



**SCHOOL OF COMPUTING, ENGINEERING AND DIGITAL  
TECHNOLOGIES  
ICA REPORT**

**CIS4008-N-BF1 Big Data and Business Intelligence**

**Global Superstore**

**Sruthi Dubba**

**S3544788**

**MODULE LEADER: Dr Ala Alkafri**

**Academic year:2025-2026**

**January 2026**

## CONTENTS

1.EXECUTIVE SUMMARY	-----	01
1.1 Introduction	-----	01
1.2 Key Findings	-----	01
1.3 Recommendations	-----	02
2. INTRODUCTION.	-----	02
2.1 Data Source	-----	03
2.2 BI Questions	-----	04
3.FINDINGS BASED ON ANALYSIS AND EVALUATION	-----	06
Q1. How is the business performing overall, and is growth stable or volatile over time and across sectors?	-----	06
Q2. Which factors control profitability, and where are the major profit risks in the business?	-----	14
Q3. How do geography and product categories influence profit contribution and overall business performance?	-----	21
Q4. What are the key drivers behind profit variations and profit loss, and how can AI techniques help identify root causes affecting profitability?	-----	23
Q5 How can machine learning techniques be used to predict profit behaviour and support profit optimisation decisions?	-----	27
4. CONCLUSIONS AND RECOMMENDATIONS	-----	33
4.1 Summary	-----	33
4.2 Recommendations	-----	33
4.3 Personal Conclusion	-----	34
APPENDIX: BI DESIGN	-----	34
1.DATA PRE-PROCESSING AND DATA CLEANING	-----	36
1.1 Load and Clean Dataset	-----	36

1.2 Creating Dimensional Tables for Relationships:	-----	36
2.DATA MODELLING	-----	38
3.DAX AND M-LANGUAGE	-----	39
3.1 DAX	-----	39
3.2 Maching learning with R	-----	40
4.DASHBOARD DESIGN	-----	45
4.1 Home Page	-----	45
4.2 Overall Business Performance & Profitability Overview	-----	46
4.3 Profit & Performance Deep-Dive Analysis	-----	48
4.4 Geographical and Category-Level Profit Insights	-----	49
4.5 AI-Driven Profitability Drivers & Root Cause Analysis	-----	49
4.6 Predictive Analytics for Profit Optimisation Using Machine Learning	-----	50
5.REFERENCE	-----	52

# 1. EXECUTIVE SUMMARY

To achieve sustainable business performance, organizations must put data-driven insights at the center of their decision-making processes. While many factors affect sales and profitability, meaningful analysis of historical data provides the basis for identifying trends, patterns and performance drivers. Such analysis is essential to guide strategic decisions and support long-term business planning.

## 1.1. Introduction

This report is a case study of a retail business that uses business intelligence and advanced analytics techniques to analyse sales performance, profitability, delivery efficiency and customer behavior. The goal of the organization is to understand historical performance trends and identify factors that affect profits, sales growth and operational efficiency. The following key questions were addressed in this report:

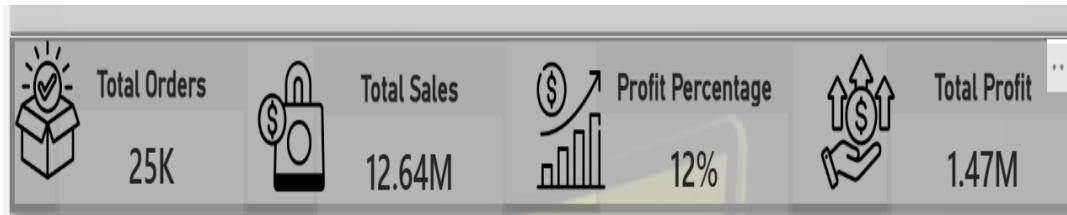
- What is the overall sales and profit performance of the business over time?
- Which areas, categories and subcategories contribute most and least to profitability?
- Are there products causing losses despite high sales volumes?
- How do discounts and delivery performance affect profitability?
- Can machine learning techniques be used to predict profits and aid decision making?

The dataset used for this case study was obtained from [Kaggle](#) and was chosen because it met the requirements of the business questions and allowed the use of Business Intelligence, Artificial Intelligence and Machine Learning techniques. The original flat dataset was transformed through data cleansing and modeling processes, including the creation of multiple dimension tables and measures to support interactive analysis in Microsoft Power BI.

## 2.2 Key Findings

- The dataset includes several years of transactional sales data across regions, categories, subcategories and customer segments.
- Total sales show a positive growth trend over time, which indicates continued expansion of business operations.
- Almost certain subcategories contribute disproportionately to overall profits, while many products generate consistent losses.
- Higher discount levels are strongly associated with lower profit margins, confirming discounts as a key driver of profit erosion.
- Some sectors generate high sales revenues but relatively low profits, highlighting inefficiencies in cost management or pricing strategies.
- Delivery performance varies widely between regions, with some regions recording significantly higher average delivery times.
- Products and sectors with long delivery times show weak profitability.

- Regression-based machine learning analysis reveals a positive linear relationship between sales and profits, supporting the use of predictive analytics.
- Machine learning models implemented using R in Power BI successfully demonstrate the ability to predict business performance trends.
- Overall, the findings suggest that profitability optimization depends on better discount control, assessment of product mix and improvements in operational efficiency.



## 2.3 Recommendations

Based on the findings from this analysis, the following recommendations are proposed

- Review and optimize discount strategies by introducing discount limits that protect profitability while encouraging sales growth.
- Focus on high-performing product subcategories and reevaluate or terminate products that consistently lose money.
- Improve delivery efficiency in areas with excessively long delivery times by rethinking logistics partners and shipping methods.
- Implement result-focused KPIs together with sales KPIs to ensure that revenue growth does not lead to a decline in profitability.
- Leverage machine learning models for profit forecasting to support proactive decision making and scenario analysis.
- Extend the use of AI-powered graphics to continuously monitor key profit drivers and detect emerging risks early.
- Overall, this project demonstrates that the combination of business intelligence, AI visualization and machine learning provides a powerful framework for understanding business performance and supporting data-driven strategic decisions.

## 2. INTRODUCTION

In today's data-driven business environment, organizations increasingly rely on Business Intelligence (BI) systems to monitor performance, identify inefficiencies and assist in strategic decision making. While traditional BI focuses on descriptive reporting, modern analytical solutions integrate advanced analytics and artificial intelligence (AI) to uncover deep insights and predict future outcomes. Therefore, it has become necessary to analyze sales, profitability and operating results in order to understand the company's sustainability and competitive advantage.

This project presents a comprehensive BI and analytics solution developed using Microsoft Power BI, supported by machine learning technologies implemented through DAX, M language (Power Query) and R scripts. A global sales dataset was analyzed to evaluate business performance across multiple dimensions, including products, regions, discounts, shipping modes and delivery performance. The analysis addresses key business issues related to sales sustainability, profit drivers, operational inefficiencies and the impact of discounting strategies, while demonstrating the application of predictive analytics in a BI environment.

To ensure efficient analysis and best practice data modeling, the original flat data set was transformed into a star schema data model, consisting of a central fact table linked to several dimension tables. Interactive dashboards were then designed to visualize insights using a variety of charts, AI-powered visuals and machine learning models. By combining traditional BI techniques with AI and ML, this project highlights how advanced analytics can improve business understanding, support data-driven decisions and provide actionable recommendations.

## 2.1 Data Source

The dataset used for this project is the Global Superstore dataset, obtained from [Kaggle](#). This dataset was chosen because it closely represents a real retail organization and contains sufficient transactional, customer, product and logistics information necessary to answer advanced business intelligence and analytical questions.

The original data set was provided as a single flat table in Excel format. To improve data organization, analytical performance, and support advanced Power BI features such as AI visualizations and machine learning integration, the dataset was transformed into a star schema-based data model. In total, five tables were created from the original data set, which ensures a clear separation of facts and dimensions.

### **Table Creation Description:**

In this model, Globalstore acts as a fact table, as it contains important transactional metrics such as sales, profit, volume, discounts, shipping costs, and delivery days. Surrounding the fact table are several dimension tables, including customer, product, date, and shipment mode, which provide descriptive attributes used to partition, filter, and aggregate data across different analytical approaches.

Detail description of the Fact and dimension tables given below:

<b>Table Name</b>	<b>Column Name</b>	<b>Description</b>
Globalstore (Fact table)	Order ID	Unique identifier for each customer order
	Order Date	Date on which the order was placed
	Ship Date	Date on which the order was shipped
	Sales	Total sales value of each order
	Profit	Profit generated from each order
	Quantity	Number of items ordered
	Discount	Discount applied to the order
	Shipping Cost	Cost incurred to ship the order
	Delivery Days	Calculated difference between Order Date and Ship Date
	Customer ID	Foreign key linking to Customer table
	Product ID	Foreign key linking to Product table
	Region	Geographical region of the order
	Market	Market in which the sale occurred
	Segment	Customer segment classification
	Ship Mode	Mode of shipment used
Customer (Dimension Table)	Customer ID	Unique identifier for each customer
	Customer Name	Name of the customer
	Segment	Segment to which the customer belongs
	City	City City of the customer
	State	State of the customer
	Country	Country of the customer
	Profit Margin %	Calculated measure representing customer profitability
Product(Dimension Table)	Product ID	Unique identifier for each product
	Product Name	Name of the product
	Category	High-level product classification

Date (Dimension Table)	Date	Calendar date
	Year	Year of the order
	Quarter	Quarter of the year
	Month	Month name
	Month Number	Numeric representation of the month
	YearMonth	Combined Year and Month field for trend analysis
Ship Mode (Dimension Table)	Ship Mode	Type of shipment used (Standard, First Class, etc.)

Bi Questions:

Q1. How is the business performing overall, and is growth stable or volatile over time and across sectors?

#### **Sub questions**

- a. How have total sales and profit evolved over time?
- b. Which sectors and product categories generate high sales but have a negative impact on profitability?
- c. How does the profit contribution vary across product subcategories?
- d. What is the current overall business sales in terms different factors?

Q2. Which factors control profitability, and where are the major profit risks in the business?

#### **Sub questions**

- a. Which categories and subcategories generate the highest and lowest profit margins?
- b. Which months had the lowest profits and what patterns explain this decline?
- c. What is the impact of different shipping modes on order volume and profitability?

Q3. How do geography and product categories influence profit contribution and overall business performance?

- a. How do different product categories contribute to overall profit growth or decline?
- b. How are sales and profit distributed geographically across global markets?
- c. How do multiple performance metrics compare across different regions?

Q4. What are the key drivers behind profit variations and profit loss, and how can AI techniques help identify root causes affecting profitability?

- a. What factors contribute most to profit variation across categories, sub-categories, and regions?
- b. Which factors most strongly influence profit increase or profit loss?
- c. How do sales and profit interact across products, shipping modes, and regions?

Q5 How can machine learning techniques be used to predict profit behaviour and support profit optimisation decisions?

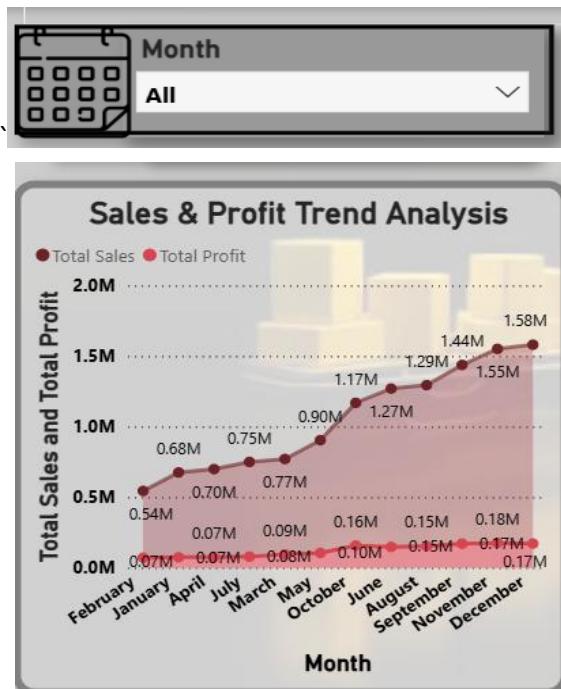
- a. How accurately can profit be predicted based on sales using a regression model?
- b. How does discount level impact profitability across orders?
- c. Are there abnormal profit patterns or outliers that require business attention?

### 3.FINDINGS BASED ON ANALYSIS AND EVALUATION

Q1. How is the business performing overall, and is growth stable or volatile over time and across sectors?

To assess whether an organization is growing steadily or experiencing volatile ups and downs, it is important to understand the overall business performance. This question focuses on analyzing long-term sales and profit trends to assess financial health, identify growth patterns and spot potential risks early. By examining both revenue generation and profitability over time, management can determine whether growth is driven by real value creation or by short-term factors such as aggressive discounting. The insights from this analysis support strategic planning, forecasting and investment decisions.

**a. How have total sales and profit evolved over time?**



Field	Description
Title	Sales and Profit Trend Analysis
Chart Type	Line Chart
Measures Used	Total Sales, Total Profit
Dimensions Used	Order Year
Reason for Selection	Line charts are ideal for identifying long-term trends and growth stability
Key Findings	Sales and profit demonstrate consistent upward growth over time
Filter	Month

Figure 1: Sales and Profit Trend Analysis

**Detailed description:**

This view presents the trend of total sales and total profit over several years, scaled by a moving average to smooth out short-term fluctuations. The results indicate a continued upward trend in both metrics, suggesting sustainable business growth rather than volatile performance. Correspondence between sales and profit trends means that income growth is not achieved at

the expense of profitability. Such stability reflects effective pricing strategies, cost management and consistent customer demand, placing the business well positioned for long-term expansion.

**b. Which sectors and product categories generate high sales but have a negative impact on profitability?**



Field	Description
Title	Sales vs Profit by Region and Product Category
Chart Type	Scatter Chart
Measures Used	Total Sales, Total Profit
Dimensions Used	Region, Category
Reason for Selection	Scatter charts reveal trade-offs and outliers between revenue and profitability
Key Findings	Some regions generate high sales but low or negative profit
Filter	Country
Tooltip	Total Sales by Category

Figure 2: Sales vs Profit by Region and Product Category

**Detailed description:**

This scatter plot evaluates the relationship between sales and profits in different regions and product categories. Each bubble represents a product or category, so areas of high performance and low performance can be identified. Interactive tooltips were added to provide additional context when hovering over each data point and reveal the product name, discount level, and profit contribution.

Tooltip details mentioned below:

**Tooltip Details:**

Field	Description
Title	Total Sales by Category
Chart Type	Bar Chart
X-Axis	Category
Y-Axis	Total Sales
Tooltip Field	Total Sales

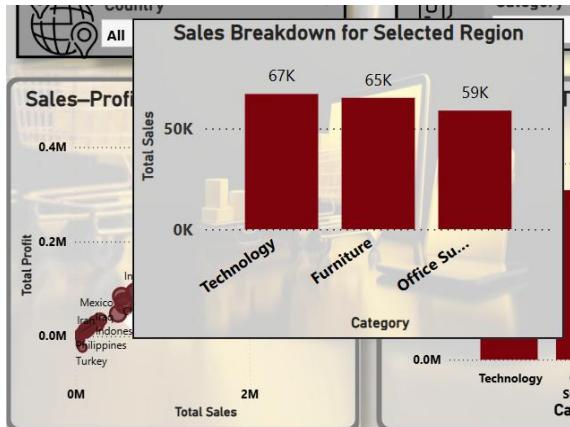


Figure 2.1 Sales break down for selected region(tooltip)

### C .How does the profit contribution vary across product subcategories?

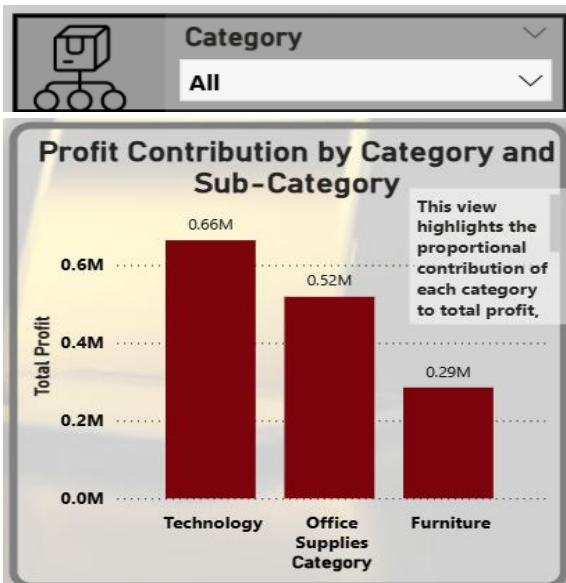


Figure 4: Profit contribution by category and sub category

#### Detailed description:

It analyzes the profit distribution across product categories and their respective subcategories using visual stacked bar charts. The stacked structure allows users to compare both overall category-level profits and the relative contribution of each subcategory within those categories. To increase analytical depth, a pie chart was embedded as a tooltip, showing subcategory-wise profit contribution when users hover over any bar segment. This interactive design enables users to examine the profit structure in more detail without adding visual clutter to the main chart. The analysis shows that although some categories generate high overall profits, they are also highly dependent on a small number of subcategories, indicating potential dependency risks. Other categories show more evenly distributed profit contributions, which suggests balanced product performance.

