**COVID-19 VACCINE GAP – INVESTIGATION OF THE**

**UNDERLYING CAUSES**

**EPHRAIM TOCHUKWU AMANZE**

**STUDENT ID: 77278209**

**RESEARCH PAPER**

**Submitted in partial fulfillment of the requirements of Leeds Beckett University for the Degree of MSc. Data Science**

**SCHOOL OF BUILT ENVIRONMENT,**

**ENGINEERING, AND COMPUTING**

**LEEDS BECKETT UNIVERSITY**

**JANUARY 2022**

**Abstract**

Covid 19 vaccination gap has been a major concern to the World Health Organization in the face of the prevailing pandemic. To evaluate the driving factors behind this, three datasets on global vaccination count, infection and death rate and IMF report on GDP were merged and analysed with the aim of understanding and providing a recommendation of remedial action that could be taken to address the gap. The result of the study found that enough vaccine for percentage of a country’s population had the most significant correlation with GDP. A polynomial regression model has also been design to predict changes in enough vaccine for percentage of country’s population as a result of increase or decrease in GDP. The recommendation from this study is to provide financial aid to counties in the lower cadre of poverty level to purchase vaccine for their population size and to encourage countries on high GDP to redistribute and donate vaccine to poorer countries.

Table of Contents

1. Introduction……………………………………………………………………4

2. Aim and Objective:…………………………………………………………….5

2.1 Aim……………………………………..……………………………..5

2.2 Objective…………………………...………………………………..5

2.3 Research question:……..…………………….………………….…5

1. Literature Review……………………………………………………………7
2. Methodology………………………………………………………………….-9

4.1 Tools and Techniques…… ………………………….………………..9

4.2 Data selection and Collection……………..…………….........…….10

4.3 Data Cleaning and Merging ………………………………………10

4.4 Visualization and Modeling…………………………………………..11

4.5 Dataset Characteristics………………….……………….................11

1. Findings and Discussion……………………………………………………14

5.1 Level 1: Descriptive Statistical Analysis…………………………….14

5.1.1 Bar and Pie Chat Discussion……………..………………………29

5.2 Level 2: Inferential Statistical Analysis……………………………..30

5.3 Level 3: Use of Machine Learning – regression model…………37

5.3.1 Multivariate regression: GDP , Cumulative cases and Doses administered………………………………………………………………….37

5.3.2 Simple Linear regression: GDP and Enough\_for\_ percentage\_of\_population………………………………………………….39

5.3.3 Polynomial Regression……………………………………………41

5.4 Error with data………………………………………………………..44

1. Recommendations and Conclusion……………………………………….44

7. Reference………………………………………………….……………….47

**1. Introduction**

The current coronavirus disease (COVID-19) is recognised as one of the worst outbreak in human history.as reported by World Health Organisation (WHO, 2020). Nevertheless, the pandemic reaction has also equally exemplified rapid mobilisation and vaccine technology at its best. ( Josep, 2021)

Faced with a never-before-seen crisis, governments, international organisations, the health-care business, and global health experts have banded together on an unprecedented scale. As a result, several vaccinations have been developed in record time, with more than eight billion doses administered worldwide.( Bloomberg, 2021).

It's a fantastic accomplishment that demonstrates what we can achieve when we work together in a spirit of cooperation and unity.

The health situation is still out of hand; the delta variant and the new entrant omicron is spreading over the world, resulting in hundreds of additional fatalities every day, widening inequalities and rising poverty (WHO, 2021). According to International Monetary Fund (IMF, 2021); "Vaccine access has emerged as the principal fault line along which the global recovery splits into two blocks: those who can look forward to further normalisation of activity later this year (nearly all advanced economies) and those who will still face resurgent infections and rising COVID death tolls,"

It is this principal fault line that forms the basis of the rationale for this report. This has given rise to what is term the vaccine gap. The gap has brought forth the earning income (GDP) of countries as they battle the pandemic. The goal then is to use investigative analytics approaches on open data to understand what has given rise to this divide (Abhay,, 2020). The GDP and the covd-19 infection rate of each region and countries are used as indicators with the purpose of determining the relationship between these and the covid-19 vaccination progress of each country.

**2. Aim and Objective:**

**2.1 Aim**

The aim of this report is to utilise data on covid-19 vaccination, infection rate and economic income of nations to evaluate the vaccine gap and justify if financial and economic enhancement will bring to World Health Organization minimum level of covid-19 vaccination campaign of countries in the tail end of the vaccine administration. With this aim in mind this report will extablish the following objectives below;

**2.2 Objective:**

1. To carry out critical analysis of other reports and publications on the covid-19 vaccination gap.
2. To select the appropriate data, prepare and clean it for further analysis.
3. To look at the GDP, infection rate and death recorded and how they have impacted on vaccination drive.
4. To use the required algorithm to establish the appropriate decision that will address the vaccination gap.

**2. 3 Research question:**

1. Level 1: Descriptive Statistical Analysis
2. Which countries were among the top 20 in the vaccination campaign?
3. Which countries were among the top 20 in the infection and death rate
4. Which country/region among the top 20 in the economic earning/GDP chart?
5. Level 2: Inferential Statistical Analysis
6. Which of these criteria, GDP, infection and death rate has the strongest correlation with vaccine administration?
7. Level 3: Use of Machine Learning – regression model
8. Can change in economic earning/GDP have an impact on the vaccination campaign?

