1. COVID Analysis

Import libraries ¶

In [1]:
```
import pandas as pd
import numpy as np
from sklearn.impute import SimpleImputer as si
```

Loading dataset

In [2]:
```
df = pd.read_csv('country_vaccinations.csv')
df.head()
```

Out[2]:

	country	iso_code	date	total_vaccinations	people_vaccinated	people_fully_vaccinated	daily_vaccinat
0	Afghanistan	AFG	2021-02-22	0.0	0.0	NaN	
1	Afghanistan	AFG	2021-02-23	NaN	NaN	NaN	
2	Afghanistan	AFG	2021-02-24	NaN	NaN	NaN	
3	Afghanistan	AFG	2021-02-25	NaN	NaN	NaN	
4	Afghanistan	AFG	2021-02-26	NaN	NaN	NaN	

In [3]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 86512 entries, 0 to 86511
Data columns (total 15 columns):
 #   Column                                Non-Null Count  Dtype  
---  ------                                --------------  -----  
 0   country                               86512 non-null  object 
 1   iso_code                              86512 non-null  object 
 2   date                                  86512 non-null  object 
 3   total_vaccinations                    43607 non-null  float64
 4   people_vaccinated                     41294 non-null  float64
 5   people_fully_vaccinated               38802 non-null  float64
 6   daily_vaccinations_raw                35362 non-null  float64
 7   daily_vaccinations                    86213 non-null  float64
 8   total_vaccinations_per_hundred        43607 non-null  float64
 9   people_vaccinated_per_hundred         41294 non-null  float64
 10  people_fully_vaccinated_per_hundred   38802 non-null  float64
 11  daily_vaccinations_per_million        86213 non-null  float64
 12  vaccines                              86512 non-null  object 
 13  source_name                           86512 non-null  object 
 14  source_website                        86512 non-null  object 
dtypes: float64(9), object(6)
memory usage: 9.9+ MB
```

Checking null values

In [4]:
```
df.isnull().sum()
```

Out[4]:
```
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
```

Dealing with Null Values

CASE 1 - Remove null values

1. Using Function

In [5]:
```
df_new =  df.dropna()

df_new.isnull().sum()
```

Out[5]:
```
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
```

In [6]:
```
df_new.to_csv('WP - 10 outputs/Remove_null_values_function.csv',index=False)
```

2. From Scratch

In [7]:
```
def isnan(data):
    return type(data) == float and not float('-inf') < data < float('inf')

data = df.values

new_data = []

for rows in data:
    flag = False

    for element in rows:
        flag = isnan(element)
        if flag: break

    if flag: continue

    new_data.append(rows)

new_df = pd.DataFrame(new_data,columns = df.columns)
```

In [8]:
```
new_df.isnull().sum()
```

Out[8]:
```
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
```

In [9]:
```
new_df.to_csv('WP - 10 outputs/Remove_null_values_scratch.csv',index=False)
```

CASE 2 - Deal null values using Simple Imputer

In [10]:
```
# collecting columns which present some null value
total_vaccinations                  = df.total_vaccinations
people_vaccinated                   = df.people_vaccinated
people_fully_vaccinated             = df.people_fully_vaccinated
daily_vaccinations_raw              = df.daily_vaccinations_raw
daily_vaccinations                  = df.daily_vaccinations
total_vaccinations_per_hundred      = df.total_vaccinations_per_hundred
people_vaccinated_per_hundred       = df.people_vaccinated_per_hundred
people_fully_vaccinated_per_hundred = df.people_fully_vaccinated_per_hundred
daily_vaccinations_per_million      = df.daily_vaccinations_per_million

# Converting pandas series to numpy array of each column
total_vaccinations                  = total_vaccinations.values
people_vaccinated                   = people_vaccinated.values
people_fully_vaccinated             = people_fully_vaccinated.values
daily_vaccinations_raw              = daily_vaccinations_raw.values
daily_vaccinations                  = daily_vaccinations.values
total_vaccinations_per_hundred      = total_vaccinations_per_hundred.values
people_vaccinated_per_hundred       = people_vaccinated_per_hundred.values
people_fully_vaccinated_per_hundred = people_fully_vaccinated_per_hundred.values
daily_vaccinations_per_million      = daily_vaccinations_per_million.values

