Supermarket Sales Data Analysis Project Documentation

# 1. Project Overview

This project focuses on analyzing a supermarket's sales data to uncover trends, track performance, and support business decisions. The process included:  
- Data cleaning using Python  
- Data transformation using Power Query in Excel  
- Data extraction and querying using SQL  
- Interactive dashboard creation in Power BI  
  
The final product is a report-ready dashboard offering deep insights into product sales, customer behavior, and branch performance.

# 2. Project Objectives

- Clean and prepare raw sales data for analysis  
- Structure and enrich data using Power Query  
- Extract insights using SQL queries  
- Visualize results with interactive dashboards in Power BI  
- Recommend actions to improve business outcomes

# 3. Tools & Technologies

Python (Pandas, NumPy) – Data cleaning and preprocessing  
Excel Power Query – Data transformation and shaping  
SQL (MySQL / SQLite) – Data extraction and querying  
Power BI – Dashboard development and visualization  
Microsoft Excel – Exploratory data analysis

# 4. Data Cleaning (Python)

The raw dataset included missing values, inconsistent formats, and duplicate entries. Python (with Pandas) was used to clean the data.  
  
Key Steps:  
- Removed duplicates  
- Converted columns to correct data types (e.g., date, price)  
- Handled null values  
- Created calculated fields (e.g., Total = Quantity × Unit\_Price)  
  
A black and white screen

AI-generated content may be incorrect.

# 5. Data Transformation (Power Query)

Power Query in Excel was used to further transform and shape the data.  
  
Steps:  
- Merged and appended datasets  
- Renamed and reordered columns  
- Created new calculated columns  
- Grouped and summarized data as needed

# 6. Data Extraction (SQL)

SQL queries were used to explore and summarize the data from a structured database.  
  
Example Query:  
*SELECT TOP 10 PRODUCTCODE, SUM(SALES) AS total\_sales FROM invoice\_fact\_table\_sql GROUP BY PRODUCTCODE ORDER BY total\_sales DESC;*A screenshot of a computer

AI-generated content may be incorrect.

# 7. Power BI Dashboard

Power BI was used to build an interactive and visual dashboard.  
  
Key Visuals:  
- KPI Cards: Total Revenue, Quantity Sold, Profit  
- Bar Charts: Top 5 Products   
- Line Graph: Sales trend over time  
- Pie Chart: Customer payment distribution  
- Filters: Date range, product category, location A screenshot of a computer screen

AI-generated content may be incorrect.A screenshot of a graph

AI-generated content may be incorrect.

A screenshot of a graph

AI-generated content may be incorrect.

# 8. Data Model

The Power BI data model includes the following relationships:  
invoice\_fact\_table(PRODUCTCODE FK)🡪 product\_dimension(PRODUCTCODE PK)

invoice\_fact\_table (DATE\_ID FK) 🡪 date\_dimension(DATE\_ID PK)

invoice\_fact\_table (COUNTRY FK) 🡪 country\_dimension (COUNTRY PK)

invoice\_fact\_table (CUSTOMERNAME FK) 🡪 customer\_dimension (CUSTOMERNAME PK)

A screenshot of a computer

AI-generated content may be incorrect.

# 9. Insights & Findings

- Product classic cars and motorcycles had the highest sales revenue  
- USA outperformed others in both quantity and revenue  
- Most purchases occurred during forth quarter  
- trains had the largest profit margin

# 10. Challenges Faced

- Incomplete and messy raw data  
- Learning curve with Power BI and Power Query  
- Matching and merging data from different sources  
- Maintaining consistent formatting across tools

# 11. Conclusion

This project demonstrated the full data analysis pipeline from raw data to actionable business insights. It highlights how tools like Python, SQL, Excel, and Power BI can work together to support data-driven decision-making in the retail industry.