**3. Literature Review**

Economic Income and Vaccine Gap

A study of the vaccine gap as a result of economic income is capture in this research by (Bawa, 2021); according to Singh, in poor and least developed countries, healthcare facilities have been severely underfunded. Accessibility, affordability, and equity of healthcare facilities have become a key concern, particularly for low-income nations in the context of covid-19 vaccine due to substandard healthcare facilities, abject poverty, and insufficient research and development expenditures. According to the study, countries in this low income regions are economically handicapped to acquire the vaccine or produce it. It is noted in the study that developed countries, mainly China, the United States, the United Kingdom, Russia, and India, created the vaccines. At the same time, it is vital to note that there is low or no distribution of the same vaccine among the poor countries. Singh based his study from WHO publication on vaccine production and distribution, showing that over three-quarters of COVID-19 vaccine doses has been delivered only in the world's richest 10 countries, which account for around 60% of global GDP. The study stated that majority of COVID-19 vaccines were reportedly stockpiled in wealthy countries. A similar result was seen in a report by conducted by (Barbara Nel, 2021) using data from Bloomberg and Tony Blair Institute for Global Change, The study stated that over 8billion of covd-19 vaccine has been produced, but the coronavirus, on the other hand, has exposed long-standing imbalances that perpetuate poor health and well-being in already vulnerable region. Using Africa as the base for the study, the report stated that fewer than 2% of those billions of vaccines have been distributed throughout Africa.

(Josep Borrell, 2021) used report and data from IMF and One World Data to analyze the vaccination inequality, the study stated that just 3.1 percent of persons in low-income nations had received at least one dose. Going further, the study stated that although global vaccination rates have increased, there is still a North-South gap and considerable differences in immunisation rates. The study made a time base observation of global covid -19 vaccinnation progress, stating on September 16, 31% of the world's population had been fully immunised. This was the case for 61 percent of the adult population in the EU. The study stated that the decision to purchase vaccines in bulk at the EU level paid off in the end. North America had 45 percent, Latin America and the Caribbean had 36 percent, and Asia had 34 percent, while Africa had only 3.8 percent. (Anna Rouw, 2021) conducted an in-depth study on the covine-19 vaccine disparity relying on vaccine procurement report from Duke Global Health Innovation Center. The study gave a succinct report on purchase with anticipation that there will not be enough vaccine doses to cover the world's population until at least 2023, according to a study that tracks COVID-19 vaccine purchases and reveals that high-income countries now own more than half of all worldwide doses purchased. Analyzing data from the Duke Global Health Innovation Foundation to compute the share of doses purchased by country income category versus their share of the global adult population (focused on adults, aged 18+, because most COVID-19 vaccines are now only available for adults), the study discovered that, despite accounting for only 19% of the worldwide adult population, high-income nations have purchased more than half (54 percent, or 4.6 billion) of all vaccination doses purchased so far. Low and Middle Income Countries (LMICs), who account for 81 percent of the global adult population, have acquired 33 percent of the remaining dosages, while COVAX has purchased 13 percent (Anna Rouw, 2021). Lower-middle-income countries had the greatest gap between doses purchased and population share (37 percent of world population vs. 12 percent of purchased doses, or 989 million doses), closely followed by upper-middle-income countries (37 percent vs. 18 percent , or 1.5 billion doses). For low-income nations, the disparity is lower (3 percent vs. 7 percent , or 263 million doses). The study concluded that the low-middle-income will experience further disparity in the year 2023 as multiple variances of the disease emerge unless a global holistic approach is adopted or a financial aid is put in place.

**4. Methodology**

4.1 Tools and Techniques:

Python programming language has remain the most applied tool in data science and has sufficient libraries to deploy at every stage (Abhay, 2020). Python libraries, pandas and numpy was used in the cleaning, repairing and merging of the datasets. While matplot and seaborn was used for the graphical display. Python module, statsmodel was utilze for the regression model. According to (Mahmood, 2020); statsmodel provides top-of-the-line statistics and econometric capabilities that have been validated against other statistics software. StatsModels includes a wide range of linear regression models, polynomial and mixed linear models regression with discrete defendant variables,

4.2 Data selection and Collection

Determination of goals for which a required data will be used is the first step in the data mining process (Carlos, 2019), The goal of this research is to understand the factors responsible for global covid-19 vaccine gap and how best to reduce the divide. As a result, vaccination track dataset and covid-19 global stats from Bloomberg and World Health Organisation respectively are necessary for this research. The most mentioned fact from the literature review was earning income as a predominant factor. GDP of countries was obtained from Wikipedia. The data being examined is covid-19 infection rate from outbreak till date. Dataset was analysed and ongoing vaccination till date was pulled for analysis .The World Health Organisation, Bloomberg and Wikipedia have previously collected and organised the data, thus it is secondary data.

4.3 Data Cleaning and Merging

After that, the three dataset was set up for cleaning deploying pandas and numpy. The impurities was extracted using regex, i loc and loc (Mahmood, 2020). Column names were adjusted to ensure no data was lost at the point of merging. Inner merging was used to join the three dataset .together using pandas.