This combination of stacked bar and tooltip visualization supports both high-level comparisons and granular exploration, making it a powerful tool for strategic product and profitability analysis. Tooltip details given below:

### Tooltip Details:

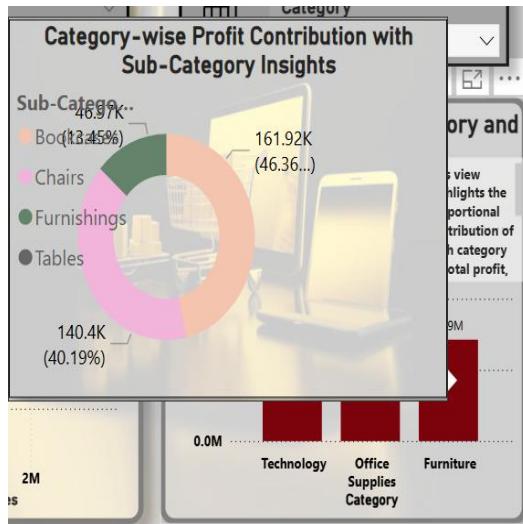


Figure 4.1: Category-wise Profit Contribution with sub-category insights

Field	Description
Title	Category-wise Profit Contribution with sub-category insights
Chart Type	Pie Chart
Legend	Sub-Category
Values	Total Profit
Tooltip Fields	Sub-Category, Total Profit
Reason	To understand how profit is distributed across different product sub-categories
Key Findings	Profit contribution is uneven across sub-categories, with a few driving the majority of profits

### d.What is the current overall business sales in terms different factors?

Filed	Description
Title	Top 10 Customers by Sales (CLV)
Chart Type	Horizontal Bar Chart
Y-Axis	Customer Name
X-Axis	Customer Lifetime Value (Total Sales)
Measures Used	Total Sales

A.

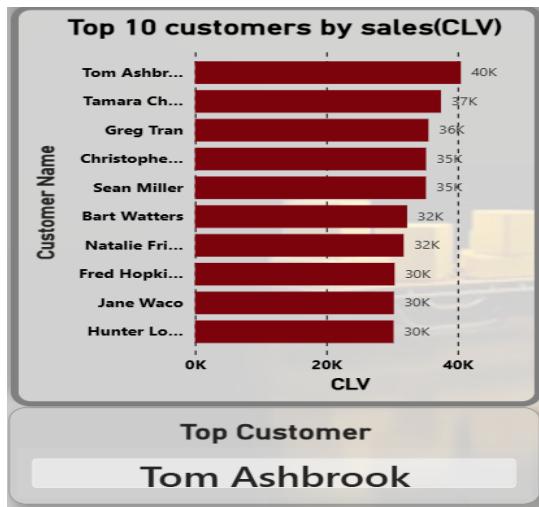


Figure 5: Top 10 Customers by Sales (CLV)

<b>Reason for Chart</b>	To identify high-value customers contributing most to revenue
<b>Findings</b>	A small group of customers contributes a significant share of total sales, highlighting customer concentration risk

### Detailed description:

This view presents the top 10 customers ranked by customer lifetime value (CLV), measured using total sales contribution. The diagram highlights how the revenue is distributed between the most valuable customers and helps to identify customer concentration in the business. By focusing on high-value customers, the organization can assess dependency risk and opportunities for targeted retention strategies. The results show that a small group of customers contributes a disproportionately large share of the total sales, which suggests that maintaining relationships with these customers is important for maintaining the overall income result.

Item	Description
<b>Title</b>	Profit and Average Discount Monthly Trend
<b>Chart Type</b>	Line Chart
<b>X-Axis</b>	Order Month
<b>Y-Axis</b>	Total Profit, Average discount
<b>Measures Used</b>	Total Profit, Average discount
<b>Reason for Chart</b>	To examine the relationship between discounting strategies and profitability
<b>Findings</b>	Increased discount levels are associated with

B.

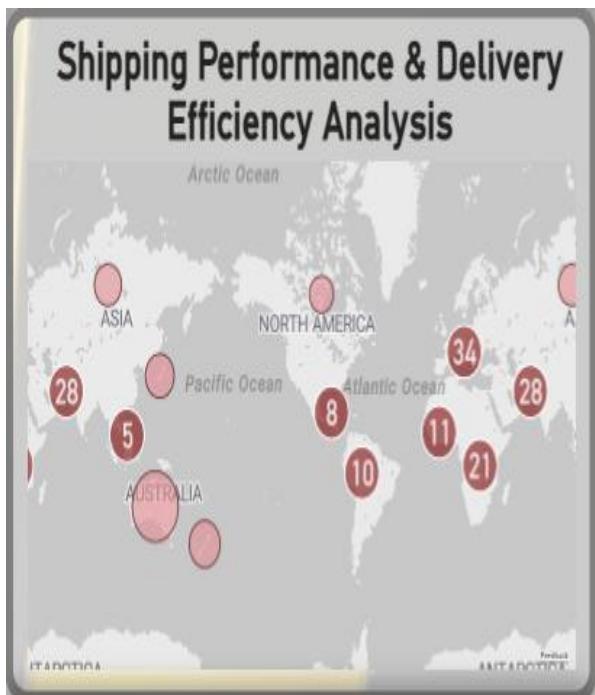


declining profits,  
indicating margin erosion

Figure 6: Profit and Average Discount Monthly Trend

### Detailed description:

This view shows the monthly trend of total profit along with average discount levels to analyze the relationship between discount strategies and profitability over time. Line charts make it easy to compare between high discount periods and related bottom line performance. The analysis reveals an inverse relationship between discount levels and profits, where higher discounts correspond to declining profits. This suggests that aggressive discounting has a negative impact on profit margins and highlights the need for a more balanced pricing and marketing strategy to maintain profitability.



Item	Description
Title	Shipping Performance & Delivery Efficiency Analysis
Chart Type	Map Visual
Location	Country / Region
Measures Used	Order Count / Delivery Volume
Reason for Chart	To analyse geographical distribution of orders and shipping activity
Findings	High order concentration is observed in specific regions, impacting logistics planning and delivery efficiency

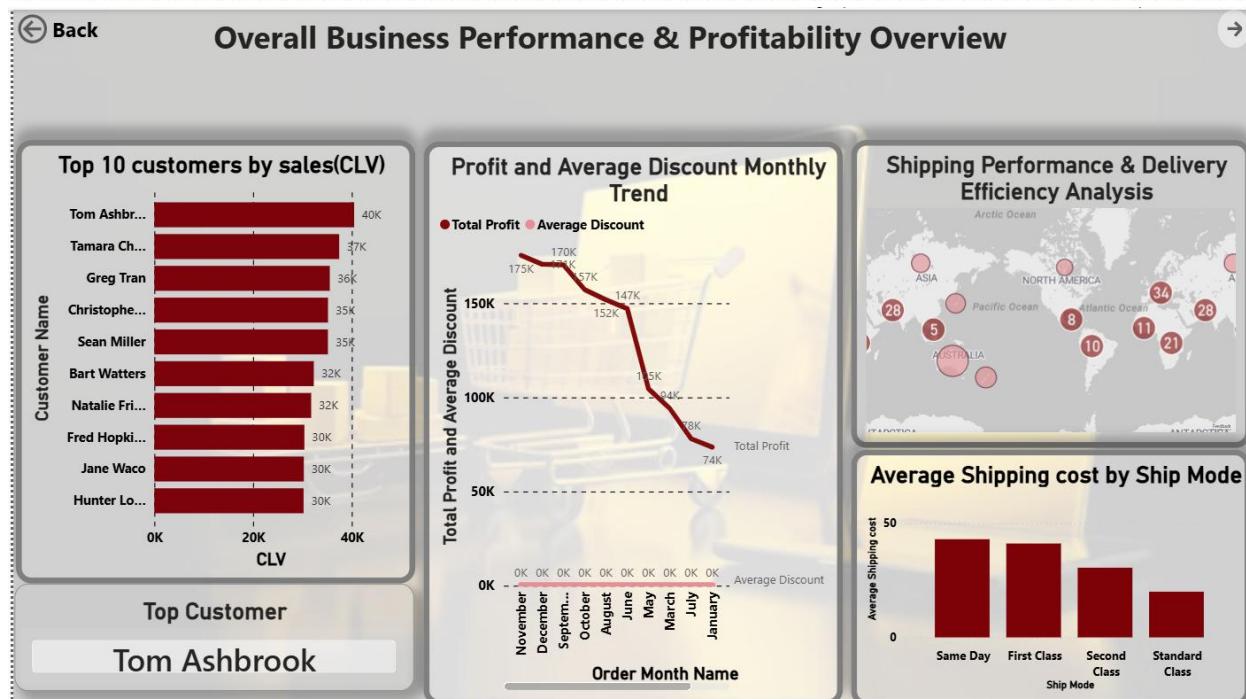
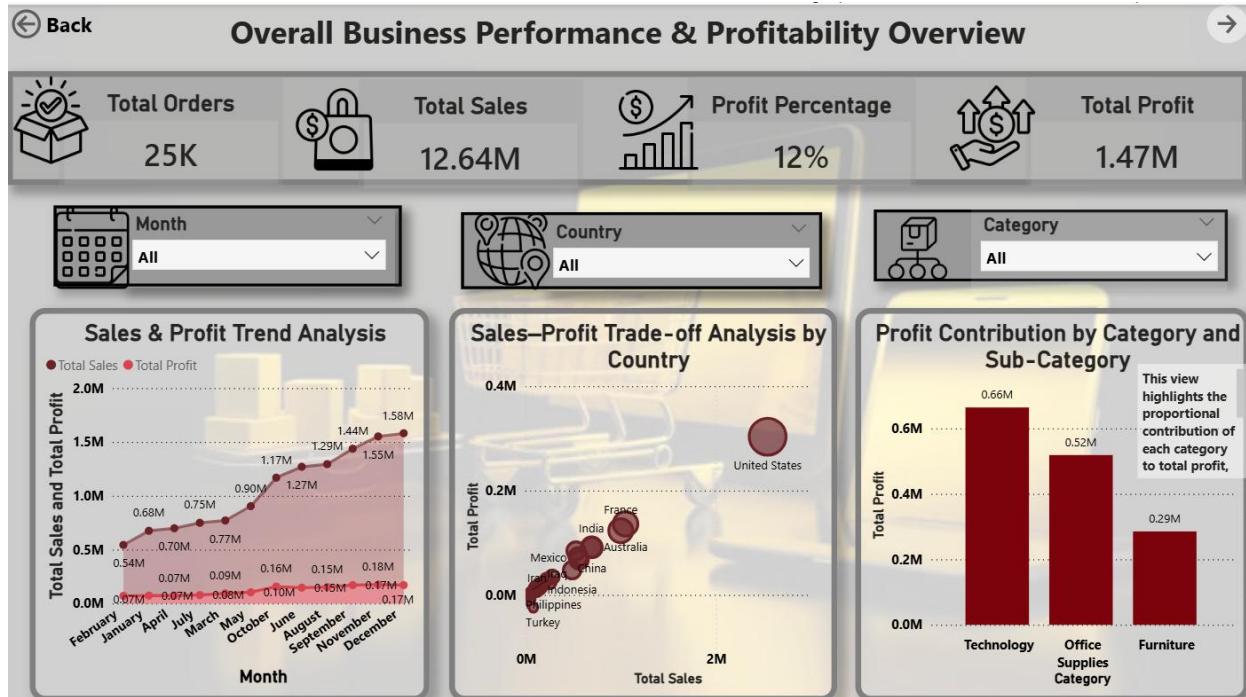
Figure 7: Shipping Performance & Delivery Efficiency Analysis

### Detailed description:

This map view shows the geographic distribution of order volume across different regions, providing insight into delivery performance and delivery demand around the world. By visualizing order concentration geographically, businesses can identify potential pressure points in key markets and delivery functions that drive logistics activity. The results show high order density in specific regions, indicating that shipping resources and operational planning must be prioritized to maintain delivery efficiency and customer satisfaction.

### Overall Summary – Page 1: Business Performance & Profitability Overview :

Page 1 provides a comprehensive overview of the organisation's overall business performance by analyzing sales growth, profitability, customer contribution and operational efficiency. The analysis shows that while sales are driven by a small group of high-value customers, profitability is highly sensitive to periodic discounting strategies. Regional and shipping analyzes show further differences in distribution efficiency and cost structures, which affect overall profit margins. In addition, variations in shipping costs and order concentration in different regions indicate opportunities for operational optimization. Overall, the insights on this page establish a strong foundation for understanding the key revenue drivers and cost-related challenges that impact business performance.



Q2. Which factors control profitability, and where are the major profit risks in the business?

This question focuses on understanding how discounting strategies affect bottom line performance across products, regions, and time. By analyzing profit percentage, discount behavior and profit distribution, a company can identify whether discounts are driving sustainable growth or reducing margins. Advanced analysis and visualization techniques are used to highlight risk areas and optimization opportunities.

#### **Sub questions:**

a.Which categories and subcategories generate the highest and lowest profit margins?

Sales and Profit Margin Comparison by Category and Sub-Category			
Category	Total Sales	Total Profit	Percentage of Profit
Technology	47,44,557.50	6,63,778.73	14%
Accessories	7,49,237.02	1,29,626.31	17%
Copiers	15,09,436.27	2,58,567.55	17%
Phones	17,06,824.14	2,16,717.01	13%
Machines	7,79,060.07	58,867.87	8%
Office Supplies	37,87,070.23	5,18,473.83	14%
Paper	2,44,291.72	59,207.68	24%
Labels	73,404.03	15,010.51	20%
Envelopes	1,70,904.30	29,601.12	17%
Binders	4,61,911.51	72,449.85	16%
Art	3,72,091.97	57,953.91	16%
Appliances	10,11,064.30	1,41,680.59	14%
Fasteners	83,242.32	11,525.42	14%
Storage	11,27,085.86	1,08,461.49	10%
Supplies	2,43,074.22	22,583.26	9%
Furniture	41,10,874.19	2,85,204.72	7%
Furnishings	3,85,578.26	46,967.43	12%
Bookcases	14,66,572.24	1,61,924.42	11%
Chairs	15,01,681.76	1,40,396.27	9%
Tables	7,57,041.92	-64,083.39	-8%
Total	1,26,42,501.91	14,67,457.29	12%

Field	Description
Title	Sales and Profit Margin Comparison by Category and Sub-Category
Chart Type	Table
Measures Used	Total Sales, Total Profit, Profit Percentage
Dimensions	Category, Sub-Category
Formatting Used	Conditional formatting on Profit & Profit %
Reason for Selection	To identify sub-categories with strong revenue but weak or negative profitability
Key Findings	While Technology and Office Supplies show strong profit margins, Furniture—particularly Tables—records negative profit despite high sales. This indicates pricing, discounting, or cost issues within specific sub-categories.

Figure 8: Sales and Profit Margin Comparison by Category and Sub-Category

### Detailed explanation:

This table provides a detailed overview of the profitability of drilling from category to subcategory level. Conditional formatting highlights areas of concern, enabling quick identification of loss-making products. The analysis shows that high sales volume alone does not guarantee profitability, and emphasizes the importance of margin analysis when evaluating product performance.

b.Which months had the lowest profits and what patterns explain this decline?

