Retail Business Performance & Profitability Analysis

🗓️ Project Overview

This project aims to analyze the performance and profitability of a retail business using Python, SQL, and Tableau. By exploring sales, inventory, seasonal trends, and profitability, the report provides strategic insights to improve business decisions.

💡 Tools & Technologies Used

- **Python (Pandas, Matplotlib, Seaborn)** for EDA and profitability analysis
- **SQLite** for SQL-based category-wise profit and margin calculations
- **Tableau** for interactive dashboard and visual storytelling

---

1. Data Preparation

Steps Performed (Using Python Pandas):

- Imported dataset and converted the 'Date' column to datetime format for time-series analysis.
- Cleaned column names to remove spaces and special characters to ensure compatibility with SQL queries.
- Checked for missing/null values and duplicate records.
- Result: No significant missing values or duplicates were found, ensuring a clean dataset for analysis.

```
df = pd.read_csv("retail_store_inventory.csv")

df['Date'] = pd.to_datetime(df['Date'])

df.columns = df.columns.str.strip().str.replace(' ', '_').str.replace('/', '_').str.replace('-', '_')
```

---

2. Exploratory Data Analysis (EDA)

2.1 Dataset Overview

Used .info(), .describe(), .isnull().sum(), and .duplicated().sum() to:

- Understand data structure
- View summary statistics
- Identify missing or duplicate values

2.2 Visual Exploration

- **Histograms:** Showed skewed distributions for Inventory Level and Demand Forecast.
- **Category Distribution:** Clothing and Groceries had the highest product count.
- **Region Distribution:** South and North had more transactions compared to East and West.
- **Correlation Heatmap:**
    - Strong positive correlation between Units Ordered and Demand Forecast (business logic confirmed).
    - Weak correlation between Inventory Level and Profitability suggests other factors impact profits.

2.3 Seasonal Trends

- **Line Chart of Daily Sales:** Seasonal spikes observed, especially in April and October.
- **Bar Plot of Units Sold by Season:** Winter had the lowest average units sold, pointing to weaker performance.

---

3. SQL-Based Profitability Analysis

3.1 Query to Calculate Profitability by Category

Used SQLite to calculate:

- Total Revenue
- Total Profit (adjusted for discount and competitor pricing)
- Profit Margin (%)

```
SELECT
    Category,
    ROUND(SUM(Price * Units_Sold), 2) AS Total_Revenue,
    ROUND(SUM((Price - Competitor_Pricing) * (1 - Discount/100.0) * Units_Sold), 2) AS
Total_Profit,
```

```
    ROUND((SUM((Price - Competitor_Pricing) * (1 - Discount/100.0) * Units_Sold) * 100.0) /
        NULLIF(SUM(Price * Units_Sold), 0), 2) AS Profit_Margin_Percent
FROM retail
GROUP BY Category
ORDER BY Profit_Margin_Percent ASC;
```

3.2 Key Takeaways

- Only **Toys** category had a positive net profit.
- **Clothing**, **Groceries**, and **Electronics** showed significant negative profit margins, indicating potential overstock, mispricing, or poor demand.

---

4. Inventory Efficiency & Profitability Analysis (Python)

4.1 Feature Engineering

Calculated key business metrics:

- **Profit per unit** = Adjusted for price, discount, and competitor pricing
- **Total Profit** = Units Sold × Profit per unit
- **Inventory Days** = Inventory Level ÷ Average Daily Sales (assumed 30 days)

```
df['Profit_per_unit'] = (df['Price'] - df['Competitor_Pricing']) * (1 - df['Discount'] / 100)
df['Total_Profit'] = df['Units_Sold'] * df['Profit_per_unit']
df['Inventory_Days'] = df['Inventory_Level'] / df['Units_Sold'].replace(0, 1)
```

4.2 Correlation Analysis

- Checked correlation between Inventory Days and Total Profit.
- Result: Very weak correlation (r = **0.001**) → Inventory duration is not a strong predictor of profit.

4.3 Identifying Inefficiencies

- **Slow Movers:** Products with high inventory days but negative profits → Need clearance or removal.
- **Overstocked Items:** High inventory with very low units sold → Indicate storage cost issues or poor demand.

---

5. Tableau Dashboard Summary

Dashboard Panels:

- **Region-Wise Profitability:** North performs worst; South performs near breakeven.
- **Profit by Category:** Only Toys are profitable.
- **Sales Trend by Season:** Winter shows lowest sales.
- **Inventory Days vs Profit:** Scatter plot confirms no strong correlation.
- **Top Overstocked Products:** Product IDs like P0017 and P0014 have excessive inventory and need review.

---

6. Key Insights

- ✅ **Toys** are the only profitable category.
- ❌ **Clothing** category suffers major losses and needs urgent attention.
- 📉 **North region** incurs the highest losses, while the **South region** is almost breakeven.
- 🔍 **Inventory Days** does not significantly influence profit—profitability likely driven by pricing, seasonality, or category.
- 📦 **Overstocked items** should be liquidated or promoted to reduce holding costs.

---

7. Strategic Recommendations

- 🧹 **Clear slow-moving and overstocked inventory** using targeted discounts and promotions.
- 📦 **Revisit procurement and pricing strategies** especially in loss-making categories like Clothing and Groceries.
- 🌍 **Realign regional strategies**, with focused efforts to improve profitability in the North region.
- ❄️ **Seasonal Promotions in Winter** to boost demand during low-performing months.
