Online Retail II Data Analysis Report

An Exploratory Analysis of Sales performance, Customers, and Cancellations

Ibrahim

Data Analyst

Project Purpose:
This report examines the Online Retail II dataset to extract actionable insights on sales trends, customer behavior, and cancellations, demonstrating analytical and visualization skills for portfolio presentation.

Tools: Python (Pandas, Matplotlib)

January 2025

Contents

1 Introduction

1.1 Project Overview

This project analyzes sales performances and customer purchasing behavior using the Online Retail II dataset, which contains two years of transaction data from a UK-based online retailer. The goal is to uncover insights into sales trends, customer behavior, and cancellations, supporting data-driven decision-making.

1.2 Dataset Summary

The Online Retail II dataset contains transactional data from a UK-based online retailer, providing a comprehensive view of sales activities over a two-year period. Below is a detailed overview:

- **Time Period:** December 2009 – December 2011

- **Scope:** This dataset records retail transactions mainly but does include some bulk transactions, capturing diverse sales scenarios from individual customers to large-scale orders.

- **Key Attributes:**

  - **Invoice:** Unique identifier for each transaction, representing a sale or a cancellation (e.g., "536365" for a sale or "C536366" for a cancellation).
  - **StockCode:** A five-character identifier for products, which may include prefixes or suffixes (e.g., "85123A").
  - **Description:** A textual description of the product (e.g., "WHITE HANGING HEART T-LIGHT HOLDER").
  - **Quantity:** Number of units per product in each transaction.
  - **UnitPrice:** Price per unit in GBP.
  - **InvoiceDate:** Date and time of the transaction.
  - **CustomerID:** A five-digit unique identifier for each customer.
  - **Country:** Geographic location of the customer.

- **Focus of Analysis:** The report aims to explore:

  - **Sales Performance:**
    * Revenue patterns, trends and growth rates across retail and bulk transactions.
    * Identify top-performing products and regions.
  - **Customer Behavior:** top-performance, retention, and segmentation.
  - **Cancellations:** Frequency and impact of transaction reversals.

- **Relevance:** This dataset is valuable for examining business performance, customer engagement, and operational challenges in an online retail context.

2 Methodology Summary

2.1 Data Pre-processing:

The dataset underwent a rigorous cleaning process to ensure accuracy and relevance for analysis:

- **Duplicate Removal:** Identified and removed duplicate entries to avoid double-counting transactions.

- **Handling Missing Values:**

  - Rows with missing or null product prices were treated appropriately, ensuring consistency in revenue calculations.
  - Rows with multiple critical missing values (e.g., missing CustomerID and Product Details) were dropped, as they offered limited analytical value.

- **Filtering Non-Transactional Data:** Excluded rows that did not represent actual transactions, such as administrative entries, adjustments, or placeholders.

- **Outlier Detection:** Reviewed outliers in quantity and price (e.g., excessively high or negative values) and handled them through segmentation or exclusion.

- **Ensuring Consistency:** Standardized data formats for key attributes like dates and identifiers to facilitate smooth segmentation and analysis.

2.2 Segmentation

The dataset was segmented to allow for tailored analysis of distinct transaction types and customer behaviors:

- **Retail Transactions:** Regular retail transactions, making up the largest segment and representing smaller, individual customer purchases, were isolated to capture typical consumer purchasing behaviors.

- **Retail Cancellations:** Cancellations within retail transactions were identified separately, enabling the analysis of cancellation trends and their effect on sales and customer retention.

- **Bulk Orders:** Transactions involving large quantities (e.g., over 50 units per product) were categorized as bulk orders, making up a small segment, providing insights into high-volume purchasing patterns.

- **Bulk Cancellations:** Identified cancellations associated with bulk orders to analyze their impact on revenue and customer behavior, focusing on trends specific to large-scale transactions.

2.3 Feature Engineering

Derived key metrics, such as revenue ($Quantity \times UnitPrice$), to measure transactions revenue. Additionally, the data was organized into time-based aggregations, including monthly, quarterly, and yearly summaries, to enable trend analysis and deeper temporal insights.

2.4 Tools

- **Excel:** Used for initial data inspection and exploration.

- **Python (Pandas, Matplotlib):** Utilized for advanced data cleaning, analysis, and creating insightful visualizations.

2.5 Analytical Methods:

- **Trend Analysis:** Calculated multiple KPI's, Monitored sales trends (MoM, QoQ) and customer behaviors.

- **Segmentation Analysis:** Examined regular sales, bulk orders, and their corresponding cancellations.

- **Visualization:** Used a variety of plots (bar, line, treemap, heatmap, piechart) to convey insights clearly.

3 Results and Insights

3.1 Sales Performance

3.1.1 Retail

This section focuses on retail sales with a quantity of fewer than 50 units per product, excluding bulk purchases (Quantity $\geq$ 50) and cancellations, providing insights into retail sales trends and performance.

Overview of Retail Sales:

Covering the period from December 2009 to December 2011, we have:

- **Total Revenue**: £14,283,212.07

- **Total Number of Transactions**: 970,147

- **Total number of Orders(Invoice)**: 39,287

General Key Metrics:

- **Average Order Value (AOV):** £363.56, which represents the average revenue per order. The high AOV suggests customers are making significant investments per order.

- **Average Basket Size (ABS):** 24.69 items, which indicates the average number of products per transaction. The high ABS reflects customers purchasing multiple items, effective bundling and large orders.
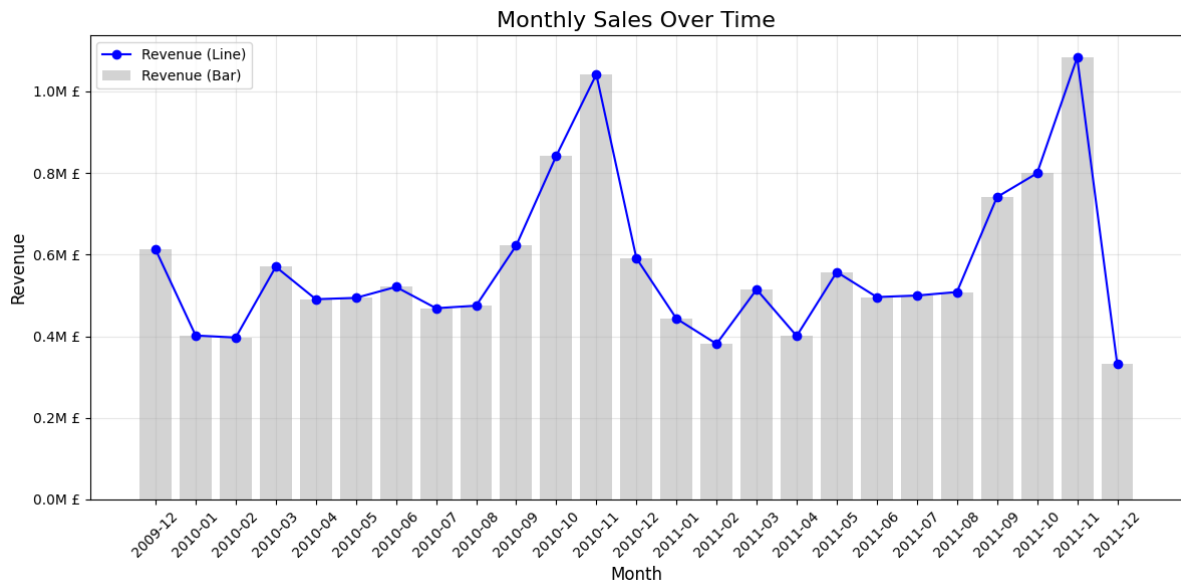
Temporal Analysis:

MoM:

Figure 1: Monthly Sales Over Time

The sales data exhibits clear seasonal fluctuations, with a notable peak in November likely due to the upcoming holiday season, followed by a decline in the subsequent months, suggesting the presence of seasonality. Additionally, the general trend shows a gradual decrease in sales post-November. Similarly, sales growth shows a steady positive growth from July to November, followed by a sharp decline in December, further supporting the hypothesis of a seasonal effect and possible underlying trend.
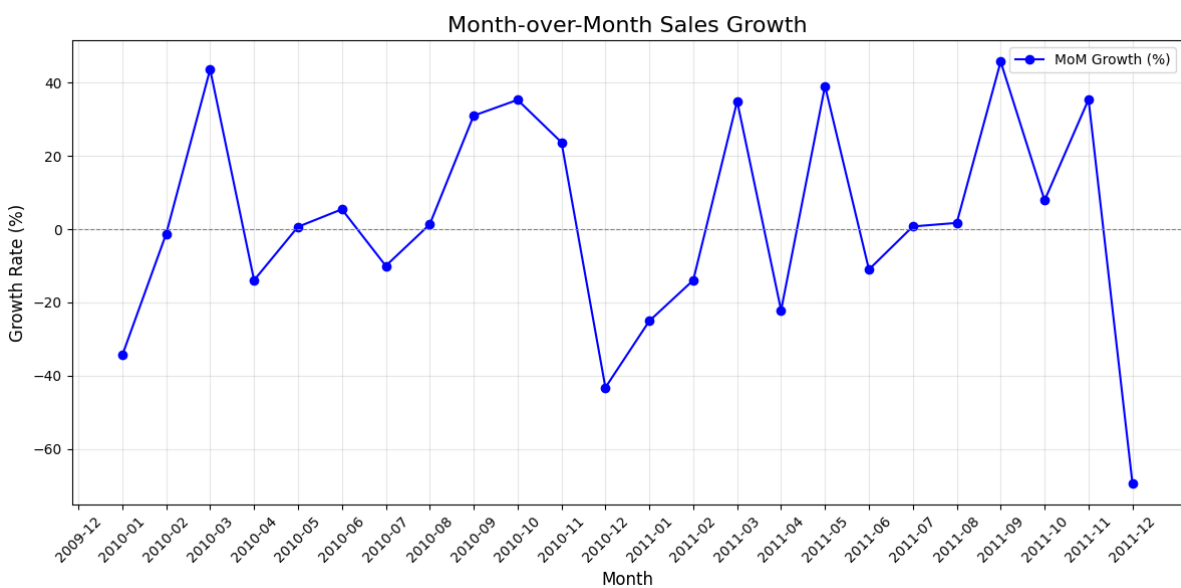
Figure 2: Month-over-Month Sales Growth

However, following this dip, sales stabilize between the month of March and August, indicating a period of relative stability post-holiday season.
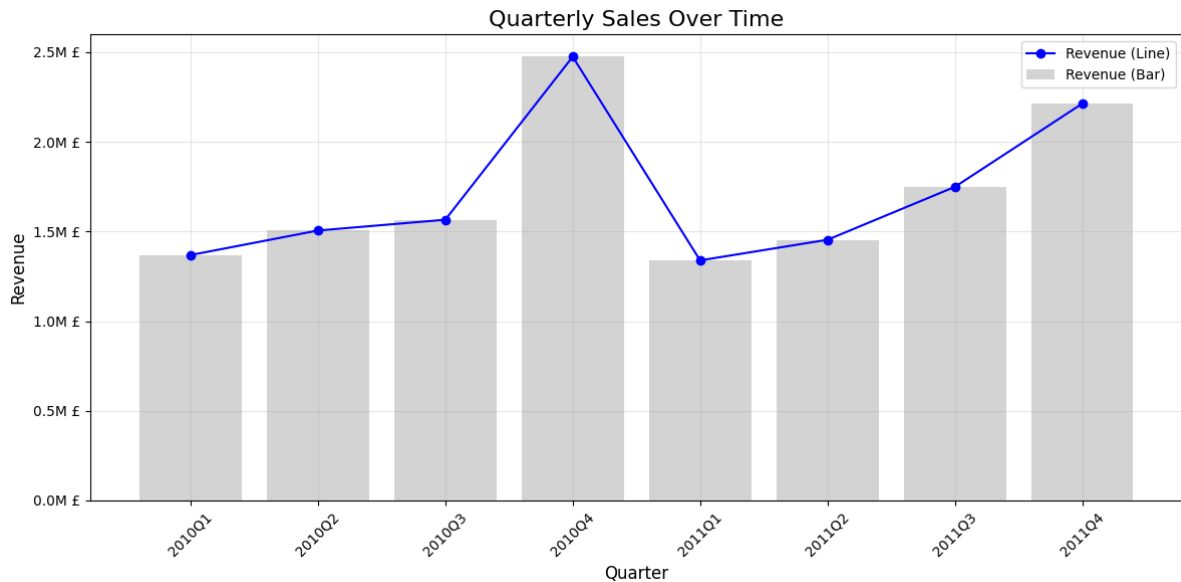
QoQ:
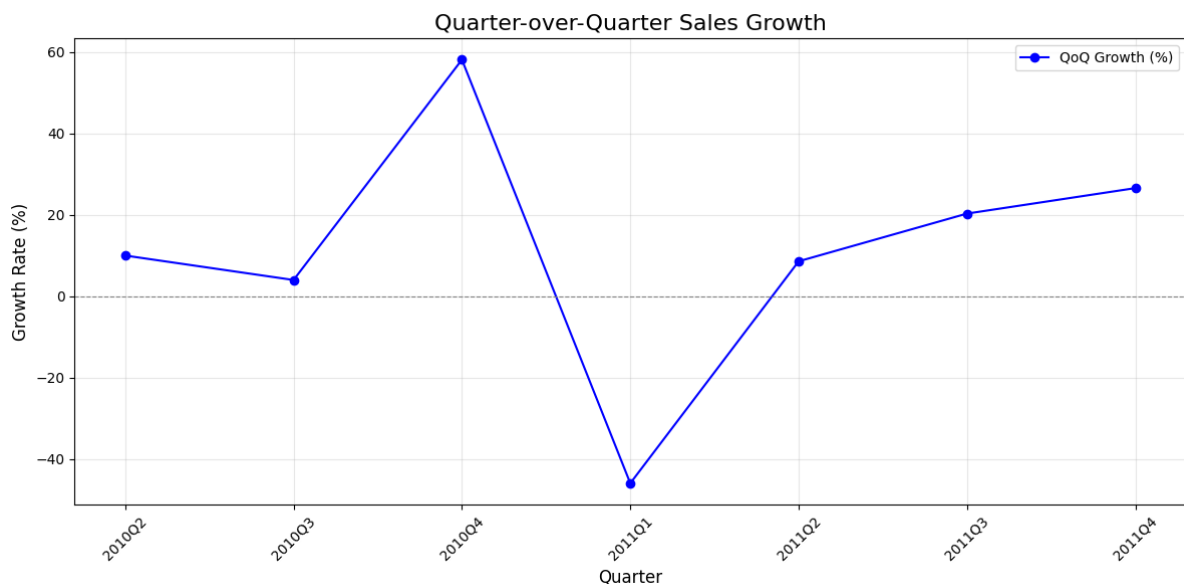

Figure 3: Quarterly Sales Over Time

Figure 4: Quarter-over-Quarter Sales Growth

The quarter-over-quarter analysis supports the earlier monthly analysis, reinforcing the pattern of strong sales growth in Q4, particularly in 2010Q4, followed by a decline in 2011Q1. This suggests that the peak in sales during Q4 is likely driven by seasonal factors, similar to the trends observed in the monthly data. The growth in subsequent quarters (Q2 and Q3 of both years) reflects a return to more stable, yet positive, sales

figures, with the 2011Q1 dip indicating the post-holiday season slowdown.

