SALES DATA ANALYSIS SYSTEM

1. Project Definition.

Sales Data Analysis System analyzes business sales data to:

- Track revenue
- Identify top products & categories
- Analyze monthly sales trends
- Generate statistical and visual reports

2. Tools and Technologies used

Tool	Purpose
SQLite	Lightweight database to store sales data
Python (Pandas, SQLite3)	Data cleaning & calculations
R (ggplot2, dplyr)	Statistical analysis & visualization

3. Dataset(raw_sales.csv)

order_id,order_date,product,category,quantity,price,region
101,2024-01-05,Laptop,Electronics,2,55000,South
102,2024-01-10,Mobile,Electronics,1,20000,North
103,2024-02-01,Chair,Furniture,4,3500,West
104,2024-02-15,Table,Furniture,1,12000,South
105,2024-03-05,Headphones,Electronics,3,2500,East
106,2024-03-20,Laptop,Electronics,1,60000,North

4. SQLite Database

```
import sqlite3

conn = sqlite3.connect("sales.db")
cursor = conn.cursor()

cursor.execute("SELECT SUM(revenue) FROM sales")
print(cursor.fetchone())

conn.close()
```

5. Python: Data Cleaning & Database Operations

- Read CSV data
- Clean & transform data
- Calculate revenue
- Store data in SQLite
- Export processed data

```
import pandas as pd

import sqlite3

# Load raw sales data

df = pd.read_csv("../data/raw_sales.csv")



# Convert date column

df['order_date'] = pd.to_datetime(df['order_date'])



# Create revenue column

df['revenue'] = df['quantity'] * df['price']



# Connect to SQLite database

conn = sqlite3.connect("../database/sales.db")



# Store data in SQLite

df.to_sql("sales", conn, if_exists="replace", index=False)



# SQL Queries using Python

query_total_sales = """

SELECT SUM(revenue) AS total_revenue FROM sales;

"""
```

```
query_category_sales = """
SELECT category, SUM(revenue) AS category_revenue
FROM sales
GROUP BY category;
"""

total_sales = pd.read_sql(query_total_sales, conn)
category_sales = pd.read_sql(query_category_sales, conn)

print("Total Sales Revenue:")
print(total_sales)

print("\nCategory-wise Sales:")
print(category_sales)

# Export cleaned data for R
df.to_csv("../data/cleaned_sales.csv", index=False)
conn.close()
print("\nData processing completed successfully.")
```

6. R: Statistical Analysis & Visualization

- Summary Statistics
- Revenue distribution
- Monthly sales trends
- Category wise comparison

```
library(ggplot2)

library(dplyr)

# Load cleaned data

sales <- read.csv("C:/Users/LENOVO/Desktop/mtech\salesdata/cleaned_sales.csv")
```

- Summary Statistics

```
summary(sales$revenue)
```

- Revenue by Category

```
category_sales <- sales %>%

  group_by(category) %>%

  summarise(total_revenue = sum(revenue))

print(category_sales)
```

- BAR PLOT: Category-wise Revenue

```
ggplot(category_sales, aes(x = category, y = total_revenue, fill = category)) +

  geom_bar(stat = "identity") +

  labs(

    title = "Category-wise Revenue Comparison",

    x = "Product Category",

    y = "Total Revenue"

  ) +

  theme_minimal()
```

- Revenue Distribution (Histogram)

```
ggplot(sales, aes(x = revenue)) +
  geom_histogram(binwidth = 5000, fill = "steelblue", color = "black") +
  labs(
    title = "Revenue Distribution",
    x = "Revenue Amount",
    y = "Frequency"
  ) +
  theme_minimal()
```

- Monthly Sales Trend

```
sales$order_date <- as.Date(sales$order_date)
sales$month <- format(sales$order_date, "%Y-%m")
monthly_sales <- sales %>%
  group_by(month) %>%
  summarise(total_revenue = sum(revenue))
print(monthly_sales)
```

- LINE GRAPH: Monthly Trend

```
ggplot(monthly_sales, aes(x = month, y = total_revenue, group = 1)) +
  geom_line(color = "blue", size = 1) +
  geom_point(size = 3) +
  labs(
    title = "Monthly Sales Trend",
    x = "Month",
    y = "Total Revenue"
  ) +
theme_minimal()
```