# Reshape in 2D array of each column
total_vaccinations                  = total_vaccinations.reshape(-1, 1)
people_vaccinated                   = people_vaccinated.reshape(-1, 1)
people_fully_vaccinated             = people_fully_vaccinated.reshape(-1, 1)
daily_vaccinations_raw              = daily_vaccinations_raw.reshape(-1, 1)
daily_vaccinations                  = daily_vaccinations.reshape(-1, 1)
total_vaccinations_per_hundred      = total_vaccinations_per_hundred.reshape(-1, 1)
people_vaccinated_per_hundred       = people_vaccinated_per_hundred.reshape(-1, 1)
people_fully_vaccinated_per_hundred = people_fully_vaccinated_per_hundred.reshape(-1, 1)
daily_vaccinations_per_million      = daily_vaccinations_per_million.reshape(-1, 1)
```

1. Replace wiith mean values

In [11]:
```
mean = si(strategy='mean')

# Creating new dataframe and adding column which are not null
new_df = pd.DataFrame()

new_df['country'] = df.country
new_df['iso_code'] = df.iso_code
new_df['date'] = df.date
```

Fit and Transformation with null values

In [12]:
```
# total_vaccinations

mean.fit(total_vaccinations )

new_df['total_vaccinations'] = mean.transform(total_vaccinations).reshape(1,-1)[0]

#-------------------------------------------------------------------------------------

# people_vaccinated

mean.fit(people_vaccinated )

new_df['people_vaccinated'] = mean.transform(people_vaccinated).reshape(1,-1)[0]

#-------------------------------------------------------------------------------------

#people_fully_vaccinated

mean.fit(people_fully_vaccinated )

new_df['people_fully_vaccinated'] = mean.transform(people_fully_vaccinated).reshape(1,-1

#-------------------------------------------------------------------------------------

#daily_vaccinations_raw

mean.fit(daily_vaccinations_raw )

new_df['daily_vaccinations_raw'] = mean.transform(daily_vaccinations_raw).reshape(1,-1)[

#-------------------------------------------------------------------------------------

# daily_vaccinations

mean.fit(daily_vaccinations )

new_df['daily_vaccinations'] = mean.transform(daily_vaccinations).reshape(1,-1)[0]

#-------------------------------------------------------------------------------------

# total_vaccinations_per_hundred

mean.fit(total_vaccinations_per_hundred )

new_df['total_vaccinations_per_hundred'] = mean.transform(total_vaccinations_per_hundred

#-------------------------------------------------------------------------------------

# people_vaccinated_per_hundred

mean.fit(people_vaccinated_per_hundred )

new_df['people_vaccinated_per_hundred'] = mean.transform(people_vaccinated_per_hundred).

#-------------------------------------------------------------------------------------

# people_fully_vaccinated_per_hundred
```

```
mean.fit(people_fully_vaccinated_per_hundred )

new_df['people_fully_vaccinated_per_hundred'] = mean.transform(people_fully_vaccinated_p

#-------------------------------------------------------------------------------------

# daily_vaccinations_per_million

mean.fit(daily_vaccinations_per_million )

new_df['daily_vaccinations_per_million'] = mean.transform(daily_vaccinations_per_millior
```

In [13]:
```
# Adding remain columns in new dataframe
new_df['vaccines']       = df.vaccines
new_df['source_name']    = df.source_name
new_df['source_website'] = df.source_website

del mean
#checking null values
new_df.isnull().sum()
```

Out[13]:
```
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
```

In [14]:
```
new_df.to_csv('WP - 10 outputs/Replace_null_values_imputer_mean.csv',index=False)
```

2. Replace with median values

In [15]:
```
median = si(strategy='median')

# Creating new dataframe and adding column which are not null
new_df = pd.DataFrame()

new_df['country'] = df.country
new_df['iso_code'] = df.iso_code
new_df['date'] = df.date
```

Fit and Transformation with null values

In [16]:
```
# total_vaccinations

median.fit(total_vaccinations )

new_df['total_vaccinations'] = median.transform(total_vaccinations).reshape(1,-1)[0]

#-------------------------------------------------------------------------------------

