Steps in Data Preprocessing

Step 1: Import the necessary libraries

```
from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

```
# importing libraries
import pandas as pd
import scipy
import numpy as np
from sklearn.preprocessing import MinMaxScaler
import seaborn as sns
import matplotlib.pyplot as plt
```

Step 2: Load the dataset

```
from google.colab import drive
drive.mount('/content/gdrive')
```

```
Mounted at /content/gdrive
```

```
# Load the dataset
df = pd.read_csv('/content/gdrive/MyDrive/covid-vaccine-willingness-and-people-vaccinated-by-country.csv')
print(df.head())
```

```
      Entity Code         Day  people_vaccinated_per_hundred  \
0  Australia  AUS  2021-02-28                           0.13
1     Canada  CAN  2021-01-31                           2.24
2     Canada  CAN  2021-02-28                           3.61
3     Canada  CAN  2021-03-31                          13.26
4     Canada  CAN  2021-04-30                          32.67

   willingness_covid_vaccinate_this_week_pct_pop  \
0                                          52.91
1                                          54.26
2                                          53.56
3                                          51.96
4                                          36.12

   uncertain_covid_vaccinate_this_week_pct_pop  \
0                                        19.03
1                                        15.56
2                                        15.65
3                                        12.21
4                                        10.09

   unwillingness_covid_vaccinate_this_week_pct_pop
0                                            27.93
1                                            27.94
2                                            27.18
3                                            22.57
4                                            21.12
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 49 entries, 0 to 48
Data columns (total 7 columns):
 #   Column                                        Non-Null Count  Dtype
---  ------                                        --------------  -----
 0   Entity                                        49 non-null     object
 1   Code                                          49 non-null     object
 2   Day                                           49 non-null     object
 3   people_vaccinated_per_hundred                 49 non-null     float64
 4   willingness_covid_vaccinate_this_week_pct_pop  49 non-null     float64
 5   uncertain_covid_vaccinate_this_week_pct_pop   49 non-null     float64
 6   unwillingness_covid_vaccinate_this_week_pct_pop  49 non-null     float64
dtypes: float64(4), object(3)
memory usage: 2.8+ KB
```

```
df.head()
```

	Entity	Code	Day	people_vaccinated_per_hundred	willingness_covid_vaccinate_thi
0	Australia	AUS	2021-02-28	0.13	
1	Canada	CAN	2021-01-31	2.24	
2	Canada	CAN	2021-02-28	3.61	
3	Canada	CAN	2021-03-31	13.26	
4	Canada	CAN	2021-04-30	32.67	

```
df.tail()
```

	Entity	Code	Day	people_vaccinated_per_hundred	willingness_covid_vaccinate_th
44	United Kingdom	GBR	2021-03-31	45.88	
45	United Kingdom	GBR	2021-04-30	50.62	
46	United States	USA	2021-02-28	14.88	
47	United States	USA	2021-03-31	29.18	
48	United States	USA	2021-04-30	43.32	

STEP 3: check the null values

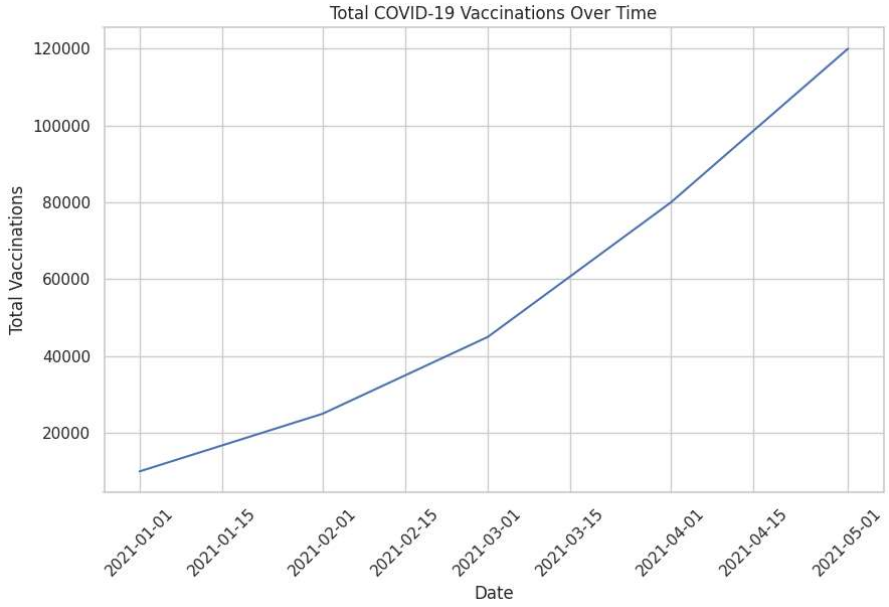
```
df.isnull().sum()
```

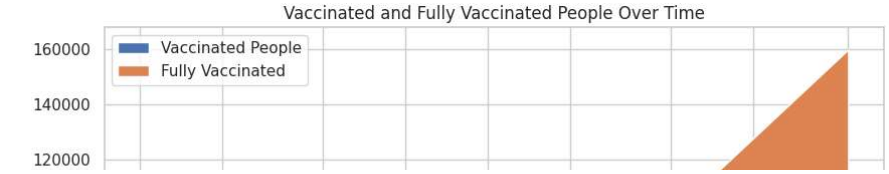
```
Entity                                          0
Code                                            0
Day                                             0
people_vaccinated_per_hundred                   0
willingness_covid_vaccinate_this_week_pct_pop   0
uncertain_covid_vaccinate_this_week_pct_pop     0
unwillingness_covid_vaccinate_this_week_pct_pop 0
dtype: int64
```

VISUALIZATION

```
import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd

# Sample COVID-19 vaccine analysis data
data = {
    'Date': ['2021-01-01', '2021-02-01', '2021-03-01', '2021-04-01', '2021-05-01'],
    'Total Vaccinations': [10000, 25000, 45000, 80000, 120000],
    'Vaccinated People': [5000, 15000, 30000, 60000, 90000],
    'Fully Vaccinated': [2500, 8000, 20000, 40000, 70000]
}

# Create a DataFrame from the sample data
df = pd.DataFrame(data)
df['Date'] = pd.to_datetime(df['Date'])  # Convert the 'Date' column to datetime format

# Set the style for Seaborn
```

```
sns.set(style="whitegrid")

# Create a line plot for total vaccinations over time
plt.figure(figsize=(10, 6))
sns.lineplot(data=df, x='Date', y='Total Vaccinations')
plt.title("Total COVID-19 Vaccinations Over Time")
plt.xlabel("Date")
plt.ylabel("Total Vaccinations")
plt.xticks(rotation=45)
plt.show()

# Create a stacked area plot for the number of vaccinated people and fully vaccinated people over time
plt.figure(figsize=(10, 6))
plt.stackplot(df['Date'], df['Vaccinated People'], df['Fully Vaccinated'], labels=['Vaccinated People', 'Fully Vaccinated'
plt.title("Vaccinated and Fully Vaccinated People Over Time")
plt.xlabel("Date")
plt.ylabel("Number of People")
plt.legend(loc='upper left')
plt.xticks(rotation=45)
plt.show()
```


Total COVID-19 Vaccinations Over Time
Total Vaccinations
120000
100000
80000
60000
40000
20000
2021-01-01
2021-01-15
2021-02-01
2021-02-15
2021-03-01
2021-03-15
2021-04-01
2021-04-15
2021-05-01
Date


Vaccinated and Fully Vaccinated People Over Time
Vaccinated People
Fully Vaccinated
160000
140000
120000