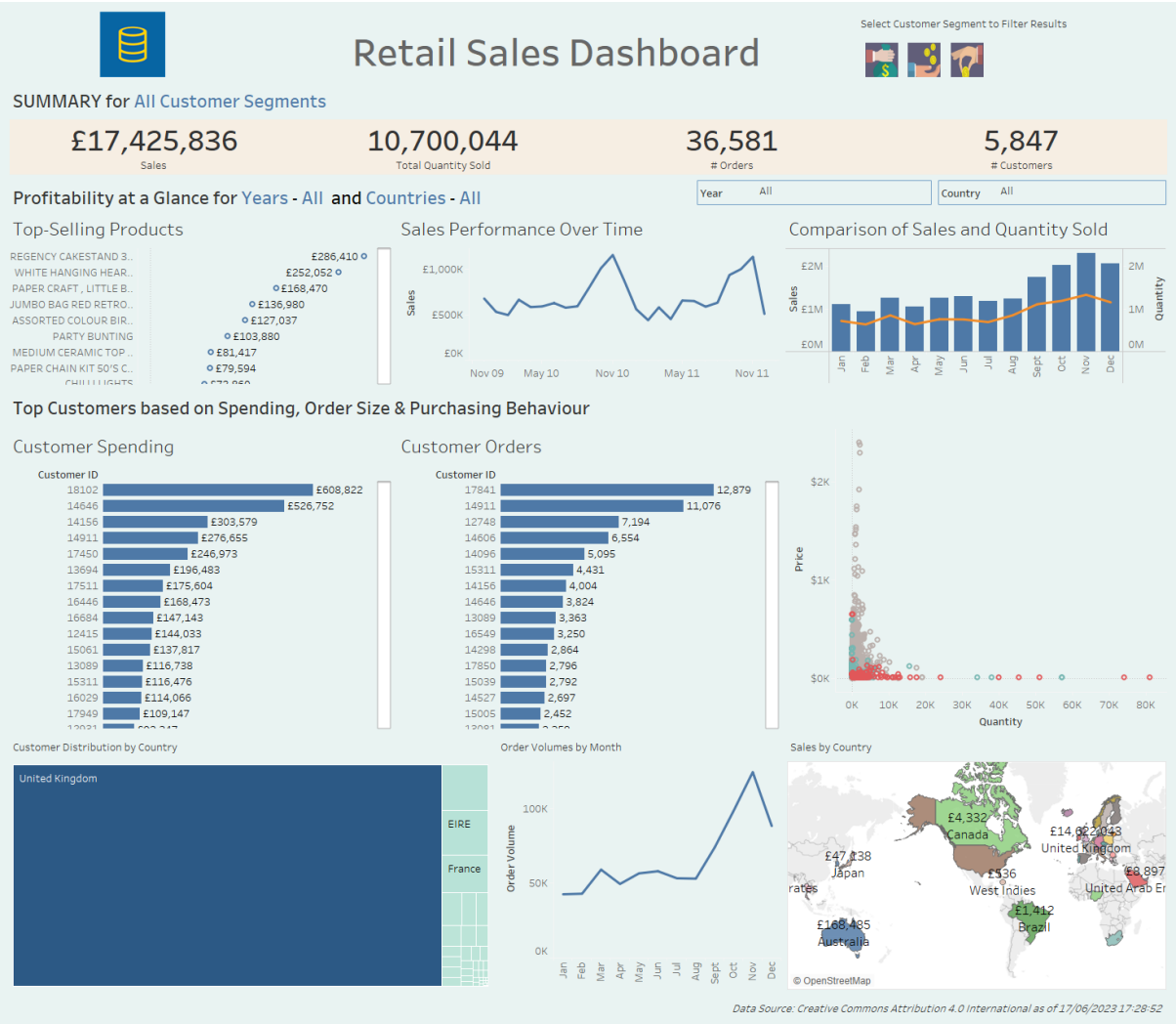
SALES PERFORMANCE ANALYSIS PORTFOLIO PROJECT

Exploring Sales Data to Drive Business Growth and Success



by
Sachin Dahiya
Master of Applied Statistics, Louisiana State University, 2023

Table of Contents

Abstract

This portfolio project focuses on the comprehensive analysis of sales performance to drive business growth and enhance sales effectiveness. The objective is to leverage data insights and statistical techniques to uncover patterns, identify opportunities, and optimise sales strategies. The project employs a combination of quantitative analysis, statistical modelling, and data visualisation to provide valuable insights into sales trends, customer behaviour, and market dynamics.

