5 July 2025
Niranjan Kumar

UK Retail Analytics

Data Cleaning & Exploration Report

Table of Contents

01

INTRODUCTION

UK Retail Analytics – Data Understanding and Cleaning Report
The *UK Retail Analytics* project uses the **Online Retail dataset** from the *UCI Machine Learning Repository*. This dataset contains detailed transaction records from a UK-based online retail company, spanning **over 540,000 rows** from **December 2010 to December 2011**.

The data captures each sale or return made by customers across different countries, and provides rich information for understanding **sales performance**, **product trends**, and **customer behavior** — making it an ideal real-world dataset to practice data cleaning, exploration, and business-driven analysis.

Key Initial Details:

- **Total Rows:** ~542,000 transactions.
- **Initial Columns (8 main fields):**

Column Name	**Description**
InvoiceNo	Unique identifier for each transaction. Alphanumeric; prefix 'C' indicates a cancelled invoice.
StockCode	Unique product or item code; includes numbers, text, or special codes (e.g., POST for postage).
Description	Text description of the product sold. May contain blanks or junk text.
Quantity	Number of units purchased (positive) or returned (negative).
InvoiceDate	Date and time when the transaction occurred.
UnitPrice	Price per unit of the product (decimal).
CustomerID	Unique ID to identify the custome. May contain missing values.
Country	County of the custome, used for geographical analysis

The raw dataset is **typical of real-world e-commerce data** — it includes missing values, inconsistent text, and mixed data types, which makes it ideal for demonstrating practical **data cleaning** and **wrangling**.

02

PROJECT OVERVIEW

The **UK Retail Analytics** project aims to transform messy, raw retail data into clear, actionable insights for business decision-making. Using over **540,000 real-world transactions**, this project covers the entire data workflow: **data understanding, data cleaning, data exploration, and final dashboarding.**

Key Business Goals:

1. **Identify top-selling and underperforming products** to optimize stock levels and pricing strategies.
2. **Detect patterns in cancelled or returned invoices** to minimize revenue loss.
3. **Analyze customer buying behavior** — typical quantities, price points, and peak times — for better marketing and promotions.
4. **Segment customers** by frequency and value to plan loyalty programs or targeted campaigns.
5. **Evaluate sales by country** to find new expansion opportunities and grow market share outside the UK.
6. **Deliver a clean, interactive Power BI dashboard** to present insights clearly to both technical and non-technical stakeholders.

Through this project, you will demonstrate practical data analytics skills — from handling messy data to generating business-focused insights — and deliver a visually compelling Power BI dashboard that supports real-world decision-making.

03

DATA UNDERSTANDING

The dataset used in this project is the **UCI Online Retail Dataset**, a popular benchmark for practicing real-world retail analytics. It contains detailed information about **online transactions** made by customers between **December 2010 and December 2011**.

Total Records:
~ **542,000 rows** of transactional data.

Initial Columns
Below are the original columns with a brief description for each:

Column Name	Description
InvoiceNo	Unique ID for each transaction. Some invoices have a "C" prefix indicating cancellations.
StockCode	Unique code assigned to each product sold.
Description	Textual description of the produc
Qunatity	Number of product units sold (can be negative for returns).
InvoiceDate	Date and time when the transaction occurred.
UnitPrice	Price per unit of the product (in GBP).
CustomerID	Unique numeric ID for the customer placing the order (some records have missing IDs).
Country	Country where the customer is based.

Key Characterstics:

1. Mixed Data types: Numeric IDs, text descriptions, dates, and prices.
2. Contains missing values and inconsistencies, such as:
    - Null **Description** or **CustomerID**
    - Cancellations and returns (negative **Quantity**)
    - Potential duplicates and inconsistent product codes.

This understanding forms the **baseline** for designing targeted cleaning steps, ensuring that only reliable, meaningful information feeds into the analysis and dashboard.

04

DATA EXPLORATION

In this phase, the dataset was thoroughly explored to understand its structure, detect data quality issues, and gather insights that shaped the cleaning plan. Below is a **column-wise summary** of **all** data exploration steps performed:

4.1 InvoiceNo

- **Pattern Check:**
  - Contains both purely numeric and alphanumeric values.
  - Invoices with a prefix **'C'** indicate cancelled transactions.
- **Null Check:**
  - No missing values found.
- **Duplicates**
  - Valid to have multiple rows with the same InvoiceNo since a single invoice can contain multiple products.

Action: Extracted first character to create an Is_Cancalled flag for further analysis of cancellation

4.2. StockCode

- **Patten Check**
  - Contains purely numeric codes for regular products.
  - Some special codes include:
    - Single letters → Possible special items/services.
    - 'POST' → Postage fees.
    - 'Gift_xxx_xx' → Gift cards or promotional items.
- **Null Check:**
  - No missing values found.
- **Special Pattern:**
  - Identified patterns to categorize StockCodes into Product, Variant, Postage, or Gift for better analysis.

Action: Split StockCode, extracted token, calculated code length and created a Stock_Code_Type classification.

4.3. Description

- **Pattern Check:**
  Some rows had blanks, true nulls, or placeholder junk text (like ?, ??, missing).
- **Nulls:**
  Found ~5 true nulls and some blanks/empty strings.
- **Junk Text:**
  Quickly scanned for obvious placeholder values.

Action :

- Trimmed leading/trailing spaces using Format > Trim.
- Replaced blanks and placeholder junk like ? or missing with **null**.
- Did **not** recreate the DescriptionLength helper this time.
- Directly replaced remaining nulls with **"Unknown"** (did **not** fallback to StockCode in this version).
- Applied **Proper Case** formatting for consistency.

4.4 Quantity

- **Pattern Check:**
  No nulls. Min: -24 (valid returns), Max: 216, Average: ~11.52.
- **SpecialValue:**
  Negative = valid returns, zeros = none found.
- **Distribution:**
  Common bulk pack sizes confirmed valid.

Action : Kept negative values as valid → enforced **whole number** type → created **Is_Return** column (Yes for negative, No for positive) to simplify return analysis.

4.5 InvoiceDate

- **Pattern Check:**
  No nulls. Range: 01-Dec-2010 → 09-Dec-2011. Time part available for peak-hour insights.

Action: Split into **Invoice_Date** (date) and **Invoice_Time** (time) → ensured correct data types → flagged for duplicate check on combined fields.

4.6 UnitPrice

- **Pattern Check:**
  No nulls. Min: £0.72, Max: £20.79, cluster at £0.83–£0.85. No zeros or negatives.
- **Distribution:**
  Mostly low-value items — normal for retail.

Action: Confirmed as **decimal** → rounded to **2 decimal places** → rechecked for invalid values post-rounding.

4.7 CustomerID

- **Pattern Check:**
  Found **47 nulls**. Range: 12576–18248. Mostly one-time buyers.
- **No Suspicious values:**
  All numeric, no gaps.

Action : Replaced nulls with **99999** placeholder → confirmed type as **whole number** → verified no blanks remained.

4.8 Country

- **Pattern Check**
  Found 231 nulls. UK dominates. No casing or spelling issues.
- **Frequency:**
  UK is 90%+ of data.

Action :
Replaced nulls with "Unknown" → trimmed spaces → final consistency check for blanks.

05

DATA CLEANING

In this phase, the raw dataset was cleaned and standardized to ensure reliable and consistent analysis. Each column was reviewed, transformed, and validated according to the issues found during data exploration.

Below is a **column-wise summary** of all data cleaning actions performed:

1. **InvoiceNo**
    - Some values were purely numeric while others were alphanumeric (prefix **'C'** means cancelled invoices).
    - Created new column **Is_Cancelled** → flagged rows starting with **'C'** as **Yes**, all others as **No**.
    - Changed **Invoice_No** data type to **Text** for consistency.
    - Changed **Is_Cancelled** data type to **Text**.
    - Renamed **InvoiceNo** to **Invoice_No**.
2. **Stock_Code**
    - Original data contained mixed numeric + alphabetic codes → converted **Stock_Code** to **Text** for consistency.
    - Applied **UPPERCASE** transformation to standardize similar codes (e.g., 19001a → 19001A).
    - Trimmed any extra spaces to remove hidden duplicates.
    - Created **Stock_Code_Type** column:
      Gift → if code starts with Gift
      Service/Postage → if code equals POST
      Product → all standard product codes.
    - Grouped values to check distribution: **Product: 539,889**, **Gift: 34**, **Service: 1,986**.
3. **Description**
    - Removed placeholder text entries: **?**, **??**, **???**, **?missing**, **?? missing**, **???missing**, etc
    - Applied **UPPERCASE** to unify inconsistent text.
    - Trimmed leading/trailing spaces to clean up empty prefixes/suffixes.
    - Replaced all junk placeholders with **null**.
    - Filled remaining nulls with **UNKNOWN PRODUCT** for completeness.
    - Finally, formatted text using **Capitalize Each Word** for a readable, consistent style.
4. **Quantity**
    - Verified no invalid values → negative quantities valid for returns.
    - Confirmed **Quantity** is whole number type.
    - Created **Is_Return** column → flagged Yes for negative values, No for positive sales.
    - Flag distribution: Yes → 531,285 | No → 10,624.
5. **Invoice_Date & Invoice_Time**
    - Confirmed **InvoiceDate** is **DateTime** type.
    - Split into **Invoice_Date** (date only) and **Invoice_Time** (time only) for time-based analysis.
    - Removed duplicates based on combined **Invoice_Date + Invoice_Time** to ensure unique transactions.

6. **Unit_Price**
    - Verified **Unit_Price** is decimal number type.
    - Confirmed no blanks, zero, or negative prices exist.
    - Rounded all prices to **2 decimal places**.
    - Double-checked for negative/zero prices post-rounding.
    - Reviewed max price (20.79) → no unrealistic spikes.
    - Validated typical prices clustered between **0.83 – 0.85**, confirming consistent pricing.
7. **CustomerID**
    - Found **47 missing values** → replaced with placeholder **99999** to keep records traceable.
    - Ensured **CustomerID** is whole number for joins and calculations.
    - Final check → all missing IDs correctly labeled as **99999**.
    - No unexpected nulls remain.
8. **Country**
    - Found **231 missing values** → replaced with **Unknown** for transparency.
    - Verified all blanks replaced, no stray nulls left.
    - Trimmed accidental spaces to keep country names consistent.

06

KEY DECISIONS & ASSUMPTIONS

This section highlights critical decisions made during data preparation and the assumptions taken to ensure meaningful, reliable analysis.

1. **Treating Cancelled Invoices**
   - Decision: Created **Is_Cancelled** flag based on invoice numbers starting with **'C'**.
   - Assumption: All invoices with a 'C' prefix represent legitimate cancellations, following the dataset's known pattern.
2. **Stock Code Categorization**
   - Decision: Classified **Stock_Code** into **Product**, **Gift**, and **Service/Postage**.
   - Assumption: Codes with Gift or POST truly represent gift items or postage fees. Single-character codes were reviewed and treated as variants only if clearly separate from standard products.
3. **Handling Missing Descriptions**
   - Decision: Filled blank **Description** fields with UNKNOWN PRODUCT or the **Stock_Code** if available.
   - Assumption: A missing description does not mean a missing product — the Stock_Code alone is sufficient to identify the item.
4. **Negative Quantities as Returns**
   - Decision: Negative **Quantity** values flagged as **Returns** via **Is_Return**.
   - Assumption: Negative values correctly indicate product returns/cancellations and are not data entry errors.
5. **Missing Customer IDs**
   - Decision: Replaced blank **CustomerID** with 99999 to retain transaction records.
   - Assumption: It's better to preserve transactions with unknown customers than to drop potentially valuable data.
6. **Country Field Nulls**
   - Decision: Missing **Country** entries replaced with Unknown.
   - Assumption: This ensures no records are lost, and missing geographic data can be separately analyzed if needed.
7. **Rounding Unit Prices**
   - Decision: Rounded **Unit_Price** to **2 decimal places** for standard currency formatting.
   - Assumption: Minor rounding does not materially impact sales figures given the low price range of most products.

07

DAX MEASURES & CALCULATIONS

This section lists all key DAX measures created for analyzing the UK Retail dataset. Each measure is linked directly to the business goals and enables the final dashboard to deliver clear, actionable insights.

1. Sales & Revenue Measures

```
Total Revenue =
CALCULATE(
    SUMX(
        'Online Retail',
        'Online Retail'[Quantity] * 'Online Retail'[UnitPrice]
    ),
    'Online Retail'[Quantity] > 0
)

Total Sales =
SUMX(
    'Online Retail',
    'Online Retail'[Quantity] * 'Online Retail'[UnitPrice]
)

Net Revenue =
CALCULATE(
    SUMX(
        'Online Retail',
        'Online Retail'[Quantity] * 'Online Retail'[UnitPrice]
    ),
    'Online Retail'[Is_Cancalled] = "No",
    'Online Retail'[Is_Return] = "No"
)

Returned Revenue =
CALCULATE(
    SUMX(
        'Online Retail',
        'Online Retail'[Quantity] * 'Online Retail'[UnitPrice]
    ),
    'Online Retail'[Is_Return] = "Yes"
)
```

```
Revenue UK =
CALCULATE(
    [Total Revenue],
    'Online Retail'[Country] = "United Kingdom"
)

Revenue Non UK =
CALCULATE(
    [Total Revenue],
    'Online Retail'[Country] <> "United Kingdom"
)

Sales by Country =
SUMX(
    'Online Retail',
    'Online Retail'[Quantity] * 'Online Retail'[UnitPrice]
)
```

2. Order & Quantity Measures

```
Total Orders = DISTINCTCOUNT('Online Retail'[Invoice_No])

Total Cancelled Orders =
CALCULATE(
    COUNTROWS('Online Retail'),
    'Online Retail'[Is_Cancalled] = "Yes"
)

Cancelled Invoices =
CALCULATE(
    DISTINCTCOUNT('Online Retail'[Invoice_No]),
    'Online Retail'[Is_Cancalled] = "Yes"
)

Cancelled Orders % = DIVIDE([Total Cancelled Orders], [Total Orders], 0)

Total Returned Orders =
CALCULATE(
    COUNTROWS('Online Retail'),
    'Online Retail'[Is_Return] = "Yes"
)

Returned Orders % = DIVIDE([Total Returned Orders], [Total Orders], 0)
```

```
Total Units Sold = SUM('Online Retail'[Quantity])

Total Returns Units =
CALCULATE(
    SUM('Online Retail'[Quantity]),
    'Online Retail'[Is_Return] = "Yes"
)

Average Order Qty = AVERAGE('Online Retail'[Quantity])

Average Order Value = DIVIDE([Total Sales], [Total Orders])

Average Unit Price = AVERAGE('Online Retail'[UnitPrice])
```

3. Customer Measures

```
Total Customers = DISTINCTCOUNT('Online Retail'[CustomerID])

Revenue per Customer = DIVIDE([Total Revenue], [Total Customers], 0)

Customer Total Revenue =
CALCULATE(
    SUMX(
        'Online Retail',
        'Online Retail'[Quantity] * 'Online Retail'[UnitPrice]
    ),
    FILTER(
        'Online Retail',
        'Online Retail'[CustomerID] = EARLIER('Online Retail'[CustomerID])
    )
)

Customer Order Count =
CALCULATE(
    DISTINCTCOUNT('Online Retail'[Invoice_No]),
    ALLEXCEPT('Online Retail', 'Online Retail'[CustomerID])
)

Customer Segment =
SWITCH(
    TRUE(),
    'Online Retail'[Customer Total Revenue] >= 1000, "High Value",
    'Online Retail'[Customer Total Revenue] >= 500, "Mid Value",
    'Online Retail'[Customer Total Revenue] >= 100, "Low Value",
    "Very Low Value"
```

```
)

High Value Customers =
CALCULATE(
   DISTINCTCOUNT('Online Retail'[CustomerID]),
   'Online Retail'[Customer Segment] = "High Value"
)

Mid Value Customers =
CALCULATE(
   DISTINCTCOUNT('Online Retail'[CustomerID]),
   'Online Retail'[Customer Segment] = "Mid Value"
)

Low Value Customers =
CALCULATE(
   DISTINCTCOUNT('Online Retail'[CustomerID]),
   'Online Retail'[Customer Segment] = "Low Value"
)
```

4. Products & Miscellaneous

```
Distinct Products Sold = DISTINCTCOUNT('Online Retail'[StockCode])

Top Product Revenue =
CALCULATE(
   MAXX(
      VALUES('Online Retail'[Description]),
      CALCULATE([Total Revenue])
   )
)

Top Product Units =
CALCULATE(
   MAXX(
      VALUES('Online Retail'[Description]),
      CALCULATE([Total Units Sold])
   )
)

Countries Active = DISTINCTCOUNT('Online Retail'[Country])

Total Rows = COUNTROWS('Online Retail')

Total Return Value =
```

```
CALCULATE(
    SUMX(
        'Online Retail',
        'Online Retail'[Quantity] * 'Online Retail'[UnitPrice]
    ),
    'Online Retail'[Is_Return] = "Yes"
)
```

08

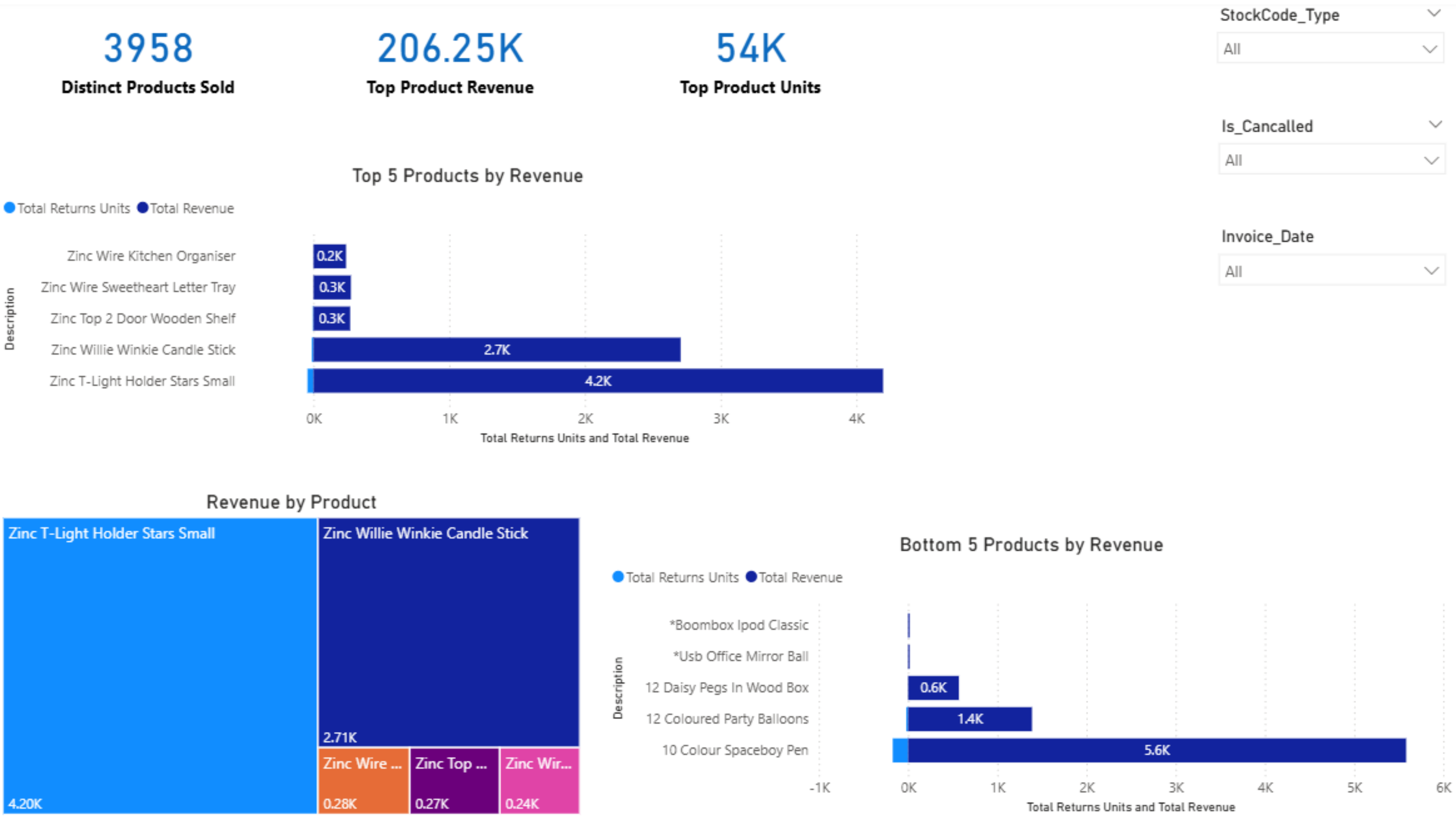
DATA ANALYSIS

1. **<u>Sales Overview</u>**

<u>Observations :-</u>

- **Seasonal Spike:**
  Revenue is stable for most months but rises sharply from **September to November**, peaking in **November (~1.9M)** — indicates **holiday shopping surge** (typical for retail).
- **Lowest Revenue Month:**
  April is the lowest revenue month (~0.52M) — possible seasonal dip or business cycle impact.
- **Consistency Mid-Year Plateau:**
  May to August shows stable revenue (~0.72M–0.77M) — steady sales during mid-year.
- **Early Year Moderate:**
  January starts at ~0.69M — healthy post-holiday sales carryover.
- **December Drop:**
  After peaking in November, **December revenue dips slightly to ~1.45M** — likely due to early shopping before Christmas, or late orders invoiced in January
- **Overall Upward Trends:**
  Despite fluctuations, there's clear upward momentum from mid-year to year-end — good sign of sales ramp-up.

2. Product Performance



Observation :-

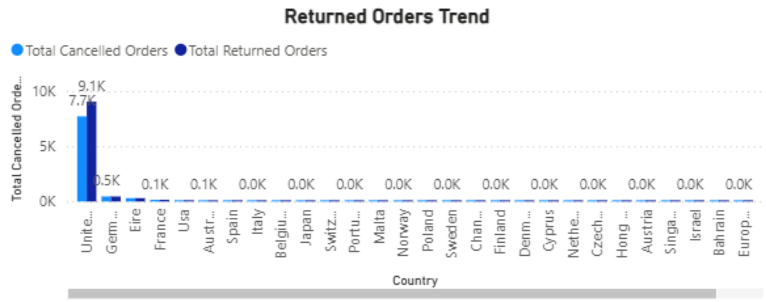
- **Wide Product Range:**
  About **3,958 distinct products** sold — shows a broad, diversified catalog.
- **Top Revenue Products:**
  Zinc T-Light Holder Stars Small is the highest earner (~4.2K) — a strong single-product performer.
- **Second Strong Seller:**
  Zinc Willie Winkie Candle Stick follows with solid revenue (~2.7K) — supports main top-seller, likely same category.
- **Top Seller Category Focous:**
  Top products are mostly **home decor or small gift items** — clear indication of strong demand in this niche.
- **Return Present in Top Products:**
  Even top sellers show returns — e.g., ~0.2–0.3K returns for individual best sellers — normal but worth monitoring for quality or packaging improvements
- **Bottom Performers Exists:**
  Bottom products like *Boombox Ipod Classic*, *USB Office Mirror Ball*, *12 Daisy Pegs In Wood Box* generate very low revenue (~0–1.4K) — contribute little to total sales.
- **Negative Returns for some SKUs:**
  A few low sellers have **negative net units or revenue** — more returns than sales — these are clear loss-making products.
- **Revenue Concentration:**
  Revenue is **top-heavy** — a small group of products drives the majority of total product revenue.

- **Long Tail Products:**
  Large number of products have minimal sales — classic long-tail scenario — may need pruning or better promotion.
- **Portfolio Opportunity:**
  Focus on scaling up mid-tier products that show demand but haven't reached top seller status yet.
- **Clear Quantity/ Porfitability Watch:**
  Monitor repeat return patterns on popular SKUs — balancing high volume with low return rates protects profit.

3. Returns & Cancellation

9139	0.35	10K	0.40	-881.63K
Total Cancelled Orders	Cancelled Orders %	Total Returned Orders	Returned Orders %	Returned Revenue

Invoice_Date
All



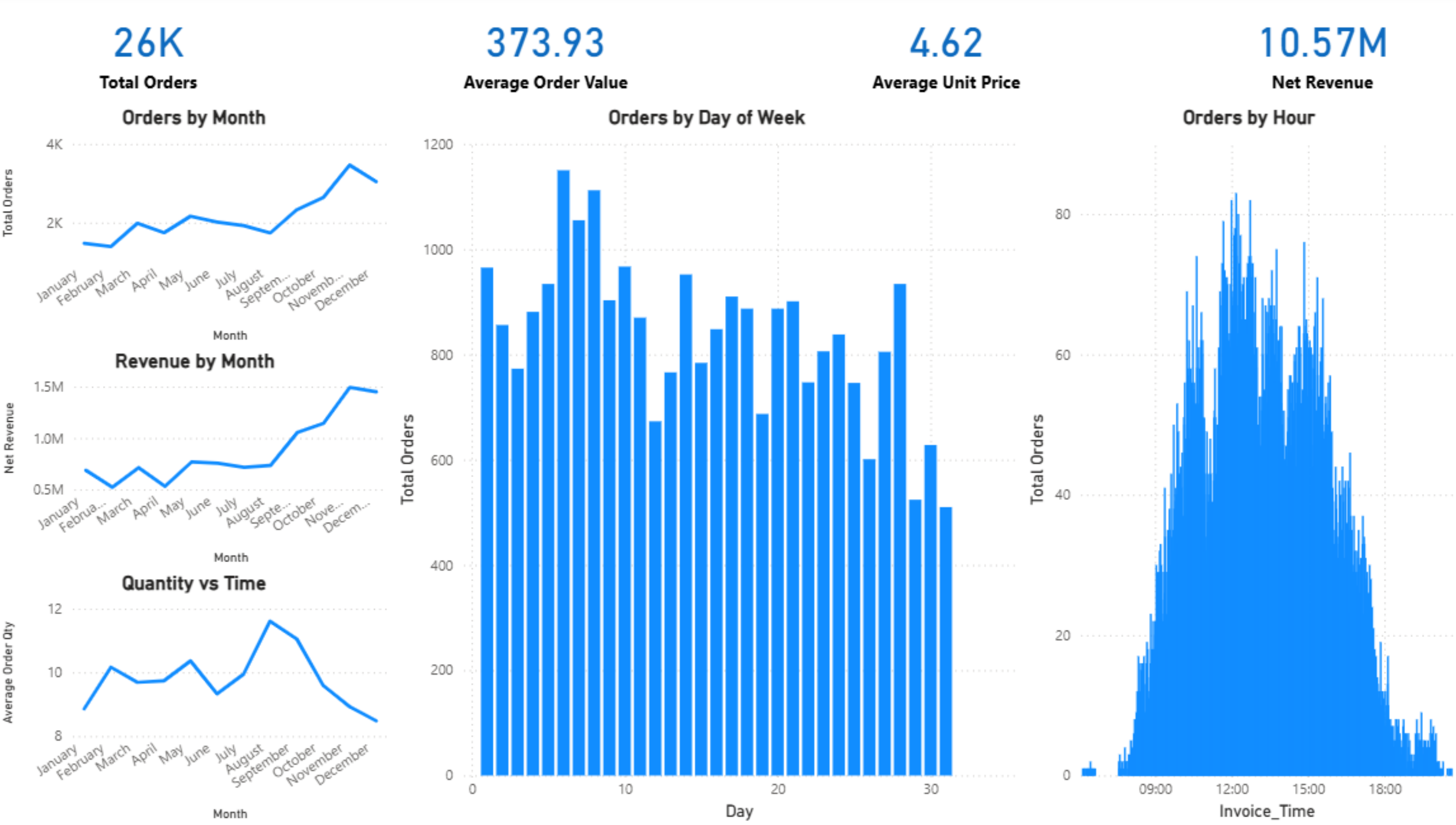
Description	StockCode	Total Cancelled Orders	Total Returned Orders
Manual	M	223	223
Regency Cakestand 3 Tier	22423	180	180
Postage	POST	124	124
Jam Making Set With Jars	22960	87	87
Set Of 3 Cake Tins Pantry Design	22720	73	73
Discount	D	65	65
Roses Regency Teacup And Saucer	22699	54	54
Strawberry Ceramic Trinket Box	21232	52	52
Recipe Box Pantry Yellow Design	22666	47	47
Total		**9139**	**10475**

Observation:

- **Low Cancellation Rate Overall**:
  Cancelled Orders represent only **0.35%** of total orders — indicates **strong customer satisfaction** and **efficient order fulfillment**.
- **Low Return Rate Overall**:
  Returned Orders make up **0.40%** of total orders — a positive indicator of **product quality** and **customer alignment**.
- **Peak Returns and Cancellations**:
- Both cancellation and return volumes **spike sharply from September to November**, peaking in **October** (225 cancelled, 242 returned) — likely due to **holiday season ordering volume**, which stresses logistics and increases errors.
- **Revenue Impact Significant**:
  Total **Returned Revenue exceeds £881K**, which is a substantial **profit drain** despite low return rates — suggests that **high-value items** are being returned more often.

- **UK Market Drives Returns**:
  UK accounts for **over 9,000 returns/cancellations**, far more than any other country — understandable due to sales volume, but also suggests need for **local process optimization**.
- **High-Risk Products Identified**:
  Items like **Manual (M)**, **Cake Stand**, and **Postage (POST)** have the **highest return and cancellation counts** (100+ each) — signals need for **product review or clarification** in listing, pricing, or delivery terms.
- **Postage & Discounts Often Involved**:
  Returns and cancellations frequently include **POST (shipping)** and **D (discounts)** — may suggest **customer confusion**, unmet delivery expectations, or **refund issues**.
- **Minimal Returns Outside UK**:
  Most countries outside the UK (e.g., Japan, Austria, Norway) show **near-zero returns** — could indicate **low order volume**, better satisfaction, or **incomplete return tracking** internationally.
- **Parallel Trends in Returns and Cancellations**:
  Return and cancellation timelines follow **very similar patterns**, indicating they might be driven by **common factors** such as seasonality, stock issues, or promotional cycles.

4. Order Patterns



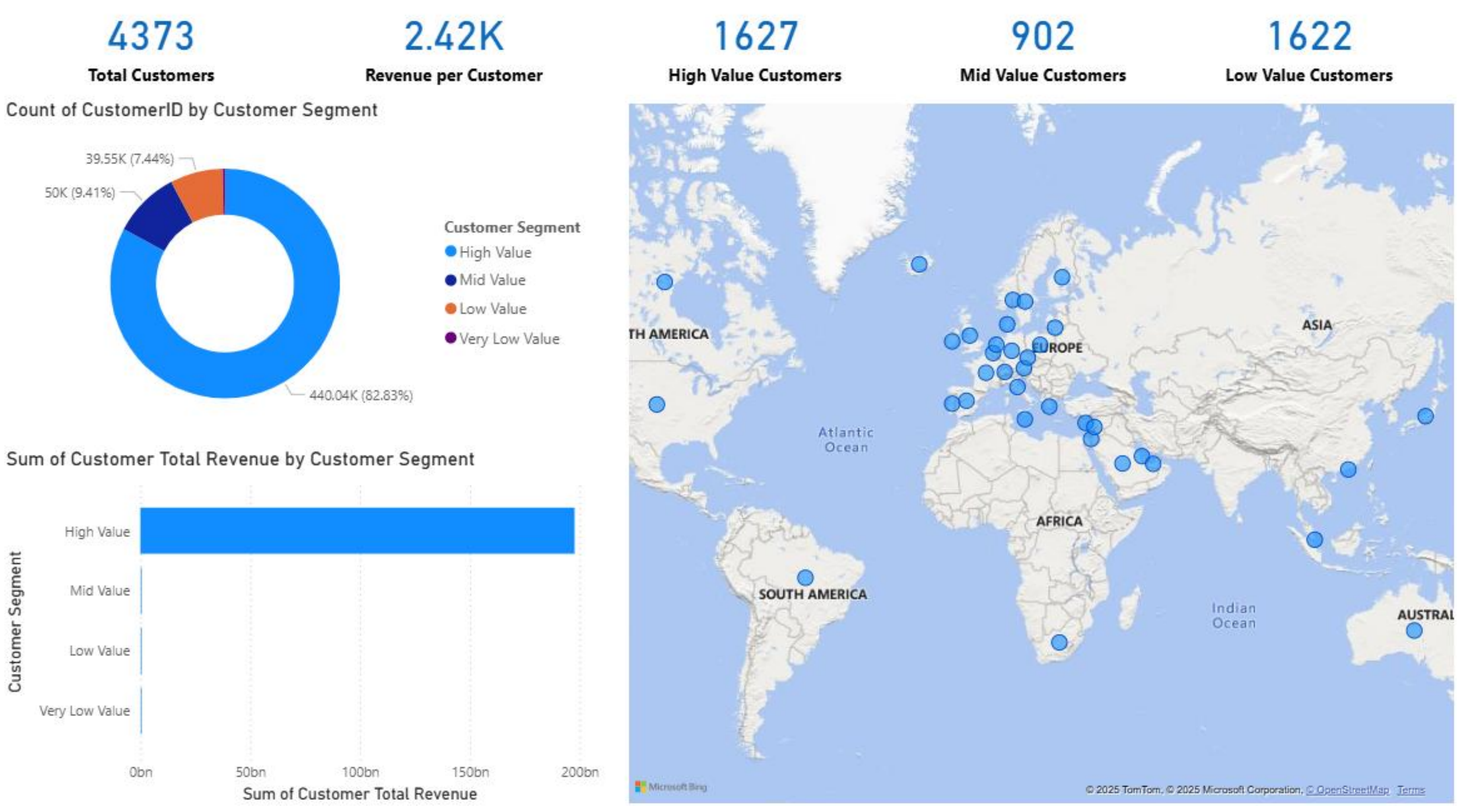
Observation:

- **Peak Month for Orders**:
  November shows the highest order volume (~3.7K orders) — likely due to **holiday season promotions** or year-end demand spike.
- **Consistent Growth Over Year**:

Total orders **increase steadily** from May to November — shows **strong mid-to-late year sales momentum**.

- **Lowest Order Month**:
  February sees the lowest number of orders (~1.5K) — possibly due to **post-holiday slowdown** or shorter month.
- **Revenue Mirrors Order Volume**:
  Revenue trends closely follow order trends, peaking in **November** (~£1.6M) and dipping in **April** — indicates **stable average order value** throughout the year.
- **Quantity Ordered Peaks Mid-Year**:
  July to September shows highest **average quantity per day** (~11–13 units) — indicates **bulk purchases** or **seasonal stock-ups**.
- **Order Distribution by Day of Month**:
  Orders are **highest in the first 10 days of each month**, especially on the **5th and 6th (~1.1–1.2K orders)** — likely due to **monthly budgets, pay cycles**, or **scheduled promotions**.
- **Lowest Order Days**:
  Last 2–3 days of each month (e.g., 30th–31st) show significantly fewer orders — potential area to explore for **end-of-month incentive campaigns**.
- **Peak Ordering Hours**:
  Most orders are placed between **10:00 AM and 3:00 PM**, with a sharp spike around **12:00 PM (noon)** — indicates customer activity during **mid-day hours**, possibly during **lunch breaks** or **business hours**.
- **Off-Hours Have Minimal Activity**:
  Very few orders occur before **8 AM** or after **6 PM** — implies most purchases are done during **standard working hours**.

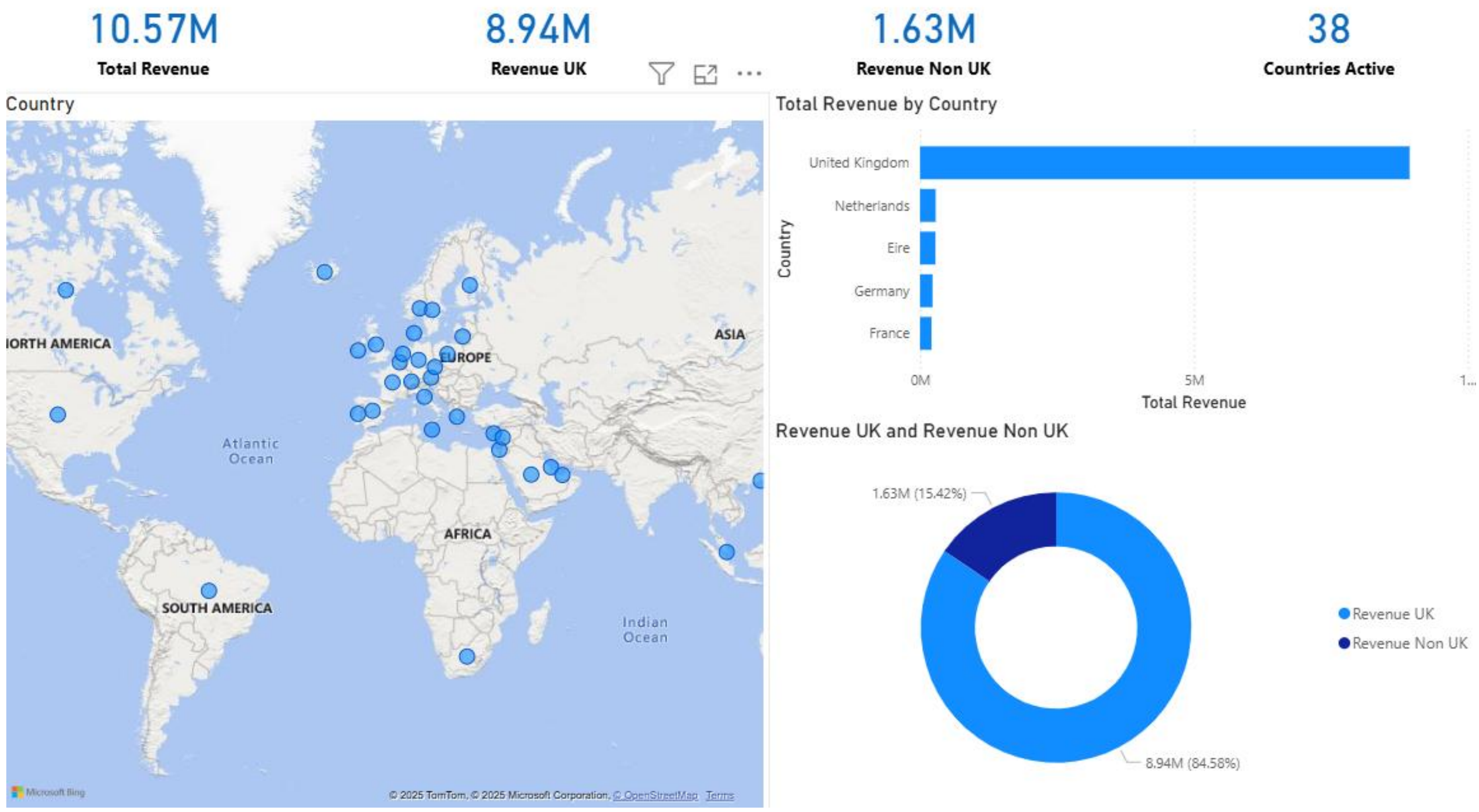
5. <u>**Customers Insights**</u>



<u>Observation:</u>

- **Majority Are High-Value Customers**:
  High Value segment makes up 82.8% of total revenue (440.04K) with **1,627 customers (37% of total)** — strong indication that a **core group of loyal, repeat buyers** is driving business success.
- **Revenue Heavily Concentrated**:
  The **High Value segment alone contributes nearly all revenue** — other segments (Mid, Low, Very Low) have **minimal contribution**, despite making up over 60% of total customers.
- **Revenue per Customer Is Healthy**:
  Average Revenue per Customer = £2.42K — suggests good **customer spend behavior**, particularly within high-value segments.
- **Low and Very Low Value Segments Are Large but Inefficient**:
  1,622 Low Value and **902 Mid Value** customers collectively generate **minimal revenue** — indicates an opportunity to:
  - Upsell or re-engage these users
  - Refine targeting strategies
  - Analyze why they aren't converting
- **Very Low Segment Tiny in Revenue**:
  Very Low Value segment contributes only **~7.44% of revenue** — potential churn risk or low engagement. Could be:
  - One-time buyers
  - Test transactions
  - Incomplete customer journeys
- **Customer Reach is Global but Clustered**:
  Customers span across **Europe, Asia, Americas, and Africa**, but the **majority are concentrated in Europe**, especially **UK, Germany, France, Netherlands, and Italy**.
- **Emerging Market Opportunities**:
  Visible presence in **Asia, Australia, and Africa**, though smaller — signals **potential growth zones** for targeted marketing or localized product offerings.

6. Geographic Sales



Observation:

- **UK Dominates Revenue**:
  The **United Kingdom alone contributes £8.94M**, which is **84.6% of total revenue** — clearly the **primary market** and growth driver.
- **Limited Global Penetration**:
  Only **15.4% of revenue (£1.63M)** comes from **non-UK markets**, even though the business operates in **38 countries** — indicates **underutilized international potential**.
- **Top Non-UK Contributors**:
  After the UK, the leading revenue sources are:
  - **Netherlands**
  - **Ireland**
  - **Germany**
  - **France**

  These contribute marginally — each generating under £0.3M.
- **Global Distribution Exists, But with Low Yield**:
  The **map shows wide international coverage** across **Europe, Middle East, Asia, and Americas**, but revenue concentration is almost entirely within UK and a few European countries — strong **reach but weak monetization** outside home market.
- **Revenue Disparity Signals Strategic Gap**:
  The significant difference between UK and non-UK revenue points to:
  - **Potential lack of localized marketing**
  - **Pricing barriers**
  - **Limited logistics support**
  - Or **low brand awareness** in foreign markets

- **Opportunities in Europe & Asia-Pacific**:
  Countries with visible sales activity but low revenue (e.g., Italy, Spain, Australia, Japan) may benefit from **targeted growth campaigns**, local promotions, or partnerships.