TASK 7: Get Basic Sales Summary from a Tiny SQLite Database using Python

Objective

To generate and analyze a larger dataset of sales information using SQL within Python, and visualize it through various chart types including bar, pie, and line charts to gain insights into quantity sold, revenue by product, region-wise performance, and monthly trends.

Tools Used

- Python 3
- SQLite3
- Pandas
- Matplotlib
- Jupyter Notebook or .py file

Step-by-Step Execution:

Step 1: Create and Populate SQLite Database

Connected to a database using sqlite3 and created a new table named "sales". Generated over 150 random sales records with fields: product, category, quantity, price, region, and sale_date. Used random selection from predefined lists and inserted the data using cursor.execute in a loop.

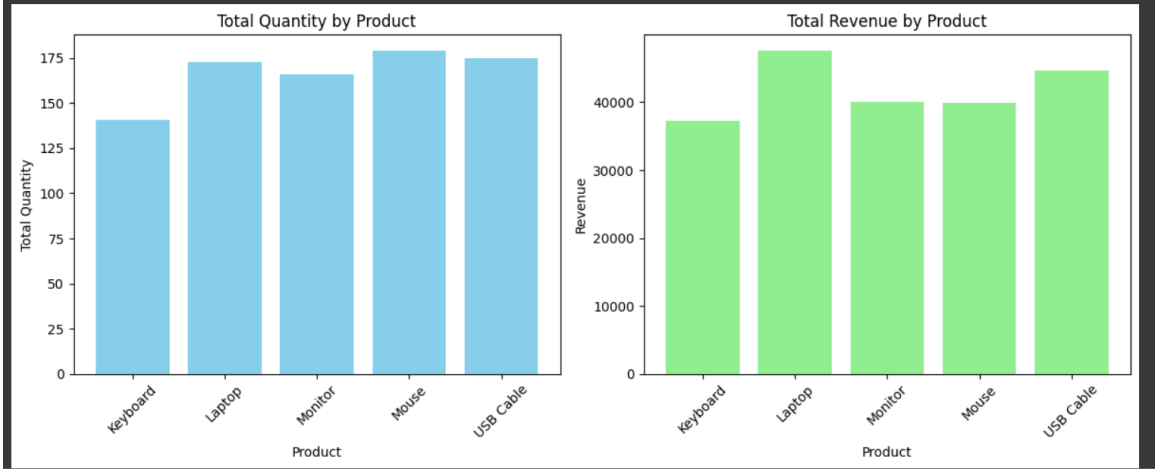
Step 2: View Full Sales Data

Fetched the complete dataset using:
SELECT * FROM sales
and loaded it into a pandas DataFrame.

Step 3: Product-wise Summary (Quantity and Revenue)

Executed SQL query to calculate total quantity sold and revenue per product using:
GROUP BY product
Visualized the result with two bar charts placed side-by-side, showing quantity and revenue.

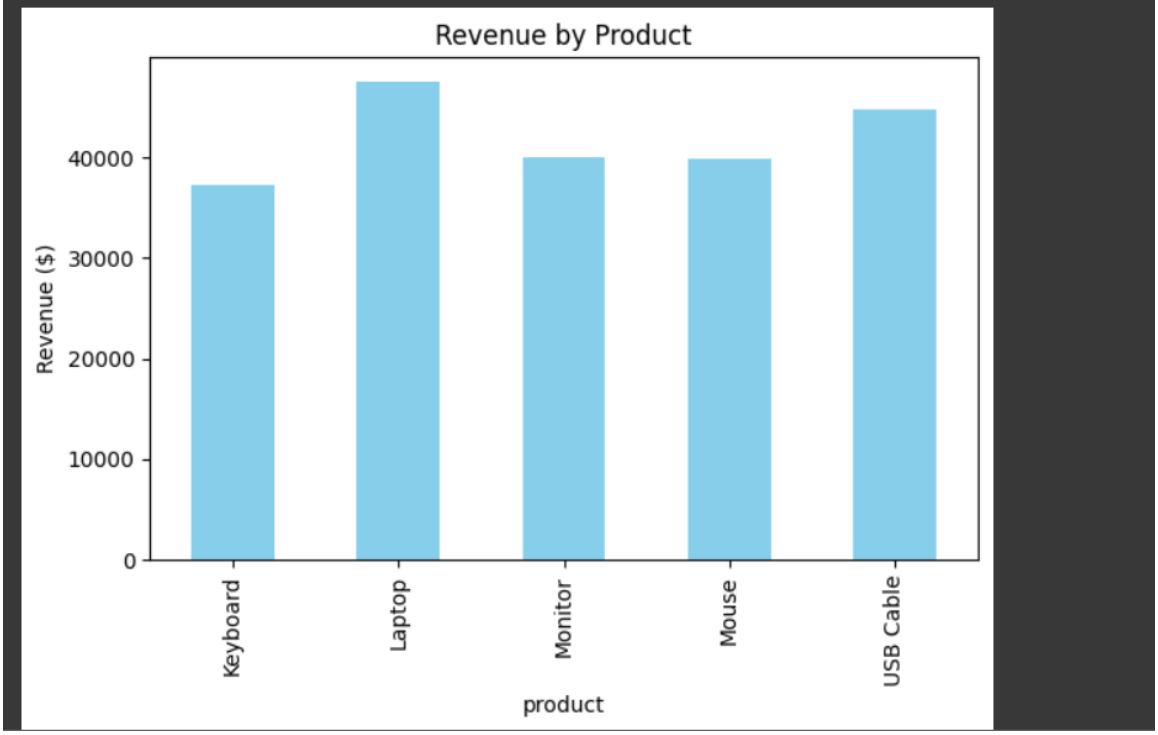
```
     product  total_qty   revenue
0   Keyboard        141  37212.19
1     Laptop        173  47570.19
2    Monitor        166  40066.80
3      Mouse        179  39929.31
4  USB Cable        175  44730.48
```



