

Project Summary – Retail Sales Analysis

Objective

The objective of this project was to analyze retail sales data to uncover customer behavior, product performance, and store profitability. By applying data cleaning, exploratory analysis, SQL queries, and advanced visualization, the goal was to answer key business questions, improve decision-making, and demonstrate end-to-end data analytics capabilities.

Tools Used

- **Excel** → Initial data inspection, identifying errors, duplicates, and missing values.
 - **PyCharm (Jupyter Notebook)** → Data cleaning, preprocessing, and exploratory data analysis (EDA) using **Python (Pandas, NumPy, Matplotlib, Seaborn)**.
 - **MS SQL Server** → Stored cleaned data, created ER diagram with **Primary & Foreign Keys**, solved 10+ business questions, and derived key metrics.
 - **Power BI** → Created interactive dashboards and visualizations, including **RFM-based customer segmentation** and profitability reports.
-

Steps Taken

1. **Data Inspection:**
 - Checked raw CSV datasets for missing values, incorrect entries, and datatype mismatches.
 - Identified duplicates and inconsistencies across customers, sales, returns, and store data.
2. **Data Cleaning (Python):**
 - **Store ID Handling:** Assigned "0" for online sales (missing store IDs) to maintain consistent datatypes and allow proper relationships.
 - **Customer Dataset:** Filled missing ages with **median values** (statistical method) instead of dropping rows, ensuring better customer segmentation.
 - **Returns Dataset:** Cleaned null values and standardized dates.

- **Products Dataset:** Derived new columns (sold_units, return_quantity, return_rate, profit) for deeper product performance analysis.
- Exported cleaned datasets into new CSV files (*_cleaned1.csv).

3. SQL Database Integration:

- Faced challenges in importing CSVs directly; solved by modifying datatypes during load.
- Created **tables with PK & FK constraints** to establish relationships.
- Built **ER diagram** for structured data modeling.
- Wrote SQL queries to answer business questions (e.g., top-selling products, customer lifetime value, return analysis, store profitability).

4. Exploratory Data Analysis (EDA):

- Analyzed sales trends, return behavior, and customer demographics.
- Applied **statistical summaries, groupings, and visual checks** to validate data.

5. Power BI Dashboards:

- Built interactive reports covering **sales performance, product returns, store profitability, and customer segmentation**.
- Used **RFM (Recency, Frequency, Monetary, Return Rate)** model for customer segmentation.
- Added KPIs for revenue, profit, and return performance to support decision-making.

Key Outcomes

- Cleaned and structured multiple retail datasets into a relational database.
- Derived meaningful metrics (return_rate, profit, age_group) to simplify business analysis.
- Solved 10+ business problems using SQL queries.
- Developed interactive Power BI dashboards to visualize customer behavior, product performance, and store profitability.
- Demonstrated an end-to-end data analytics pipeline: **raw data → cleaning → database modeling → SQL queries → visualization → insights**.

File structure:

Retail_Sales_Capstone/

```
├── data/
│   ├── raw/
│   │   ├── sales_data.csv
│   │   ├── customers.csv
│   │   ├── products.csv
│   │   ├── returns.csv
│   │   └── stores.csv
│   └── cleaned/
│       ├── sales_cleaned1.csv
│       ├── customers_cleaned1.csv
│       ├── products_cleaned1.csv
│       ├── returns_cleaned1.csv
│       └── stores_cleaned1.csv
├── notebooks/
│   └── data_cleaning_eda.ipynb
├── sql/
│   ├── business_questions.sql
│   └── ER_diagram.png
├── dashboard/
│   └── Retail_Sales_Report.pbix
├── docs/
│   ├── final_report.pdf
│   └── data_cleaning_steps.pdf
```

 **Author:** VENKAT BONASU

 **Contact:** venkatbonasu1@gmail.com