Through rigorous data exploration, this project explores various dimensions of sales performance, including revenue generation, customer acquisition and retention, product performance, and sales team effectiveness. By analysing historical sales data and applying advanced statistical techniques, patterns and correlations are discovered, enabling the identification of key drivers and predictors of sales success.

The findings and recommendations derived from this project aim to provide valuable insights for businesses seeking to optimise their sales performance, improve decision-making processes, and achieve sustainable growth. By harnessing the power of data and statistical analysis, businesses can gain a competitive edge, adapt to changing market dynamics, and make informed strategic decisions that align with their goals and objectives.

This portfolio project serves as a testament to the author's expertise in sales performance analysis and highlights the practical application of statistical techniques in driving business success. It showcases the author's ability to transform raw data into actionable insights and provides a foundation for continued research and exploration in the field of sales analytics.

Keywords: sales performance analysis, data insights, statistical modelling, predictive analytics, business growth, sales strategies, decision-making, competitive advantage.

1. Introduction

1.1. Background

I have chosen the topic of "Sales Performance Analysis" due to its significance in assessing the financial well-being of a business and gaining valuable insights into customer preferences and market dynamics. The objective of this project is to analyse the sales performance of a company using an online retail dataset. By examining sales trends, conducting product analysis, segmenting customers, analysing geographic patterns, and evaluating sales channels, the project aims to provide actionable insights for optimising sales strategies and driving business growth.

The chosen dataset provides a wealth of information on customer transactions, product details, and demographics, making it an ideal resource for comprehensively understanding different aspects of sales performance.

1.2. Objectives

Throughout this project, I will accomplish the following outcomes:

1. Gain a comprehensive understanding of the company's sales performance and trends.
2. Identify the top-selling products and analyse sales distribution across different product categories.
3. Uncover meaningful customer segments based on purchasing behaviour and assess their contribution to overall sales.
4. Determine the regions or countries that generate the most sales and identify potential marketing opportunities.
5. Provide actionable insights and recommendations to optimise sales strategies and improve business performance.

By conducting an in-depth analysis of the sales performance, this project aims to provide valuable recommendations for enhancing sales effectiveness, targeting specific customer

segments, and maximising revenue. The results will drive data-driven decision-making and contribute to the growth and success of the company.

2. Methodology

2.1. Data Description

The dataset for this project will be sourced from an online retail dataset, the "Online Retail II" dataset available on the UCI Machine Learning Repository or from Kaggle. Specifically, the dataset contains detailed information about customer transactions, product details, customer demographics, and other relevant variables.

The Online Retail II dataset includes all the transactions occurring for a UK-based and registered non-store online retail company between 01/12/2009 and 09/12/2011. The company specialises in selling unique all-occasion gift ware, and it serves both retail customers and wholesalers. The dataset provides a comprehensive view of the company's sales activities and can be used to analyse various aspects of sales performance.

Here is a breakdown of the attributes present in the dataset:

1. InvoiceNo: This attribute represents the invoice number associated with each transaction. It is a nominal variable consisting of a 6-digit integral number. If the code starts with the letter 'c', it indicates a cancellation.
2. StockCode: This attribute corresponds to the product (item) code. It is a nominal variable represented by a 5-digit integral number uniquely assigned to each distinct product.
3. Description: This attribute provides the name of the product (item). It is a nominal variable.
4. Quantity: This attribute denotes the quantities of each product (item) purchased in a transaction. It is a numeric variable.
5. InvoiceDate: This attribute represents the date and time when a transaction was generated. It is a numeric variable.

6. UnitPrice: This attribute indicates the unit price of a product. It is a numeric variable denoting the product price per unit in sterling (£).
7. CustomerID: This attribute represents the customer number associated with each transaction. It is a nominal variable consisting of a 5-digit integral number uniquely assigned to each customer.
8. Country: This attribute denotes the name of the country where a customer resides. It is a nominal variable.

The dataset provides a comprehensive set of information that allows for a holistic analysis of the company's sales performance. By leveraging these attributes, the project will gain insights into sales amounts, product categories, customer demographics, geographic patterns, and other factors crucial for understanding and optimising sales strategies.

