## Phase 3

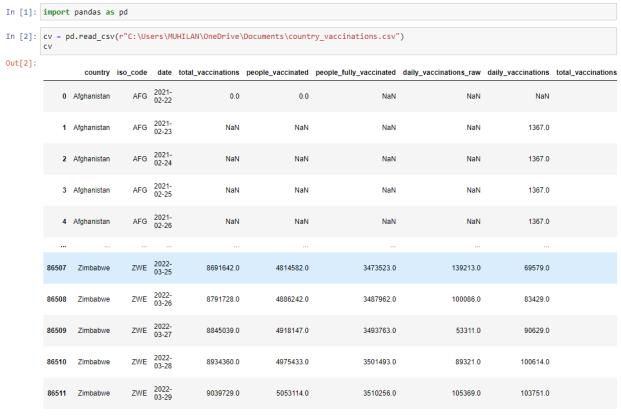
#### **COVID Vaccine Analysis**

To load and preprocess a COVID Vaccine dataset for analysis, you can follow these general steps using Python and Pandas. Make sure you have a COVID Vaccine dataset in a suitable format available.

**1. Import Libraries:** Start by importing the necessary Python libraries, including Pandas, to load and preprocess the dataset.

Import pandas as pd

2. Load the COVID-19 Dataset: Load the COVID Vaccine dataset into a Pandas DataFrame. You can use pd. read\_csv() for CSV files, but the method may vary depending on the file format.



86512 rows x 15 columns

**3. Data Inspection:** Before preprocessing, inspect the data to understand its structure and identify any potential issues.

```
In [3]: print(cv.head())
               country iso code
                                        date total vaccinations
                                                                  people vaccinated
          Afghanistan
                            AFG 2021-02-22
                                                             0.0
           Afghanistan
                            AFG 2021-02-23
                                                             NaN
                                                                                NaN
           Afghanistan
                            AFG 2021-02-24
                                                             NaN
                                                                                NaN
        3 Afghanistan
                            AFG
                                 2021-02-25
                                                             NaN
                                                                                NaN
           Afghanistan
                            AFG 2021-02-26
                                                             NaN
                                                                                NaN
           people_fully_vaccinated
                                     daily_vaccinations_raw
                                                             daily_vaccinations \
        0
                                                        NaN
                                                                            NaN
        1
                                NaN
                                                        NaN
                                                                         1367.0
        2
                               NaN
                                                        NaN
                                                                         1367.0
        3
                               NaN
                                                        NaN
                                                                         1367.0
        4
                               NaN
                                                        NaN
                                                                         1367.0
           total_vaccinations_per_hundred
                                           people_vaccinated_per_hundred
        0
                                       0.0
                                                                      0.0
        1
                                       NaN
                                                                      NaN
        2
                                       NaN
                                                                      NaN
        3
                                       NaN
                                                                      NaN
        4
                                       NaN
                                                                      NaN
           people_fully_vaccinated_per_hundred
                                                daily_vaccinations_per_million
        0
                                            NaN
        1
                                            NaN
                                                                           34.0
        2
                                            NaN
                                                                           34.0
        3
                                            NaN
                                                                           34.0
        4
                                            NaN
                                                                           34.0
                                                     vaccines \
           Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi...
           Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi...
        1
        2
           Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi...
           Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi...
          Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi...
                         source name
                                                 source website
          World Health Organization https://covid19.who.int/
           World Health Organization https://covid19.who.int/
           World Health Organization https://covid19.who.int/
           World Health Organization https://covid19.who.int/
        4 World Health Organization https://covid19.who.int/
```

## In [5]: print(cv.isnull().sum())

country	0
iso_code	0
date	0
total_vaccinations	42905
people_vaccinated	45218
people_fully_vaccinated	47710
daily_vaccinations_raw	51150
daily_vaccinations	299
total_vaccinations_per_hundred	42905
people_vaccinated_per_hundred	45218
people_fully_vaccinated_per_hundred	47710
daily_vaccinations_per_million	299
vaccines	0
source_name	0
source_website	0
dtype: int64	

