***PHASE:03***

PROJECT:Covid-19 cases

***INTRODUCTION:***

***In this phase,the given dataset can be cleansed and***

***the data can be splitted and virtualize the data using***

***datascience***

***import pandas as pd***

***import pandas as pd***

***import numpy as np***

***import matplotlib.pyplot as plt***

***import seaborn as sns***

***import geopandas as gpd***

***#data cleansing***

***data = pd.read\_csv('Covid\_19\_cases4.CSV')***

***data = data.drop\_duplicates()***

***data['Day'] = data['month'].int('desired\_data\_type')***

***data = data[(data['Cases'] >= lower\_threshold) & (data[‘death'] <= upper\_threshold)]***

***from sklearn.preprocessing import StandardScaler, MinMaxScaler***

***scaler = StandardScaler()***

***data['year'] = scaler.fit\_transform(data[[‘day']])***

***data['cases'] = data['countriws&territories '].str.strip()***

***data[‘cases'] = data['cases'].str.upper()***

***data.to\_csv('cleaned\_20140711.CSV', index=False)***

***#data visualization***

***Day = [5,10,15,20,25,3]***

***Month = [2,4,6,8,10,12]***

***plt.plot(Day,Month)***

***plt.Xlabel("Day")***

***plt.Ylabel("Month")***

***plt.show()***

***#geospatial analysis***

***gdf = gpd.read\_file('Covid\_19\_cases4.CSV')***

***print(gdf.head())***

***print(gdf.crs)***

***gdf.plot()***

***plt.title('Geospatial Data Visualization')***

***plt.show()***

***#data mining***

***from sklearn.datasets import load\_iris***

***from sklearn.model\_selection import train\_test\_split***

***from sklearn.tree import DecisionTreeClassifier***

***from sklearn.metrics import accuracy\_score***

***data = load\_iris()***

***Day = data.data***

***Day\_train, Day\_test, Year\_train, Year\_test = train\_test\_split(Day, Month , test\_size=0.3, random\_state=42)***

***classifier = DecisionTreeClassifier()***

***classifier.fit(Day\_train, Month\_train)***

***Year \_pred = classifier.predict(Year\_test)***

***accuracy = accuracy\_score(Year\_test, Year\_pred)***

***print("Accuracy:", accuracy)***