EX.NO: 1.b Analyze and visualize the distribution of various data science roles

AIM:

To analyze and visualize the distribution of various Data Science roles such as Data Analyst, Data Engineer, and Data Scientist using a dataset to understand their relative demand.

ALGORITHM:

- *Import required libraries – pandas, matplotlib, seaborn.*
- *Load the dataset containing job postings or role information.*
- *Filter and clean the data for relevant Data Science job roles.*
- *Count the frequency of each unique job role (e.g., Data Analyst, Data Engineer, Data Scientist).*
- *Visualize the distribution using bar and pie charts.*
- *Interpret the results to identify which roles are most in demand.*

PROGRAM:

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
data = {
   "Role": ["Data Analyst", "Data Scientist", "Data Engineer", "ML Engineer", "Business Analyst",
         "Data Scientist", "Data Engineer", "Data Analyst", "ML Engineer", "Data Scientist"]
}

df = pd.DataFrame(data)

role_counts = df['Role'].value_counts()
print(role_counts)

plt.figure(figsize=(7,5))
sns.barplot(x=role_counts.index, y=role_counts.values, palette='viridis')
plt.title("Distribution of Data Science Roles")
plt.xlabel("Job Role")
plt.ylabel("Count")
plt.xticks(rotation=30)
```

```
plt.grid(True)
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("Proportion of Data Science Roles")
plt.show()
```

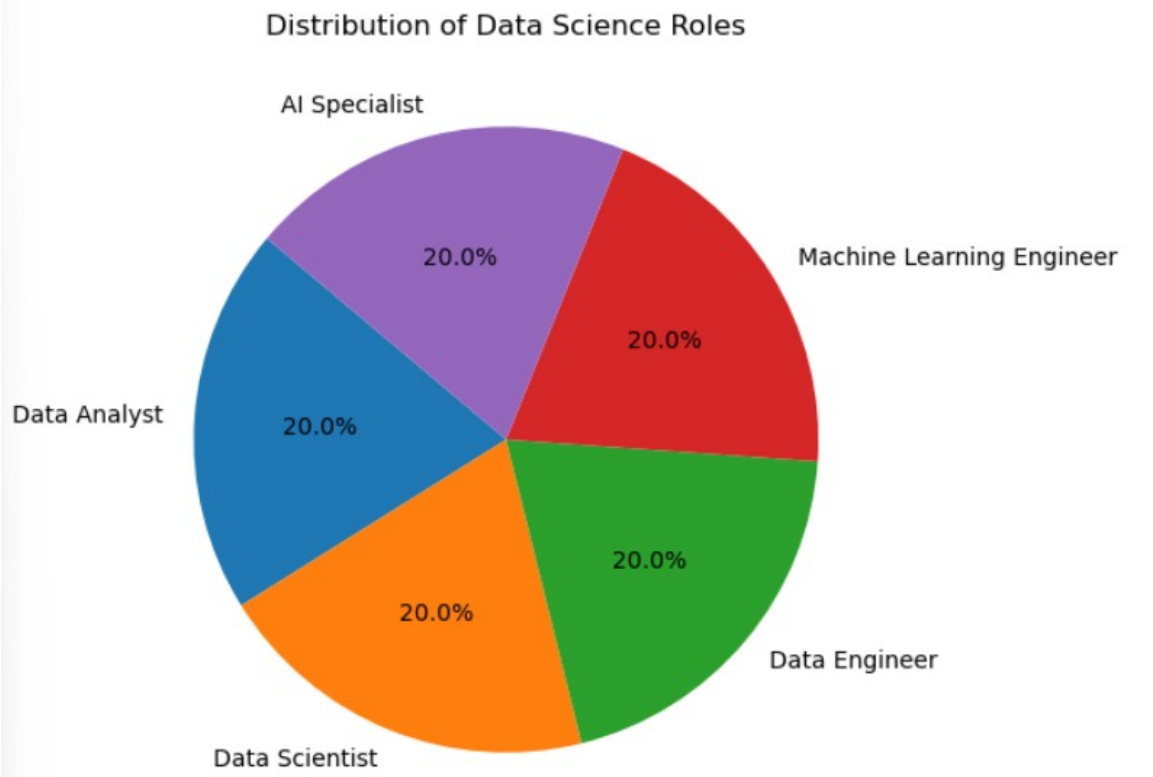
OUTPUT:

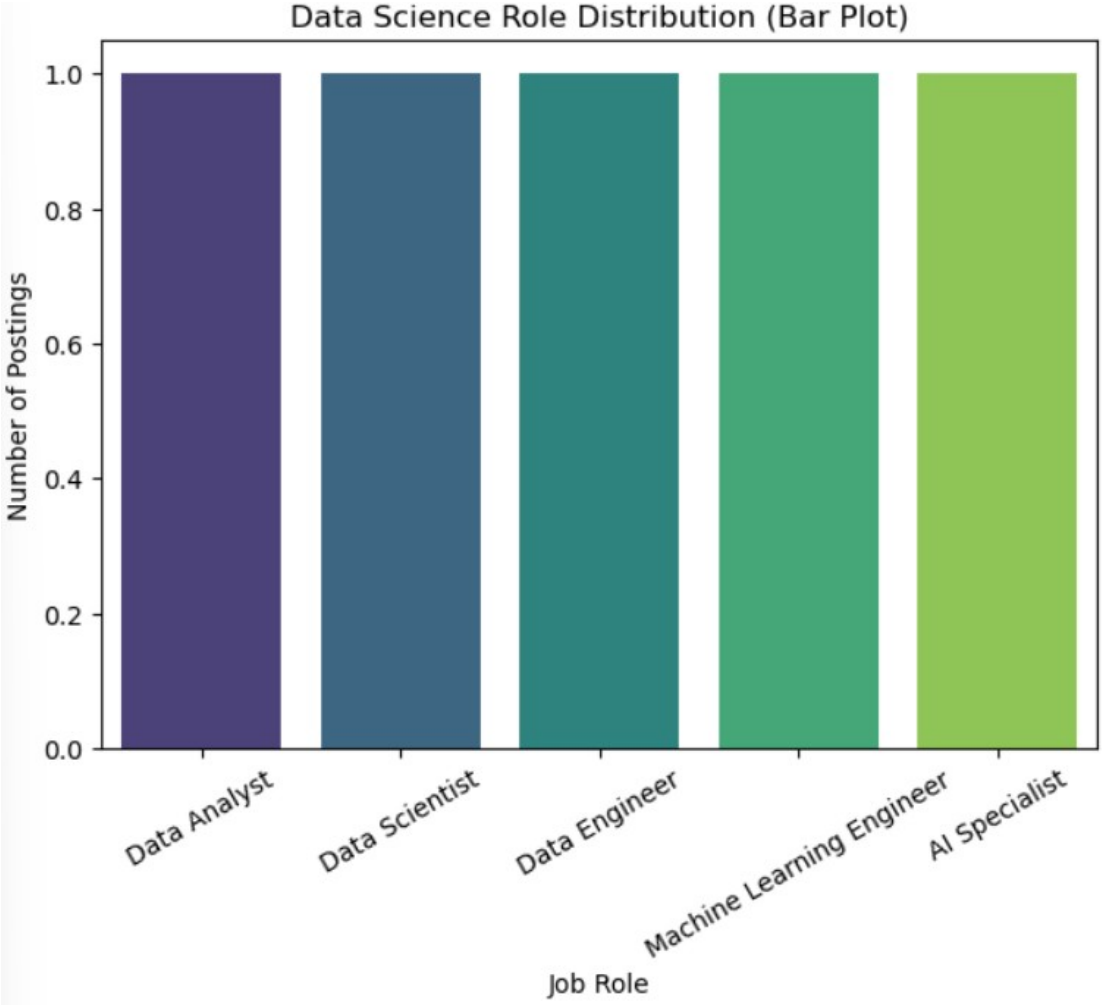
[12]:

	Job Title
0	Data Analyst
1	Data Scientist
2	Data Engineer
3	Machine Learning Engineer
4	AI Specialist

[13]:

```
Job Title
Data Analyst                1
Data Scientist              1
Data Engineer               1
Machine Learning Engineer   1
AI Specialist               1
Name: count, dtype: int64
```


Distribution of Data Science Roles
AI Specialist
Machine Learning Engineer
Data Analyst
Data Engineer
Data Scientist
20.0%
20.0%
20.0%
20.0%
20.0%



RESULT:

The analysis shows that Data Scientist and Data Engineer roles are the most common, followed by Data Analyst and Machine Learning Engineer positions. This indicates a balanced but high demand for core Data Science roles across the industry.