EXPERIMENT NO:1(A)

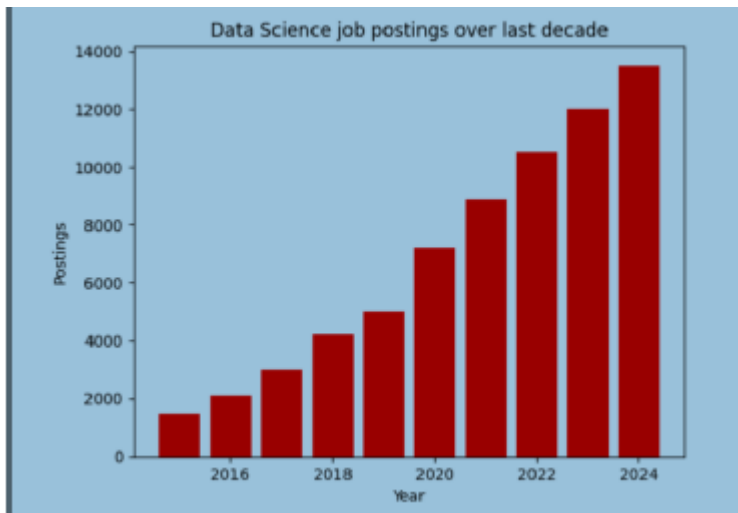## MATPLOTLIB LIBRARY – DATA VISUALIZATION

AIM :

To analyze the trends of data science job postings over the last decade

PROCEDURE:

•   Import Numpy and Matplotlib

•   Give some dummy data and make it as Data Frame

•   Use Bar plot for year over postings

•   Give attributes like xlabel,ylabel,title ,etc

•   Finally show the bar plot

•

PROGRAM :

```python
import matplotlib.pyplot as plt
import pandas as p
data={
    'Year':[2015,2016,2017,2018,2019,2020,2021,2022,2023,2024],
    'Postings':[1500,2100,3000,4200,5000,7200,8900,10500,12000,13500]
}
df=p.DataFrame(data)
yr=df['Year']
post=df['Postings']
plt.bar(yr,post,color='red')
plt.xlabel("Year")
plt.ylabel("Postings")
plt.title("Data Science job postings over last decade")
plt.show()
```

Data Science job postings over last decade

RESULT :

Thus the python program to visualize data using bar plot is executed and verified

EXPERIMENT-1(B)

MATPLOT LIBRARY – DATA VISUALIZATION

AIM :

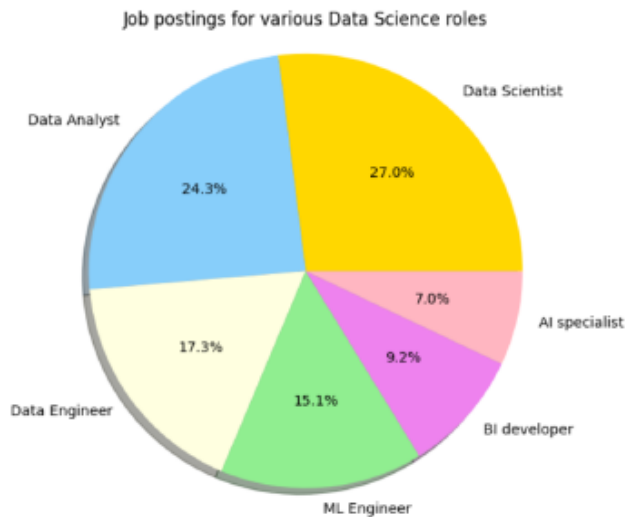To analyze and visualize distribution of various roles of Data Science

PROCEDURE :

