CS-23334 FUNDAMENTALS OF DATA SCIENCE

Abenanthan P 240701005

EXPERIMENT 1

1.A Analyze the trend of data science job postings over the last decade

AIM:

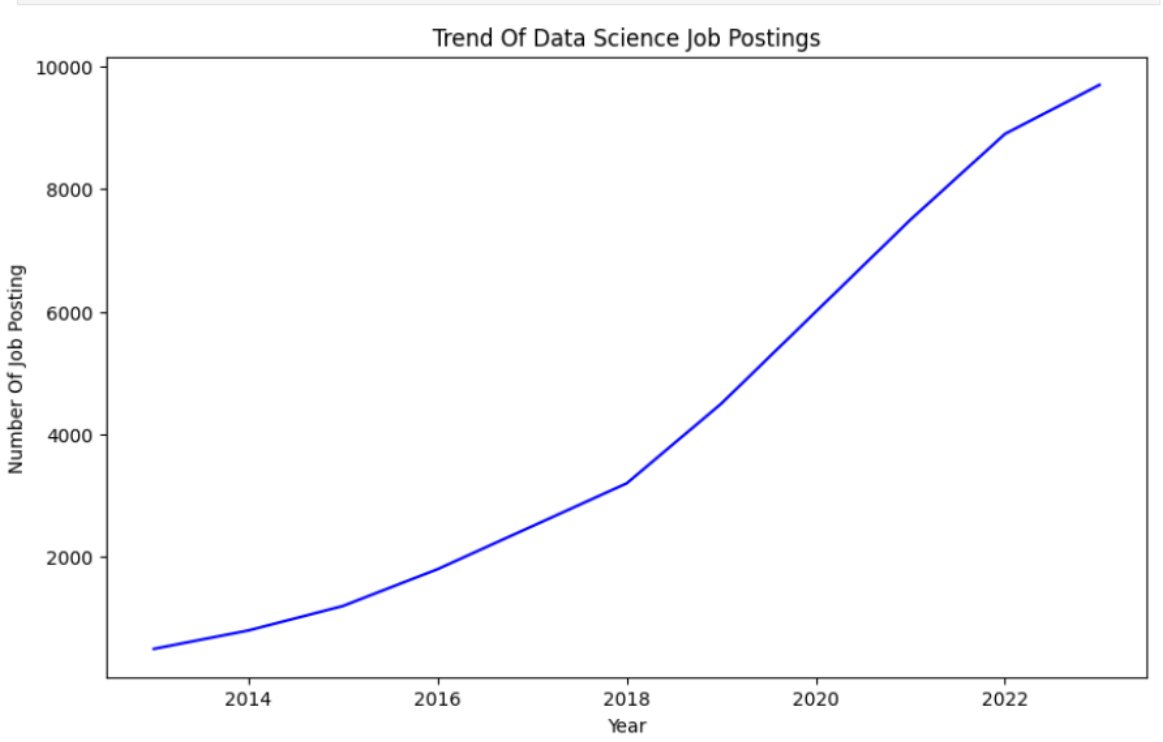
To analyze and visualize the distribution of various data science roles using a bar chart

```
In [1]: import pandas as pd
        import matplotlib.pyplot as plt

        data={'Year':list(range(2013,2024)),
              'Job_Postings':[500,800,1200,1800,2500,3200,4500,6000,7500,8900,9700]
             }

        df=pd.DataFrame(data)
        plt.figure(figsize=(10,6))
        plt.plot(df['Year'],df['Job_Postings'],color='blue')
        plt.title('Trend Of Data Science Job Postings')
        plt.xlabel('Year')
        plt.ylabel('Number Of Job Posting')
        plt.show()
```

Output:



RESULT:

The line graph shows a consistent and significant increase in data science job postings from 2013 to 2023, indicating growing demand in the field

1.B. Analyze and visualize the distribution of various data science roles (Data Analyst, Data Engineer, Data Scientist, etc.) from a dataset

AIM:

