ShopOnline
Business Analysis Project

Xavier Zen
Xaizhen0808@gmail.com

Table of Contents

1.0 Introduction

ShopOnline is an E-commerce enterprise that specializes in the merchandising of gifts. This report is a general annual sales analysis for the company. The report will contain many textual analysis, tables, charts, and interactive dashboards for the reader to visualize and understand in detail.

This report comprises business insights from the analysis of data, and business opportunities and recommendations that can be observed from the analysis. In addition, there is a description of how the data preparation and analysis and visualizations were conducted. Finally, a professional evaluation will evaluate the visualization and provide recommendations to enable ShopOnline to improve its data analysis and visualization in the future.

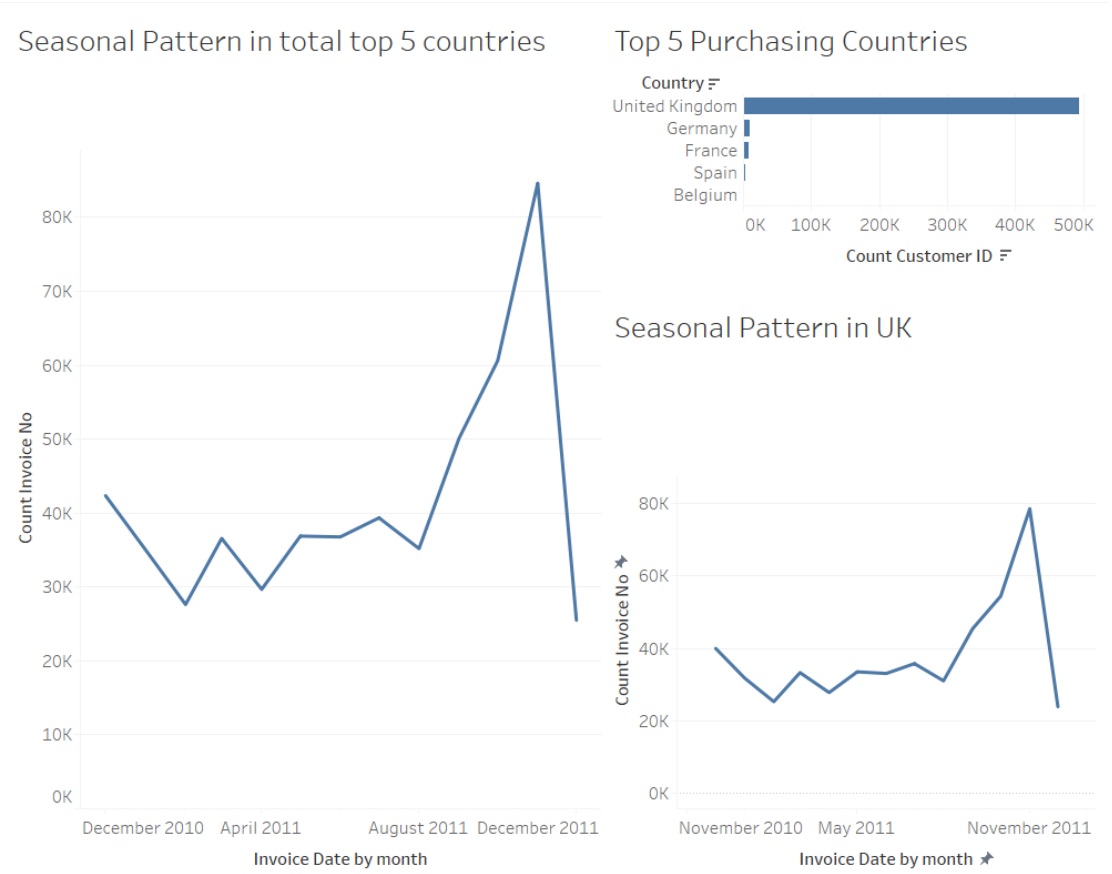
2.0 Business Insights

2.a Probability that customers will want to become repeat customers

First, it should take some action on how to find the number of customers, hence, it needs to use Tableau to find the number of customers. As the below graph shows, it could understand there are 4373 customer duplicates from the original database. Furthermore, it needs to find out the customers who are willing to repeat shopping more than once. As the below graph shows, there are 1283 customers only shopping once in this store, which means other customers would like to become repeat customers, and the repeat rate should be 70.7%. Hence, it could be easy to find out that customers would like to become repeat customers. Therefore, it could be said that people would like to become repeat customers after they buy once.

