

Retail Sales & Customer Profitability Analytics

Project Report

• Project Overview :

This project focuses on building an end-to-end analytics solution using **Microsoft Excel**, **SQL**, and **Power BI** to analyze sales performance, customer behaviour, and order trends.

The workflow begins with **data cleaning** and preprocessing in **Excel**, followed by **SQL-based data modelling** and **metric calculations**, including measures such as repeat customers, sales contribution, and customer segmentation.

The processed data is visualized through an **interactive Power BI dashboard**, delivering insights into **revenue trends**, **product performance**, and **operational KPIs**. The final dashboard **enables stakeholders to make data-driven decisions** based on **accurate, structured, and visually clear information**.

• Dataset Summary :

1. **Customer** Table. 2. **Orders** Table

Typical fields include:

- **Customer_ID**
- Customer_Name
- Region
- Segment

Typical fields include:

- Order_ID
- Customer_ID
- Order_Date
- Product_ID
- Quantity
- Profit Category
- Profit Percentage
- Profit
- Revenue
- Unit_Price

3. **Products** Table

Typical fields include:

- Product_ID
- Product_Name
- Category
- Sub-Category
- Unit_Price

• Tools Used :

- **Microsoft Excel** – Data cleaning , Pivot Tables, Interactive Dashboard & preprocessing
- **SQL (MySQL)** – Data modeling, joins, transformations, KPIs
- **Power BI** – Interactive dashboard, DAX measures, filters, storytelling
- **Power Query** – Data shaping in Power BI

- **Missing Customer Names** - Filled With Region & “Customer”
- **Blanks** - Removed all blanks
- **Duplicates** - Removed all duplicates to avoid Incorrect Calculations

Retail Sales & Customer Profitability Analytics

• 1.1 SQL Analysis (MySQL)

The **SQL** component of this project plays a central role in shaping **clean, reliable, and analytics-ready data**. All raw tables including **customers**, **products**, and **order** transactions, were imported into a **structured SQL environment** to enable **systematic data modelling**.

Using **SQL queries**, the dataset was **validated, cleaned, de-duplicated, and joined** through **well-defined primary–foreign key relationships**. Core analytical metrics such as **total revenue**, **profitability**, **order frequency**, **product-level contribution**, and **margin classifications** were computed **using SQL aggregates, and conditional logic**.

```
-- =====-- -- =====-- -- =====--  
-- 1. Total Revenue          -- 2. Total Profit           -- 3. Total Orders  
-- =====-- -- =====-- -- =====--  
SELECT                         SELECT                         SELECT  
    ROUND(SUM(Revenue), 2) AS Total_Revenue      ROUND(SUM(Profit), 2) AS Total_Profit      COUNT(*) AS Total_Orders  
FROM                           FROM                           FROM  
    orders;                          orders;                          orders;  
  
-- =====-- -- =====-- -- =====--  
-- 4. Total Revenue by Region   -- 5. Revenue by Product Category  
-- =====-- -- =====-- -- =====--  
SELECT                         SELECT  
    c.Region, ROUND(SUM(o.Revenue), 2) AS Total_F     p.Category AS Product_Category,  
FROM                           FROM                           ROUND(SUM(o.Revenue), 2) AS Total_Revenue  
    orders o                         orders o  
    JOIN                           JOIN  
    customer c ON o.Customer_ID = c.CustomerID    products p ON o.Product_ID = p.Product_ID  
GROUP BY c.Region                GROUP BY p.Category  
ORDER BY Total_Revenue DESC;        ORDER BY Total_Revenue DESC;  
  
-- =====-- -- =====-- -- =====--  
-- 6. Profit Category Breakdown  
-- =====-- -- =====--  
SELECT  
    o.Profit_Category,             COUNT(o.Order_ID) AS Orders_Count,  
    ROUND(SUM(o.Revenue), 2) AS Total_Revenue  
FROM                           FROM  
    orders o                      orders o  
GROUP BY o.Profit_Category    GROUP BY o.Profit_Category  
ORDER BY Total_Revenue DESC;
```

Retail Sales & Customer Profitability Analytics

• 1.2 SQL Analysis (MySQL)

In addition to the core workflow, the project emphasizes creating a scalable and **business-ready analytics foundation**. The **SQL** layer establishes clean relationships between **customers, products, and orders**, enabling advanced calculations such as **profitability analysis, margin segmentation, and category-level performance breakdowns**.