Following that, the data was examined for any errors and Nan values. Additionally, the data was purified to remove some of the results that were unnecessary for the following level of analysis. For example, the GDP for 2021 was required which was reported by IMF. The report from world bank was for 2019 so this information was excluded from the data set. There are other anomalies in the merged data which produce Nan values as a result of non-uniform country list per dataset, there were 238 countries in the covid-19 stats from WHO, 232 countries in the vaccine track dataset from Bloomberg and 209 countries in the GDP dataset from Wikipedia, there was a record of Nan value at the point of cleaning the data as most countries has missing data in the various dataset. To ensure the result is not skewed by countries with missing data if included, the rows with Nan values were dropped after final merging.

4.4 Visualization and Modeling

The data will be analysed and visualised, with matplot and seaborn to depict the data's trends (Jake, 2020). The research focus for this study is on the vaccine gap in relationship to GDP, infection and death rate. Pearson's correlation coefficients will be used to determine the importance of the relationship between the variables. Finally, the data will be subjected to machine learning analysis. Linear and polynomial regression model will be deployed in this study.. This model was selected because it allows for a model that can assess the statistical significance of the independent variables, GDP, to the dependent variable, vaccination count. This type of regression model also allows for the prediction of the dependent variable for future measurement (Jake, 2020). For vaccination gap, this regression model can be used to understand if financial aid will bridge the gap to the WHO minimum vaccination requirement per country.

4.5 Dataset Characteristics

The final resulting dataset following data cleaning and preparation can be seen below

The final dataset has 14 variables and 159 countries (observations)

Each of the variables are describe below:

Date\_reported

The date of 2December 2021 was the date the dataset was collected for reported infection rate and vaccination count. Therefore the analysis is made on incidence recorded till this date.

Country

Country consist of member nations that was pulled from the 3 dataset.

New\_cases

Covid-19 cases reported on 2 December 2021

Cumulative\_cases

Covid-19 cases recorded from the inception of infection breakout till 2 December 2021

New\_deaths

Deaths reported on 2 December 2021

Cumulative\_deaths

Deaths recorded from the inception of infection breakout till 2 December 2021

Doses\_administered

Doses of vaccine that has been administered per country

Enough\_for\_per\_of\_people

Covid-19 vaccine secure per country for their population size. This is vaccine already secured measured against each country population size.

Percentage of population with 1+ dose

Percentage of population that has taken one dose vaccine per country.

Percentage of population fully vaccinated

Percentage of population with double vaccination per country.

Daily rate of doses administered

Daily rate of doses administered per country

GDP\_by\_IMF

GDP estimate per capital from IMF was used as this present up till date report

Year as per IMF

The year of IMF report taken for the study is 2021

**5. Findings and Discussion**

1. Level 1: Descriptive Statistical Analysis

A1. Which countries were among the top 20 in the vaccination campaign?

1. The top 20 countries with double vaccination

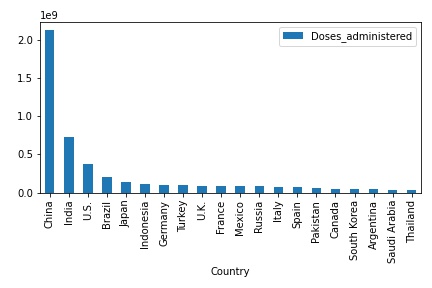


Fig 1: The top 20 countries with the highest doses administered

The fig 1 above shows the top 20 country with highest number of doses administer. From the bar chart, it is clear that China leads the group with 2.12billion with the closest figure of 0.73billion and 0.37billion coming from India and US respectively the least in top 20 is Thailand with 0.03billion of doses

1. The bottom 20 countries with number of doses administered

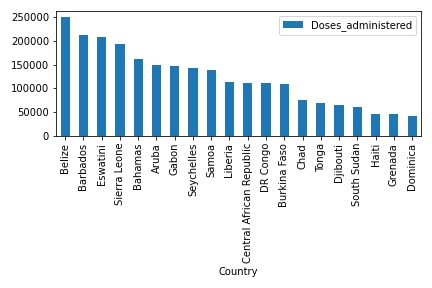


Fig 2: The top 20 countries with the highest doses administered

The fig 2 shows that Dominica, Grenada and Haiti are at the bottom chart with 0.000043billion, 0.000045billion and 0.000046bilion respectively. Belize and Barbados lead the bottom 20 by 0.00025 and 0.00021 respectively

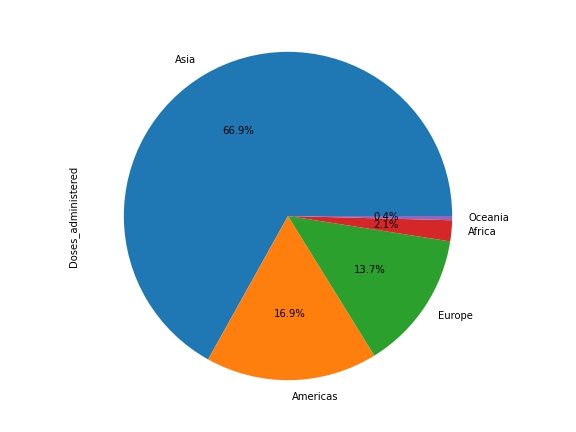


Fig a. Regional representaion of doses administred

1. The top 20 countries with doses enough for percentage of population

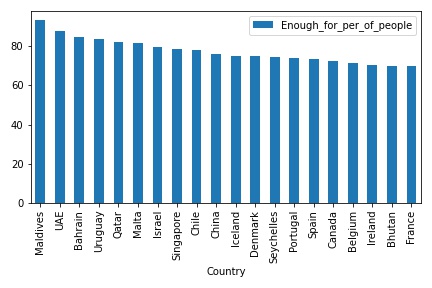


Fig 3: The top 20 countries with doses enough for percentage of population

Maldives and UAE lead this group with 93% and 89% respectively. The last 2 countries in the group are Bhutan and France with 73% and 72% respectively