Regional Performance:

The retail data subset include 43 countries, the top 10 in terms of revenue generated are:
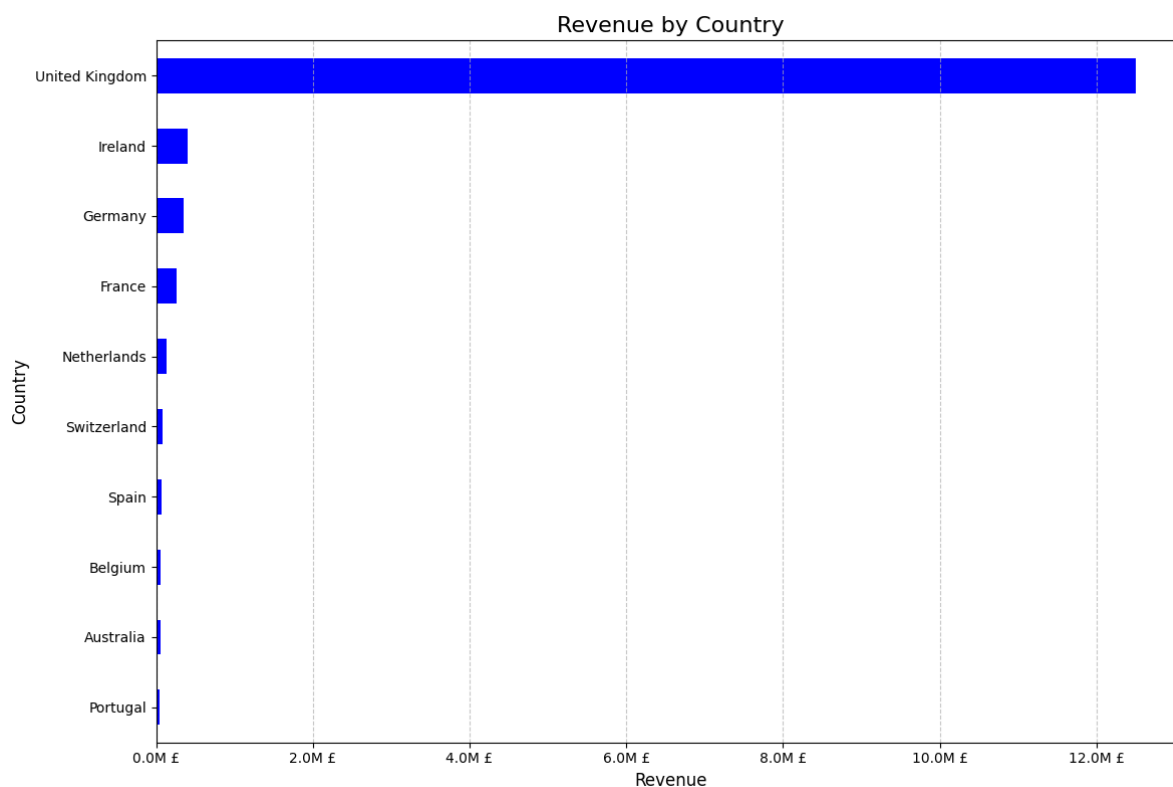

Figure 5: Top 10 Countries by Revenue

The UK dominates revenue at £12.5M, followed by Ireland (£399K) and Germany (£344K). Other countries contribute smaller amounts, highlighting the UK's leading role and potential growth opportunities in key markets like Ireland and Germany.
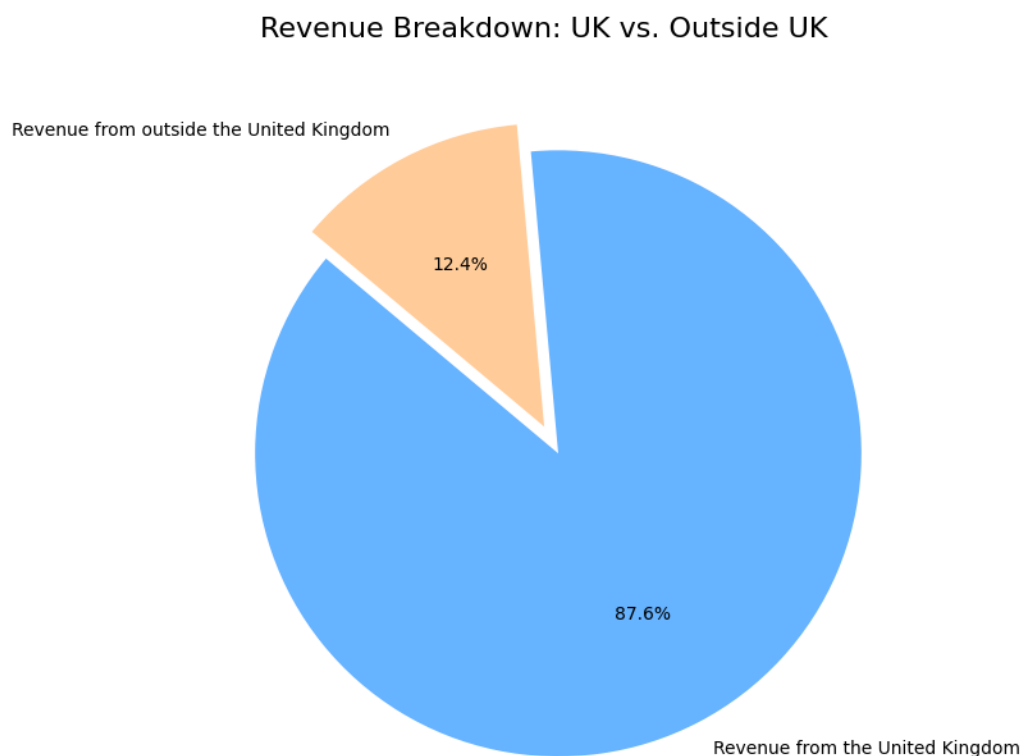
Figure 6: Revenue Breakdown UK vs Outside The UK

The pie chart illustrates the dominant contribution of the United Kingdom to the total retail revenue, accounting for an overwhelming 87.6%. This highlights the UK's central role in driving sales, while all other regions collectively make up just 12.4%, emphasizing a significant concentration of revenue generation in a single market.

Product Insights:

The retail data subset include 4883 distinct Products, top 20 in terms of revenue generated are:
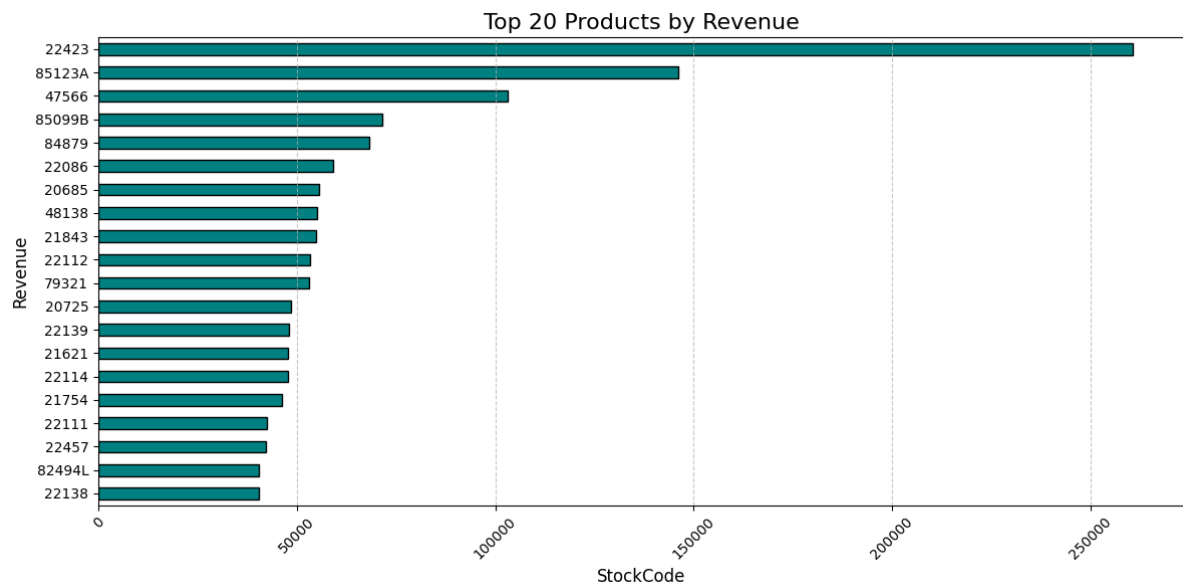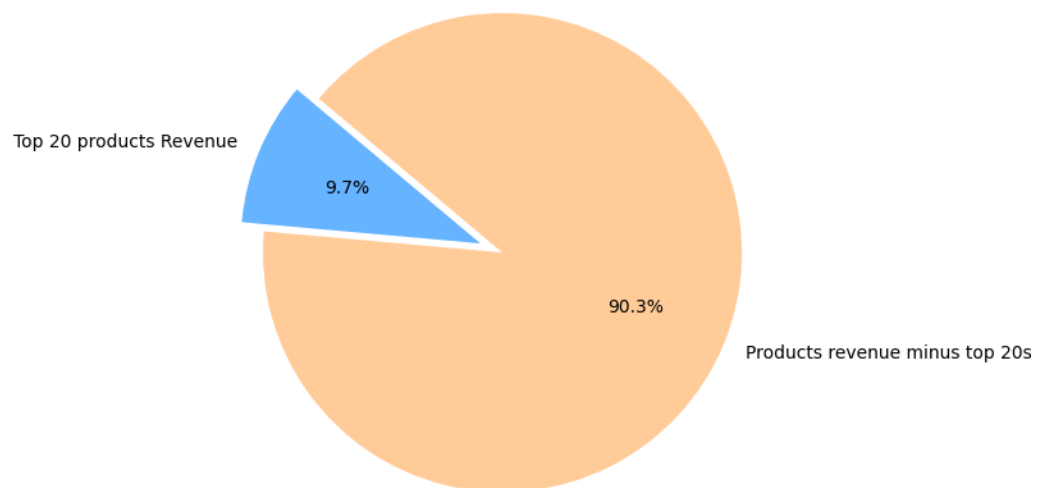
Figure 7: Top 20 Products by Revenue

Figure 8: Revenue Breakdown: Top 20 products vs The rest of the products

The top 20 products contribute significantly to the overall revenue, with the highest-selling product for example (22423) generating £260,775.07. comprising only 0.4% of the total 4,883 products, they generate 9.7% of total revenue. This highlights their outsized contribution and strategic importance.

Summary of Retail Insights:

The analysis highlights a strong seasonal pattern, with December seeing a significant sales peak, followed by a post-holiday decline. This suggests a clear seasonal effect, with sales stabilizing in the months after the holidays.

Regionally, the United Kingdom accounts for 87.6% of total revenue, indicating its dominant role in the market, while other regions contribute a smaller share. This concentration presents opportunities for growth in other markets such as Ireland and Germany.

In product performance, the top 20 products, despite representing just 0.4% of the total product range, contribute 9.7% of total revenue. This disproportionate share emphasizes their importance in the product strategy.

These insights suggest that leveraging the UK's strength while exploring expansion in other regions, alongside optimizing the product portfolio, could drive further growth.

3.1.2 Bulk

This section focuses on Bulk sales with a quantity of over 50 units per product(Quantity > 50), providing insights into bulk sales trends and performance.

Overview of Bulk Sales:

Covering the period from December 2009 to December 2011, we have:

- **Total bulk Revenue**: £5802651.11

- **Total Number of Bulk Transactions**: 24350

- **Total number of Orders That Included Bulk Transactions(Invoice)**: 8504

The total revenue, combining both retail and bulk, is £20,085,863.18, with bulk making up 28.9% of this. Despite individual bulk transactions being smaller (24,350 rows compared to 970,147 rows for retail), the bulk transactions are still significant, representing a considerable portion of total sales.
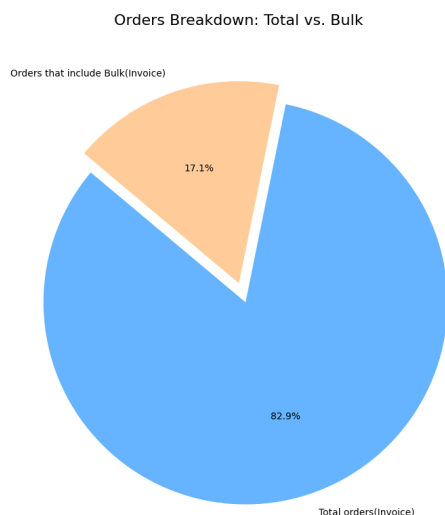
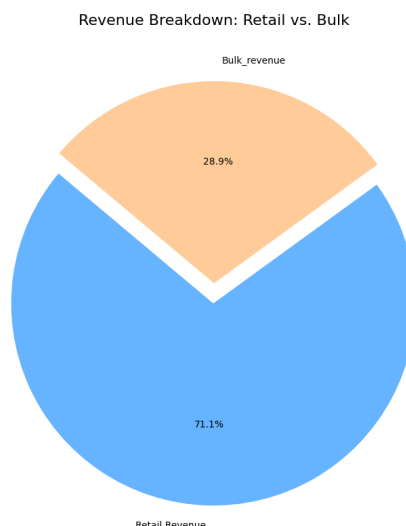Figure 9: Orders Breakdown: Retail vs Bulk

Figure 10: Revenue Breakdown: Retail vs Bulk

With 8,504 orders that include bulk purchases out of 41297 total orders, the volume of bulk transactions indicates a steady revenue contribution.

General Key Metrics

Average Order Value (AOV): £682.34. The AOV of £682.34 for bulk purchases reflects the higher spending typical of bulk transactions, where customers invest more per order due to buying larger quantities. This is normal for bulk purchases. **Temporal Analysis**
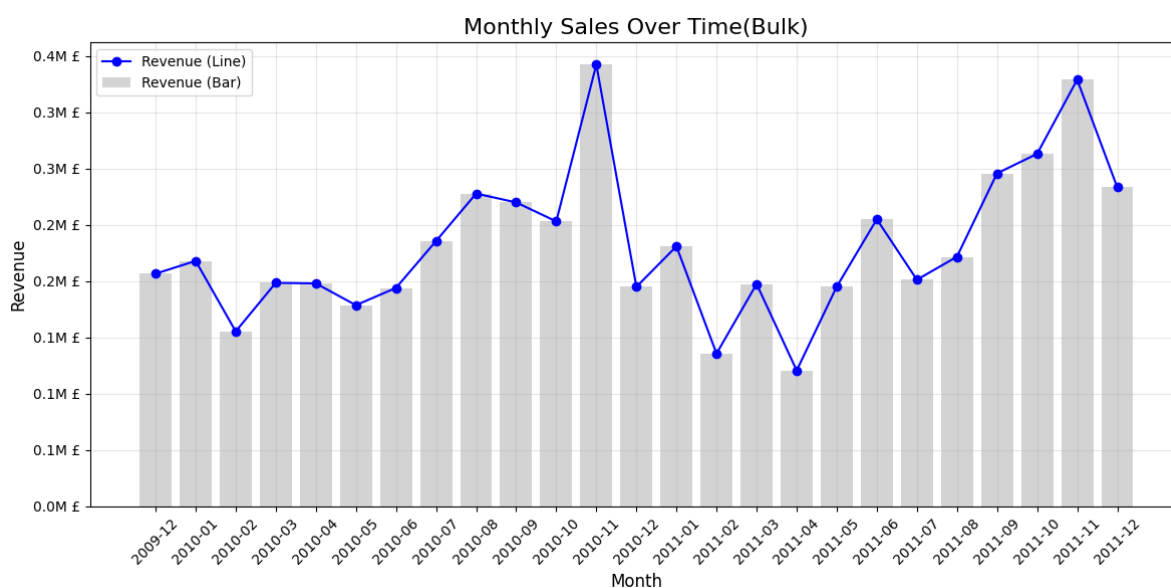
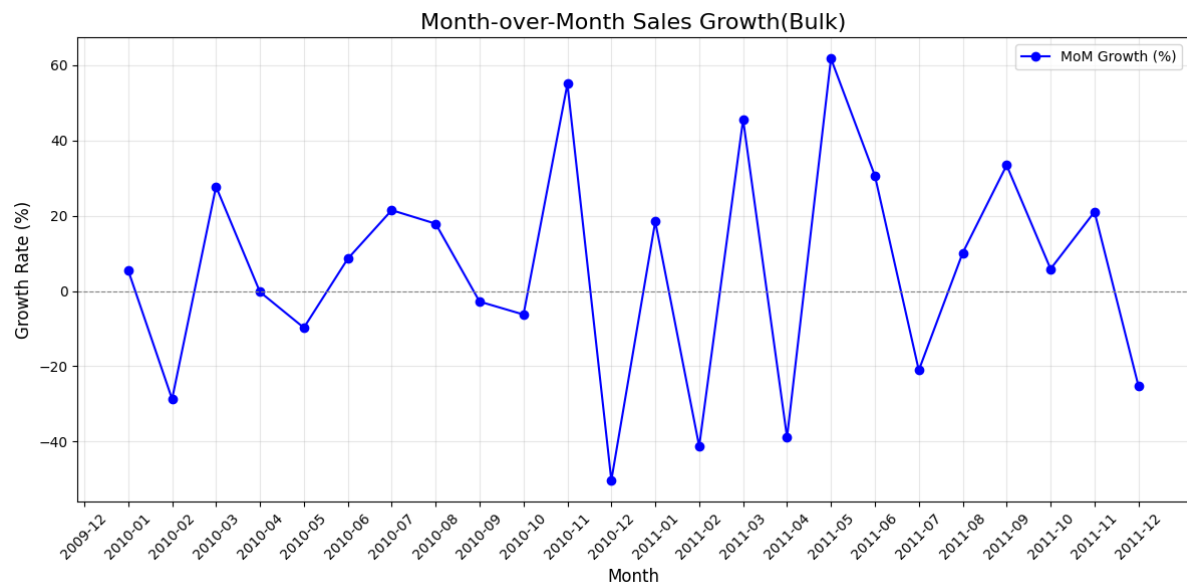**MoM:**