# people_vaccinated

median.fit(people_vaccinated )

new_df['people_vaccinated'] = median.transform(people_vaccinated).reshape(1,-1)[0]

#-------------------------------------------------------------------------------------

#people_fully_vaccinated

median.fit(people_fully_vaccinated )

new_df['people_fully_vaccinated'] = median.transform(people_fully_vaccinated).reshape(1,

#-------------------------------------------------------------------------------------

#daily_vaccinations_raw

median.fit(daily_vaccinations_raw )

new_df['daily_vaccinations_raw'] = median.transform(daily_vaccinations_raw).reshape(1,-1

#-------------------------------------------------------------------------------------

# daily_vaccinations

median.fit(daily_vaccinations )

new_df['daily_vaccinations'] = median.transform(daily_vaccinations).reshape(1,-1)[0]

#-------------------------------------------------------------------------------------

# total_vaccinations_per_hundred

median.fit(total_vaccinations_per_hundred )

new_df['total_vaccinations_per_hundred'] = median.transform(total_vaccinations_per_hundr

#-------------------------------------------------------------------------------------

# people_vaccinated_per_hundred

median.fit(people_vaccinated_per_hundred )

new_df['people_vaccinated_per_hundred'] = median.transform(people_vaccinated_per_hundred

#-------------------------------------------------------------------------------------

# people_fully_vaccinated_per_hundred
```

```
median.fit(people_fully_vaccinated_per_hundred )

new_df['people_fully_vaccinated_per_hundred'] = median.transform(people_fully_vaccinated

#-------------------------------------------------------------------------------------

# daily_vaccinations_per_million

median.fit(daily_vaccinations_per_million )

new_df['daily_vaccinations_per_million'] = median.transform(daily_vaccinations_per_milli
```

In [17]:
```
# Adding remain columns in new dataframe
new_df['vaccines']       = df.vaccines
new_df['source_name']    = df.source_name
new_df['source_website'] = df.source_website

del median
#checking null values
new_df.isnull().sum()
```

Out[17]:
```
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
```

In [18]:
```
new_df.to_csv('WP - 10 outputs/Replace_null_values_imputer_median.csv',index=False)
```

3. Replace with most frequent values

In [19]:
```
mode = si(strategy='most_frequent')

# Creating new dataframe and adding column which are not null
new_df = pd.DataFrame()

new_df['country'] = df.country
new_df['iso_code'] = df.iso_code
new_df['date'] = df.date
```

Fit and Transformation with null values

In [20]:
```
# total_vaccinations

mode.fit(total_vaccinations )

new_df['total_vaccinations'] = mode.transform(total_vaccinations).reshape(1,-1)[0]

#-------------------------------------------------------------------------------------

# people_vaccinated

mode.fit(people_vaccinated )

new_df['people_vaccinated'] = mode.transform(people_vaccinated).reshape(1,-1)[0]

#-------------------------------------------------------------------------------------

#people_fully_vaccinated

mode.fit(people_fully_vaccinated )

new_df['people_fully_vaccinated'] = mode.transform(people_fully_vaccinated).reshape(1,-1

#-------------------------------------------------------------------------------------

#daily_vaccinations_raw

mode.fit(daily_vaccinations_raw )

new_df['daily_vaccinations_raw'] = mode.transform(daily_vaccinations_raw).reshape(1,-1)[

#-------------------------------------------------------------------------------------

# daily_vaccinations

mode.fit(daily_vaccinations )

new_df['daily_vaccinations'] = mode.transform(daily_vaccinations).reshape(1,-1)[0]

#-------------------------------------------------------------------------------------

# total_vaccinations_per_hundred

mode.fit(total_vaccinations_per_hundred )

new_df['total_vaccinations_per_hundred'] = mode.transform(total_vaccinations_per_hundred

#-------------------------------------------------------------------------------------

# people_vaccinated_per_hundred

mode.fit(people_vaccinated_per_hundred )

new_df['people_vaccinated_per_hundred'] = mode.transform(people_vaccinated_per_hundred).

