Leena Cherian

Retail Business Performance & Profitability Analysis

📌 Introduction

The retail industry generates large volumes of transactional data that, when properly analysed, can uncover key business insights. This project aims to analyse retail sales data to identify unprofitable product categories, inventory inefficiencies, and seasonal trends in order to guide strategic decision-making.

📄 Abstract

This project focuses on understanding retail business performance using real-world sales data. Key objectives include calculating profit margins, identifying slow-moving or overstocked products, and examining the correlation between discounting, sales, and profitability. Through SQL, Python, and Power BI, we cleaned and analysed the data, created visual dashboards, and extracted actionable insights.

🛠 Tools Used

- **SQL (MySQL):** Data cleaning and aggregation
- **Python (Pandas, Seaborn):** Correlation and statistical analysis
- **Power BI:** Interactive dashboard for visual storytelling

---

🔧 Steps Involved in Building the Project

1. **Data Import & Cleaning (SQL):**
    - Loaded superstore_sales.csv into MySQL
    - Removed null/missing values and duplicate records
2. **SQL Analysis:**
    - Calculated profit margins by category and sub-category
    - Identified high-discount, low-profit items
3. **Python Analysis:**
    - Correlation analysis between discount, profit, sales, and quantity
    - Used Pandas and Seaborn to visualize insights
4. **Power BI Dashboard:**
    - Created interactive visuals (bar, line, scatter plot) to show profit, sales, and discount trends.
    - Added filters for region and category to explore data dynamically.
    - Identified top categories, low-profit items, and monthly sales patterns.
    - Included a reset button to clear all slicers easily.

✅ Conclusion

The analysis revealed that certain sub-categories with high discounts consistently yield low profits. Additionally, correlations showed that deep discounts do not always translate into higher sales or profit. The final dashboard allows business users to explore profit patterns interactively, enabling better inventory and pricing decisions.

SQL code used

```sql
create database Sales;
use sales;
CREATE TABLE superstore_sales (Row_ID INT, Order_ID VARCHAR(20), Order_Date DATE, Sh
DELETE FROM superstore_sales
WHERE Profit IS NULL OR Sales IS NULL OR Category IS NULL;
SELECT * FROM superstore_sales
WHERE Profit IS NULL OR Sales IS NULL OR Category IS NULL
   OR Sub_Category IS NULL OR Order_Date IS NULL;
SELECT ï»¿order_id, COUNT(*) AS cnt
FROM superstore_sales_
GROUP BY ï»¿order_id
HAVING cnt > 1;
DELETE s1
FROM superstore_sales s1
JOIN superstore_sales s2
ON s1.`ï»¿order_id` = s2.`ï»¿order_id`
AND s1.Row_ID > s2.Row_ID;
select * from superstore_sales;
ALTER TABLE superstore_sales
ADD COLUMN Row_ID int NOT NULL AUTO_INCREMENT PRIMARY KEY;
ALTER TABLE superstore_sales
drop  Row_ID ;
SELECT `ï»¿order_id`, COUNT(*) AS cnt
```

```sql
SELECT `ï»¿order_id`, COUNT(*) AS cnt
FROM superstore_sales
GROUP BY `ï»¿order_id`
HAVING cnt > 1;
SELECT
    category,
    sub_category,
    ROUND(SUM(profit), 2) AS total_profit,
    ROUND(SUM(sales), 2) AS total_sales,
    ROUND(SUM(profit) / SUM(sales) * 100, 2) AS profit_margin_percent
FROM superstore_sales
GROUP BY category, sub_category
ORDER BY profit_margin_percent ASC;
select * from superstore_sales;
```

Python (Panda code used)

```python
# 📦 Import libraries
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# 📄 Load dataset
df = pd.read_csv('sql final project.csv')

# 🔍 View data structure
print(df.head())
print(df.info())

# ✅ Check for missing values
print(df.isnull().sum())

# 🏷️ Unique categories
print(df['category'].unique())
```