Figure 11: Monthly Bulk Sales Over Time

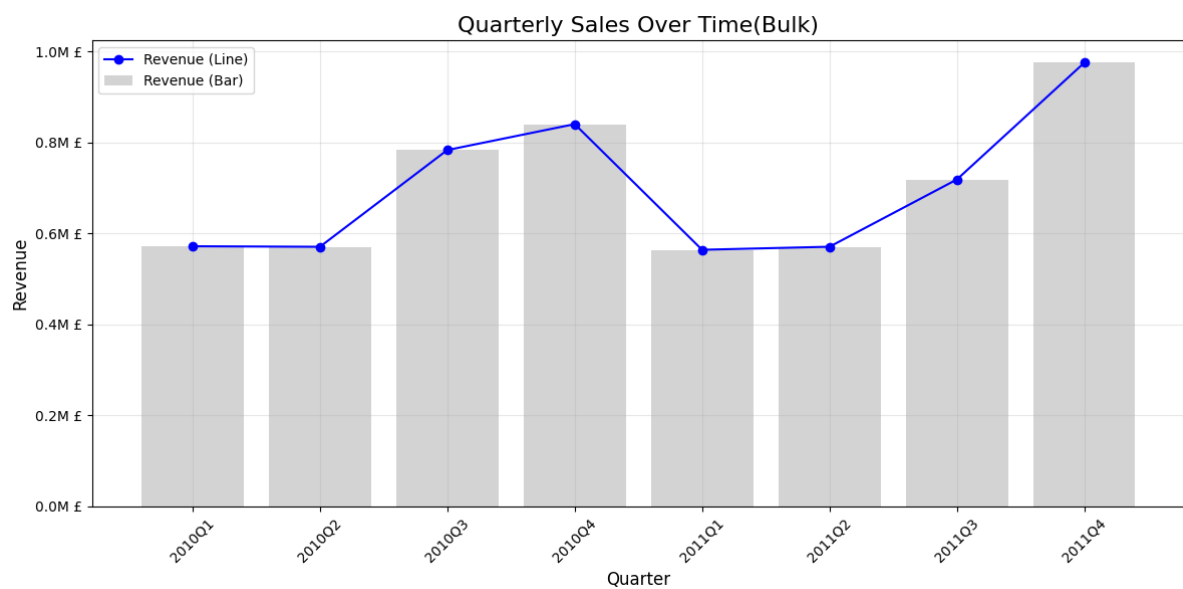Figure 12: Monthly Bulk Sales Growth Over Time

QoQ:

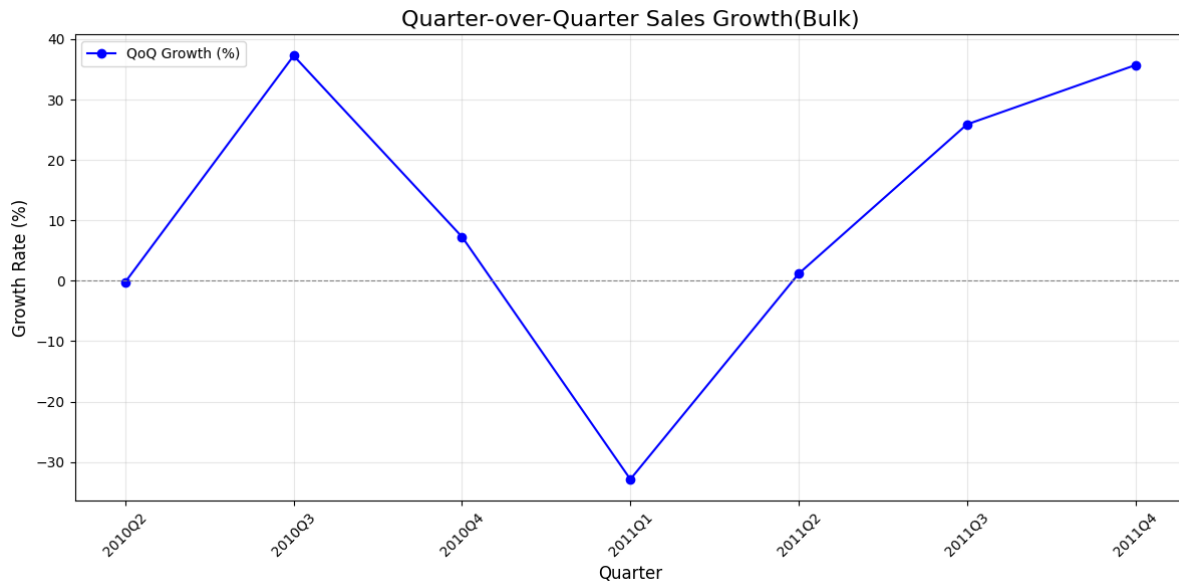Figure 13: Quarterly Bulk Sales Over Time

Figure 14: Quarterly Bulk Sales Growth Over Time

The bulk sales show similar trends to retail in both MoM and QoQ analyses. In November (2010Q4), bulk sales experience a sharp increase,, followed by a decrease in December(2011Q1). From March (2011Q1) through August (2011Q3), bulk sales stabilize, reflecting a return to more consistent performance. These patterns indicate that, like retail sales, bulk transactions are seasonal and exhibit a stable demand after the holiday peak.

Regional Performance:

The bulk data subset include 34 distinct countries, top 10 in terms of revenue generated are:
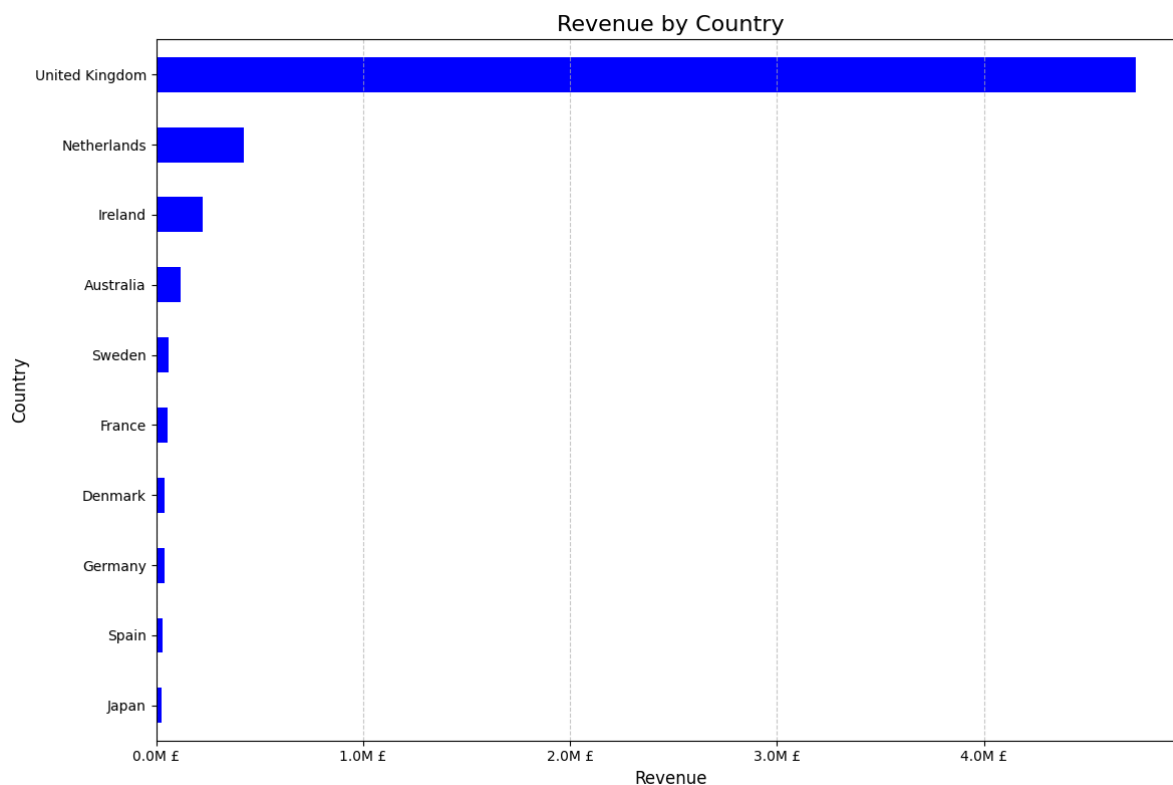
Figure 15: Top 10 Revenues by Country(Bulk)

Like retail sales, the UK heavily dominates bulk sales, with Ireland and Germany also appearing in the top ranks for both. Bulk sales, however, see stronger contributions from the Netherlands, Australia, and Sweden, unlike retail, where France plays a larger role. This suggests differing market demands between the two categories.

Product Insights:

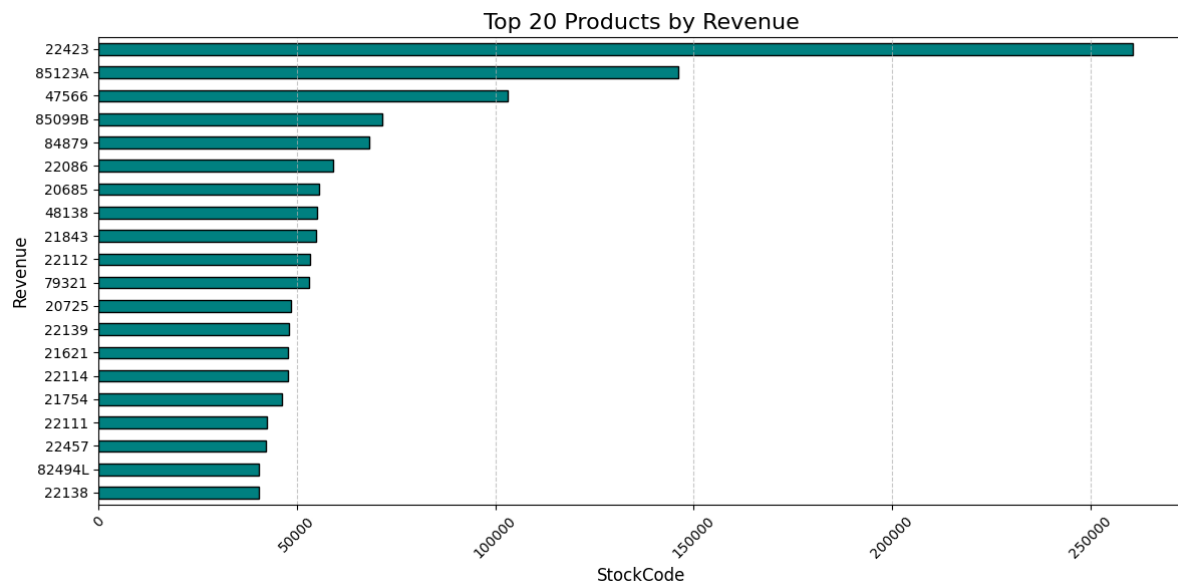The retail data subset include 2691 distinct Products, top 20 in terms of revenue generated are:
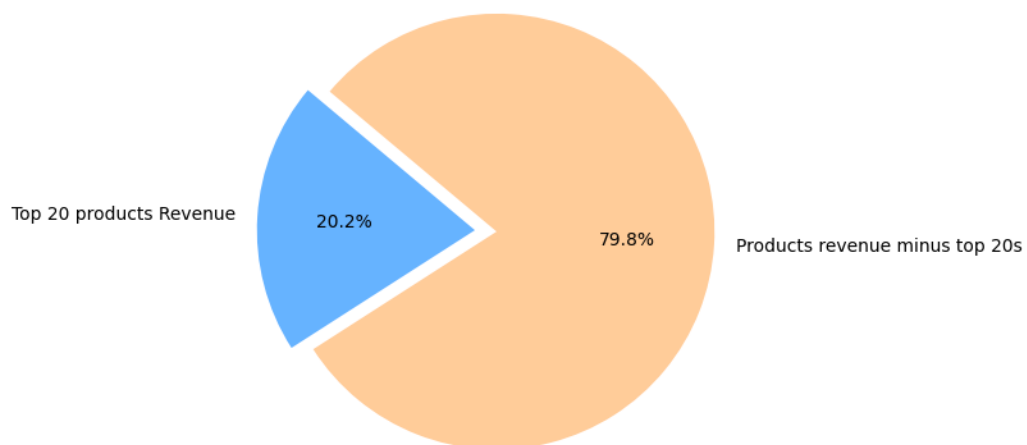
Figure 16: Top 20 Products by Revenue(Bulk)

Figure 17: Revenue Breakdown: Top 20 products vs The rest of the products(Bulk)

Similar to retail, the top 20 products contribute significantly to the overall revenue, with the highest-selling product (23843) generating £168469.60

The top 20 products, representing just 0.7% of the 2,691 total products, contribute a substantial 20.2% of the total revenue. This reinforces their critical role in driving sales and underscores the strategic importance of focusing on high-performing products.

Summary of Bulk Insights:

Bulk sales, while smaller in volume (24,350 transactions vs. 970,147 for retail), contribute significantly to overall revenue, comprising 28.9% of the total £20,085,863.18. Bulk transactions show high value, with an Average Order Value (AOV) of £682.34 in terms of bulk transactions, reflecting larger purchases typical of this category.

Seasonality trends in bulk sales align closely with retail, featuring a November peak (2010Q4) followed by stabilization in the first half of 2011. Regional performance is similarly dominated by the UK, though bulk sales see stronger contributions from the Netherlands, Australia, and Sweden, contrasting with the secondary retail's higher reliance on France.

Product concentration is evident in both categories; the top 20 bulk products, accounting for just 0.7% of offerings, generate 20.2% of bulk revenue, mirroring the outsized importance of high-performing products in retail. These patterns underscore the strategic importance of bulk sales as a steady, high-value revenue source.

3.2 Customer Analysis

This section analyzes customer behavior within the dataset by examining key metrics such as customer segmentation, purchase frequency, customer lifespan, Customer lifetime value, churn rate, and retention rate. These insights help us understand customer loyalty, purchasing trends, and overall business sustainability.

3.2.1 Retail

This sub-section focuses on retail Customers with a quantity of fewer than 50 units per product sold, excluding bulk purchases (Quantity ≥ 50) and cancellations.

Customer Analysis Overview:

Covering the period from December 2009 to December 2011, we have:

- **Total Known Customers**: 5,758 Customers

- **Total Number of Transactions with a known customer**: 743,119 transaction

- **Transactions by Unknown Customers**: 227,028 transaction

- **Total Number Of Orders(Invoice)**: 39,287 order

- **Orders By Unknown Customers**: 4,293 order
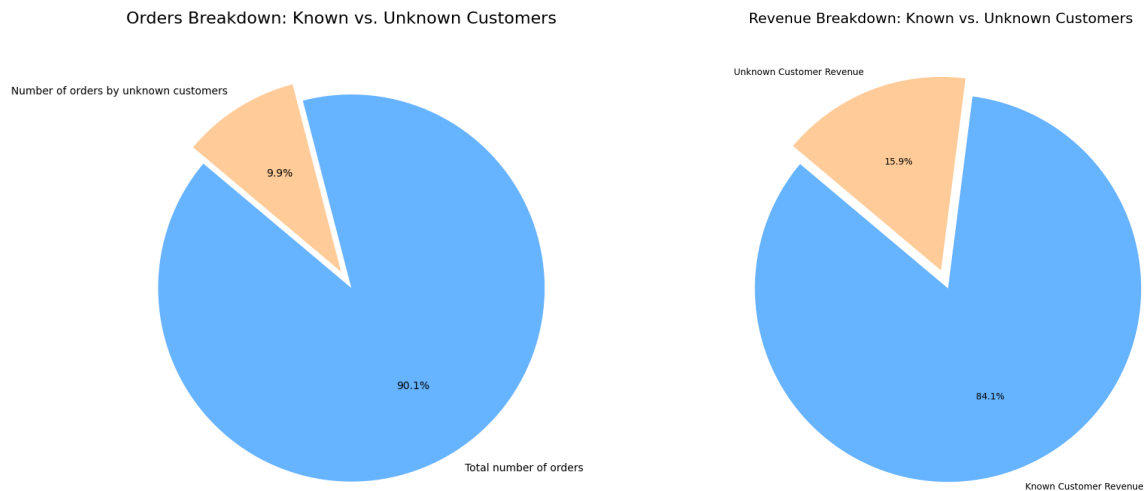
Known vs. Unknown Customers:

Figure 18: Orders Breakdown: Known vs Unknown Customers

Figure 19: Revenue Breakdown: Known vs Unknown Customers

The fact that 15.9% of total retail revenue 9.9% of orders comes from unidentified customers, raises a concern about the significant portion of revenue from unidentifiable Customers. This gap highlights the importance of improving customer identification processes, as understanding these customers could uncover valuable insights and opportunities for targeted strategies and relationship-building.

