120A3051
Shreya Idate
Batch: E3

EXPERIMENT NO. 4

Aim : Data Visualization / Exploratory Data Analysis for the selected data set using Seaborn.

a) Create a bar graph, contingency table using any 2 variables.

b) Create normalized histogram

c) Describe what this graphs and tables indicate.

Theory:

Seaborn is a data visualization library built on top of matplotlib and closely integrated with pandas data structures in Python. Visualization is the central part of Seaborn which helps in exploration and understanding of data.
Seaborn offers the following functionalities:

- Dataset oriented API to determine the relationship between variables.
- Automatic estimation and plotting of linear regression plots.
- It supports high-level abstractions for multi-plot grids.
- Visualizing univariate and bivariate distribution. Using Seaborn we can plot wide.

Varieties of plots like:

- Distribution Plots
- Pie Chart & Bar Chart
- Scatter Plots
- Pair Plots
- Heat maps

Data Visualization:

Data visualization is the graphical representation of information and data. The data visualization is one of the most important fundamental toolkits of a data scientist. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.
Visualizations are the easiest way to analyze and absorb information. Visuals help to easily understand the complex problem. They help in identifying patterns, relationships, and outliers in data. It helps in understanding business problems better and quickly. It helps to build a compelling story based on visuals. Insights gathered from the visuals help in building. strategies for businesses. It is also a precursor to many high-level data analyses for Exploratory Data Analysis(EDA) and Machine Learning(ML).

Exploratory Data Analysis:

Exploratory Data Analysis (EDA) is an analysis approach that identifies general patterns in the data. These patterns include outliers and features of the data that might be unexpected.

EDA is an important first step in any data analysis. Understanding where outliers occur and how variables are related can help one design statistical analyses that yield meaningful results. In biological monitoring data, sites are likely to be affected by multiple stressors.
Thus, initial explorations of stressor correlations are critical before one attempt to relate stressor variables to biological response variables. EDA can provide insights into candidate causes that should be included in a causal assessment.

Bar Graph

A bar plot or bar chart is a graph that represents the category of data with rectangular bars with lengths and heights that is proportional to the values which they represent. The bar plots can be plotted horizontally or vertically. A bar chart describes the comparisons between the discrete categories. One of the axes of the plot represents the specific categories being compared, while the other axis represents the measured values corresponding to those categories.

Contingency Table

Contingency Table is one of the techniques for exploring two or more variables. It is basically a tally of counts between two or more categorical variables. Contingency Tables give clear correlation values between those variables, thus making it much more useful to understand the data for further information extraction.

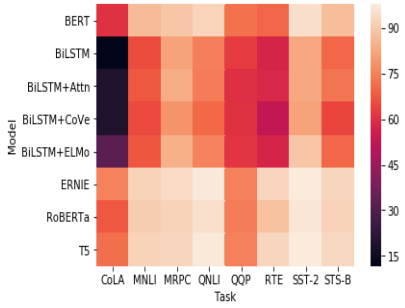
Normalized Histogram

Normalized Histogram: A histogram is a frequency distribution that depicts the frequencies of different elements in a dataset. This graph is generally used to study frequencies and determine how the values are distributed in a dataset. Normalization of histogram refers to mapping the frequencies of a dataset between the range [0, 1] both inclusive.

Program:

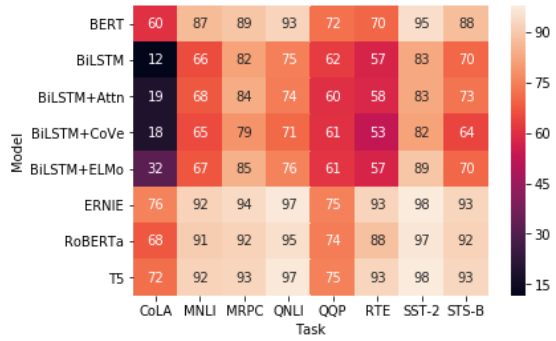
```
In [9]: import seaborn as sns
        %matplotlib inline

        glue = sns.load_dataset("glue").pivot("Model","Task","Score")
        sns.heatmap(glue)
```

```
Out[9]: <matplotlib.axes._subplots.AxesSubplot at 0x2436a45cb70>
```