1. The bottom 20 countries with doses enough for percentage of population

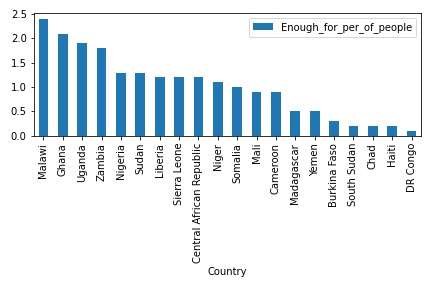


Fig 4: The bottom 20 countries with the lowest doses administered

The last 2 countries at the bottom are DR Congo and Haiti with 0.15% and 0.24%. The top of this group are Malawi and Ghana with 2.42% and 2.12% respectively

## C:\Users\ANDYTECH\Documents\DAV\image14.jpeg

Fig b. Regional representation of Doses enough for percentage of population

A2. Which countries were among the top 20 in the infection and death rate

1. The top 20 countries with the highest infection rate

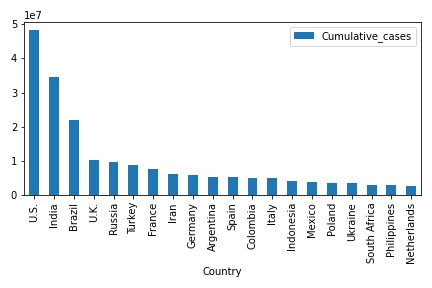


Fig 5: The top 20 countries with highest infection cases

The fig5 above shows the top 20 countries with the highest infection rate. US and India top the group with 4.8million and 3.4million respectively. The last 2 countries are Philippines and Netherlands with 2.8million and 2.6million respectively

1. The bottom 20 countries with the lowest infection rate

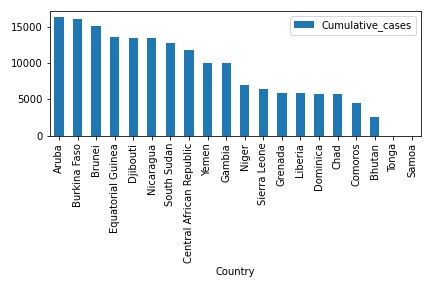


Fig 6: The bottom 20 countries with lowest infection cases

The fig 6 above shows the bottom 20 countries with the lowest infection rate. Tonga and Samoa are at the bottom of the group with 1 case each. Aruba and Burkina Faso top the group with 16,354 cases and 16,000 cases respectively

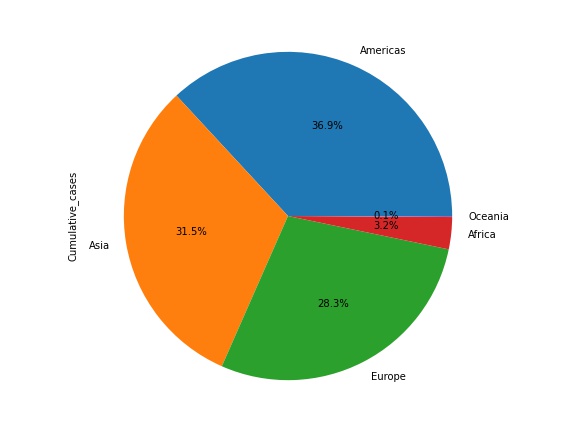


Fig c. Regional representation of cumulative cases

1. The top 20 countries with the highest death rate

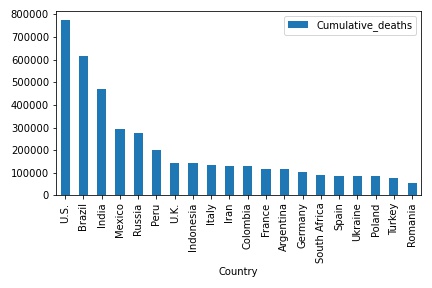


Fig 7: The top 20 countries with highest death rate

USA and Brazil top the top with 780,000 and 610,000 respectively. Turkey and Romania are at the foot with 82,000 and 64,000 respectively.