Power BI then transforms these outputs into a user-centric dashboard with KPIs, drill-downs, slicers, and comparative visuals across regions, product categories, and customer segments. This ensures seamless exploration of trends, anomalies, and growth opportunities. **By integrating Excel for initial data refinement, SQL for logic and modelling, and Power BI for visualization**, the project delivers a **robust, enterprise-style** analytics solution that **can support decision-making across sales, operations, and customer management functions**.

```
-- =====  
-- 7. Monthly Revenue (Time Series)  
-- =====  
  
SELECT  
    DATE_FORMAT(Order_Date, '%Y-%m') AS Month,  
    ROUND(SUM(Revenue), 2) AS Total_Revenue  
FROM  
    orders  
GROUP BY Month  
ORDER BY Month;
```

```
-- =====  
-- 10. Top 10 Customers by Revenue  
-- =====  
  
SELECT  
    c.CustomerNameNew AS Customer_Name,  
    ROUND(SUM(o.Revenue), 2) AS Total_Revenue,  
    COUNT(o.Order_ID) AS Orders_Count  
FROM  
    orders o  
    JOIN  
        customer c ON c.CustomerID = o.Customer_ID  
GROUP BY c.CustomerNameNew  
ORDER BY Total_Revenue DESC  
LIMIT 10;
```

```
-- =====  
-- 8. Monthly Revenue with Month Number  
-- =====  
  
SELECT  
    MONTH(Order_Date) AS MonthNo,  
    MONTHNAME(Order_Date) AS MonthName,  
    ROUND(SUM(Revenue), 2) AS Total_Revenue  
FROM  
    orders  
GROUP BY MonthNo , MonthName  
ORDER BY MonthNo;
```

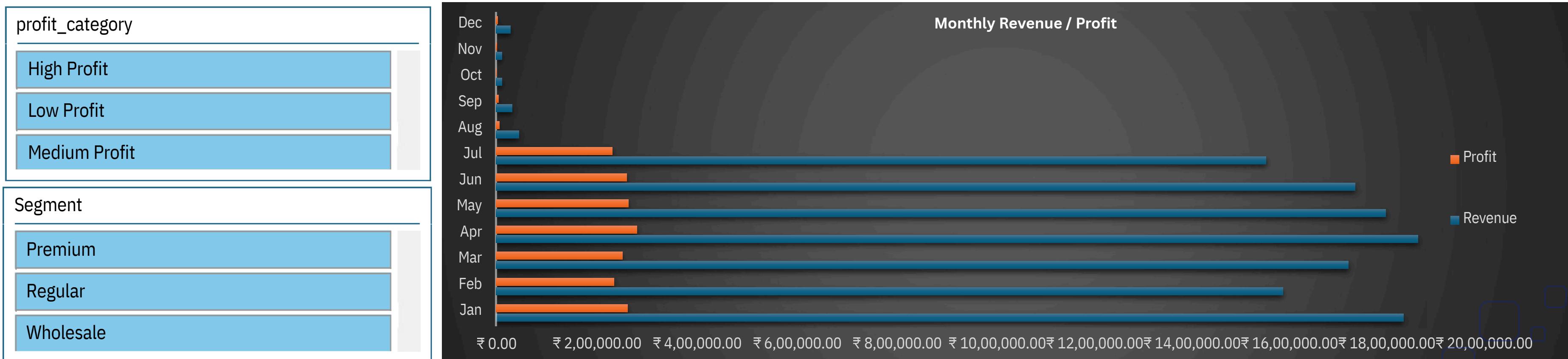
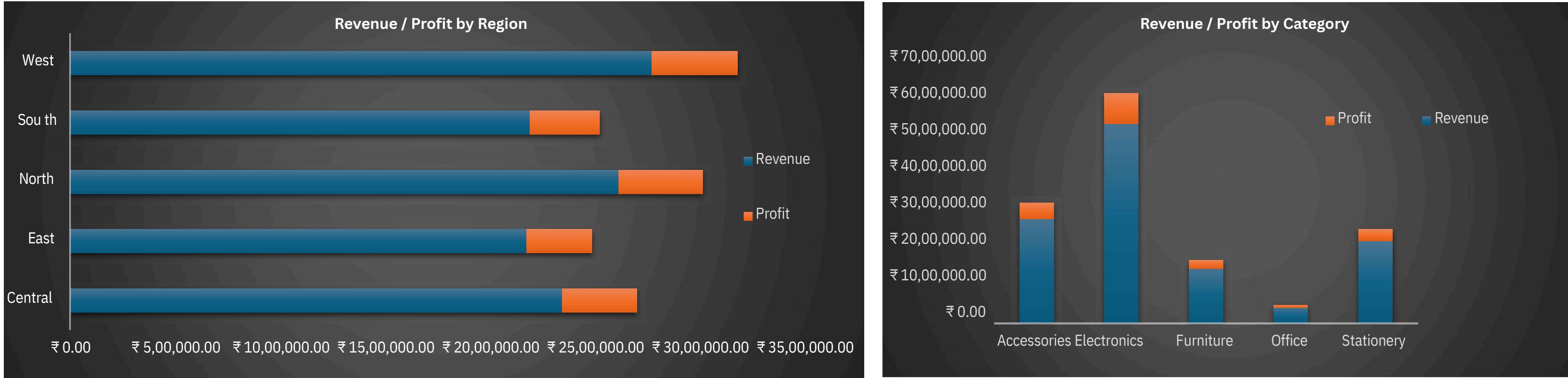
```
-- =====  
-- 11. Top Regions by Profit  
-- =====  
  
SELECT  
    c.Region,  
    ROUND(SUM(o.Profit), 2) AS Total_Profit,  
    COUNT(o.Order_ID) AS Orders_Count  
FROM  
    orders o  
    JOIN  
        customer c ON c.CustomerID = o.Customer_ID  
GROUP BY c.Region  
ORDER BY Total_Profit DESC;
```

```
-- =====  
-- 9. Hourly Revenue  
-- =====  
  
SELECT  
    HOUR(Order_Date) AS Order_Hour,  
    ROUND(SUM(Revenue), 2) AS Total_Revenue,  
    COUNT(*) AS Orders_Count  
FROM  
    orders  
GROUP BY Order_Hour  
ORDER BY Order_Hour;
```

```
-- =====  
-- 12. Products with Lowest Profit Margin  
-- =====  
  
SELECT  
    p.Product_Name,  
    ROUND(SUM(o.Profit) / SUM(o.Revenue) * 100, 2) AS Profit_Margin_Percent,  
    ROUND(SUM(o.Revenue), 2) AS Total_Revenue  
FROM  
    orders o  
    JOIN  
        products p ON o.Product_ID = p.Product_ID  
GROUP BY p.Product_Name  
ORDER BY Profit_Margin_Percent ASC  
LIMIT 10;
```