2.b Analysis for whether there are seasonal patterns in purchasing behaviour

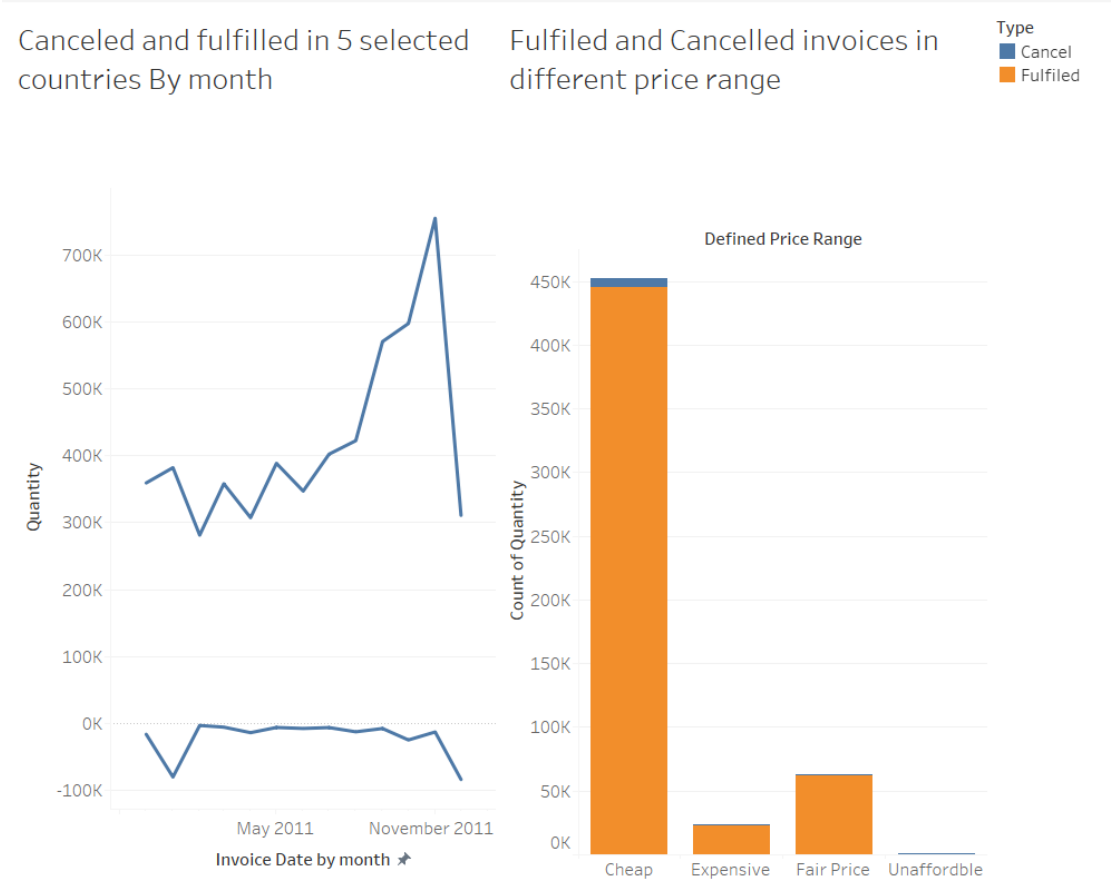
For the question about seasonal patterns in purchasing behaviour. First, it needs to be understood which country is in the top 5 purchasing countries to narrow the samples, then it could be easy to mention that the United Kingdom would be the biggest market. Hence, the seasonal patterns in purchasing behaviour would be listed below. It could be easy to find the fact that customers would like to purchase from August to November, and November would be the highest point (84,626 transaction Invoices). This means people would like to make purchasing decisions in Q4. Hence, there is a seasonal pattern that exists in purchasing behaviour.



