**Project2-Group1**

**Goal**

The team goal was to look at COVID-19 vaccinations in relations to countries of the world populations and their GDP percapita.

**Data Extraction**

[**https://www.kaggle.com/kamal007/global-covid19-vaccination**](https://www.kaggle.com/kamal007/global-covid19-vaccination)

[**https://www.kaggle.com/eng0mohamed0nabil/population-by-country-2020**](https://www.kaggle.com/eng0mohamed0nabil/population-by-country-2020)

Data was extracted from Kaggle.com, where the team agreed to go with the following Data sets.

* Covid vaccination numbers
* GDP per Capita
* Population numbers

**Tools**

* **Excel**
* **Pandas**
* **SQL Postgres**

**Data Import and Cleaning**

After downloading the data sets from Kaggle**, Isha**, imported these Data sets into pandas, which was read, as a team we discussed some of the flaws of these data set, and realised some data sets has multiple columns for the same type of data from different sources there were some NAN Values within the data sets, some columns had more NaN values than others, that needed to be dropped.

**Data Merging**

**Tom** headed the Data Manipulation by merging the three sets of data frames for the team to achieve the Final data sets as we had discussed previously in the README File, On the Data transformation Section as follows. A correspondence spreadsheet was used to match the countries names between the data sets.

1. Country name

2. Population

3. GDP per capita

4. Number of vaccinations

5. % of population vaccinated (1 dose)

6. % of population vaccinated (2 doses)

7. Region

The final data frame head looked like below:

![Table

Description automatically generated]()

**Database Loading**

After achieving the data transformation to suit our goal**, Choki** led the creation of the database base, where as a team we had agreed to go with SQL Postgres. First a schema was done which was to create a table named Covid Vaccination report. To import Data into the SQL Database, we used Postgres SQL import csv file. We also used pandas to update the table.

Our final database had one table, with the following schemata:

|  |  |  |
| --- | --- | --- |
| **Column** | **Type** | **Notes** |
| Country | String (Varchar) | Not null, primary key |
| Population | Number (Integer) |  |
| GDP\_per\_capita | Number (Decimal) |  |
| Doses | Number (Decimal) |  |
| Pc\_1\_dose | Number (Decimal) |  |
| Pc\_2\_doses | Number (Decimal) |  |
| World\_region | String (Varchar) |  |

**Report Writing**

**Belinda** headed the report writing for the team, here we had to chronologically outline the steps, actions, tools, and decisions we took to attain out project objectives.