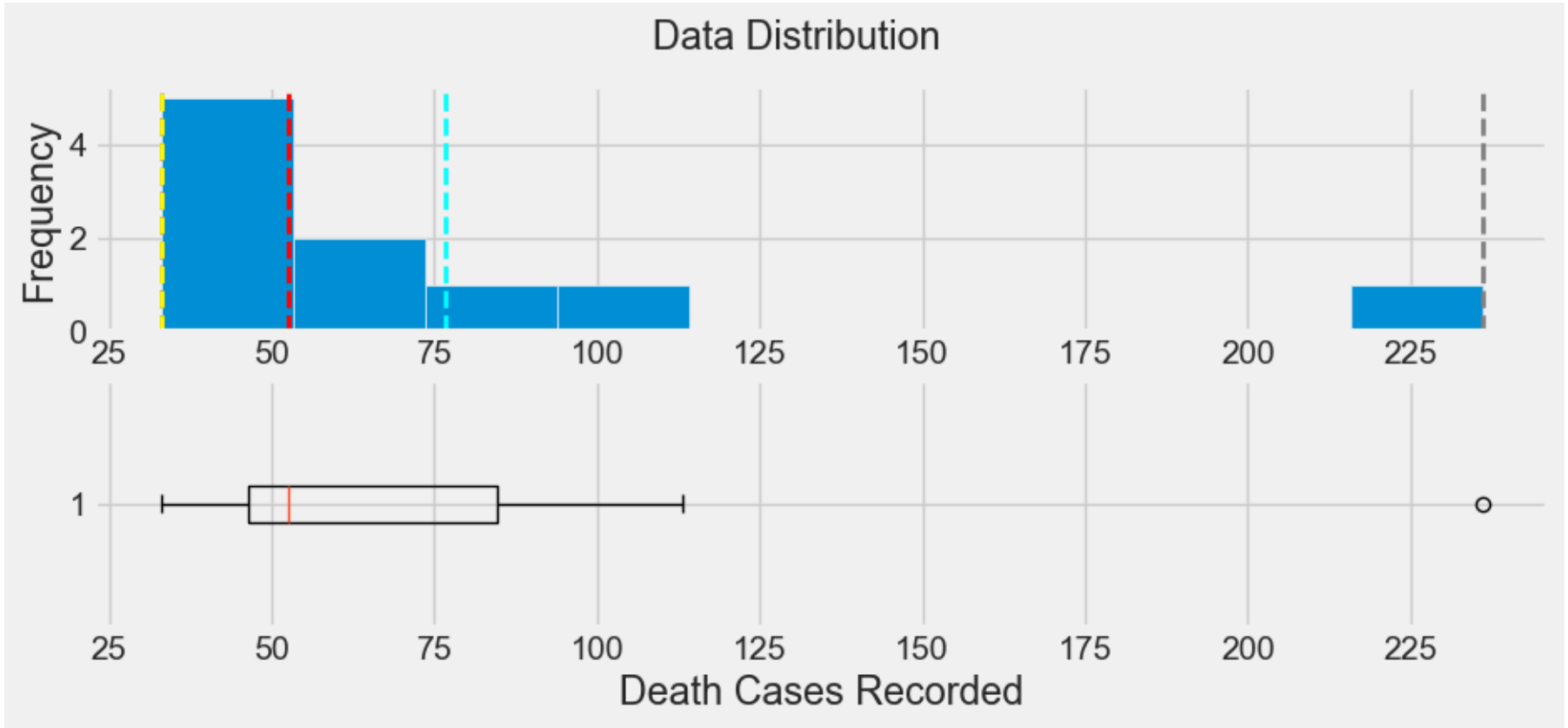

Distribution Density for Death Rate
Density
0.0175
0.0150
0.0125
0.0100
0.0075
0.0050
0.0025
0.0000
−100
0
100
200
300
Death

From the graph above,

Death Rate recorded mostly between 0 to 100.