#-------------------------------------------------------------------------------------

# people_fully_vaccinated_per_hundred
```

```
mode.fit(people_fully_vaccinated_per_hundred )

new_df['people_fully_vaccinated_per_hundred'] = mode.transform(people_fully_vaccinated_p

#-------------------------------------------------------------------------------------

# daily_vaccinations_per_million

mode.fit(daily_vaccinations_per_million )

new_df['daily_vaccinations_per_million'] = mode.transform(daily_vaccinations_per_millior
```

In [21]:
```
# Adding remain columns in new dataframe
new_df['vaccines']       = df.vaccines
new_df['source_name']    = df.source_name
new_df['source_website'] = df.source_website

del mode
#checking null values
new_df.isnull().sum()
```

Out[21]:
```
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
```

In [22]:
```
new_df.to_csv('WP - 10 outputs/Replace_null_values_imputer_mode.csv',index=False)
```

4. Replace with zero

In [23]:
```
const = si(strategy ='constant', fill_value = 0)

# Creating new dataframe and adding column which are not null
new_df = pd.DataFrame()

new_df['country'] = df.country
new_df['iso_code'] = df.iso_code
new_df['date'] = df.date
```

Fit and Transformation with null values

In [24]:
```
# total_vaccinations

const.fit(total_vaccinations )

new_df['total_vaccinations'] = const.transform(total_vaccinations).reshape(1,-1)[0]

#-------------------------------------------------------------------------------------

# people_vaccinated

const.fit(people_vaccinated )

new_df['people_vaccinated'] = const.transform(people_vaccinated).reshape(1,-1)[0]

#-------------------------------------------------------------------------------------

#people_fully_vaccinated

const.fit(people_fully_vaccinated )

new_df['people_fully_vaccinated'] = const.transform(people_fully_vaccinated).reshape(1,-

#-------------------------------------------------------------------------------------

#daily_vaccinations_raw

const.fit(daily_vaccinations_raw )

new_df['daily_vaccinations_raw'] = const.transform(daily_vaccinations_raw).reshape(1,-1)

#-------------------------------------------------------------------------------------

# daily_vaccinations

const.fit(daily_vaccinations )

new_df['daily_vaccinations'] = const.transform(daily_vaccinations).reshape(1,-1)[0]

#-------------------------------------------------------------------------------------

# total_vaccinations_per_hundred

const.fit(total_vaccinations_per_hundred )

new_df['total_vaccinations_per_hundred'] = const.transform(total_vaccinations_per_hundre

#-------------------------------------------------------------------------------------

# people_vaccinated_per_hundred

const.fit(people_vaccinated_per_hundred )

new_df['people_vaccinated_per_hundred'] = const.transform(people_vaccinated_per_hundred)

#-------------------------------------------------------------------------------------

# people_fully_vaccinated_per_hundred
```

```
const.fit(people_fully_vaccinated_per_hundred )

new_df['people_fully_vaccinated_per_hundred'] = const.transform(people_fully_vaccinated_

#-------------------------------------------------------------------------------------

# daily_vaccinations_per_million

const.fit(daily_vaccinations_per_million )

new_df['daily_vaccinations_per_million'] = const.transform(daily_vaccinations_per_millic
```

In [25]:
```
# Adding remain columns in new dataframe
new_df['vaccines']       = df.vaccines
new_df['source_name']    = df.source_name
new_df['source_website'] = df.source_website

del const
#checking null values
new_df.isnull().sum()
```

Out[25]:
```
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
```

In [26]:
```
new_df.to_csv('WP - 10 outputs/Replace_null_values_imputer_zero.csv',index=False)
```

CASE 3 - Deal null values with some function

1. Replace with mean

In [27]:
```
# collecting columns which present some null value
mean_total_vaccinations                  = df.total_vaccinations.mean()
mean_people_vaccinated                   = df.people_vaccinated.mean()
mean_people_fully_vaccinated             = df.people_fully_vaccinated.mean()
mean_daily_vaccinations_raw              = df.daily_vaccinations_raw.mean()
mean_daily_vaccinations                  = df.daily_vaccinations.mean()
mean_total_vaccinations_per_hundred      = df.total_vaccinations_per_hundred.mean()
mean_people_vaccinated_per_hundred       = df.people_vaccinated_per_hundred.mean()
mean_people_fully_vaccinated_per_hundred = df.people_fully_vaccinated_per_hundred.mean()
mean_daily_vaccinations_per_million      = df.daily_vaccinations_per_million.mean()