For further information about the dataset, please refer to the following link: Retail Dataset.

2.2. Data Cleaning and Preprocessing

During the data cleaning and preprocessing stage, I merged the two subparts of the data, Part 1 covering the year 2009-2010 and Part 2 covering the year 2010-2011. After merging the data, I encountered an issue with null values in the CustomerID column. To address this, I made the decision to remove the records where customer IDs were null. This approach ensured that incomplete or unavailable customer information was eliminated, leading to a more accurate assessment of sales performance and customer behaviour.

Additionally, irrelevant/cancelled orders information was also deleted to maintain focus on assessing the financial well-being and sales performance of the online retail company. With the data cleaned and refined, I proceeded to create a column for customer segmentation based on spending behaviour. This process classified customers into distinct categories, providing valuable insights into different spending patterns.

The segmentation process revealed the following customer segments:

- "High Spenders": Customers demonstrating remarkable sales amounts surpassing £600 during the analysed period.
- "Moderate Spenders": Customers with moderate levels of expenditure, exhibiting sales amounts ranging between £200 and £600.
- "Standard Spenders": Customers with relatively lower spending patterns, as their sales amounts fell below the £200 threshold.

By identifying and categorising customers into these segments, I gained a deeper understanding of their spending behaviours and their impact on the company's sales performance. This information serves as a foundation for further analysis and strategic decision-making, enabling the optimisation of sales strategies and the enhancement of customer experiences.

After completing the data cleaning and preprocessing steps, I proceeded to import the cleaned data into Tableau for further analysis. Prior to creating visualisations, I ensured that the data types of each variable were correctly identified in Tableau to facilitate accurate calculations and visual representation.

To enhance the analysis, I created a new column called "Sales" by multiplying the "Quantity" and "UnitPrice" variables. This additional column provided a consolidated view of the sales amount for each transaction, enabling a more comprehensive understanding of the revenue generated.

With the data prepared and the necessary columns created, everything was ready to create visualisations in Tableau.

2.3. Data Analysis and Modelling

In the project, I employed various analytical techniques to analyse the data and derive meaningful insights. Statistical analysis played a crucial role in understanding the key variables, namely "Quantity," "Price," and "Sales." Here are the statistics for these variables:

- Quantity: The dataset contains 802,172 records with a mean quantity of 13.34. The sample standard deviation is 144.59, and the sample variance is 20,906.41. The minimum quantity is 1, while the maximum reaches 80,995, resulting in a range of 80,994.
- Price: The dataset includes 802,172 price values, with a mean price of £2.9303. The sample standard deviation is £4.2764, and the sample variance is £18.2878. The minimum price is £0, and the maximum price is £649.5, resulting in a range of £649.5.
- Sales: The dataset comprises 802,172 sales amounts, with a mean of £21.7233. The sample standard deviation is £222.6349, and the sample variance is £49,566.2806. The minimum sales amount is £0, while the maximum is £168,469.6, resulting in a range of £168,469.6.

	Quantity	*Price*	*Sales*
count	802172	802172	802172
mean	£ 13.34	£ 2.93	£ 21.72
sample standard deviation	£ 144.59	£ 4.28	£ 222.63
sample variance	£ 20,906.41	£ 18.29	£ 49,566.28
minimum	£ 1.00	£ -	£ -
maximum	£ 80,995.00	£ 649.50	£ 168,469.60
range	£ 80,994.00	£ 649.50	£ 168,469.60

Figure 2.1. Descriptive Statistics

To derive customer insights, I performed customer segmentation based on predefined spending thresholds rather than using clustering techniques. The segmentation resulted in three distinct customer segments: High Spenders, Moderate Spenders, and Standard Spenders.

The chosen approach of defining customer segments based on predefined spending thresholds allows for a straightforward and practical understanding of customer behaviour and facilitates targeted marketing strategies.

Overall, statistical analysis and customer segmentation based on spending thresholds have been suitable techniques for this online retail project. These approaches provide valuable insights into sales performance, customer behaviour, and the identification of different customer segments, enabling effective decision-making and tailored marketing strategies.

2.4. Exploratory Data Analysis (EDA)

a. Customer Segmentation: Through a meticulous analysis of the dataset, customers were classified into distinct categories based on their spending behaviour.