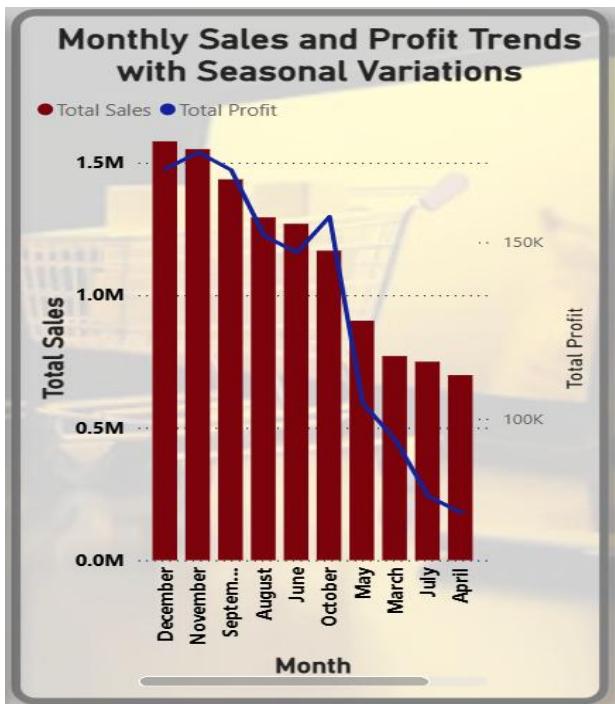


Figure 8: Monthly Sales and Profit Trends with Seasonal Variations

Field	Description
Title	Monthly Sales and Profit Trends with Seasonal Variations
Chart Type	Clustered Column & Line Chart (Combo Chart)
Measures Used	Total Sales, Total Profit
Dimensions	Month
Reason for Selection	To compare revenue and profit trends across time and detect seasonal fluctuations
Key Findings	Sales peak during the year-end months, while profits fluctuate more sharply, indicating periods where increased sales are accompanied by reduced margins due to discounts or higher operational costs.

### Detailed explanation:

This visualization highlights the difference between sales growth and profit development over time. While sales remain relatively strong for some months, profits do not always follow the same pattern. This suggests that the effects of seasonal discounts, promotions or cost variations affect profitability independently of revenue.

c.What is the impact of different shipping modes on order volume and profitability?



Field	Description
Title	Profitability Impact of Shipping Modes
Chart Type	Horizontal Bar Chart
Measures Used	Total Profit, Order Count
Dimensions	Ship Mode
Reason for Selection	To evaluate operational efficiency and profitability across shipping methods
Key Findings	Standard Class generates the highest profit and order volume, while Same Day shipping contributes the least to overall profitability, despite higher costs.

Figure 9: Profitability Impact of Shipping Modes

### Detailed explanation:

This chart assesses logistics performance by comparing profit contribution across shipping modes. Analysis shows that faster shipping options do not necessarily improve profitability and may increase operating costs. So optimizing shipping strategies can significantly increase profit margins.

c. Can profit be predicted based on sales and discount patterns using **machine learning(with R)?**

**R Script code:**

```
# dataset <- data.frame(Total Profit, Discount, Total Orders)
# dataset <- unique(dataset)
model <- lm(`Total Profit` ~ `Total Sales` + Discount, data = dataset)
plot(dataset$`Total Sales`, dataset$`Total Profit`,
     main="Profit Prediction based on Sales & Discount",
     xlab="Total Sales",
     ylab="Total Profit",
     pch = 25,
     col = "blue"
)
abline(model, col="#7C0208", lwd=4)
```

Field	Description
Title	Profit Prediction Using Linear Regression (R)
Chart Type	R Script Visual – Scatter Plot with Regression Line
Machine Learning Technique	Linear Regression
Input Variables (Independent)	Total Sales, Discount
Target Variable (Dependent)	Total Profit
Reason for Selection	To understand whether profit can be statistically predicted based on sales volume and discount behaviour
Key Findings	The regression line shows a strong positive relationship between sales and profit, indicating that sales is a strong predictor of profit, while discounts introduce variability in profit outcomes.

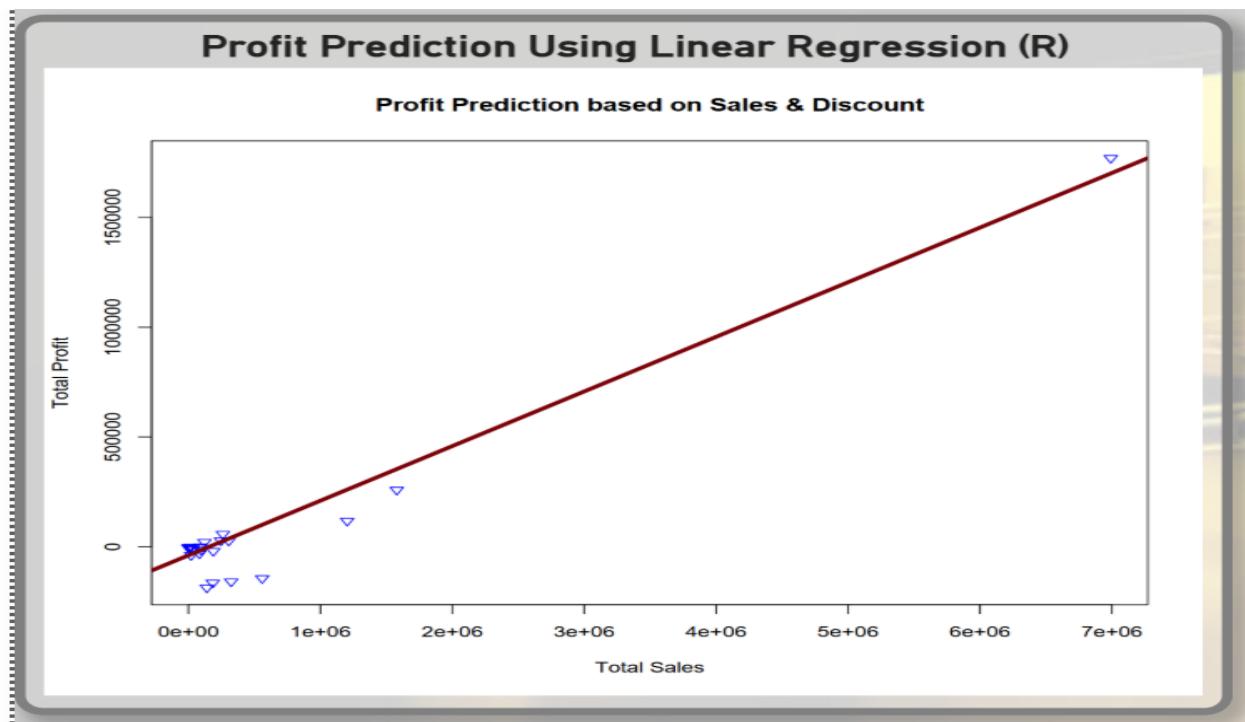


Figure 10: Profit Prediction Using Linear Regression (R)

### Detailed Explanation (ML Visual):

This visual applies a linear regression model using R to predict total profit based on sales and discount values. Each point represents aggregated business data, while the regression line represents the predicted profit trend. The positive slope confirms that higher sales generally lead to higher profits; however, dispersion around the line suggests that discounts and other operational factors influence final profitability. This model demonstrates how machine learning techniques can be integrated into BI tools to enhance predictive decision-making.

### Supporting Table :

Field	Description
Title	Sales & Profit Breakdown Supporting ML Model
Chart Type	Table
Measures Used	Total Sales, Total Profit, Sum of Discount, Profit Percentage
Dimensions	Category
Formatting Used	Conditional formatting on Sales, Profit, and Profit %
Reason for Selection	To support and validate the regression model by showing category-level contribution to sales and profit
Key Findings	Technology contributes the highest profit with relatively controlled discounting, while Furniture shows lower profit margins despite high sales, reinforcing the regression insights.

Sales & Profit Breakdown Supporting ML Model				
Category	Total Sales	Sum of Discount	Total Profit	Percentage of Profit
Technology	47,44,557.50	1,372.51	6,63,778.73	14%
Office Supplies	37,87,070.23	4,297.19	5,18,473.83	14%
Furniture	41,10,874.19	1,660.03	2,85,204.72	7%
Total	1,26,42,501.91	7,329.73	14,67,457.29	12%

The supporting table complements the regression model by showing how individual product categories contribute

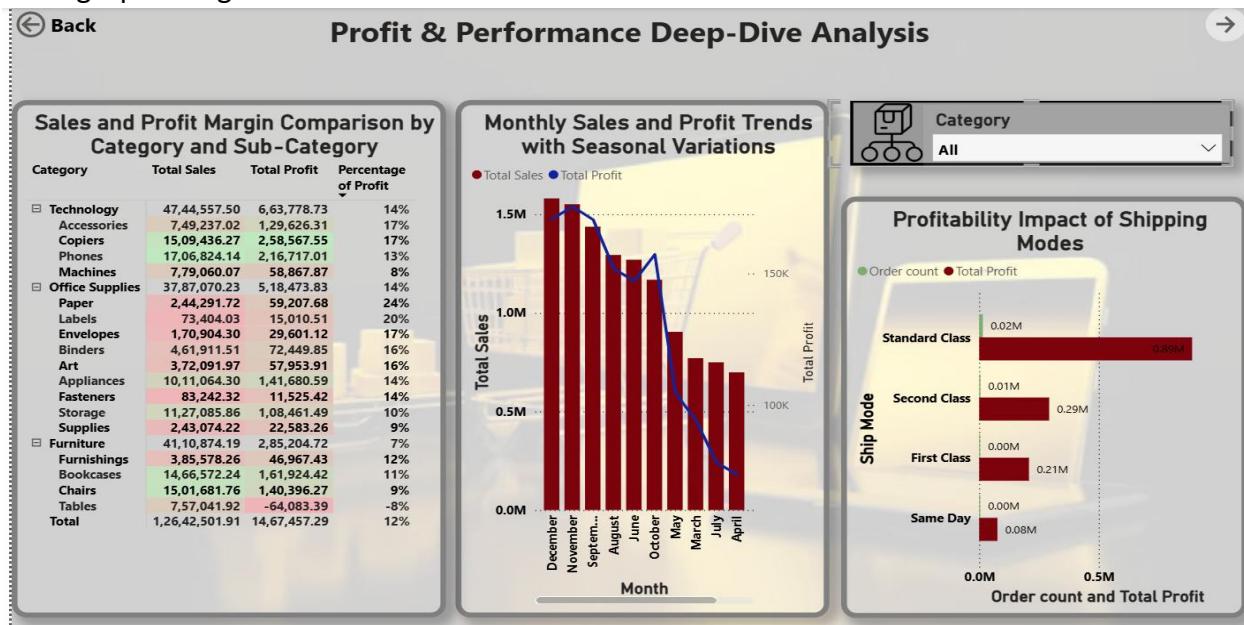
Figure 10.1 :Sales & Profit Breakdown Supporting ML Model

## Detailed Explanation (Supporting Table):

This table complements the regression model by providing category-level context. It highlights how different product categories contribute to overall sales and profit while applying varying levels of discount. The table helps interpret why some data points in the regression plot deviate from the prediction line, especially for categories where discounts significantly reduce profitability.

## Overall Summary – Page 2: Profit & Performance Deep-Dive Analysis:

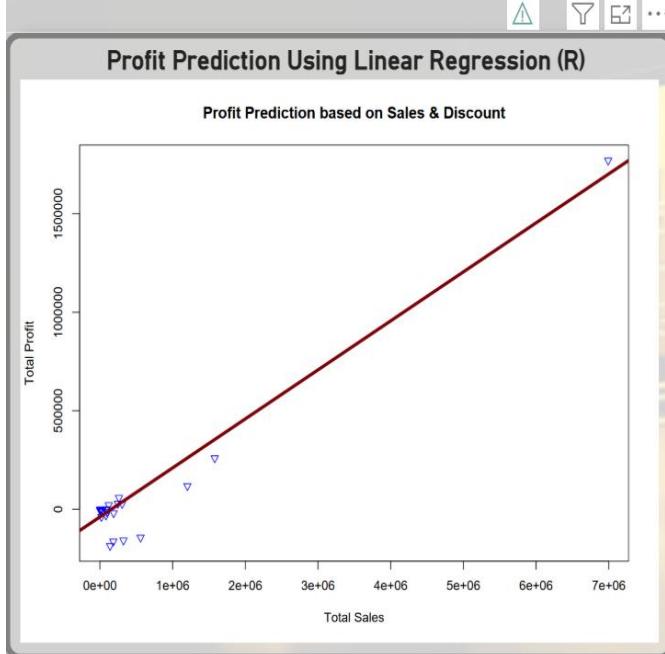
Page 2 provides comprehensive profitability by combining descriptive, diagnostic and predictive analyses. Analysis at product level shows that higher sales do not always result in stronger margins. Seasonal trends highlight periods where profits diverge from revenue growth. Shipping mode valuation highlights operational inefficiencies that affect profitability. Finally, a machine learning regression model shows that sales are a strong predictor of profits, while discounts add volatility. Together, these insights support data-driven decisions about pricing, logistics and strategic planning.



Back

## Profit & Performance Deep-Dive Analysis

→



### Sales & Profit Breakdown Supporting ML Model

Category	Total Sales	Sum of Discount	Total Profit	Percentage of Profit
Technology	47,44,557.50	1,372.51	6,63,778.73	14%
Office Supplies	37,87,070.23	4,297.19	5,18,473.83	14%
Furniture	41,10,874.19	1,660.03	2,85,204.72	7%
Total	1,26,42,501.91	7,329.73	14,67,457.29	12%

The supporting table complements the regression model by showing how individual product categories contribute

### R script editor

⚠ Duplicate rows will be removed from the data. X

```
6 | 
7 model <- lm(`Total Profit` ~ `Total Sales` + Discount, data = dataset)
8 plot(dataset$`Total Sales`, dataset$`Total Profit`,
9       main="Profit Prediction based on Sales & Discount",
10      xlab="Total Sales",
11      ylab="Total Profit",
12      pch = 25,
13      col = "blue"
14      )
15 abline(model, col="#7C0208", lwd=4)
16
```

Q3. How do geography and product categories influence profit contribution and overall business performance?

#### Sub question

a. How do different product categories contribute to overall profit growth or decline?

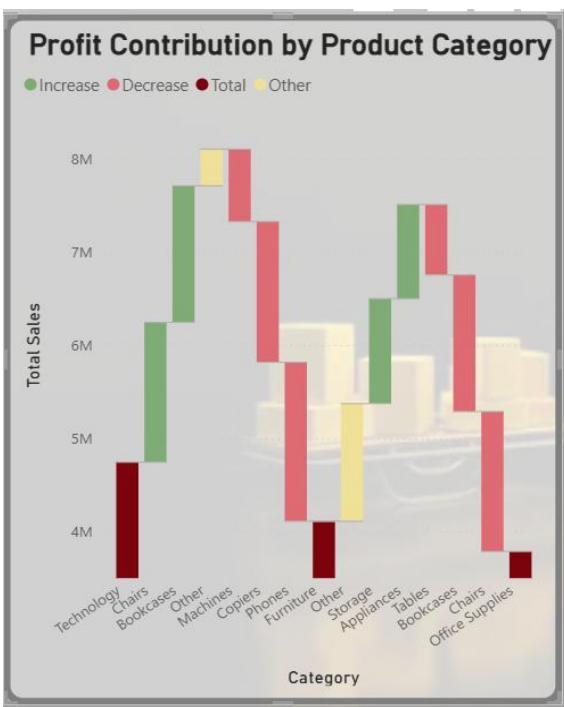


Figure 11: Profit Contribution by Product Category

Field	Description
Title	Profit Contribution by Product Category
Chart Type	Waterfall Chart
X-Axis	Product Category / Sub-Category
Y-Axis	Total Sales (Profit Contribution Effect)
Legend	Increase, Decrease, Total
Reason for Selection	To clearly visualise how individual product categories incrementally increase or decrease overall profit
Key Findings	Categories such as Technology and Chairs contribute positively, while Furniture-related sub-categories show notable profit decline

#### Detailed explanation:

Waterfall charts show the cumulative effect of individual product categories on overall profitability. Positive bars indicate categories that add value, while negative bars highlight areas that cause reduced profits. This visualization makes it easy to find poorly performing categories that may require pricing, cost optimization or discount control strategies.

b.How are sales and profit distributed geographically across global markets?



