PRESIDENCY UNIVERSITY

GAIN MORE KNOWLEDGE
REACH GREATER HEIGHTS

Data Visualization Group Project

Project Report

**Based on Monthly Annual Mean Temperature
Dataset used: Countries Dataset**

Done by:
N.DHIVYAPRAKASH-20181CSE0453
NIKITHA AS-20181CSE0487
M.REVANTH NAGA VENKATA SAI-20181CSE0444
M DEVENDRA REDDY-20181CSE0451
MUSKAN-20181CSE0450

<u>Index:</u>

Dataset Explanation

Annual Monthly Mean Temperature: 1991-2020

The dataset used here is based on the "Annual Monthly Mean Temperature" from the year 1991-2020.Here in this dataset at least data of 20 countries are taken. The 20 countries are- India, Canada, Argentina , UAE, Maldives, Greenland, Denmark, US ,UK, Afghanistan,Fiji,etc .Here 20 datasets have been combined into 1 dataset with only the main readings from each country and made into a new file.

It compares the average temperature recorded monthly and even the annual temperature is collected which is the average of maximum and minimum temperature and Mean temperature is the average temperature of the month together.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
	Country	Annual	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
	Afghanistan	13.68	0.91	3.02	8.43	14.13	19.21	23.9	25.92	24.43	19.87	13.96	7.43	2.94
	Australia	22.14	28.43	27.61	25.79	22.51	18.34	15.37	14.74	16.4	19.97	23.31	25.79	27.39
	Albania	12.04	2.41	3.49	6.68	10.61	15.33	19.35	21.95	22.19	17.62	12.87	8.22	3.8
	Argentina	14.88	21.49	20.4	18.39	14.7	11.07	8.49	7.69	9.79	12.42	15.46	18.18	20.51
	Bangladesh	25.53	18.6	21.23	25.47	27.87	28.4	28.58	28.38	28.52	28.41	27.29	23.8	19.87
	Brazil	25.58	26.17	26.12	26.01	25.71	24.81	24.12	24.05	24.94	25.94	26.54	26.32	26.23
	Central African Republic	25.31	24.48	26.58	27.53	27.32	26.35	25.08	24.31	24.02	24.24	24.7	24.8	24.25
	Cambodia	27.51	25.86	27	28.39	29.36	28.92	28.25	27.67	27.61	27.35	27.05	26.82	25.77
	Canada	-4.23	-21.2	-20.04	-15	-6.42	2.26	9.38	13.02	11.6	6.03	-1.7	-10.97	-17.68
	Denmark	8.76	1.31	1.35	3.41	7.42	11.49	14.63	17.12	17.08	13.79	9.5	5.46	2.6
	Egypt	23.04	13.62	15.24	18.59	22.97	27.09	29.57	30.34	30.6	28.67	25.26	19.5	15.01
	Fiji	24.3	25.63	25.99	25.89	25.37	23.97	23.1	22.41	22.61	22.97	23.76	24.66	25.22
	Greenland	-15.84	-22.37	-24.23	-25.95	-19.9	-11.96	-5.46	-2.22	-4.7	-11.26	-17.24	-21.72	-23.05
	India	24.83	17.72	20.36	24.6	28.36	30.48	29.6	27.48	26.8	26.69	25.36	21.88	18.58
	Luxembourg	10.02	1.67	2.62	5.98	9.64	13.57	16.72	18.77	18.31	14.63	10.26	5.51	2.55
	Laos	23.76	19.54	21.15	23.48	25.63	26.24	26.27	25.83	25.58	25.28	24.1	22.17	19.87
	Maldives	27.8	27.57	28.26	28.21	28.49	28.17	27.96	27.55	27.57	27.51	27.36	27.57	27.42
	United Kingdom	9.13	3.9	4.04	5.65	7.8	10.7	13.43	15.32	15.19	12.94	9.78	6.51	4.27
	United Arab Emirates	27.63	18.95	20.17	23.04	27.48	31.72	33.54	34.68	34.4	32.62	29.26	24.66	21.02
	United States	9.47	-2.64	-0.71	3.29	8.62	14.34	19.15	21.82	20.89	16.82	10.15	3.22	-1.35

Python packages/libraries used

Pandas :-

Pandas is a Python package providing fast, flexible, and expressive data structures designed to make working with data both easy and intuitive. It aims to be the fundamental high-level building block for doing practical, real-world data analysis in Python.

Pyplot :-

   Pyplot is a plotting library used for 2-D graphics using python programming language. It can be used in python scripts,shell,web application servers and other GUI's toolkits.

