Written Report

Author: Huyen Phan

Date: 01/11/23

1 - Initial Overview

Retail Data is the chosen dataset in the scope of this project.

The dataset has totally 8 columns and 541.909 rows with the descriptive details as belows:

Column	Description	Datatype_Excel	Datatype_SQL	Datatype_PowerBI
InvoiceNo: Categorial	a 6-digit integral number uniquely assigned to each transaction. If this code starts with letter 'c', it indicates a cancellation	TEXT	NVARCHAR	TEXT
StockCode: Categorial	a 5-digit integral number uniquely assigned to each distinct product	TEXT	NVARCHAR	TEXT
Description: Categorial	product name	TEXT	NVARCHAR	TEXT
Quantity: Integer	the quantities of each product (item) per transaction	NUMBER	INTEGER	WHOLE NUMBER
InvoiceDate: Date	the day and time when each transaction was generated	CUSTOM	DATETIME	DATE/TIME
UnitPrice: Continuos	product price per unit	NUMBER	FLOAT	DECIMAL
CustomerID: Categorial	a 5-digit integral number uniquely assigned to each customer	TEXT	NVARCHAR	TEXT
Country: Categorial	the name of the country where each customer resides	TEXT	NVARCHAR	TEXT

Table 1 - Data Type Overview

2 - Project Objective

The author will play as a data analyst role, who will process the dataset from start to end to

- provide valuable Power BI dashboard serving for strategic purpose of the business.
- propose the most appropriated strategies/ recommendations to boost the company growth.

3 - Data Wrangling

At this stage, an analysis of a transactional dataset from a non-retail UK company called UCI, conducted using SQL queries, revealed that 96.86% of the rows contained sales data (524,917 rows), 1.71% of the rows contained cancellation data (9,251 rows), 0.97% of the rows were duplicates (5,268 rows), 0.46% of the rows were test cases (2,470 rows), and 3 rows contained vague values labeled "Adjust bad debt." The data was collected from December 1, 2010 to December 9, 2011.

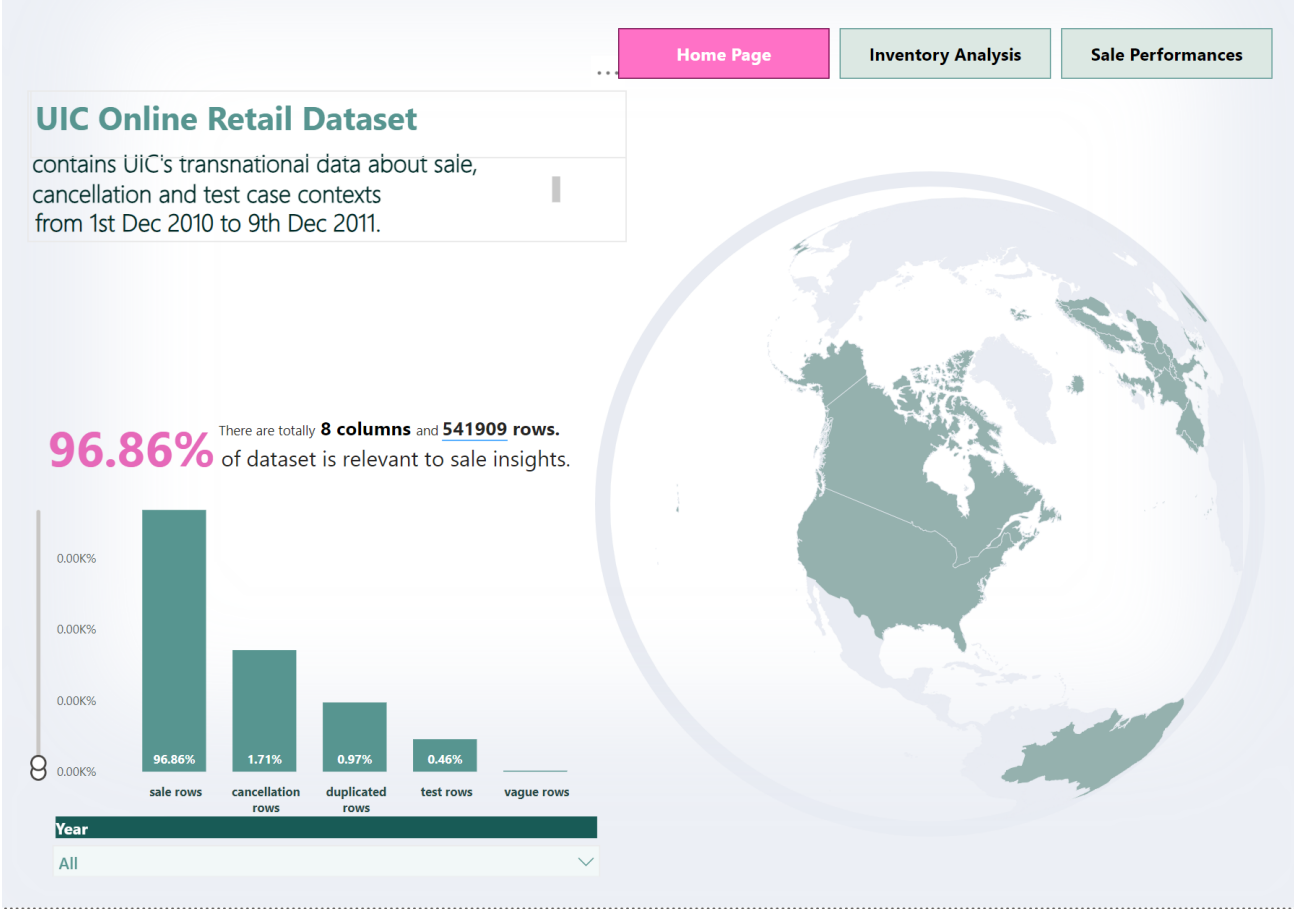

Figure 1 - HomePage Dashboard, which summarize the result from data cleaning process

To categorize the whole data set into 5 groups including sales, cancellation, dupplicated, test case and vague rows, the DA went through the following steps:

- Review the raw dataset from Excel format to get more farmiliar with the dataset and take note any points of the weirdness, such as data type issue, missing data, empty

cells, contaminated data, invalid data. This step can be done by using Filter and Comment functions on Excel.
- Set the right data type for each column from Excel.
- Import the data excel file to SQL after finishing the Excel review and set up stage.
- Conduct an exploratory data analysis (EDA) in SQL based on the key notes beforehands at the Excel review stage.
- Conclude and group the main categories of the dataset: sale_view and cancellation_view, which are excluded duplicated rows, test case rows and vague rows.
- Decide next steps to process for each data category.

You can locate the SQL script of the initial data cleaning process here.

Since the initial result from this EDA stage indicate that the dataset mainly involve to sale and cancelled rows, the DA can shape the main research and sub research questions of this project in relation to inventory management and sale/ marketing promotions in the next part.