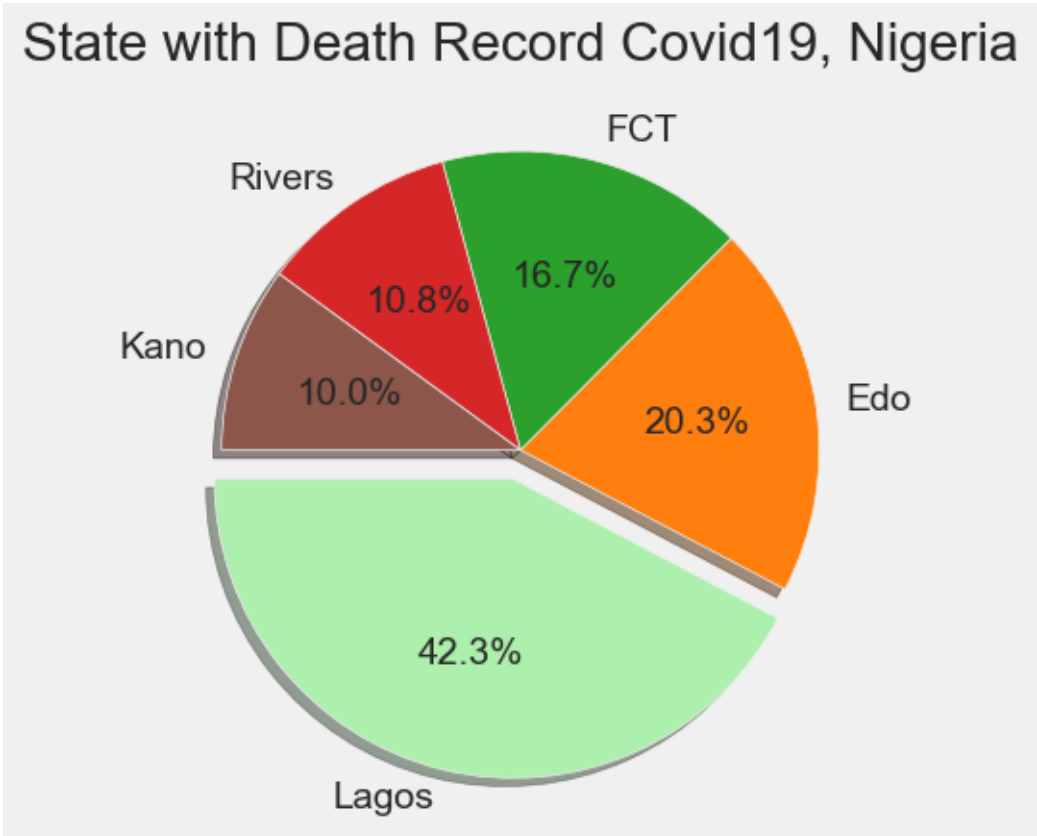

Nigeria Covid19 Analysis
Number of Death
200
150
100
50
0
Lagos
Edo
FCT
Rivers
Kano
Delta
Kaduna
Oyo
Plateau
Ogun
Affected State

The bar chart above shows that Lagos State has the highest death rate while Ogun and Plateau States recorded the least death rate in Nigeria.


Data Distribution
Frequency
0
2
4
25
50
75
100
125
150
175
200
225
1
Death Cases Recorded

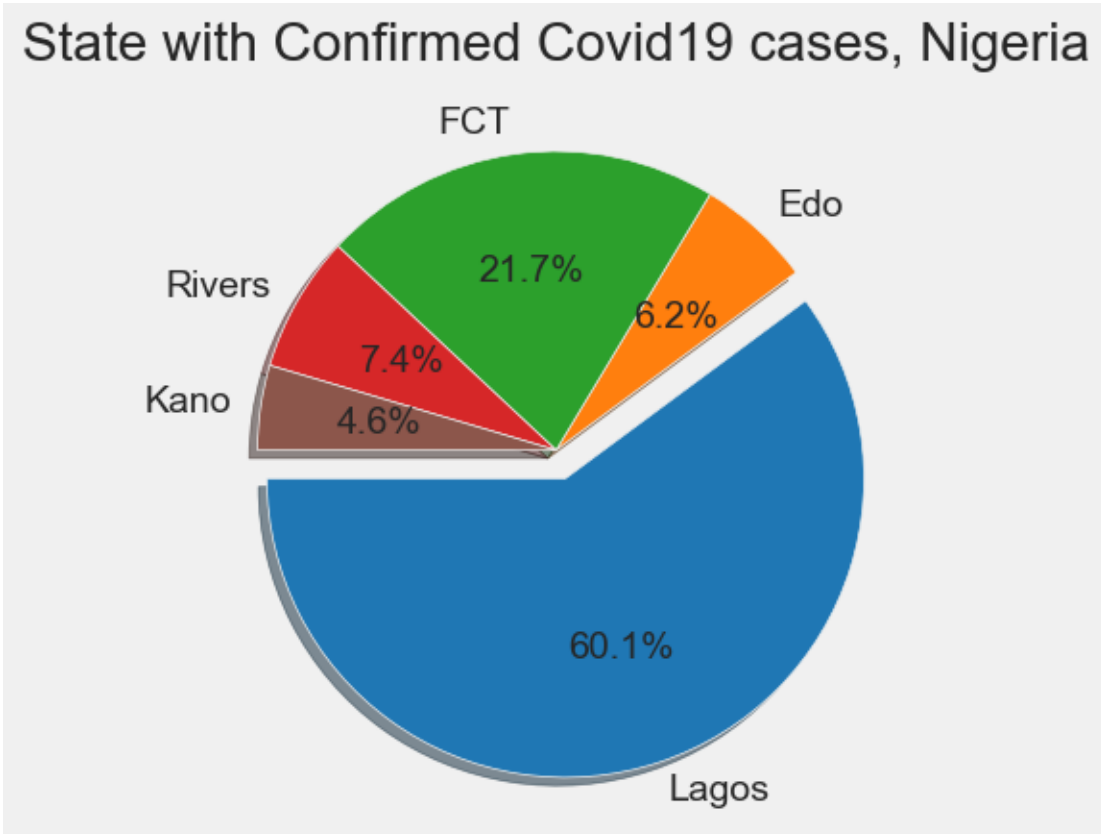
Comments:

