College code:4212
Register number: 421221243006

COVID VACCINES ANALYSIS
PHASE 3-DATA ANALYTICS WITH COGNOS GROUP 2
DEVELOPMENT PART 1

This phase involves designing of the steps that define in each phase of the previous documentation this involves importing necessary functions, data processing and so on. In this phase we have to begin our project by loading and preprocessing the dataset.

The IBM suggests using the jupyter notebook for loading and preprocess the dataset:

Here for this project title we need to define the loading of the libraries, understand the data and visualize the missing values. For this certain inputs are defined for this project.in this phase each of the input lines of the project is given as follows:

DATASET:

Dataset link:
https://www.kaggle.com/datasets/gpreda/covid-world-vaccination-progress/

The primary data source for this COVID-19 vaccine analysis project is the Kaggle dataset titled "COVID-19 World Vaccination Progress." It contains comprehensive and reliable information on global COVID-19 vaccination efforts, sourced from government health agencies, the World Health Organization, and other official records. The dataset includes key attributes like vaccine details, country-specific vaccination statistics, and population data. Preliminary data cleaning and preprocessing have been performed to ensure data quality. This dataset will serve as the foundation for subsequent analysis stages, including data exploration and modeling for insights into vaccine efficacy, distribution, and adverse effects.

dac-cva-phase3

October 17, 2023

[1]:
```
#1. Download the dataset from Kaggle: https://www.kaggle.com/datasets/gpreda/
 ↪covid-world-vaccination-progress/

#2.Import the required libraries:

import pandas as pd
```

[2]:
```
#3.Load the dataset into a Pandas DataFrame:

df = pd.read_csv('D:/country_vaccinations.csv')

df
```

[2]:

	country	iso_code	date	total_vaccinations	\
0	Afghanistan	AFG	2021-02-22	0.0	
1	Afghanistan	AFG	2021-02-23	NaN	
2	Afghanistan	AFG	2021-02-24	NaN	
3	Afghanistan	AFG	2021-02-25	NaN	
4	Afghanistan	AFG	2021-02-26	NaN	
…	…	…	…	…	
86507	Zimbabwe	ZWE	2022-03-25	8691642.0	
86508	Zimbabwe	ZWE	2022-03-26	8791728.0	
86509	Zimbabwe	ZWE	2022-03-27	8845039.0	
86510	Zimbabwe	ZWE	2022-03-28	8934360.0	
86511	Zimbabwe	ZWE	2022-03-29	9039729.0	