1. The bottom 20 countries with the lowest death rate

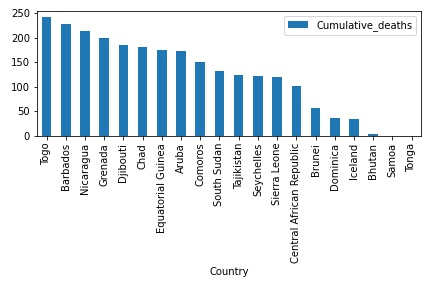


Fig 8: The bottom 20 countries with lowest death rate

The fig 8 above shows the bottom 20 countries with the lowest death rate. Tonga and Samoa are at the bottom of the group with 0 case each. Togo and Barbados top the group with 343 cases and 229 cases respectively

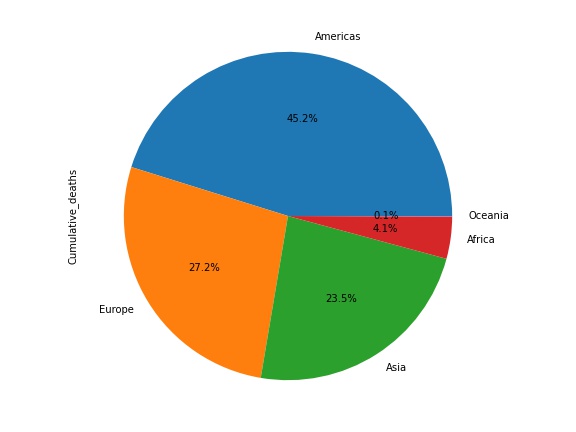


Fig d. Regional representation of cumulative deaths

A3. Which countries were among the top 20 in the economic earning/GDP chart?

1. The top 20 countries with the highest GDP

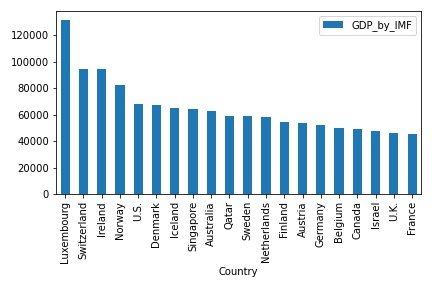


Fig 9: The top 20 countries with highest GDP

The fig 9 above shows the top 20 countries with the highest GDP. Topped by Luxembourg and Switzerland by $131,782 and $94,698 respectively.

1. The bottom 20 countries with the lowest GDP

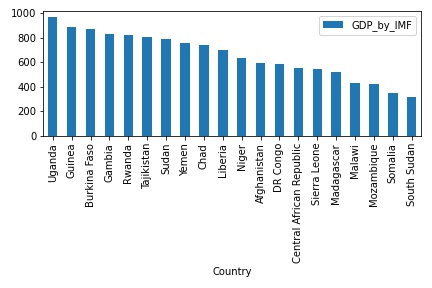


Fig 10: The bottom 20 countries with lowest GDP

The fig 10 above shows the bottom 20 countries with the lowest GDP. Somalia and South Sudan are at the bottom of the group with $347 and $315 respectively.

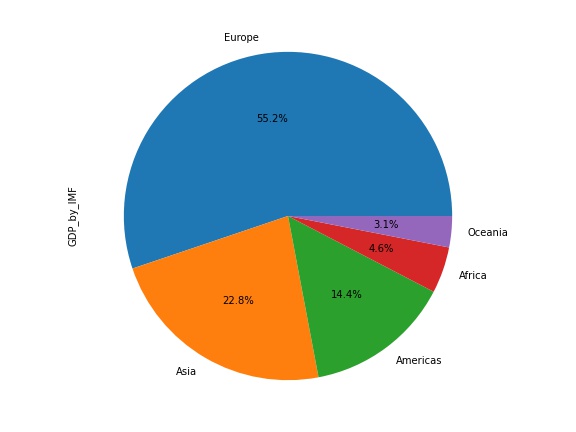


Fig e. Regional representation of GDP

5.1.1 Bar and Pie Chat Discussion:

It can be inferred from Fig1 that China has the highest doses administered by 2.12billion but was missing among the top 20 richest country in Fig 9. In fact, only 5 countries among the top 20 with highest administered doses appeared in the top 20 richest country. But this was in contrast to the countries with the highest infection rate in Fig 5 where16 countries with the highest administered doses recorded the highest infection rate. Similar pattern was also observed in Fig 7 where the top countries with highest administered doses occupied 14 spots among the top 20 with death cases.

It can also be observed that out of the top 20 countries with the highest administered doses, only 4 appeared among the top 20 countries in the bar chart of enough for percentage of their population in Fig 3. This is also evident in Fig a, where the continent of Asia had 66.9% of the regional pie chart of doses administered while Europe had 13.7%. But the reverse is the case in pie chart for doses enough for population in Fig b, where Europe had 36.8% while Asia had 31.0%

While the poorer countries in Africa and Oceania recorded low in doses administered and enough for population, they also recorded low in infection and death rate.

It can be seen from the chart that countries who administered doses the most were more frequent in the top column of infection and death rate while countries with enough vaccine for population appear more in the top cadre of GDP.

**5.2 Level 2: Inferential Statistical Analysis**

1. Which of these criteria, GDP, infection rate and death rate has the strongest correlation with vaccination ?
2. Correlation between GDP, Cumulative case, Cumulative death and Doses Administered

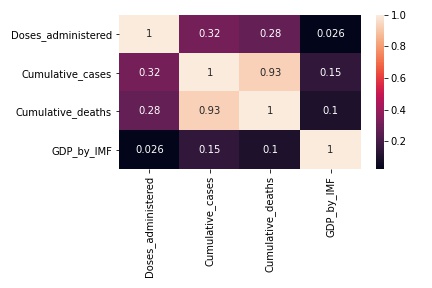


Fig11. Heat map of correlation

It can be seen from fig 11 that GDP has the weakest correlation value of 0.026 with doses administered while cumulative death and cumulative cases have a better correlation value of 0.28 and 0.32 respectively with doses administered. The weak correlation value of GDP has lay credence to the bar chart in fig 1 and fig 9 where countries with highest vaccination doses appear less among the richest country. The doses administered has more be driven by impact of death and infection rate rather the GDP.

