SUPERSTORE DATA CLEANING AND EDA

MySQL, Power BI

Abstract

-Exploratory Data Analysis and Report generation for a fictional Global Super Store Dataset.

SRI RAM SANKAR T S

tssriramsankar@gmail.com

https://github.com/sriramsankarts/superstore-data-cleaning-EDA

📌 Executive Summary

The **Superstore Sales Analysis Project** was undertaken to derive actionable insights from a fictional retail dataset containing 24 columns spanning across sales, customer, shipping, and order-related attributes. Using **MySQL Workbench**, the raw CSV data was cleaned, structured, and queried to perform exploratory data analysis (EDA). Key metrics like profit, sales trends, shipping modes, and regional performance were extracted through complex SQL queries. These insights were then visualized using **Power BI**, where interactive dashboards were built with filters, slicers, heatmaps, and time-based trend charts. The objective was to uncover areas of high profitability, identify underperforming categories, and improve business decision-making through data storytelling. The analysis concluded that regions like **South and West** generated strong profits, **Standard Class shipping** was the most efficient, and categories like **Tables** contributed to low-profit margins, thus offering scope for strategic improvement.

🗂️ Project Overview

This project focuses on analyzing sales performance using a fictional Superstore dataset. The data includes order details, customer info, shipping, sales, profit, and discounts. It was first cleaned and structured in MySQL Workbench using SQL queries. Key business metrics were derived through exploratory data analysis. Finally, interactive dashboards were built in Power BI to uncover trends, evaluate performance across regions, and support data-driven decision-making.