	people_vaccinated	people_fully_vaccinated	daily_vaccinations_raw	\
0	0.0	NaN	NaN	
1	NaN	NaN	NaN	
2	NaN	NaN	NaN	
3	NaN	NaN	NaN	
4	NaN	NaN	NaN	
…	…	…	…	
86507	4814582.0	3473523.0	139213.0	
86508	4886242.0	3487962.0	100086.0	
86509	4918147.0	3493763.0	53311.0	
86510	4975433.0	3501493.0	89321.0	

```
86511          5053114.0                3510256.0                 105369.0

       daily_vaccinations  total_vaccinations_per_hundred  \
0                     NaN                            0.00
1                  1367.0                             NaN
2                  1367.0                             NaN
3                  1367.0                             NaN
4                  1367.0                             NaN
…                     …                               …
86507             69579.0                           57.59
86508             83429.0                           58.25
86509             90629.0                           58.61
86510            100614.0                           59.20
86511            103751.0                           59.90

       people_vaccinated_per_hundred  people_fully_vaccinated_per_hundred  \
0                               0.00                                  NaN
1                                NaN                                  NaN
2                                NaN                                  NaN
3                                NaN                                  NaN
4                                NaN                                  NaN
…                                …                                    …
86507                          31.90                                23.02
86508                          32.38                                23.11
86509                          32.59                                23.15
86510                          32.97                                23.20
86511                          33.48                                23.26

       daily_vaccinations_per_million  \
0                                 NaN
1                                34.0
2                                34.0
3                                34.0
4                                34.0
…                                 …
86507                          4610.0
86508                          5528.0
86509                          6005.0
86510                          6667.0
86511                          6874.0

                                                 vaccines  \
0      Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi…
1      Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi…
2      Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi…
3      Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi…
4      Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi…
```

```
…                                                     …
86507  Oxford/AstraZeneca, Sinopharm/Beijing, Sinovac…
86508  Oxford/AstraZeneca, Sinopharm/Beijing, Sinovac…
86509  Oxford/AstraZeneca, Sinopharm/Beijing, Sinovac…
86510  Oxford/AstraZeneca, Sinopharm/Beijing, Sinovac…
86511  Oxford/AstraZeneca, Sinopharm/Beijing, Sinovac…

                    source_name  \
0      World Health Organization
1      World Health Organization
2      World Health Organization
3      World Health Organization
4      World Health Organization
…                            …
86507         Ministry of Health
86508         Ministry of Health
86509         Ministry of Health
86510         Ministry of Health
86511         Ministry of Health

                                       source_website
0                                https://covid19.who.int/
1                                https://covid19.who.int/
2                                https://covid19.who.int/
3                                https://covid19.who.int/
4                                https://covid19.who.int/
…                                                     …
86507  https://www.arcgis.com/home/webmap/viewer.html…
86508  https://www.arcgis.com/home/webmap/viewer.html…
86509  https://www.arcgis.com/home/webmap/viewer.html…
86510  https://www.arcgis.com/home/webmap/viewer.html…
86511  https://www.arcgis.com/home/webmap/viewer.html…

[86512 rows x 15 columns]
```

[3]:
```
#DATA PREPROCESSING:

#1.Handling Missing Values:

# Check for missing values
df.isnull().sum()
```

[3]:
```
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
```

[4]:
```
# Fill missing values with appropriate values (e.g., mean, median, or a␣
 ↪specific value)
df.fillna({'total_vaccinations': 0,
           'people_vaccinated': 0,
           'people_fully_vaccinated':0,
           'daily_vaccinations_raw':0,
           'daily_vaccinations':0,
           'total_vaccinations_per_hundred': 0,
           'people_vaccinated_per_hundred': 0,
           'people_fully_vaccinated_per_hundred':0,
           'daily_vaccinations_per_million':0}, inplace=True)

df.isnull().sum()
```

[4]:
```
country                                0
iso_code                               0
date                                   0
total_vaccinations                     0
people_vaccinated                      0
people_fully_vaccinated                0
daily_vaccinations_raw                 0
daily_vaccinations                     0
total_vaccinations_per_hundred         0
people_vaccinated_per_hundred          0
people_fully_vaccinated_per_hundred    0
daily_vaccinations_per_million         0
vaccines                               0
source_name                            0
source_website                         0
dtype: int64
```

[5]:
```
#2.Data Type Conversion:

df['date'] = pd.to_datetime(df['date'])
```

```
df
```

[5]:

	country	iso_code	date	total_vaccinations	people_vaccinated
0	Afghanistan	AFG	2021-02-22	0.0	0.0
1	Afghanistan	AFG	2021-02-23	0.0	0.0
2	Afghanistan	AFG	2021-02-24	0.0	0.0
3	Afghanistan	AFG	2021-02-25	0.0	0.0
4	Afghanistan	AFG	2021-02-26	0.0	0.0
…	…	…	…	…	…
86507	Zimbabwe	ZWE	2022-03-25	8691642.0	4814582.0
86508	Zimbabwe	ZWE	2022-03-26	8791728.0	4886242.0
86509	Zimbabwe	ZWE	2022-03-27	8845039.0	4918147.0
86510	Zimbabwe	ZWE	2022-03-28	8934360.0	4975433.0
86511	Zimbabwe	ZWE	2022-03-29	9039729.0	5053114.0