Matplotlib:-

   Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, wxPython,etc.

Seaborn:-

   Seaborn is an open-source Python library built on top of matplotlib. It is used for data visualization and exploratory data analysis. Seaborn works easily with dataframes and the Pandas library. The graphs created can also be customized easily.

Numpy:-

NumPy is a Python library used for working with arrays.It also has functions for working in domain of linear algebra, fourier transform, and matrices.NumPy stands for Numerical Python.

Word Cloud:-

It is a data visualization technique used for representing text data in which the size of each word indicates its frequency or importance.Particular text datas can be highlighted using this word cloud.

Data Visualization- graph representation and interpretation

1. HISTOGRAM

A histogram is a chart that plots the distribution of a numeric variable's values as a series of bars.

```
climate=data['Annual']
plt.hist(climate,edgecolor="black",color="blue")
plt.xlabel("Average temp")
plt.ylabel("No of countries")
plt.title("Average temp")
```
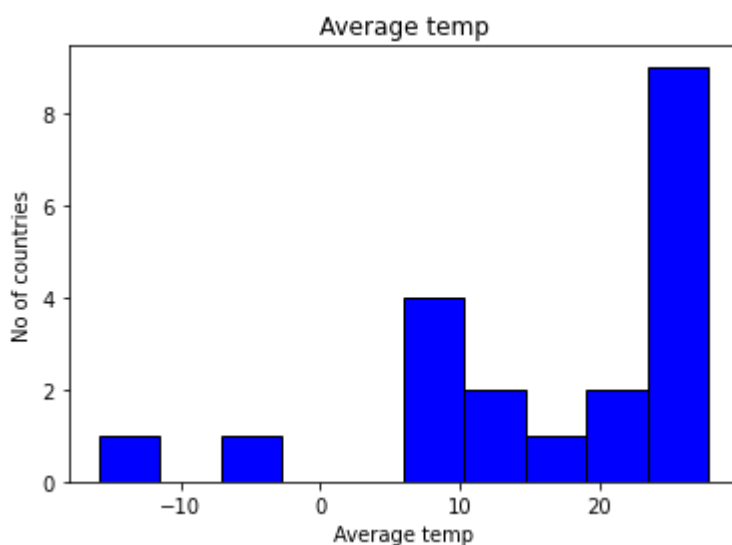

2. SCATTER PLOT

A scatter plot is a data visualization that displays the values of two different variables as points. The data for each point is represented by its horizontal (x) and vertical (y) position on the visualization.

```
rain=data['Apr']
country=data['Country']
plt.figure(figsize=(30,5))
plt.scatter(country,rain,cmap="Blues",marker="x",label="Avg")
plt.xlabel("Mean avg Temperature")
plt.ylabel("Countries")
plt.legend()
plt.show()
```
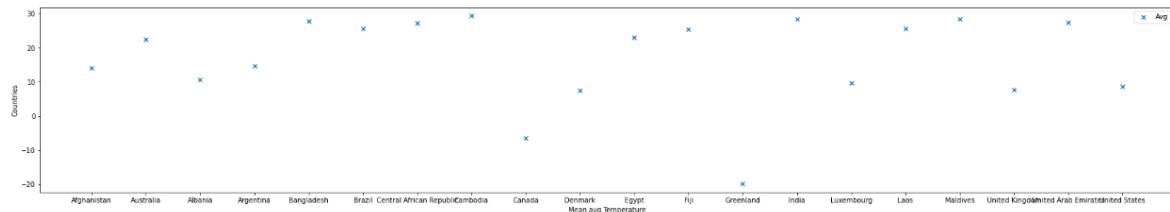
3. PIE CHART

Pie Charts help show proportions and percentages between categories, by dividing a circle into proportional segments. Each arc length represents a proportion of each category, while the full circle represents the total sum of all the data, equal to 100%.
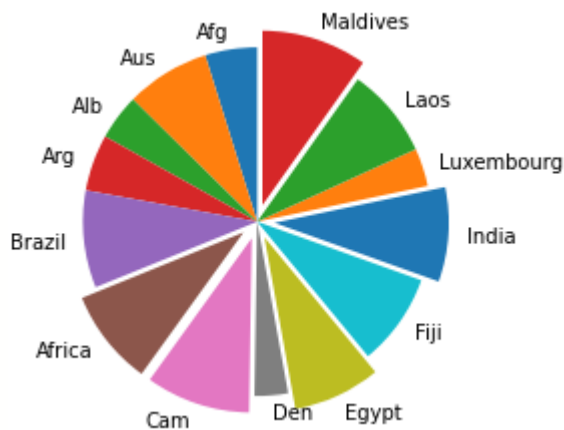
```
mean=[13.68,22.14,12.04,14.88,25.58,25.31,27.51,8.76,23.04,24.3,24.83,
10.02,23.76,27.8]
Designation=["Afg","Aus","Alb","Arg","Brazil","Africa","Cam","Den","Eg
ypt","Fiji","India","Luxembourg","Laos","Maldives"]
explodes = [0, 0, 0, 0,0, 0.1, 0.1, 0,0.1, 0, 0.1, 0,0, 0.1]
plt.pie(mean,labels=Designation,explode=explodes,startangle=90)
plt.show()
```