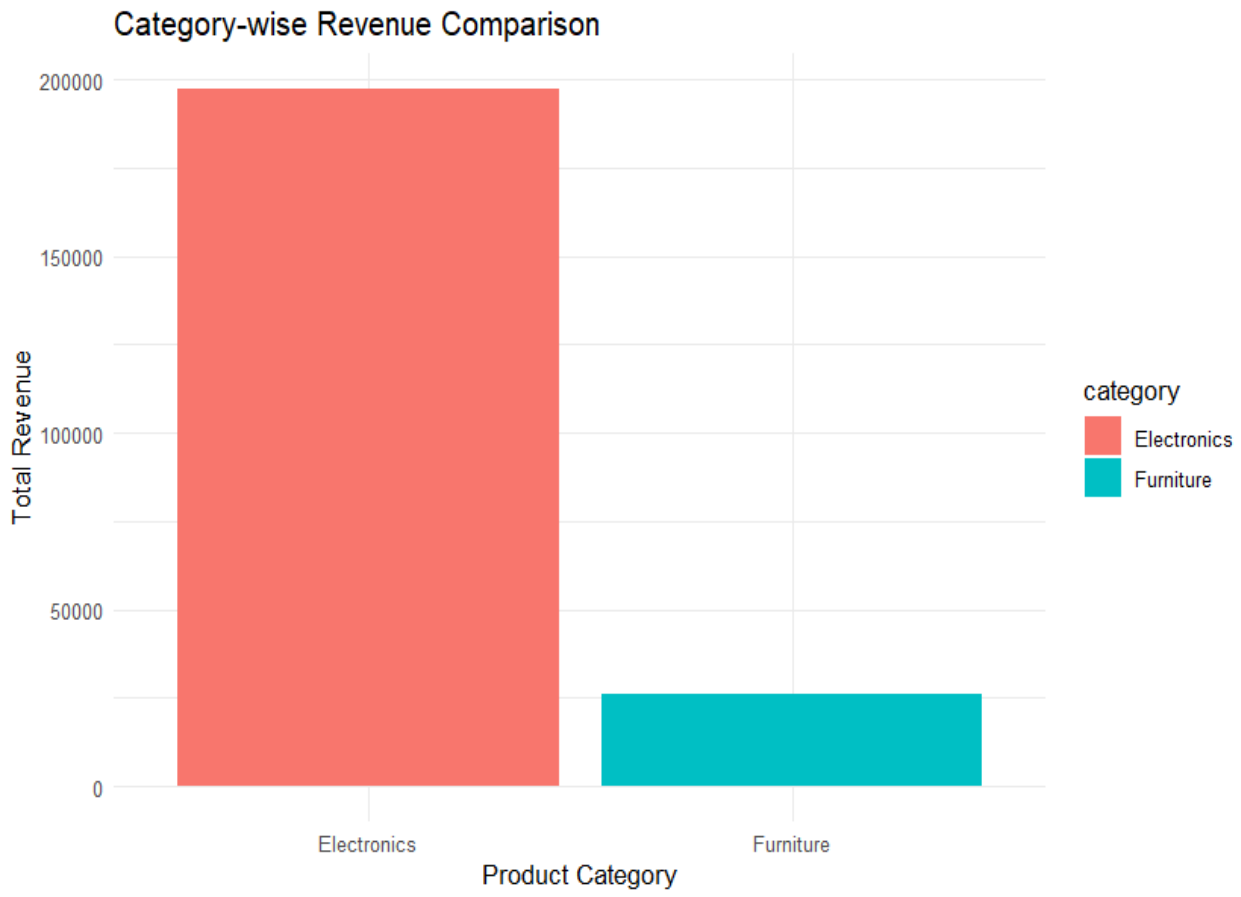
- BAR GRAPH: Monthly Revenue

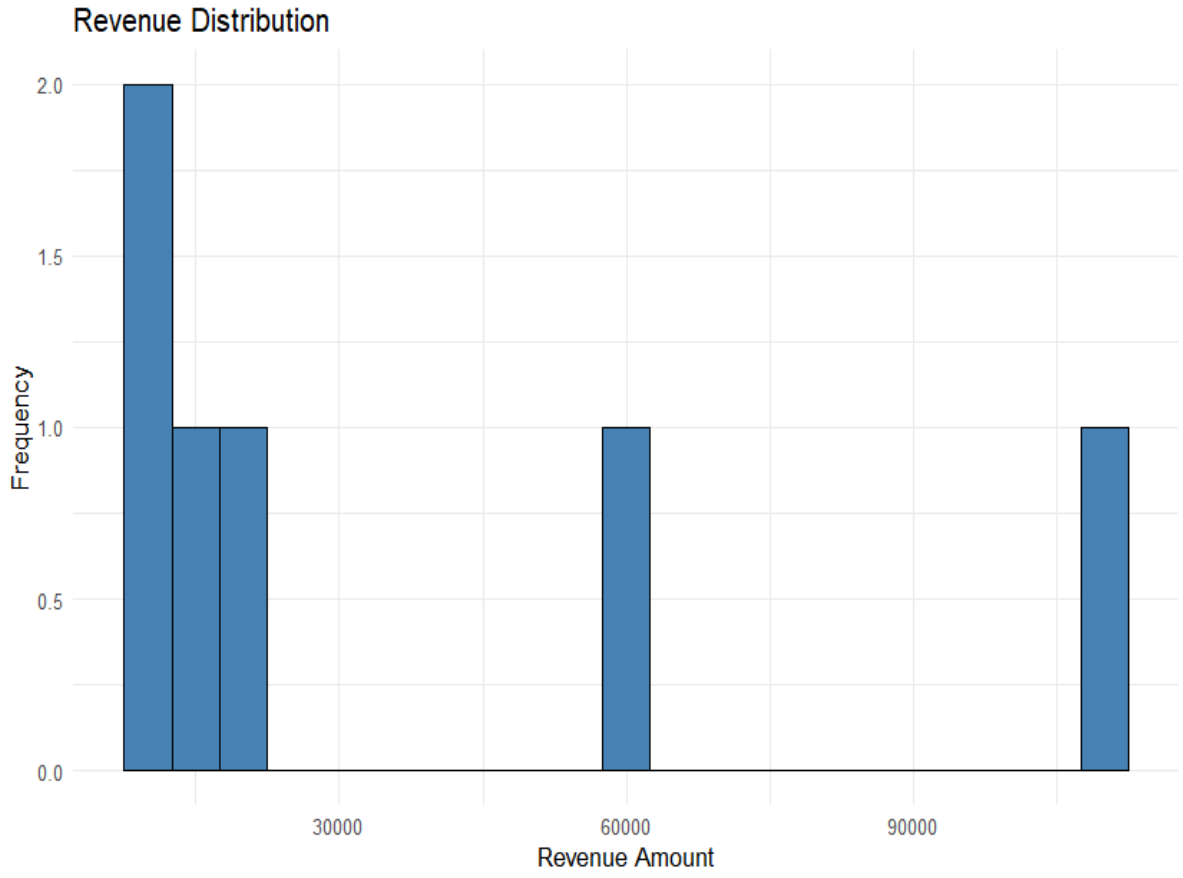
```
ggplot(monthly_sales, aes(x = month, y = total_revenue, fill = month)) +
  geom_bar(stat = "identity") +
  labs(
    title = "Monthly Revenue Comparison",
    x = "Month",
    y = "Total Revenue"
  ) +
  theme_minimal()
```