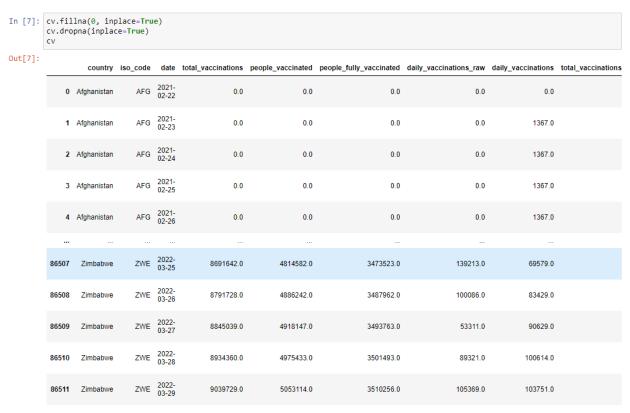
# In [6]: print(cv.dtypes)

country iso_code date total_vaccinations people_vaccinated people_fully_vaccinated daily_vaccinations_raw daily_vaccinations total_vaccinations_per_hundred people_vaccinated_per_hundred people_fully_vaccinated_per_hundred daily_vaccinations_per_million vaccines source_name	object object float64 float64 float64 float64 float64 float64 float64 object
source_name source_website dtype: object	object

## 4. Data Preprocessing:

## a. Data Cleaning:

 Handle missing values by either imputing them or removing rows with missing data.



86512 rows x 15 columns

#### b. Data Transformation:

• If necessary, transform the data to suit your analysis objectives. For instance, you may want to aggregate data by date or region.



483 rows x 3 columns

#### c. Data Filtering:

Filter the data to focus on a specific time frame or specific regions of interest.

```
In [13]: start date = '2021-06-01'
           end date = '2021-12-31'
           cv = cv[(cv['date'] >= start date) & (cv['date'] <= end date)]</pre>
Out[13]:
                        date
                                                                   country total_vaccinations
             181 2021-06-01 AfghanistanAlbaniaAlgeriaAndorraAngolaAnguilla...
                                                                                 2.016435e+09
             182 2021-06-02 AfghanistanAlbaniaAlgeriaAndorraAngolaAnguilla...
                                                                                 2.045128e+09
             183 2021-06-03 AfghanistanAlbaniaAlgeriaAndorraAngolaAnguilla...
                                                                                 2.081324e+09
                 2021-06-04 AfghanistanAlbaniaAlgeriaAndorraAngolaAnguilla...
                                                                                 2.041895e+09
                 2021-06-05 AfghanistanAlbaniaAlgeriaAndorraAngolaAnguilla...
                                                                                 2.142456e+09
            390 2021-12-27 AfghanistanAlbaniaAlgeriaAndorraAngolaAnguilla...
                                                                                 8.012818e+09
            391 2021-12-28 AfghanistanAlbaniaAlgeriaAndorraAngolaAnguilla...
                                                                                 8.308256e+09
            392 2021-12-29 AfghanistanAlbaniaAlgeriaAndorraAngolaAnguilla...
                                                                                 7.585792e+09
            393 2021-12-30 AfghanistanAlbaniaAlgeriaAndorraAngolaAnguilla...
                                                                                 8.070097e+09
                 2021-12-31 AfghanistanAlbaniaAlgeriaAndorraAngolaAnguilla...
                                                                                 7.627917e+09
```

214 rows x 3 columns

**5. Save the Preprocessed Data (Optional):** If you want to save the preprocessed data for future analysis, you can use Pandas to save it to a new CSV file.

```
In [14]: cv.to_csv('preprocessed_cv_data.csv', index=False)
```

These are the general steps to load and preprocess a COVID Vaccine dataset using Python and Pandas. Remember that the specific preprocessing steps and operations may vary depending on the structure of your dataset and your analysis objectives.