Figure 12: Geographical Distribution of Sales and Profit

Field	Description
Title	Geographical Distribution of Sales and Profit
Chart Type	Map (Bubble Map)
Location Field	Country / Region
Bubble Size	Total Sales
Bubble Colour	Total Profit
Tooltips Used	Country, Total Sales, Total Profit
Reason for Selection	To identify high-performing and underperforming regions spatially
Key Findings	North America and parts of Europe generate higher sales and profit, while some regions show high sales with relatively lower profit

### Detailed explanation:

Map view highlights regional concentration of sales and profitability in global markets. Larger bubbles represent higher sales volume, while color intensity indicates profit levels. This enables rapid identification of areas where strong sales do not necessarily translate into strong profitability, indicating potential costs or logistics inefficiencies.

c.How do multiple performance metrics compare across different regions?

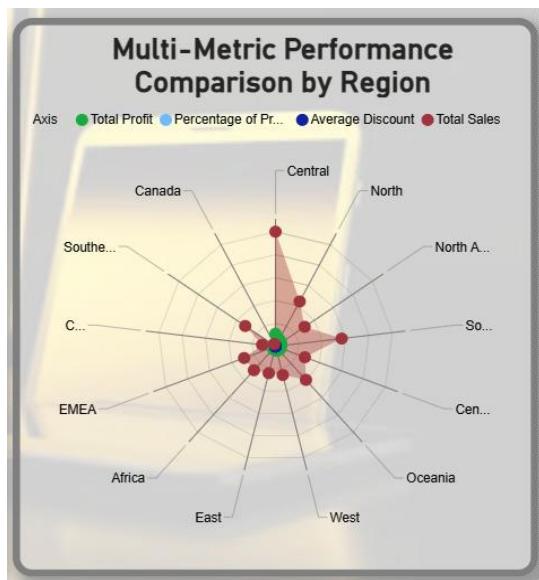


Figure 12: Multi-Metric Performance Comparison by Region

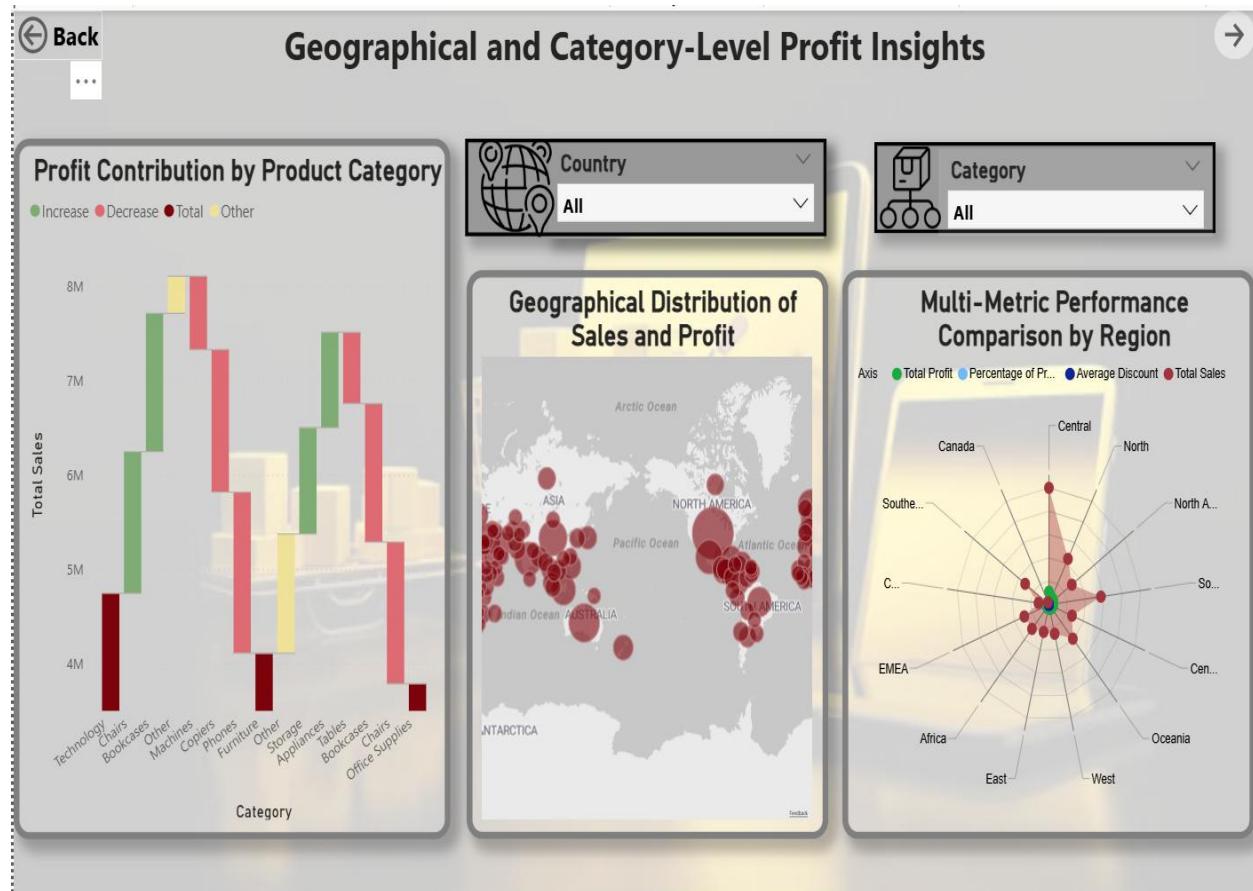
Field	Description
Title	Multi-Metric Performance Comparison by Region
Chart Type	Radar Chart
Dimensions	Region
Metrics Used	Total Sales, Total Profit, Profit Percentage, Average Discount
Legend	Performance Metrics
Reason for Selection	To compare multiple KPIs simultaneously across regions
Key Findings	Central and North regions show balanced performance, while some regions exhibit high discounts with lower profit efficiency

## Detailed explanation:

Radar charts allow simultaneous comparison of multiple performance indicators for each sector. This reveals an imbalance between sales, profit and discount strategies. Areas of extended discounts but weak profit margins become apparent, supporting data-driven regional strategy optimization.

## Overall Summary – Page 3: Business Performance & Profitability Overview :

Page 3 provides a spatial and category-focused insight into the company's performance. Waterfall analysis reveals which product categories increase or decrease profitability. The geographical map highlights regional differences between sales and profits. Finally, the radar chart compares several KPIs across regions, highlighting inefficiencies related to rebates and regional cost structures. Together, these insights support targeted regional strategies and category-level optimization.



Q4. What are the key drivers behind profit variations and profit loss, and how can AI techniques help identify root causes affecting profitability?

This site focuses on advanced, AI-powered analytical techniques to uncover the underlying factors that influence earnings performance. Using Power BI's built-in AI visualizations like decomposition trees, key influencers, scatter analysis, and waterfall charts, this section explains why profits increase or decrease across categories, subcategories, regions, discounts, and shipping modes, enabling deeper root cause analysis beyond descriptive trends.

#### Sub-Questions

a. What factors contribute most to profit variation across categories, sub-categories, and regions?

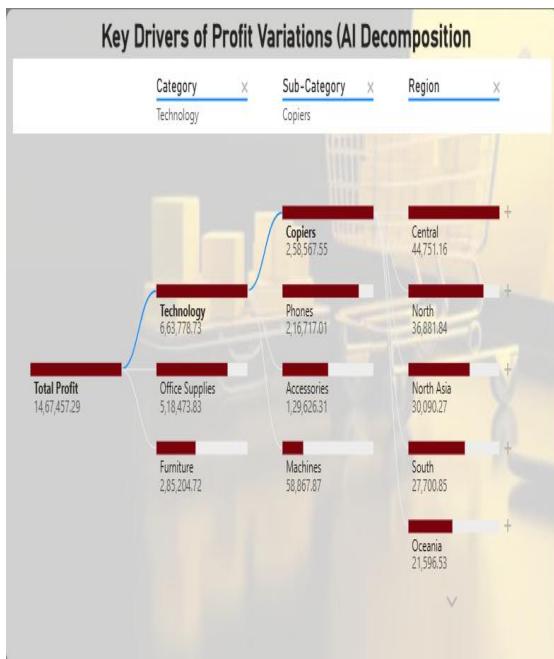


Figure 13: Key Drivers of Profit Variations (AI Decomposition)

Field	Description
Title	Key Drivers of Profit Variations (AI Decomposition)
Chart Type	Decomposition Tree
Measure Analyzed	Total Profit
Dimensions Used	Category → Sub-Category → Region
Reason for Chart	To break down total profit hierarchically and identify which dimensions contribute most to profit changes
Key Findings	Technology contributes the highest profit, with Copiers and Phones driving major gains, particularly in Central and North regions

#### Detailed explanation:

The decomposition tree starts with the total profit and systematically goes down to category, subcategory and sector to identify key contributors. Technology has emerged as the most profitable category, driven mainly by copiers and telephones. Regionally, the Middle and North regions show strong profit contributions, which indicates that both product mix and geography significantly affect profitability.

b: Which factors most strongly influence profit increase or profit loss?

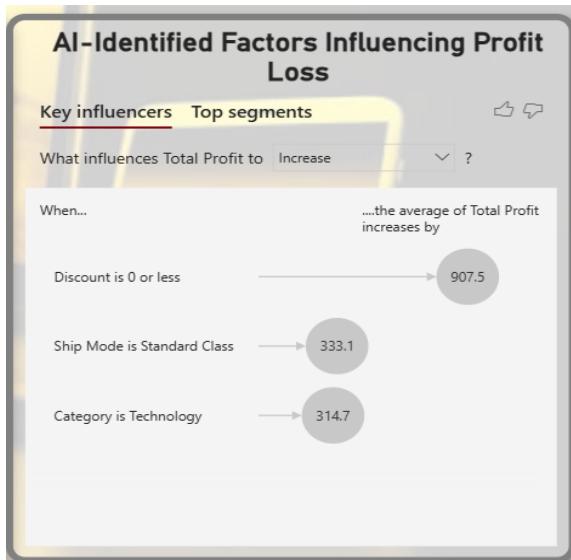


Figure 14: AI-Identified Factors Influencing Profit Loss

Field	Description
Title	AI-Identified Factors Influencing Profit Loss
Chart Type	Key Influencers (AI Visual)
Target Variable	Total Profit
Analysis Type	Increase in Total Profit
Reason for Chart	To automatically detect conditions that most influence profit increase or decrease
Key Findings	Lower discounts, Standard Class shipping, and Technology category significantly increase profitability

### Detailed explanation:

Key Influencer Visual uses machine learning algorithms to evaluate which conditions have the greatest impact on profits. The analysis shows that when discounts are zero or minimal, profits increase significantly. In addition, standard line freight and technology line products consistently generated higher profits, highlighting pricing discipline and operational efficiency as key profit drivers.

c. How do sales and profit interact across products, shipping modes, and regions?

Field	Description
Title	Sales vs Profit Dynamics by Product and Shipping Mode Across Regions
Chart Type	Scatter Chart with Play Axis
X-Axis	Total Sales
Y-Axis	Total Profit
Legend	Category
Play Axis	Region
Reason for Chart	To analyze the relationship between sales and profit dynamically across regions
Key Findings	High sales do not always translate to high profit, especially in Furniture and high-discount regions

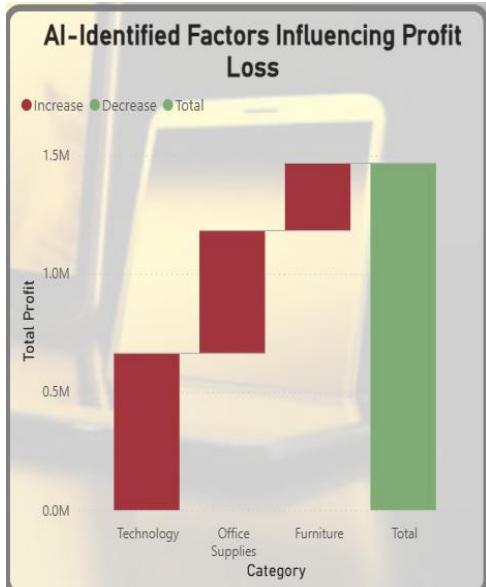


Figure 15: Sales vs Profit Dynamics by Product and Shipping Mode Across Regions

### Detailed explanation:

This scatter chart visualizes the relationship between sales and profit at a detailed level and enables regional comparisons using play axes. This highlights that technology products generally maintain positive sales-profit ratios, while furniture and some office supplies show low or negative profits despite moderate sales, indicating cost inefficiencies or discounts in some areas.

d. How does each product category contribute to overall profit and loss?



Field	Description
Title	Profit Contribution and Loss by Category
Chart Type	Waterfall Chart
Measure Analyzed	Total Profit
Category Breakdown	Technology, Office Supplies, Furniture
Reason for Chart	To visualize incremental profit contributions and losses by category
Key Findings	Technology adds the largest positive contribution, while Furniture shows weaker or negative impact

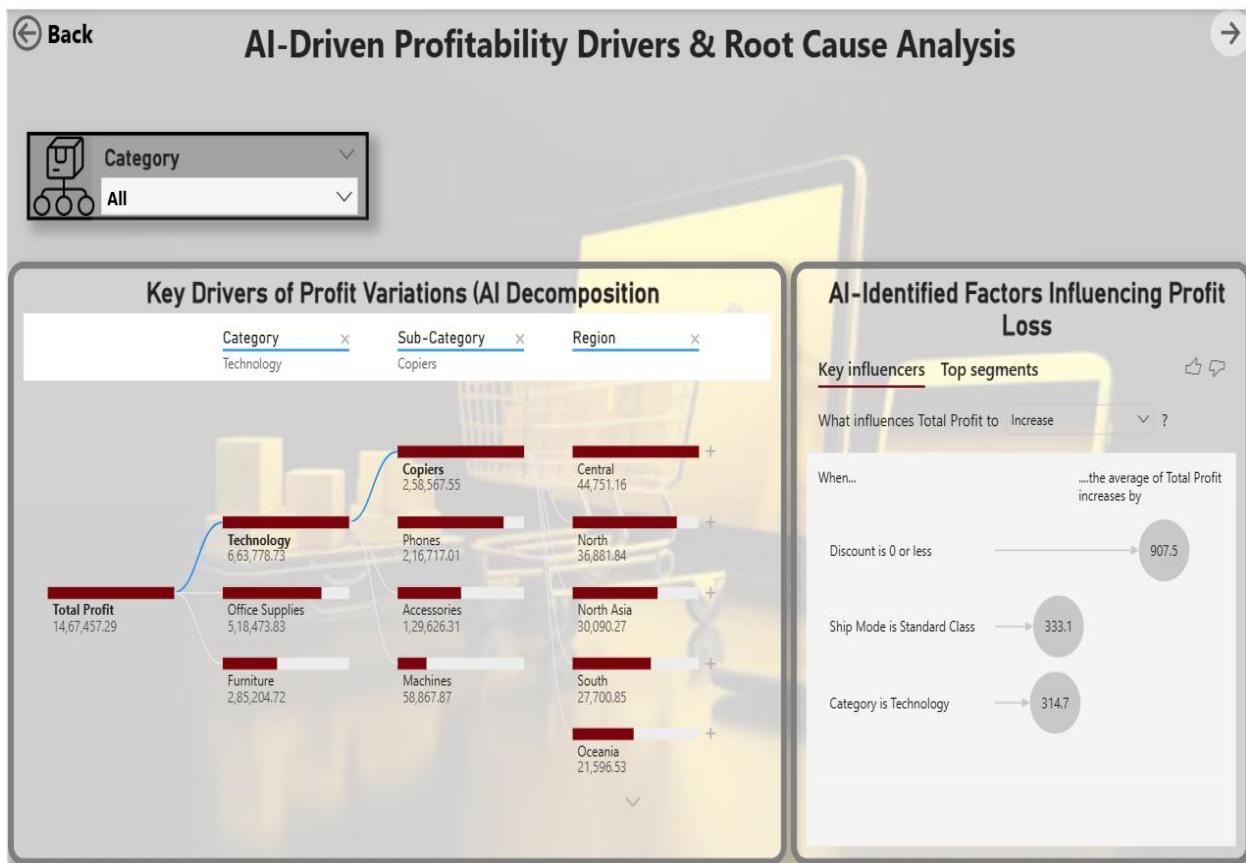
Figure 16: Profit Contribution and Loss by Category

## Detailed explanation:

A waterfall chart shows how each category incrementally contributes to the total profit. Technology contributes the largest positive growth followed by office supplies, while furniture shows a relatively small or even negative impact on profitability. This visualization clearly identifies categories that strengthen overall performance versus those that require strategic intervention.