Top Customers:

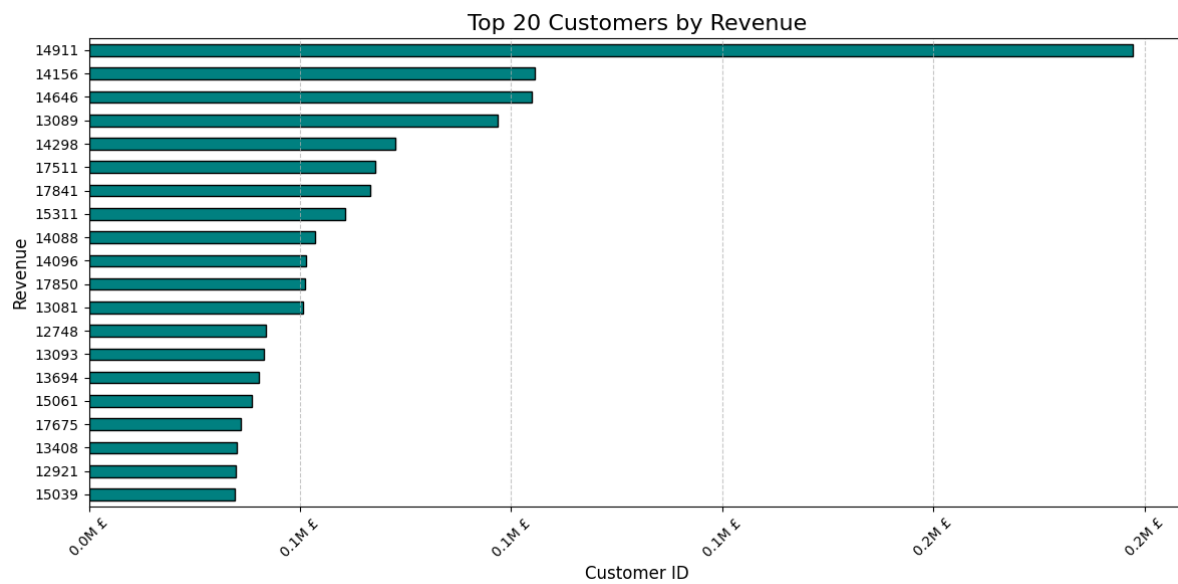Top 20 customers in terms of revenue generated are:
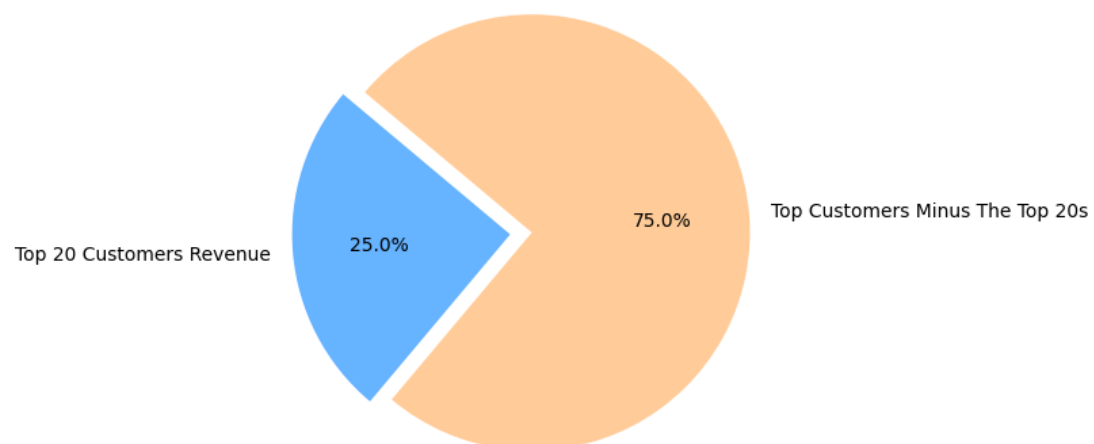
Figure 20: Top 20 Customers by Revenue

Figure 21: Revenue Breakdown: Top 20 customers vs The rest of the customers

The top 20 customers, collectively contributes a significant portion of the total revenue(25%), underscoring the importance of these key clients to the business.

Purchase Frequency

With an average purchase frequency of **6.82** over the period from 12-2009 to 12-2011, customers are making nearly 7 purchases or approximately 3.41 purchases a year.

Customer Lifetime (Lifespan)

The average customer lifetime span of **0.747** years indicates that, on average, customers engage with the business for a period of approximately 9 months before their purchasing activity declines or stops.

Customer Lifetime Value (CLV)

A Customer Lifetime Value (CLV) of **£1,853.44** indicates that, on average, each customer is expected to generate £1,853.44 in revenue throughout their entire relationship with the business.

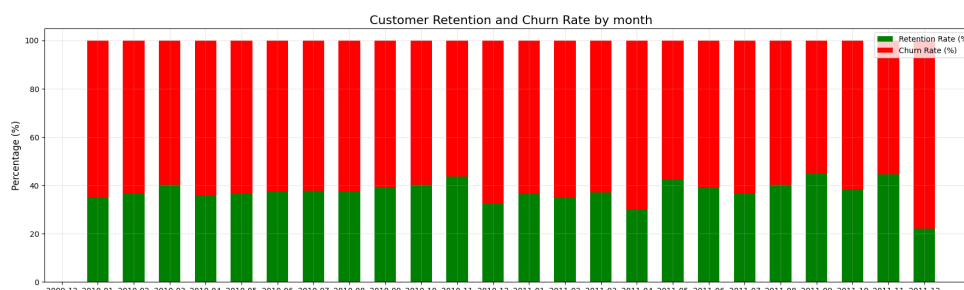**Churn and Retention Rates**

MoM:

Figure 22: Customer Retention and Churn Rate by month

The retention rate analysis reveals fluctuations over the observed period, with notable peaks (e.g., 44.7% in September 2011) and dips (e.g., 22.0% in December 2011). While some variability is present, a decrease in retention towards the end of the period is not uncommon for online retail businesses, especially during seasonal transitions or following peak shopping periods as evident by the earlier month over month sales insights. The churn rate, which is the complement of the retention rate, highlights the number of customers lost during each month.
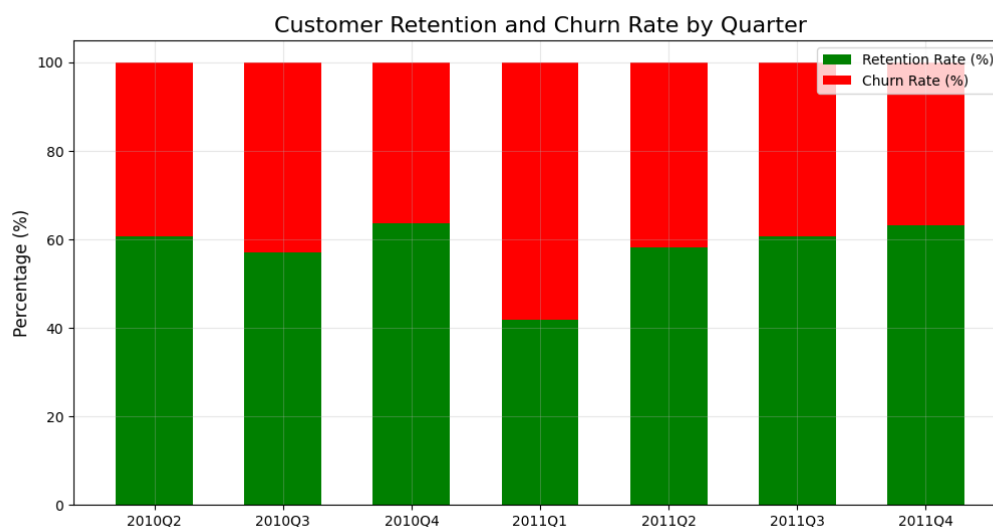
QoQ:

Figure 23: Customer Retention and Churn Rate by quarter

The quarterly retention and churn rates show more stability compared to the monthly data, reflecting longer-term customer behavior. Notably, retention rates are higher each quarter, particularly in Q4 of both 2010 and 2011, which suggests stronger customer loyalty during the seasonal peaks. The drop in retention during Q1 2011 indicates typical post-season customer disengagement.
Overall, the quarterly data highlights the importance of longer-term trends in understanding customer retention, with more consistent engagement over each quarter compared to monthly fluctuations.

Summary of Customers Insights:

Customer analysis reveals important insights into customer behavior. Known customers, totaling 5,758, generated the majority of transactions. However, unknown customers accounted for 15.9% of revenue and 9.9% of orders. This highlights the need to improve customer identification processes. The top 20 customers rsponsable for quarter of the retail revenue play a significant role in driving forward the business growth, underscoring their importance.

With an average purchase frequency of 6.82, well above the industry standard, customers demonstrate strong engagement. However, the average customer lifetime of 9 months and a CLV of £1,853.44 suggest that retaining customers over the long term is crucial for maximizing value.

Retention rates fluctuate monthly, with some months dipping below 30%, though retention rates above this threshold are considered solid for online retail. Quarterly retention data shows more stability and higher customer loyalty, particularly during seasonal peaks, highlighting the importance of longer-term customer retention strategies.

3.2.2 Bulk

This sub-section focuses on bulk Customers with a quantity of over 50 units per product sold.

Customer Analysis Overview:

Covering the period from December 2009 to December 2011, we have:

- **Total bulk Known Customers**: 1,971 Customers

- **Total Number of bulk Transactions with a known customer**: 23,013 transaction

- **Bulk transactions by Unknown Customers**: 1,337 transaction

- **Bulk Orders By Unknown Customers(Invoices that include Bulk)**: 696 order
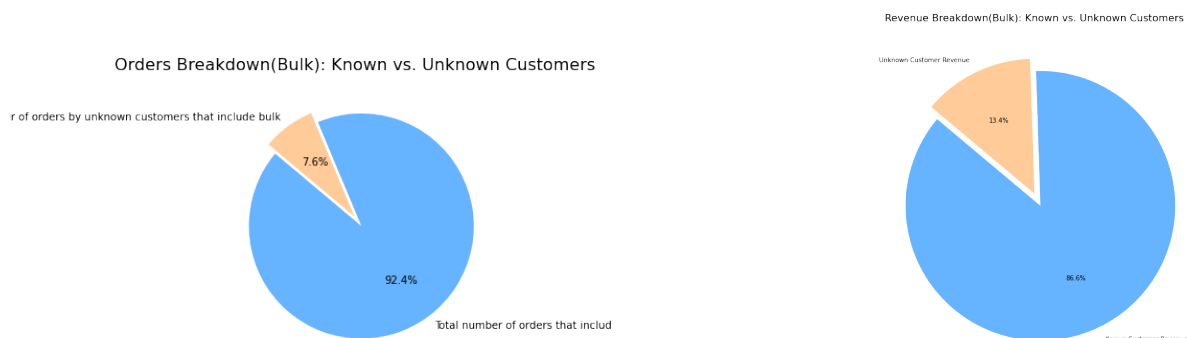
Known vs. Unknown Customers:

Figure 24: Orders Breakdown(Bulk): Known vs Unknown Customers

Figure 25: Revenue Breakdown(Bulk): Known vs Unknown Customers

In the bulk segment, 13.4% of total revenue and 7.6% of orders come from unidentified customers, which is a smaller proportion compared to retail. While the issue is less pronounced in bulk, it still highlights a notable gap in customer identification. Addressing this could provide opportunities to better understand and engage these unidentified customers, potentially uncovering strategies to enhance revenue and build stronger business relationships in the bulk segment.

Top Customers:

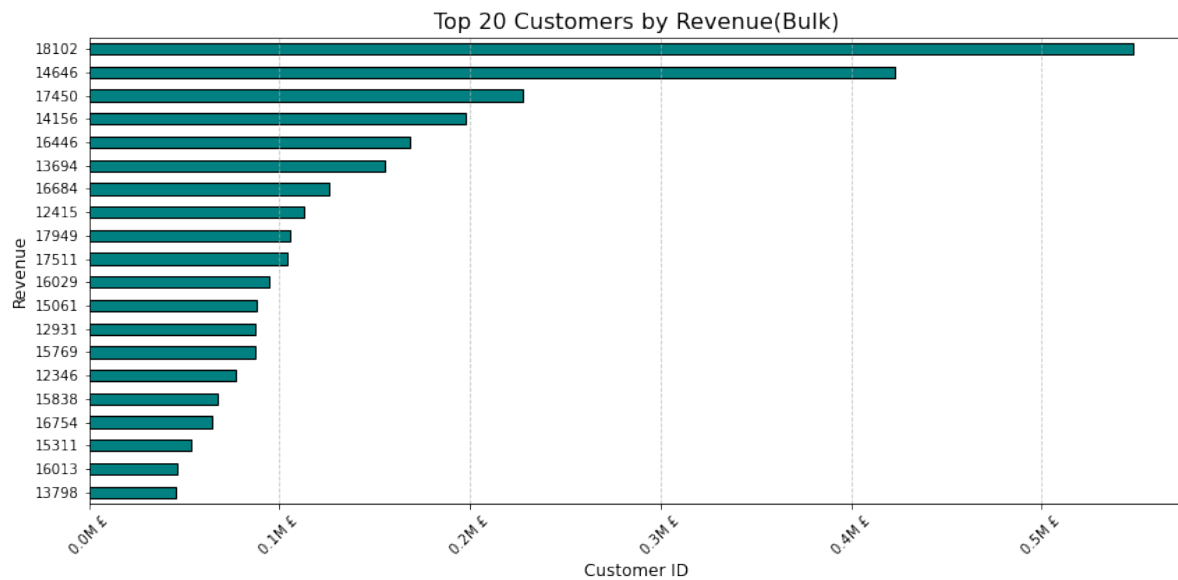Top 20 customers in terms of bulk revenue generated are:

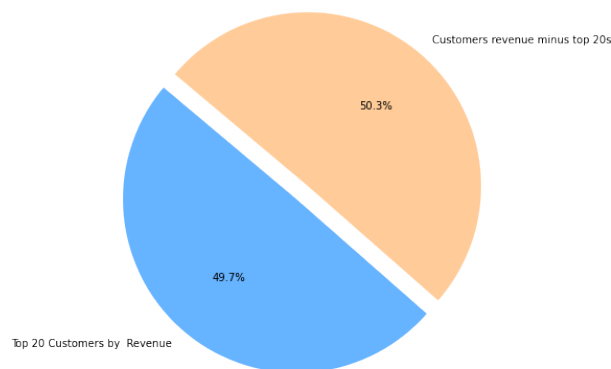Figure 26: Top 20 Customers by Revenue(Bulk)

Figure 27: Revenue Breakdown(Bulk): Top 20 customers vs The rest of the customers

In the retail segment, the top 20 customers contribute significantly to revenue but do not dominate the overall business, highlighting a more diversified customer base. In contrast, the bulk segment is heavily reliant on its top 20 customers, who account for nearly half (49.7%) of total revenue. This indicates a more concentrated revenue structure in bulk operations, emphasizing the critical importance of retaining and nurturing these key customers to maintain revenue stability.

Purchase Frequency:

In the bulk segment, the average purchase frequency is **4.31**, which is notably lower compared to the retail segment's **6.82**. This indicates that bulk customers engage with the business less frequently over the observed period.

Customer Lifetime (Lifespan):

Bulk customers have a shorter average lifespan of approximately **0.477 years** (just under 6 months), compared to retail customers who maintain activity for about 9 months (**0.747 years**).

