|  |  |
| --- | --- |
| Text  Description automatically generated  **Data Visualisation and Narratives**  **Visual Analytics project/reflection on COVID-19 Global Vaccination** |  |

**Student Names: IDs: Submission date:**

Ruoyang WANG 14121545 02/05/2021

Xilin HUANG 14035606

Yukai GUO 13656436

Introduction

It has been two unusual years for people living under the COVID-19 pandemic. Vaccination has been examined and produced in patches in 2021. This report will explore vaccination(s) performance of different countries from various brands and visualise the relationship between vaccine distribution and Human Development Index level and the vaccination performance and people's willingness to COVID-19 vaccinate after the international vaccination coverage.

Data Exploration

The project for data exploration mainly follows the ETL pipeline, which refers to extracting data, transforming data and loading it into an optimised format file (e.g., CSV).

To provide a cross-comparison on COVID-19 pandemic data and vaccination status, data are collected from 4 different public resources, World Health Organisation (WHO), Our World in Data (OWID), OxCGRT and YouGov. The following data exploratory analysis is fundamental for exploring the COVID-19 vaccination and vaccine willingness starting 2020-12-04 up to 2021-04-26.

The second step is data transforming; the mainstream is below:

1. Combing data from different sources
2. Converting data types
3. Parsing dates
4. Dealing with missing values
5. Removing outliers
6. Engineering features

Here are some transforming processes in detail.

Data sets from OWID and WHO are firstly merged by location name and published date. Then, we acquire features from this optimised format source, including country name, date, number for daily and total vaccines administrated, vaccine used, human development index (HDI) values and population, etc. It is noted that observations related to the number of total vaccinations with missing values in OWID datasets have been first filled with the vaccine data from WHO. If both data sources have not recorded this, we then filled it up with zero.  After several steps for data cleaning, the percentage for vaccination missing values has been reduced from 89% to 0.3%.

Data sets for vaccination policy and willingness are separately analysed without any missing value. At the same time, it is only available for a select number of countries that report these two necessary datasets. Therefore, we might focus on working with these observations already recorded.

Finally, we saved all transformed datasets into CSV format as our data warehouse for reporting and data analysis.

Narrative 1

What vaccines are top used in countries? What is the vaccine rate by different Human Development Index (G1-G5)?

Chart, funnel chart

Description automatically generated

From the Vaccines Production Distribution plot, we can see that the vaccines such as AstraZeneca, Pfizer BioTech, SII, Moderna, and Beijing CNBG are expected to be used widely in the world.

Moving on now to consider how are the vaccines distributed by different Human Development Index. United Nations Development Programme defines the Human Development Index (HDI) as the index to the summary measure of average achievement in key dimensions of human development: a long and healthy life, being knowledgeable and have a decent standard of living(http://hdr.undp.org/en/content/human-development-index-hdi). It should be the ultimate criteria for assessing a country's development rather than only emphasising the economic growth rate (*Human Development Index (HDI) | Human Development Reports n.d.*).

The followed picture shows that HDI requires analysis of the life expectancy index, education index, and GNI index indicators to understand a country's level of human development.

Diagram

Description automatically generated

Figure from (<http://hdr.undp.org/en/content/human-development-index-hdi>)

To better understand HDI, we classified the countries into five different levels:

|  |  |  |
| --- | --- | --- |
| Classification | Threshold | Level in HDI |
| G1 | >0.9 | High |
| G2 | 0.8-0.9 | Upper-middle |
| G3 | 0.7-0.8 | Middle |
| G4 | 0.6-0.7 | Lower-middle |
| G5 | <0.6 | Low |

Graphical user interface, application, Teams

Description automatically generated

From the plot, some interesting facts are easily noticeable:

1. G1, despite the small size of the population, has the highest percentage of vaccines.
2. Almost 40% of the total vaccines will deliver to G1 as expected. A higher level in HDI means decent and wealth in living. As a result, the expenditure of vaccines is more affordable for the countries in G1.
3. G1 has the highest vaccine token rate, which 26%. However, G5 only has 0.45%.

Our final report will discuss the correlations between the factors of development indicators and the numbers of vaccinations in the example. (Australia in G1 vs. Afghanistan in G5).

Narrative 2

After the beginning of the vaccine process worldwide, how the infection trend moves, and when can we achieve Herd Immunity?

Map

Description automatically generated

Map

Description automatically generated

Chart, line chart

Description automatically generated

According to the geographic map from OWI (*'Daily COVID-19 vaccine doses administered' n.d.*) and the daily vaccination rate, we can see that:

1. China is the first country that starts "Mass Vaccination" in December 2020.
2. Though the US started second, in terms of daily vaccinations, it stands at the top one.
3. Following that, we have the United Kingdom, which has a dramatically increase.
4. India comes in the 4th position in terms of daily vaccination percentage.
5. Australia has the lowest vaccination rate of the five countries.

Chart

Description automatically generated

Herd Immunity can be defined as the indirect protection from an infectious disease that happens when a population is immune either through vaccination or immunity developed through previous infection (*Coronavirus disease (COVID-19): Herd immunity, lockdowns and COVID-19 n.d.*). To achieving 'herd immunity' through vaccination, the threshold is about 80% which means about 80% of a population need to be vaccinated against COVID-19 (*Coronavirus disease (COVID-19): Herd immunity, lockdowns and COVID-19 n.d.*).

The above Herd Immunity Target plot shows the threshold: 80% of the total population. It looks like the vaccination process still in the early stage that only 15% of the population is vaccinated at least one dose till mid-2021. The expected time to achieve the optimistic scenario is 2022-12-29 with the current projection rate. However, the vaccine projection rate depends on various factors, such as the vaccine manufacturing and distribution, the infection rates, political lean (anti-polio campaign), and the people's willingness to the vaccine.