- 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.


State with Death Record Covid19, Nigeria
FCT
Rivers
16.7%
10.8%
Kano
10.0%
Edo
20.3%
42.3%
Lagos

Comment:

- Lagos recorded the highest death of 42.3% follow by Edo with 20.3% while Kano has the least record of 10%.
- Therefore it is unsafe to live in Lagos.

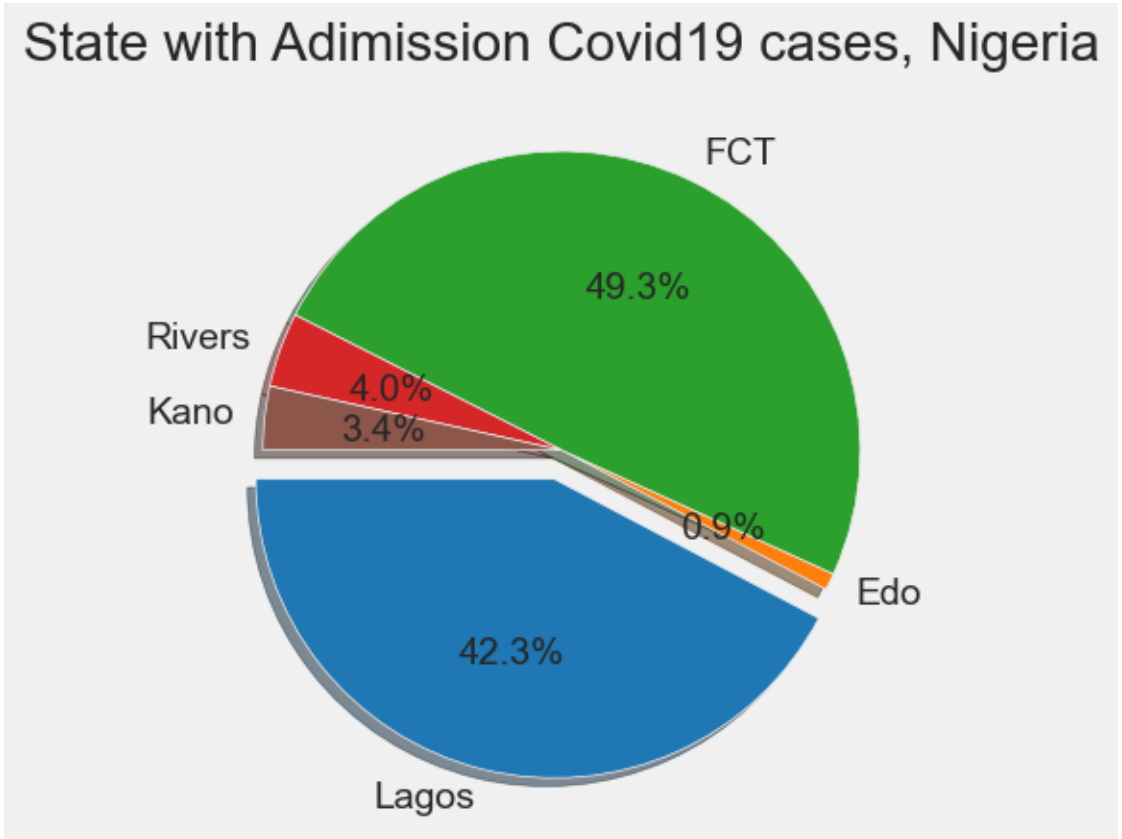

State with Confirmed Covid19 cases, Nigeria
FCT
Edo
Rivers
Kano
Lagos
21.7%
6.2%
7.4%
4.6%
60.1%

Base on the Pie Chart shown above

- Lagos has the highest confirmed case record of 60.1% follow by FCT with 21.7% while Kano has the least confirmed case record.