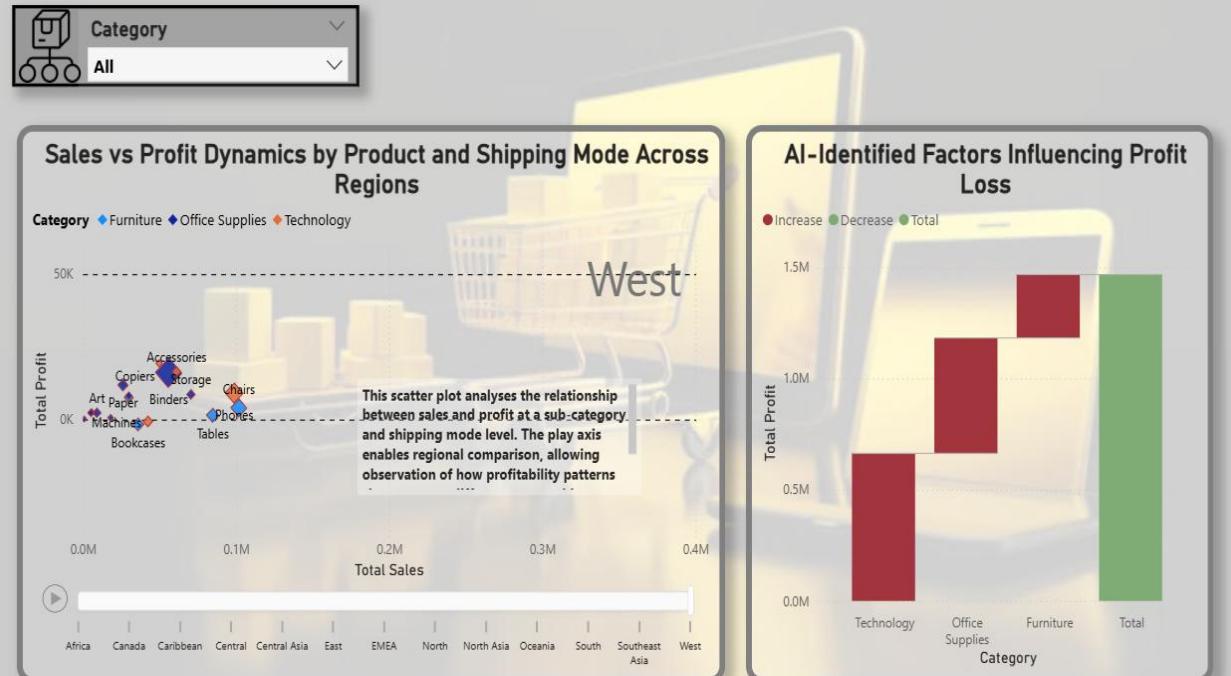
## Overall Summary – Page 4: AI-Driven Profitability Drivers & Root Cause Analysis:

Page 4 uses AI-powered graphics to highlight the reasons behind profit variations. The decomposition tree highlights technology and copiers as key profit drivers, while key influencers show that lower discounts and standard class shipping increase profits significantly. Scatter analysis shows that high sales do not always guarantee high profits, especially in furniture. Waterfall charts consolidate this insight by showing performance contribution at the category level. Overall, this site provides useful information to improve pricing, product mix and operational strategies.



[Back](#)

## AI-Driven Profitability Drivers & Root Cause Analysis

[→](#)

**Q5 How can machine learning techniques be used to predict profit behaviour and support profit optimisation decisions?**

This page focuses on applying machine learning concepts using R to predict profit patterns and identify areas of risk that impact profitability. Regression-based models are used to understand the relationships between sales, discounts and profits, while anomaly detection logic highlights unusual transactions. This insight supports data-driven decision making for pricing, discount strategies and risk control.

### Sub Questions

a. How accurately can profit be predicted based on sales using a regression model?

Field	Description
Title	Profit Prediction Using Machine Learning (Regression Model)
Chart Type	Scatter Plot with Linear Regression Line (R Script Visual)
X-Axis	Sales
Y-Axis	Profit
Technique Used	Linear Regression
Reason for Selection	To model and predict profit behaviour based on sales values
ML Component	Supervised Learning – Regression

Findings	A positive relationship exists between sales and profit, indicating that higher sales generally result in higher profits
----------	--

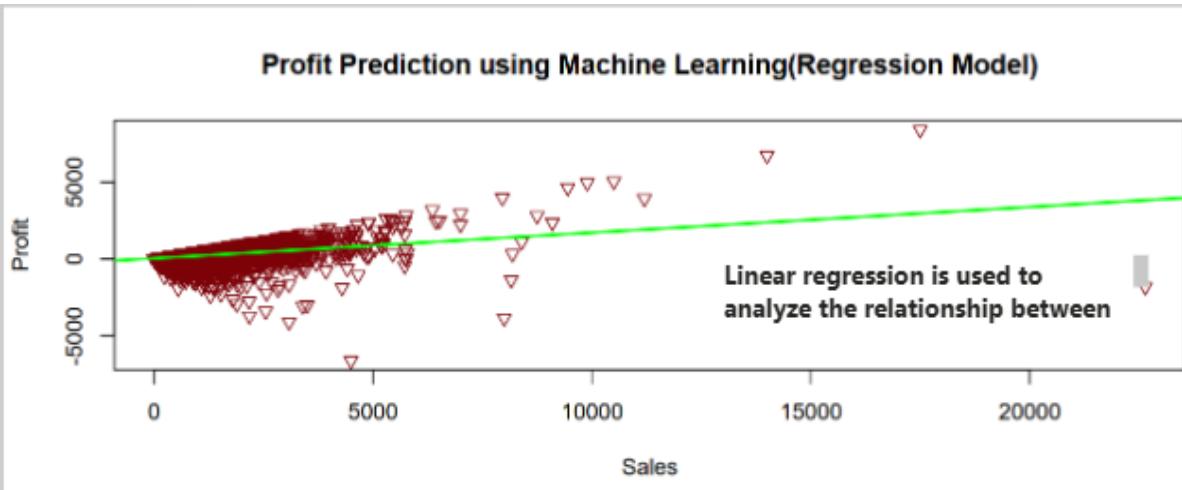


Figure 17: Profit Prediction Using Machine Learning (Regression Model)

**Supporting Table:**

Category	Total Profit	Total Sales	% of Profit
Furniture	2,85,082.73	41,10,451.90	7%
Office Supplies	5,18,595.83	37,87,492.51	14%
Technology	6,63,778.73	47,44,557.50	14%
Total	14,67,457.29	1,26,42,501.91	12%

This is added because

1. validates regression results
2. Shows contribution to profit at the category level
3. ML connects predictions to actual business results

**Detailed explanation (section):**

This view uses a linear regression model that uses R to analyze the relationship between sales and profit. Each data point represents an order, while the regression line indicates the estimated profit trend. The upward slope of the regression line confirms that sales are a strong predictor of profit. This model helps companies calculate expected profit levels for future sales scenarios and supports revenue forecasting decisions.

b. How does discount level impact profitability across orders?

Field	Description
Title	Impact of Discount on Profit
Chart Type	Scatter Plot with Regression Line (R Visual)
X-Axis	Discount
Y-Axis	Profit
Technique Used	Linear Regression
Reason for Selection	To analyse how increasing discount levels affect profit
ML Component	Regression Analysis
Findings	Higher discount levels are associated with declining profit and frequent losses

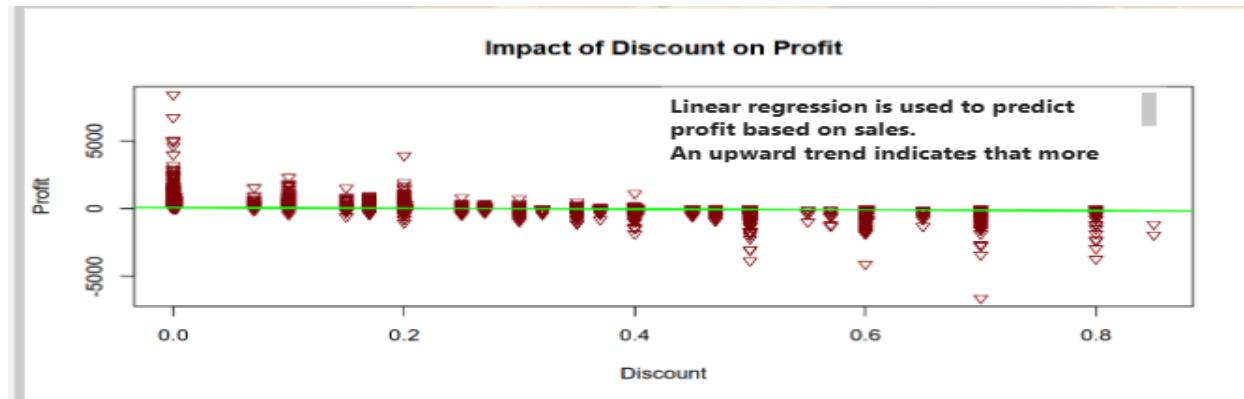


Figure 18: Impact of Discount on Profit

Supporting Table:

Discount type	Avg Profit	Avg Sales	Count of Order
No Discount	61.04	241.04	29009
Low Discount	46.69	339.62	10953
Medium Discount	-56.19	217.80	7156
High Discount	-98.89	89.08	4172
Total	28.61	246.49	51290

This is added because

- a. clearly quantifies the impact of the discount
- b. Numerically confirms the regression trend
- c. Strengthens business rationale

### **Detailed explanation:**

This visualization shows how discount levels affect profitability using regression logic. As the discount value increases, profits begin to decline, with many data points falling into negative profit territory. The regression trend highlights the risk of aggressive discounting, and shows that excessive discounting reduces margins. This insight helps decision makers optimize discount strategies while maintaining profitability.

c.Are there abnormal profit patterns or outliers that require business attention?

Field	Description
Title	Detection of Profit Outliers Using Machine Learning Logic
Chart Type	Scatter Plot (Outlier Detection)
X-Axis	Sales
Y-Axis	Profit
Technique Used	Outlier Identification using Regression Residuals
Reason for Selection	To identify unusual profit or loss patterns
ML Component	Anomaly Detection Logic
Findings	Extreme profit and loss values exist that deviate from normal business behaviour

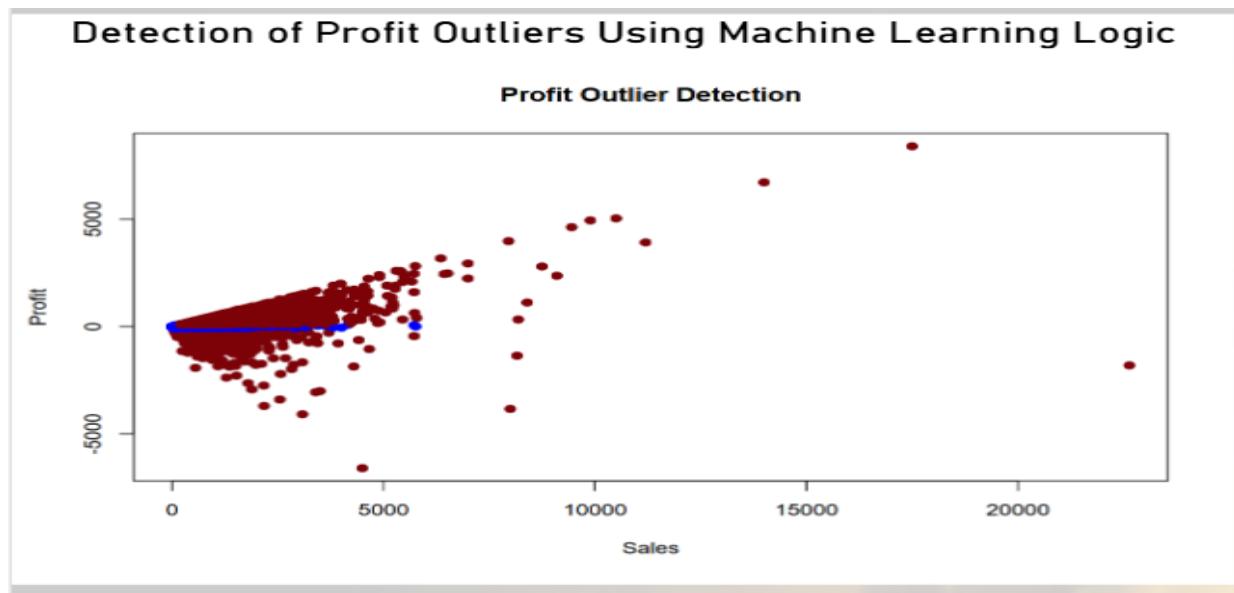


Figure 19 :Detection of Profit Outliers Using Machine Learning Logic

### **Supporting Table:**

<b>Category</b>	<b>Total Profit</b>	<b>Total Sales</b>	<b>% of Profit</b>
Furniture	2,85,082.73	41,10,451.90	7%
Office Supplies	5,18,595.83	37,87,492.51	14%
Technology	6,63,778.73	47,44,557.50	14%
<b>Total</b>	<b>14,67,457.29</b>	<b>1,26,42,501.91</b>	<b>12%</b>

This is added because

- a. Helps contextualize outliers
- b. Shows whether irregularities are range driven
- c. supports management control

#### **Detailed explanation :**

This view identifies profit drivers by analyzing deviations from expected profit behavior. Data points that are off the regression trend indicate unusual transactions, such as unusually high profits or extreme losses. These outliers may indicate pricing errors, data quality issues or operational inefficiencies and should be investigated further to minimize risk.

#### **Overall Summary – Page 5: Predictive Analytics for Profit Optimisation Using Machine Learning**

Page 5 shows the use of machine learning techniques to support profit optimization decisions. Regression models reveal a strong positive relationship between sales and profits, while discount analysis highlights the negative effect of excessive discounts. Support tables validate model results across product categories. External investigation uncovers unusual profit patterns that require further investigation. Overall, this site empowers predictive decision making using data-driven AI techniques.

[Back](#) **Predictive Analytics for Profit Optimisation Using Machine Learning** [→](#)

Profit Prediction using Machine Learning(Regression Model)

Linear regression is used to analyze the relationship between Profit and Sales.

Category	Total Profit	Total Sales	% of Profit
Furniture	2,85,082.73	41,10,451.90	7%
Office Supplies	5,18,595.83	37,87,492.51	14%
Technology	6,63,778.73	47,44,557.50	14%
Total	14,67,457.29	1,26,42,501.91	12%

Impact of Discount on Profit

Linear regression is used to predict profit based on sales. An upward trend indicates that more profit is generated as the discount increases.

Discount type	Avg Profit	Avg Sales	Count of Order
No Discount	61.04	241.04	29009
Low Discount	46.69	339.62	10953
Medium Discount	-56.19	217.80	7156
High Discount	-98.89	89.08	4172
Total	28.61	246.49	51290

[Back](#) **Predictive Analytics for Profit Optimisation Using Machine Learning**

Detection of Profit Outliers Using Machine Learning Logic

Profit Outlier Detection

Category	Total Profit	Total Sales	% of Profit
Furniture	2,85,082.73	41,10,451.90	7%
Office Supplies	5,18,595.83	37,87,492.51	14%
Technology	6,63,778.73	47,44,557.50	14%
Total	14,67,457.29	1,26,42,501.91	12%

## 4.CONCLUSIONS AND RECOMMENDATIONS

### 4.1 Summary

This project presented a comprehensive business intelligence and predictive analytics solution that uses the Global Superstore dataset to evaluate overall business performance, profitability drivers and future opportunities for profit optimization. Through structured data modeling, interactive dashboards and advanced analytics techniques, key patterns were identified in sales, profits, geography, product categories, shipping modes and discount strategies. The integration of descriptive, clinical and predictive analysis enabled a holistic understanding of both historical trends and future benefit behaviour.

