Practical 6

Querying Big Data Using SQL

Aim: To perform SQL queries on a dataset stored in a database or CSV file.

Steps:

1. **Set up a SQLite database with a library like sqlite3. Command (pip install sqlite3)**

2. **Load the dataset into a database table.**

3. **Write SQL queries to extract specific insights (e.g., top 5 records, group by, etc.).**

4. **Display the query results.**

Code:

```
import sqlite3

import pandas as pd

# Step 1: Create SQLite connection and load data from CSV

conn = sqlite3.connect("big_data.db")

# Sample data (this is the same as the contents of sample_dataset.csv)

data = {

    "ID": [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15],

    "Name": [
```

```python
        "John Doe", "Jane Smith", "Emily Johnson",
"Michael Brown", "Emma Davis",
        "David Miller", "Liam Wilson", "Olivia Moore",
"Lucas Taylor", "Sophia Anderson",
        "Charlotte Thomas", "Amelia Jackson", "James
White", "Benjamin Harris", "Isabella Martin"
    ],
    "Category": [
        "Electronics", "Clothing", "Electronics",
"Furniture", "Electronics",
        "Clothing", "Clothing", "Furniture", "Clothing",
"Furniture",
        "Electronics", "Furniture", "Clothing",
"Electronics", "Clothing"
    ],
    "Value": [100, 50, 150, 200, 120, 80, 60, 220, 40, 210, 90,
240, 55, 130, 95]
}
# Step 2: Create DataFrame from the sample data
df = pd.DataFrame(data)
# Step 3: Insert data into the SQLite database (creating
the table)
df.to_sql("data_table", conn, if_exists="replace",
index=False)
```

 # Step 4: Write SQL query to count records by category

query = "SELECT Category, COUNT(*) as Count FROM data_table GROUP BY Category ORDER BY Count DESC LIMIT 5"

 # Step 5: Execute SQL query and fetch results

result = pd.read_sql(query, conn)

 # Step 6: Print the result of the query

print(result)

 # Step 7: Close the connection to the database

conn.close()

Output:

```
        Category   Count
0       Clothing      6
1    Electronics      5
2      Furniture      4
```