Step 4: Revenue Bar Chart by Product

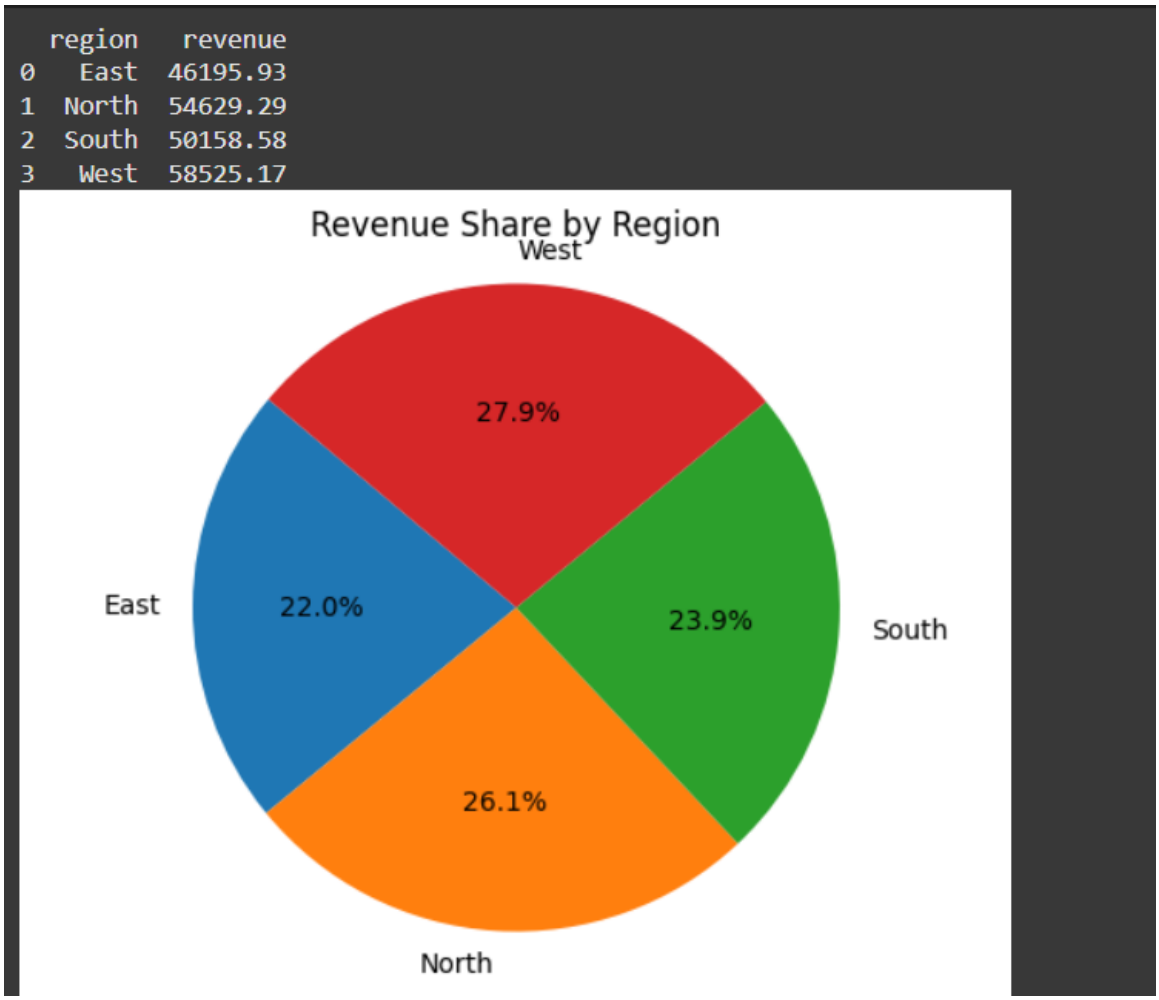
Plotted a separate bar chart to represent revenue per product using matplotlib and pandas DataFrame plotting.

```
# Bar chart of revenue per product
df_product_summary.plot(kind='bar', x='product', y='revenue', legend=False, color='skyblue')
plt.title("Revenue by Product")
plt.ylabel("Revenue ($)")
plt.tight_layout()
plt.show()
```