The analysis revealed that although overall sales growth is strong, profitability varies significantly by category, region and operating factors. Technology consistently emerged as the most profitable category, while furniture showed low profit margins despite high sales volumes. Regional analysis highlighted that North America and parts of Asia contribute strongly to profits, while some regions generate high sales but relatively low returns. Evaluation of shipping modes indicates that Standard Class contributes the highest profit volume, while Same Day Shipping increases operating costs with limited profit impact.

In addition, the incorporation of machine learning techniques using R enhanced the analytical depth of the project. Linear regression models showed a strong positive relationship between sales and profit, while discount-based analysis showed that higher discount levels were strongly associated with a decrease in profit. Anomaly detection techniques identify excessive profit and loss transactions, thereby identifying abnormal business behaviour that can affect overall performance. AI-powered visualizations such as decomposition trees

### 4.2 Recommendations

Based on the insights generated from the dashboard and predictive models, the following recommendations are suggested:

1. Optimize discount strategy

The analysis clearly shows that medium and high discount levels have a negative impact on profitability. Companies should limit aggressive discounting and adopt data-driven discount limits, especially for low-margin categories such as furniture. Targeted discounts should be applied selectively to high performing products and customer segments.

2. Focus on high profit categories and subcategories

Technology and select office supplies subcategories consistently deliver high profit margins. Strategic investments, inventory prioritization and marketing efforts should focus on these areas to maximize overall profitability.

3. Improve regional profitability management

Areas that generate high sales but low profits should be closely monitored. Cost optimization strategies, regional price adjustments and logistics improvements should be initiated to convert strong sales volumes into sustainable profits.

#### 4. Rationalize the use of shipping methods

Standard class shipping contributes the most to total profit, while expedited shipping modes increase costs without commensurate profit gains. Encouraging customers to choose cost-effective shipping options can significantly improve margins.

#### 5. Leverage predictive analytics to make decisions

Machine learning models should be incorporated into business planning processes to predict profit results under various sales and discount scenarios. This proactive approach enables management to make decisions before implementation.

### 4.3 Personal conclusion

This project provided valuable hands-on experience in applying Business Intelligence and Machine Learning techniques to real-world datasets. It strengthened my understanding of data modeling, DAX calculations, interactive dashboard design and integration of R-based predictive analytics in Power BI. The use of artificial intelligence, regression models and pivot tables improved my ability to translate complex analytical results into meaningful business insights.

Overall, the project demonstrated how data-driven decision making can significantly improve organizational performance when supported by a strong analytical framework. The combination of descriptive analysis, advanced visualization and predictive modeling has reinforced the importance of analysis in the modern business environment and has contributed positively to my technical and analytical skill development.

# **Big Data and Business Intelligence**



---

## **APPENDIX:BI DESIGN**

---



# 1. DATA PRE-PROCESSING AND DATA CLEANING

## 1.1 Load and Clean Dataset

The primary dataset used in this project was obtained from Kaggle (Global Superstore Dataset) and loaded into Power BI using the Get Data function. The dataset was originally available in a flat-table Excel format that included transaction, customer, product, shipping and geographic information.

A screenshot of the Power Query Editor interface. At the top, there is a formula bar with the text '= Table.RemoveColumns(#"Renamed Columns", {"Postal Code"})'. Below the formula bar is a table with 25 rows and 5 columns. The columns are labeled 'Order ID', 'Order Date', 'Ship Date', 'Ship Mode', and 'Customer ID'. The data in the table includes various order IDs, dates, shipping modes, and customer IDs. Row 25 is highlighted in grey, indicating it is the current row being edited.

	Order ID	Order Date	Ship Date	Ship Mode	Customer ID
1	CA-2012-124891	31-07-2012	31-07-2012	Same Day	RH-19495
2	IN-2013-77878	05-02-2013	07-02-2013	Second Class	JR-16210
3	IN-2013-71249	17-10-2013	18-10-2013	First Class	CR-12730
4	ES-2013-1579342	28-01-2013	30-01-2013	First Class	KM-16375
5	SG-2013-4320	05-11-2013	06-11-2013	Same Day	RH-9495
6	IN-2013-42360	28-06-2013	01-07-2013	Second Class	JM-15655
7	IN-2011-81826	07-11-2011	09-11-2011	First Class	TS-21340
8	IN-2012-86369	14-04-2012	18-04-2012	Standard Class	MB-18085
9	CA-2014-135909	14-10-2014	21-10-2014	Standard Class	JW-15220
10	CA-2012-116638	28-01-2012	31-01-2012	Second Class	JH-15985
11	CA-2011-102988	05-04-2011	09-04-2011	Second Class	GM-14695
12	ID-2012-28402	19-04-2012	22-04-2012	First Class	AJ-10780
13	SA-2011-1830	27-12-2011	29-12-2011	Second Class	MM-7260
14	MX-2012-130015	13-11-2012	13-11-2012	Same Day	VF-21715
15	IN-2013-73951	06-06-2013	08-06-2013	Second Class	PF-19120
16	ES-2014-5099955	31-07-2014	03-08-2014	Second Class	BP-11185
17	CA-2014-143567	03-11-2014	06-11-2014	Second Class	TB-21175
18	ES-2014-1651774	08-09-2014	14-09-2014	Standard Class	PJ-18835
19	IN-2014-11763	31-01-2014	01-02-2014	First Class	JS-15685
20	TZ-2014-8190	05-12-2014	07-12-2014	Second Class	RH-9555
21	PL-2012-7820	08-08-2012	10-08-2012	First Class	AB-600
22	CA-2011-154627	29-10-2011	31-10-2011	First Class	SA-20830
23	IN-2011-44803	02-05-2011	03-05-2011	First Class	JK-15325
24	ES-2013-2860574	27-02-2013	01-03-2013	Second Class	LB-16795
25					

During the data loading phase, several cleaning steps were performed using Power Query (M language). These include removing duplicate records, checking for null or missing values in key fields such as sales, profit and order date, and ensuring correct data types for numeric, categorical and date attributes. Date fields were converted to standard date formats to support time series analysis, while numeric fields were validated to avoid aggregation errors in the measures.

## 1.2 Creating dimension tables for relationships

To support efficient analysis and reduce redundancies, the original flat data set was transformed into a snowflake-style data model. Separate dimension tables were created for key entities such as customer, product, date, and shipping mode, while the central fact table retained transactional measures including sales, profit, volume, discounts, and shipping costs.

This approach improved model clarity, increased performance, and allowed meaningful filtering across scenes. Each dimension table was created by referencing the master dataset and maintaining only the relevant attributes, ensuring consistency and integrity throughout the model.

### Description of Dimension tables

#### 1.customer

The Customer table stores unique customer-related information and is used to analyze sales and performance behavior across different customer segments.

Customer	...
W	City
	Country
	Customer ID
	Customer Name
	Segment
	State
	Profit Margin%
Collapse ^	

Use:

Separating customer data can avoid duplication and accurately analyze a customer's contribution to sales and profits across multiple orders.

#### 2.product

The Product table contains product classification details that support product-level performance analysis.

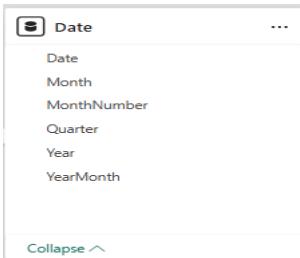
Product	...
	Category
	Product Name
	Sub-Category
Collapse ^	

Use:

This table enables profitability to be compared across product categories and subcategories and supports AI visualizations such as decomposition trees.

#### 3.Date

The Date table enables time-based analysis such as annual trends, monthly performance and seasonality.

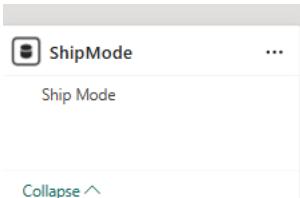


Use:

Using a separate date table allows precise time information, trend analysis and aggregation over different time periods.

#### 4. ShipMode:

The ShipMode table stores shipping-related information to analyze delivery performance and cost effectiveness.

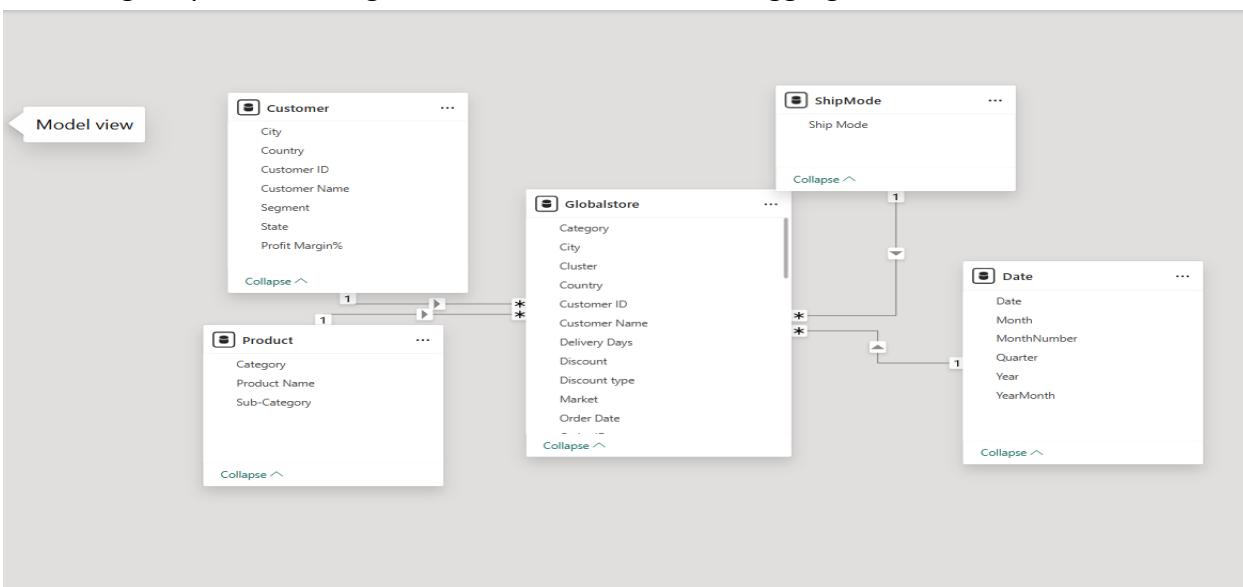


Use:

This enables analysis of how different shipping methods affect sales, profits and delivery efficiency.

## 2. DATA MODELLING

A structured data model was created in Power BI by defining relationships between fact tables and dimension tables. One-to-many relationships were established using appropriate primary and foreign keys, with filter guidelines to ensure accurate aggregation.



The completed data model enabled seamless interaction across visuals, slicers, and advanced analytics features such as decomposition trees, key influencers, and machine learning-based insights.

## 3.DAX AND M-LANGUAGE

### 3.1 DAX

Data analysis expressions (DAX) were used extensively to create calculated measurements and increase analytical depth. Key measures include:

Category	Measure Name	DAX Formula	Purpose / Description
Sales Metrics	Total Sales	Total Sales = SUM(Globalstore[Sales])	Calculates total revenue generated
	Avg Sales	Avg Sales = AVERAGE(Globalstore[Sales])	Calculates average sales per order
Profit Metrics	Total Profit	Total Profit = SUM(Globalstore[Profit])	Calculates total profit
	Avg Profit	Avg Profit = AVERAGE(Globalstore[Profit])	Calculates average profit
	Percentage of Profit	Percentage of Profit = DIVIDE([Total Profit], [Total Sales])	Calculates profit margin as a percentage
Order Metrics	Total Orders	Total Orders = DISTINCTCOUNT(Globalstore[Order ID])	Counts unique orders
	Count of Order	Count of Order = COUNT(Globalstore[Order ID])	Counts total order records
Discount Metrics	Average Discount	Average Discount = AVERAGE(Globalstore[Discount])	Calculates average discount applied
	Avg Discount	Avg Discount = AVERAGE(Sheet1[Discount])	Calculates average discount from source table

<b>Shipping Metrics</b>	Average Delivery Days	Average Delivery Days AVERAGE(Globalstore[Delivery Days])	=	Calculates average delivery duration
	Average Shipping Cost	Average Shipping Cost AVERAGE(Globalstore[Shipping Cost])	=	Calculates average shipping cost
<b>Customer Metrics</b>	Top Customer	Top Customer = CALCULATE( SELECTEDVALUE(Customer[Customer Name]), TOPN(1, ALL(Customer), [CLV], DESC))		Identifies customer with highest lifetime value

These measures enabled dynamic calculations in filters and slicers, and supported comparative and trend-based analysis across the dashboard.

### 3.2 Machine learning with R

Machine learning techniques were implemented using R scripts embedded in Power BI visualizations. Linear regression models were developed to analyze the relationship between sales, discounts and profits. In addition, R-based reasoning was used to identify profit drivers and understand unusual profit behaviour.

The integration of R enables predictive analytics directly in Power BI, increasing the analytical sophistication of the dashboard and supporting data-driven decision making.

#### 1.R Script for Profit Prediction Using Linear Regression (R):

Code:

```
# dataset <- data.frame(Total Profit, Discount, Total Orders)
# dataset <- unique(dataset)
model <- lm('Total Profit' ~ 'Total Sales' + Discount, data = dataset)
plot(dataset$`Total Sales`, dataset$`Total Profit`,
     main="Profit Prediction based on Sales & Discount",
     xlab="Total Sales",
     ylab="Total Profit",
     pch = 25,
     col = "blue"
)
abline(model, col="#7C0208", lwd=4)
```

#### 2.R Script Profit Prediction using Machine Learning Regression Model

Code:

```
# dataset <- data.frame(Sales, Profit, Discount, Quantity)
# dataset <- unique(dataset)
```

```

# Paste or type your script code here:
#removing the Missing values
head(dataset)
dataset <- na.omit(dataset)
#Buliding Regression model(ML model)
model <- lm(Profit ~ Sales + Discount + Quantity, data = dataset)
# Step 3: Plot the data
plot(dataset$Sales, dataset$Profit,
      main = "Profit Prediction using Machine Learning(Regression Model)",
      xlab = "Sales",
      ylab = "Profit",
      pch = 25,
      col = "#7C0208")

# Step 4: Add prediction line
abline(model, col = "green", lwd = 2)

3. R Script of Impact of Discount on Profit Code:
# dataset <- data.frame(undefined, undefined.1, undefined.2)
# dataset <- unique(dataset)
data <- na.omit(dataset)
model <- lm(Profit ~ Discount, data = data)
plot(
  data$Discount,
  data$Profit,
  pch = 25,
  col = "#7C0208",
  xlab = "Discount",
  ylab = "Profit",
  main = "Impact of Discount on Profit"
)
abline(model, col = "green", lwd = 2)

4. R Script of Impact of Profit Outlier Detection
# dataset <- data.frame(Profit, Sales)
# dataset <- unique(dataset)
dataset <- na.omit(dataset)

```

```

Q1 <- quantile(dataset$Profit, 0.25)
Q3 <- quantile(dataset$Profit, 0.75)
IQR <- Q3 - Q1
dataset$outlier <- dataset$Profit < (Q1 - 1.5*IQR) |
    dataset$Profit > (Q3 + 1.5*IQR)
plot(dataset$Sales, dataset$Profit,
    col = ifelse(dataset$outlier, "#7C0208", "blue"),
    pch = 19,
    main = "Profit Outlier Detection",
    xlab = "Sales",
    ylab = "Profit")

```

## 4.DASHBOARD DESIGN

### 4.1 Home Page

The home page serves as the navigation hub for the report, providing quick access to all analytical sections. It includes interactive buttons and clear headings to guide users through various analytical approaches.



The figure shows a summary page of a dashboard. On the left, there is a vertical sidebar with ten menu items, each with a small icon and text. The items are:

- 1.1 Overall Business Performance & Profitability Overview
- 1.2 Overall Business Performance & Profitability Overview
- 2.1 Profit & Performance Deep-Dive Analysis
- 2.2 Profit & Performance Deep-Dive Analysis
- 3. Geographical and Category-Level Profit Insights
- 4.1 AI-Driven Profitability Drivers & Root Cause Analysis
- 4.2 AI-Driven Profitability Drivers & Root Cause Analysis
- 5.1 Predictive Analytics for Profit Optimisation Using Machine Learning
- 5.2 Predictive Analytics for Profit Optimisation Using Machine Learning

The sidebar items 2.1, 2.2, 3, 4.1, 4.2, 5.1, and 5.2 are highlighted with yellow rectangular overlays. The main area of the dashboard features a photograph of a modern supermarket at night. The supermarket has a two-story facade with large glass windows and doors. A prominent blue sign on the upper level reads "Global Superstore". The building is illuminated from within, showing shelves stocked with products. The overall design is clean and professional.