# Creating new dataframe and adding all column
new_df = pd.DataFrame(df.values , columns = df.columns)
```

In [28]:
```
new_df.total_vaccinations.fillna(mean_total_vaccinations, inplace=True)

new_df.people_vaccinated.fillna(mean_people_vaccinated, inplace=True)

new_df.people_fully_vaccinated.fillna(mean_people_fully_vaccinated, inplace=True)

new_df.daily_vaccinations_raw.fillna(mean_daily_vaccinations_raw, inplace=True)

new_df.daily_vaccinations.fillna(mean_daily_vaccinations, inplace=True)

new_df.total_vaccinations_per_hundred.fillna(mean_total_vaccinations_per_hundred, inplac

new_df.people_vaccinated_per_hundred.fillna(mean_people_vaccinated_per_hundred, inplace=

new_df.people_fully_vaccinated_per_hundred.fillna(mean_people_fully_vaccinated_per_hundr

new_df.daily_vaccinations_per_million.fillna(mean_daily_vaccinations_per_million, inplac


#checking null values
new_df.isnull().sum()
```

Out[28]:
```
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
```

In [29]:
```
new_df.to_csv('WP - 10 outputs/Replace_null_values_function_mean.csv',index=False)
```

2. Replace with median

In [30]:
```
# collecting columns which present some null value
median_total_vaccinations                  = df.total_vaccinations.median()
median_people_vaccinated                   = df.people_vaccinated.median()
median_people_fully_vaccinated             = df.people_fully_vaccinated.median()
median_daily_vaccinations_raw              = df.daily_vaccinations_raw.median()
median_daily_vaccinations                  = df.daily_vaccinations.median()
median_total_vaccinations_per_hundred      = df.total_vaccinations_per_hundred.median()
median_people_vaccinated_per_hundred       = df.people_vaccinated_per_hundred.median()
median_people_fully_vaccinated_per_hundred = df.people_fully_vaccinated_per_hundred.medi
median_daily_vaccinations_per_million      = df.daily_vaccinations_per_million.median()

# Creating new dataframe and adding all column
new_df = pd.DataFrame(df.values , columns = df.columns)
```

In [31]:
```
new_df.total_vaccinations.fillna(median_total_vaccinations, inplace=True)

new_df.people_vaccinated.fillna(median_people_vaccinated, inplace=True)

new_df.people_fully_vaccinated.fillna(median_people_fully_vaccinated, inplace=True)

new_df.daily_vaccinations_raw.fillna(median_daily_vaccinations_raw, inplace=True)

new_df.daily_vaccinations.fillna(median_daily_vaccinations, inplace=True)

new_df.total_vaccinations_per_hundred.fillna(median_total_vaccinations_per_hundred, inpl

new_df.people_vaccinated_per_hundred.fillna(median_people_vaccinated_per_hundred, inplac

new_df.people_fully_vaccinated_per_hundred.fillna(median_people_fully_vaccinated_per_hur

new_df.daily_vaccinations_per_million.fillna(median_daily_vaccinations_per_million, inpl


#checking null values
new_df.isnull().sum()
```

Out[31]:
```
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
```

In [32]:
```
new_df.to_csv('WP - 10 outputs/Replace_null_values_function_median.csv',index=False)
```

3. Replace with mode

In [33]:
```
# collecting columns which present some null value
mode_total_vaccinations                  = df.total_vaccinations.mean()
mode_people_vaccinated                   = df.people_vaccinated.mean()
mode_people_fully_vaccinated             = df.people_fully_vaccinated.mean()
mode_daily_vaccinations_raw              = df.daily_vaccinations_raw.mean()
mode_daily_vaccinations                  = df.daily_vaccinations.mean()
mode_total_vaccinations_per_hundred      = df.total_vaccinations_per_hundred.mean()
mode_people_vaccinated_per_hundred       = df.people_vaccinated_per_hundred.mean()
mode_people_fully_vaccinated_per_hundred = df.people_fully_vaccinated_per_hundred.mean()
mode_daily_vaccinations_per_million      = df.daily_vaccinations_per_million.mean()