• Import Pandas and Matplotlib

• Create a dataset

• Visualize through pie chart for job postings for various Data Science roles

• Use some attributes like colors ,title ,figure size etc

PROGRAM :

```
import matplotlib.pyplot as plt
import pandas as pd
data={
    'Roles':['Data Scientist','Data Analyst','Data Engineer','ML Engineer','BI developer','AI specialist'],
    'Post':[5000,4500,3200,2800,1700,1300]
}
df=pd.DataFrame(data)
role=df['Roles']
post=df['Post']
colors=['gold','lightskyblue','lightyellow','lightgreen','violet','lightpink']
plt.figure(figsize=(6,6))
plt.pie(post,labels=role,colors=colors,shadow=True,startangle=0,autopct='%1.1f%%')
plt.title('Job postings for various Data Science roles')
plt.axis('equal')
plt.show()
```

Job postings for various Data Science roles

RESULT :

Thus the python program to analyze and visualize postings of data science roles is executed and verified

EXPERIMENT-1(C)

DISPLAY THE STRUCTURED , UN STRUCTURED AND SEMI-STRUCTURED DATA

Aim :

　　　To differentiate structured , unstructured and semi structured data.

Procedure :

•　　　Import Matplotlib and Pandas

•　　　Create a DataFrame for structured data and print it

•　　　Likewise create Data Frame for unstructured and normalize it using json attribute and print it

•　　　Now create unstructured data and print it

Program:

```python
import matplotlib.pyplot as plt
import pandas as pd
structured={
    'Jerseyno':[45,7,18,33],
    'Name':['Rohit','MS','Virat','Hardik'],
    'Age':[38,44,36,31]
}
df=pd.DataFrame(structured)
print("Structured Data:\n")
print(df)
semistructured=[
    {'Name':'Rohit','Jerseyno':45,'Age':38},
    {'Name':'MS','Jerseyno':7,'Age':44},
    {'Name':'Virat','Jerseyno':18,'Age':36},
    {'Name':'Hardik','Jerseyno':33,'Age':31},
]
dff=pd.json_normalize(semistructured)
print("\nSemistructured Data:\n")
print(dff)
unstructured=['Rohit is a cricketer with jersey no 45.His age is 38',
              'MS is a cricketer with jersey no 7.  His age is 44',
              'Virat is a cricketer with jersey no 18.His age is 36',
              'Hardik is a cricketer with jersey no 33.His age is 31']
print("Unstructured Data:\n")
print(unstructured)
count=[len(structured['Age']),len(semistructured),4]
data=['Strcutured','Semistructured','Unstructured']
plt.bar(data,count,color=['gold','cornflowerblue','lightskyblue'])
plt.xlabel("Category of data")
plt.ylabel("No of data")
plt.title("No of data entry in each categories")
plt.show()
```
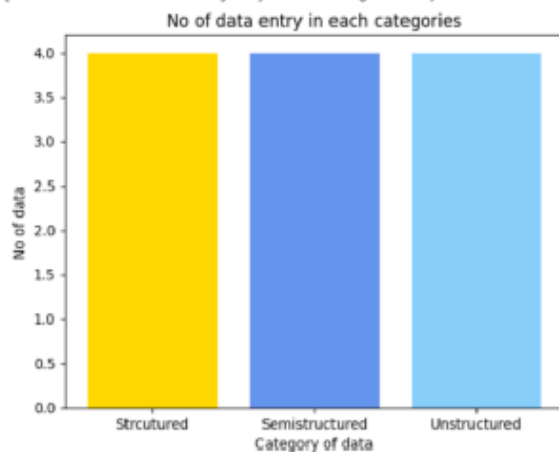
Structured Data:

```
   Jerseyno    Name  Age
0        45   Rohit   38
1         7      MS   44
2        18   Virat   36
3        33  Hardik   31
```

Semistructured Data:

```
     Name  Jerseyno  Age
0   Rohit        45   38
1      MS         7   44
2   Virat        18   36
3  Hardik        33   31
```
Unstructured Data:

['Rohit is a cricketer with jersey no 45.His age is 38', 'MS is a cricketer with jersey no 7. His age is 44', 'Virat is a cricketer with jersey no 18.H

Result:

Thus the python program to differentiate the characteristics of structured , unstructured and semi-structured data