RETAIL REVENUE & CUSTOMER SEGMENTATION ANALYSIS

A Data-Driven Business Evaluation Using SQL & Power BI

1. Project Objective

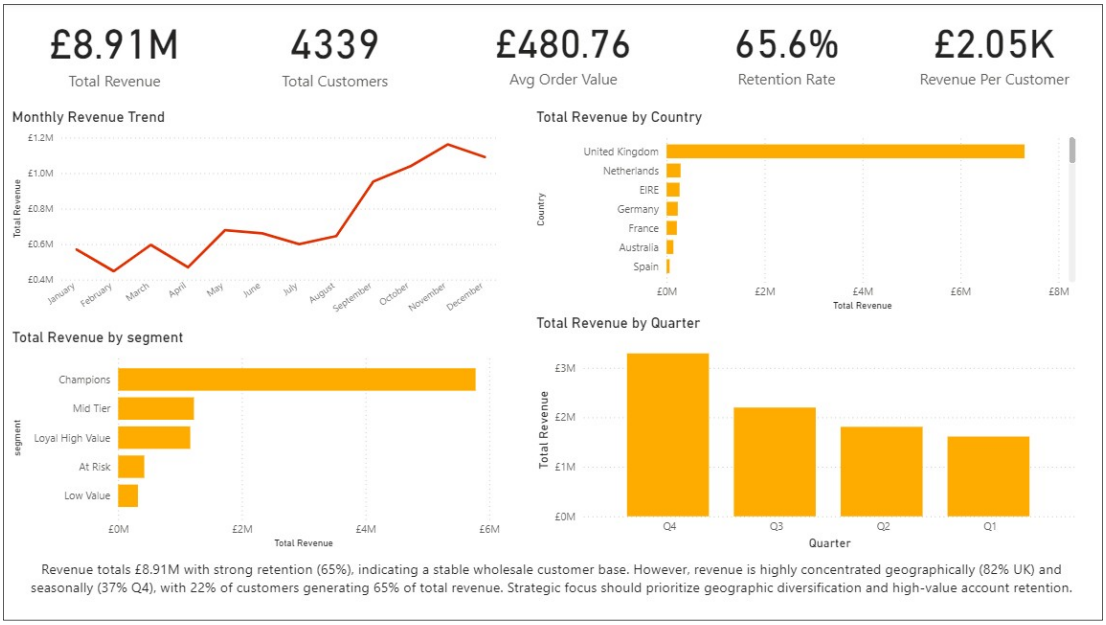
The goal of this project was to analyze a large retail transaction dataset to identify:

- Revenue patterns
- Customer behavior
- Concentration risks
- Potential strategic improvements

Dataset used: **541,909 raw retail transactions.**

After cleaning: 397,924 valid records.

PowerBI Dashboard Screenshot



Key Observations

- Total Revenue: £8.91M
- 4,339 unique customers
- 65.6% repeat purchase rate
- Average Order Value: £480
- 37% revenue generated in Q4

Initial insight suggests the business operates with strong repeat purchasing behavior but may be concentrated in specific areas.

---

2. Data Cleaning & Approach

Data Cleaning Steps

1) Removed records with missing CustomerID
2) Excluded cancelled invoices
3) Removed negative/return transactions
4) Created Revenue column (Quantity × UnitPrice)

```
23 •  SELECT *
24    FROM retail_raw
25    WHERE CustomerID IS NOT NULL
26      AND CustomerID <> ''
27      AND InvoiceNo NOT LIKE 'C%'
28      AND CAST(Quantity AS SIGNED) > 0;
29
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	12/1/2010 8:26	2.55	17850	United Kingdom
536365	71053	WHITE METAL LANTERN	6	12/1/2010 8:26	3.39	17850	United Kingdom
536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	12/1/2010 8:26	2.75	17850	United Kingdom

```
30 •  SELECT
31        InvoiceNo,
32        CustomerID,
33        CAST(Quantity AS SIGNED) AS Quantity,
34        CAST(UnitPrice AS DECIMAL(10,2)) AS UnitPrice,
35        (CAST(Quantity AS SIGNED) *
36         CAST(UnitPrice AS DECIMAL(10,2))) AS Revenue
37    FROM retail_clean;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

InvoiceNo	CustomerID	Quantity	UnitPrice	Revenue
536365	17850	6	2.55	15.30
536365	17850	6	3.39	20.34
536365	17850	8	2.75	22.00
536365	17850	6	3.39	20.34

Analytical Techniques Used

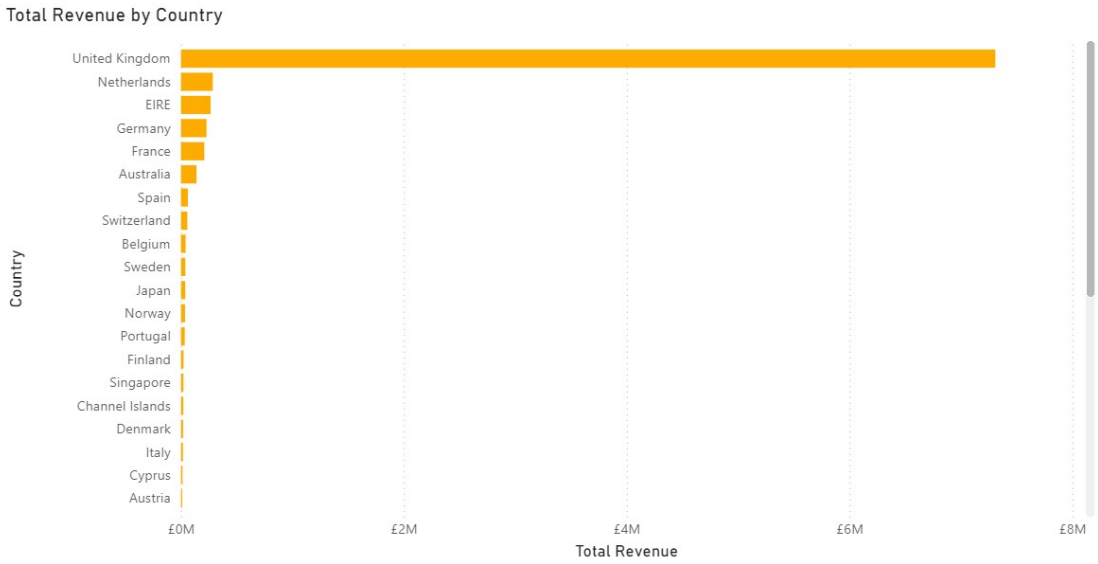
- SQL aggregation (SUM, GROUP BY, HAVING)
- Date transformations (STR_TO_DATE)
- Window functions (NTILE for RFM scoring)
- Revenue modeling
- DAX measures in Power BI

This project involved handling a large dataset and transforming raw transactional data into meaningful business insights.

---

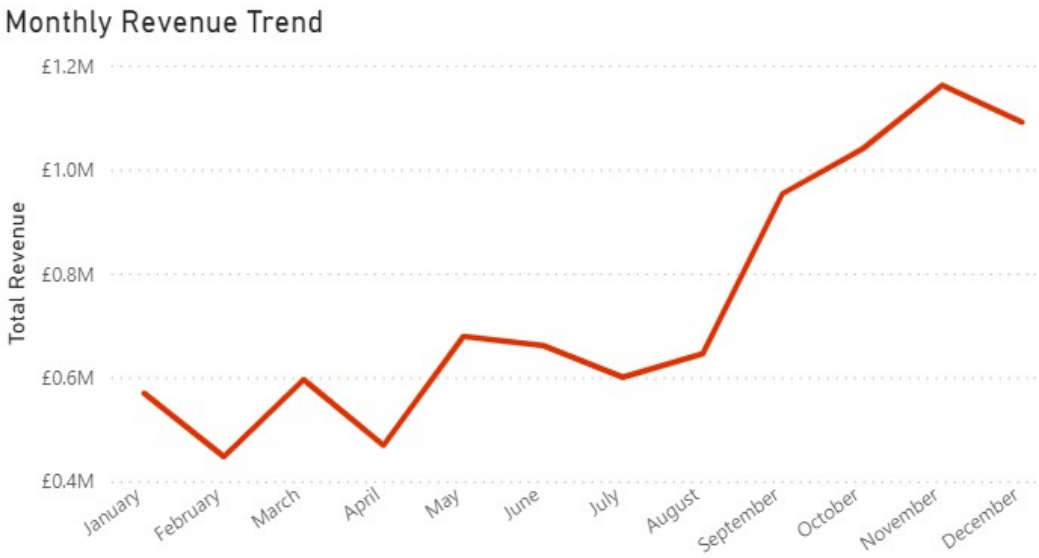
3. Revenue Analysis

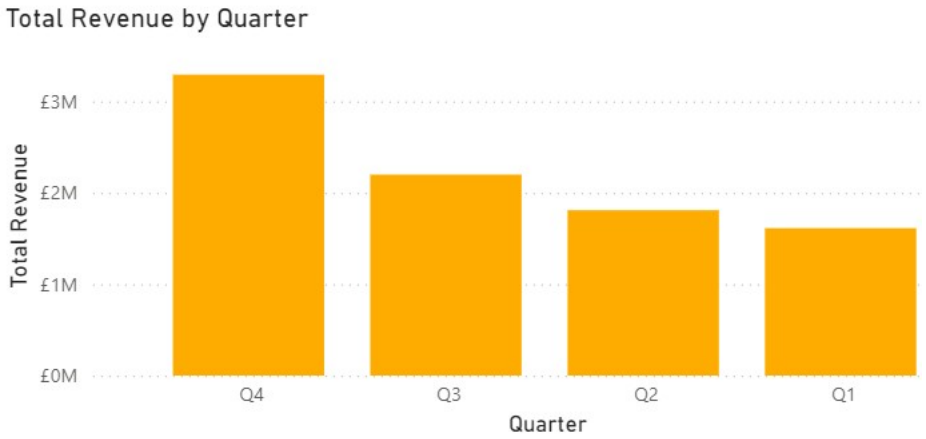
Geographic Distribution



Observation: Approximately 82% of revenue comes from the UK, indicating geographic concentration.

Seasonal Pattern

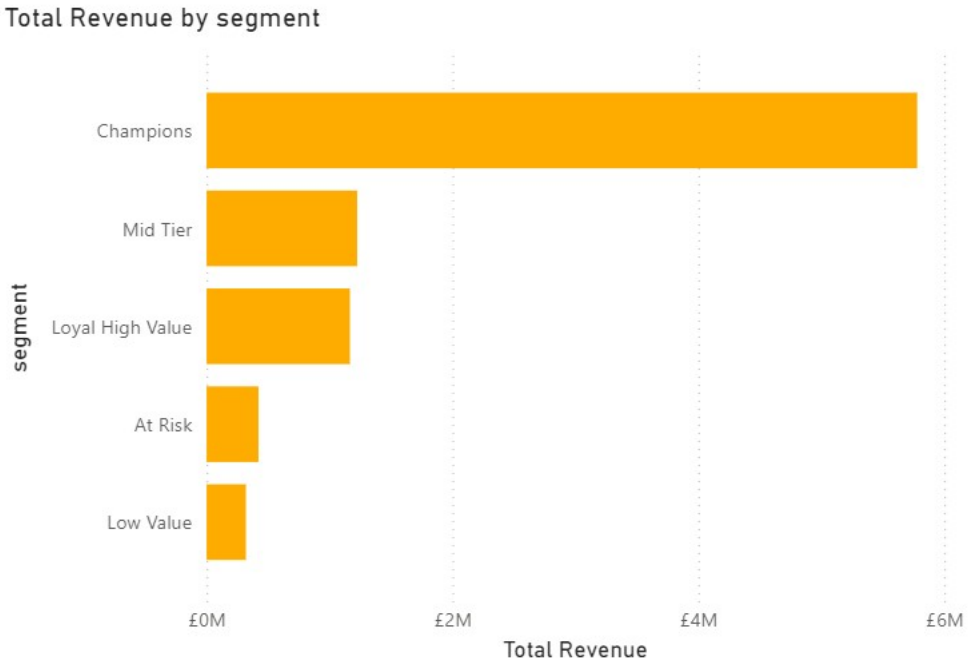




Observation: Q4 accounts for 37% of annual revenue, suggesting strong seasonality.

---

4. Customer Segmentation



Findings:

- Top 22% of customers generate around 65% of total revenue
- 65.6% repeat purchase rate suggests strong customer loyalty
- High-value customers represent a major share of total revenue

---

5. Recommendation

Based on the analysis, potential strategies could include:

- Expanding into new geographic markets to reduce dependency on the UK
- Creating retention programs for high-value customers
- Running off-season promotions to reduce revenue volatility
- Monitoring “at-risk” customers using recency data