This customer segmentation approach allows for a comprehensive understanding of customer spending habits and facilitates targeted strategies tailored to each segment. By recognising the variations in customer spending behaviours, businesses can develop customised marketing campaigns, optimise pricing strategies, and prioritise resources to cater to the distinct needs of each customer segment.

b. Analyse Customer Contribution: The analysis of customer contribution revealed the following insights:

- "High Spenders" accounted for a total sales amount of £1,880,100.
- "Moderate Spenders" contributed £2,049,934 in sales.
- "Standard Spenders" represented the majority with £13,495,803 in sales.
- Overall, the total sales for the analysed period amounted to £17,425,836.

Figure 2.2. Customer Segment Contribution

c. Customer Behaviour Analysis: Key statistics derived from the customer behaviour analysis are as follows:

- The total quantity of products sold during the analysed period was 10,700,044.
- The number of unique orders placed by customers reached 36,581.
- The dataset included a total of 5,847 unique customers.

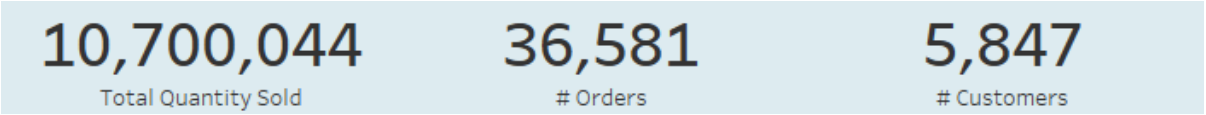


Figure 2.3. Customer Behaviour

d. Visualise Customer Segments: Visualisation of the data yielded valuable insights into sales patterns and customer segments:

- Analysis of sales data from December 2009 to December 2011 revealed November as the peak sales month, while February consistently exhibited the lowest sales. These patterns provide opportunities for targeted marketing and promotional campaigns during peak sales months.

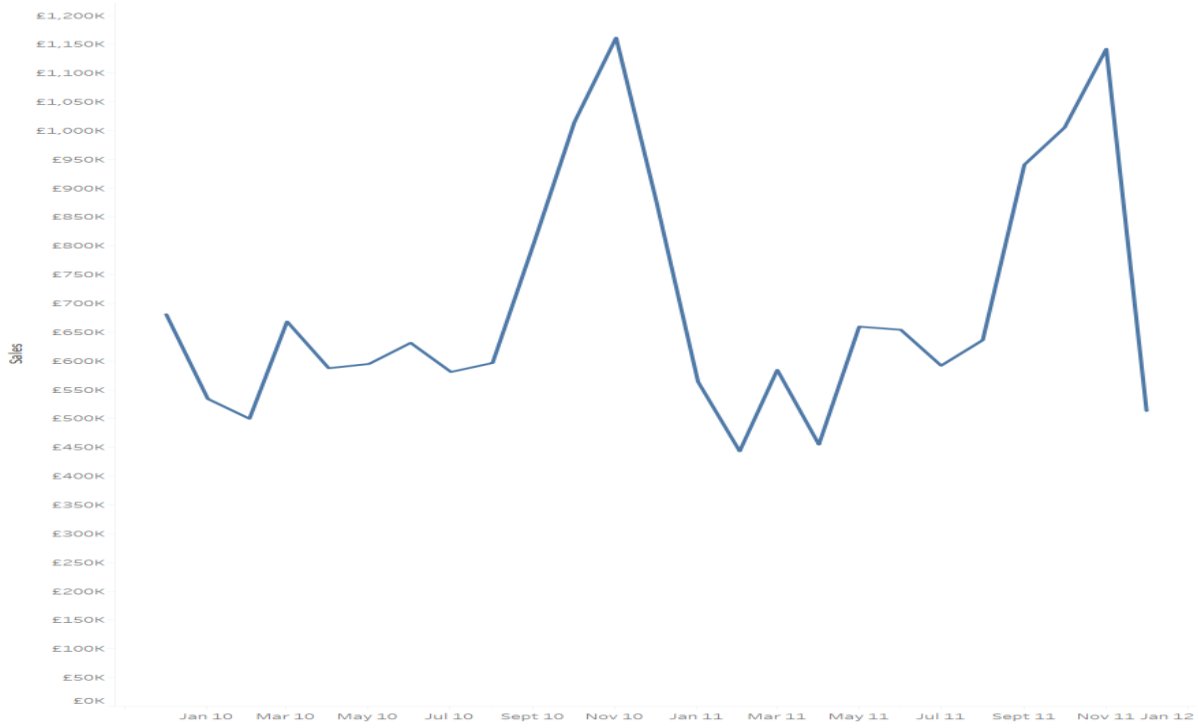


Figure 2.4. Sales over Time

- 10.79% of high-spending customers greatly impact revenue, emphasising the need for personalised experiences and targeted marketing to foster loyalty and attract similar high-value customers.