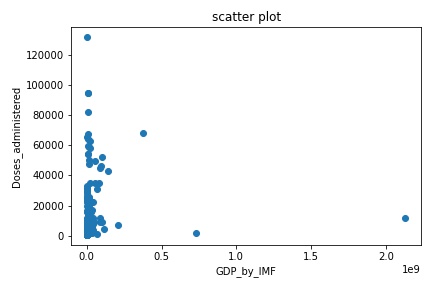


Fig 12 scatter plot between GDP and Doses administered

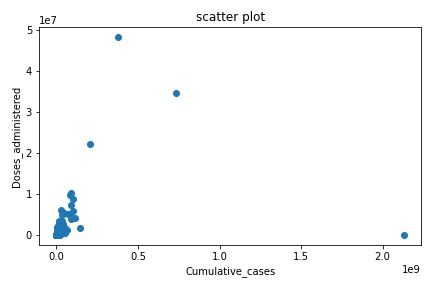


Fig 13 scatter plot between Cumulative cases and Doses administered

It can be seen from the scatter plot that cumulative case has a better liner re0lationship with doses administered than GDP.

1. Correlation between GDP, Cumulative case, Cumulative death and Enough vaccine for percentage of population

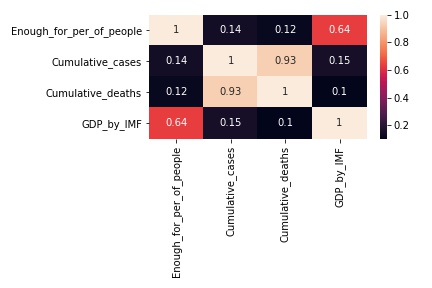


Fig14. Heat map of correlation

In contrast to fig 11, fig 14 has shown that GDP has the strongest correlation value of 0.646 with enough vaccine for percentage of population while cumulative death and cumulative cases have a weaker correlation value of 0.14 and 0.12 respectively with enough vaccine for percentage of population. This is an interesting part of the analysis, To secure enough vaccination for any country’s population size is more dependent on the country’s economic earning/GDP. This is illustrated in bar chart fig 1 and 3. If a country has secure a high dose volume of 1million, this will not be enough if the population size is high, another country with a lower population size with same 1mjllion dose will vaccinate more population size. The richer countries has average population size and more money to secure enough vaccine for this population size, but the poorer countries cannot afford the luxury of securing enough vaccine for their population size, Most of them appeared at the bottom in fig 4 bar chart where 17 out of the 20 lowest countries with enough vaccine for percentage of population are African countries.

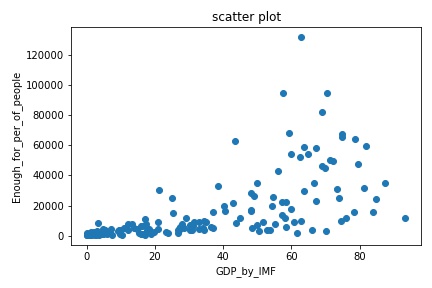


Fig 15 scatter plot between GDP and Enough for percentage of population

It can be seen from the scatter plot in fig 15 that GDP has a better liner relationship with Enough for percentage of population

.

5.3 Level 3: Use of Machine Learning – regression model

Can increase in economic earning/GDP impact on the volume of vaccine administered?

* + 1. Multivariate regression: GDP and Doses administered

From the correlation analysis conducted above, it can be observed that GDP has minimal to zero correlation with Doses administered. We shall further examine this with OLS linear regression model

OLS Regression Results

==============================================================================

Dep. Variable: Doses\_administered R-squared: 0.100

Model: OLS Adj. R-squared: 0.089

Method: Least Squares F-statistic: 8.709

Date: Thu, 30 Dec 2021 Prob (F-statistic): 0.000260

Time: 13:14:28 Log-Likelihood: -1043.4

No. Observations: 159 AIC: 2093.

Df Residuals: 156 BIC: 2102.

Df Model: 2

Covariance Type: nonrobust

====================================================================================

coef std err t P>|t| [0.025 0.975]

------------------------------------------------------------------------------------

Intercept 20.0529 17.071 1.175 0.242 -13.668 53.773

GDP\_by\_IMF -0.1896 0.630 -0.301 0.764 -1.434 1.055

Cumulative\_cases 0.0112 0.003 4.159 0.000 0.006 0.016

==============================================================================

Omnibus: 337.772 Durbin-Watson: 2.022

Prob(Omnibus): 0.000 Jarque-Bera (JB): 137342.432

Skew: 11.801 Prob(JB): 0.00

Kurtosis: 145.035 Cond. No. 6.76e+03

==============================================================================

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Fig 17. Multivariate model, GDP, Cummulative case and Doses adninstred

From fig 17 above, we can derive the following from the model

1. This is a multivariate regression analysis to evaluate how GDP and Cumulative\_cases affect Doses\_administered. Cumulative\_deaths was dropped due to high multicollinearity
2. Taking a significant level ( of 0.05, the GDP has a Pvalue of 0.764 which is greater than 0.05 and shows that GDP has no significant impact on Doses\_administered.
3. Cumulative\_cases has a Pvalue of 0.00 which is lower than 0.05 and shows that Cumulative\_cases has a significant impact on Doses\_administered.
4. The R-square is 0.1 which is very low and shows that the model can not be effective