Customer Lifetime Value (CLV):

The Customer Lifetime Value (CLV) in the bulk segment is **£1,404.90**, lower than the retail segment's **£1,853.44**. This indicates that bulk customers generate less revenue over their relationship with the business.
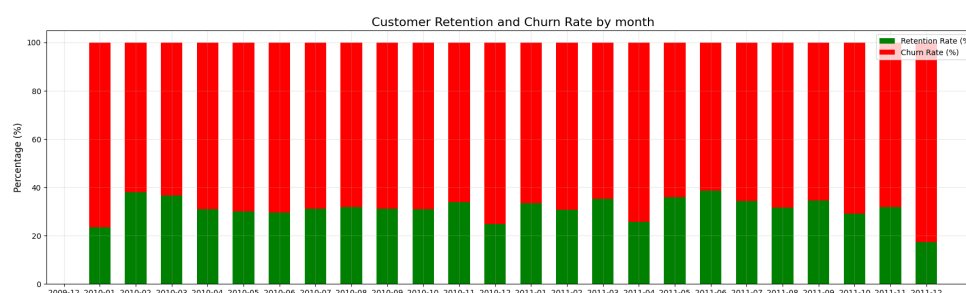
Churn and Retention Rates:

MoM:

Figure 28: Customer Retention and Churn Rate by month(Bulk)

Bulk retention rates range from 17.24% to 38.71%, with churn rates inversely peaking at 82.76%. Compared to retail's higher and steadier retention (40-50%), bulk shows greater volatility and a less loyal customer base, as visualized in the stacked bar chart.
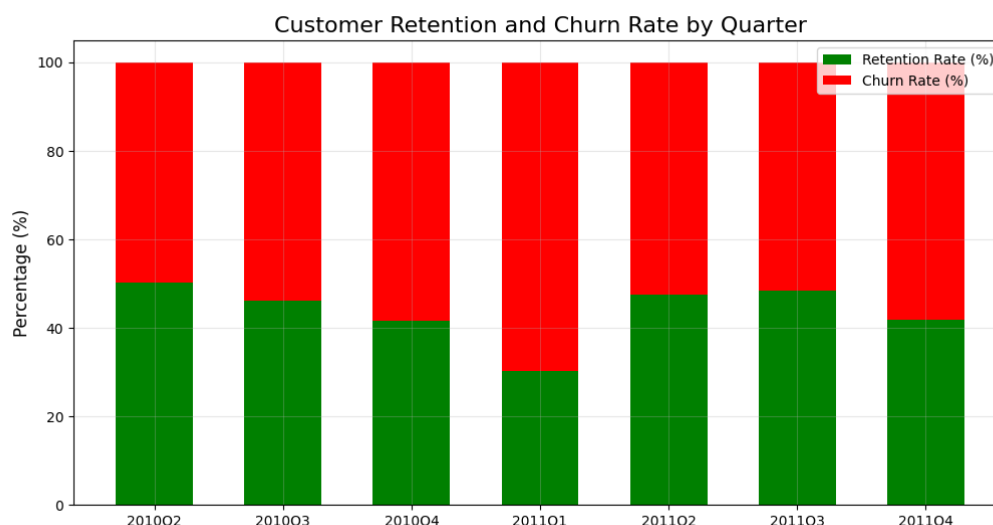
QoQ:

Figure 29: Customer Retention and Churn Rate by Quarter(Bulk)

Compared to monthly figures, bulk quarterly retention rates (30.29% to 50.25%) and churn rates (49.75% to 69.71%) exhibit greater stability, smoothing out month-to-month fluctuations. In contrast, retail quarterly retention rates are consistently higher (47.95% to 62.93%), reflecting a more loyal customer base, while churn rates are lower (37.07% to 52.05%), highlighting stronger customer engagement in the retail segment. This comparison underscores the relative stability of bulk metrics at the quarterly level but also emphasizes that retail retains a stronger overall customer relationship dynamic.

Summary of Customers Insights:

The bulk customer segment displays unique characteristics compared to retail, with 1,971 known customers generating 23,013 transactions during the observed period. Unidentified customers account for 13.4% of revenue, a smaller proportion than in retail, but still significant enough to warrant efforts to improve customer identification. Bulk revenue is highly concentrated, with nearly 50% generated by the top 20 customers, highlighting the critical need to retain and nurture these key accounts.

Bulk customers engage less frequently than retail customers, with a lower purchase frequency (4.31 vs. 6.82) and a shorter average lifespan (6 months vs. 9 months). Retention rates in bulk are also lower and more volatile, ranging from 17.24% to 38.71%, compared to retail's steadier 40–50%. However, quarterly metrics show slightly greater stability in bulk retention and churn rates, smoothing out monthly fluctuations.

Despite these challenges, bulk customers remain a valuable segment, with a customer lifetime value of £1,404.90. Strategic efforts to enhance loyalty, expand engagement, and reduce revenue reliance on top customers could significantly improve long-term performance.

3.3   Cancellation Analysis

This section delves into the analysis of cancellations within the retail and bulk datasets, examining cancellation rates, trends over time, and their impact on sales. Insights into canceled products and the customers responsible for the highest cancellation rates will provide valuable context for improving operational strategies.

3.3.1   Retail

This subsection focuses on retail cancellations with a quantity of fewer than 50 units per product, providing insights into retail cancellation trends and impact.

Retail Cancellations Overview:

Covering the period from December 2009 to December 2011, we have:

- **Total Retail Cancellations loss Revenue**: £256,097.23£

- **Number of Retail Canceled Transactions**: 17,219

- **Number of Retail Canceled Orders(Invoice)**: 7,216
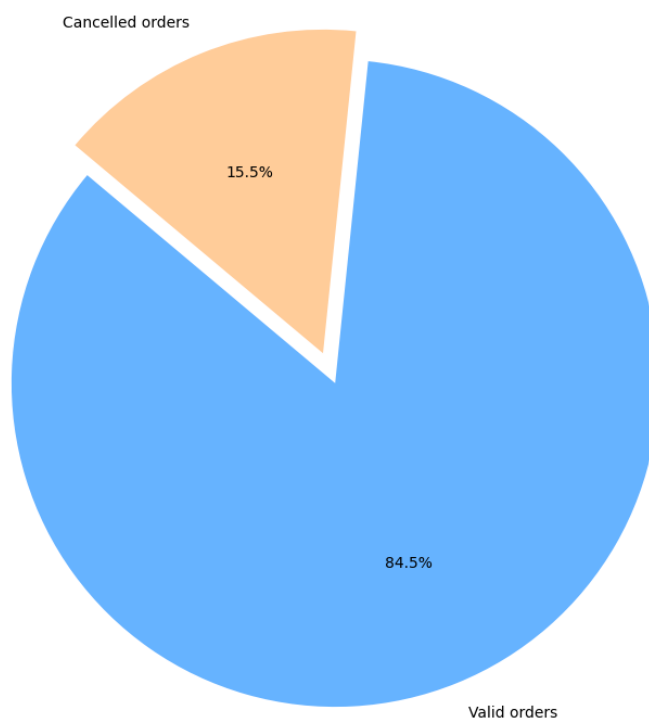
Cancellations vs. Sales:

Figure 30: Invoice Breakdown: Sold Orders vs Cancelled Orders

Although the raw cancellation rate may seem notable, it's important to consider the context. The average number of items per canceled order is significantly lower than that of retail orders, suggesting a different transactional dynamic. This observation invites a deeper look at how cancellations are distributed and their influence on overall sales performance.

Cancellation rate: The raw cancellation rate of 15.51% provides an initial perspective on the impact of cancellations; however, it doesn't fully capture the transactional dynamics due to the discrepancy in the average number of items per order between canceled and retail sales. Canceled orders typically contain fewer items, with an average of 2.39 items per cancellation compared to 24.69 items per retail order.

To adjust for this difference and offer a more accurate cancellation rate, a correction factor was applied by multiplying the raw rate by the ratio of the average number of items per canceled order to the average number of items per retail order. This adjustment ensures a more meaningful representation of cancellations in the context of overall sales, providing a clearer view of the real impact on revenue and customer behavior.

$$\text{Adjusted Cancellation Rate} = \text{Cancellation Rate} \times \left(\frac{\text{Average Transaction per Cancellation Order}}{\text{Average Transaction per Retail Order}} \right)$$
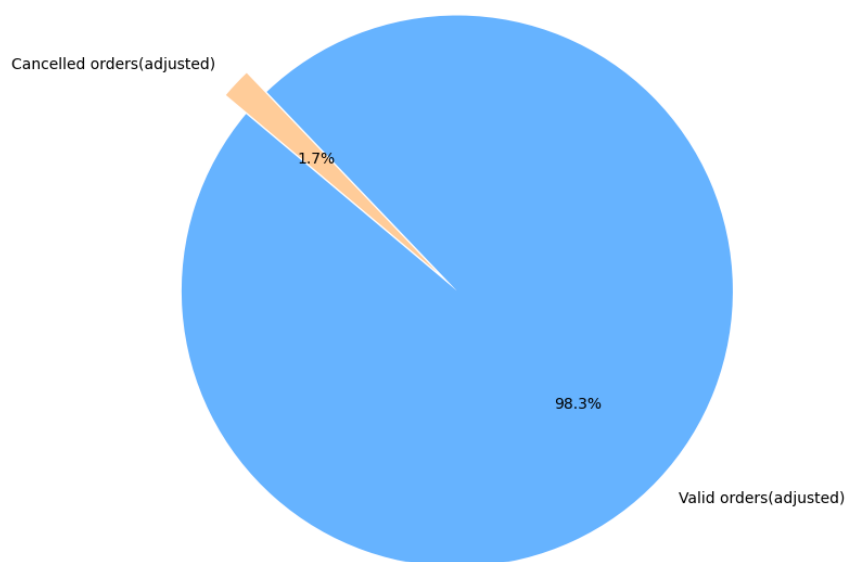

Figure 31: Invoice Breakdown(adjusted to the avrage basket size): Sold Orders vs Cancelled Orders

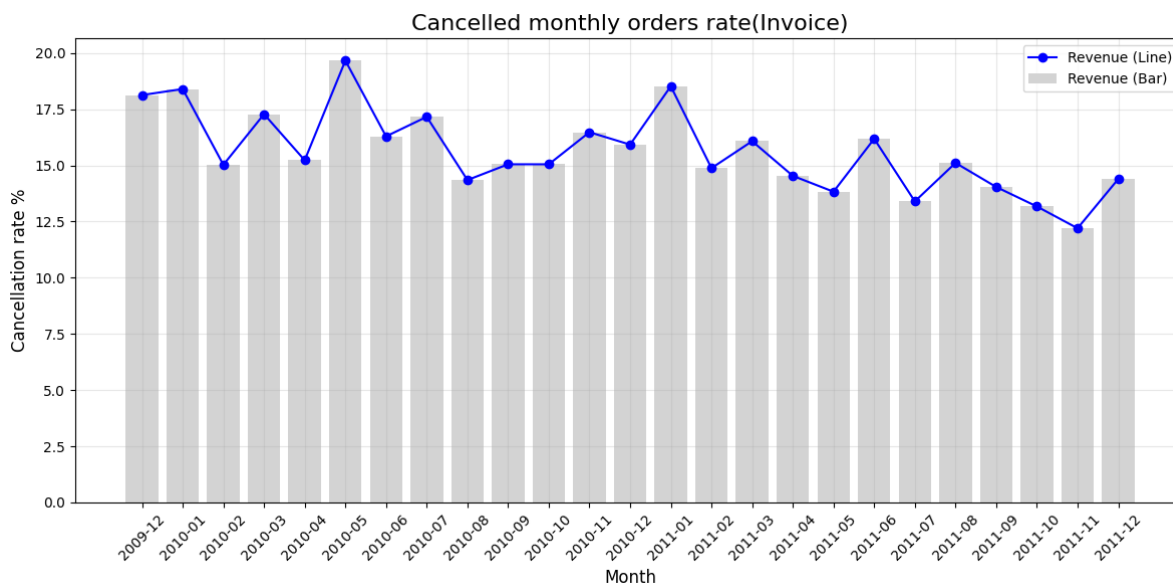**Cancellations Over Time:**

MOM:

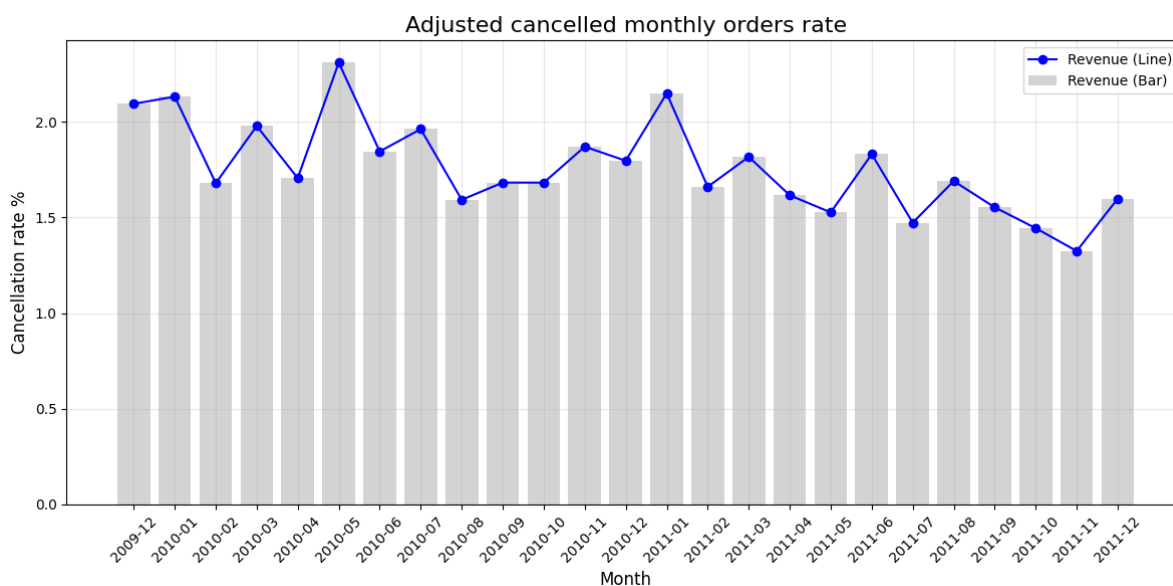Figure 32: Cancelled monthly orders rate(Invoice)

Figure 33: Adjusted cancelled monthly orders rate(Invoice)

The monthly cancellation rates reflect fluctuations in the percentage of canceled orders with a slow overall decrease over time, and a noticeable peak in May 2010 at 19.66% as well as a lower point in November 2011 at 12.20%. When adjusting the rates, the cancellation pattern follows the same trend with significantly lower rates.
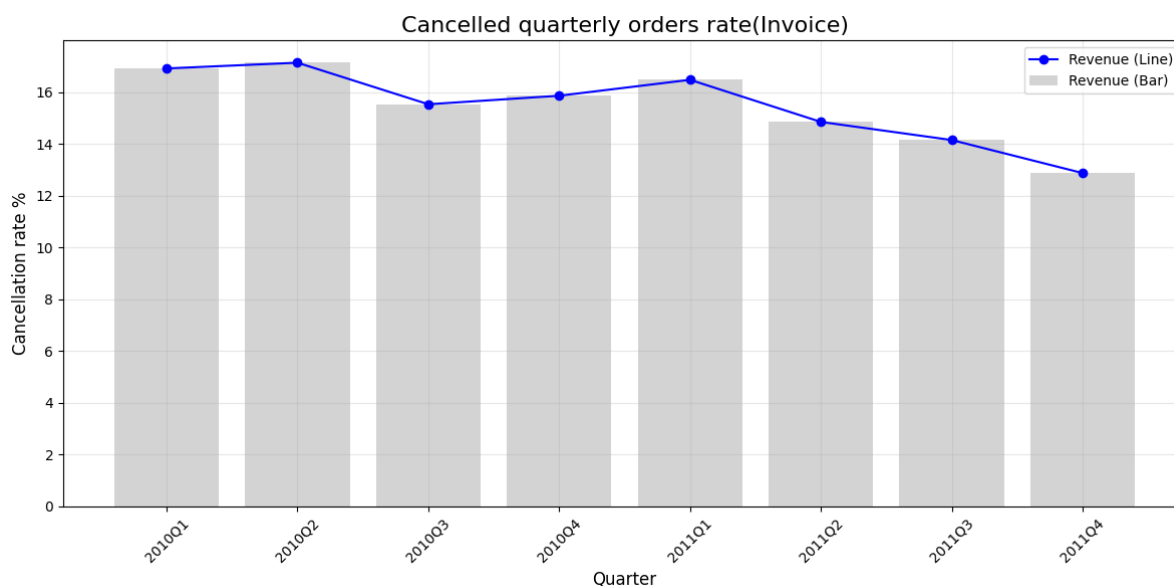
QoQ:

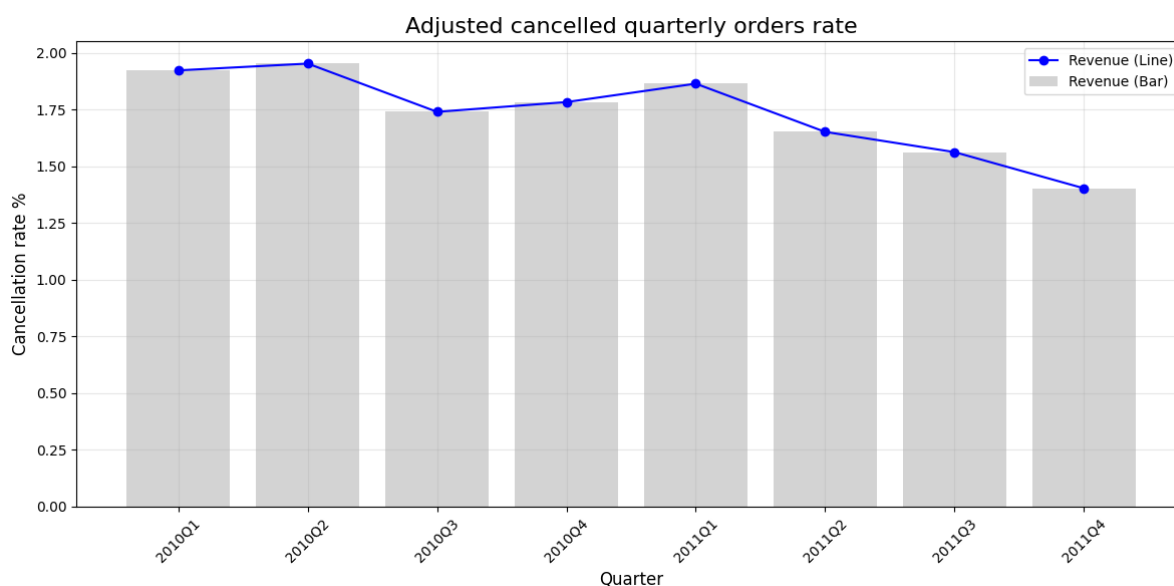Figure 34: Cancelled quarterly orders rate(Invoice)

Figure 35: Adjusted cancelled quarterly orders rate(Invoice)

Similar to the monthly cancellations, the quarterly cancellation rates show a generally decreasing trend from 16.91% in Q1 2010 to 12.87% in Q4 2011, with Q2 2010 peaking at 17.13%. This indicates that cancellations were more frequent in the earlier quarters,

gradually tapering off toward the end of 2011. When adjusting the rates, the cancellation pattern follows the same trend with significantly lower rates. The decline in both the raw and adjusted rates suggests a steady improvement in customer purchasing behavior over time.

Top Canceled Products:
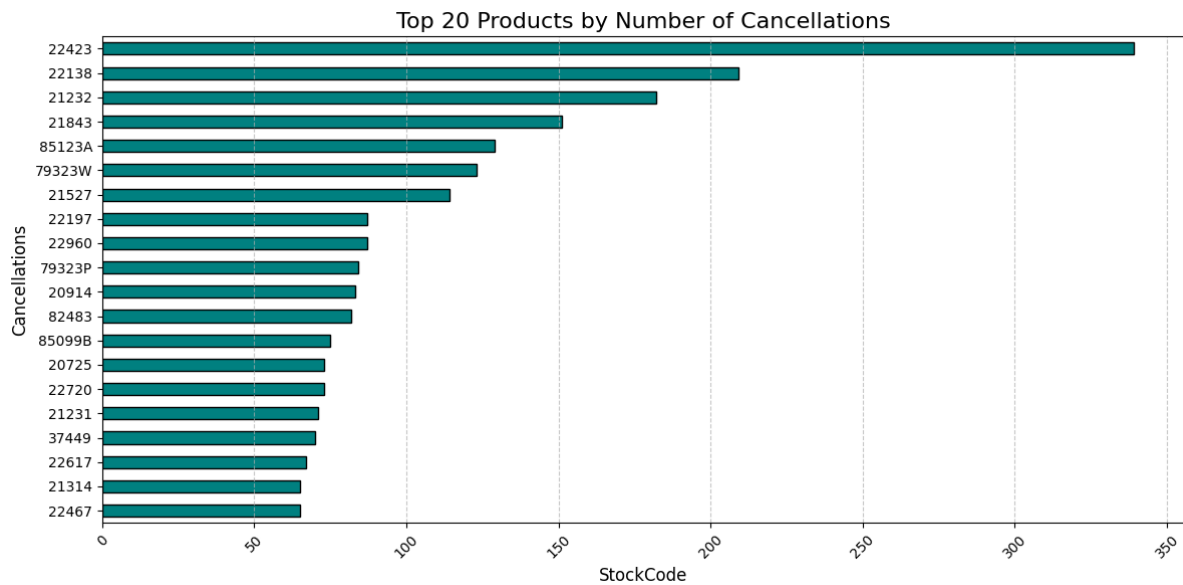
The top 20 canceled products are:

Figure 36: Top 20 Products by Number of Cancellations

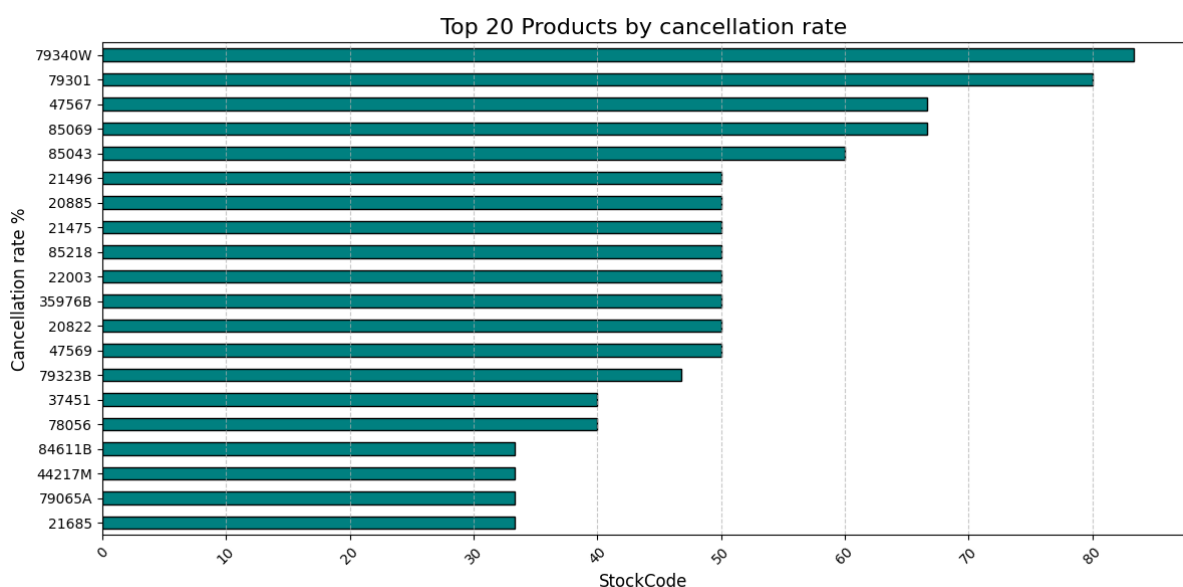Meanwhile the highest 20 cancellation rates in products are:

Figure 37: Top 20 Products by Cancellation Rate

Some items like 79340W or 79301 (83.33% and 80% cancellation rate) show disproportionately high rates, signaling potential issues with these products.

Most Cancelling Customers:
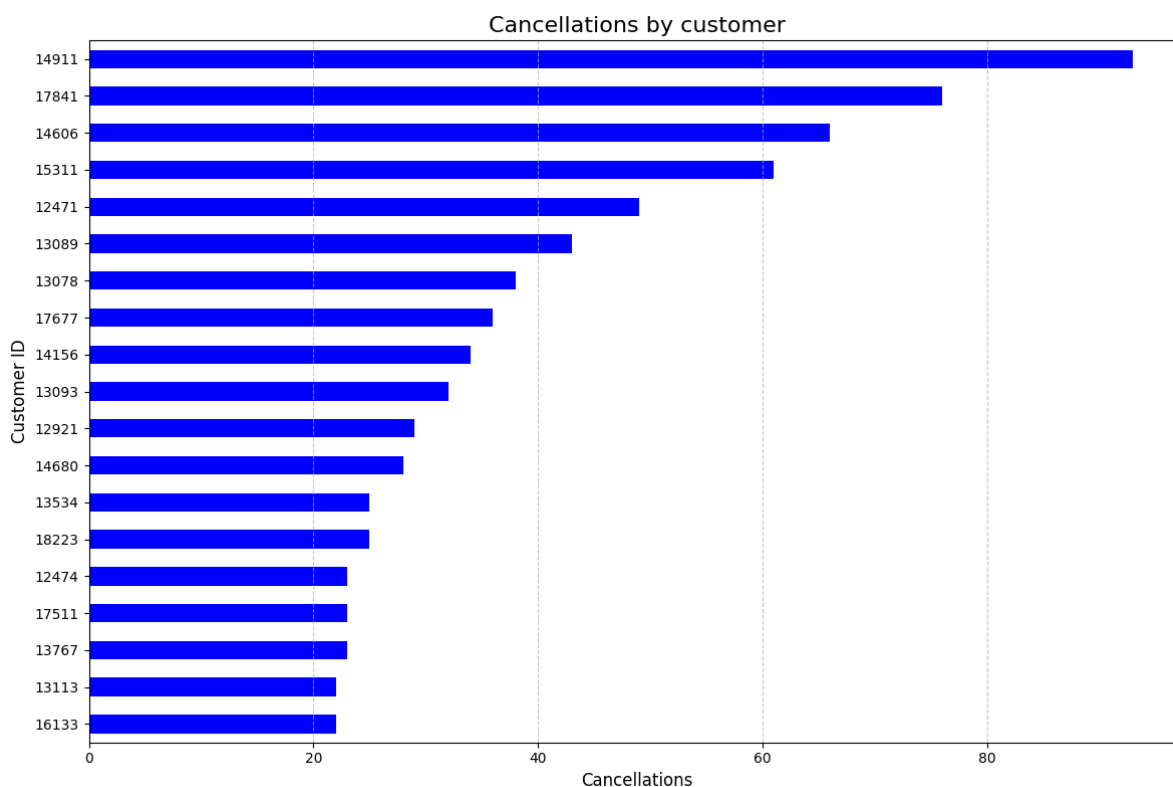
The 20 most cancelling customers are:

Figure 38: Top 20 cancellations Number by Customer

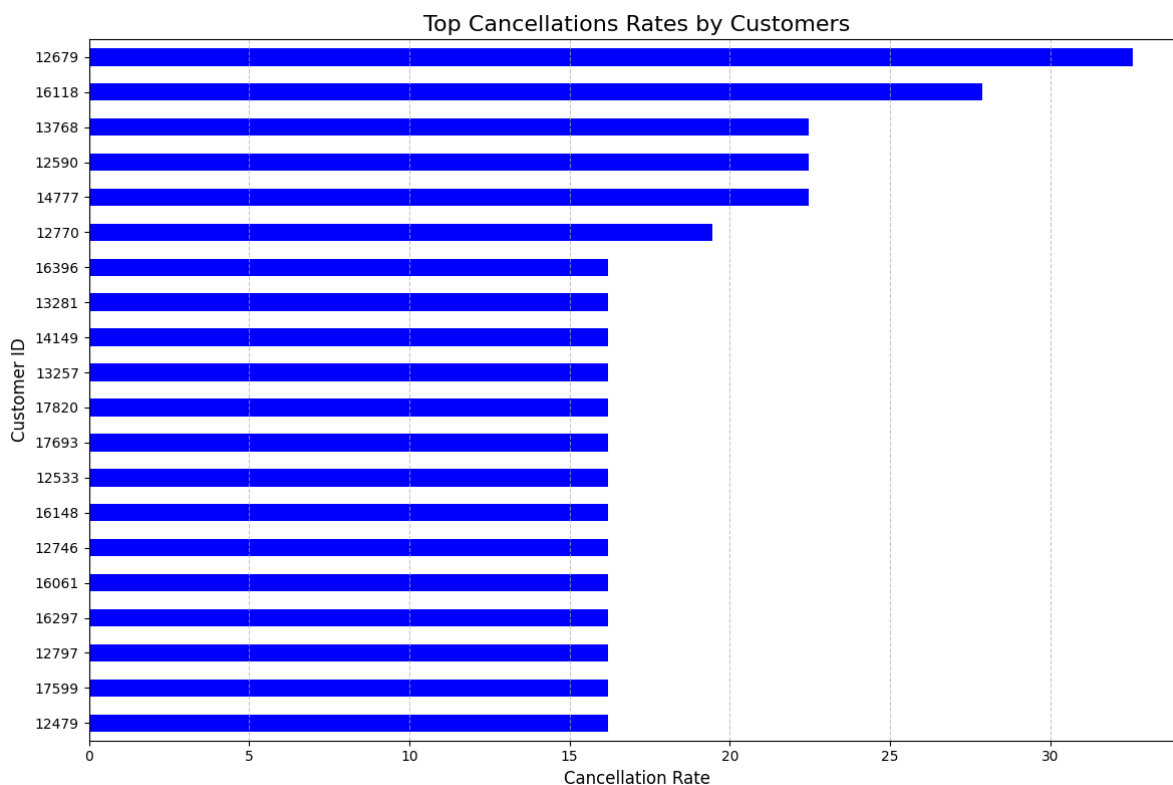Meanwhile the highest cancellation rates by customers are:

Figure 39: Top 20 Cancellations Rates by Customer

The top 20 customers by cancellations show that a few customers, such as Customer ID 14911, have a high number of cancellations. Some high-cancellation customers, like Customer ID 12679, also have notably high cancellation rates, indicating potential issues to address in order to reduce cancellations.

Revenue Loss Due to Cancellations:

The total revenue loss from retail cancellations amounts to £256,097.23, which represents 1.8% of the total retail revenue. While this may seem a modest proportion, it underscores the impact cancellations can have on overall sales performance.
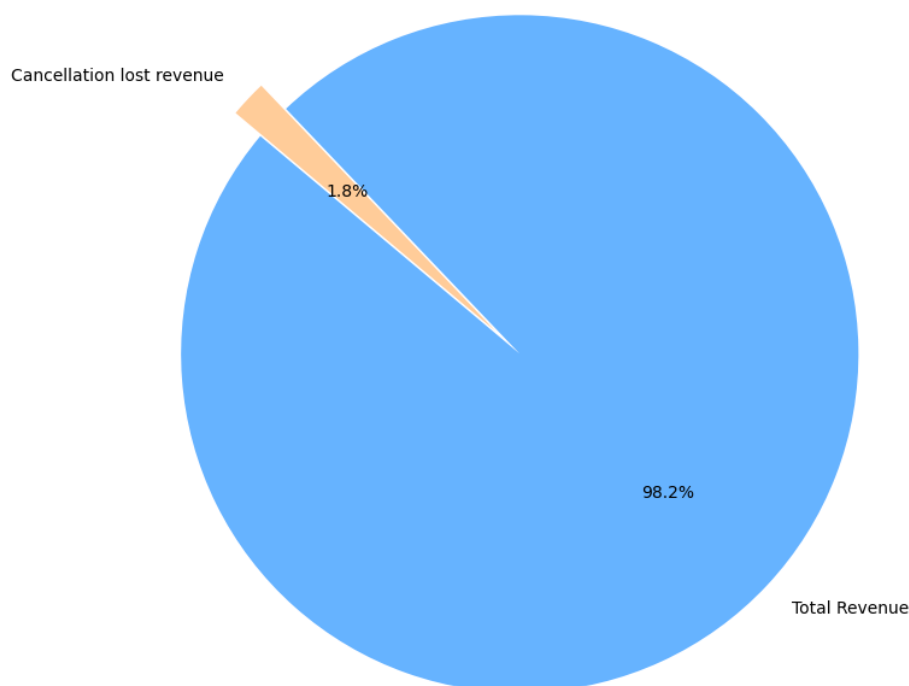
Revenue Breakdown: Total vs. Cancellation loss

Figure 40: Revenue Breakdown: Total vs Cancellation loss

Summary of Retail Cancellation Insights:

The retail cancellations analysis reveals a reasonable revenue loss, with a total of £256,097.23 lost from 17,219 canceled transactions over a two-year period. While the raw cancellation rate is 15.51%, it is adjusted to account for the smaller average basket size of canceled orders (2.39 items) compared to retail orders (24.69 items). This adjustment alongside the revenue loss ratio provides a clearer picture of cancellations' true effects.