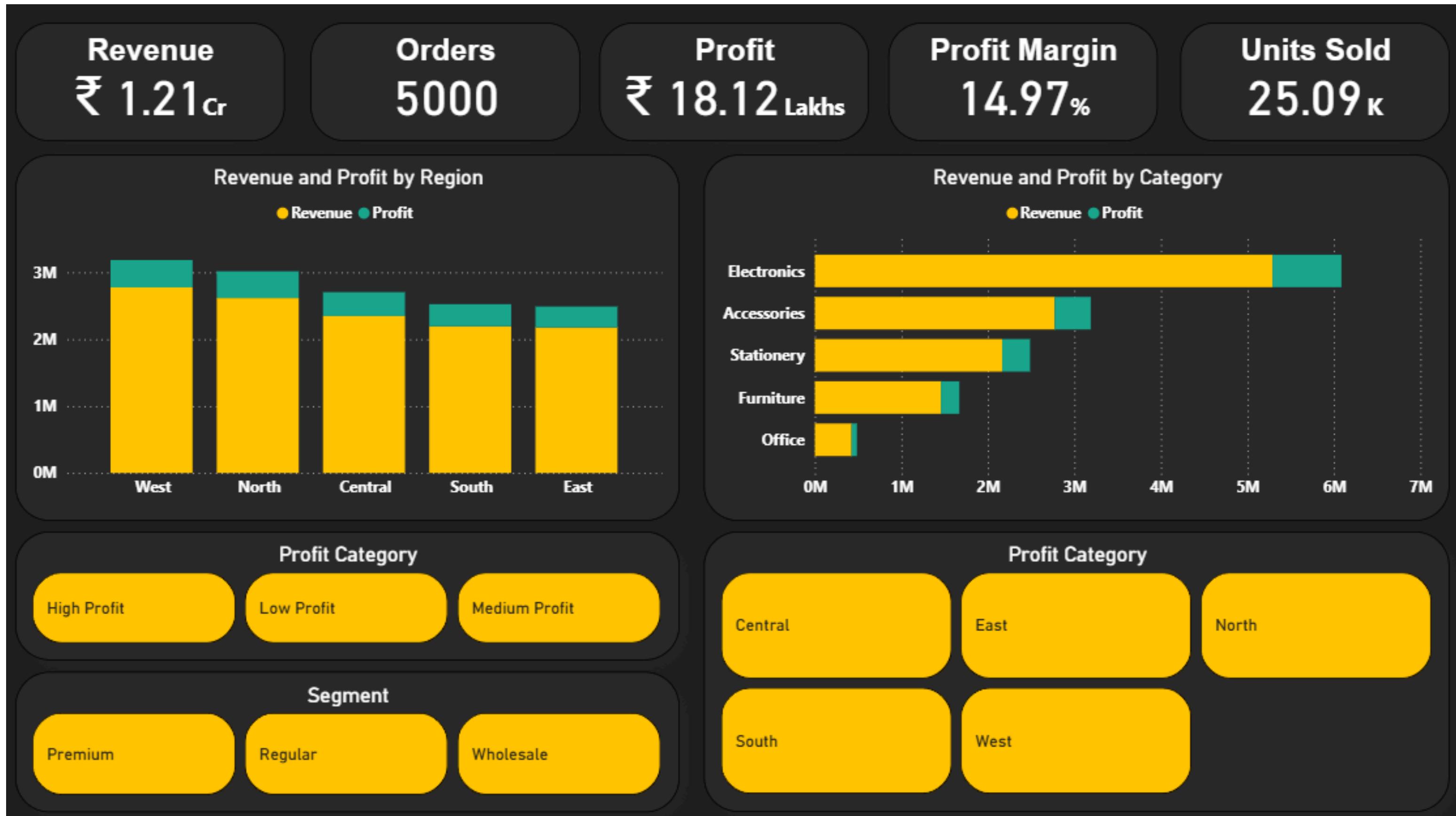
Retail Sales & Customer Profitability Analytics

