Code Documentation

Assignment-3

-By Gursimar Singh Bedi

Task in hand

To create a pivot table from a dataset and plot a heat map of the same.

Procedure

# Installing Libraries

**In [1]:pip install** **pandas**

**In [2]:pip install** **seaborn**

**In [3]:pip install** **matplotlib**

Pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series. We will be the data-frame data structure to read the dataset in hand.

Seaborn is a Python data visualization library based on [matplotlib](https://matplotlib.org/). It provides a high-level interface for drawing attractive and informative statistical graphics. Seaborn will simplify the process of creating a heat map.

# Importing Libraries

**In [1]:import** **pandas as pd**

**In [2]:import** **seaborn as sns**

**In [3]:import** **matplotlib.pyplot as plt**

pd and sns are optional aliases for both the libraries and help reducing the length of the code. Plt is being used to plot the heat map and view the results.

# Importing the Dataset

**In [4]: df=** **pd.read\_csv("https://raw.githubusercontent.com/resbaz/r-novice- gapminder-files/master/data/gapminder-FiveYearData.csv")**

Using pandas we will store the dataset in a variable named df.

# Analyzing the Dataset

**In [5]: print(df.head())**

**In [6]: print(df.describe())**

**In [7]: print(df.isnull.sum())**

head() returns the first five rows of the dataset. We will use this to find out the names of the column and get a brief idea about the dataset.

Describe() is used to return a statistical summary of all columns in a dataset. We can use this to gain insights and look for anomalies.

.isnull() checks for null values and sum() returns the column-wise sum, hence in amalgamation the command returns the number of null values in each row. This was used to ensure that the data was relatively clean

# Creating the pivot table

A Pivot Table is used to summarise, sort, reorganise, group, count, total or average data stored in a table. The objective here is to create a pivot table using the dataset with years along the x-axis and countries along the y-axis. The life expectancy will be the data being entered in the pivot table.

**In [8]: table= pd.pivot\_table(df, values=‘lifeExp’, index=[‘country’], columns=[‘year’],)**

The first argument is the dataset that needs to be explored for the creation of the pivot table. Index are the columns of the data set to be used as rows and values include the columns to use as data for the formation of pivot table.

# Plotting the Heat Map

**In [9]: sns.heatmap(table)**

**In [10]: plt.show()**

Using the Seaborn library a heat map of any table can easily be created. plt.show() is used to view the plot in a separate window

# Result

