**Assignment 3**

**Name: Akshaya Balaji**

**Email ID:** [akshayavb99@gmail.com](mailto:akshayavb99@gmail.com)

**Question:**

Create a HeatMap Plot for the following plot with Seaborn:

1. Download the DataSet from the GitHub link: <https://raw.githubusercontent.com/resbaz/r-novice-gapminder-files/master/data/gapminder-FiveYearData.csv> to solve this problem.

The dataset has 5 columns namely: **country,year,pop,continent,lifeExp**and**gdpPercap**

2. Create a pivot table dataframe with **year** along x-axes, **continent** along y-axes and **lifeExp** filled within cells.

3. Plot a heatmap using seaborn for the pivot table that was just created.

**Solution:**

The given dataset is first loaded as a dataframe by using the read\_csv() function. The describe() is used on the newly created dataframe to find the mean value of population which will be helpful later during the plotting of the heatmap. Next, the dataframe is checked for NaN values and these values are removed, if any. Originally the dataframe is consisting of 1704 rows and 142 columns which is too much data to be plotted on the heatmap. In such cases, we can group them by the X-axis and Y-axis parameters we need to plot the heatmap, which in this case are the columns year and continent. We also take the mean of the values of the dataframe (wherever possible), in order to reduce the number of datapoints to be plotted, while at the same time, ensure that the overall data with respect to a given continent for a given year can be represented.

The pivot table is created with ‘year’ along the X-axis, ‘continent’ along the Y-axis and the data values of the cells equal to the ‘lifeExp’ column of the dataframe. This table has 5 rows and 12 columns, which provides us with enough reduction in the number of datapoints without compromising accuracy.

To plot the heatmap, we first set the figure size of plotting area as 5x12 in order to clearly display all the datapoints being plotted. Then the function heatmap() is invoked with the pivot table as an argument, while setting the other parameters for display as well. The savefig() function also saves a copy of the heatmap as an image with .png extension after the heatmap is plotted. The show() function of the matplotlib library is then used to display the heatmap during runtime.

**Code:**

**import** pandas **as** pd  
**import** matplotlib.pyplot **as** plt  
**import** seaborn **as** sns  
  
df=pd.read\_csv(**'D:\PycharmProjects\Python Practice\gapminder-FiveYearData.csv'**)  
*#print(df.describe()) #Checking to see if csv file is parsed successfully  
  
#Checking for data duplication and NaN values*df=df.dropna(how=**'all'**)  
df1=df.groupby([**'year'**,**'continent'**],as\_index=**False**).mean()  
*#print(df1)  
  
#Creating the pivot table*pvt\_tbl=df1.pivot(**'continent'**,**'year'**,**'lifeExp'**)  
print(pvt\_tbl)  
  
*#Creating heatmap using seaborn*plt.figure(figsize=[5,12])  
fig=sns.heatmap(pvt\_tbl,fmt=**'f'**,linewidth=0.1,center=60,square=**True**).get\_figure().savefig(**'assignment\_output.png'**)  
plt.show()

**Output:**

