E-Commerce Sales Performance Analysis Report

(UK-Based Online Retail Dataset | Dec 2009 – Dec 2010)

1. Introduction (What & Why)

This project was completed as part of a real-world data analytics internship task, where the objective was to work with raw e-commerce sales data and convert it into meaningful business insights.

Instead of directly creating charts, the focus was first on understanding the dataset, identifying data quality issues, and deciding what data is actually relevant for sales analysis. The final outcome of this task is an interactive dashboard that helps business stakeholders quickly understand sales performance, demand patterns, and key revenue drivers.

2. Understanding the Dataset (Initial Observations)

The dataset contains online retail transaction data from a UK-based e-commerce business, covering the period from December 2009 to December 2010.

During the initial exploration, I observed:

- Multiple non-sales records such as adjustments, bad debt entries, bank charges, and manual corrections.
- Negative and zero quantities, which indicate returns, cancellations, or internal adjustments.
- Several high-value invoices with missing customer IDs.
- The year 2009 contains only December, while 2010 contains full monthly data.

These observations directly influenced how I approached data cleaning and analysis.

3. Data Cleaning & Preprocessing (How & Why)

This was the most important phase of the task.

3.1 Removal of Non-Sales Transactions

Rows with descriptions such as:

- Adjustment
- Adjust bad debt
- Amazon fees
- Bank charges
- Manual entries

were removed.

Why?

These records do not represent actual customer purchases and would distort revenue, quantity, and product performance analysis.

3.2 Handling Negative and Zero Quantity

All rows with:

- Quantity = 0
- Quantity < 0

were removed.

Reason:

Negative quantities usually represent returns or corrections, not real demand. Since the task focuses on sales performance, keeping these rows would misrepresent actual customer behavior.

3.3 Date & Time Processing

The original datetime column was split into:

- Invoice Date
- Invoice Time
- Hour (derived from time)

This enabled:

- Monthly trend analysis
- Hourly sales and quantity analysis

3.4 Revenue Calculation

A new column Revenue was created:

- Revenue = Quantity × Unit Price

This became the core metric for most analyses.

3.5 Customer ID Handling

Some invoices with blank customer IDs were retained.

Why?
Even without customer identification, these transactions still:

- Generated revenue
- Contributed to demand trends

Removing them would understate total sales.

4. Exploratory Data Analysis (EDA)

After cleaning, I performed EDA to understand patterns and outliers.

Key EDA Findings:

- The dataset effectively represents full-year behavior only for 2010.
- Monthly trends were analyzed across available data, with awareness of partial data in 2009.
- United Kingdom dominates in:
    - Number of invoices
    - Quantity sold
    - Revenue generated

5. Key Business Insights

5.1 Product Performance

- Highest revenue-generating product: *Regency Cake Stand 3 Tier*
- Highest quantity sold product: *White Hanging Heart T-Light Holder*

This shows that revenue leaders and volume leaders are not always the same, which is important for pricing and inventory strategy.

5.2 Time-Based Insights

- November recorded the highest revenue and quantity sold.
- 12 PM is the peak hour for revenue.
- 1 PM is the peak hour for quantity sold.

This suggests strong midday purchasing behavior, useful for promotions and staffing decisions.

5.3 Invoice-Level Insight

Some invoices with a missing customer ID generated unusually high revenue, indicating bulk or wholesale purchasing behavior.

5.4 Country Performance

Only the top 5 countries were analyzed to keep insights focused.

- United Kingdom is the best-performing market by a large margin.

6. Dashboard Design (What I Chose & Why)

The dashboard was designed to be:

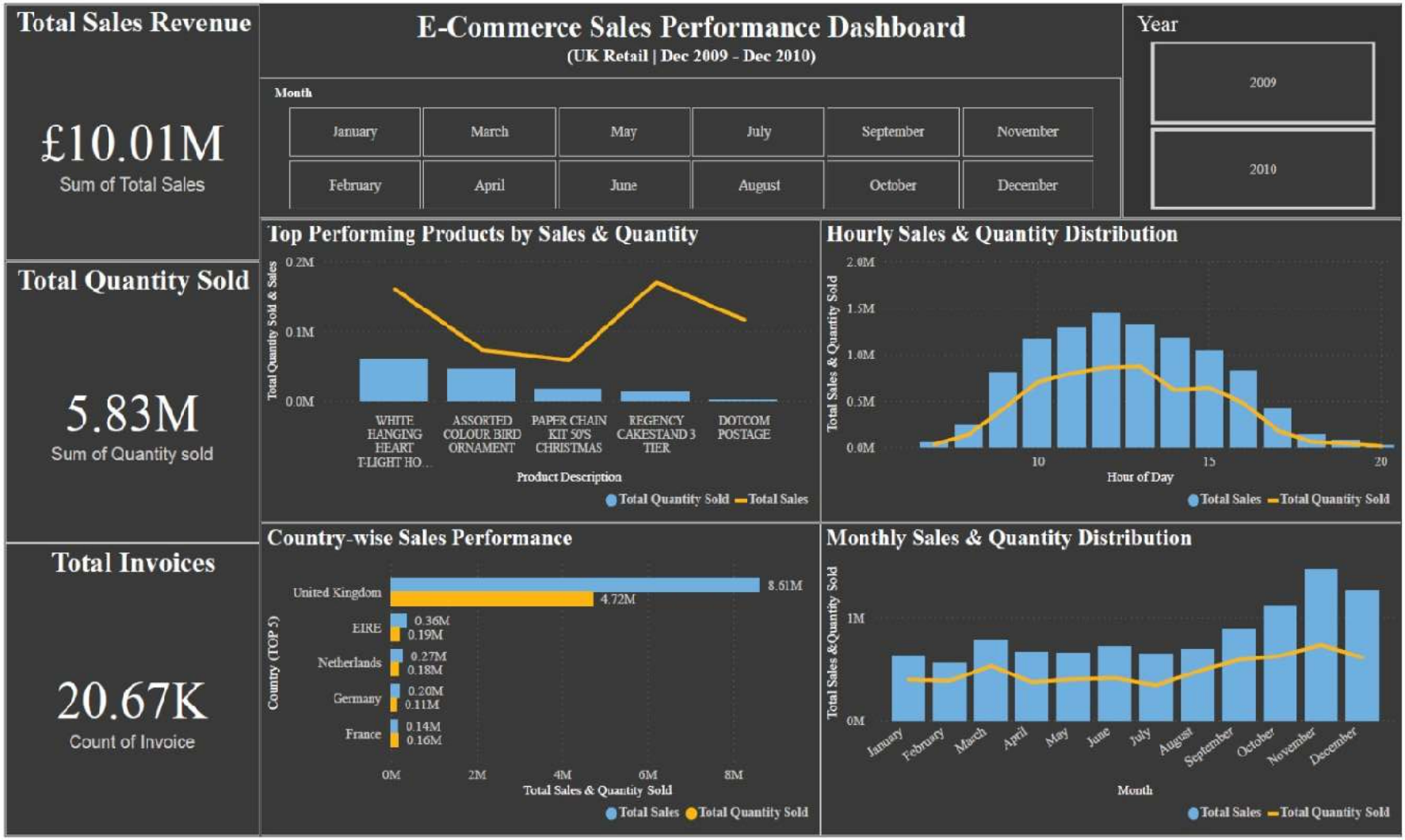
- Simple
- Non-technical user friendly
- Insight-focused

Instead of KPIs, Cards were used to clearly show:

- Total Sales Revenue
- Total Quantity Sold
- Total Invoices

Visuals include:

- Monthly Sales & Quantity Distribution (include both the year)
- Hourly Sales vs Quantity Distribution
- Top Performing Products by Sales & Quantity
- Country-wise Performance (Top 5)


Total Sales Revenue
£10.01M
Sum of Total Sales
E-Commerce Sales Performance Dashboard
(UK Retail | Dec 2009 - Dec 2010)
Month
January
March
May
July
September
November
February
April
June
August
October
December
Year
2009
2010
Total Quantity Sold
5.83M
Sum of Quantity sold
Top Performing Products by Sales & Quantity
Total Quantity Sold & Sales
0.2M
0.1M
0.0M
WHITE HANGING HEART T-LIGHT HO...
ASSORTED COLOUR BIRD ORNAMENT
PAPER CHAIN KIT 50'S CHRISTMAS
REGENCY CAKESTAND 3 TIER
DOTCOM POSTAGE
Product Description
Total Quantity Sold
Total Sales
Hourly Sales & Quantity Distribution
Total Sales & Quantity Sold
2.0M
1.5M
1.0M
0.5M
0.0M
10
15
20
Hour of Day
Total Sales
Total Quantity Sold
Total Invoices
20.67K
Count of Invoice
Country-wise Sales Performance
Country (TOP 5)
United Kingdom
8.61M
4.72M
EIRE
0.36M
0.19M
Netherlands
0.27M
0.18M
Germany
0.20M
0.11M
France
0.14M
0.16M
0M
2M
4M
6M
8M
Total Sales & Quantity Sold
Total Sales
Total Quantity Sold
Monthly Sales & Quantity Distribution
Total Sales &Quantity Sold
1M
0M
January
February
March
April
May
June
July
August
September
October
November
December
Month
Total Sales
Total Quantity Sold

7. Conclusion

This project demonstrates the end-to-end analytics workflow:

- Understanding messy real-world data
- Cleaning with business logic
- Performing meaningful EDA
- Translating insights into a dashboard

The task strengthened my ability to think like a business-focused data analyst, not just a tool user.