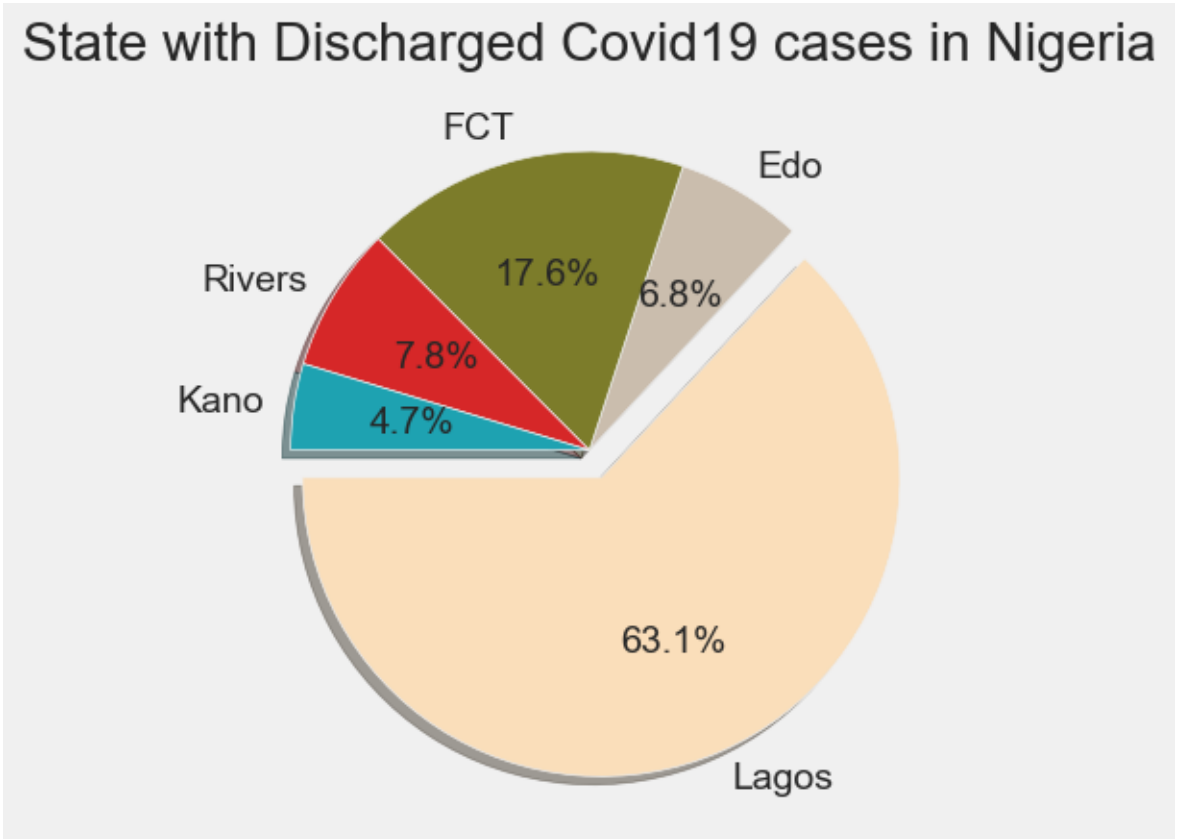
Conclusion:

- Covid19 virus tends to spread more in Lagos than any other state in Nigeria
- It is more safe to live in Kano than in Lagos.


State with Adimission Covid19 cases, Nigeria
FCT
49.3%
Rivers
4.0%
Kano
3.4%
0.9%
Edo
42.3%
Lagos