4 - Main Reseach Question and Sub-questions

4.1 Main Reseach Question

Which business strategies can be intepreted from this dataset in the most benefitial ways to support UCI - an online retail non-store UK company in promote its business growth in the next year?

4.2 Sub questions

- S1 - How can we optimize inventory management strategy to meet demand without running out of stock or carrying excess supply in the next year?
- S2 - How can we optimize marketing and sale strategy in 2012 based on the sales insights from December 1, 2010 to December 9, 2011 ?

5 - Findings and Discussion

So the descriptive data analysis will go through the break down idea of each sub research question to develop further the final answer for each sub question.

5.1 - Findings

In one way, based on the available features from the sale and cancellation rows, including InvoiceNo, StockCode, Description, Quantity, InvoiceDate, UnitPrice, CustomerID, Country; and the theory of inventory management (citation), the author can conduct an inventory analysis to evaluate the quantity of reorder on each product and the frequency of those reorders.

In this inventory analysis, some new features were identified to gain the target purpose (evaluate the quantity of reorder on each product and the frequency of those reorders), as follows:

- “Demand” of each StockCode
- “MeanDemand” of each StockCode

- “Demand per Quarter” of each StockCode
- “StvDevDemand” of each StockCode
- “Revenue” of each StockCode
- “RevenuePercentitle” of each StockCode
- “ABCCategory” of each StockCode
- “ServiceLevel” of each StockCode
- “Zscore” of each StockCode
- “SafetyStock” of each StockCode
- “ReorderPoint” of each StockCode
- “ReorderQuantity” of each StockCode
- “ReorderFrequency” of each StockCode

In another way, the available features in the sale_view can help the author build a sale performance analysis to propose possible sale and marketing strategies iin the next year to boost the company growth.

> In this sale performance analysis, a new dim_date table with new columns including “dow” (date of week) and “hour” was created to help the author gain more insight of purchased behaviour by different dimensions of time.

The EDA SQL query of this analysis can be found here.

Then two Power BI dashboards, including Inventory Management Dashboard and Sale Performance Dashboard were built from those Inventory Analysis and Sale Analysis to help the author gain more valueable insights in relation to propose inventory managment strateggy and sale/markeing promotion on each product, which boost company growth in the next year.

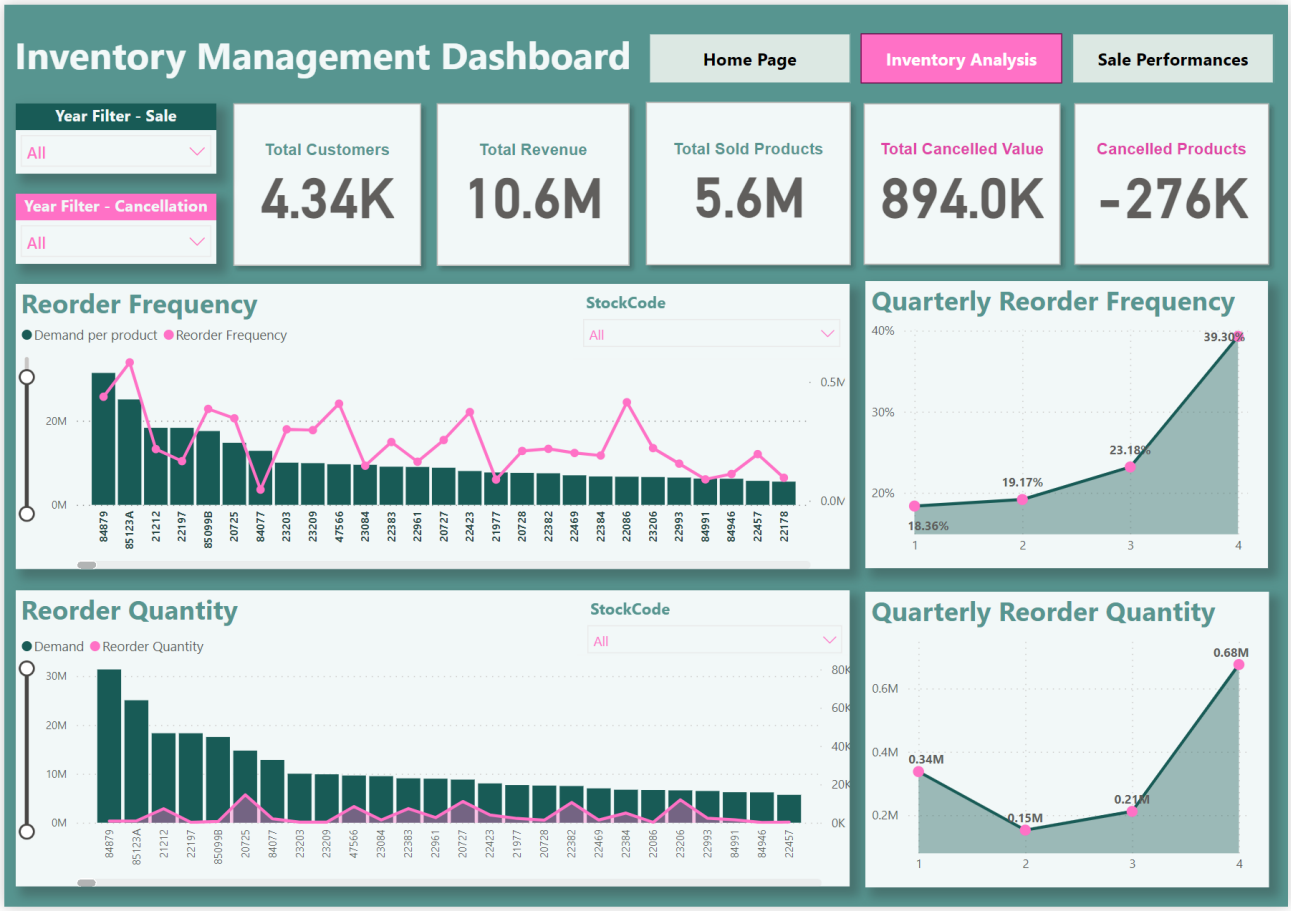

Figure 2 - Inventory Management Dashboard, which visualize the evaluation of the reorder frequency and reorder quantity of each product on each quarter of the period from December 1, 2010 to December 9, 2011