	people_fully_vaccinated	daily_vaccinations_raw	daily_vaccinations
0	0.0	0.0	0.0
1	0.0	0.0	1367.0
2	0.0	0.0	1367.0
3	0.0	0.0	1367.0
4	0.0	0.0	1367.0
…	…	…	…
86507	3473523.0	139213.0	69579.0
86508	3487962.0	100086.0	83429.0
86509	3493763.0	53311.0	90629.0
86510	3501493.0	89321.0	100614.0
86511	3510256.0	105369.0	103751.0

	total_vaccinations_per_hundred	people_vaccinated_per_hundred
0	0.00	0.00
1	0.00	0.00
2	0.00	0.00
3	0.00	0.00
4	0.00	0.00
…	…	…
86507	57.59	31.90
86508	58.25	32.38
86509	58.61	32.59
86510	59.20	32.97
86511	59.90	33.48

	people_fully_vaccinated_per_hundred	daily_vaccinations_per_million
0	0.00	0.0
1	0.00	34.0
2	0.00	34.0
3	0.00	34.0
4	0.00	34.0

```
…                                    …                                 …
86507                                23.02                              4610.0
86508                                23.11                              5528.0
86509                                23.15                              6005.0
86510                                23.20                              6667.0
86511                                23.26                              6874.0

                                                      vaccines  \
0       Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi…
1       Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi…
2       Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi…
3       Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi…
4       Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi…
…                                                     …
86507  Oxford/AstraZeneca, Sinopharm/Beijing, Sinovac…
86508  Oxford/AstraZeneca, Sinopharm/Beijing, Sinovac…
86509  Oxford/AstraZeneca, Sinopharm/Beijing, Sinovac…
86510  Oxford/AstraZeneca, Sinopharm/Beijing, Sinovac…
86511  Oxford/AstraZeneca, Sinopharm/Beijing, Sinovac…

                    source_name  \
0       World Health Organization
1       World Health Organization
2       World Health Organization
3       World Health Organization
4       World Health Organization
…                             …
86507          Ministry of Health
86508          Ministry of Health
86509          Ministry of Health
86510          Ministry of Health
86511          Ministry of Health

                                          source_website
0                                  https://covid19.who.int/
1                                  https://covid19.who.int/
2                                  https://covid19.who.int/
3                                  https://covid19.who.int/
4                                  https://covid19.who.int/
…                                                        …
86507  https://www.arcgis.com/home/webmap/viewer.html…
86508  https://www.arcgis.com/home/webmap/viewer.html…
86509  https://www.arcgis.com/home/webmap/viewer.html…
86510  https://www.arcgis.com/home/webmap/viewer.html…
86511  https://www.arcgis.com/home/webmap/viewer.html…

[86512 rows x 15 columns]
```

