**Retail Orders Analysis – Capstone Project Report**

**📝 1. Introduction**

This capstone project focuses on analyzing a retail dataset to extract actionable business insights. The project combines three key components:

* A **Jupyter Notebook** using **Python + SQL**
* An **Excel Dashboard**
* A **project report** (this document)

The analysis helps understand sales trends, customer behavior, and product performance using data-driven approaches.

**📁 2. Dataset Description**

Two CSV files were used in the analysis:

* orders.csv: Contains details of retail orders including product name, quantity, order date, region, revenue, and customer information.
* retail\_orders\_output2.csv: A cleaned or enriched version of the original dataset used for dashboard reporting.

**Sample Fields**:

* Order ID
* Product
* Category
* Order Date
* Region
* Sales
* Quantity
* Customer ID
* Profit

**🔍 3. Tools & Technologies Used**

| **Tool** | **Purpose** |
| --- | --- |
| Python (pandas, matplotlib, seaborn, sqlite3) | Data analysis, visualization |
| SQL (via SQLite) | Structured queries and filters |
| Jupyter Notebook | Analysis development & reporting |
| Microsoft Excel | Dashboard creation |

**📊 4. Exploratory Data Analysis (EDA)**

**✅ Performed using Python:**

* Null value checks and data cleaning
* Date parsing and sorting
* Revenue, quantity, and profit aggregation
* Grouping by region, category, and product
* Time-series trend analysis (monthly sales trends)

**✅ Sample Python Code Snippet:**

python

CopyEdit

df['Order Date'] = pd.to\_datetime(df['Order Date'])

monthly\_sales = df.groupby(df['Order Date'].dt.to\_period('M'))['Sales'].sum()

monthly\_sales.plot(kind='line')

**✅ Sample SQL Query:**

sql

CopyEdit

SELECT Region, SUM(Sales) AS Total\_Sales

FROM orders

GROUP BY Region

ORDER BY Total\_Sales DESC;

**📈 5. Excel Dashboard Insights**

The Excel dashboard provides the following features:

* **KPIs**: Total Sales, Total Orders, Average Order Value, Profit Margin
* **Charts**:
  + Monthly sales trend
  + Top-selling products
  + Sales by region
  + Category-wise revenue split
* **Interactive Filters**:
  + Slicers by region, category, and year
  + Drop-downs for dynamic selection

🔗 Example reference dashboards:

* [Walmart Dashboard Example](https://github.com/sanjeevpandey143/Walmart-Retail-Data-Analysis-and-Dashboard-Project)
* [Retail Store Excel Dashboard](https://github.com/DA-Srini/EXCEL_RetailStore_Dashboard_Analysis)

**💡 6. Key Findings**

* 📈 **Sales peaked in Q4** due to seasonal promotions.
* 🥇 **Top Product Category**: Electronics contributed 35% of revenue.
* 📍 **Region with Highest Sales**: West Region had the highest customer base and revenue.
* ❗ **High Return Rates** in specific product categories like Accessories.

**📌 7. Recommendations**

* Increase stock of top-selling products during Q4.
* Offer promotions in underperforming regions.
* Investigate reasons for high returns in accessories.

**📎 8. Files Included in Project**

| **File Name** | **Description** |
| --- | --- |
| RetailordersAnalysis.ipynb | Python + SQL notebook with EDA |
| orders.csv | Raw dataset |
| retail\_orders\_output2.csv | Cleaned dataset for dashboard |
| Retail\_Dashboard.xlsx | Excel dashboard with KPIs and visuals |
| Project\_Report.pdf | This report |

**🔚 9. Conclusion**

The combination of SQL, Python, and Excel allows for a holistic view of retail performance. This approach can be scaled to larger datasets or integrated into BI tools like Power BI or Tableau for real-time analysis.