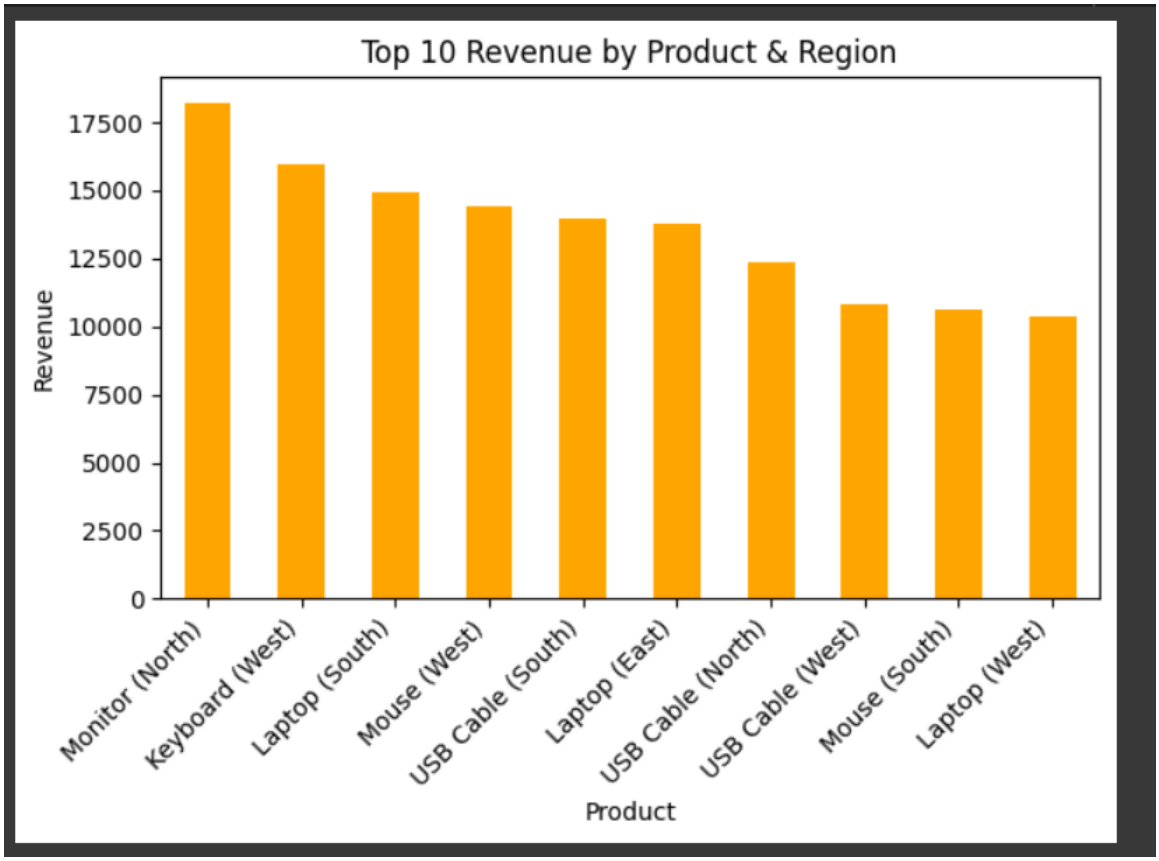
Step 5: Revenue Share by Region (Pie Chart)

Executed a GROUP BY query on the region column to sum up revenue by region.
Visualized the result using a pie chart to show revenue distribution across regions.