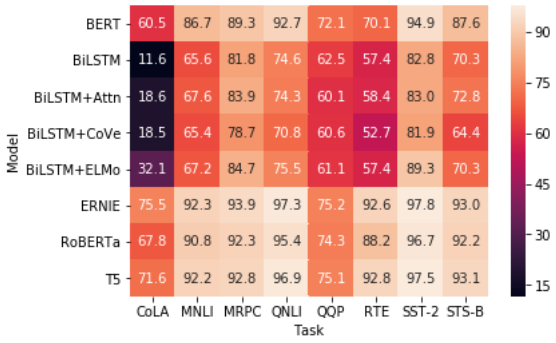
```
In [10]: sns.heatmap(glue,annot=True)
```

```
Out[10]: <matplotlib.axes._subplots.AxesSubplot at 0x2436a4c98d0>
```



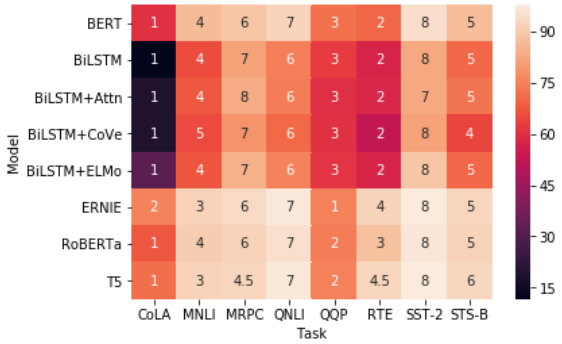

```
In [11]: sns.heatmap(glue, annot=True,fmt=".1f")
```

```
Out[11]: <matplotlib.axes._subplots.AxesSubplot at 0x2436a45c5c0>
```

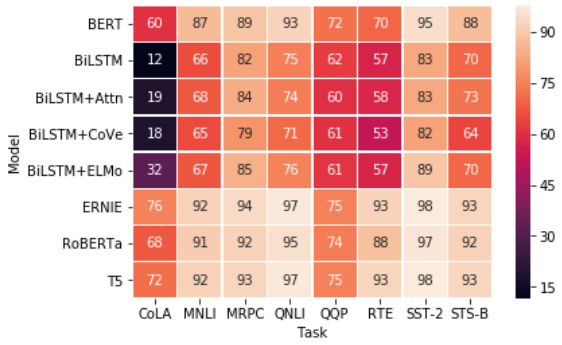



```
In [12]: sns.heatmap(glue, annot=glue.rank(axis="columns"))
```

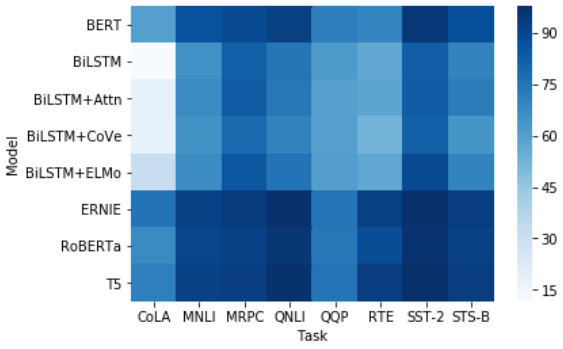
```
Out[12]: <matplotlib.axes._subplots.AxesSubplot at 0x2436a56db00>
```



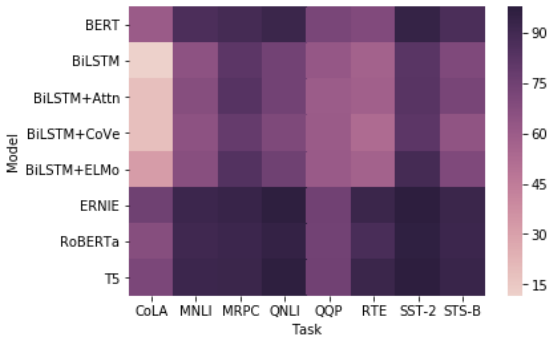
```
In [13]: sns.heatmap(glue, annot=True, linewidth=.5)
Out[13]: <matplotlib.axes._subplots.AxesSubplot at 0x2436a822e80>
```




```
In [15]: sns.heatmap(glue, cmap="Blues")
Out[15]: <matplotlib.axes._subplots.AxesSubplot at 0x2436a830fd0>
```

