Sales Data Analysis Report

Introduction

The dataset used in this project was obtained from the Kaggle website and contains sales data of an e-commerce company over several years. It includes information such as product categories, sales, profit, shipping cost, delivery mode, and order priority.

The objective of this project is to analyse sales performance and profitability to answer key business questions related to profit margins, shipping efficiency, and product performance. The analysis helps identify low-performing areas and opportunities to improve overall sales and profit.

This project involves data cleaning and feature selection using Microsoft Excel, data analysis using MySQL, and interactive dashboard development using Power BI to generate meaningful business insights and support data-driven decision-making.

Problem Statement

The e-commerce company generates a large volume of sales data over multiple years. However, high sales do not always guarantee high profitability. The company wants to understand which product categories and sub-categories are underperforming in terms of profit margin, how shipping costs affect overall profitability, and whether delivery priorities are priced efficiently.

Objectives

The main objectives of this project are:

- To compare category-level profit margins with the overall business profit margin.
- To identify sub-categories that are contributing to low profitability.
- To analyse the relationship between shipping cost and profit margin.
- To evaluate the profitability of different order priority levels.

Dataset Description

The dataset used in this project was obtained from Kaggle and contains sales data of an e-commerce company.

The dataset includes information such as:

- Row ID
- Order ID
- Order Date
- Ship Date
- Ship Mode
- Customer ID
- Customer Name
- Segment
- City
- State
- Country-
- Product ID
- Category
- Sub-Category
- Product Name
- Sales
- Quantity
- Discount
- Shipping Cost
- Order Priority

The dataset contains 51285 records and 20 columns, covering sales transactions over the period 2012 to 2016.

This dataset provides detailed transactional-level data that allows analysis of revenue generation, cost structure, and profitability trends.

Methodology

This project followed a structured approach consisting of three stages: data cleaning in Microsoft Excel, data analysis in MySQL, and data visualization in Power BI. Excel was used to prepare and refine the dataset, MySQL was used to perform analytical queries and generate insights, and Power BI was used to create a dynamic dashboard for visualizing business performance.

Data Preparation Using Microsoft Excel

The raw dataset obtained from Kaggle was first cleaned and transformed using Microsoft Excel to ensure data quality and relevance.

Data Cleaning Steps

- Removed duplicate records to prevent incorrect aggregation.
- Identified and eliminated null or missing values.
- Deleted unnecessary rows that were not relevant to the company's analytical requirements.

Feature Selection

To focus only on relevant business variables, the following columns were removed:

- Row ID
- Customer ID
- Customer Name
- Product Name
- Country
- Discount

These columns were excluded because they were not required for profitability, shipping, and performance analysis.

Feature Engineering (New Columns Created)

To enhance analytical capability, the following calculated columns were added:

- Shipment Days – Calculated as the difference between Ship Date and Order Date to measure delivery efficiency.
- Profit Margin (%) – Calculated using the formula (Profit / Sales) × 100.
- Unit Price – Derived by dividing Sales by Quantity (where applicable).

These transformations ensured that the dataset was structured, relevant, and ready for advanced SQL analysis.

Data Analysis Using MySQL

After cleaning and restructuring the dataset, it was imported into MySQL for detailed querying and analysis.

MySQL was used to perform structured data analysis and answer the business questions provided by the organization. Various SQL techniques were applied to ensure accurate calculations and logical interpretation.

SQL Techniques Used

Aggregate Functions

Functions such as SUM(), AVG(), COUNT(), and ROUND() were used to calculate total sales, total profit, average delivery days, and profit margins.

GROUP BY Clause

The GROUP BY clause was used to analyse data at different levels, including:

- Category
- Sub-category
- Shipping mode
- Order priority

This allowed comparison of profitability across various business segments.

Control Flow Statements

CASE WHEN statements were used to classify categories and shipping modes as above-average or below-average performers.

Common Table Expressions (CTEs)

CTEs were used to simplify complex queries and break calculations into logical steps, improving readability and efficiency.