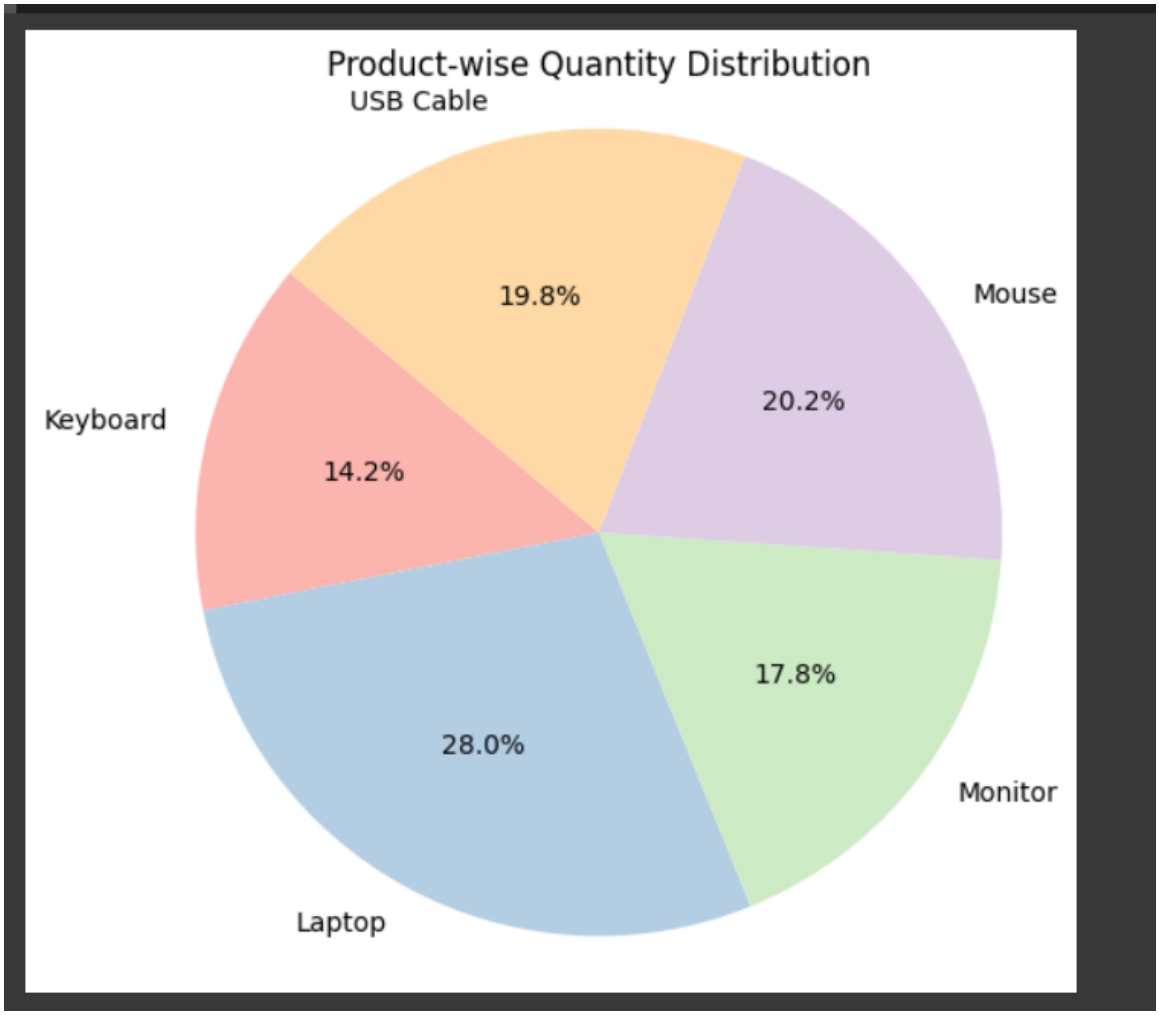
Step 6: Top 10 Product-Region Revenue Combinations

Executed a more advanced GROUP BY query to get revenue grouped by both product and region.
Sorted the result and displayed the top 10 combinations using a bar chart with labeled bars.