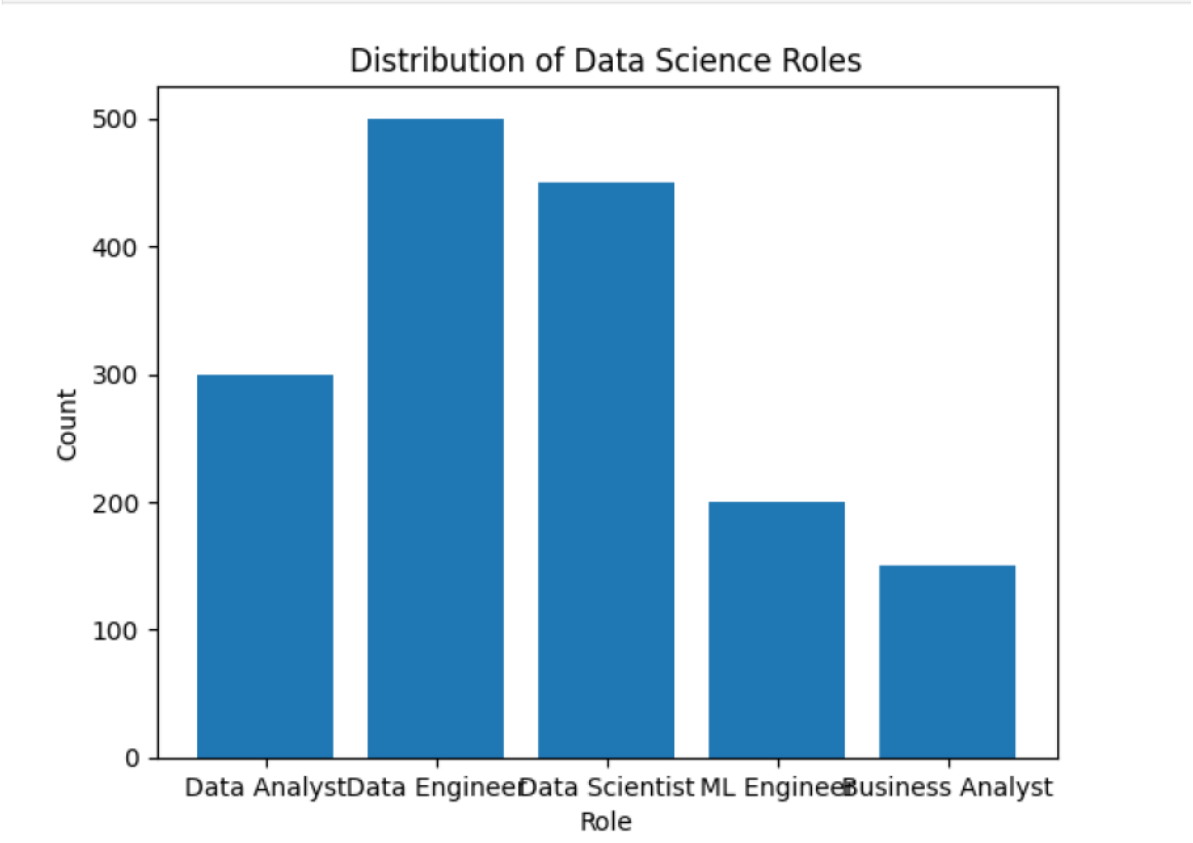
To analyze and visualize the distribution of various data science roles using a bar chart.

```
import pandas as pd
import matplotlib.pyplot as plt

roles = ['Data Analyst', 'Data Engineer', 'Data Scientist', 'ML
Engineer', 'Business Analyst']
counts = [300, 500, 450, 200, 150]

plt.bar(roles, counts)
plt.title('Distribution of Data Science Roles')
plt.xlabel('Role')
plt.ylabel('Count')
plt.show()
```

Output:



RESULT:

The bar chart reveals that Data Engineer and Data Scientist roles are the most prevalent, followed by Data Analyst, ML Engineer, and Business Analyst.

1.C. Conduct an experiment to differentiate Structured , Un-structured and Semi structured data

```
# Structured data example

structured_data = pd.DataFrame({
'ID': [1, 2, 3],
'Name': ['Alice', 'Bob', 'Charlie'], 'Age': [25, 30, 35]
})
print("Structured Data:\n", structured_data)
```

```
# Unstructured data example

unstructured_data = "This is an example of unstructured data. It can
be a piece of text, an image, or a video file."
print("\nUnstructured Data:\n", unstructured_data)
```

```
# Semi-structured data example (JSON)
semi_structured_data = {'ID': 1, 'Name': 'Alice', 'Attributes':
{'Height': 165, 'Weight': 68}}
print("\nSemi-structured Data:\n", semi_structured_data)
```

Output:

```
Structured Data:
    ID     Name  Age
0   1    Alice   25
1   2      Bob   30
2   3  Charlie   35

Unstructured Data:
 This is an example of unstructured data. It can be a piece of text,
an image, or a video file.

Semi-structured Data:
 {'ID': 1, 'Name': 'Alice', 'Attributes': {'Height': 165, 'Weight':
68}}
```

RESULT:

- Structured Data: Tabular format with defined schema (e.g., DataFrame with ID, Name, Age).
- Unstructured Data: Free-form text without predefined structure.
- Semi-structured Data: JSON-like format with nested attributes.

1.D. Conduct an experiment to encrypt and decrypt given sensitive data.

AIM:

To encrypt and decrypt sensitive data using the Fernet symmetric encryption method from the cryptography library.

```
from cryptography.fernet import Fernet
key = Fernet.generate_key()
f = Fernet(key)
token = f.encrypt(b"Abenanthan 240701005")
token
b'...'
f.decrypt(token)
b'Abenanthan 240701005'
key = Fernet.generate_key()
cipher_suite = Fernet(key)
plain_text = b"Abenanthan 240701005"
cipher_text = cipher_suite.encrypt(plain_text)
decrypted_text = cipher_suite.decrypt(cipher_text)
print("Original Data:", plain_text)
print("Encrypted Data:", cipher_text)
print("Decrypted Data:", decrypted_text)
```

Output:

```
Original Data: b'Abenanthan 240701005'
Encrypted Data:
b'gAAAAABo63XjX4by2WLWfqIDOt_JABlo6QlRY7UFPlF7imBNNTjF6vJNQhST0w0hzNjW
4_dSL-BvwiD6Jipje3GY8Ni3gpgwDn1xyqusL1Jb4YXVEN-Nao4='
Decrypted Data: b'Abenanthan 240701005'
```

RESULT:

The original data ("Abenanthan 240701005") was successfully encrypted into a secure token and decrypted back to its original form, demonstrating effective data protection.