**DATA EXPLORATION AND VISUALIZATION**

**DATA EXPLORATION REPORT**

***“COVID-19 WORLD VACCINATION PROGRESS”***

****

**Name: Sai Suraj Enumula**

**ABSTRACT**

COVID-19 is a global pandemic, which as of September 30, 2021, has affected about 217.9 million people and claimed over 4.5 million deaths globally (“COVID-19 Map,” n.d.). After originating in Wuhan, China, it spread to other countries mainly through people travelling in and out of China. Scientific community across the globe raced to find answers in terms of therapeutics and vaccines to control SARS-CoV-2. (Vandana Gupta, 2020). Prior to the COVID-19 Pandemic (Wikipedia contributors, n.d.), scientists were able to establish knowledge about the diseases caused by a family of RNA Viruses called coronaviruses like severe acute respiratory syndrome (2002-2004 SARS outbreak) and Middle East respiratory syndrome (MERS). This is attributed to the fact that SARS-CoV-2 uses the same receptor as SARS-CoV on the host cell i.e., human Angiotensin Converting Enzyme 2 (hACE2) and is approximately 79% similar genetically to SARS-CoV which caused 2002-2004 SARS outbreak.

1. **INTRODUCTION**

This study aims to explore the global vaccination progress until August 2021 through following questions:

1. *Vaccination Rate:* These questions aim to provide answers on what percentage of people have been vaccinated fully and what percentage of population have received at least one dose with bifurcation by country, age group. This question also aims to answer vaccination rates in comparison to diabetes prevalence, smoking habits, handwashing facilities, hospital beds, income group and human development index
2. *Manufacturer wise usage:* These questions aim to provide answers on which vaccine manufacturer was widely adopted by country, age group and how effective was a vaccine on countries decline of covid cases.
3. *Vaccine preference by country:* These questions aim to infer the preference of a particular vaccine and how does it affect future of vaccination progress.

It is essential to understand why we have more than one manufacturer for vaccine for a single disease. As of August 2021- there are 22 vaccines authorized for use by national governments with six vaccines being approved for emergency or full use by at least one WHO-recognised regulatory authority (Wikipedia contributors, n.d.-b). There are two main types of vaccines:

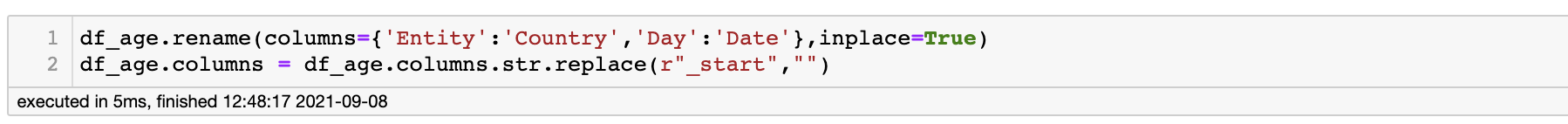
1. *RNA Vaccines:* These vaccines use RNA to simulate an immune response. When a vaccine containing RNA is introduced into the body, it acts as a messenger RNA which causes cells to build the SARS-CoV-2 spike protein. This trains the body how to identify and destroy the pathogen when it enters the body. Examples of this type of vaccines include Pfizer–BioNTech, Moderna.
2. *Adenovirus vector vaccines:* These vaccines use an adenovirus shell containing the DNA that encodes a SARS-CoV-2 protein which does not make new virus particles but rather produce the antigen which elicits a systemic immune response. Examples of this type of vaccines include Oxford-AstraZeneca, Sputnik V, Janssen COVID-19 vaccine (Johnson & Johnson vaccine)

Further, each vaccine has different FDA approval dates. (“Historic Dates and Events Related to Vaccines and Immunization,” n.d.)

1. **DATA WRANGLING**
2. *Description of Data Sources:*

Covid Datasets from OWID daily updated dumps in GitHub: <https://github.com/owid/covid-19-data/tree/master/public/data/vaccinations> and Covid Datasets from WHO public downloadable dumps: <https://covid19.who.int/info/>

1. *Tools Used:* Python*,* Excel*,* Tableau
2. *Cleaning and Transformations:*
3. In the dataset containing age which contains the data related to vaccination percentage by age, Country name is mentioned as Entity and date as Day. To maintain consistency across datasets, both the columns have been renamed:



1. In the dataset provided by OWID, total number of columns are 32 with only 5 columns smoothed (removed noise). Hence, dropped all columns and kept only columns from which have been smoothed. The date column has been converted to datetime format.

Graphical user interface, text, application, email

Description automatically generated

1. In the dataset provided by WHO for vaccinations used by different countries, we have single column indicating a long string of all the vaccines used in country. For better visualising, the column has been split and a meaningful primary key has been created vac\_id has been created for better visualising.