Narrative 3

How does the vaccine willingness change in developed countries？

Chart, line chart

Description automatically generated

The previous finding shows that the countries in G1 (high Human Development Index) expected to have more vaccinations. The above figure reveals that the variability of people's vaccine willingness between 8 countries in G1. Most people in 8 countries are willing to vaccinate till Feb 2021, despite a slight decrease of vaccine willingness in the earlier stage. The most likely causes of the increasing desire are education, confidence in the government/public health system, low barriers to getting the vaccine (*Willingness to get the COVID-19 vaccine with and without emergency use authorisation - PubMed n.d.).* On the other hand, the concerns of the people who are not willing to inject the COVID-19 vaccine mostly are worrying about the pace of vaccine development and potential danger in side-effects (*Sustained COVID-19 vaccine willingness in Denmark following the rare cases of blood clots | EurekAlert! Science News n.d.*). According to the researchers, the increased chances of blood clots cases that occur through the AstraZeneca vaccine injection leads to a low level of confidence in vaccination safety (*Sustained COVID-19 vaccine willingness in Denmark following the rare cases of blood clots | EurekAlert! Science News n.d.*).

It is interesting to note that Australia, as a successful country in suppressing the COVID-19, has a moderate willingness to have the vaccine. The United States with the highest infection rate has the lowest willingness.

Reference

*Coronavirus disease (COVID-19): Herd immunity, lockdowns and COVID-19* n.d., viewed 2 May 2021, <<https://www.who.int/news-room/q-a-detail/herd-immunity-lockdowns-and-covid-19>>.

'Daily COVID-19 vaccine doses administered' n.d., *Our World in Data*, viewed 2 May 2021, <<https://ourworldindata.org/grapher/daily-covid-19-vaccination-doses>>.

*Home* n.d., viewed 2 May 2021, <<https://www.who.int>>.

*Human Development Index (HDI) | Human Development Reports* n.d., viewed 2 May 2021, <<http://hdr.undp.org/en/content/human-development-index-hdi>>.

NW, 1615 L. St, Suite 800Washington & Inquiries, D. 20036USA202-419-4300 | M.-857-8562 | F.-419-4372 | M. 2021, 'Growing Share of Americans Say They Plan To Get a COVID-19 Vaccine – or Already Have', *Pew Research Center Science & Society*, viewed 2 May 2021, <<https://www.pewresearch.org/science/2021/03/05/growing-share-of-americans-say-they-plan-to-get-a-covid-19-vaccine-or-already-have/>>.

*Sustained COVID-19 vaccine willingness in Denmark following the rare cases of blood clots | EurekAlert! Science News* n.d., viewed 2 May 2021, <<https://www.eurekalert.org/pub_releases/2021-04/au-scv040921.php>>.

*Willingness to get the COVID-19 vaccine with and without emergency use authorisation - PubMed* n.d., viewed 2 May 2021, <<https://pubmed.ncbi.nlm.nih.gov/33227323/>>.

Appendix

|  |  |  |  |
| --- | --- | --- | --- |
| Source | File Name | Observations Counts | Missing Values Percentage (Vaccination related) |
| OWID (Our World in Data) | vaccinations.csv | 16545 | Location: 0   * Date: 0 * Total\_vaccinations: 37.8% * People\_vaccinated: 41.8% |
|  | owid-covid-data.csv | 84097 | * Location: 0 * Date: 0 * Total\_cases: 2.4% * New\_cases: 2.4% * Total\_vaccinations: 89% * People\_vaccinated: 90% * Population: 0.64% * Gdp\_per\_capita: 9.5% * Human\_development\_index: 8.9% |
| WHO (World Health Organisation) | vaccination-data.csv | 212 | * Country: 0 * Total\_vaccinations: 0.5% * Persons\_vaccinated\_1plus  \_dose: 12.7% * Vaccines\_used: 16% |
| OxCGRT (Oxford Coronavirus Government Response Tracker) | covid-vaccination-policy.csv | 88345 | * Entity: 0 * Day: 0 * Vaccination\_policy: 0 |
| YouGov (The Imperial College London YouGov Covid-19 Behaviour Tracker Data Hub) | covid-vaccine-willingness.csv | 83 | * Entity: 0 * Day: 0 * Willingness\_covid\_  vaccinate\_this\_week: 0 |

Reflection

The team have already decided which domain we are going to focus on making analysis, which is the COVID 19 Vaccinations. We did great teamwork with each other and made significant contributions to the project, including sharing ideas, helping others with programming and reporting, building visuals etc. We almost have a group meeting once in four days to make sure everyone on track. Honestly, everyone did their best to devote themselves to this assignment which resulted in efficient progress.

My main job in this program so far has three parts:

1. GitHub repository administration – I have created a GitHub repository that offers distributed version control and source code management. Thus, the team could share their datasets, visuals and codes with others.
2. Python code programming and reviewing, making visualisations and technique supports – I first provided some visual demos with several brainstorming insights for the team by using Python and packages like pandas, plotly and sklearn. After the team decided on the ideas that we are going to dig in more, I simply applied knowledge and techniques learnt from the courses to make visuals more informative and less complex to the audience. Problems during data transforming and visuals programming have troubled me a lot, especially dealing with columns in Date type. Some of them are stored in DateTime, while some of them are in Date format.
3. Ideas sharing, and progress report written.

For the next stages of this project, we have planned in the structure following:

1. Focusing on existing research fields we have already decided, grabbing more interesting and informative ideas from the datasets available.
2. Making more intuitive and less complex visualisations to support our research questions and then share with our stakeholders.
3. Acquiring more reports, news, policies and research through various channels to make the program in-depth, esp. we hope to provide a broader view for the universe to have a clear awareness of the current situation of the COVID pandemic.