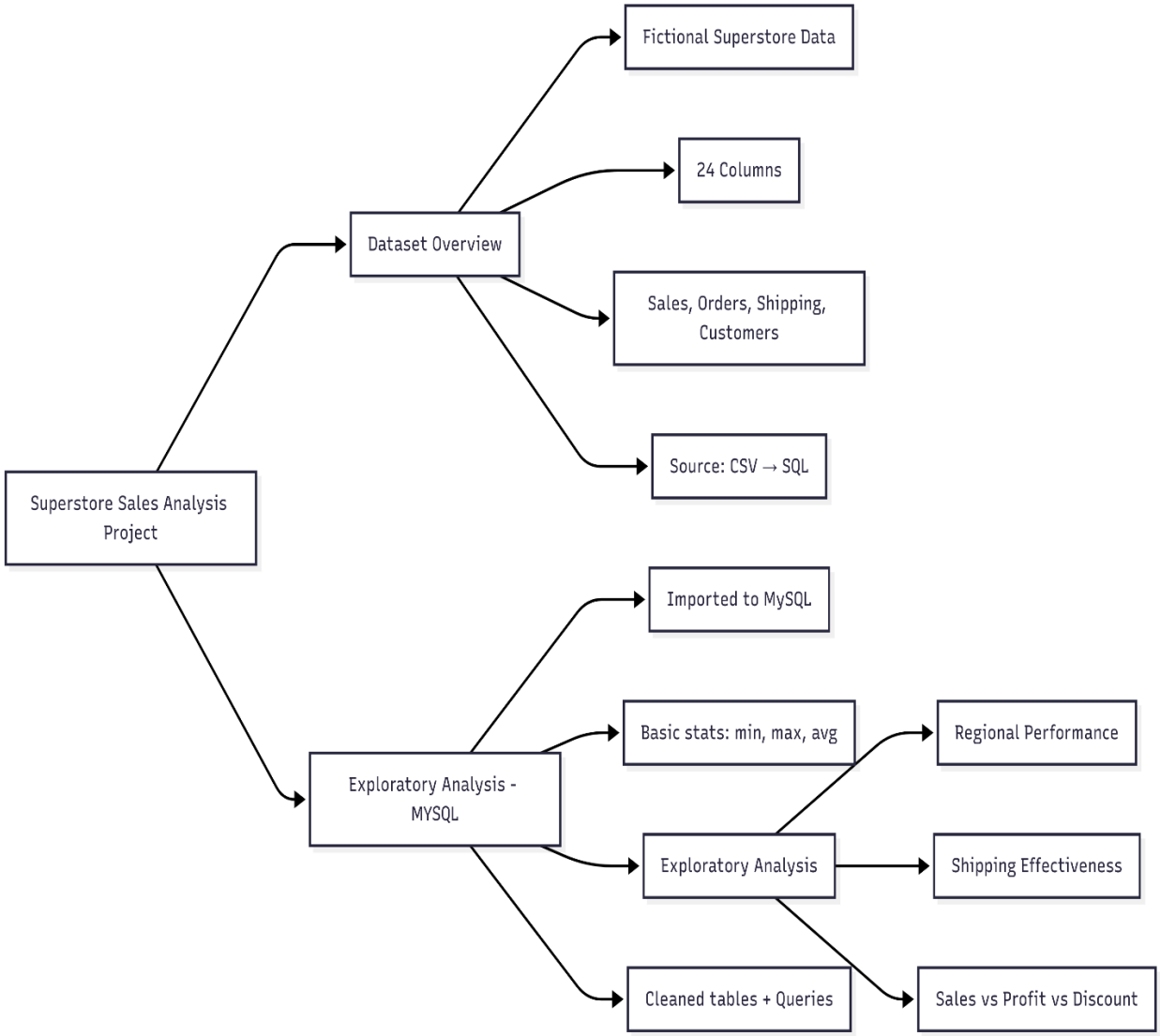
🛠️ Tools Used

- **MySQL Workbench** – For data import, cleaning, and SQL-based exploratory analysis
- **Power BI** – For building interactive dashboards and visual storytelling
- **Microsoft Excel** – For preliminary data checks and exports

---

🔁 Project Workflow

The below mind map illustrates the complete workflow of the Superstore Sales Analysis project, covering dataset overview, data engineering, visual analytics, and final insights.



🔁 Workflow Continues.....

🧹 Data Cleaning & Preprocessing

To ensure a reliable foundation for analysis, the dataset was imported into MySQL Workbench and meticulously cleaned. Unnecessary columns like Row_ID were dropped, and column names were standardized for readability. Date fields were converted into appropriate DATE formats, while missing or inconsistent values (like empty postal codes) were handled gracefully. Text fields were normalized to uppercase for consistency.

New feature columns such as profit_status and discount_status were created to better understand profitability and discount trends. Duplicate records were removed, and outliers like invalid discounts were identified. This process ensured a clean, analysis-ready dataset.

🔍 Exploratory Data Analysis (SQL)