Window Functions

Window functions such as DENSE_RANK() were used to rank shipping modes and compare performance metrics without losing detailed information.

Joins

Joins were applied wherever necessary to combine related data and ensure comprehensive analysis.

Analytical Queries and Insights (MySQL)

1. Category Profit Margin Analysis

Objective

To identify which product categories are performing above or below the overall business profit margin (22%).

Analysis Approach

Each category's profit margin was calculated and compared with the overall company profit margin of 22%.

```
SELECT
  s.Category AS category,
  ROUND(overall_pm.overall_profit_margin_pct, 2) AS overall_profit_margin_pct,
  ROUND((SUM(s.Profit) / SUM(s.Sales)) * 100, 2) AS category_profit_margin_pct,
  CASE
    WHEN (SUM(s.Profit) / SUM(s.Sales)) * 100
      >= overall_pm.overall_profit_margin_pct
    THEN 'Above Average'
    ELSE 'Below Average'
  END AS profitability_status
FROM sales AS s
CROSS JOIN (
  -- Overall business profit margin
  SELECT
    (SUM(Profit) / SUM(Sales)) * 100 AS overall_profit_margin_pct
  FROM sales
) AS overall_pm
GROUP BY
  s.Category,
  overall_pm.overall_profit_margin_pct
ORDER BY
  category_profit_margin_pct;
```

category	overall_profit_margin_pct	category_profit_margin_pct	profitability_status
Furniture	21.89	19.64	Below Average
Technology	21.89	23.15	Above Average
Office Supplies	21.89	23.26	Above Average

Findings

The analysis shows that the **Furniture** category has a profit margin of **21%**, which is below the overall business margin. Although this category generates strong sales revenue, its profitability is relatively low.

Business Interpretation

High sales with low margin indicates pricing inefficiency or higher procurement costs within the category.

Recommendation

- Review and optimize selling prices.
- Negotiate with suppliers to reduce the cost of buying goods or services.

---

2. Sub-Category Margin Leakage Analysis

Objective

To identify specific sub-categories contributing to low profit margins within underperforming categories.

Analysis Approach

Sub-category profit margins were calculated and compared against the overall business margin benchmark of 20%.

```
SELECT
  s.sub_category,
  b.net_profit_margin,
  ROUND((SUM(s.profit) / SUM(s.sales)) * 100) AS profit_margin,
  CASE
    WHEN ROUND((SUM(s.profit) / SUM(s.sales)) * 100) >= b.net_profit_margin
    THEN 'Profitable'
    ELSE 'Less Profitable'
  END AS profitability
FROM sales AS s
CROSS JOIN (
  SELECT ROUND((SUM(profit) / SUM(sales)) * 100) AS net_profit_margin
  FROM sales
  where category="Furniture"
) AS b
GROUP BY s.sub_category, b.net_profit_margin
order by profit_margin;
```

sub_category	net_profit_margin	profit_margin	profitability
Bookcases	20	18	Less Profitable
Chairs	20	20	Profitable
Machines	20	20	Profitable
Tables	20	20	Profitable
Phones	20	22	Profitable
Labels	20	22	Profitable
Appliances	20	23	Profitable
Storage	20	23	Profitable
Accessories	20	23	Profitable
Paper	20	23	Profitable
Supplies	20	23	Profitable
Art	20	24	Profitable
Envelopes	20	24	Profitable
Fasteners	20	24	Profitable
Binders	20	25	Profitable
Copiers	20	26	Profitable
Furnishings	20	27	Profitable

Findings

The following sub-categories were found to have margins below the benchmark:

- Bookcases – 18%
- Chairs – 20%
- Tables – 20%
- Machines – 20%

Although these sub-categories generate high sales volumes, their profit margins remain below average.

Business Interpretation

This indicates margin leakage due to either high purchasing costs or suboptimal pricing strategies.