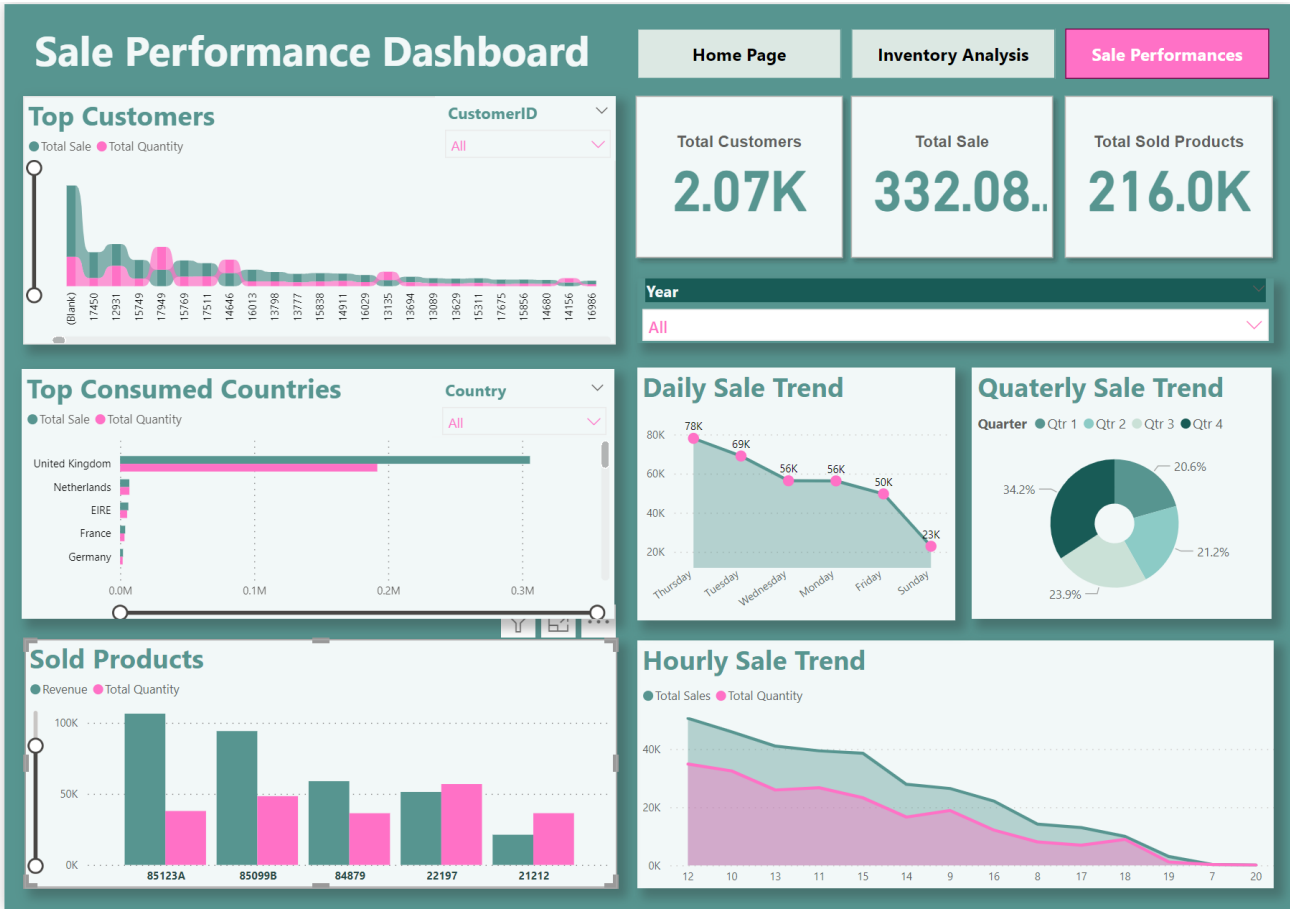

Figure 3 - Sale Performance Dashboard, which visualize the purchase behaviours of customer during different dimissions of time series from December 1, 2010 to December 9, 2011

5.2 Discussion

As you can easily observe from these two dashboards, the insights obtained can be varied and complex, depending on each specific StockCode. In this report, the author will discuss small examples of one or a few products and analyze those specific StockCodes according to the given sub-research questions to explicitly explain how the analysis and dashboards are beneficial to business growth.

Let's look at top 5 products have the highenest numbers of demand with the details as the following table:

StockCode	Demand per product	Product Name	Reorder Frequency	Revenue	Sold Quantity
84879	*31355844*	*ASSORTED COLOUR BIRD ORNAMENT*	*437276*	*58927.62*	*36362*
85123A	*25055541*	*CREAM HANGING HEART T-LIGHT HOLDER*	*581777*	*106415.23*	*37933*
21212	*18309225*	*PACK OF 72 RETROSPOT CAKE CASES*	*217095*	*21246.45*	*36396*
22197	*18301666*	*POPCORN HOLDER*	*166447*	*51334.47*	*56898*
85099B	*17542496*	*JUMBO BAG RED RETROSPOT*	*386087*	*94159.81*	*48371*

Table 2 - Top 5 products have the highenest numbers of demand by quarter from December 1, 2010 to December 9, 2011

In the context of the top 5 products, it's crucial to note the distinction between demand and the sold revenue or sold quantity of a product. For instance, the StockCode "84879" (ASSORTED COLOUR BIRD ORNAMENT) stands out as the highest demanding product, whereas StockCode "85123A" (CREAM HANGING T-LIGHT ORNAMENT) takes the lead in terms of sold revenue, and StockCode "85099B" (JUMBO BAG RED RETROSPOT) leads in sold quantity.

This difference arises because the "Demand" feature considers both Sale_view and Cancellation_view, whereas the other metrics only take Sale_view into account. This variation is essential for accurate calculations of features such as Reorder Frequency and Reorder Quantity. For example, a product may generate high revenue but also have a considerable number of cancellations. In such cases, considering both features becomes imperative for the Data Analyst (DA) to make informed decisions, avoiding over- or under-stock situations. The incorporation of both Sale_view and Cancellation_view ensures a more precise evaluation of a product's demand, facilitating more reliable conclusions for inventory management strategies.