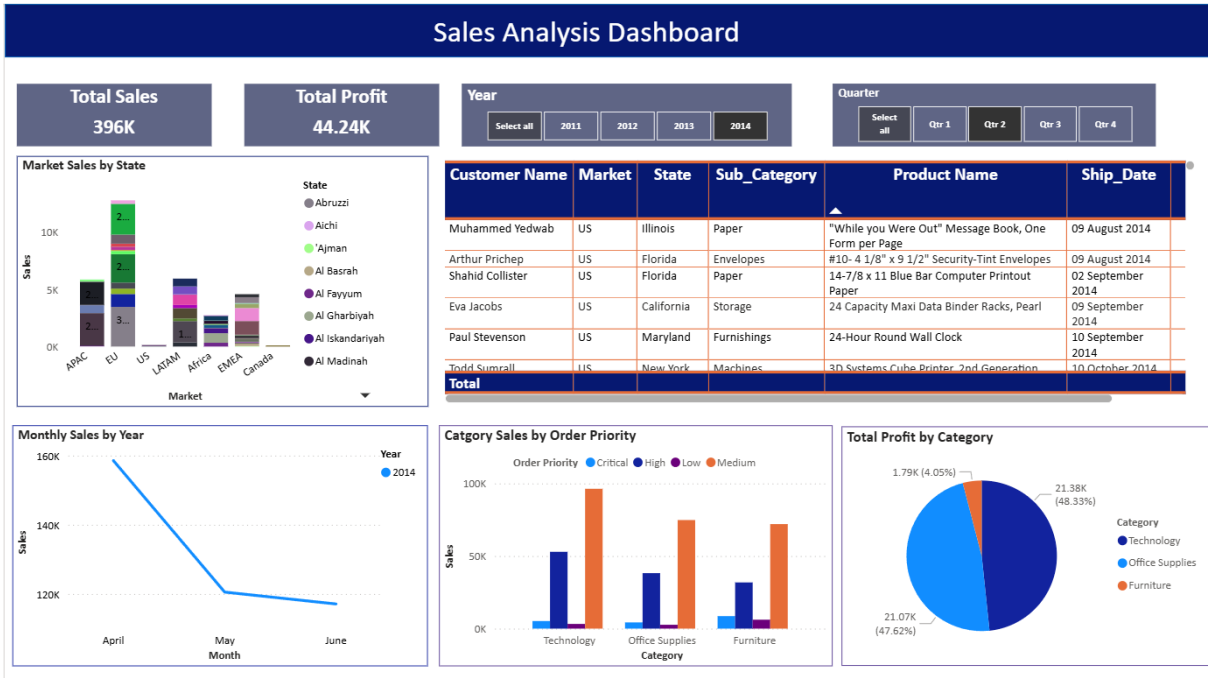
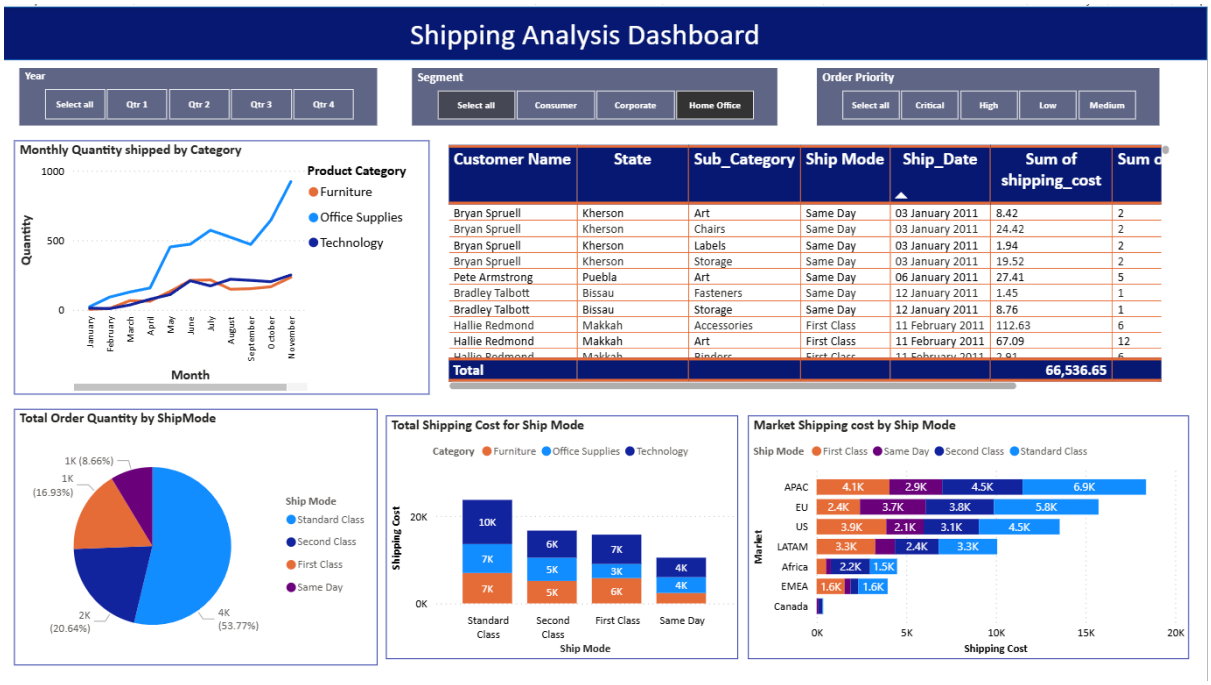
Extensive SQL queries were used to extract descriptive statistics like average, min, max, and standard deviation for key metrics (Sales, Profit, Quantity, Shipping Cost). EDA also included:

1. Regional trends (sales & profit by country/region)
2. Segment-wise customer analysis
3. Shipping mode efficiency
4. Top-selling products
5. Time-series trends (monthly sales/profit)
6. Delivery time patterns
7. Discount impact on profitability

These insights laid the groundwork for building interactive dashboards and business recommendations in Power BI.

📈 Dashboard Snippets

1. Sales Analysis Dashboard



2. Shipping Analysis Dashboard