[6]:
```
#3.Handling Duplicates:

df.drop_duplicates(inplace=True)

df
```

[6]:

```
           country iso_code       date  total_vaccinations  people_vaccinated  \
0      Afghanistan      AFG 2021-02-22                 0.0                 0.0
1      Afghanistan      AFG 2021-02-23                 0.0                 0.0
2      Afghanistan      AFG 2021-02-24                 0.0                 0.0
3      Afghanistan      AFG 2021-02-25                 0.0                 0.0
4      Afghanistan      AFG 2021-02-26                 0.0                 0.0
…              …      …         …                    …                   …
86507     Zimbabwe      ZWE 2022-03-25           8691642.0           4814582.0
86508     Zimbabwe      ZWE 2022-03-26           8791728.0           4886242.0
86509     Zimbabwe      ZWE 2022-03-27           8845039.0           4918147.0
86510     Zimbabwe      ZWE 2022-03-28           8934360.0           4975433.0
86511     Zimbabwe      ZWE 2022-03-29           9039729.0           5053114.0

       people_fully_vaccinated  daily_vaccinations_raw  daily_vaccinations  \
0                          0.0                     0.0                 0.0
1                          0.0                     0.0              1367.0
2                          0.0                     0.0              1367.0
3                          0.0                     0.0              1367.0
4                          0.0                     0.0              1367.0
…                          …                       …                   …
86507                3473523.0                139213.0             69579.0
86508                3487962.0                100086.0             83429.0
86509                3493763.0                 53311.0             90629.0
86510                3501493.0                 89321.0            100614.0
86511                3510256.0                105369.0            103751.0

       total_vaccinations_per_hundred  people_vaccinated_per_hundred  \
0                                0.00                           0.00
1                                0.00                           0.00
2                                0.00                           0.00
3                                0.00                           0.00
4                                0.00                           0.00
…                                …                              …
86507                           57.59                          31.90
86508                           58.25                          32.38
86509                           58.61                          32.59
86510                           59.20                          32.97
86511                           59.90                          33.48

       people_fully_vaccinated_per_hundred  daily_vaccinations_per_million  \
0                                     0.00                             0.0
```

```
1                                    0.00                              34.0
2                                    0.00                              34.0
3                                    0.00                              34.0
4                                    0.00                              34.0
…                                    …                                 …
86507                                23.02                             4610.0
86508                                23.11                             5528.0
86509                                23.15                             6005.0
86510                                23.20                             6667.0
86511                                23.26                             6874.0

                                                      vaccines  \
0      Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi…
1      Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi…
2      Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi…
3      Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi…
4      Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi…
…                                                        …
86507  Oxford/AstraZeneca, Sinopharm/Beijing, Sinovac…
86508  Oxford/AstraZeneca, Sinopharm/Beijing, Sinovac…
86509  Oxford/AstraZeneca, Sinopharm/Beijing, Sinovac…
86510  Oxford/AstraZeneca, Sinopharm/Beijing, Sinovac…
86511  Oxford/AstraZeneca, Sinopharm/Beijing, Sinovac…

                    source_name  \
0      World Health Organization
1      World Health Organization
2      World Health Organization
3      World Health Organization
4      World Health Organization
…                            …
86507         Ministry of Health
86508         Ministry of Health
86509         Ministry of Health
86510         Ministry of Health
86511         Ministry of Health

                                          source_website
0                                https://covid19.who.int/
1                                https://covid19.who.int/
2                                https://covid19.who.int/
3                                https://covid19.who.int/
4                                https://covid19.who.int/
…                                                       …
86507  https://www.arcgis.com/home/webmap/viewer.html…
86508  https://www.arcgis.com/home/webmap/viewer.html…
86509  https://www.arcgis.com/home/webmap/viewer.html…
```

```
86510  https://www.arcgis.com/home/webmap/viewer.html…
86511  https://www.arcgis.com/home/webmap/viewer.html…

[86512 rows x 15 columns]
```

[7]:
```
#4.Data Scaling and Normalization:

from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
df[['total_vaccinations', 'people_vaccinated']] =scaler.
 ↪fit_transform(df[['total_vaccinations', 'people_vaccinated']])

df
```

