CS23334 -Fundamental Of Data Science

240701463

Analysis and Visualization of Data Science Role Distribution

Exp :1B

Date: 29-07-2025

Aim:

To **analyze and visualize the distribution of various Data Science roles** (Data Analyst, Data Engineer, Data Scientist, etc.) from a dataset, using **bar plots and pie charts** for visualization.

Algorithm:

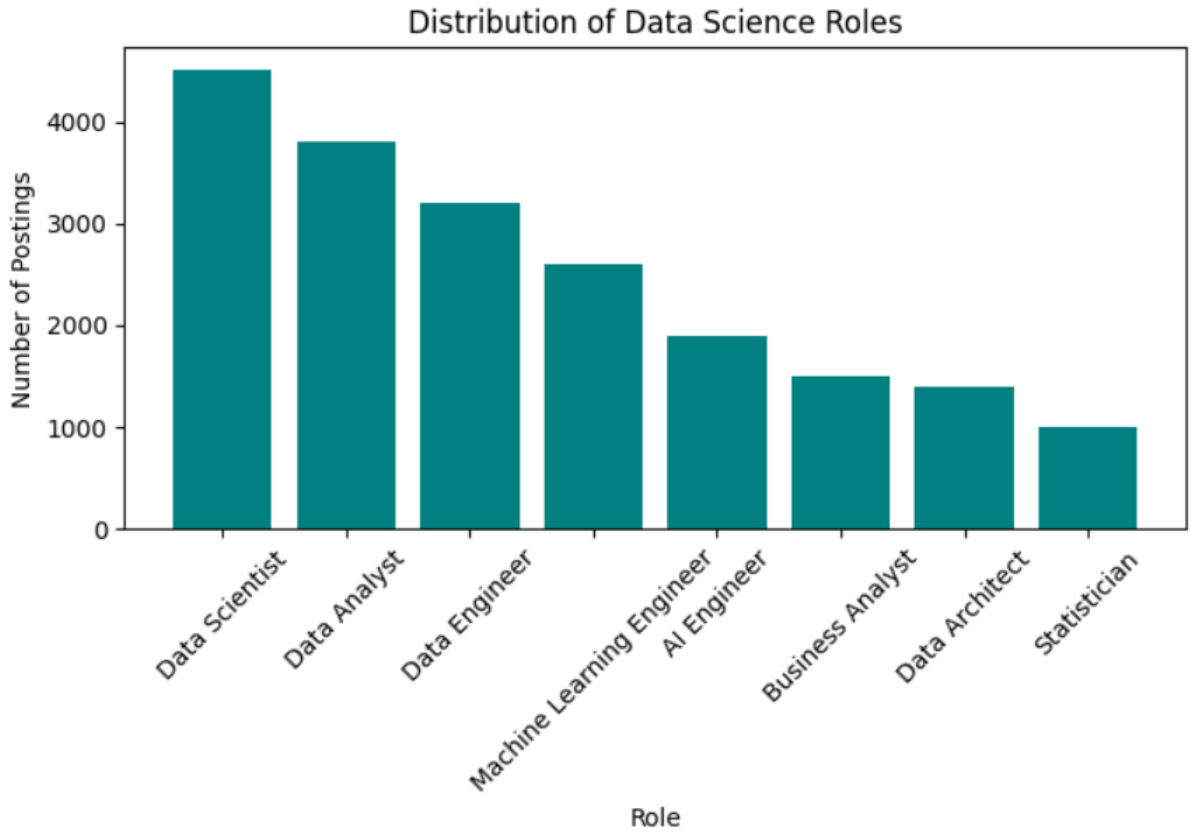
1. Load the data_science_roles.csv dataset into a pandas DataFrame.
2. Generate a **bar plot** to visualize the raw count of postings per role.
3. Label the bar plot with a title, axis labels, and rotate x-ticks for readability.
4. Generate a **pie chart** to visualize the percentage distribution of roles.
5. Label the pie chart with percentages (autopct) and a title.
6. Display both visualizations and confirm the output.

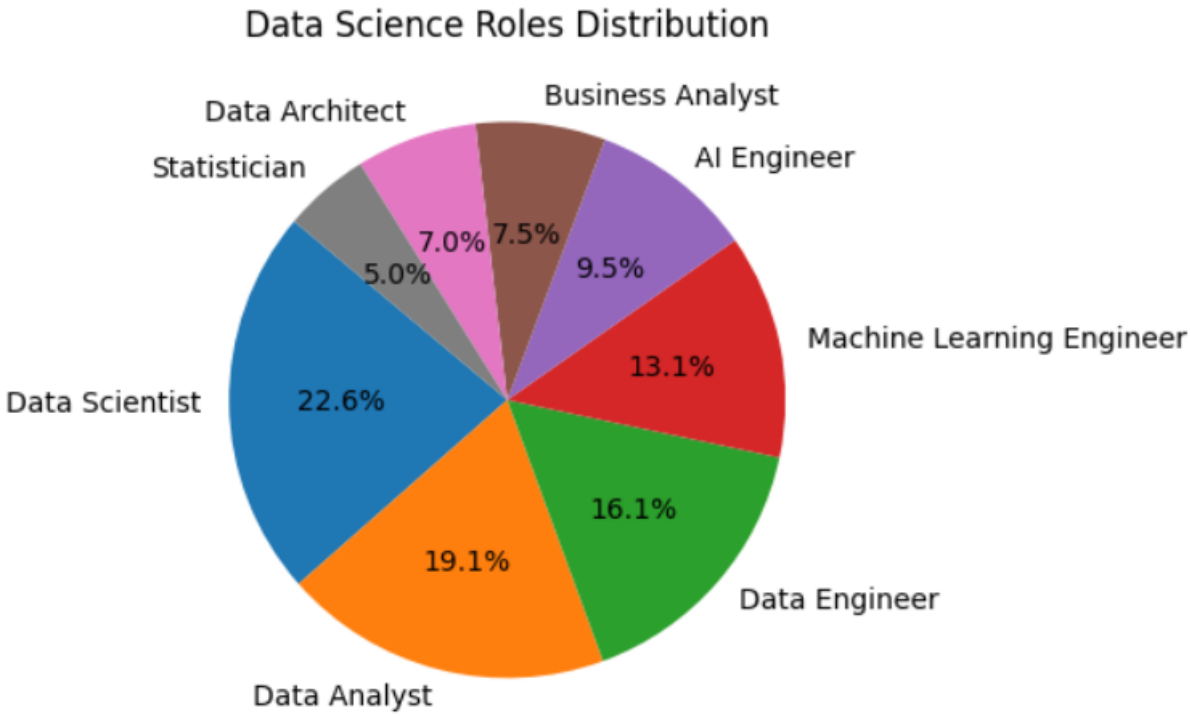
Code:

```
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv("data_science_roles.csv")
plt.figure(figsize=(7,5))
plt.bar(df['Role'], df['Count'], color='teal')
plt.title('Distribution of Data Science Roles')
plt.xlabel('Role')
plt.ylabel('Number of Postings')
plt.xticks(rotation=45)
```

```
plt.tight_layout()
plt.show()
plt.figure(figsize=(6,6))
plt.pie(df['Count'], labels=df['Role'], autopct='%1.1f%%', startangle=140)
plt.title('Data Science Roles Distribution')
plt.tight_layout()
plt.show()
```

Output:





Result:

The analysis shows that **"Data Scientist" is the most frequently posted role**, accounting for the largest percentage of job postings in the dataset. Thus the python program was executed successfully, and the output is verified.