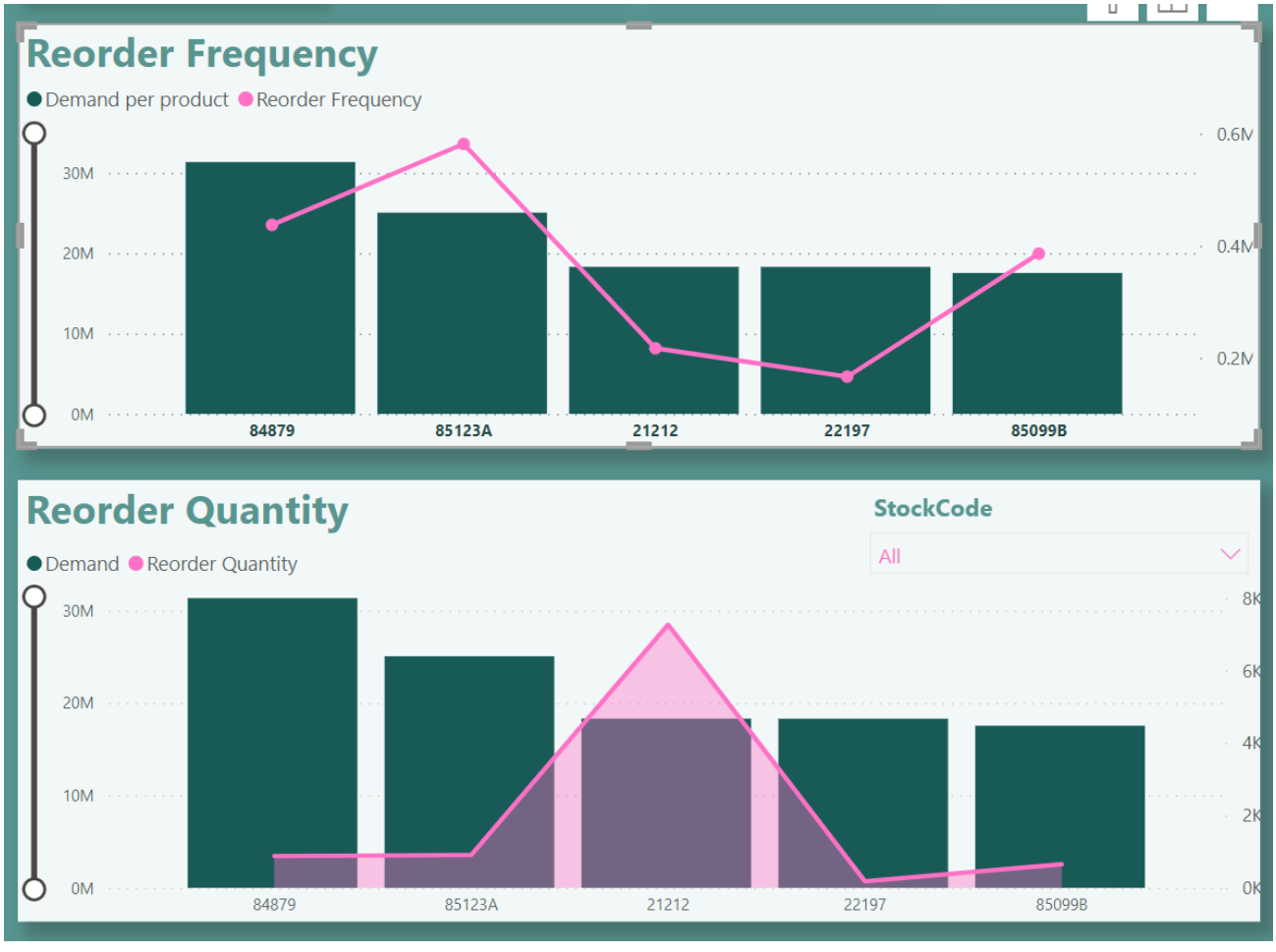

Figure 4 - Top 5 products have the highenest numbers of demand by quarter from December 1, 2010 to December 9, 2011

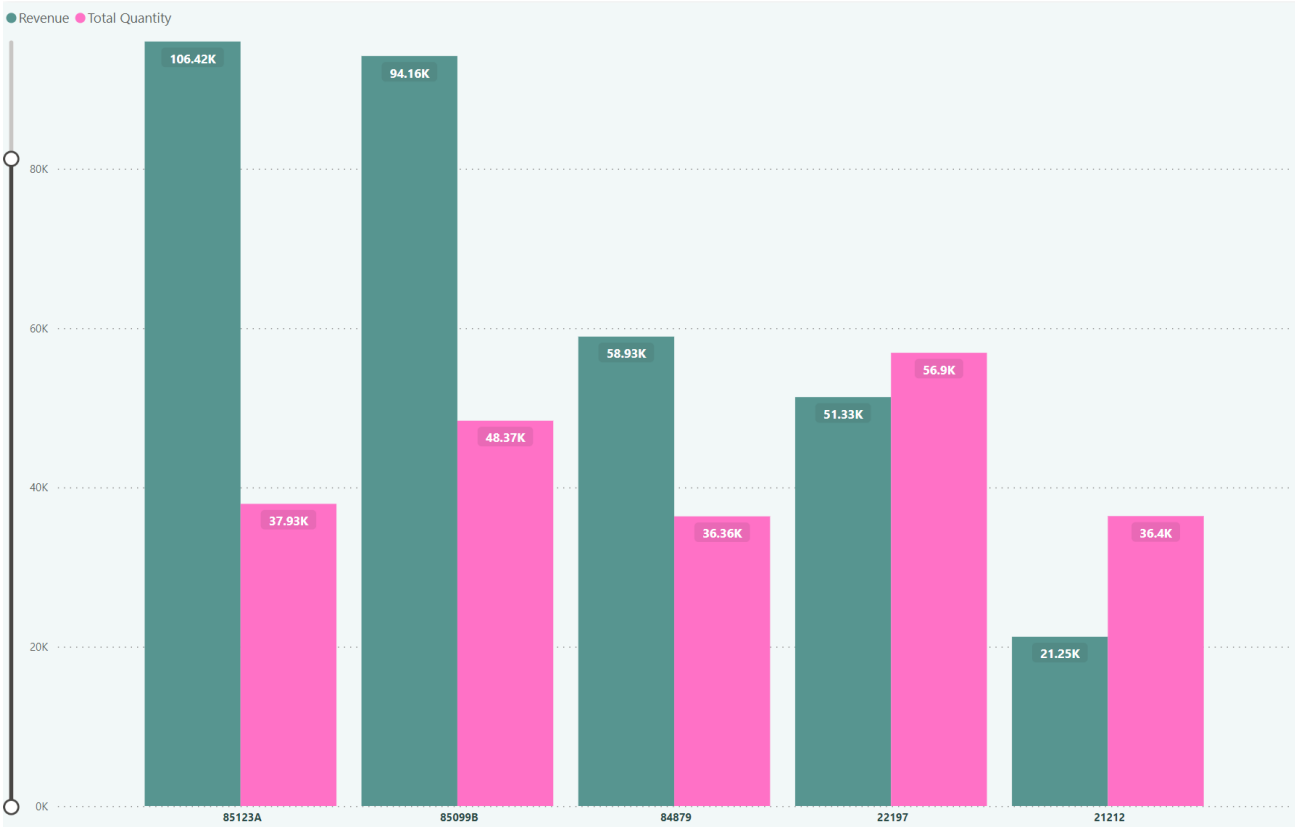

Figure 5 - Top 5 products have the highest numbers of revenue and sold quantity *from December 1, 2010 to December 9, 2011*

- **S1 - How can we optimize inventory strategy to meet demand without overstocking in the next year?**

StockCode "84879" (ASSORTED COLOR BIRD ORNAMENT) was the highest-demanding product during the period of December 1, 2010 and December 9, 2011. Anticipated quarterly reorder quantities in the next year are 297.26 units in quarter 1, 214.25 units in quarter 2, 69.49 units in quarter 3, and 297.26 units in quarter 4.