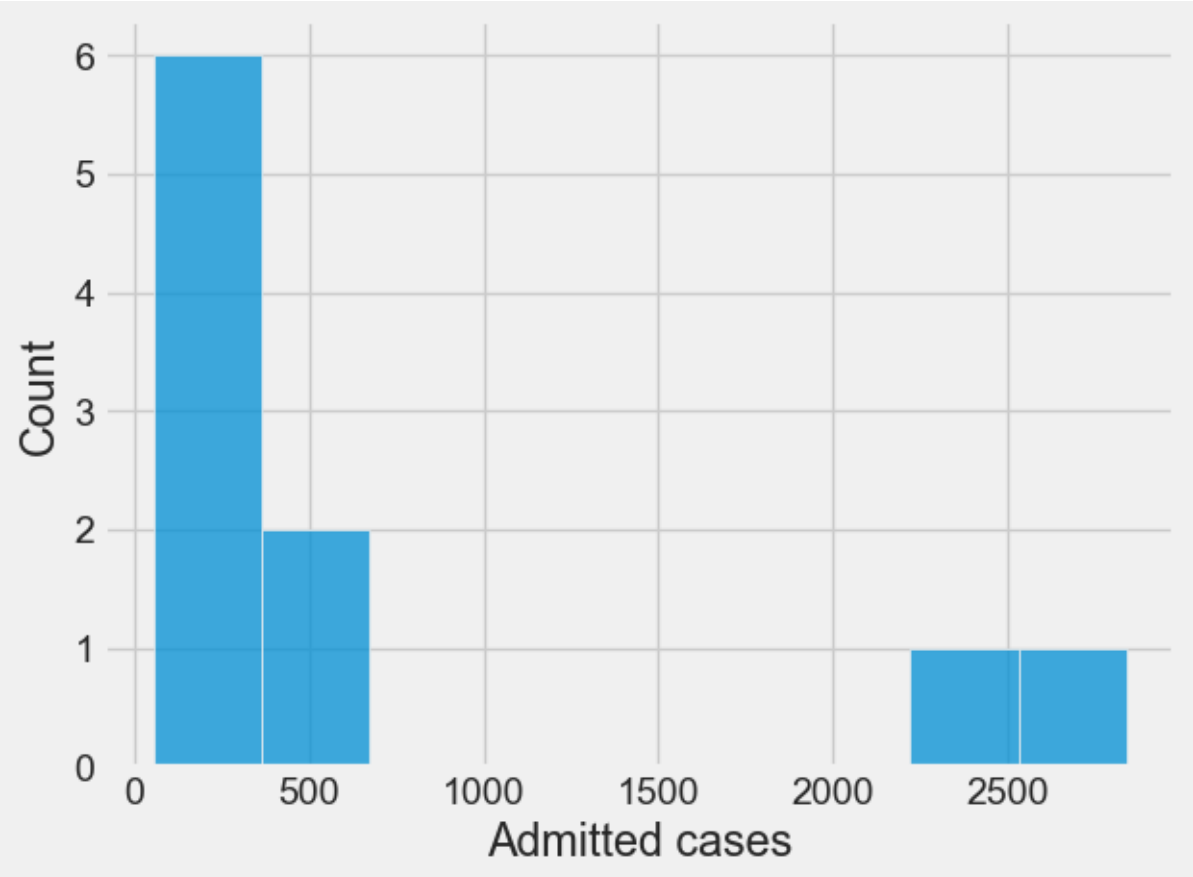
Comment:

- Abuja has the highest number of admitted case record of 49.3% follow by Lagos with 42.3%% while Edo has the minimum number of admitted case record of 0.9%.


State with Discharged Covid19 cases in Nigeria
FCT
Edo
Rivers
17.6%
6.8%
7.8%
Kano
4.7%
63.1%
Lagos

The above pie chart shows that:

Lagos is the state with the highest record of discharged with 63.1% follow by FCT with 17.6% while Kano has the least record of discharged record.


Count
6
5
4
3
2
1
0
0
500
1000
1500
2000
2500
Admitted cases

From the graph above, it can be concluded that:

- Below 500 Patients were admitted and affected by covid-19 virus mostly.
- No admitted record cases for patients between 1000 and 2000.