[7]:
```
           country iso_code       date  total_vaccinations  people_vaccinated  \
0      Afghanistan      AFG 2021-02-22           -0.143704          -0.170046
1      Afghanistan      AFG 2021-02-23           -0.143704          -0.170046
2      Afghanistan      AFG 2021-02-24           -0.143704          -0.170046
3      Afghanistan      AFG 2021-02-25           -0.143704          -0.170046
4      Afghanistan      AFG 2021-02-26           -0.143704          -0.170046
…              …      …         …                   …                  …
86507     Zimbabwe      ZWE 2022-03-25           -0.089753          -0.073170
86508     Zimbabwe      ZWE 2022-03-26           -0.089132          -0.071728
86509     Zimbabwe      ZWE 2022-03-27           -0.088801          -0.071086
86510     Zimbabwe      ZWE 2022-03-28           -0.088247          -0.069933
86511     Zimbabwe      ZWE 2022-03-29           -0.087593          -0.068370

       people_fully_vaccinated  daily_vaccinations_raw  daily_vaccinations  \
0                          0.0                     0.0                 0.0
1                          0.0                     0.0              1367.0
2                          0.0                     0.0              1367.0
3                          0.0                     0.0              1367.0
4                          0.0                     0.0              1367.0
…                          …                       …                   …
86507                3473523.0                139213.0             69579.0
86508                3487962.0                100086.0             83429.0
86509                3493763.0                 53311.0             90629.0
86510                3501493.0                 89321.0            100614.0
86511                3510256.0                105369.0            103751.0

       total_vaccinations_per_hundred  people_vaccinated_per_hundred  \
0                                0.00                           0.00
1                                0.00                           0.00
2                                0.00                           0.00
3                                0.00                           0.00
4                                0.00                           0.00
…                                …                              …
```

```
86507                          57.59                          31.90
86508                          58.25                          32.38
86509                          58.61                          32.59
86510                          59.20                          32.97
86511                          59.90                          33.48

       people_fully_vaccinated_per_hundred  daily_vaccinations_per_million  \
0                                     0.00                             0.0
1                                     0.00                            34.0
2                                     0.00                            34.0
3                                     0.00                            34.0
4                                     0.00                            34.0
…                                        …                               …
86507                                23.02                          4610.0
86508                                23.11                          5528.0
86509                                23.15                          6005.0
86510                                23.20                          6667.0
86511                                23.26                          6874.0

                                                vaccines  \
0      Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi…
1      Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi…
2      Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi…
3      Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi…
4      Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi…
…                                                    …
86507  Oxford/AstraZeneca, Sinopharm/Beijing, Sinovac…
86508  Oxford/AstraZeneca, Sinopharm/Beijing, Sinovac…
86509  Oxford/AstraZeneca, Sinopharm/Beijing, Sinovac…
86510  Oxford/AstraZeneca, Sinopharm/Beijing, Sinovac…
86511  Oxford/AstraZeneca, Sinopharm/Beijing, Sinovac…

                    source_name  \
0      World Health Organization
1      World Health Organization
2      World Health Organization
3      World Health Organization
4      World Health Organization
…                            …
86507         Ministry of Health
86508         Ministry of Health
86509         Ministry of Health
86510         Ministry of Health
86511         Ministry of Health

                                      source_website
0                                https://covid19.who.int/
```

```
1                              https://covid19.who.int/
2                              https://covid19.who.int/
3                              https://covid19.who.int/
4                              https://covid19.who.int/
…                                                    …
86507  https://www.arcgis.com/home/webmap/viewer.html…
86508  https://www.arcgis.com/home/webmap/viewer.html…
86509  https://www.arcgis.com/home/webmap/viewer.html…
86510  https://www.arcgis.com/home/webmap/viewer.html…
86511  https://www.arcgis.com/home/webmap/viewer.html…

[86512 rows x 15 columns]
```

[8]:
```
#5.Encoding Categorical Variables:

df = pd.get_dummies(df, columns=['country', 'vaccines'], drop_first=True)


df
```