Recommendation

- Re-evaluate pricing structure.
- Increase procurement volume to reduce buying costs.
- Focus on higher-margin product variants.

---

3. Delivery Priority and Profitability Analysis

Objective

To evaluate whether high-priority deliveries generate proportionally higher profit margins.

Analysis Approach

The average delivery days and profit margins were calculated for each order priority level.

```
WITH order_priority_cte AS (
  SELECT
    order_priority AS order_priority,
    ROUND(AVG(shipment_days)) AS avg_shipment_days,
    ROUND((SUM(profit) / SUM(sales)) * 100, 2) AS profit_margin_pct
  FROM sales
  GROUP BY order_priority
)

SELECT
  order_priority,
  avg_shipment_days,
  profit_margin_pct,
  DENSE_RANK() OVER (ORDER BY avg_shipment_days DESC) AS
shipment_days_rank,
  DENSE_RANK() OVER (ORDER BY profit_margin_pct ASC) AS
profit_margin_rank
FROM order_priority_cte;
```

order_priority	avg_shipment_days	profit_margin_pct	shipment_days_rank	profit_margin_rank
Low	6	19.05	1	1
High	3	20.69	3	2
Medium	5	21.93	2	3
Critical	2	26.79	4	4

Findings

High-priority orders are delivered in an average of **3 days**, demonstrating fast service. However, the average profit margin for high-priority orders is **21%**, which is lower than that of medium-priority orders.

Business Interpretation

High-priority deliveries involve faster processing and higher logistics costs. Ideally, they should generate higher profit margins. The current situation suggests pricing misalignment and reduced profitability.

Recommendation

- Increase delivery charges for high-priority orders.
- Align pricing with service-level commitments.
- Ensure premium services generate premium profits.

---

4. Shipping Cost vs Profit Margin Analysis

Objective

To analyze how shipping costs impact profit margins across different shipping modes.

Analysis Approach

Shipping cost percentage and profit margins were calculated for each shipping mode and compared.

```
WITH ship_mode_summary AS (
  SELECT
    ship_mode,
    ROUND(SUM(shipment_days) / COUNT(shipment_days)) AS avg_days,
    ROUND((SUM(shipping_cost) / SUM(sales)) * 100, 2) AS shipping_cost_pct,
    ROUND((SUM(profit) / SUM(sales)) * 100, 2) AS profit_margin
  FROM sales
  GROUP BY ship_mode
)

SELECT
  ship_mode,
  avg_days,
  shipping_cost_pct,
  profit_margin,
  DENSE_RANK() OVER (ORDER BY shipping_cost_pct DESC) AS
shipping_cost_rank,
  DENSE_RANK() OVER (ORDER BY profit_margin ASC) AS profit_margin_rank
FROM ship_mode_summary
ORDER BY shipping_cost_rank;
```

ship_mode	avg_days	shipping_cost_pct	profit_margin	shipping_cost_rank	profit_margin_rank
Same Day	0	21.21	24.01	1	3
First Class	2	18.73	25.73	2	4
Second Class	3	11.73	21.89	3	2
Standard Class	5	8.19	21.17	4	1

Findings

- Same Day shipping has the highest shipping cost percentage but does not generate the highest profit margin.
- First Class shipping generates the highest profit margin despite high shipping costs.
- Standard and Second Class shipping modes are cost-efficient but produce comparatively lower margins.

Business Interpretation

Same Day shipping may be underpriced relative to its operational cost. First Class shipping demonstrates effective cost recovery. Standard and Second Class shipping generate lower profitability per order.

Recommendation

- Re-evaluate pricing for Same Day shipping.
- Maintain pricing strategy for First Class shipping.
- Optimize revenue strategies for Standard and Second Class shipping.

Final Conclusion

The analysis highlights pricing inefficiencies and cost structure challenges affecting overall profitability. By optimizing category pricing, reducing procurement costs, and aligning shipping charges with service levels, the organization can significantly improve financial performance.