Ex 2 EDA-Data Import and Export

AIM:

To perform exploratory data analysis by importing data from various sources such as CSV, Excel, SQL, and web scraping, and export DataFrames into Excel and CSV formats using Python.

1. Importing data from CSV, Excel, SQL databases, and web scraping

- **CSV**

```
import pandas as pd

df = pd.read_csv('/content/suv_data.csv')

df.head()
```

OUTPUT:

	User ID	Gender	Age	EstimatedSalary	Purchased
0	15624510	Male	19	19000	0
1	15810944	Male	35	20000	0
2	15668575	Female	26	43000	0
3	15603246	Female	27	57000	0
4	15804002	Male	19	76000	0

- **EXCEL**

```
!pip install openpyxl

df.to_excel('suv_data.xlsx', index=False)

df2 = pd.read_excel('suv_data.xlsx')

df2.head()
```

OUTPUT:

	User ID	Gender	Age	EstimatedSalary	Purchased
0	15624510	Male	19	19000	0
1	15810944	Male	35	20000	0
2	15668575	Female	26	43000	0
3	15603246	Female	27	57000	0
4	15804002	Male	19	76000	0

- **SQL DB**

```
import sqlite3
import pandas as pd
conn = sqlite3.connect('mydata.db')
df = pd.read_csv('suv_data.csv')  # Your existing data
df.to_sql('suv_table', conn, if_exists='replace', index=False)  # Store to SQL
df_sql = pd.read_sql_query("SELECT * FROM suv_table", conn)
df_sql.head()
conn.close()
```

OUTPUT:

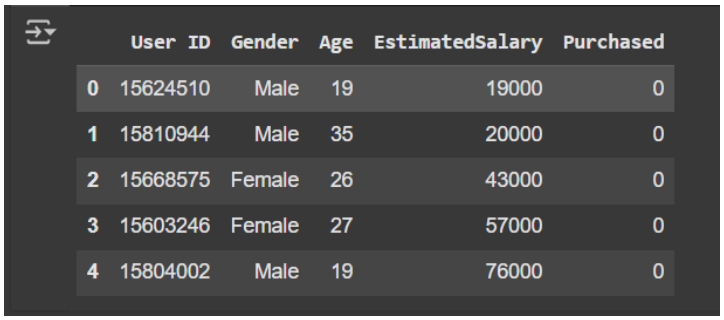
	User ID	Gender	Age	EstimatedSalary	Purchased
0	15624510	Male	19	19000	0
1	15810944	Male	35	20000	0
2	15668575	Female	26	43000	0
3	15603246	Female	27	57000	0
4	15804002	Male	19	76000	0

- **WEB SCRAPING**

```
import pandas as pd
```

```
url = 'https://en.wikipedia.org/wiki/List_of_countries_by_GDP_(nominal)'
tables = pd.read_html(url)  # This will return a list of tables
print(len(tables))  # See how many tables were found
df_web = tables[1]  # You can try 0, 1, 2, etc.
df_web.head()
```

OUTPUT:



	User ID	Gender	Age	EstimatedSalary	Purchased
0	15624510	Male	19	19000	0
1	15810944	Male	35	20000	0
2	15668575	Female	26	43000	0
3	15603246	Female	27	57000	0
4	15804002	Male	19	76000	0

2. Handling different data formats

- **JSON**

```python
import json

data = {
    "name": ["Meenakshi", "Rahul", "Aisha"],
    "age": [20, 21, 22],
    "department": ["AIML", "CSE", "ECE"]
}

# Convert to DataFrame
df = pd.DataFrame(data)

# Save to JSON
df.to_json('students.json', orient='records', lines=True)

# Read it back
df_back = pd.read_json('students.json', lines=True)
df_back
```

	name	age	department
0	Meenakshi	20	AIML
1	Rahul	21	CSE
2	Aisha	22	ECE

- **XML**

```python
import pandas as pd

df = pd.read_csv('suv_data.csv')
```

```python
import pandas as pd

# Step 1: Load the CSV
df = pd.read_csv('suv_data.csv')

# Step 2: Rename columns to remove spaces (XML tags can't have spaces)
df.columns = [col.replace(" ", "_") for col in df.columns]

# Step 3: Save to XML
df.to_xml('suv_data.xml', index=False)
```

```python
df_xml = pd.read_xml('suv_data.xml')
df_xml.head()
```

	User_ID	Gender	Age	EstimatedSalary	Purchased
0	15624510	Male	19	19000	0
1	15810944	Male	35	20000	0
2	15668575	Female	26	43000	0
3	15603246	Female	27	57000	0
4	15804002	Male	19	76000	0

- **PYTHON DICTIONARY**

```python
data = {
    'Name': ['Alice', 'Bob'],
    'Age': [25, 30],
    'City': ['Delhi', 'Chennai']
}

df_dict = pd.DataFrame(data)
df_dict
```

	Name	Age	City
0	Alice	25	Delhi
1	Bob	30	Chennai

3. Export a DataFrame to an Excel file.

```python
import pandas as pd

df = pd.read_csv('suv_data.csv')  # Or use any DataFrame you've created
```

```python
df.to_excel('suv_data_exported.xlsx', index=False)
```

```python
df_excel = pd.read_excel('suv_data_exported.xlsx')
df_excel.head()
```

	User ID	Gender	Age	EstimatedSalary	Purchased
0	15624510	Male	19	19000	0
1	15810944	Male	35	20000	0
2	15668575	Female	26	43000	0
3	15603246	Female	27	57000	0
4	15804002	Male	19	76000	0