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

1. In the dataset provided by OWID for age group bifurcation of vaccines issued, we have several age groups (For example:0-17,12-17,18-24,18-30) as various countries followed different reprint methodology. These groups have been combined to reduce the clutter. (For example:0-17,12-17 have been combined into 0-17 group). This has been done using tableau.

Graphical user interface

Description automatically generated

1. Consistency in vaccine names: Many vaccines are traded as different names in different countries,

For example: Vaccine developed by Oxford-AstraZeneca is traded as Covishield in India and Janssen - Ad26.COV 2.5 is manufactured by Johnson & Johnson and traded as Janssen

To avoid multiple names of the same manufacturer, following transformation has been done:

Text

Description automatically generated

Text

Description automatically generated

1. **DATA CHECKING**
2. *Duplicate and missing values:*

* Duplicates: No duplicates have been found in any of the datasets provided by OWID or WHO.
* Missing values: Empty string has been placed where we had the missing values.

1. *Insufficient Data:* 
   * 1. In the dataset provided by both OWID & WHO, vaccination bifurcation on age group is limited to only 29 countries and all which are in EU region. This is because most of the countries are not reporting Age data.
     2. Following countries have not reported the type of vaccine used: Bonaire, French Guiana, Saba, Sint Eustatius. This are special municipality, overseas department/region, single territorial of a country.
2. **DATA EXPLORATION**

*Questions:*

1. *How many people have been vaccinated?*
2. *What is percentage of people have been vaccinated fully?*

In the current dataset which has data range of 1 year 7 months from January 1, 2020 – August 31, 2021, only around 16% of world population is fully vaccinated. Only 30 out of 100 people have been fully vaccinated.

Table

Description automatically generated with medium confidence

***Global Fully Vaccinated Vs Total Vaccinations Trend (Animated):***

A picture containing graphical user interface

Description automatically generated

1. *Which country has highest vaccinations?*

China, United States, and India have the highest number of fully vaccinated population occupying first three spots respectively. In terms of Total vaccinations issued, China, India and United States occupy top 3 spots respectivelyChart

Description automatically generated with medium confidence

Chart, treemap chart

Description automatically generated

1. *Which age group has mostly vaccinated?*

Many Countries have initially started vaccination in a phased manner that prioritize those at high risk of complications such as the elderly and those at high risk of exposure and transmission such as healthcare workers, people working at COVID test centres, residents, and staff of nursing homes (Wikipedia contributors). CDC’s Advisory committee on Immunization Practices (ACIP) recommended that the 2nd phase of distribution include persons aged ≥75 years and non-healthcare frontline workers such as those employed in grocery stores, restaurants, military, law enforcement, fire departments, retail, sanitation, school, public transportation, news media etc (“ACIP COVID-19 Vaccine Recommendations | CDC,” n.d.). However, different countries have various rollout plans and age group bifurcation.

Graphical user interface, application

Description automatically generated

Trends from the above show that 80+ Age group have highest number of people fully vaccinated per 100 followed by 70-80 Age group.

1. *What is the Percentage of people fully vaccinated/percentage of people received at least one dose?*

Gibraltar has highest percentage of fully vaccinated population followed by Malta, Cayman Islands. It can be observed that countries with low population and with small number of people living in an area (i.e., population density) have higher percentage of fully vaccinated percentage. In current scenario, key challenge for any country is to secure enough vaccines. Small countries have gained advantage of better governance and efficient supply chain thereby contributing to faster delivery of vaccines within its boundaries.

Chart

Description automatically generated

Chart

Description automatically generated

1. *Bifurcation of countries with diabetes prevalence, smoking habits, handwashingfacilities, hospital\_beds\_per\_thousand,human deveoplment index*

It can be observed from the image below that United States with high diabetes prevalence has highest new vaccinations. Countries with high smokers like Croatia, Cyprus have low new vaccinations, however this data related to smoking cannot be inferred towards vaccination progress as the population density of those countries is low.

HDI has significant impact on the total population which is fully vaccinated and new vaccinations. Countries with higher HDI show either high fully vaccinated population or high number of new vaccines issued.Chart

Description automatically generated

1. *Manufacturer wise usage*

*-* *Which manufacturer (Pfizer/BioNTech, Oxford/AstraZeneca) was mostly used by country?*

Most countries used more than 1 vaccine which depend on on availability and procurement with countries like Mexico using 7 vaccine types.

Chart, scatter chart

Description automatically generated

Pfizer is most widely used vaccine with more than 300M doses used followed by Moderna (150M) and AstraZeneca(52M). Countries like china have mostly home-grown vaccines like Sinovac, Sinopharm which have been also exported in significant quantities to countries like Pakistan etc.