# Creating new dataframe and adding all column
new_df = pd.DataFrame(df.values , columns = df.columns)
```

In [34]:
```
new_df.total_vaccinations.fillna(mode_total_vaccinations, inplace=True)

new_df.people_vaccinated.fillna(mode_people_vaccinated, inplace=True)

new_df.people_fully_vaccinated.fillna(mode_people_fully_vaccinated, inplace=True)

new_df.daily_vaccinations_raw.fillna(mode_daily_vaccinations_raw, inplace=True)

new_df.daily_vaccinations.fillna(mode_daily_vaccinations, inplace=True)

new_df.total_vaccinations_per_hundred.fillna(mode_total_vaccinations_per_hundred, inplac

new_df.people_vaccinated_per_hundred.fillna(mode_people_vaccinated_per_hundred, inplace=

new_df.people_fully_vaccinated_per_hundred.fillna(mode_people_fully_vaccinated_per_hundr

new_df.daily_vaccinations_per_million.fillna(mode_daily_vaccinations_per_million, inplac


#checking null values
new_df.isnull().sum()
```

Out[34]:
```
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
```

In [35]:
```
new_df.to_csv('WP - 10 outputs/Replace_null_values_function_mode.csv',index=False)
```

4. Replace with zero

In [36]:
```
# Creating new dataframe and adding all column
new_df = pd.DataFrame(df.values , columns = df.columns)
```

In [37]:
```
new_df.total_vaccinations.fillna(0, inplace=True)

new_df.people_vaccinated.fillna(0, inplace=True)

new_df.people_fully_vaccinated.fillna(0, inplace=True)

new_df.daily_vaccinations_raw.fillna(0, inplace=True)

new_df.daily_vaccinations.fillna(0, inplace=True)

new_df.total_vaccinations_per_hundred.fillna(0, inplace=True)

new_df.people_vaccinated_per_hundred.fillna(0, inplace=True)

new_df.people_fully_vaccinated_per_hundred.fillna(0, inplace=True)

new_df.daily_vaccinations_per_million.fillna(0, inplace=True)


#checking null values
new_df.isnull().sum()
```

Out[37]:
```
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
```

In [38]:
```
new_df.to_csv('WP - 10 outputs/Replace_null_values_function_zero.csv',index=False)
```

CASE 4 - Deal null values from scratch

In [39]:
```
nan_columns = [ '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',      ]
```

1.Replace with mean

In [40]:
```
def mean( data ):

    tot   = 0
    count = 0
    for i in data:
        # condition for nan values
        if not float('-inf') <= i <= float('inf'): continue

        tot   += i
        count += 1

    return tot//count

data = df.values

means = []

for i in nan_columns:

    val   = df[i].values
    means.append( mean(val) )

for index in range(len(data)):
    for columns in range(len(data[0])):
        val = data[index][columns]
        if type(val) == float and not float('-inf') <= val <= float('inf'):
            data[index][columns] = means[columns-3]

new_df = pd.DataFrame(data,columns = df.columns)
new_df.isnull().sum()
```

Out[40]:
```
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
```

In [41]:
```
new_df.to_csv('WP - 10 outputs/Replace_null_values_scratch_mean.csv',index=False)
```

1.Replace with median

In [42]:
```
def median( data ):


    data = [ i for i in data if float('-inf') <= i <= float('inf')]

    data.sort()

    length = len(data)

    if length % 2 == 1:
        return data[ length//2 + 1 ]
    else:
        median_1 = data[ length//2     ]
        median_2 = data[ length//2 + 1 ]
        return (median_1 + median_2) // 2

data = df.values

medians = []

for i in nan_columns:

    val   = df[i].values
    medians.append( median(val) )

for index in range(len(data)):
    for columns in range(len(data[0])):
        val = data[index][columns]
        if type(val) == float and not float('-inf') <= val <= float('inf'):
            data[index][columns] = medians[columns-3]

new_df = pd.DataFrame(data,columns = df.columns)
new_df.isnull().sum()
```

Out[42]:
```
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
```

In [43]:
```
new_df.to_csv('WP - 10 outputs/Replace_null_values_scratch_median.csv',index=False)
```