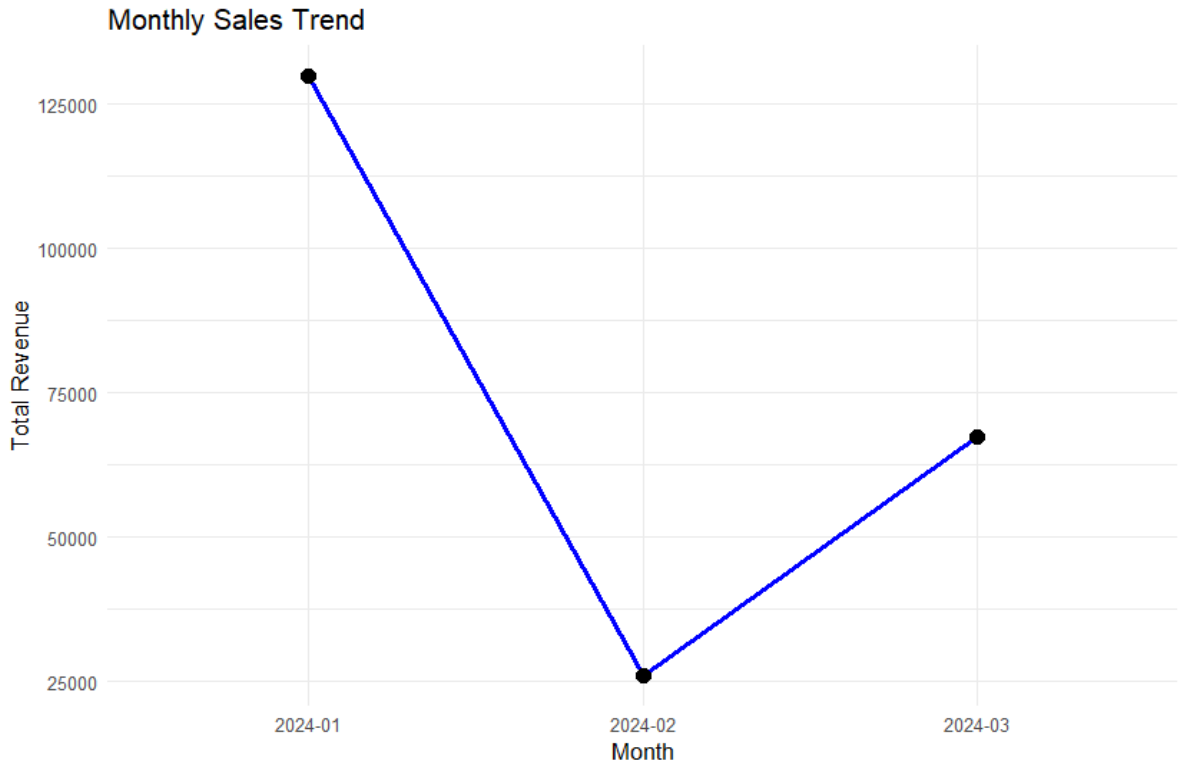
7. OUTPUT

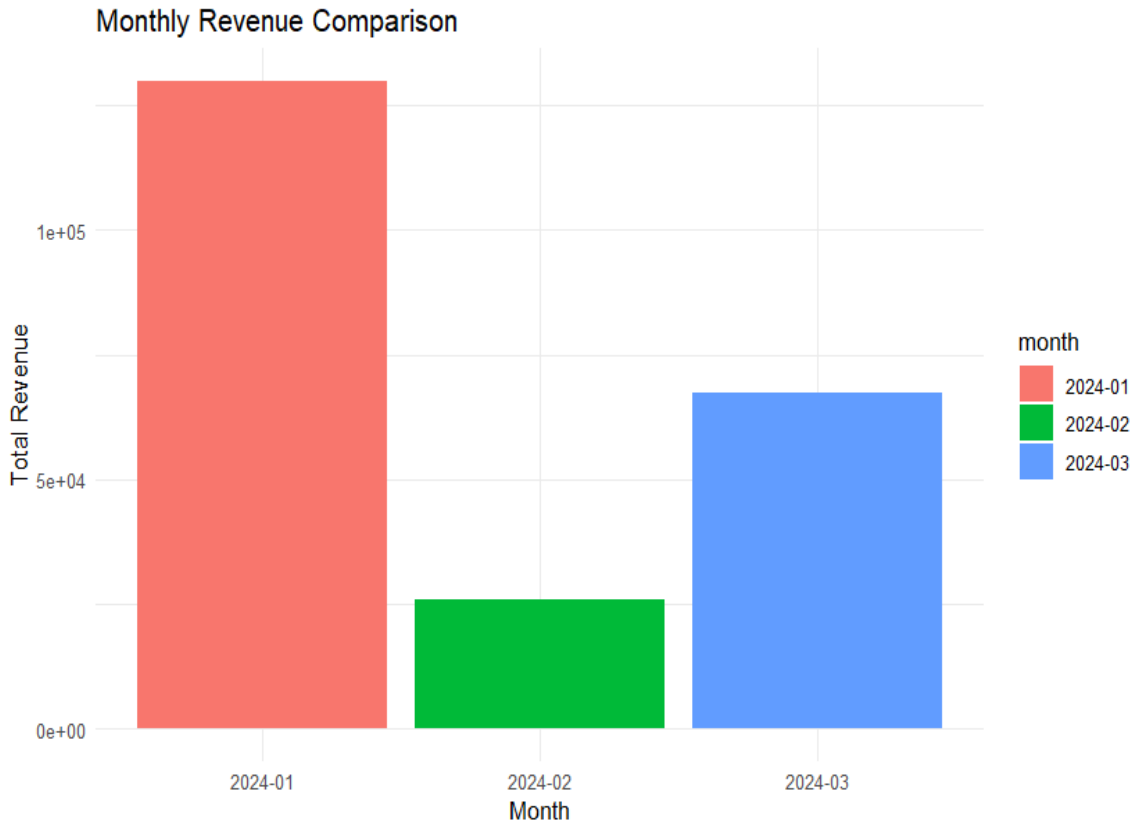
1) Cleaned_sales

```
order_id,order_date,product,category,quantity,price,region,revenue
101,2024-01-05,Laptop,Electronics,2,55000,South,110000
102,2024-01-10,Mobile,Electronics,1,20000,North,20000
103,2024-02-01,Chair,Furniture,4,3500,West,14000
104,2024-02-15,Table,Furniture,1,12000,South,12000
105,2024-03-05,Headphones,Electronics,3,2500,East,7500
106,2024-03-20,Laptop,Electronics,1,60000,North,60000
```


Category-wise Revenue Comparison
200000
150000
100000
50000
0
Total Revenue
Electronics
Furniture
Product Category
category
Electronics
Furniture


Revenue Distribution
2.0
1.5
1.0
0.5
0.0
Frequency
30000
60000
90000
Revenue Amount


Monthly Sales Trend
125000
100000
75000
50000
25000
Total Revenue
2024-01
2024-02
2024-03
Month


Monthly Revenue Comparison
1e+05
5e+04
0e+00
Total Revenue
2024-01
2024-02
2024-03
Month
month
2024-01
2024-02
2024-03