Data Science with Python

Day 01 - Activity - 02

01. Answer the following questions using the “countries of the world.csv” and “geo coordinates.csv” files

I. Use Pandas to read the "countries of the world.csv" file and display the first five rows of the DataFrame.
II. Filter the rows of the DataFrame to show only the countries with a population greater than 100 million
III. Filter the columns of the DataFrame to show only the name, region, and population columns
IV. Sort the country details DataFrame in descending order based on Population.
V. Rename the "Pop. Density (per sq. mi.)" column to "Population Density" and the "GDP ($ per capita)" column to "GDP" in the DataFrame
VI. Group the DataFrame by the "Region" column and get the maximum population values of each region
VII. Merge the "countries of the world.csv" file with the "geo coordinates.csv" file and load it into a DataFrame

02. Use the “countries of the world.csv” dataset to answer the following questions

I. How can we check if there are any missing values in the DataFrame?
II. Identify the percentages of missing values in each column of the DataFrame?
III. Drop the rows from the DataFrame that have missing values in the "Agriculture" column?
IV. Fill the missing values in the "Climate" column with the mode value.
V. Identify the number of duplicate rows in the DataFrame.
VI. Drop the above identified duplicate rows from the DataFrame.
VII. Replace the value "EASTERN EUROPE" in the "Region" column with "EUROPE (EASTERN)"?
VIII. Use a lambda function to create a new column in the DataFrame called ‘Population Category’, that categorizes countries based on the ’population’ as below.

- High: population greater than 100 million
- Medium: population between 10 million and 100 million
- Low: population less than 10 million