**# Q1) PART A- REPORT:**

* Imported the pandas library and reads data from an Excel file, targeted specifically rows that contain relevant energy information.
* Skiped the first 17 rows and limited the read to 227 rows while selecting only the specified columns, which include 'Country', 'Energy Supply','Energy Supply per Capita', and '% Renewable'.
* Assigned the read data to a DataFrame named energy, with clear column names for easier data manipulation and analysis.

**# Q1) PART B- REPORT:**

* Replaced all instances of "..." in the energy DataFrame with pd.NA, marked them as missing values.
* Converted the 'Energy Supply' column to a numeric type, ensured that any invalid values (such as "...") are coerced into NaN and multiplied valid entries by 1,000,000 to adjust units.
* Prepared the energy DataFrame for further analysis by addressing missing data and ensuring numeric consistency in the 'Energy Supply' column.

**# Q1) PART C- REPORT:**

* Renamed specific countries in the 'Country' column of the energy DataFrame, replaced long-form names with shorter.
* Cleaned the 'Country' column by removing any text within parentheses, keeped only the main country names.
* Standardized country names for more consistent and accurate analysis in the energy DataFrame.

**# Q1) PART D- REPORT:**

* Read the world\_bank.csv file into a DataFrame, skiped the first 4 rows to remove unnecessary headers and focus on the relevant data.
* Standardized country names by renaming entries like "Korea, Rep." to "South Korea" and "Iran, Islamic Rep." to "Iran" for consistency.
* Updated the Country Name column to ensure the renamed countries are reflected correctly, prepared the data for further analysis.

**# Q1) PART F- REPORT:**

* Read GDP data from a CSV file, merged it with two previously loaded datasets (ScimEn and energy) using the country name as the key, created unified dataset merged\_data.
* Filtered and selected the top 15 countries based on their rank using the nsmallest function, set the country name as the index for further analysis.
* Created filtered dataset final\_data by selecting specific columns related to rank, documents, energy supply, and GDP data from 2006 to 2015.

**# Q2)- REPORT:**

* Calculated the average GDP for each country from 2006 to 2015 by selecting the corresponding columns, stored the result in a new column avgGDP within the dataset.
* Sorted the data in descending order, with countries that have the highest average GDP appearing at the top.
* Converted the sorted average GDP values into a DataFrame avgGDP\_df for easier display and analysis.

**# Q2) VISUALIZATION- REPORT:**

* Calculated the average GDP for each country from 2006 to 2015 by taking the mean of the GDP columns and stored the results in a new column, avgGDP.
* Selected the top 15 countries, and their average GDP, visualized using a bar chart.
* Customized the bar chart with a title, axis labels, and rotated x-axis labels to improve readability,displayed the top 15 countries by their average GDP.

**# Q3) REPORT:**

* Calculated the average GDP for each country from 2006 to 2015 and added this value as a new column, avgGDP, to the dataset.
* Identified the country with the 6th largest average GDP after sorting the countries by their average GDP in descending order.
* Calculated and returned the difference between the GDP in 2015 and 2006 for the 6th largest country by average GDP.
* Computed the average GDP for each country between 2006 and 2015 and added this as a new column, avgGDP, to the final\_data DataFrame.
* Sorted the countries by their average GDP in descending order and retrieved the GDP data for the 6th largest country, calculated the change in GDP from 2006 to 2015.
* Created a bar chart to visually represent the GDP values for the years 2006 and 2015 for the identified country, enhanced understanding of the GDP change, and displayed the plot.

**# Q4) REPORT:**

* Computed the mean energy supply per capita from the final\_data DataFrame and printed the result, provided insight into the average energy supply levels across the dataset.
* Cleaned the data by removing rows with missing values in the 'Energy Supply per Capita' column, and added a new column indicating whether each entry is above or below the mean value.
* Calculated and provided the minimum and maximum energy supply per capita values from the cleaned dataset, allowed for quick identification of the range of energy supply levels among the countries.

**# Q4) VISUALIZATION- REPORT:**

* Calculated the mean energy supply per capita from the final\_data DataFrame, Created a new DataFrame, final\_data\_clean,by removing any rows with missing values in the 'Energy Supply per Capita' column.
* Added a new column, 'Above Mean',to final\_data\_clean, indicating whether each country's energy supply per capita is above the calculated mean. Used visualization different colors ('purple' for above mean and 'yellow' for below)to represent this distinction.
* Generated a bar chart to display energy supply per capita by country, with a green dashed line indicating the mean value.

**# Q5) REPORT:**

* Defined a function, plot\_max\_renewable\_country, which calculates the country with the maximum percentage of renewable energy from the given dataset (final\_data) and visualized this information using a bar chart.
* Displayed the percentage of renewable energy for each country by bar chart, highlighted the country with the highest percentage in orange. Indicated the maximum renewable percentage by a horizontal line for enhancing visual clarity.
* Printed the country with the maximum renewable percentage and its corresponding value, provided users with immediate insights alongside the visual representation.