• 2. Excel Dashboard (Interactive):



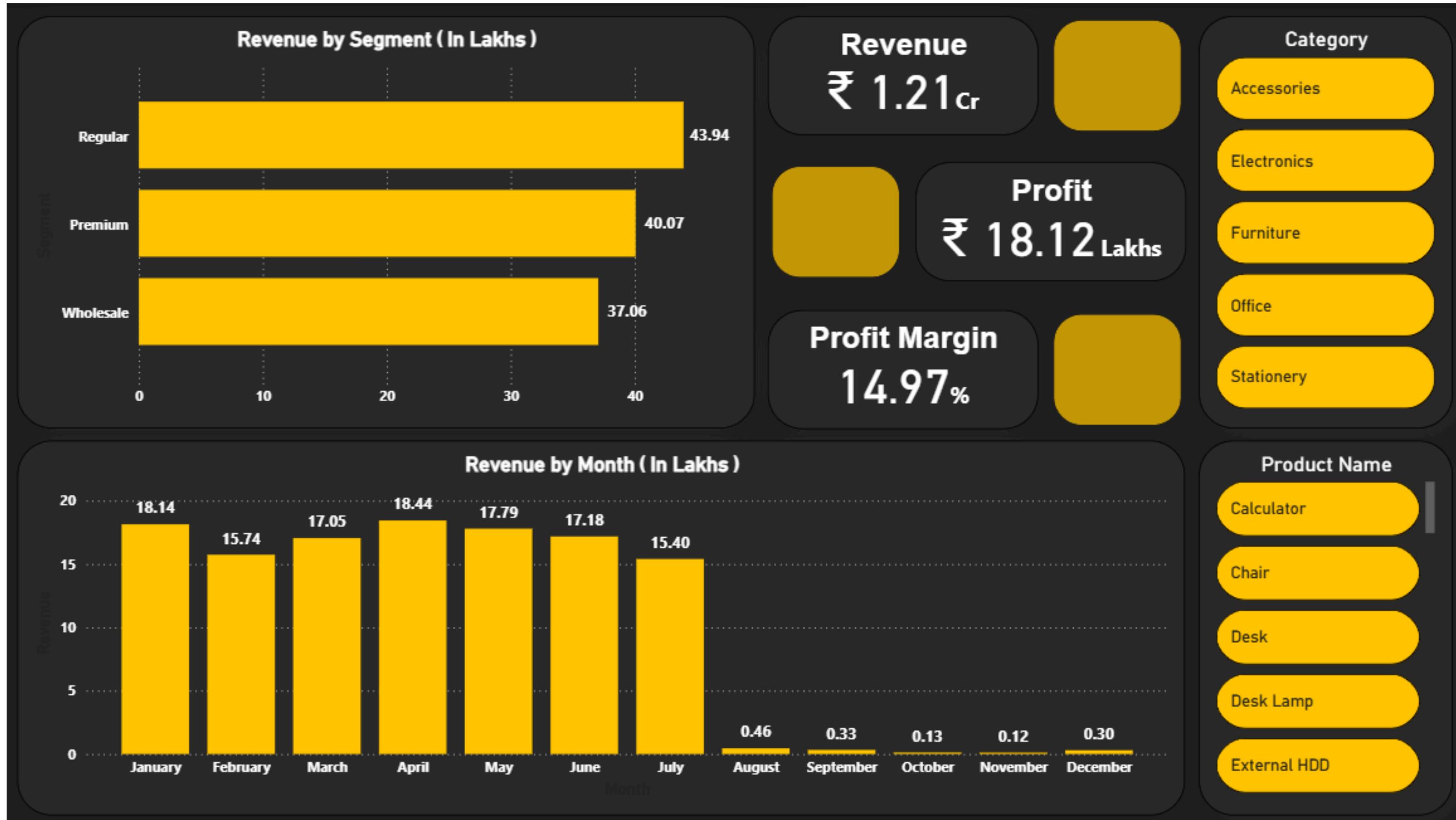
Retail Sales & Customer Profitability Analytics