Over time, the cancellation rate fluctuated but generally declined.

Top canceled products and customers highlight specific areas that need attention, such as products with exceptionally high cancellation rates like 79340W and 79301. While cancellations account for 1.8% of total retail revenue, addressing the underlying issues could help mitigate further revenue loss.

3.3.2 Bulk

This subsection focuses on bulk cancellations with a quantity of over 50 units per product sold, providing insights into bulk cancellation trends and impact.

Bulk Cancellations Overview:

Covering the period from December 2009 to December 2011, we have:

- **Total Bulk Cancellations loss Revenue**: £456,833.69£

- **Number of Bulk Canceled Transactions**: 631

- **Number of Bulk Canceled Orders(Invoice)**: 275

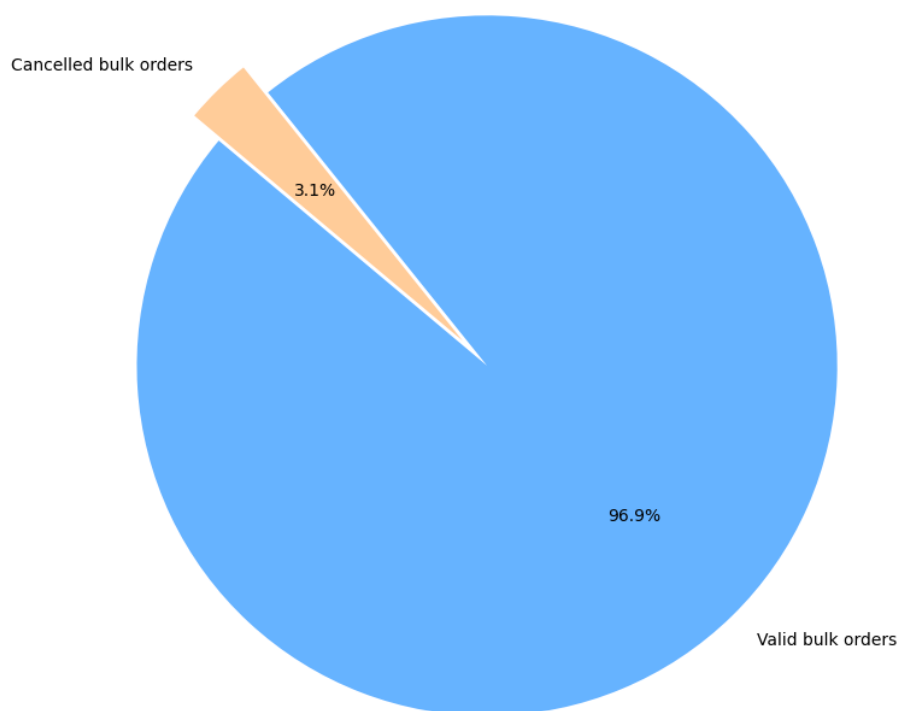**Cancellations vs. Sales(Bulk):**

Figure 41: Invoice Breakdown(Bulk): Sold orders vs Canceled orders

For bulk cancellations, there were 275 cancellations compared to 7,216 in retail, indicating a much lower volume in the bulk category. However, due to the higher quantity sales in bulk, these cancellations can still have a notable impact on revenue.

Cancellation rate(Bulk):

the cancellation rates for bulk orders show a different pattern compared to retail cancellations. The raw cancellation rate for bulk orders is 3.13%, which is lower than the retail cancellation rate of 15.51%. However, when adjusted for transaction size (average

number of items per order), the bulk cancellation rate drops further to 2.29% compared to 1.7%, indicating a more stable pattern when considering the differences in order volume. This suggests that bulk orders may be less impacted by cancellations than retail orders, likely due to the larger basket sizes or different customer behavior, highlighting distinct dynamics in cancellation trends between the two categories.
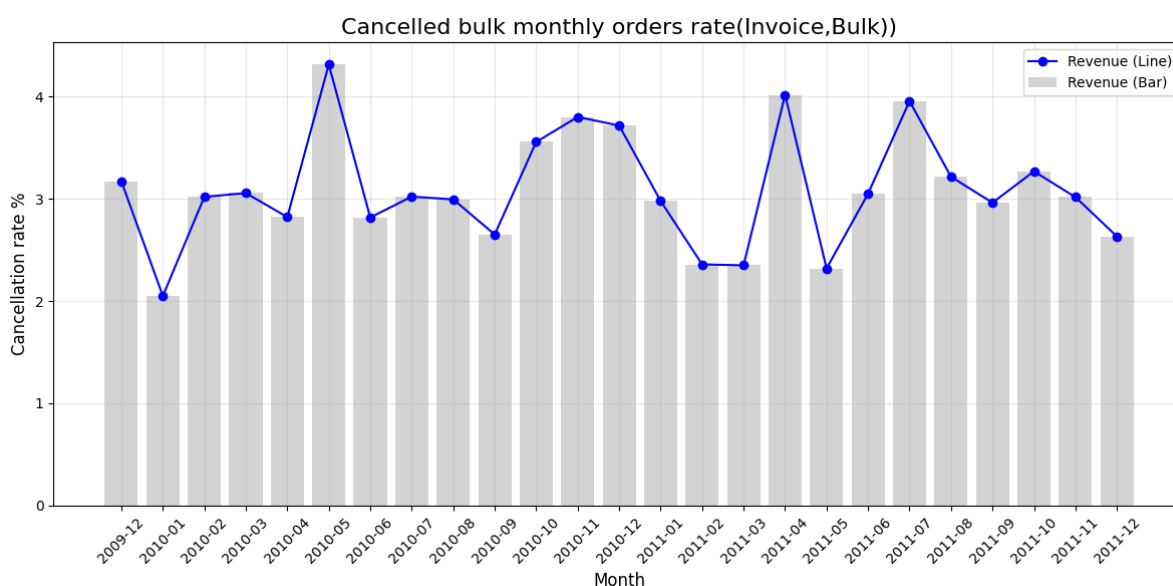
Cancellations Over Time:

MOM:

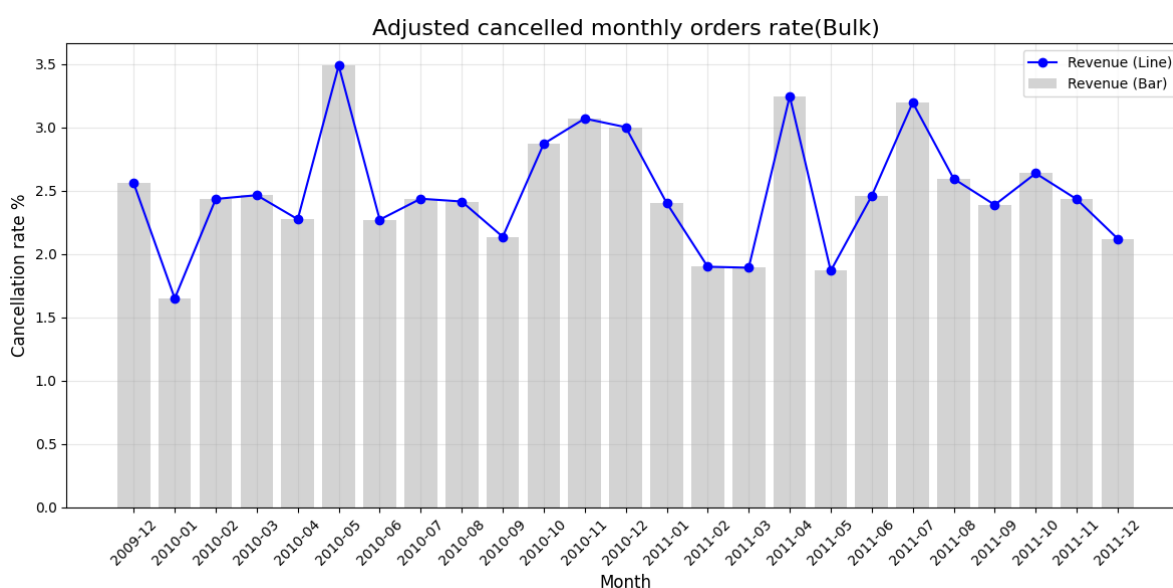Figure 42: Cancelled monthly orders rate(Invoice)

Figure 43: Adjusted cancelled monthly orders rate(Invoice)

The cancellation rates for bulk orders show moderate fluctuations over time, with a raw cancellation rate averaging around 3.13% monthly. These rates peak in May 2010 at 4.31% and fluctuate between 2.35% and 4.02% during 2011. When adjusted for transaction size, the cancellation rates decrease to around 2.29% but the adjustment numbers are not as drastic as retail cancellations rates, indicating that the smaller basket sizes in bulk orders play a significant role in shaping these figures.

QoQ:
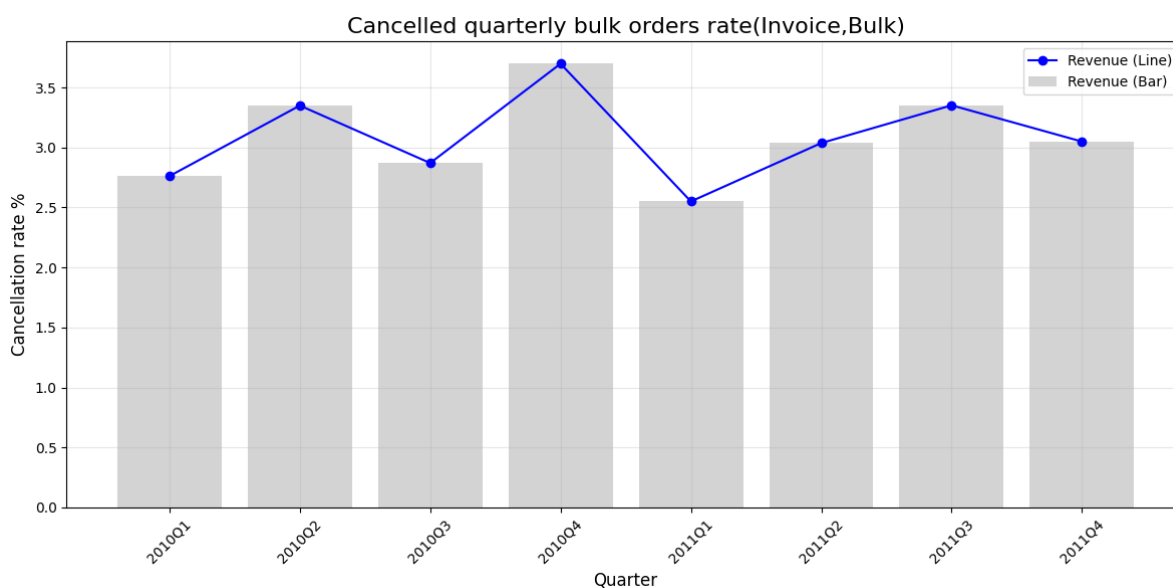
Figure 44: Canceled quarterly bulk orders rate(Invoice,Bulk)

Figure 45: Adjusted canceled monthly orders rate(Invoice,Bulk)

The quarter over quarter bulk cancellation rates are stable, ranging from 2.55% to 3.70%, with minimal fluctuation between quarters. Both raw and adjusted rates are almost similar, indicating little impact from basket size.

Top Canceled Products:

The top 20 canceled bulk products are:
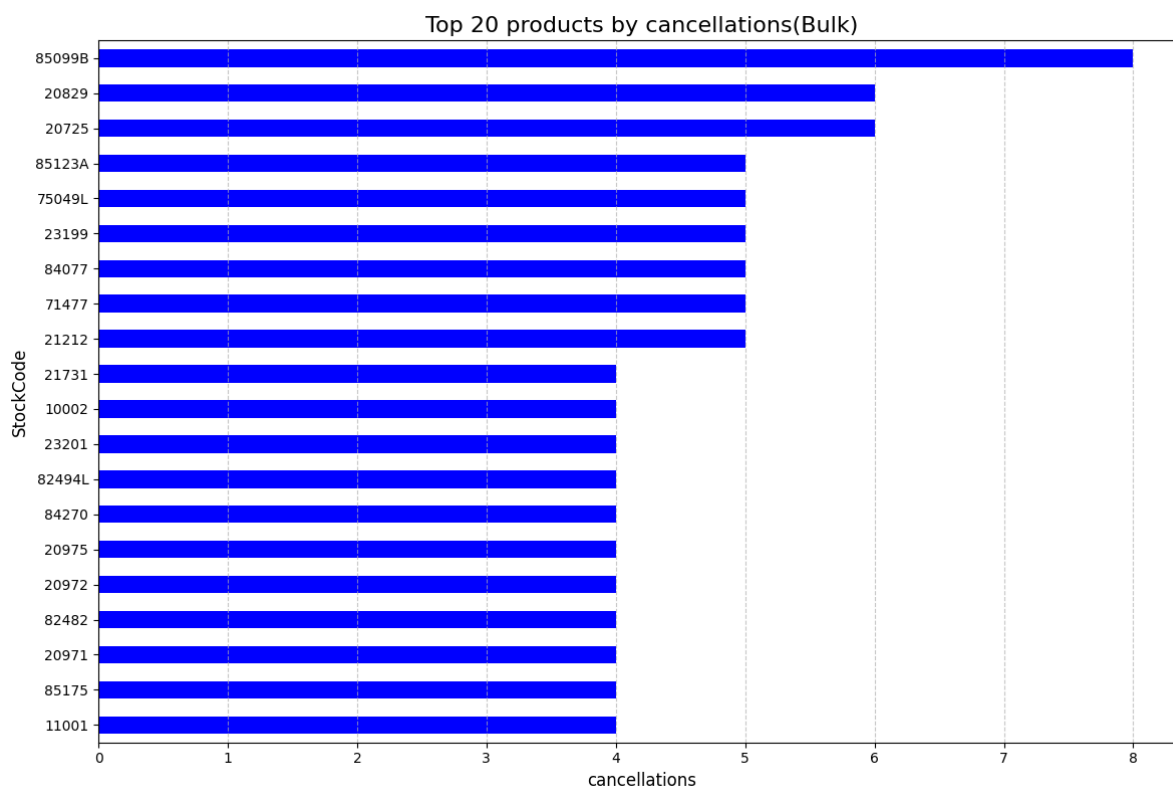
Figure 46: Top 20 Bulk Products by Number of Cancellations