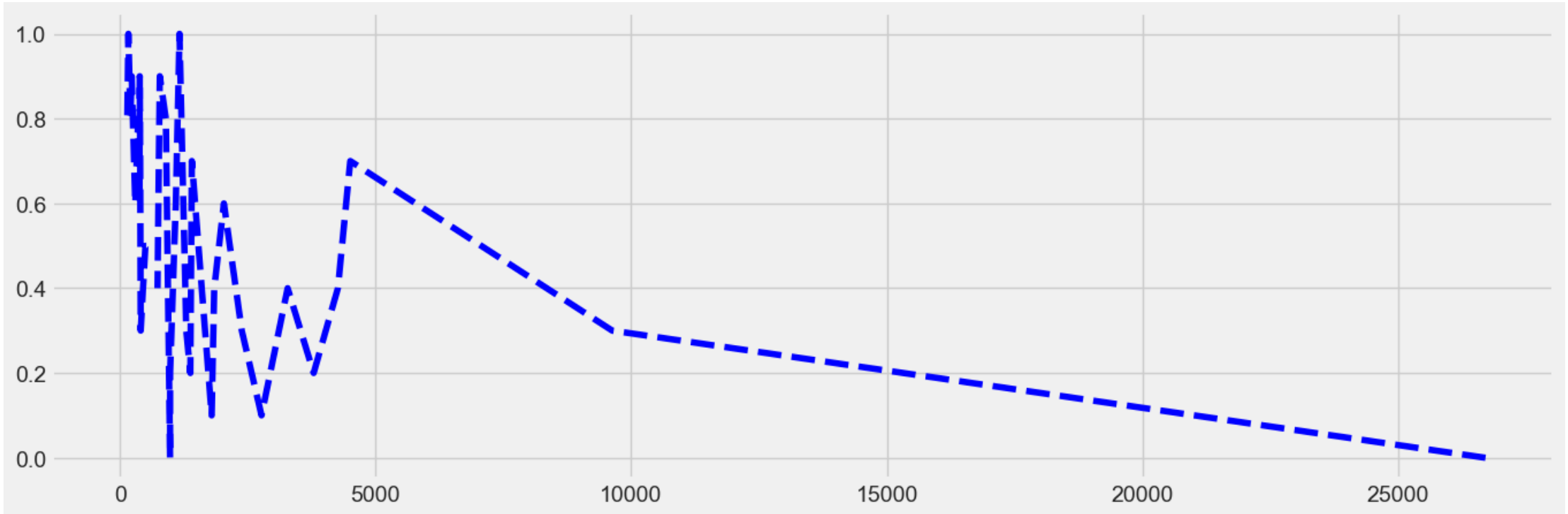
Collection of John Hopkins data

- 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

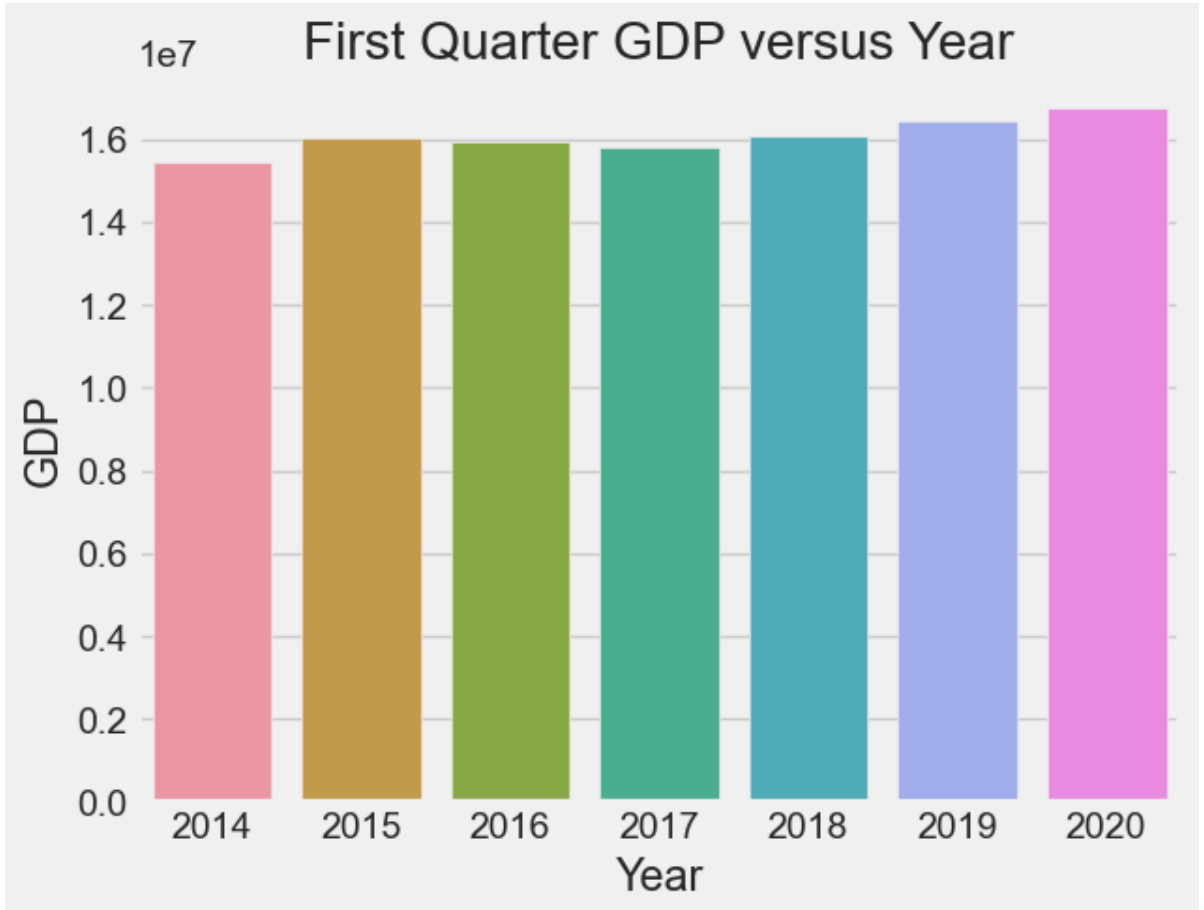
```
gdp_df = pd.read_csv('RealGDP.csv')
gdp_df
```

