Data Analyst Internship Task 7 Report

Soumen das

November 24, 2025

Objective

The objective of Task 7 was to connect to a SQLite database using Python, perform a basic sales summary analysis using SQL queries, and visualize the results using a bar chart. This task demonstrates the integration of database management (SQLite) with data analysis tools (Pandas) and visualization libraries (Matplotlib).

1 Tools and Technologies

The following tools were utilized to complete the task:

- **Python**: The primary programming language used for scripting.
- **SQLite3**: A built-in Python library used to interact with the `sales_data.db` database.
- **Pandas**: Used to store the results of the SQL query into a DataFrame for easy manipulation and tabular display.
- **Matplotlib**: Used to generate a bar chart representing the total revenue per product.

2 Methodology

2.1 1. Database Connection and Creation

A connection was established to a database file named `sales_data.db`. A table named `sales` was created with columns for `product`, `quantity`, and `price`. Sample data was inserted to simulate a real-world scenario.

2.2 2. SQL Query Execution

An SQL query was executed to aggregate the data. The query grouped sales by product and calculated:

1. Total Quantity Sold (`SUM(quantity)`)
2. Total Revenue (`SUM(quantity * price)`)

```sql
SELECT
    product,
    SUM(quantity) AS total_qty,
    SUM(quantity * price) AS revenue
FROM sales
GROUP BY product
```

Listing 1: SQL Query Used

3 Code Implementation

The Python script uses the `sqlite3` library to execute the query and `pandas` to read the output directly into a DataFrame.

```
# Load into Pandas DataFrame
df = pd.read_sql_query(query, conn)

# Print the results
print(df)

# Plot simple bar chart
df.plot(kind='bar', x='product', y='revenue', legend=False)
plt.title('Total Revenue by Product')
plt.ylabel('Revenue')
plt.show()
```

Listing 2: Python Script Snippet

4 Conclusion

The task was successfully completed. The Python script connected to the database, retrieved the aggregated sales data, and displayed both a tabular summary and a visual bar chart. This workflow is essential for basic automated reporting and data analysis pipelines.