Introduction To Data Science

Assignment One

Delivery Instructions:

1. Cheaters will be graded by ***–ve points, Don't copy any code from anywhere.***
2. Due Date: **April 20, 2022**
3. No late submission will be accepted.
4. upload your python code as well as screenshots of the outcomes in zip file **YourGroup-Student1ID-Assignment1.zip**

Question One

Load the energy data from the given file Energy Indicators.xls, which is a set of United Nations indicators of energy supply and renewable electricity production for the year 2013, into a DataFrame with the variable name Energy. make sure to exclude the footer and header information from the datafile. The first two columns are unnecessary, so you should get rid of them and you should change the column labels so that the columns are:

['Country', 'Energy Supply', 'Energy Supply per Capita', 'Renewable']

Then Perform the following manipulations:

1. Convert Energy Supply to gigajoules (Note: there are 1,000,000 gigajoules in a petajoule). For all countries which have missing data (e.g. data with "...") make sure this is reflected as np.NaN values.

2. Rename the following list of countries :

```
"Republic of Korea": "South Korea",

"United States of America": "United States",

"United Kingdom of Great Britain and Northern Ireland": "United Kingdom",

"China, Hong Kong Special Administrative Region": "Hong Kong"
```

3. There are several countries with numbers and/or parenthesis in their name. Be sure to remove these, e.g. 'Bolivia (Plurinational State of)' should be 'Bolivia'. 'Switzerland17' should be 'Switzerland'

Introduction To Data Science

Assignment One

Question Two

load the GDP data from the given file world_bank.csv, which is a csv containing countries' GDP from 1960 to 2015 from the World Bank. Call this DataFrame **GDP**

1. Rename the following list of countries (for use in later questions):

```
"Korea, Rep.": "South Korea",
"Iran, Islamic Rep.": "Iran",
"Hong Kong SAR,
China": "Hong Kong"
```

Question Three

1. Join the two datasets: GDP and Energy into a new dataset (using the intersection of country names). Use only the last 6 years (2010-2015) of GDP data.The index of this DataFrame should be the name of the country.
2. Create a function that returns the top 15 countries in terms of average GDP over the previous six years. This function should produce an 'averageGDP' Series containing 15 nations and their average GDP ordered in **descending** order.
3. Write a function to return the mean energy supply per capita.
4. Write a function to return the country that has the minimum “Renewable” and the value.