Step 7: Product-wise Quantity Distribution (Pie Chart)

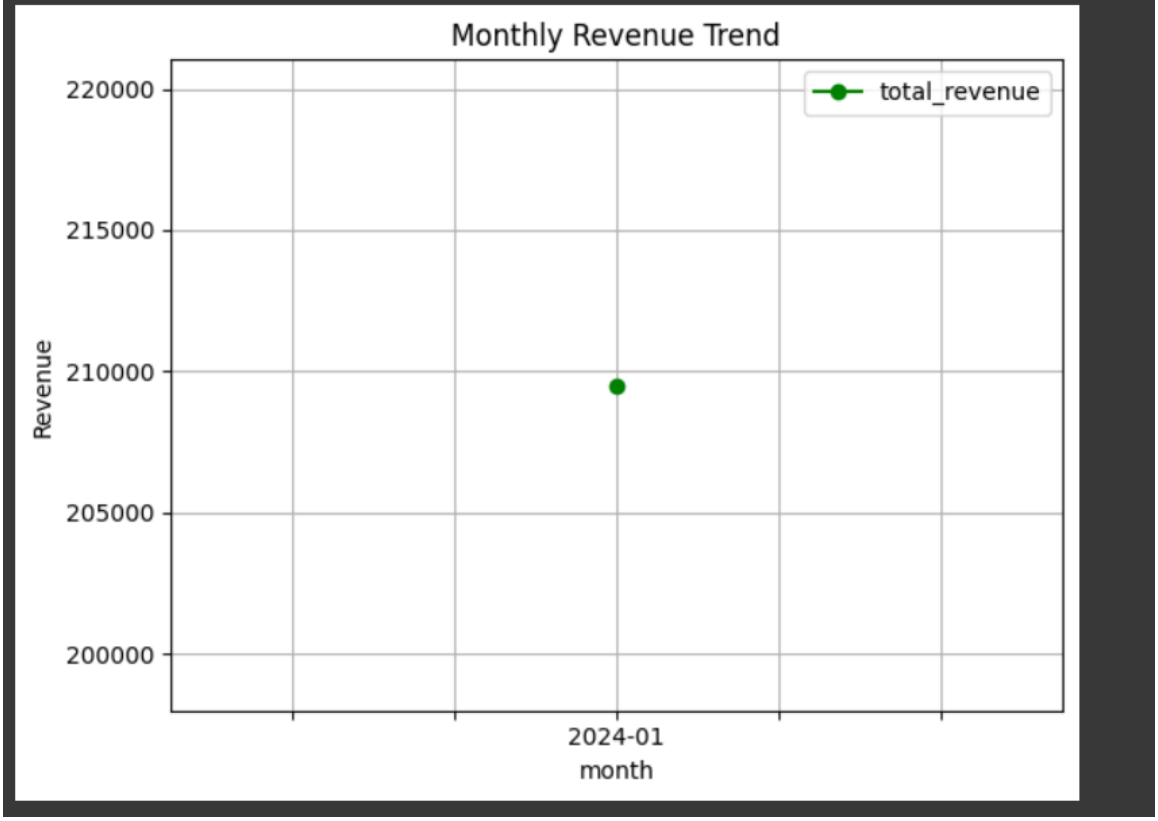
Queried the database to get total quantity sold per product and created a pie chart to show how quantity is distributed among different products.



Step 8: Monthly Revenue Trend (Line Chart)

Used SQLite's strftime function to group sales data by month, calculated monthly revenue, and visualized it using a line chart.

```
     month  total_revenue
0  2024-01      209508.97
```



Outcomes

- Generated and inserted synthetic data into SQLite.
- Mastered GROUP BY with multiple columns.
- Created multiple visualizations: bar charts, pie charts, and line charts.
- Analyzed data across multiple dimensions: product, region, and time.

Interview Preparation

- Q: How did you connect Python to the database?

A: Using sqlite3.connect('sales_data.db').

- Q: How was the data generated?

A: Used random and pandas date_range to generate 150+ synthetic sales records.

- Q: How did you group revenue and quantity by product?

A: Using SQL GROUP BY clause with SUM functions.

- Q: How was monthly trend analysis done?

A: By grouping sale_date using strftime('%Y-%m') and summing revenue.

- Q: What types of visualizations did you use?

A: Bar charts, pie charts, and line charts using matplotlib.

- Q: Why use pandas in this project?

A: To load SQL query results into DataFrames and for easy plotting and data manipulation.

- Q: How do you identify top performers?

A: By sorting aggregated revenue results and selecting the top records.

- Q: What chart was most insightful for you?

A: The line chart showing monthly revenue trend helps track business growth over time.