	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


1.0
0.8
0.6
0.4
0.2
0.0
0
5000
10000
15000
20000
25000

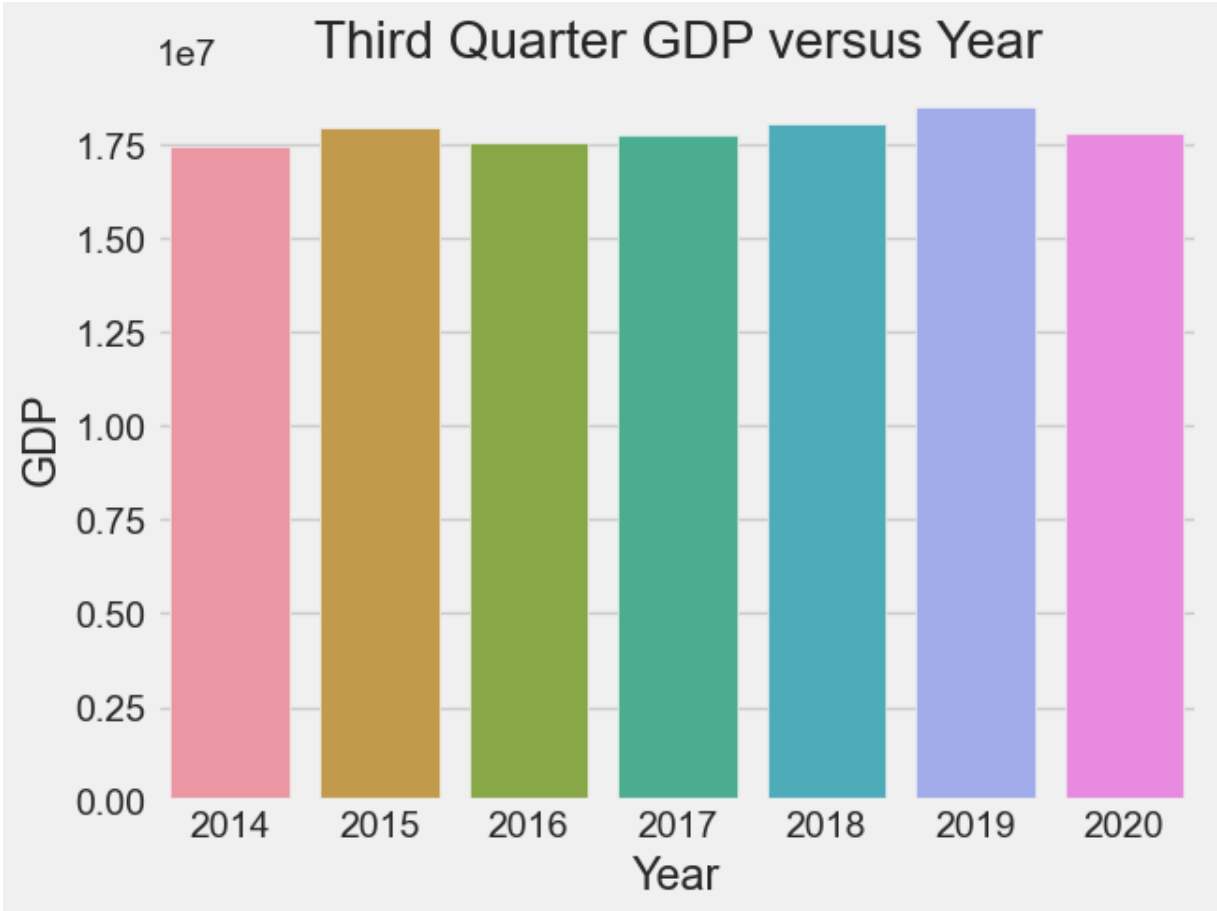
The result above shows that:

- Less dense area has the highest confirmed cases.