4. LINE GRAPH

A line chart or line graph is a type of chart which displays information as a series of data points called 'markers' connected by straight line segments

```
p=data['Country']
jan=data['Nov']
fig,ax=plt.subplots()
ax.plot(jan,p,linestyle="-.",label="Temp")
ax.legend()
ax.set_title("Annual-Nov month plotted graph")
ax.set_xlabel("Mean annual temp of Nov")
ax.set_ylabel("List of Countries")
plt.show()
```
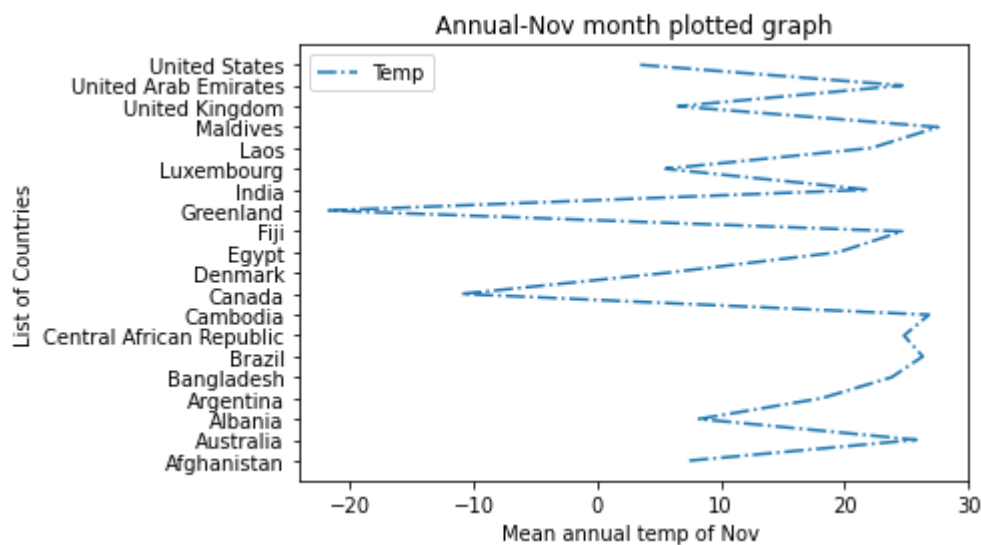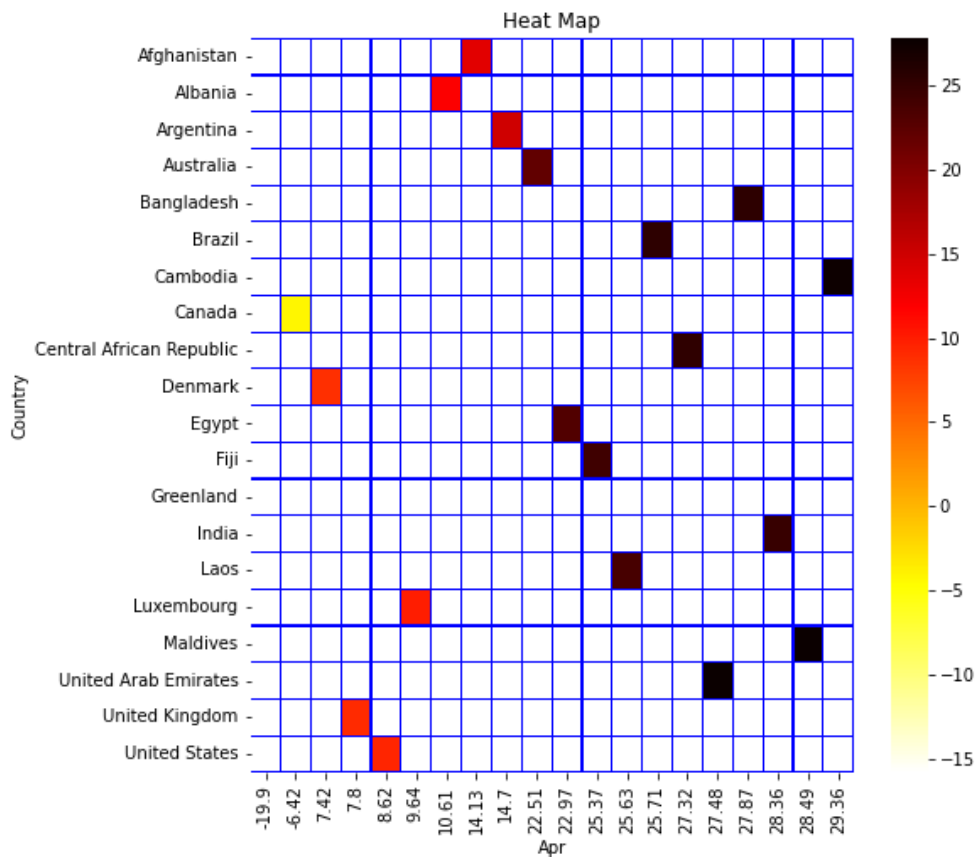