Figure: Summary page

## 4.2 Overall Business Performance & Profitability Overview

This page provides a high-level summary of business performance using KPIs, trend charts and customer-level insights. It focuses on understanding overall sales growth, profitability trends and key contributors to revenue.

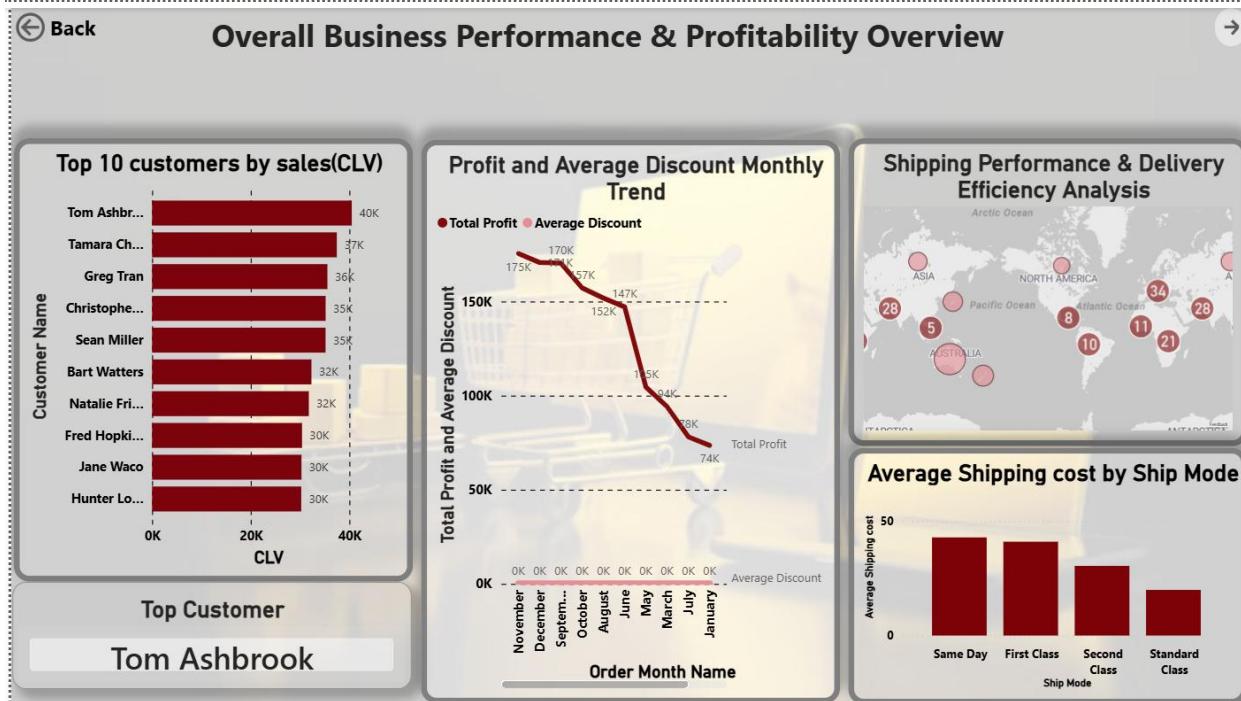
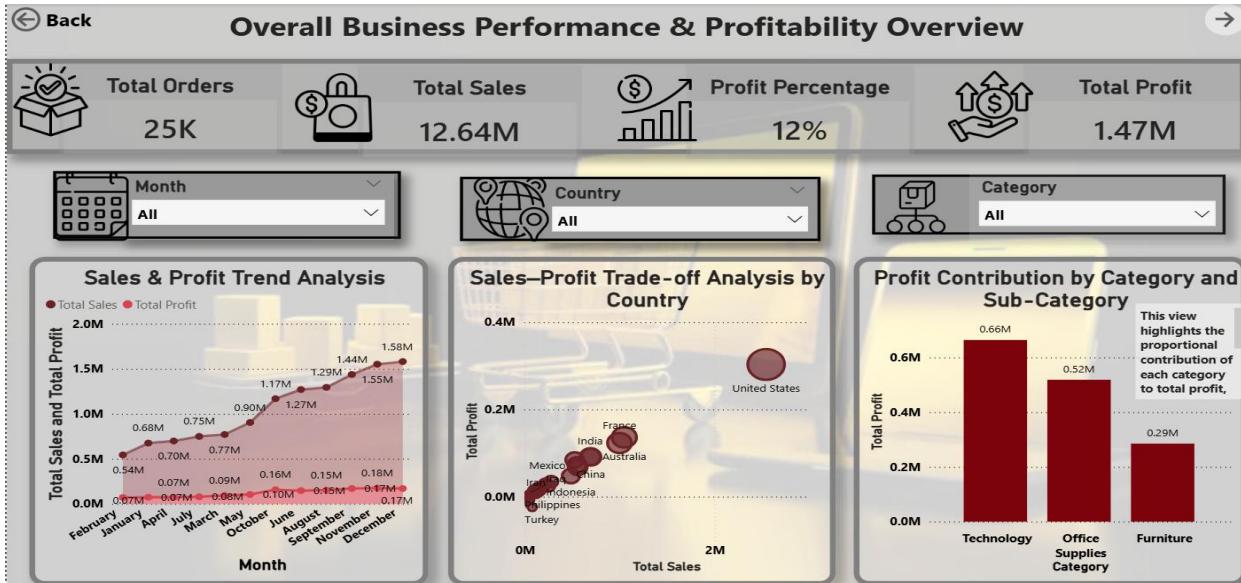


Figure: Overall Business Performance & Profitability Overview

## Tooltips:

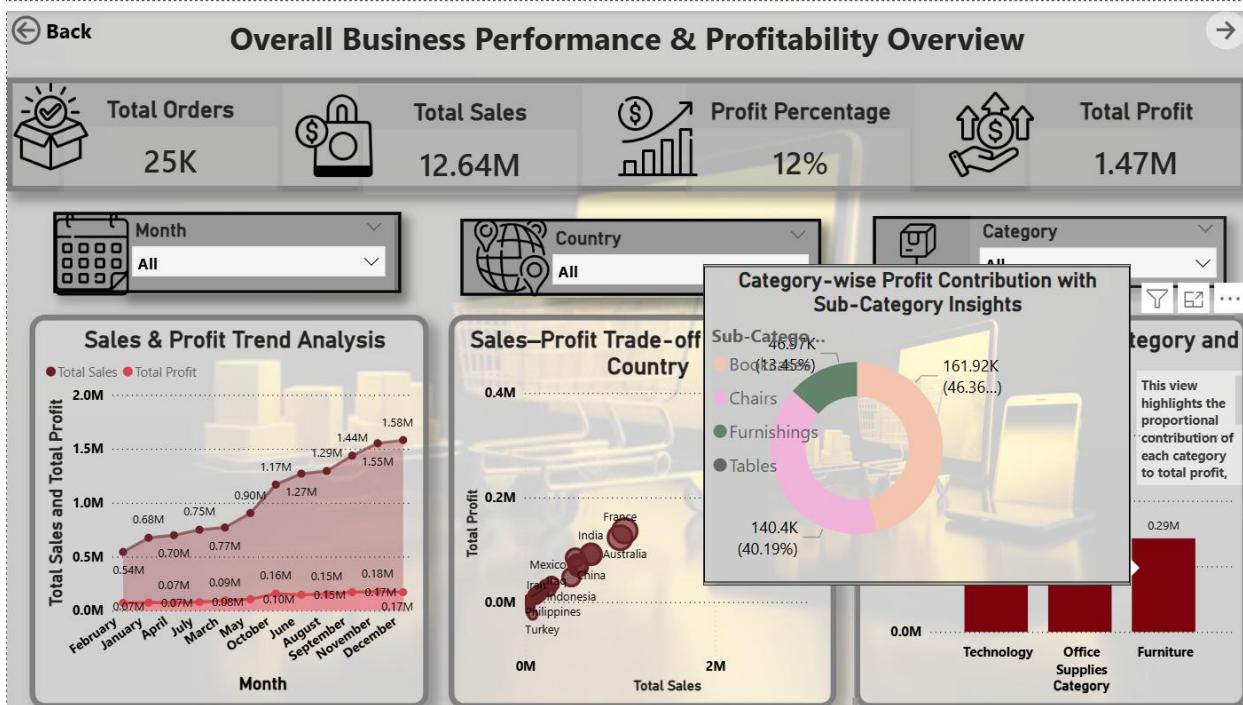
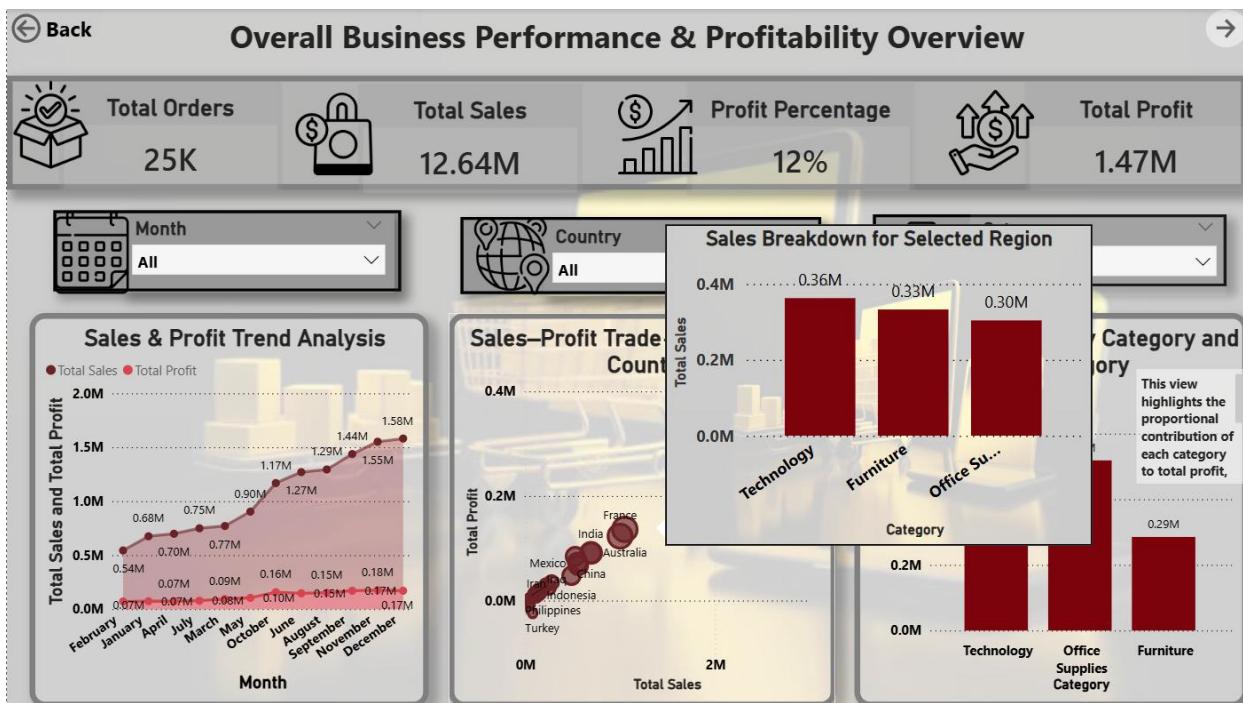


Figure :Overall Business Performance & Profitability Overview with tooltips

## 4.3 Profit & Performance Deep-Dive Analysis

This section provides a detailed comparison of sales, profits and profit margins across categories, subcategories and time periods. Helpful charts and trend views help you identify performance variations and seasonal patterns.

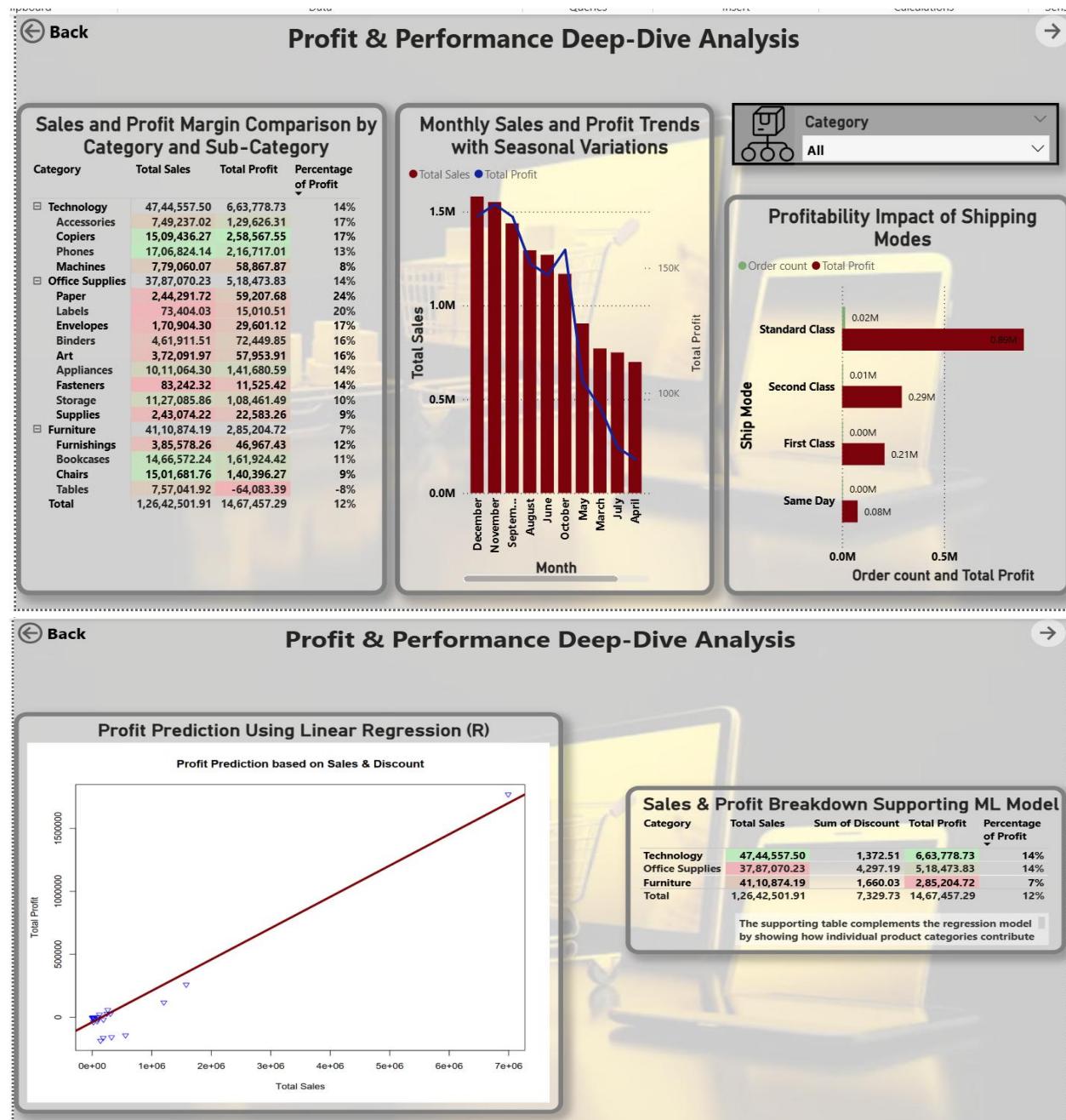


Figure Profit & Performance Deep-Dive Analysis

## 4.4 Geographical and Category-Level Profit Insights

Geographic maps, waterfall maps and multimetric visualizations were used to analyze profit distribution across regions and product categories. This page highlights regional performance differences and category-level contributions to overall earnings.

