Retail Sales Data Analysis Project

# 1. Introduction

This project involves analyzing retail sales data to extract insights about product performance and category trends. The dataset was created manually with 30 rows of transaction records, including information such as sale date, product, category, quantity, and price per unit.

# 2. Tools Used

• Python (Pandas, Matplotlib)  
• Microsoft Excel (for manual data entry)  
• Google Colab / Jupyter Notebook (for code execution)  
• CSV file format (for data storage)

# 3. Dataset Description

The dataset includes the following columns:  
• Sale\_ID – Unique identifier for each transaction  
• Date – Date of sale  
• Product – Item sold  
• Category – Type of product  
• Quantity – Number of units sold  
• Price\_per\_Unit – Cost per unit of the product  
• Total\_Sale – Calculated field (Quantity x Price\_per\_Unit)

# 4. Code Summary

The Python code performs the following steps:  
• Reads the data using Pandas  
• Creates a new column 'Total\_Sale' for revenue analysis  
• Groups data by Product and Category to identify top performers  
• Visualizes the top 5 products using a bar chart

# 5. Key Insights

• Identified top-selling products based on total revenue  
• Determined which product categories contributed most to sales  
• Gained hands-on experience with data analysis and visualization using Python

# 6. Conclusion

This mini-project helped build foundational skills in data cleaning, manipulation, and visualization using Python. It serves as a portfolio piece for showcasing practical data analysis capabilities.