Chart

Description automatically generated

* *Which manufacturer is popular with respect to age/location?*

As mentioned in earlier, due to limited data available on vaccination bifurcation on age group is limited to only 29 countries and all are in EU region. However, it can be confirmed Pfizer has been most popular across the age groups. Further, Pfizer is most popular in EURO region followed by AstraZeneca and Moderna. In Western Pacific region (WPRO)- Pfizer was used for more than 80% of the jabs followed by Moderna.

Graphical user interface, chart

Description automatically generated

Timeline

Description automatically generated

1. **STASTICAL ANALYSIS:**

Following question would be answered using statistical analysis\*:

* *Effectiveness of vaccine manufacturer by country*

Due to large data, trends have been narrowed down to Quarterly data:Q4-2020-Oct-Dec, Q1-2021-Jan-Mar, Q2-2021-April-June, Q3-2021-July-Aug.Following model has been generated:

﻿*Trend Lines Model:*

**Measure names= New Deaths, Fully Vaccinated**

Chart, line chart

Description automatically generated

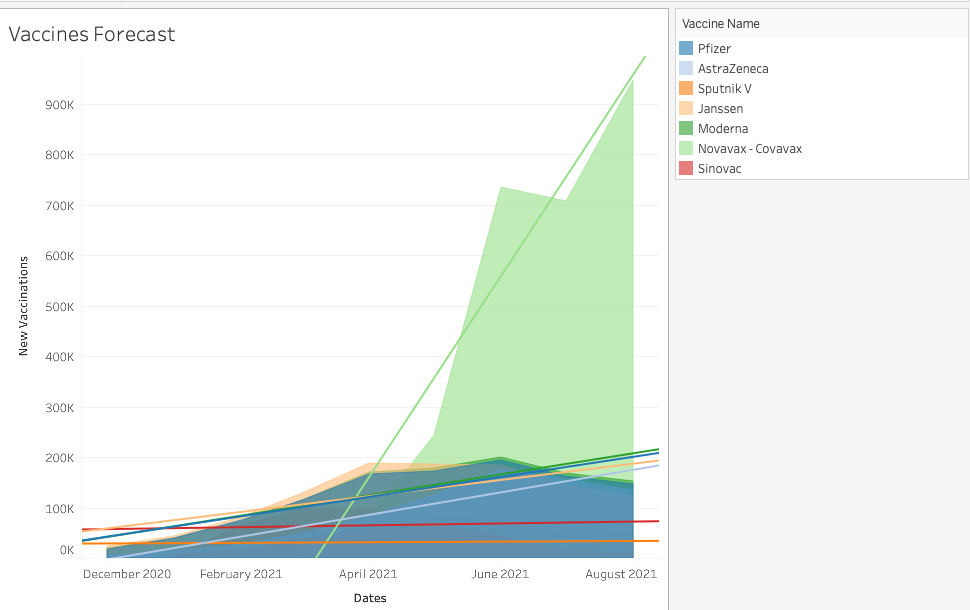
Table

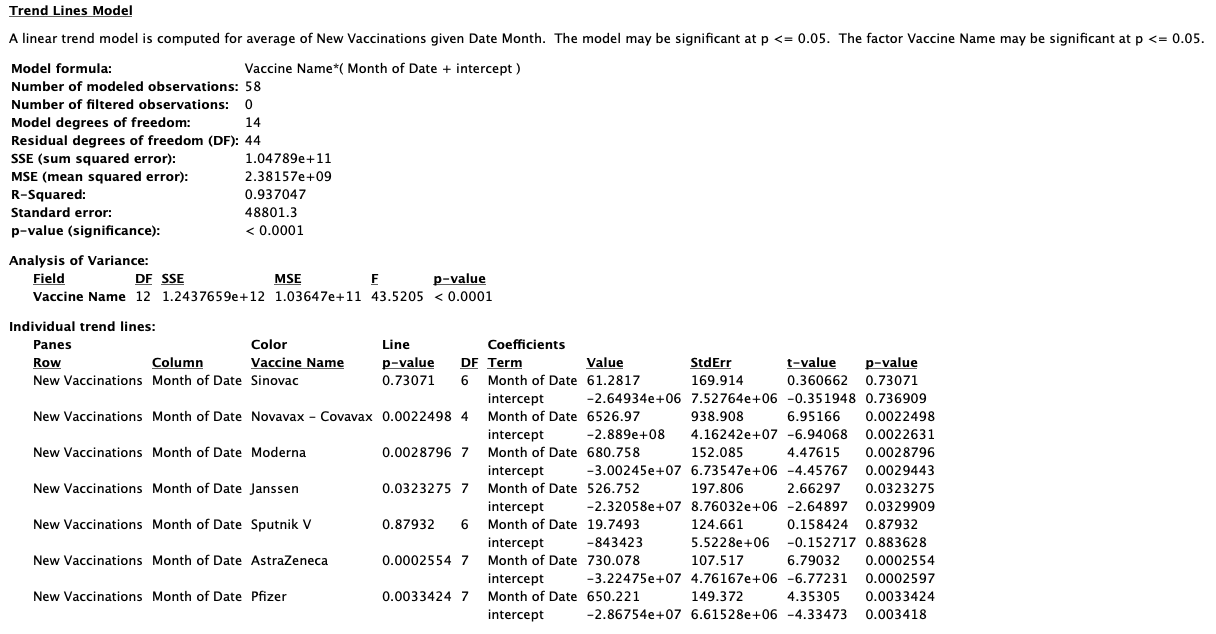
Description automatically generated

***\*Note-Due to page restrictions, analysis has been added as image***

From above, it can be inferred that the vaccines appear to be effective in preventing deaths from COVID‐19 in adults of all ages. Studies have confirmed that as COVID-19 virus mutates into variants with currently prevalent DELTA variant, vaccines are effective with more than one dose (2021 doi: 10.5694/mja2.51182).

1. *Can we infer what is preferred choice of vaccine manufacturer by country and why?*





***\*Note-Due to page restrictions, analysis has been added as image***

From the above trends, it can be inferred that most of the vaccines except Sinovac, Sputnik-V have preferences in future vaccinations in coming months, however there are many factors which include availability, socio-political factors, and awareness of vaccination among population.

1. **CONCLUSION:** From the above data exploration, it can be identified that countries with higher HDI and countries with better supply chain management and lower population density are able to fully vaccinate their population in a much faster way. Countries who have facilities to manufacture the vaccine doses themselves (like USA, CHINA, INDIA etc.) can achieve higher new vaccination rates. Effectiveness of vaccine can’t be concluded from this study as population is infected by new COVID variants, however media reports on side effects of a vaccine can have significant impact