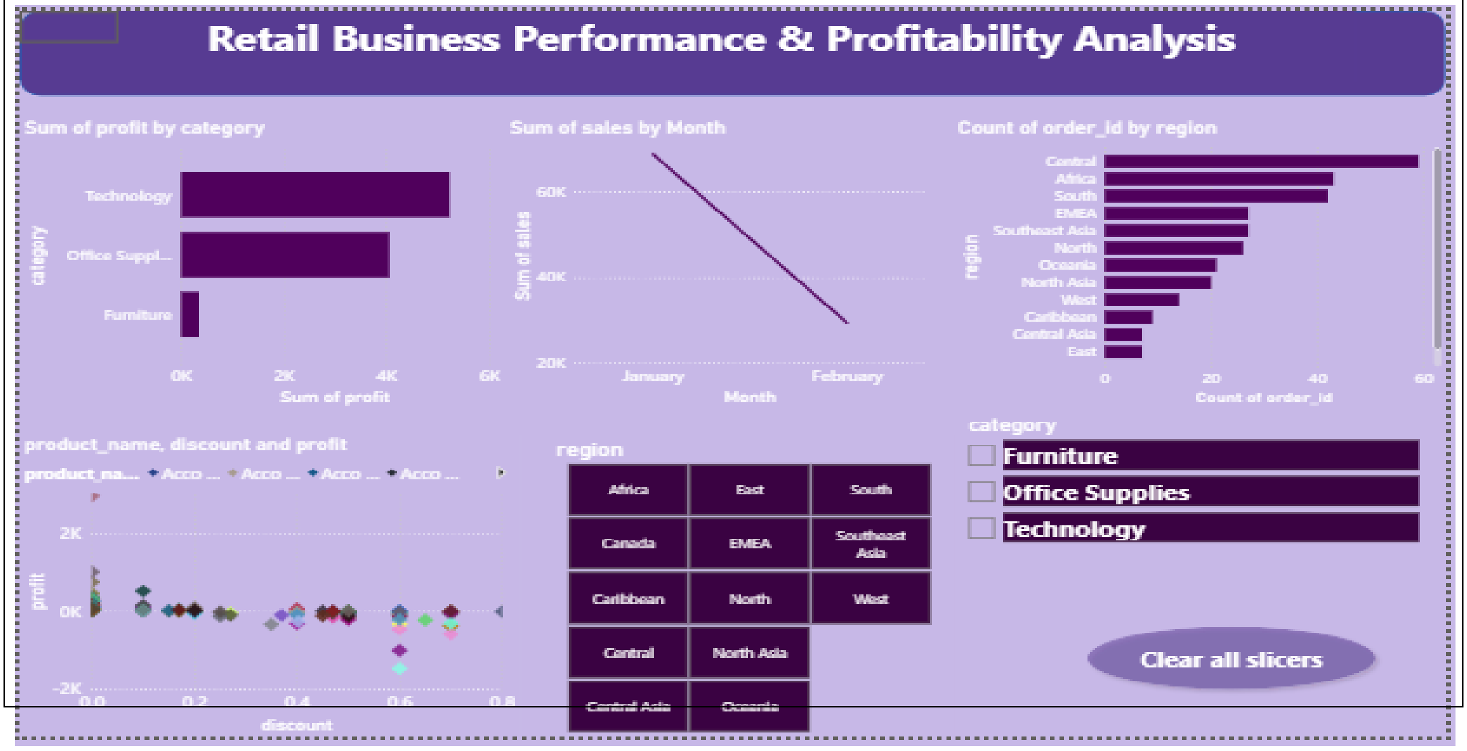
```python
# 📈 Correlation between numerical features
correlation = df[['sales', 'profit', 'quantity', 'discount']].corr()
print(correlation)

# 🔥 Heatmap of correlation
sns.heatmap(correlation, annot=True, cmap='coolwarm')
plt.title('Correlation Matrix')
plt.show()

# 📉 Scatter plot: Discount vs Profit
plt.figure(figsize=(6, 4))
sns.scatterplot(x='discount', y='profit', data=df)
plt.title('Discount vs Profit')
plt.xlabel('Discount')
plt.ylabel('Profit')
plt.grid(True)
plt.show()
```

```python
# 📊 Scatter plot: Quantity vs Profit
plt.figure(figsize=(6, 4))
sns.scatterplot(x='quantity', y='profit', data=df)
plt.title('Quantity vs Profit')
plt.xlabel('Quantity')
plt.ylabel('Profit')
plt.grid(True)
plt.show()
```

PowerBI Dashboard


Retail Business Performance & Profitability Analysis
Sum of profit by category
Technology
Office Suppl...
Furniture
category
0K
2K
4K
6K
Sum of profit
Sum of sales by Month
60K
40K
20K
Sum of sales
January
February
Month
Count of order_id by region
Central
Africa
South
EMEA
Southeast Asia
North
Oceania
North Asia
West
Caribbean
Central Asia
East
region
0
20
40
60
Count of order_id
product_name, discount and profit
product_na... Acco ... Acco ... Acco ... Acco ...
2K
0K
-2K
profit
0.0
0.2
0.4
0.6
0.8
discount
region
Africa
East
South
Canada
EMEA
Southeast Asia
Caribbean
North
West
Central
North Asia
Central Asia
Oceania
category
Furniture
Office Supplies
Technology
Clear all slicers