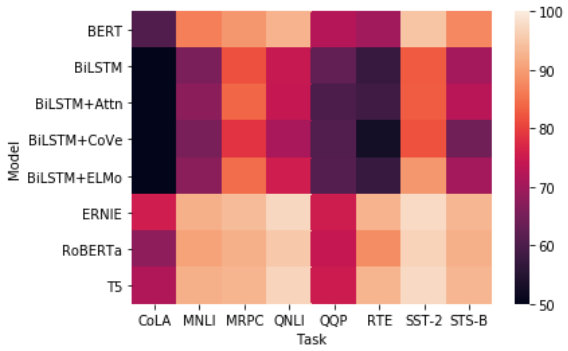



```
In [16]: sns.heatmap(glue, cmap=sns.cubehelix_palette(as_cmap=True))
Out[16]: <matplotlib.axes._subplots.AxesSubplot at 0x2436a99cf98>
```

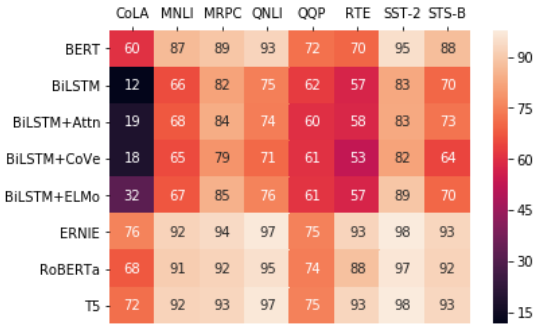


In [17]:
```
sns.heatmap(glue,vmin=50, vmax=100)
```

Out[17]: <matplotlib.axes._subplots.AxesSubplot at 0x2436a9fb0f0>

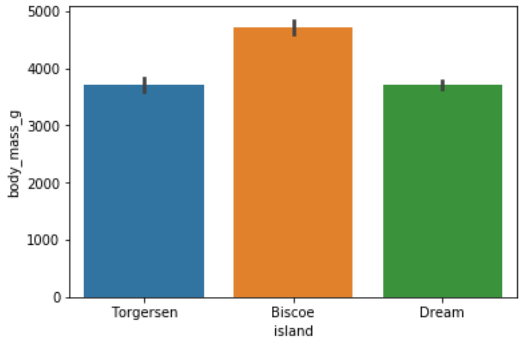


In [18]:
```
ax = sns.heatmap(glue, annot=True)
ax.set(xlabel="", ylabel="")
ax.xaxis.tick_top()
```

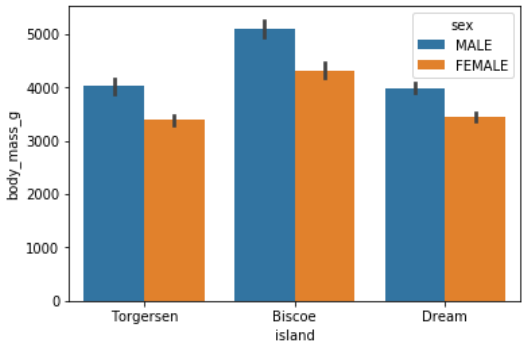


In [19]:
```
df = sns.load_dataset("penguins")
sns.barplot(data=df, x="island", y="body_mass_g")
```

Out[19]: <matplotlib.axes._subplots.AxesSubplot at 0x2436ab522b0>



```
In [21]: sns.barplot(data=df, x="island", y="body_mass_g", hue="sex")
Out[21]: <matplotlib.axes._subplots.AxesSubplot at 0x2436acaccf8>
```



Conclusion:

Successfully performed Data Visualization / Exploratory Data Analysis for the selected data set using Seaborn.