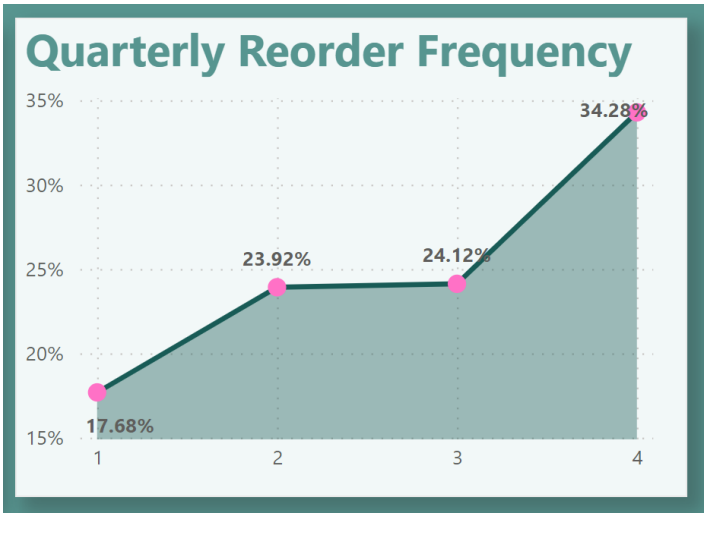

Figure 6 - Quarterly Reorder Frequency of StockCode "84879" (ASSORTED COLOR BIRD ORNAMENT)

These insights suggest that the business should prioritize StockCode "84879" when making inventory management decisions, especially in quarter 1 and quarter 4, when demand is expected to be highest.

Additionally, its reorder frequency percentages was estimated at 17.68%, 23.92%, 24.12% and 34.28%, respectively from quarter 1 to quarter 4. These figures suggest that the business should prioritize StockCode "84879" when placing orders with suppliers. For example, the business may want to place larger orders for StockCode "84879" in quarter 4 to take advantage of any economies of scale. The business may also want to negotiate with suppliers to ensure that it has access to a sufficient supply of StockCode "84879" throughout the year.

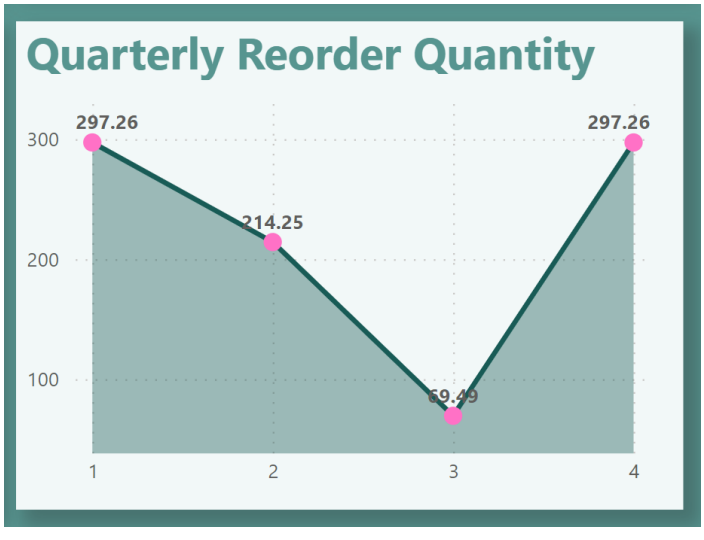

Figure 7 - Quarterly Reorder Quantity of StockCode "84879" (ASSORTED COLOR BIRD ORNAMENT)

- **S2 - How can we optimize marketing and sale strategy?**

Looking at the sale performance of StockCode "84879" (ASSORTED COLOR BIRD ORNAMENT) , total customer is 679 people , total sale (revenue) is 58.93k (unit of money , total sold products is 36.4k units .

Figure 8 - Sale Performance Figures of StockCode "84879" (ASSORTED COLOR BIRD ORNAMENT)

Its quarterly percentages of total sale were accounted for 16.8%, 20.1%, 30.5% and 32.5% from quarter 1 to quarter 4 respectively. Sales are increasing gradually from quarter 1 to

quarter 4, suggesting that the business is experiencing strong growth.

- The business is experiencing strong growth, with sales increasing gradually from quarter 1 to quarter 4. This is a positive sign, and the business should continue to invest in its marketing and sales initiatives to maintain this growth trajectory.
- Quarter 4 is the highest demand period for the business, which suggests that the business should focus its marketing and sales efforts on this quarter to maximize revenue.
- The business is able to generate a significant portion of its revenue in the first three quarters of the year, even though these quarters typically have lower demand. This suggests that the business has a diversified product portfolio and is not overly reliant on any one product or category.

Its daily total sales were reported from Monday to Sunday excepting Saturday at 9K, 9K, 11K, 18K , 7K and 5K, respectively.

- Weekday sales are higher than weekend sales, with the exception of Friday. This suggests that the business's target customers are primarily working adults who are shopping online during their lunch break or after work.
- **Thursday** is the highest sales day, while **Sunday** is the lowest sales day. The business could focus its marketing and sales efforts on Thursday to maximize revenue, and offer discounts or promotions on Sunday to drive sales.