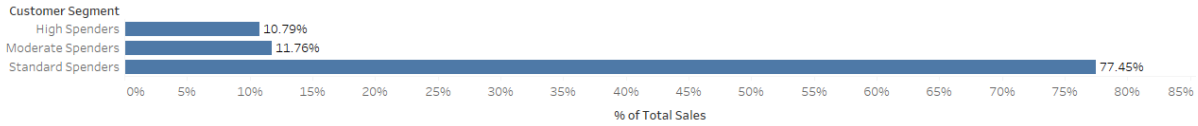


Figure 2.5. Customer Segment Contribution (in Percent)

- The dataset indicates the UK as the leading market, as it has the highest customer distribution. Additionally, November witnessed the highest order volumes, while January experienced the lowest. These insights offer valuable information for inventory management and targeted marketing strategies tailored to different regions and seasons.

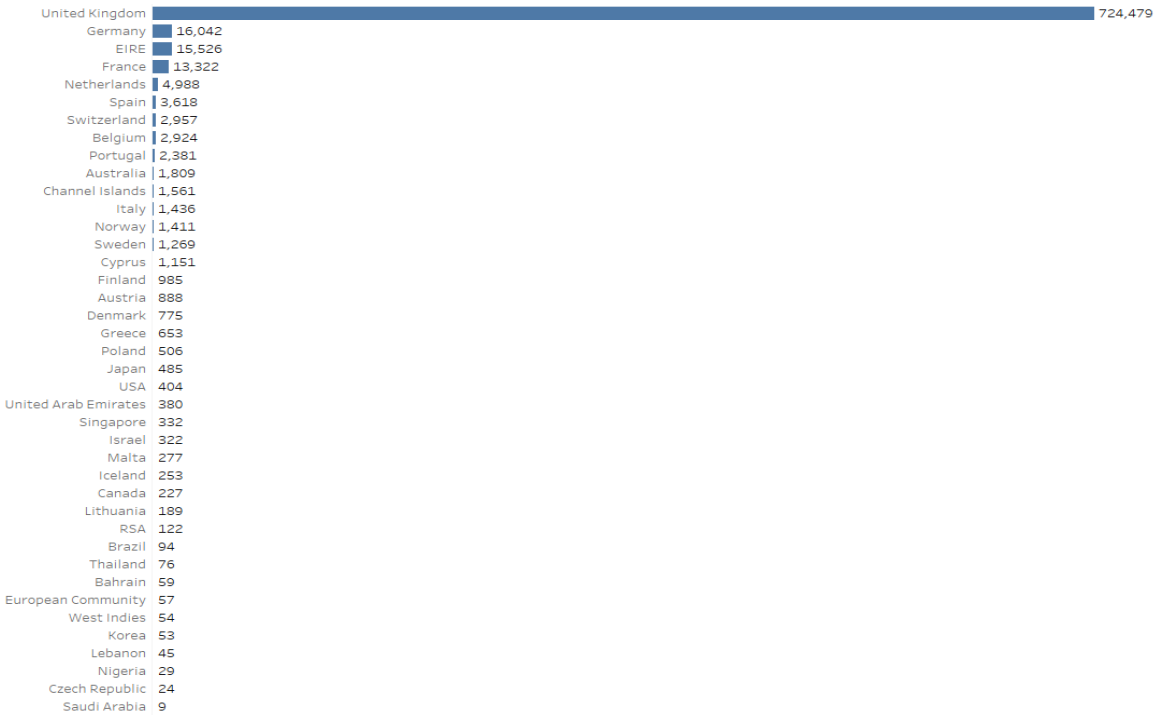


Figure 2.6. Customer Distribution

By leveraging customer segmentation and behaviour analysis, businesses can make informed decisions to optimise sales strategies, improve inventory management, and enhance customer experiences.