[8]:
```
      iso_code       date  total_vaccinations  people_vaccinated  \
0          AFG 2021-02-22           -0.143704          -0.170046
1          AFG 2021-02-23           -0.143704          -0.170046
2          AFG 2021-02-24           -0.143704          -0.170046
3          AFG 2021-02-25           -0.143704          -0.170046
4          AFG 2021-02-26           -0.143704          -0.170046
…          …          …                   …                  …
86507      ZWE 2022-03-25           -0.089753          -0.073170
86508      ZWE 2022-03-26           -0.089132          -0.071728
86509      ZWE 2022-03-27           -0.088801          -0.071086
86510      ZWE 2022-03-28           -0.088247          -0.069933
86511      ZWE 2022-03-29           -0.087593          -0.068370

       people_fully_vaccinated  daily_vaccinations_raw  daily_vaccinations  \
0                          0.0                     0.0                 0.0
1                          0.0                     0.0              1367.0
2                          0.0                     0.0              1367.0
3                          0.0                     0.0              1367.0
4                          0.0                     0.0              1367.0
…                          …                       …                   …
86507                3473523.0                139213.0             69579.0
86508                3487962.0                100086.0             83429.0
86509                3493763.0                 53311.0             90629.0
86510                3501493.0                 89321.0            100614.0
86511                3510256.0                105369.0            103751.0

       total_vaccinations_per_hundred  people_vaccinated_per_hundred  \
```

```
0                                  0.00                                  0.00
1                                  0.00                                  0.00
2                                  0.00                                  0.00
3                                  0.00                                  0.00
4                                  0.00                                  0.00
…                                     …                                     …
86507                             57.59                                 31.90
86508                             58.25                                 32.38
86509                             58.61                                 32.59
86510                             59.20                                 32.97
86511                             59.90                                 33.48

       people_fully_vaccinated_per_hundred  …  \
0                                     0.00  …
1                                     0.00  …
2                                     0.00  …
3                                     0.00  …
4                                     0.00  …
…                                        …  …
86507                                23.02  …
86508                                23.11  …
86509                                23.15  …
86510                                23.20  …
86511                                23.26  …

       vaccines_Oxford/AstraZeneca, Sputnik V vaccines_Pfizer/BioNTech  \
0                                           0                         0
1                                           0                         0
2                                           0                         0
3                                           0                         0
4                                           0                         0
…                                         …                         …
86507                                       0                         0
86508                                       0                         0
86509                                       0                         0
86510                                       0                         0
86511                                       0                         0

      vaccines_Pfizer/BioNTech, Sinopharm/Beijing  \
0                                               0
1                                               0
2                                               0
3                                               0
4                                               0
…                                             …
86507                                           0
86508                                           0
```

```
86509                                                 0
86510                                                 0
86511                                                 0

       vaccines_Pfizer/BioNTech, Sinopharm/Beijing, Sputnik V  \
0                                                        0
1                                                        0
2                                                        0
3                                                        0
4                                                        0
…                                                        …
86507                                                    0
86508                                                    0
86509                                                    0
86510                                                    0
86511                                                    0

       vaccines_Pfizer/BioNTech, Sinovac  \
0                                      0
1                                      0
2                                      0
3                                      0
4                                      0
…                                      …
86507                                  0
86508                                  0
86509                                  0
86510                                  0
86511                                  0

       vaccines_Pfizer/BioNTech, Sinovac, Turkovac  \
0                                                 0
1                                                 0
2                                                 0
3                                                 0
4                                                 0
…                                                 …
86507                                             0
86508                                             0
86509                                             0
86510                                             0
86511                                             0

       vaccines_Pfizer/BioNTech, Sputnik V  \
0                                        0
1                                        0
2                                        0
```

```
3                                          0
4                                          0
…                                          …
86507                                      0
86508                                      0
86509                                      0
86510                                      0
86511                                      0

       vaccines_QazVac, Sinopharm/Beijing, Sputnik V  \
0                                                    0
1                                                    0
2                                                    0
3                                                    0
4                                                    0
…                                                    …
86507                                                0
86508                                                0
86509                                                0
86510                                                0
86511                                                0

       vaccines_Sinopharm/Beijing  vaccines_Sinopharm/Beijing, Sputnik V
0                               0                                      0
1                               0                                      0
2                               0                                      0
3                               0                                      0
4                               0                                      0
…                               …                                      …
86507                           0                                      0
86508                           0                                      0
86509                           0                                      0
86510                           0                                      0
86511                           0                                      0

[86512 rows x 318 columns]
```

[9]:
```
#6.Feature Selection:

df = df[['iso_code', 'date', 'total_vaccinations', 'people_vaccinated']]

df
```