* + 1. Linear regression: GDP and Enough\_for\_ percentage\_of\_population

From the correllation analysis conducted , it was proven that GDP correllation with Enough\_for\_ percentage\_of\_population. We shall also examine this with OLS linear regression model

OLS Regression Results

====================================================================================

Dep. Variable: Enough\_for\_per\_of\_people R-squared: 0.404

Model: OLS Adj. R-squared: 0.400

Method: Least Squares F-statistic: 106.3

Date: Thu, 30 Dec 2021 Prob (F-statistic): 2.33e-19

Time: 15:24:36 Log-Likelihood: -706.03

No. Observations: 159 AIC: 1416.

Df Residuals: 157 BIC: 1422.

Df Model: 1

Covariance Type: nonrobust

==============================================================================

coef std err t P>|t| [0.025 0.975]

------------------------------------------------------------------------------

Intercept 21.8711 2.010 10.881 0.000 17.901 25.841

GDP\_by\_IMF 0.0008 7.43e-05 10.311 0.000 0.001 0.001

==============================================================================

Omnibus: 8.183 Durbin-Watson: 2.174

Prob(Omnibus): 0.017 Jarque-Bera (JB): 8.340

Skew: 0.560 Prob(JB): 0.0155

Kurtosis: 3.051 Cond. No. 3.32e+04

==============================================================================

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Fig 18 . Linera model, GDP and Enough\_for\_ percentage\_of\_population.

From fig 18 above, we can derive the following from the model

1. This is a simple linear regression analysis to evaluate how GDP affect Enough\_for\_ percentage\_of\_population
2. Taking a significant level ( of 0.05, the GDP has a Pvalue of 0.000 which is very lower than 0.05 and shows that GDP has a significant impact on Enough\_for\_ percentage\_of\_population
3. The R-square is 0.4 which shows that only 40% of variability in Enough\_for\_ percentage\_of\_population can be accounted for by linear regression on GDP.
4. Therefore, the model is inadequate.
   * 1. Polynomial regression: GDP and Enough\_for\_

percentage\_of\_population

Having seen the R suare value from linear regression, it is imperative that the relationship might not to perfectly linear. To further examine the relationship, polynomial regression will be used to observed if there will be improvement in the model performance.

OLS Regression Results

====================================================================================

Dep. Variable: Enough\_for\_per\_of\_people R-squared: 0.557

Model: OLS Adj. R-squared: 0.552

Method: Least Squares F-statistic: 98.17

Date: Thu, 30 Dec 2021 Prob (F-statistic): 2.51e-28

Time: 16:26:00 Log-Likelihood: -682.37

No. Observations: 159 AIC: 1371.

Df Residuals: 156 BIC: 1380.

Df Model: 2

Covariance Type: nonrobust

======================================================================================

coef std err t P>|t| [0.025 0.975]

--------------------------------------------------------------------------------------

Intercept 14.6927 1.993 7.372 0.000 10.756 18.629

GDP\_by\_IMF 0.0018 0.000 11.559 0.000 0.002 0.002

I(GDP\_by\_IMF \*\* 2) -1.302e-08 1.77e-09 -7.355 0.000 -1.65e-08 -9.52e-09

==============================================================================

Omnibus: 23.323 Durbin-Watson: 2.107

Prob(Omnibus): 0.000 Jarque-Bera (JB): 28.461

Skew: 0.985 Prob(JB): 6.60e-07

Kurtosis: 3.641 Cond. No. 2.96e+09

==============================================================================

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Fig 19 Polynomial model, GDP and Enough\_for\_ percentage\_of\_population.

From fig 19 above, we can derive the following from the model

1. This is a polynomial regression model to evaluate how GDP effect Enough\_for\_ percentage\_of\_population
2. A degree of 2 was used on the GDP variable
3. Taking a significant level ( of 0.05, the GDP has a Pvalue of 0.000 which is very low and shows that GDP has a significant impact on Enough\_for\_ percentage\_of\_population.
4. The R-square is 0.55 which shows an improvement from the linear regression and can be derived that only 55% of variability in Enough\_for\_ percentage\_of\_population can be accounted for by polynoial regression on GDP
5. The percentage though insufficient enough but the model has shown a realistic coefficient of 14.6 against the coefficient of 21.8 shown in the linear model therefore, can give a better prediction
6. It can be seen from the model that an increase in GDP will result to an increase in Enough\_for\_ percentage\_of\_population

5.4 Error with data:

Some errors observed with the data shows an imperfect liner relationship when each dataset is pair against each other especially in regard to vaccine count and purchase as each country maneuver to secure the vaccine for their population. The same applies to report of cases of infection and death with observed irregularities from countries trying to manage their figures These types of errors are called irreducible errors and can only be taken into account.

6. Recommendation and Conclusion:

From the study it is observed that why a high volume of vaccine has been administered, this was mainly driven by the impact of the covid-19 virus on countries who administered the most doses. As the top countries with highest infection and death rate had the highest doses administered. The vaccine gap perceive from this perspective was created as a response to death and infection rate but did not reflect the true holistic reason and can only be justified as a reaction equal to action.