Overall, this analysis provides actionable insights for driving sales growth, developing targeted marketing campaigns, and optimising inventory planning based on customer segments and their behaviour patterns.

3. Results and Discussion

I conducted a comprehensive analysis of the online retail dataset and visualised the findings using a dashboard. The dashboard can be accessed at the following link: Retail Sales Dashboard.

The dashboard consists of various graphs and visualisations that provide insights into different aspects of the online retail business. Here are the key findings based on the analysis:

1. Top-Selling Products: The top 5-6 products have been identified, generating the maximum revenue ranging between £100,000 and £300,000. These products play a crucial role in driving overall sales and should be given special attention.
2. Sales Performance Over Time: The sales analysis from December 2009 to December 2011 reveals that November consistently stands out as the peak sales month. On the other hand, February consistently exhibits the lowest sales. This information can guide marketing and sales strategies to maximise revenue during peak months and improve performance during slower periods.
3. Comparison of Sales and Quantity Sold: The analysis shows that November has both the highest sales and quantity, indicating a strong positive correlation between these two metrics. Conversely, February consistently exhibits the lowest sales and quantity. These insights can help in understanding customer buying behaviour and planning inventory accordingly.
4. Top Customers based on spending and order size: The top 10-15 customers have been identified, generating significant revenue between £100,000 and £700,000. Moreover, it is observed that a majority of these top spending customers also place a large number of orders. This highlights the importance of retaining and nurturing these high-value customers.

5. Customer Segmentation based on Purchasing Behavior: The analysis reveals distinct customer segments based on spending behaviour. These segments provide insights into different customer preferences and can guide personalised marketing strategies to enhance customer satisfaction and loyalty.
6. Customer Distribution by Country: The analysis indicates that the UK consistently leads in terms of customer distribution, reflecting a strong market presence. This information can help in targeting marketing efforts and expanding the customer base in other countries with potential growth opportunities.
7. Order Volumes by Month: November exhibits peak order volumes, while January experiences the lowest. This information can be leveraged for effective inventory management and staffing optimisation to meet customer demand efficiently.
8. Sales by Country: The analysis highlights that the UK, EIRE, Netherlands, Germany, France, Australia, and Spain are significant markets with high sales. This emphasises the need for tailored marketing strategies to capitalise on these markets' potential.

The dashboard provides a comprehensive overview of the online retail business, allowing stakeholders to gain valuable insights into sales performance, customer behaviour, and market trends. These findings can be used to make informed decisions, optimise marketing efforts, improve customer satisfaction, and drive overall business growth.

Please refer to the Retail Sales Dashboard for a detailed view of the visualisations and further exploration of the data.

4. Technical Skills

Throughout the project, I utilised Tableau and Excel as the primary software tools for data analysis and visualisation.

Tableau: I leveraged Tableau to create interactive and visually appealing dashboards that effectively conveyed the insights derived from the dataset. Tableau allowed me to explore and analyse the data from multiple perspectives, enabling the creation of various charts, graphs, and visualisations to showcase key findings.

Excel: I utilised Excel for data cleaning, preprocessing, and initial exploratory analysis. Excel provided a familiar and versatile platform for manipulating and organising the data, performing calculations, and generating summary statistics. It allowed me to ensure data quality, handle missing values, and prepare the dataset for further analysis.

By combining the capabilities of Tableau and Excel, I was able to perform a comprehensive analysis of the online retail data, visualise the results, and derive meaningful insights that informed decision-making and strategic recommendations.

5. Visual Presentation

For my portfolio project, I aimed to create a visually appealing and informative presentation to summarise the key findings and insights from the analysis. To achieve this, I utilised Tableau's storytelling feature to create an engaging and interactive storyline.

The Tableau story, accessible through the following link: Retail Sales Storyboard, incorporates various visuals and narratives to effectively communicate the project's outcomes.

The storyline begins by providing an overview of the project's objectives and data sources, setting the context for the analysis. I utilised a combination of text, images, and charts to capture the audience's attention and create a visually appealing introduction.

Throughout the story, I included relevant visuals to support the interpretation of the data. These visuals include graphs such as "Top-Selling Products," showcasing the revenue generated by the top 5-6 products, and "Sales Performance Over Time," illustrating the sales trends from December 2009 to December 2011.

