# subtask 1 - importing libraries

import pandas as pd

import numpy as np

# subtask 2 - Dataset url

# Name - WeatherAUS

# URL - https://www.kaggle.com/danielfesalbon/covid-19-global-reports-early-march-2022

# local machine relative address - /covid\_19\_clean\_complete\_2022.csv/

#subtask 3 - Loading dataset

df = pd.read\_csv("covid\_19\_clean\_complete\_2022.csv")

df.head()

df.shape

df.drop('Province/State', axis='columns',inplace=True)

df.head()

# subtask 4 - data preprocessing - detecting NaN values and using describe() function

df.isna().any()

df.describe()

# shape of dataset (dimensions)

df.shape

# subtask 5 - data formatting and normalization

df.dtypes

df['Date'] = df['Date'].astype('datetime64[ns]')

df['Country/Region'] = df['Country/Region'].astype('string')

df['Lat'] = df['Lat'].astype('string')

df.dtypes

# subtask 6 - handling categorical values

# dropping the categorical variable column

df = df.drop(['WHO Region','Country/Region'], axis='columns')

df.head()