on adoption of a vaccine by countries population. With the current rate of vaccination, WHO expects to on global scale it is expected to achieve full vaccination percentage of have access to 1.425 billion doses of vaccine in 2021 and milestone of two billion doses is now expected to be reached in the first quarter of 2022 (“Joint COVAX Statement on Supply Forecast for 2021 and early 2022,” 2021).
2. **REFLECTION:** From this project, I was able to develop understand types of vaccines, global vaccination numbers and was to develop what are the factors which would affect availability of vaccine doses in coming months. However, as it can be observed from above analysis – considering the dataset of worldwide vaccinations is too large to draw statistically inferences. Further, number of questions which I wanted to answer could have been less. Focusing on country specific or region specific would have helped me in inferring superior interpretations and predictions.
3. **BIBLIOGRAPHY:**
4. COVID-19 Map. (n.d.). Retrieved September 8, 2021, from <https://coronavirus.jhu.edu/map.html>
5. Vandana Gupta, S. P. K. (2020, August 13). COVID-19 Vaccine: A comprehensive status report. Retrieved September 8, 2021, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7423510/>
6. Wikipedia contributors. (n.d.). COVID-19 vaccine. Retrieved September 8, 2021, from <https://en.wikipedia.org/wiki/COVID-19_vaccine> , Wikipedia contributors. (n.d.-c). COVID-19 vaccine. Retrieved September 8, 2021, from https://en.wikipedia.org/wiki/COVID-19\_vaccine
7. COVID-19 Vaccination. (n.d.). Retrieved September 8, 2021, from [Link](file:///C:\Users\suraj\FIT5147\new\Exploration\new\1.%09https:\www.cdc.gov\coronavirus\2019-ncov\vaccines\fully-vaccinated.html#:%7E:text=In%20general%2C%20people%20are%20considered,as%20Johnson%20%26%20Johnson’s%20Janssen%20vaccine)
8. Historic Dates and Events Related to Vaccines and Immunization. (n.d.). Retrieved September 9, 2021, from <https://www.immunize.org/timeline/> , ACIP COVID-19 Vaccine Recommendations | CDC. (n.d.). Retrieved September 8, 2021, from <https://www.cdc.gov/vaccines/hcp/acip-recs/vacc-specific/covid-19.html>
9. D.A.H.M.A.J.P.S.P.P.G. (2021, August 16). Effectiveness of COVID‐19 vaccines: findings from real world studies. Retrieved September 9, 2021, from <https://www.mja.com.au/journal/2021/215/4/effectiveness-covid-19-vaccines-findings-real-world-studies>
10. Joint COVAX Statement on Supply Forecast for 2021 and early 2022. (2021, September 8). Retrieved September 9, 2021, from https://www.who.int/news/item/08-09-2021-joint-covax-statement-on-supply-forecast-for-2021-and-early-2022