**# Q6) REPORT:**

* Calculated the self-citation ratio for each country in the dataset (final\_data), identified the country with the highest ratio, returned both the country and its ratio.
* Generated a bar chart displaying the self-citation ratios for all countries, with the highest ratio highlighted in green.
* Printed the country with the highest self-citation ratio and its value, while also annotating the bar chart with the ratio,provided clear insights to the user.

**# Q7) REPORT:**

* Calculated and estimated the population for each country using energy supply data, then sorted the countries to identify the third most populous country.
* Visualized the estimated population for all countries in a bar plot, highlighted the third most populous country in a different color.
* Displayed the third most populous country both on the chart.

**# Q8) REPORT:**

* Calculated the correlation between "Energy Supply per Capita" and "Citable Documents per Capita" by estimating the population and dividing citable documents by population.
* Visualized the relationship between these two variables using a scatter plot, with energy supply on the x-axis and citable documents on the y-axis.
* Printed calculated correlation value , indicated the strength and direction of the relationship between the two metrics.

**Correlation: 0.751516595603683**

**# Q9) REPORT:**

* Sorted countries by rank, selected the top 15, and calculated the median renewable energy percentage.
* Assigned a value of 1 if a country's renewable energy percentage is greater than or equal to the median and 0 otherwise,
* created a new column called HighRenew.
* Returned a series that indicates whether each country has a renewable energy percentage higher or lower than the median among the top 15 ranked countries.
* Visualized the number of countries with high versus low renewable energy by bar chart, labeled as "Low Renewable" and "High Renewable."

**# Q10) REPORT:**

* Imported energy data from an Excel file and GDP data from a CSV file, applied necessary transformations such as renaming columns, handling missing values, and standardizing country names for consistency.
* Merged the energy and GDP datasets with a third dataset (ScimEn) based on country names, calculated estimated populations using energy supply and energy supply per capita, and mapped countries to their respective continents.
* Grouped the cleaned and merged data by continent, calculated and displayed key statistics (size, sum, mean, and standard deviation) for estimated populations across different continents.

**# Q10) VISUALIZATION- REPORT:**

* Created a bar chart to visualize the estimated population statistics by continent, utilized a figure size of 10x6 inches and a sky-blue color for the bars.
* Represented the sample size (number of countries) for each continent, with axes labeled for clarity, rotated the x-axis tick labels for better readability.
* Enhanced the visualization by adding value annotations on top of each bar, displayed the exact sample size, included a legend for reference, ensured the chart is informative and easy to interpret.

**# Q11) REPORT:**

* Defined a dictionary to map countries to their respective continents and then added a new column to the DataFrame (merged\_df) that categorizes each country by continent using this mapping.
* Segmented the % Renewable data into five bins, created a new column (% Renewable Bin) that classifies countries based on their renewable energy percentage.
* Grouped the data by continent and the renewable energy bins, counted the number of countries in each group, resulted in a structured overview of renewable energy distribution across continents.

**# Q11) VISUALIZATION- REPORT:**

* Added continent information based on a mapping dictionary (ContinentDict) and segmented the % Renewable values into five categories.
* Grouped the data by continent and renewable energy bins, counted the number of countries in each category. Converted this grouped data into a DataFrame (grouped\_df) for visualization.
* Created a bar chart to display the number of countries per continent across the renewable energy bins. Represented each continent in the chart, allowed for an easy comparison of renewable energy distribution among different continents. Clarified the chart with adding labeled axes, a title, and a legend.
* Highlighted the distribution of renewable energy adoption across different continents and how some continents may have more countries with higher or lower renewable energy percentages.

**# Q12) REPORT:**

* Checked existing of columns ('Energy Supply', 'Energy Supply per Capita', and 'Country') in the merged\_data DataFrame. Raising a KeyError if any of these columns are missing.
* Calculated the 'Estimated Population' by dividing 'Energy Supply' by 'Energy Supply per Capita', converted the result to numeric format while handling any non-numeric values by replacing them with NaN.
* Formated the 'Estimated Population' values to include a thousands separator and eight decimal places. Created a new DataFrame (PopEst) with these formatted values, used country names as the index, and returned this DataFrame.

**# Q12) VISUALIZATION- REPORT:**

* Calculated the 'Estimated Population' by dividing 'Energy Supply' by 'Energy Supply per Capita'.
* Replaced non-numeric entries with NaN and formatted them for better readability.
* Constructed a new DataFrame (PopEst) containing the estimated population values with country names as the index, allowing for easy reference and further analysis.
* Generated a bar chart displaying the top 10 countries by estimated population. Enhanced the chart by formatting the values appropriately.