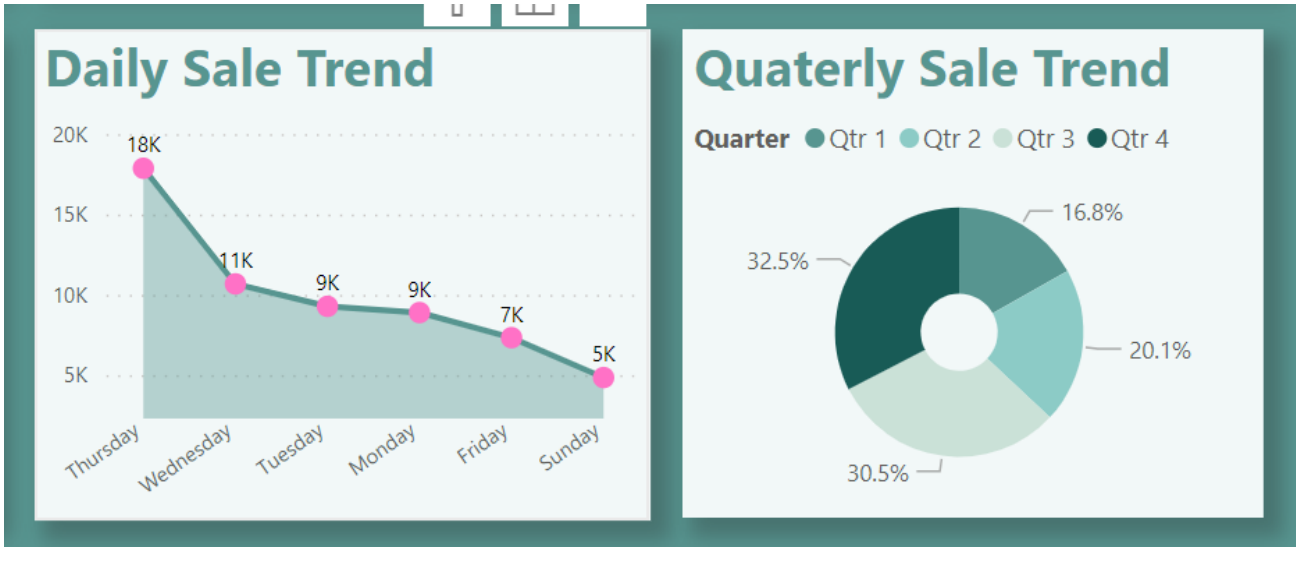

Figure 9: Daily Sale Trend and Quaterly Sale Trend of StockCode "84879" (ASSORTED COLOR BIRD ORNAMENT)

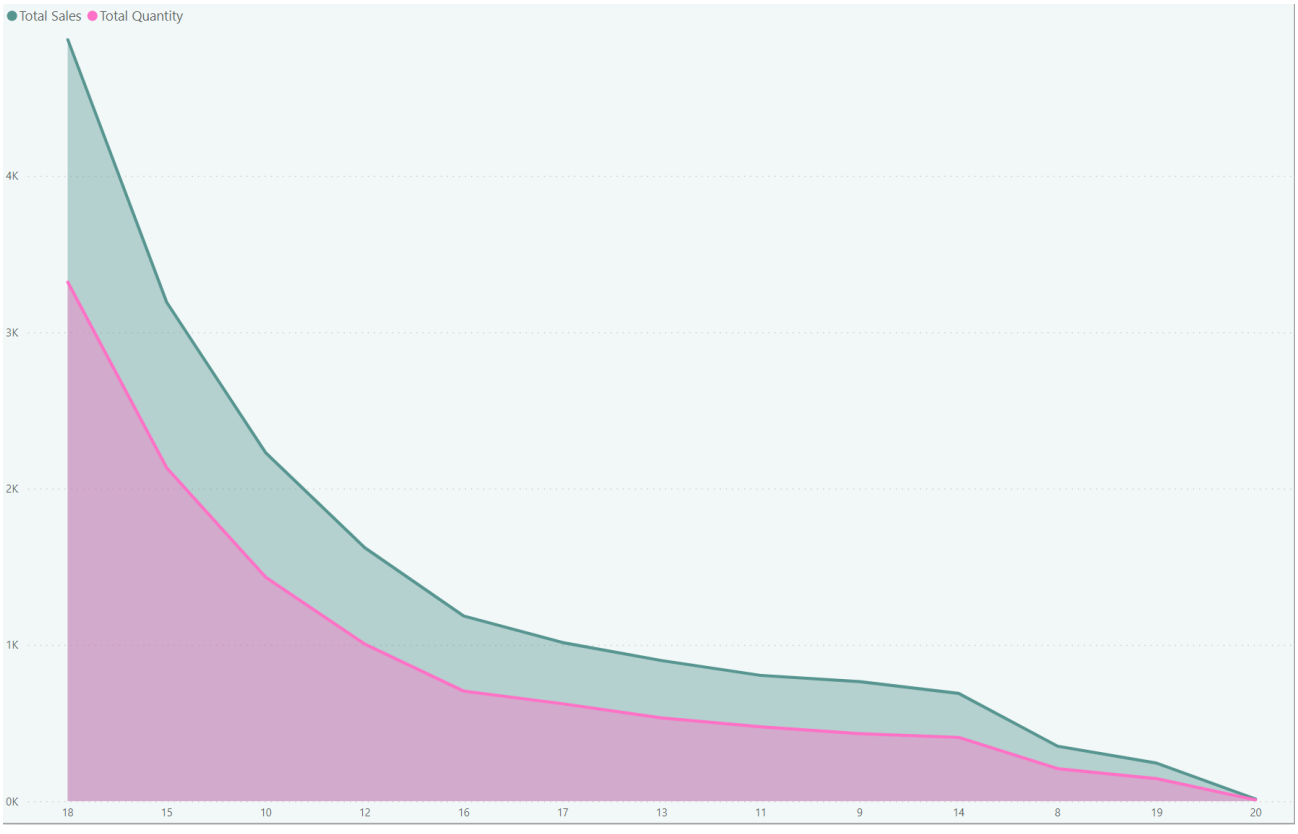

Figure 10 : Hourly sale trend on Thursday of StockCode "84879" (ASSORTED COLOR BIRD ORNAMENT)