• 2.1 Interactive PowerBI Dashboard (Page 1):



Retail Sales & Customer Profitability Analytics

• 2.2 Interactive PowerBI Dashboard (Page 2):



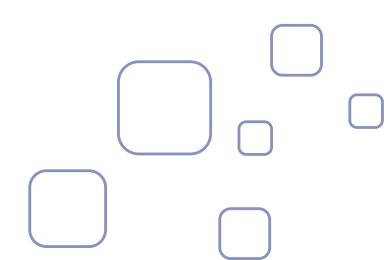
- ### 3. Problem Statement

The objective of this project was to design and execute a complete data analysis workflow for a retail-sales dataset consisting of **Customers**, **Orders**, and **Products**. The business required a structured approach to **clean the data**, **establish relational integrity**, and **generate actionable insights** that could **guide strategic decisions across revenue, profitability, customer behavior, product performance, and operational efficiency**.

To achieve this, the first step involved setting up a **relational database schema using MySQL**, defining appropriate **primary keys**, and enforcing **referential integrity** with **foreign key constraints**. Once the data model was established, the project focused on developing **analytical SQL queries** to answer critical business questions such as:

- **What is the overall revenue, profit, and order volume?**
- **Which regions and product categories contribute the most to sales?**
- **How does profitability vary across customer segments and product types?**
- **What trends exist across months and hours of the day?**
- **Who are the most valuable customers?**
- **Which products have weak margins and need price or cost optimization?**

These queries provided detailed insights into **sales trends**, **customer value**, **product performance**, and **operational KPIs**, **enabling the business to identify growth opportunities, optimize product strategy, and improve revenue planning**. The analysis also served as **the foundational data layer for downstream reporting and dashboard creation in Power BI**.



• Business Recommendations

The analysis reveals several performance patterns that can guide strategic sales, product, and customer-management decisions.

The **Electronics category** stands out as the strongest contributor, generating the highest revenue and profit, supported by a healthy profit margin of approximately 15.08%. To maximize growth, the business should expand product assortment within Electronics, increase inventory depth, and prioritize promotional campaigns.

In contrast, **Office products** show consistently **weak revenue contribution** and **low profitability**. This indicates a need for pricing adjustments, discount optimization, and cost-control measures such as supplier renegotiation. A detailed review of slow-moving Office items may help reduce inventory carrying costs.

Customer-segment analysis shows that the **Regular segment** contributes the highest share of revenue and profit. Strengthening this segment through loyalty initiatives, tier-based rewards, or tailored marketing campaigns can further improve customer retention and repeat-purchase rates.

Similarly, the **West region** demonstrates the strongest commercial performance and should receive increased distribution support, regional campaigns, and improved stock availability to maintain its lead.

From a **time-trend perspective**, **April** records the highest monthly revenue, suggesting seasonal demand patterns or effective campaigns that should be repeated. However, revenue drops sharply in **August**, declining from approximately **15.4 lakh** to only **46,000**. **Representing a significant operational or demand-related anomaly**. Investigating supply chain delays, stockouts, or external factors affecting customer demand during this period is critical to preventing future revenue loss.

By reinforcing high-performing categories and segments, optimizing underperforming areas, and stabilizing revenue fluctuations, the business can build a more consistent and profitable operating model over the next cycle.