

EXPERIMENT NO: 1(b)

Data visualization

Analyse and visualize the distribution of various Data Science roles

Aim:

To analyze and visualize the distribution of different Data Science job roles such as Data Analyst, Data Engineer, and Data Scientist using Python with Pandas and Matplotlib.

Algorithm:

1. Import required libraries (pandas, matplotlib).
2. Load or create a dataset containing job roles.
3. Group the data based on each role and count the occurrences.
4. Plot a pie chart or bar graph to show the role distribution.
5. Display and interpret the visualization.

Program:

```
[4]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
data = {
    'Job_Title': [
        'Data Analyst', 'Data Engineer', 'Data Scientist', 'ML Engineer',
        'Business Analyst', 'Data Analyst', 'Data Scientist', 'AI Engineer',
        'Data Engineer', 'Data Analyst', 'ML Engineer', 'Data Scientist',
        'Data Engineer', 'Business Analyst', 'AI Engineer'
    ]
}
df = pd.DataFrame(data)
print("Sample Dataset:")
display(df.head())
role_counts = df['Job_Title'].value_counts()
print("\nDistribution of Data Science Roles:")
display(role_counts)
plt.figure(figsize=(8, 5))
sns.barplot(x=role_counts.index, y=role_counts.values, palette="viridis")
plt.title("Distribution of Various Data Science Roles", fontsize=14, fontweight='bold')
plt.xlabel("Job Role", fontsize=12)
plt.ylabel("Number of Postings", fontsize=12)
plt.xticks(rotation=20)
plt.grid(True, axis='y', linestyle='--', alpha=0.7)
plt.show()
plt.figure(figsize=(6, 6))
plt.pie(role_counts.values, labels=role_counts.index, autopct='%1.1f%%', startangle=90, colors=sns.color_palette("viridis"))
plt.title("Percentage Distribution of Data Science Roles", fontsize=14, fontweight='bold')
plt.show()
```

```
plt.ylabel("Number of Postings", fontsize=12)
plt.xticks(rotation=20)
plt.grid(True, axis='y', linestyle='--', alpha=0.7)
plt.show()
plt.figure(figsize=(6, 6))
plt.pie(role_counts.values, labels=role_counts.index, autopct='%1.1f%%', startangle=90, colors=sns.color_palette("viridis"))
plt.title("Percentage Distribution of Data Science Roles", fontsize=14, fontweight='bold')
plt.show()
```

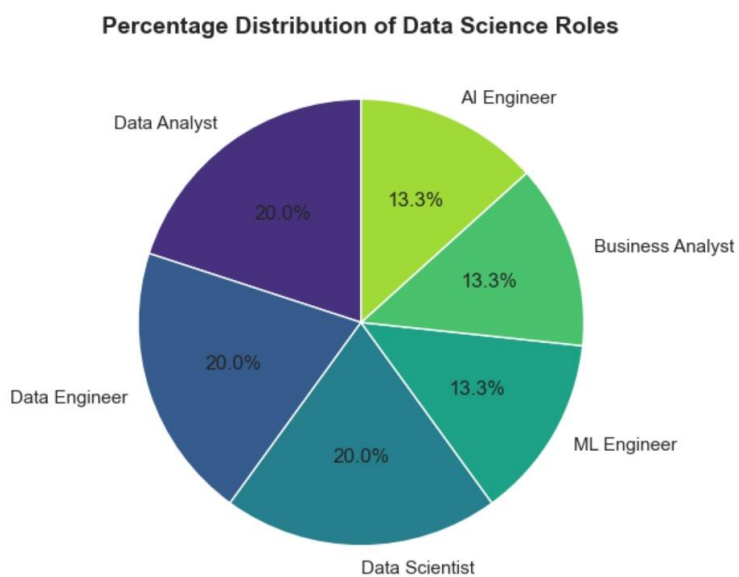
Sample Dataset:

	Job_Title
0	Data Analyst
1	Data Engineer
2	Data Scientist
3	ML Engineer
4	Business Analyst

Distribution of Data Science Roles:

Job_Title	
Data Analyst	3
Data Engineer	3
Data Scientist	3
ML Engineer	2
Business Analyst	2
AI Engineer	2

Name: count. dtype: int64



Result:

Thus, the Python code to analyze and visualize the distribution of various Data Science roles using pie chart is successfully executed.