5. HEAT MAP

A heatmap is a graphical representation of data where the individual values contained in a matrix are represented as colors.

```
import seaborn as sns

fig, ax = plt.subplots(figsize=(8,8))
df2 = pd.pivot_table(values ='Annual', index ='Country', columns ='Apr', data = data)
sns.heatmap(df2,cmap='hot_r' , linecolor ='b', linewidths = 0.2,  a
x=ax).set_title('Heat Map')
```

```
Text(0.5, 1.0, 'Heat Map')
```
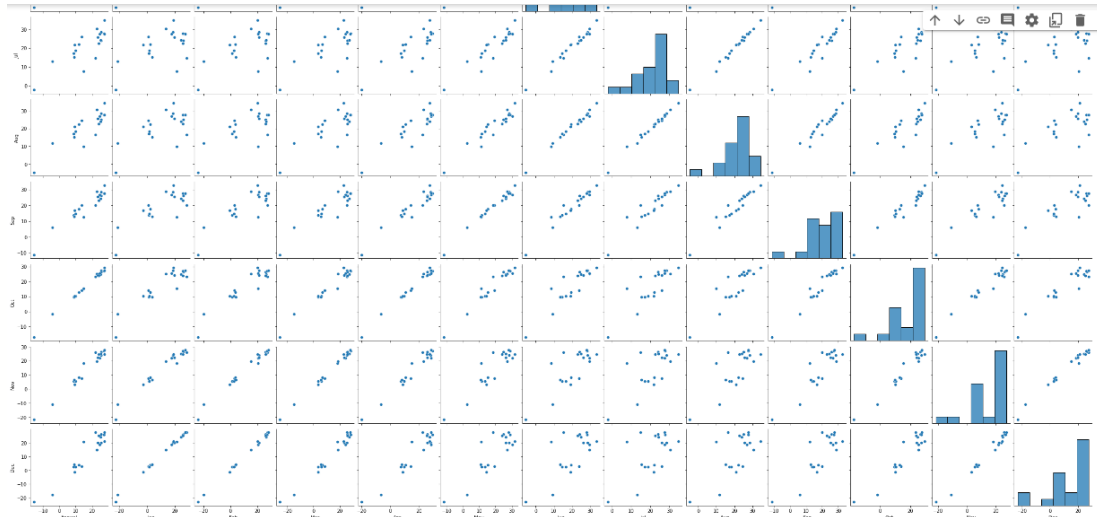