[9]:
```
      iso_code       date  total_vaccinations  people_vaccinated
0          AFG 2021-02-22           -0.143704          -0.170046
1          AFG 2021-02-23           -0.143704          -0.170046
2          AFG 2021-02-24           -0.143704          -0.170046
```

```
3          AFG 2021-02-25           -0.143704           -0.170046
4          AFG 2021-02-26           -0.143704           -0.170046
…          …          …                 …                   …
86507      ZWE 2022-03-25           -0.089753           -0.073170
86508      ZWE 2022-03-26           -0.089132           -0.071728
86509      ZWE 2022-03-27           -0.088801           -0.071086
86510      ZWE 2022-03-28           -0.088247           -0.069933
86511      ZWE 2022-03-29           -0.087593           -0.068370

[86512 rows x 4 columns]
```

[10]:
```
#7.Date-Based Features:

df['month'] = df['date'].dt.month
df['day_of_week'] = df['date'].dt.dayofweek

#8.Outlier Detection and Handling:

import numpy as np
from scipy import stats

# Detect outliers using the IQR method
Q1 = df['total_vaccinations'].quantile(0.25)
Q3 = df['total_vaccinations'].quantile(0.75)
IQR = Q3 - Q1
lower_bound = Q1 - 1.5 * IQR
upper_bound = Q3 + 1.5 * IQR

# Remove outliers
df = df[(df['total_vaccinations'] >= lower_bound) & (df['total_vaccinations']␣
 ↪<= upper_bound)]

df
```

```
C:\Users\PC\AppData\Local\Temp\ipykernel_8540\1495377907.py:3:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df['month'] = df['date'].dt.month
C:\Users\PC\AppData\Local\Temp\ipykernel_8540\1495377907.py:4:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df['day_of_week'] = df['date'].dt.dayofweek
```

[10]:
```
      iso_code       date  total_vaccinations  people_vaccinated  month  \
0          AFG 2021-02-22           -0.143704          -0.170046      2
1          AFG 2021-02-23           -0.143704          -0.170046      2
2          AFG 2021-02-24           -0.143704          -0.170046      2
3          AFG 2021-02-25           -0.143704          -0.170046      2
4          AFG 2021-02-26           -0.143704          -0.170046      2
…          …          …                 …                  …      …
86507      ZWE 2022-03-25           -0.089753          -0.073170      3
86508      ZWE 2022-03-26           -0.089132          -0.071728      3
86509      ZWE 2022-03-27           -0.088801          -0.071086      3
86510      ZWE 2022-03-28           -0.088247          -0.069933      3
86511      ZWE 2022-03-29           -0.087593          -0.068370      3

       day_of_week
0                0
1                1
2                2
3                3
4                4
…              …
86507            4
86508            5
86509            6
86510            0
86511            1

[70909 rows x 6 columns]
```

[11]:
```
#9.Data Splitting:

from sklearn.model_selection import train_test_split

# Replace 'actual_target_column_name' with the correct column name
X = df.drop('total_vaccinations', axis=1)  # Features
y = df['people_vaccinated']  # Target variable

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,␣
 ↪random_state=42)


#10.Save Preprocessed Data:

df.to_csv('preprocessed_data.csv', index=False)
```

```
df
```

[11]:

```
      iso_code       date  total_vaccinations  people_vaccinated  month  \
0          AFG 2021-02-22           -0.143704          -0.170046      2
1          AFG 2021-02-23           -0.143704          -0.170046      2
2          AFG 2021-02-24           -0.143704          -0.170046      2
3          AFG 2021-02-25           -0.143704          -0.170046      2
4          AFG 2021-02-26           -0.143704          -0.170046      2
…          …          …                 …                  …    …
86507      ZWE 2022-03-25           -0.089753          -0.073170      3
86508      ZWE 2022-03-26           -0.089132          -0.071728      3
86509      ZWE 2022-03-27           -0.088801          -0.071086      3
86510      ZWE 2022-03-28           -0.088247          -0.069933      3
86511      ZWE 2022-03-29           -0.087593          -0.068370      3

       day_of_week
0                0
1                1
2                2
3                3
4                4
…              …
86507            4
86508            5
86509            6
86510            0
86511            1

[70909 rows x 6 columns]
```

[ ]: