End-to-End Sales Analytics Project

From Messy Data to KPI-Ready Revenue – Excel, SQL & Power BI

Project Overview

This project demonstrates a real-world end-to-end sales analytics workflow, starting from messy data and ending with KPI-ready dashboards.

Key Objectives:

- Handle messy datasets with duplicates, missing values, negative or zero quantities, extreme outliers, and high discounts
- Apply statistical validation (median, IQR, outlier capping)
- Create Excel and Power BI dashboards for management insights

Problem Statement

In real companies, raw sales data can lead to wrong business decisions due to:

- Duplicate orders
- Missing customer information
- Negative or zero quantities
- Extreme selling price outliers
- High discount transactions

Project Goal: Demonstrate professional handling of messy data, KPI calculation, and reliable dashboarding.

Tools & Technologies

- Excel: Data audit, statistical analysis, KPI rules, dashboards
- SQL: Validation, aggregation, KPI calculations
- Power BI: Interactive KPI dashboards
- Statistics: Median, IQR, outlier capping, KPI eligibility rules

Excel Workflow & Raw Data Challenges

Excel Workflow

1. Raw Data Audit: Identifying duplicates, missing values, invalid entries
2. Statistical Outlier Detection: Using IQR, upper-limit capping for selling price
3. KPI Eligibility Rules:
    - Discount % threshold
    - Minimum net sales threshold

4. Pivot Tables: Calculating KPI metrics (Revenue, Orders, AOV)

5. Revenue Comparison: Raw vs Clean vs KPI-Eligible

Special Notes:

- Orders with missing customer IDs were retained as “Unknown Customer”, reflecting real scenarios
- Ensures revenue is not lost while highlighting incomplete customer info

Raw Data Challenges & Statistics

Customer Dataset:

- Total Customers: 500+ (unique)
- Missing Customer Details: 2 → marked as Unknown Customer

Order Details Dataset:

Total Orders: 5,264

Negative or Zero Quantity: 175 rows

Negative or Zero Selling Price: 76 rows

Outlier Selling Price (after IQR calculation): 102 rows flagged

Maximum Selling Price: 2,5000

Minimum Selling Price: -500

Maximum Quantity Sold: 50

High Discount Orders (≥70%): 719

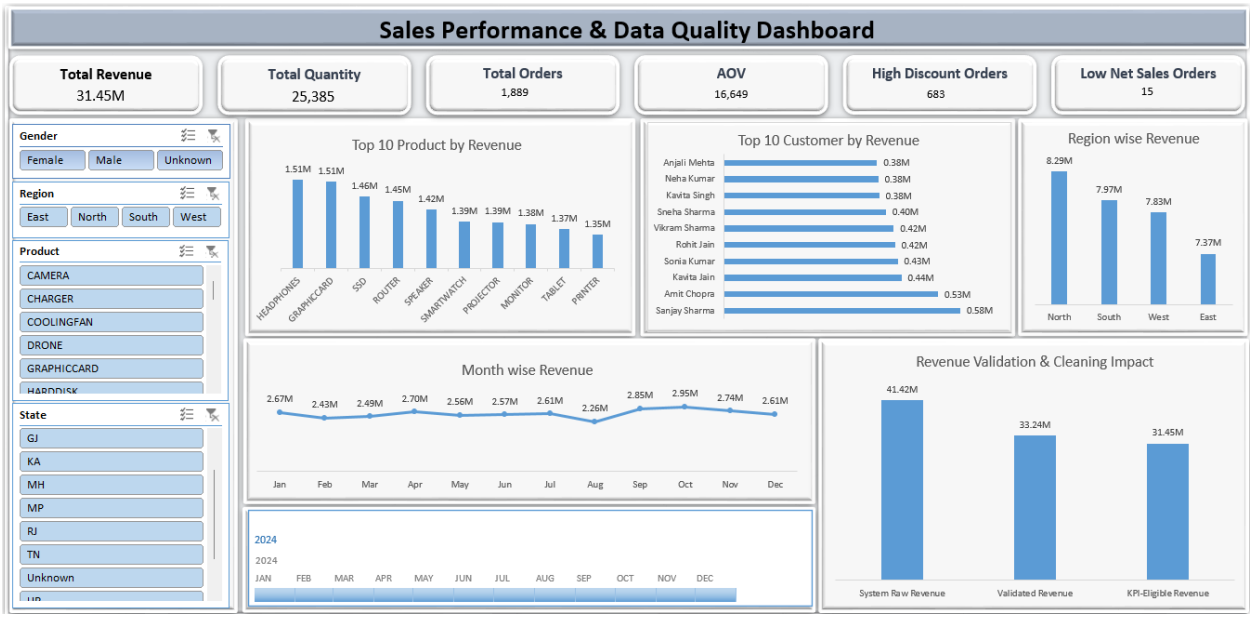
Orders Dataset:

- Duplicate Order IDs removed: 100
- Orders with missing Customer ID: 2 (retained as Unknown Customer)

Observations:

- Raw data had high variance in sales amount due to outliers
- Negative and zero values could mislead revenue reporting
- High discount orders significantly reduce net revenue if included without cleaning

Excel Dashboard Preview



SQL Workflow

SQL Tasks:

- Validate cleaned datasets
- Calculate KPI metrics (Revenue, Orders, AOV)
- Analyze by Product, Customer, Region, Month
- Cross-check Excel calculations

Examples:

- Top 10 Products by Revenue
- Top 10 Customers by Revenue
- Region-wise Revenue

Screenshot Placeholder:

	customer_name character varying (50)	customer_sales numeric
1	Sanjay Sharma	583367.30
2	Amit Chopra	528333.32
3	Kavita Jain	437904.28
4	Sonia Kumar	428709.93
5	Rohit Jain	422090.24
6	Vikram Sharma	419220.98
7	Sneha Sharma	399670.31
8	Kavita Singh	383132.44
9	Neha Kumar	380744.43
10	Anjali Mehta	377832.36

Power BI Dashboard

Built on KPI-eligible & validated data only

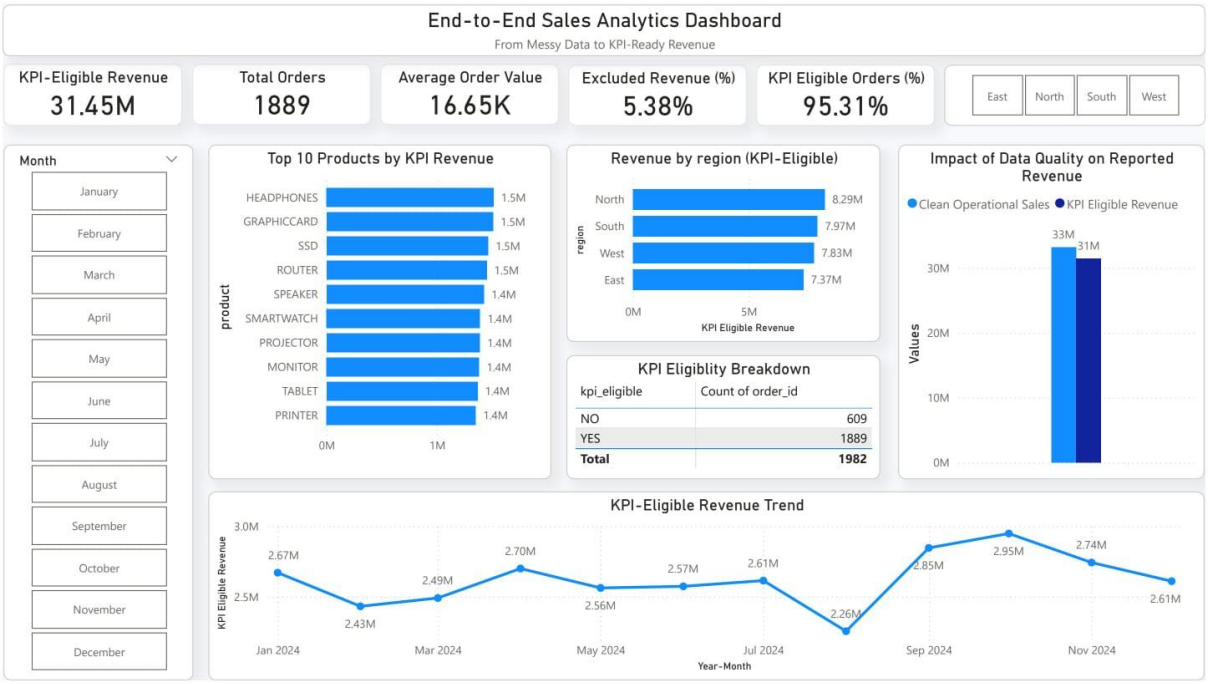
Key KPIs:

- KPI-Eligible Revenue: 31.45M
- Total Orders: 1,889
- Average Order Value (AOV): 16.65K
- Excluded Revenue (%): 5.38%
- KPI-Eligible Orders (%): 95.31%

Key Visuals:

- Top 10 Products by KPI Revenue
- Revenue by Region
- KPI-Eligible Revenue Trend (2024)
- KPI Eligibility Breakdown Table
- Data Cleaning Impact on Revenue

Screenshot Placeholder:



Business Insights & Conclusion

Business Insights:

- Over 95% of orders are KPI-eligible
- ~5% revenue excluded due to validation rules
- North region highest revenue contributor
- Headphones & Graphic Cards are top-revenue products
- October shows peak KPI-eligible revenue

About the Data:

- Datasets synthetically created to mimic real-world company data
- Intentional challenges: duplicates, missing values, negative/zero quantities, extreme outliers, high discounts
- "Unknown Customer" orders retained to reflect real business scenarios

Conclusion:

- Excel demonstrates data quality thinking & statistical validation
- SQL ensures logic accuracy & aggregation
- Power BI delivers clean, KPI-ready interactive dashboards

About the Author

Name: Md Tanbir Rja

Email: mdtanbirraza7@gmail.com

LinkedIn: https://www.linkedin.com/in/md-tanbir-rja-067561236

GitHub: https://github.com/tanbir-94

Note:

This project is a synthetic, self-created dataset designed to mimic real-world company data with duplicates, missing values, outliers, and discount challenges. It demonstrates end-to-end data cleaning, validation, and KPI reporting workflow using Excel, SQL, and Power BI.