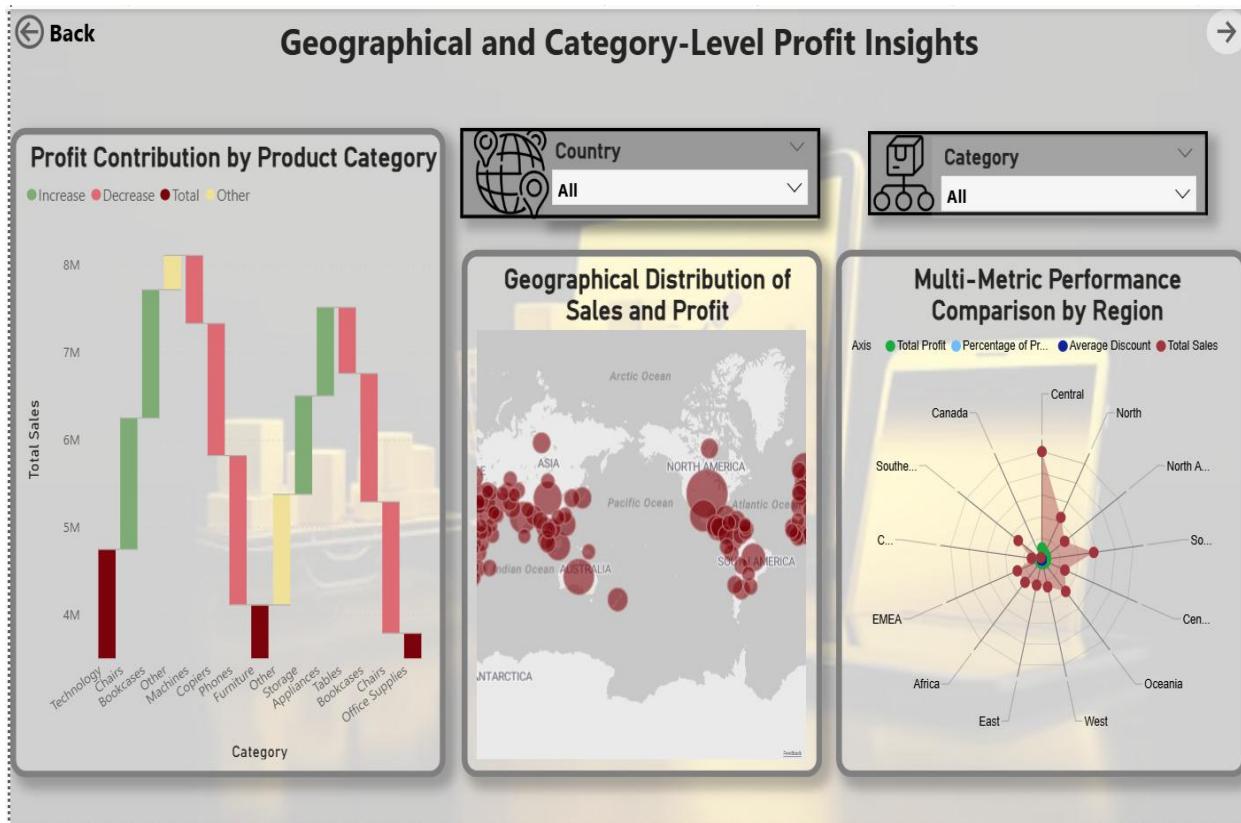


Figure Geographical and Category-Level Profit Insights

## 4.5 AI-Driven Profitability Drivers & Root Cause Analysis

Advanced AI visualizations such as decomposition trees and key influencer analysis were used to identify the main drivers of profit growth and loss. These views help explain why certain categories, regions or shipping modes affect profitability.

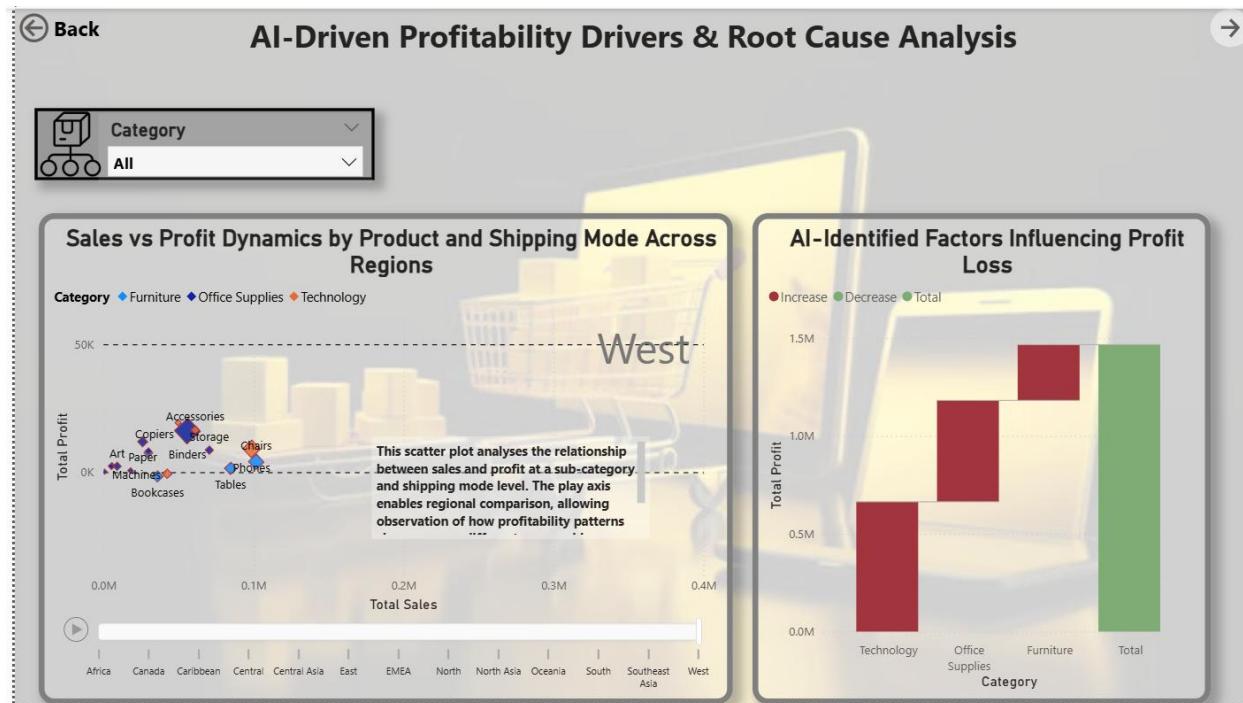
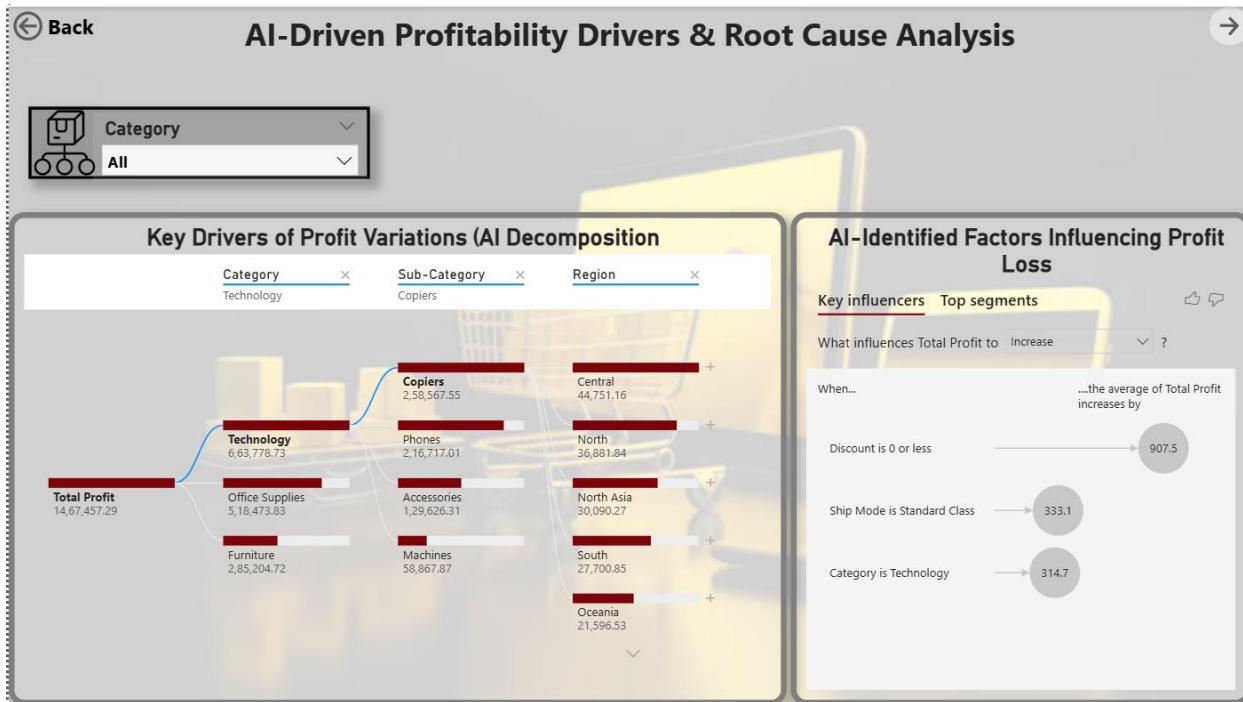


Figure AI-Driven Profitability Drivers & Root Cause Analysis

## 4.6 Predictive Analytics for Profit Optimisation Using Machine Learning

This page focuses on predictive insights using machine learning models built in R. Regression analysis was used to predict profits based on sales and discount levels, while anomaly detection identified abnormal profit behavior. Added support tables to validate and interpret model output.

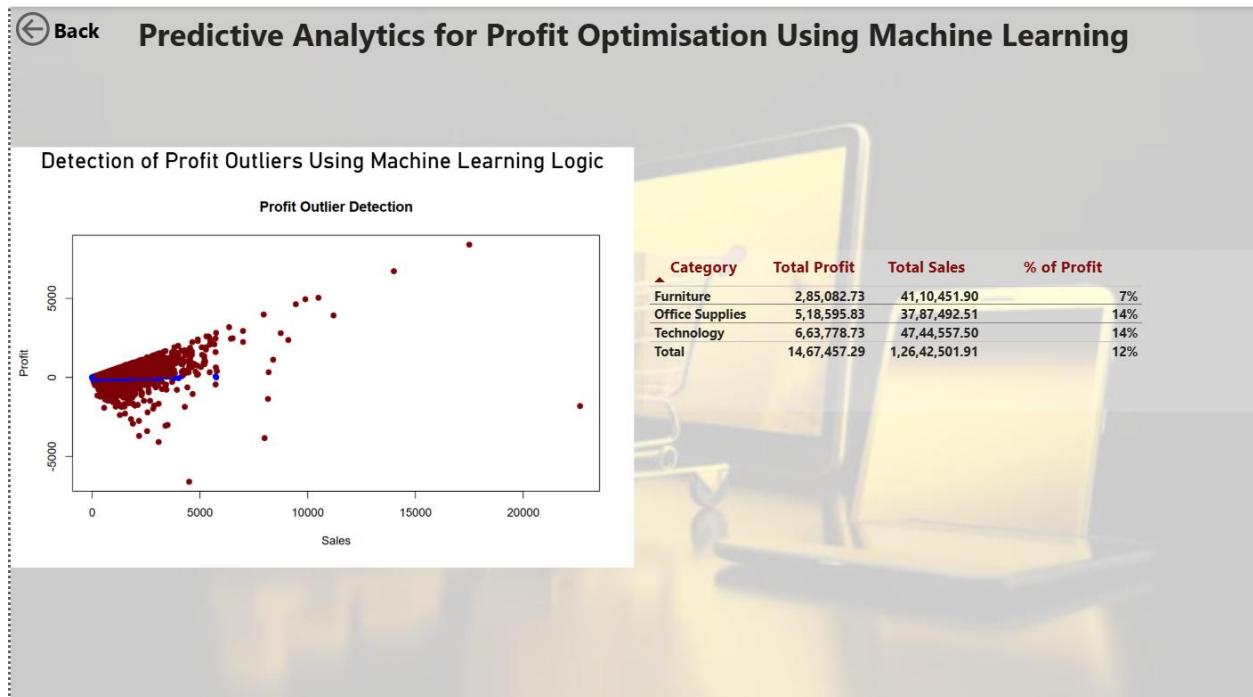
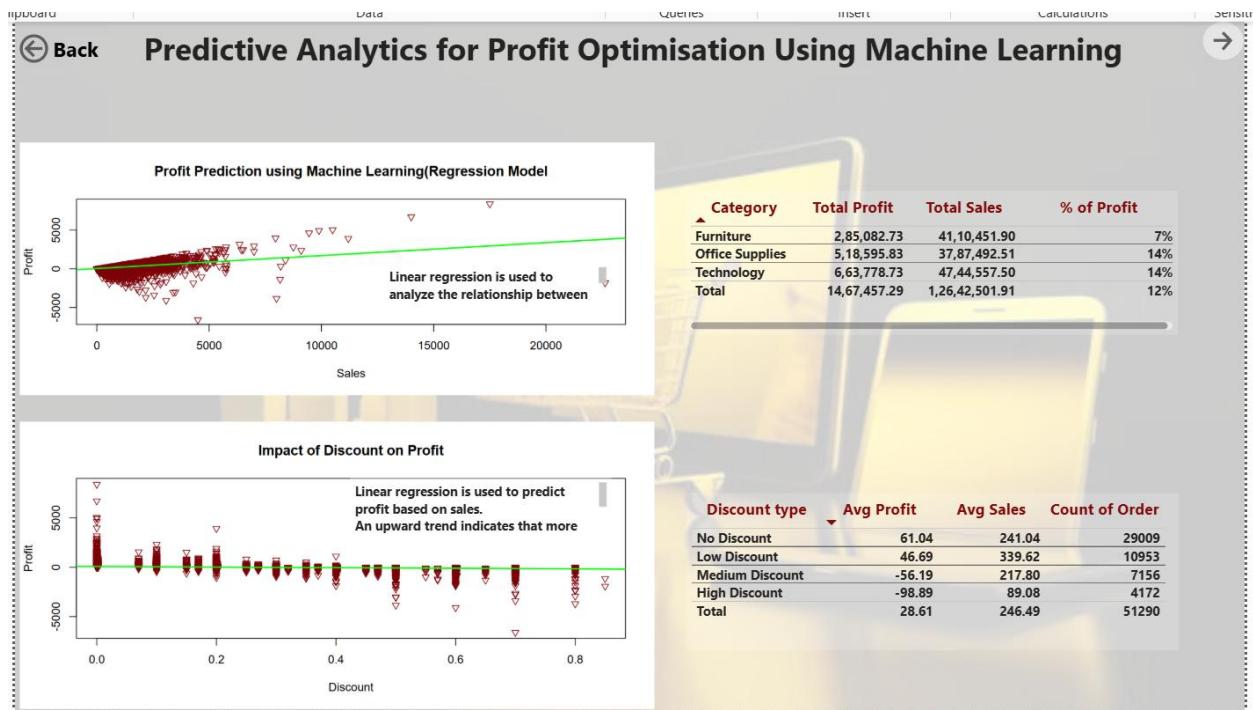


Figure: Predictive Analytics for Profit Optimisation Using Machine Learning

## 5.REFERENCE

Wang, K., Lee, F.C. and Lai, J. (2000) "Operation principles of bi-directional full-bridge DC/DC converter with unified soft-switching scheme and soft-starting capability," in *APEC 2000. Fifteenth Annual IEEE Applied Power Electronics Conference and Exposition (Cat. No.00CH37058)*, pp. 111–118 vol.1. Available at: <https://doi.org/10.1109/APEC.2000.826092>.

Singh, G. et al. (2023) "Data Visualization for Developing Effective Performance Dashboard with Power BI," in *2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA)*, pp. 968–973. Available at: <https://doi.org/10.1109/ICIDCA56705.2023.10100169>.

Vashisht, V. et al. (2021) "An Effective Approach for Integrating Microsoft Power BI Application with Python for Predictive Analytics," in D.K. Sharma et al. (eds.) *Micro-Electronics and Telecommunication Engineering*. Singapore: Springer, pp. 469–477. Available at: [https://doi.org/10.1007/978-981-33-4687-1\\_45](https://doi.org/10.1007/978-981-33-4687-1_45).

Sharif, M.S. et al. (2024) "A Comparative Study of Sales Prediction Using Machine Learning Models: Integration of PySpark and Power BI," in *2024 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*, pp. 137–142. Available at: <https://doi.org/10.1109/3ict64318.2024.10824620>.