To enhance readability, I paid careful attention to font selection, colours, and layout. I used clear and legible fonts that are easily readable on different devices. Colours were chosen strategically to create a visually cohesive and harmonious presentation, while also ensuring that important information stands out.

In addition to the visuals, I provided concise and insightful narratives to guide the audience through the project's findings. These narratives help to contextualise the visuals and highlight the main takeaways from the analysis.

The Tableau story concludes with actionable recommendations based on the insights derived from the analysis. These recommendations are designed to assist stakeholders in making informed decisions and implementing strategies to improve sales performance and customer satisfaction.

Overall, the Tableau story serves as a visually compelling and informative summary of the project. It combines engaging visuals, clear narratives, and actionable recommendations to effectively communicate the key findings and insights derived from the analysis.

6. Real-World Impact

The findings and recommendations from this analysis have significant real-world implications for decision-making and problem-solving in the online retail industry.

1. Identifying the peak sales month as November and the lowest sales month as February allows businesses to plan their resources and strategies accordingly. By allocating more resources and implementing targeted marketing campaigns during the peak season, companies can maximise their sales potential and capitalise on consumer demand. Conversely, during the low-sales period, businesses can focus on cost optimisation and explore innovative strategies to attract customers.
2. The identification of high-spending customers as key contributors to overall revenue highlights the importance of catering to their needs and preferences. By providing personalised experiences and targeted marketing initiatives, companies can enhance customer loyalty and encourage repeat purchases. This can be achieved through personalised recommendations, exclusive offers, and tailored marketing messages to maintain engagement and build long-term relationships with these valuable customers.
3. The observation that the UK has the highest customer count indicates a strong market presence in this region. Businesses can leverage this information to allocate resources, optimise logistics, and prioritise marketing efforts to further penetrate and expand their market share in the UK. This may involve tailoring marketing campaigns to the preferences and behaviours of UK customers, adapting products or services to local preferences, and establishing strategic partnerships with local businesses.
4. The insight regarding the peak order volumes in November and the lowest in January provides valuable guidance for inventory planning and management. Companies can optimise their inventory levels to meet the anticipated demand during the peak season

and avoid excess stock during the low-sales period. Additionally, targeted marketing strategies can be deployed to stimulate sales during the low-demand months, offering promotions, discounts, or exclusive deals to encourage customer purchases.

Overall, the findings and recommendations derived from this analysis empower online retail businesses to make informed decisions and implement targeted strategies to maximise sales, improve customer loyalty, optimise inventory planning, and enhance their market presence. By leveraging these insights, companies can stay competitive, drive growth, and ultimately deliver a better customer experience.

Moving forward, I would continue to explore techniques for customer segmentation and geographic analysis to further enhance the depth of insights provided. Additionally, I would consider incorporating additional data sources to gain a more comprehensive understanding of sales performance.

7. Conclusion

The Sales Performance Analysis portfolio project demonstrates the use of data analysis and visualisation techniques to gain insights into sales performance and customer behaviour. By analysing the "Online Retail II" dataset, the project provides valuable information for optimising sales strategies, improving customer satisfaction, and driving business growth. The interactive dashboards created using Tableau facilitate easy exploration of the results and enable stakeholders to make informed decisions based on the derived insights.

Through this project, businesses can unlock the power of data to understand their sales performance, identify growth opportunities, and enhance their competitive advantage in the market.

Thank you for your attention. I'm available for any questions you may have.

8. References

1. Chen, Daqing. (2019). Online Retail II. UCI Machine Learning Repository. https://doi.org/10.24432/C5CG6D.
2. Kumar, V., & Reinartz, W. (2016). Customer Relationship Management: Concept, Strategy, and Tools. Springer.
3. Verhoef, P. C., Kannan, P. K., & Inman, J. J. (2015). From multi-channel retailing to omni-channel retailing: Introduction to the special issue on multi-channel retailing. Journal of Retailing, 91(2), 174-181.
4. Rokach, L., & Maimon, O. (2014). Data mining with decision trees: theory and applications. World Scientific Publishing Company.
5. Jain, A. K., & Topchy, A. (2014). Machine learning techniques for classification problems. In Data classification: algorithms and applications (pp. 163-188). CRC Press.
6. Wedel, M., & Kannan, P. K. (2016). Marketing analytics for data-rich environments. Journal of Marketing, 80(6), 97-121.