But to ensure the rest of the world population are vaccinated, the rich countries have rather secure the large portion of the vaccine or having in possession the production capacity of the vaccine for their population size which has left the poor countries with little or no vaccine for their population size. This has created an unhealthy vaccine gap which will defeat the WHO objective of one person per minimal dose of vaccine.

To properly comprehend this gap, we look at the descriptive statistics of Enough\_for\_ percentage\_of\_population table below

Enough\_for\_per\_of\_people

count 159.0

mean 33.88427672955974

std 26.66105236556486

min 0.1

25% 9.3

50% 29.0

75% 57.9

max 93.4

Table1. descriptive statistic - Enough\_for\_per\_of\_people

It can be seen that the standard deviation of 26 is very high and a large percentage of the countries are below the mean value of 33 as 25% percentile and 50% percentile indicate.

It is apparent from the study that to secure enough vaccination for any country’s population size is more dependent on the country’s economic earning/GDP. The richer countries has more money to secure enough vaccine for their population size, but the poorer countries cannot afford the luxury of securing enough vaccine for their population size, Most of them appeared at the bottom in fig 4 bar chart where 17 out of the 20 lowest countries with Enough vaccine for percentage of population are African countries.

The following recommendation shall be made to reduce the gap.

1. Financial aid to be provided to the countries in the 25 and 50 percentile for procurement of vaccine
2. vaccine production and redistribution by countries on the 75 percentile should be encourage to imbibe donation to poorer countries in the 25 percentile
3. Economic enhancement to be facilitated among countries on the lower percentile as means of long term sustenance of a country ability to provide health facility for its population.

**7. Reference**

WHO. (2021). *Vaccine inequity undermining global economic recovery*. [online] Available at: https://www.who.int/news/item/22-07-2021-vaccine-inequity-undermining-global-economic-recovery.

Josep Borrell (2021). *Vaccination gap: From promises to action - World*. [online] Available at: https://reliefweb.int/report/world/vaccination-gap-promises-action

Bloomberg (2021). More Than 1.2 Million People Have Been Vaccinated: Covid-19 Tracker. *Bloomberg.com*. [online] Available at: https://www.bloomberg.com/graphics/covid-vaccine-tracker-global-distribution/.

IMF. (2021). *World Economic Outlook Update, July 2021: Fault Lines Widen in the Global Recovery*. [online] Available at: https://www.imf.org/en/Publications/WEO/Issues/2021/07/27/world-economic-outlook-update-july-2021

Rittikar, S. (2021). *Statistical Tests*. [online] Analytics Vidhya. Available at: https://medium.com/analytics-vidhya/statistical-tests-22e91674315b .

Singh, B. and Chattu, V.K. (2021). Prioritizing “equity” in COVID-19 vaccine distribution through Global Health Diplomacy. *Health Promotion Perspectives*, [online] 11(3), pp.281–287. Available at: https://web.a.ebscohost.com/ehost/detail/detail?vid=2&sid=08b1b7b3-4f92-4c08-9856-e3cd15a10ff2%40sdc-v-sessmgr02&bdata=JnNpdGU9ZWhvc3QtbGl2ZQ%3d%3d#AN=152369570&db=asn

Barbara Nel (2021). Bridging the vaccine gap [online] 14(2), pp. 23-14. Available at: https://www.devex.com/news/sponsored/opinion-bridging-the-vaccine-gap-threatening-africa-100447

Anna Rouw (2021). Global COVID-19 Vaccine Access: A Snapshot of Inequality, [online] 7(4), pp.182–87. Available at: https://www.kff.org/policy-watch/global-covid-19-vaccine-access-snapshot-of-inequality/?utm\_campaign=KFF-2021-The-Latest&utm\_medium=email&\_hsmi=116422786&\_hsenc=p2ANqtz84jkHstTqZ01SuX3bI8\_zTHtNwYTiutrhQtM01MJYCpIsGQeK8ScJziZIYG9fsy5QozouMOWXYFUs2tsGliMoemOvTiZRN33MAdBm7jtYnmhFea08&utm\_content=116422786&utm\_source=hs\_emil

Stephenson, J. (2021). Unequal Access to COVID-19 Vaccines Leaves Less-Wealthy Countries More Vulnerable, Poses Threat to Global Immunity. *JAMA Health Forum*, 2(3), p.e210505

Jake VanderPlas (2021). *Python Data Science Handbook | Python Data Science Handbook*. [online] Available at: https://jakevdp.github.io/PythonDataScienceHandbook/index.htm

Mito, J. from (2021). *10 Pandas Functions that make Data Science Faster*. [online] Medium. Available at: https://medium.com/trymito/9-pandas-functions-that-make-data-science-faster-72688e06ce91

Mahmood, M.S. (2021). *Simple Linear and Polynomial Regression*. [online] Medium. Available at: https://towardsdatascience.com/simply-linear-and-polynomial-regression-9c9624774f5a

Jidge, A. (2021). *Assumptions of Linear Regression*. [online] Medium. Available at: https://medium.com/@abjidge/assumptions-of-linear-regression-3e3de8eb774a.

Dawson, C. (2021). *Understanding Multiple Linear Regression*. [online] The Startup. Available at: https://medium.com/swlh/understanding-multiple-linear-regression-e0a93327e960