6. PAIR PLOT

A pairplot plot a pairwise relationships in a dataset. The pairplot function creates a grid of Axes such that each variable in data will by shared in the y-axis across a single row and in the x-axis across a single column. That creates plots as shown below.

```
g=sns.pairplot(data, palette ='hot_r')
g.fig.suptitle("Pair Plot")
g.add_legend()
```
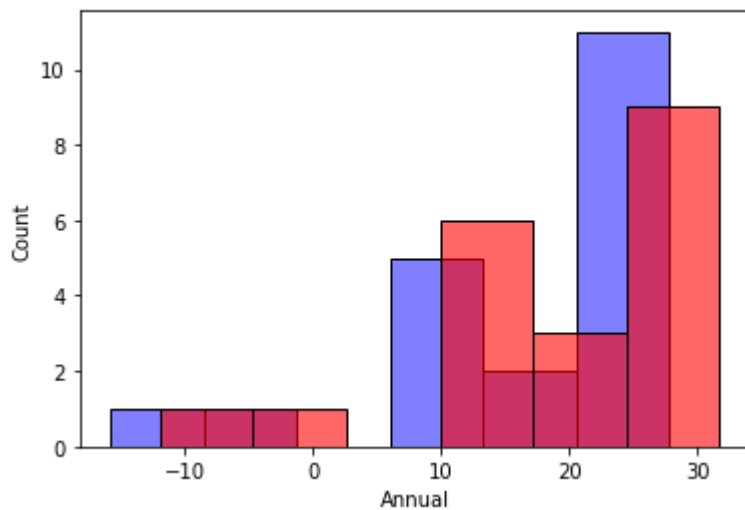
7. HIST PLOT

A hist plot is a chart that plots the distribution of a numeric variable's values as a series of bars.

```
annual=data['Annual']
monsoon=data['May']
sns.histplot(annual,color='blue',alpha=0.5)
sns.histplot(monsoon,color='red',alpha=0.6)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f2e27d50f50>
```

8. BOX PLOT

A Box Plot is also known as Whisker plot is created to display the summary of the set of data values having properties like minimum, first quartile, median, third quartile and maximum. In the box plot, a box is created from the first quartile to the third quartile, a vertical line is also there which goes through the box at the median.

```
numeric=data.select_dtypes (include="number")
num_cols=numeric.columns
plt.figure(figsize=(35,35))
for i in range(len(num_cols)):
        plt.subplot (7,7,i+1)
        sns.boxenplot (x=data[num_cols[i]])
        plt.title(num_cols[i])
plt.show()
```
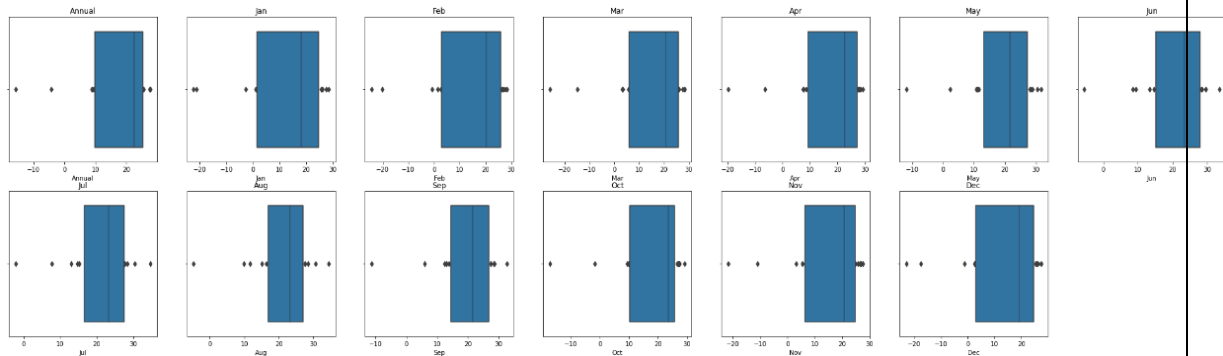

9. WORDCLOUD

Word Cloud is a data visualization technique used for representing text data in which the size of each word indicates its frequency or importance.

```
!pip install wordcloud

from PIL import Image
from wordcloud import WordCloud, STOPWORDS

#Stop Words
comment_words = ''
stopwords = set(STOPWORDS)

comment_words = data['Country'].str.lower()
comment_words = comment_words.to_string()

wc = WordCloud(width = 800, height = 800, background_color ='white', stop
words = stopwords, min_font_size = 11)
wc = wc.generate(comment_words)

plt.figure(facecolor = 'black')
```

```
1
plt.imshow(wc)
plt.axis("off")
plt.tight_layout(pad = -1)
plt.show()
```


Google Colab Project:

https://colab.research.google.com/drive/1IYtNqjgMbiDDsKXGPFomyB_wyfyrfaOr?usp=sharing

Github Repository Link:

https://github.com/Dhivyaprakash-hub/Climatic_Visualization

REFERENCE LINKS

- **https://climateknowledgeportal.worldbank.org/download-data**
- **https://www.section.io/engineering-education/seaborn-tutorial/**
- **https://public.wmo.int/en/our-mandate/climate/global-seasonal-climate-update**
- **https://www.tutorialspoint.com/python_data_science/python_heat_maps.htm**
- **https://seaborn.pydata.org/generated/seaborn.histplot.html**