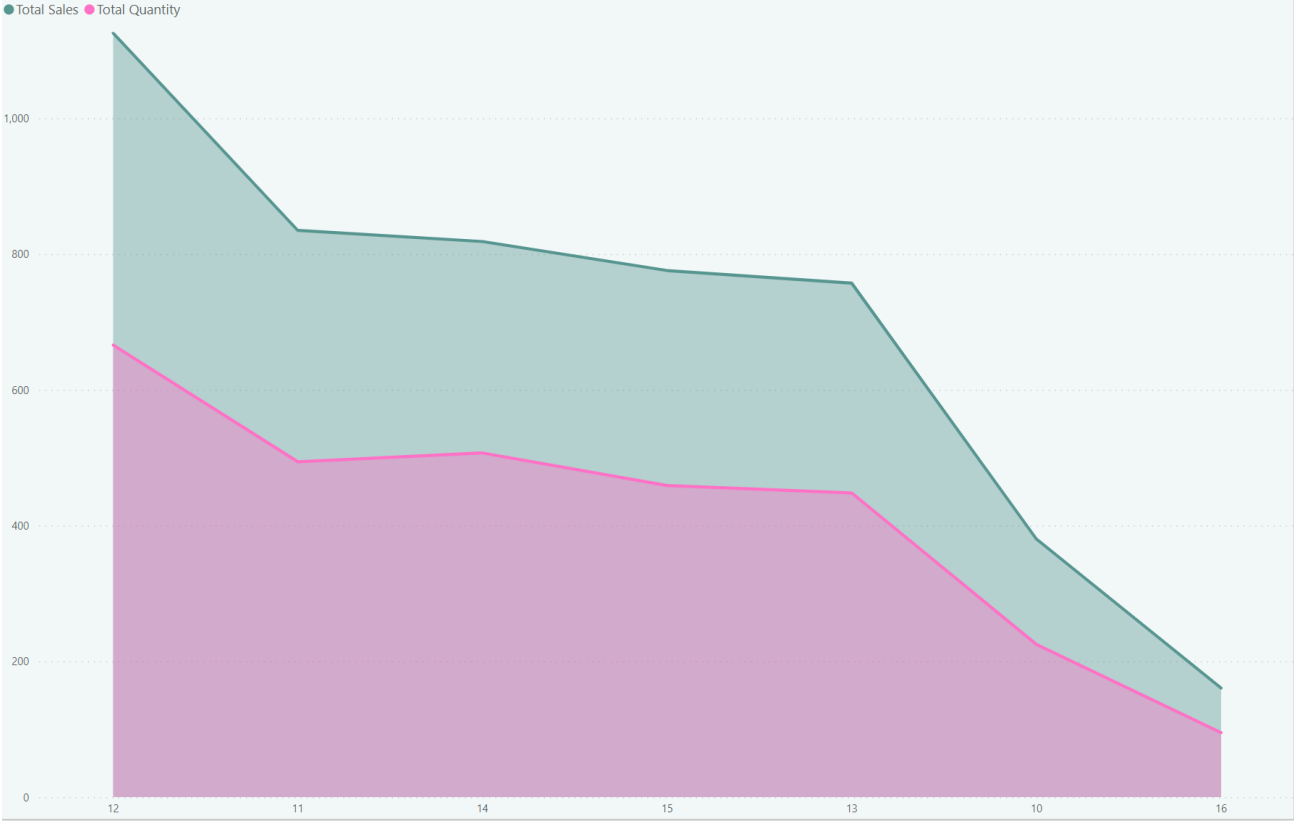

Figure 11: Hourly sale trend on Sunday of StockCode "84879" (ASSORTED COLOR BIRD ORNAMENT)

Looking at the consumtion of StockCode "84879" (ASSORTED COLOR BIRD ORNAMENT) at various countries, the United Kingdom emerges as the clear leader, boasting the highest total sales of $54,589.48 and a substantial quantity sold at 33,636 items, supported by a robust customer base of 643. This indicates a thriving market with widespread customer engagement. In contrast, other European countries such as France, Spain, and EIRE exhibit lower sales figures, suggesting potential areas for improvement or targeted marketing efforts.

To gain a deeper understanding of customer behavior, calculating the average sale per customer for each country is crucial. For instance, the United Kingdom's average sale per customer is approximately $84.95, reflecting a healthy individual spending pattern. Similarly, the average sale per quantity in the United Kingdom is approximately $1.62, providing insights into the value customers place on each item purchased. These metrics can guide strategic decisions, allowing businesses to tailor marketing strategies to different regions effectively.

Examining countries with lower sales, such as Belgium, Canada, Channel Islands, and Italy, offers an opportunity for further investigation. Understanding the reasons behind their comparatively modest performance can unveil potential challenges or untapped markets. Moreover, exploring geographic trends and product preferences in each country provides valuable context for refining marketing approaches.

Additionally, assessing customer concentration is essential to determine if a small number of clients significantly contribute to overall sales. This insight aids in risk mitigation and diversification strategies. Finally, identifying potential growth opportunities involves considering factors like low competition, a growing customer base, or a favorable market environment. By incorporating these analyses, businesses can make informed decisions to optimize their strategies, enhance customer engagement, and tap into new markets, ultimately fostering sustainable growth.

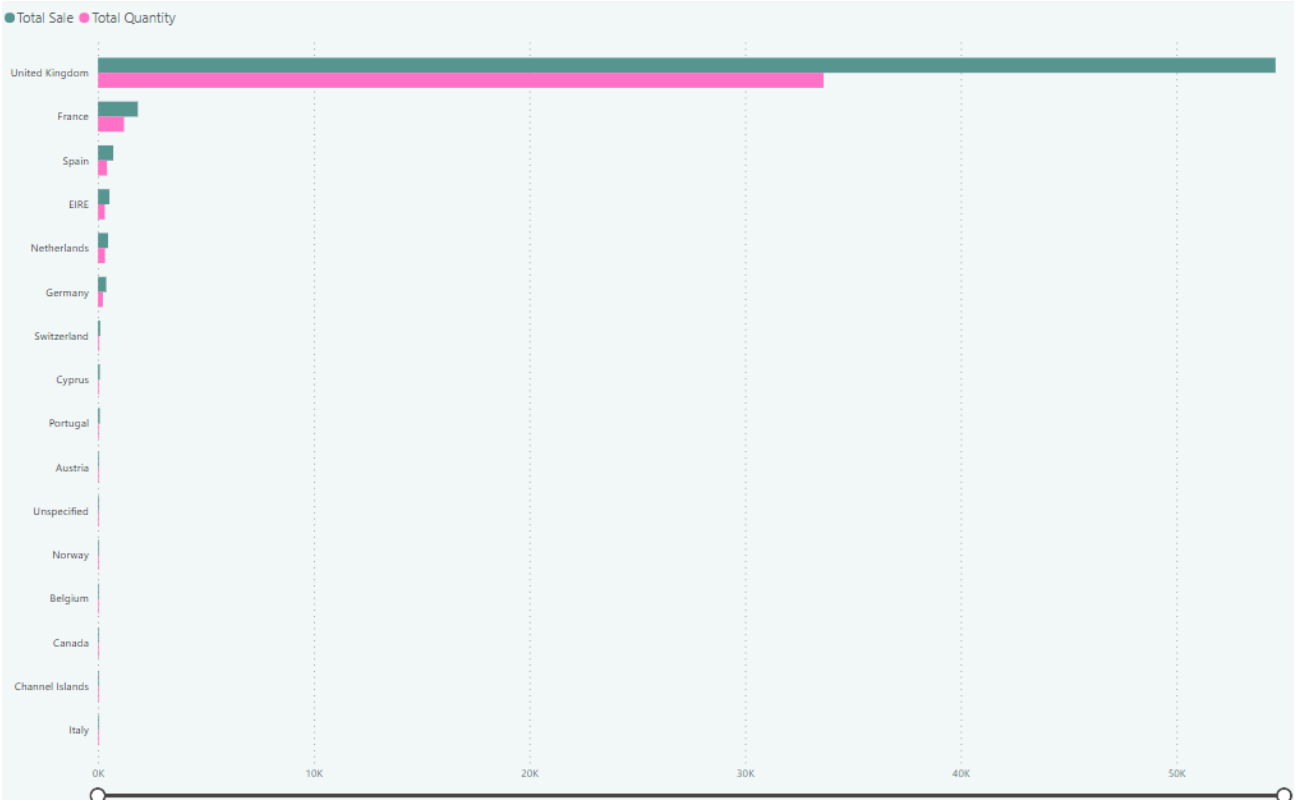

Figure 11: Top consumed contries by revenue and sold quantity

6 - Recommendation

Firstly, based on the **inventory analysis**, here are some inventory strategies that the business could implement:

- **Set higher reorder points for StockCode "84879" in quarter 1 and quarter 4.** This will help to ensure that the business does not run out of stock during these peak demand periods.
- **Allocate more warehouse space to StockCode "84879" than to other products.** This will make it easier to store and pick the product when orders are placed.
- **Consider using a just-in-time inventory system for StockCode "84879" in quarters 2 and 3.** This can help to reduce inventory costs during these periods of lower demand.
- **Monitor inventory levels of StockCode "84879" closely throughout the year.** This will help the business to identify any unexpected changes in demand and make necessary adjustments to its inventory management plans.
- The business could also consider **offering a pre-order option for StockCode "84879" in quarter 4.** This would allow the business to gauge customer demand and ensure that it has enough stock on hand to meet that demand.
- The business could also **partner with other businesses that sell complementary products to StockCode "84879."** For example, the business could partner with a Christmas tree retailer to offer a bundled discount on StockCode "84879" and Christmas trees.
- The business could also **invest in marketing and sales initiatives to increase demand for StockCode "84879" in quarters 2 and 3.** For example, the business could run

targeted ads on social media and search engines, or it could offer discounts or promotions on StockCode "84879" during these quarters.

By taking these actions, the business can improve its inventory management for StockCode "84879" and avoid stockouts, which can lead to lost sales and customer dissatisfaction.

Secondly, based on the **sale performance analysis**, here are some **sale/marketing strategies** that the business could implement especially StockCode :

Focus marketing and sales efforts on quarter 4. This is the highest demand period for the business, so it is important to make sure that the business is visible and accessible to potential customers during this time. The business could run targeted ads on social media and search engines, and offer special discounts or promotions to encourage customers to buy during this quarter.

Invest in marketing and sales initiatives to maintain growth trajectory. The business is experiencing strong growth, but it is important to continue to invest in marketing and sales to maintain this growth trajectory. The business could invest in new marketing channels, such as influencer marketing or video marketing, and it could also expand its sales team to reach more customers.

Diversify product portfolio. The business is already doing a good job of diversifying its product portfolio, but it could continue to do this by adding new products and categories to its offerings. This will help the business to reach a wider range of customers and reduce its reliance on any one product or category.

Target working adults. The business's target customers are primarily working adults, so it is important to focus marketing and sales efforts on reaching this group. The business could run targeted ads during work hours, and it could also partner with businesses that cater to working adults, such as coffee shops or gyms.

Focus marketing and sales efforts on Thursday. Thursday is the highest sales day, so it is important to focus marketing and sales efforts on this day. The business could run targeted ads on Thursday, and it could also offer special discounts or promotions on this day.

Offer discounts or promotions on Sunday. Sunday is the lowest sales day, so the business could offer discounts or promotions on this day to drive sales. The business could also offer free shipping or other incentives to encourage customers to buy on Sunday.

In addition to the above strategies, the business could also use its Sale Performance Dashboard to track its sales performance and identify areas for improvement. For example, the business could use the dashboard to track hourly sales trends and identify which days and times are the most popular for shopping. The business could then use this information to focus its marketing and sales efforts on the most popular days and times.

7 - Limitation

The presence of multiple UnitPrices for products sharing the same StockCode in a dataset contains significant challenges, including inconsistency, ambiguity, and potential inaccuracies. This scenario raises concerns about data quality, hindering reliable analyses and model training. The dataset's ambiguity can lead to confusion regarding the accurate representation of a product's UnitPrice, impacting decision-making processes and creating difficulties in reporting. Inaccurate or varied UnitPrices also pose a threat to the performance of machine learning models, as they may struggle to discern patterns or make accurate predictions in the face of such inconsistencies.

To address these issues, it is crucial to undertake a comprehensive data improvement process. This involves meticulous data cleaning to rectify errors and discrepancies, standardizing the dataset by selecting representative UnitPrices or calculating averages, implementing robust data validation to identify outliers, and thoroughly documenting the entire process for transparency and reproducibility. The significance of these improvements lies in ensuring data integrity, bolstering confidence in decision-making, enhancing model reliability, and ultimately improving operational efficiency. A clean and standardized dataset not only fosters trust in the accuracy of analyses but also streamlines the work of analysts and data scientists, allowing them to focus on extracting meaningful insights rather than dealing with data quality issues.

In essence, the necessity of improving the dataset is rooted in the fundamental principles of data integrity, reliability, and operational efficiency, all of which contribute to informed decision-making and successful data-driven endeavors.

8 - Conclusion

In conclusion, this project centered on the analysis of a retail dataset with the objective of providing valuable insights through Power BI dashboards to strategically support the growth of a non-store UK company, UCI. The data wrangling process involved a meticulous review of the dataset, SQL analysis, and categorization into 5 data groups including sales, cancellations, duplicates, test cases, and vague rows. The main research question focused on identifying beneficial business strategies for UCI's growth in the coming year, with sub-questions addressing inventory management and sales optimization.

The descriptive data analysis delved into inventory management and sales performance, utilizing features such as Demand, MeanDemand, Revenue, and others. Power BI dashboards, including Inventory Management and Sale Performance, were constructed to visualize insights derived from the analysis. The discussion highlighted the importance of considering both sales and cancellation data for precise evaluation, as exemplified by the distinction between demand and sold revenue or quantity for the highest-demand product with StockCode "84879" (ASSORTED COLOR BIRD ORNAMENT) during the period from December 1, 2010, to December 9, 2011. However, the provided dashboards can be used to analyze relevant insights for any products in the dataset, depending on the specific requirements of the business.

Despite the comprehensive analysis, the project acknowledged limitations related to inconsistent UnitPrices for products with the same StockCode in the dataset. To address these challenges, a call for a robust data improvement process was emphasized, involving meticulous cleaning, standardization, validation, and thorough documentation to ensure data integrity, boost decision-making confidence, enhance model reliability, and improve overall operational efficiency.

In essence, this project not only provided actionable recommendations for UCI's growth strategies but also underscored the critical importance of maintaining a clean and standardized dataset for reliable and effective data-driven decision-making in the business context.

9 - References

(1) Plinere, D. and Borisov, A. (2015) *Case study on inventory management improvement*, *Information Technology and Management Science*. Available at: https://doi.org/10.1515/ITMS-2015-0014 (Accessed: 01 November 2023).

(2) Nasim, Shahzad and Maaz, Syed and Ali, Faraz and Khan, Moin, Inventory Management through Lean Logistics and Warehousing Techniques (October 1, 2016). International Journal of Management Sciences and Business Research, Vol. 5, Issue 10, Oct. 2016 , Available at SSRN: https://ssrn.com/abstract=2875164

(3) chrome-extension://cipccbpjpemcnijhjcdjmkjhmhniiick/pdfViewer.html?file=https%3A%2F%2Fwww.jopafl.com%2Fuploads%2Fissue9%2FINVENTORY_MANAGEMENT_SERVICE_LEVEL_AND_SAFETY_STOCK.pdf