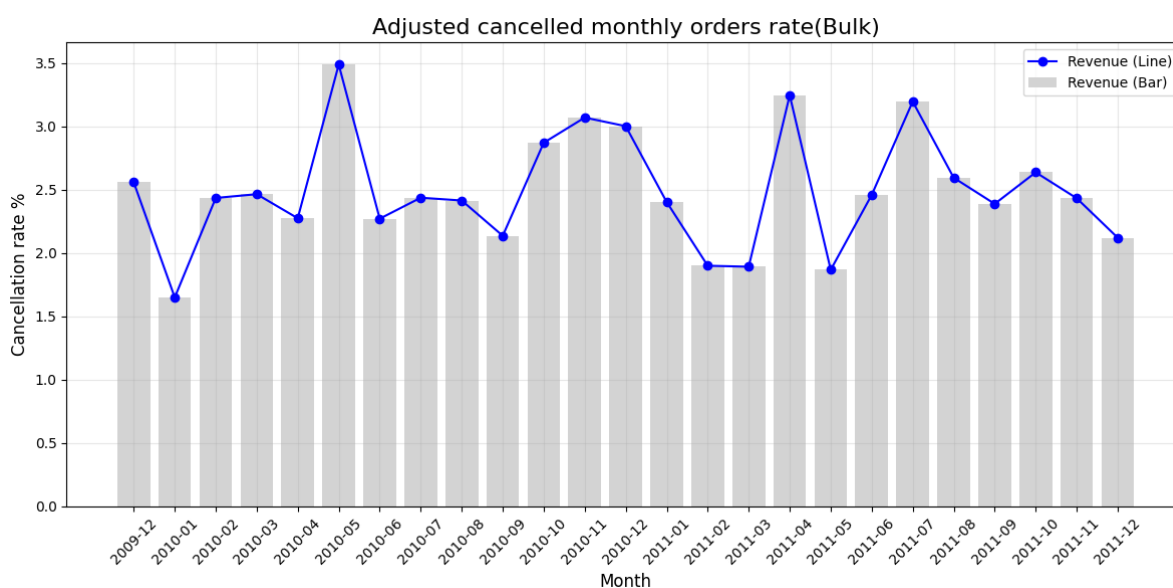Meanwhile the highest 20 cancellation rates in products are:
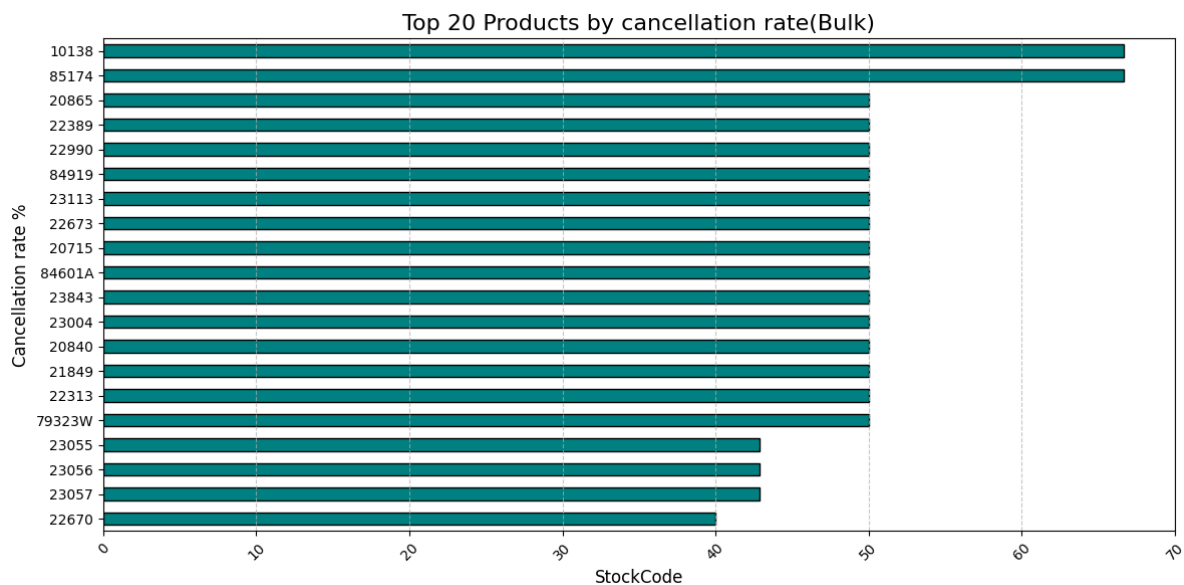
Figure 47: Top 20 Products by Cancellation Rate(Bulk)

Some items like 10138 or 85174 (66% cancellation rate) show high cancellation rates, signaling potential issues with these products.

Most Canceling Customers:

The 20 most canceling bulk customers are:

Figure 48: Top 20 cancellations Number by Customer(Bulk)

Meanwhile the highest cancellation rates by bulk customers are:
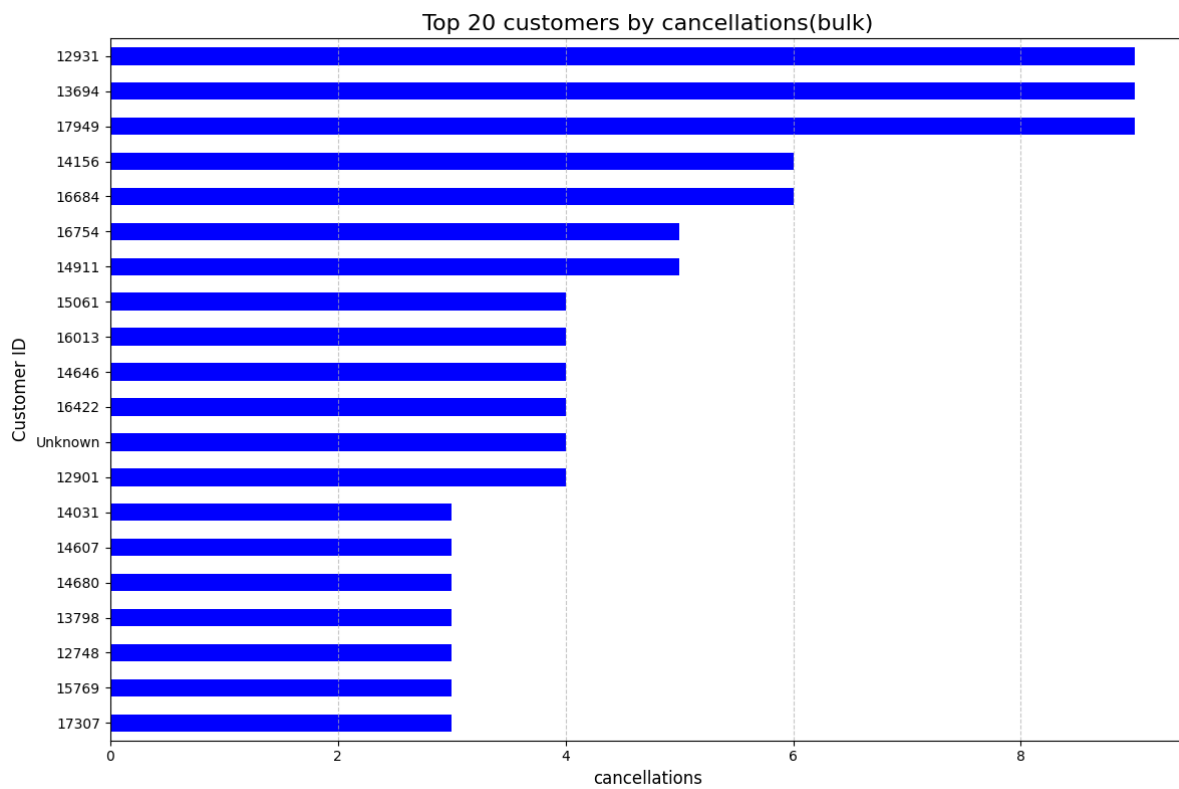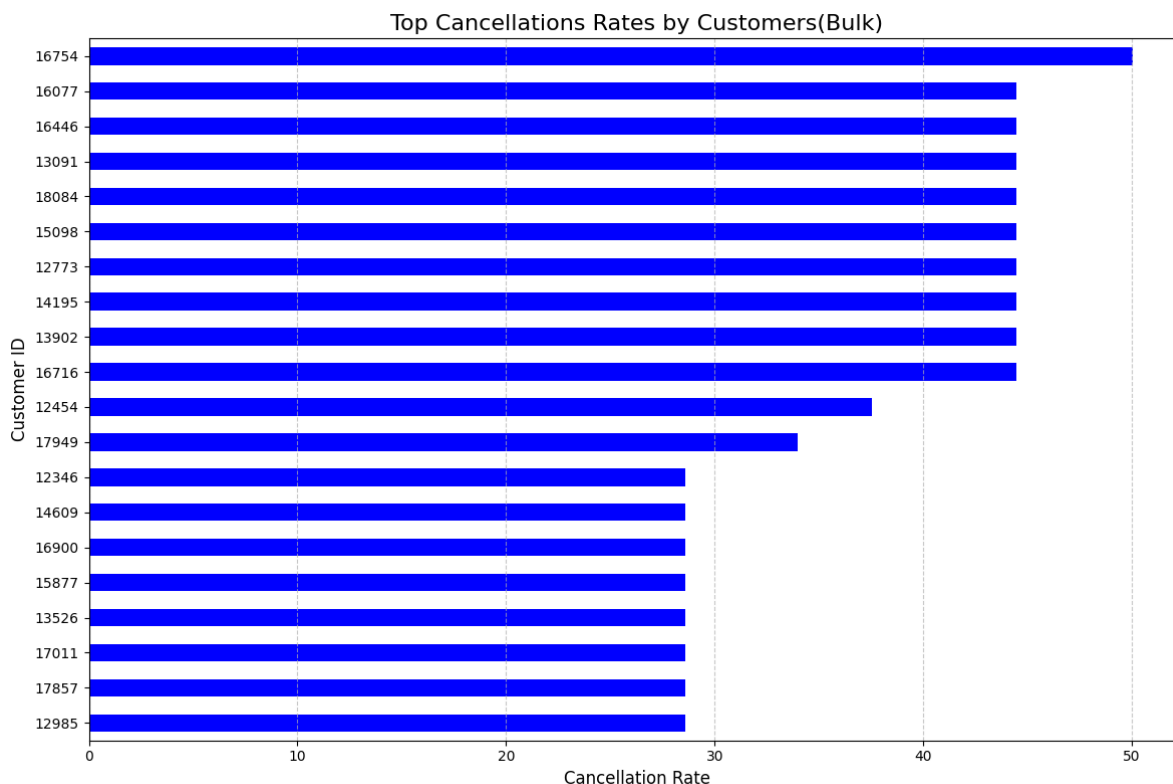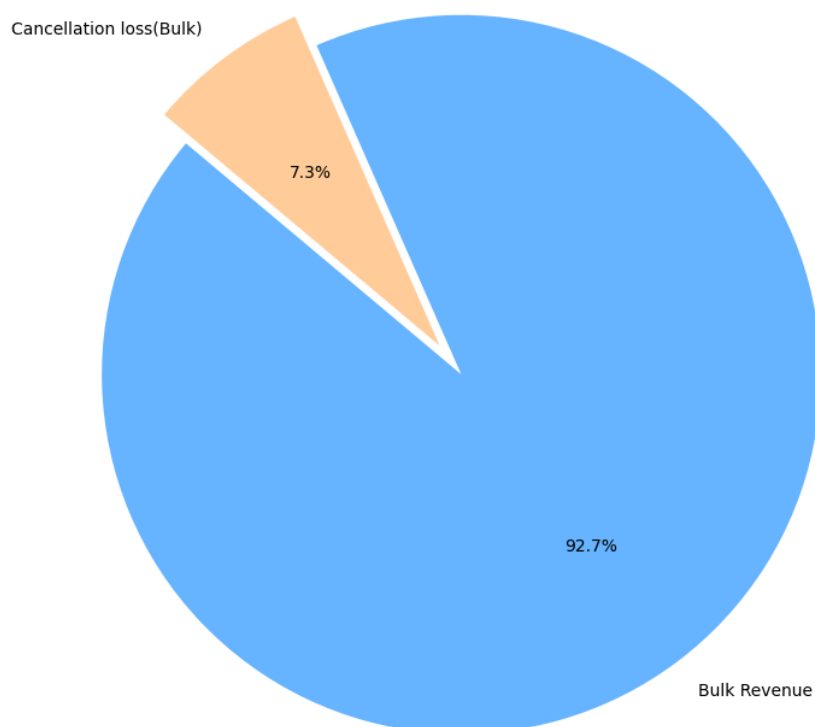

Figure 49: Top 20 Cancellations Rates by Customer(Bulk)

The top cancellations by customers show a few customers with noticeably high rates, especially with figures above 40%, indicating frequent cancellations by some particular customers.

Revenue Loss Due to Cancellations:

Revenue Breakdown(Bulk): Sales vs. Cancellation loss

Figure 50: Revenue Breakdown(Bulk): Total vs Cancellation loss

The revenue loss from bulk cancellations amounts to £456,833.69, which represents 7.3% of the total bulk revenue (£5,802,651.11). This indicates that although the overall cancellation rate is relatively low, it still has a noticeable impact on revenue, reflecting a significant cost to the business, especially considering the higher average order value in bulk transactions.

Summary of Bulk Cancellation Insights:

Bulk cancellations accounted for £456,833.69 in lost revenue, representing 7.3% of the total bulk revenue of £5,802,651.11. While cancellations in bulk are lower in volume (275 cancellations vs. 7,216 in retail), the larger average order value means cancellations still have a substantially higher financial impact.

The cancellation rate for bulk is 3.13% (raw) and 2.29% (adjusted), lower than retail's raw rate 15.51% but higher than the adjusted rate.
Compared to retail, bulk cancellations are less frequent and more manageable, with steady fluctuations and fewer disruptions. The top canceled products and customers indicate certain items and individuals contribute to a higher share of cancellations.

4 Recommendations

Based on the insights derived from the analysis, this section provides actionable recommendations tailored to the observed trends in sales, cancellations, and customer behavior. While these suggestions are grounded in the available data, their implementation should consider broader business processes and additional context.

4.1 Sales Strategy Adjustments

- **Leverage High-Performing products:** Focus efforts on products with the highest sales volume and lower cancellation rates, as identified in the analysis.

- **Sales Trends Monitoring:** Leverage identified patterns, such as seasonal variations and emerging trends, to anticipate challenges, optimize inventory management, and address recurring issues proactively.

- **Capitalize on Bulk Sales:** Given that bulk sales exhibit lower cancellation rates and contribute significantly to revenue, allocate more resources to nurturing bulk buyer relationships.

4.2 Customer Segmentation and Retention

- **Tailor segmentation strategies:** Use the identified customer segments (e.g.,frequent cancelers) to design strategies that address potential reasons for cancellations and reduce churn.

- **Reward Top Purchasers:** Engage top retail customers with exclusive offers and loyalty programs to foster long-term relationships.

- **Incentivize Bulk Buyers:** Provide tailored incentives for bulk buyers to encourage repeat orders and strengthen partnerships.

- **Activate Unknown Customers:** Introduce initiatives to convert unknown customers into identifiable accounts.

4.3 Cancellation Reduction

- **Address High-Cancellation Products:** Investigate products with the highest cancellation rates to identify potential issues such as inaccurate descriptions, quality concerns, or pricing mismatches.

- **Refine Order Processes:** Enhance order handling procedures to minimize cancellations across both retail and bulk transactions.

4.4 Operational and Analytical Enhancements

- **Cancellation Trends Monitoring:** Utilize the identified patterns (e.g., seasonal variations, trends, product-specific issues) to proactively address recurring problems.

- **Data-Driven Decisions:** Develop internal systems to better track and identify customers, enabling better data analytics and strategies.

5   Limitations of the Analysis

1. **Absence of Specific Objectives:** This report was conducted without predefined business goals or contextual guidance. The insights provided are exploratory and designed to serve as a starting point for deeper, more focused analyses.

2. **Lack of External Context:** The analysis is limited to the provided dataset and does not incorporate external market trends, competitor data, or operational factors that may influence the findings.

3. **Exploratory Depth vs. Time Constraints:** While the report covers key areas like sales trends, customer behavior, and cancellations, deeper analyses (e.g., root cause analysis for high-cancellation products or predictive modeling for future sales) were not undertaken due to the exploratory nature of the task and the time required for such analyses.

4. **Data Limitations:** Issues like missing customer identification and product-specific details limit the ability to make certain insights more precise. Additionally, the dataset may not fully capture operational nuances, such as supply chain delays or marketing efforts, that could impact trends.

5. **Static Analysis:** The report is based on historical data without dynamic tools (e.g., dashboards) for ongoing monitoring or real-time analysis.

6   Conclusion

This report analyzed the Online Retail II dataset to uncover key trends in sales performance, customer behavior, and cancellations. The findings highlighted significant differences between retail and bulk transactions, with bulk sales demonstrating higher stability and a significant revenue contribution despite lower order volumes. Retail sales, on the other hand, contributed the largest share of revenue but also exhibited higher volatility, seasonality and a greater cancellation rate.

The customer analysis underscored the importance of understanding diverse customer segments, with opportunities to foster loyalty among top purchasers and address frequent cancelers. Cancellation trends provided valuable insights into product-specific and customer-driven issues, revealing actionable opportunities for process optimization.

Based on these insights, recommendations were proposed to enhance sales strategies, reduce cancellations, and improve customer retention. While these suggestions are grounded in the available data, further context about the company's operational processes could refine these strategies.

Overall, this report underscores the value of data-driven decision-making in addressing operational challenges and identifying growth opportunities. Future analyses could explore complementary datasets or delve deeper into specific business areas to provide even more targeted recommendations.

7 References

- UCI Machine Learning Repository. (2019). *Online Retail II Dataset.* Retrieved from `https://archive.ics.uci.edu/ml/datasets/Online+Retail+II`

- The processed datasets used for this analysis, including cleaned and segmented data, are available for review and replication. Access the repository here: `https://github.com/username/processed-data`.