First Quarter GDP versus Year
1e7
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
GDP
2014
2015
2016
2017
2018
2019
2020
Year

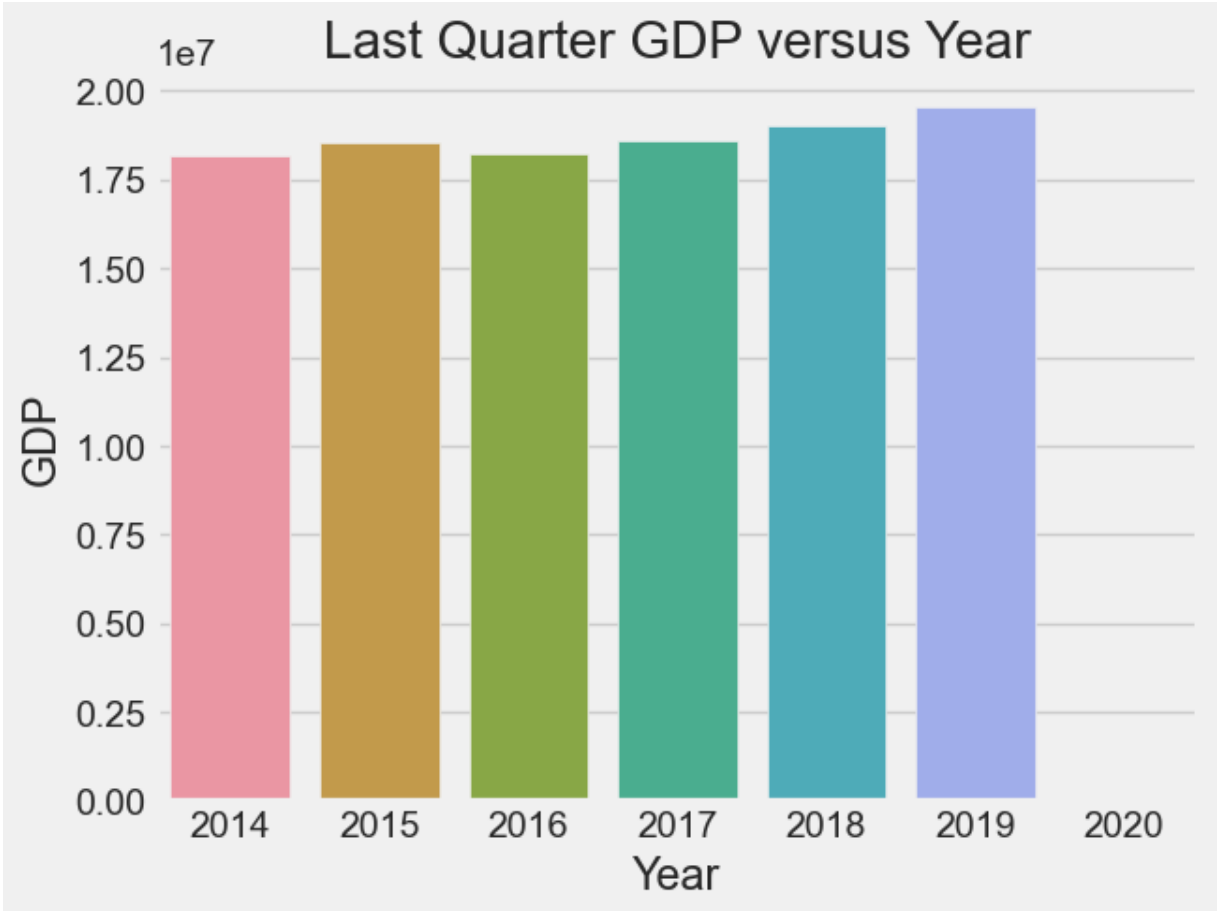
From the above graph,

It can be concluded that year 2020 recorded the highest GDP with more than 16 million while year 2014 recorded the least GDP with less than 16 million.


Third Quarter GDP versus Year
1e7
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
GDP
2014
2015
2016
2017
2018
2019
2020
Year

The graph above shows that:

In the third quarter of the year, year 2019 recorded the highest GDP close to 20 million while year 2014 and 2016 recorded the least GDP of 17.5 million.


Last Quarter GDP versus Year
1e7
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
GDP
2014
2015
2016
2017
2018
2019
2020
Year

The graph shows that, in the last quarter of the year,

- No GDP was recorded in the 2020,2019 was the year recorded the highest GDP approximately 20million while 2014 and 2016 recorded the least GDP.

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.