2.c Differences between cancelled orders and fulfilled orders

First, it should be considered the Cancelled and fulfilled orders by time measurement, which takes research into time patterns in cancelled and fulfilled orders. Hence, it takes 5 selected countries as samples, and it shows the fact that the cancelled orders will increase in January. Assuming this situation, it may be a situation where consumers are always shopping at the end of the year, and they often give up shopping in January because of their financial situation. Besides, there is another fact that could approve the assumption, which is the price of these products. According to the data, we decided to group different products into separate price ranges. Cheap products have lower prices than 5 per unit, and Unaffordable products have higher prices than 100 per unit. Also, we found an interesting thing: people cancelled 225 unaffordable orders and fulfilled 814 orders, of which the cancelled rate would be 21.65%. Furthermore, it could be mentioned that cheap products only have around a 1.3% cancelled rate.

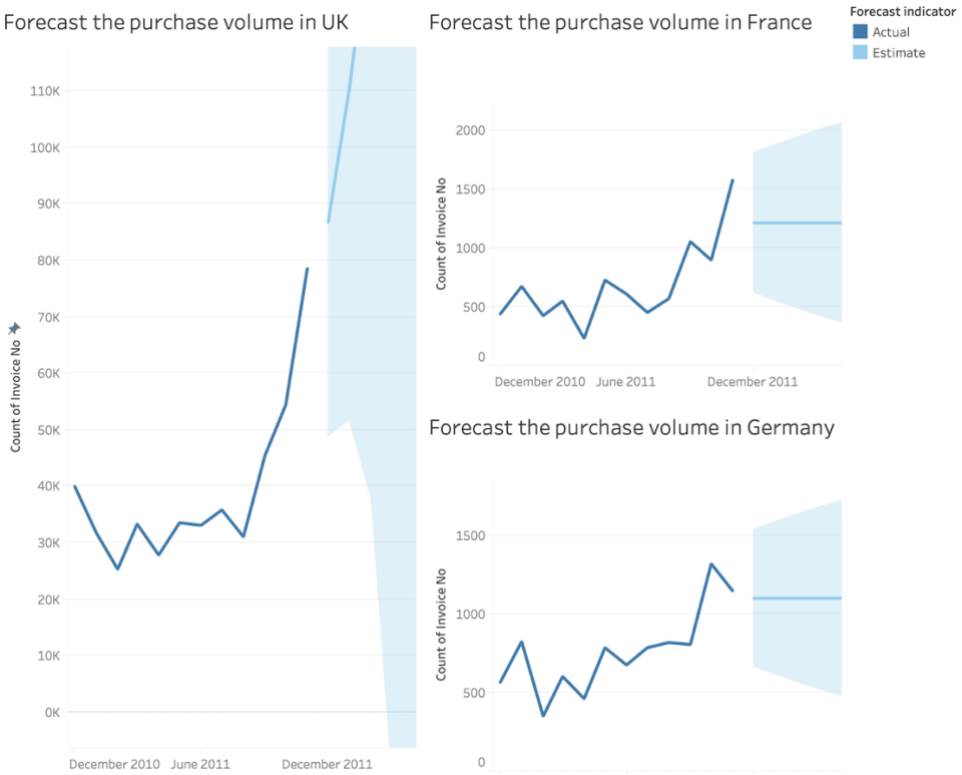
In conclusion, people will cancel orders because of the high price, and it always happens in January.



2.d Analysis of whether the country can be used as an indicator to predict purchase volume

Based on the assumption, it needs to mention that purchase volume could be predicted, but there is a requirement that needs to mention that it must promise there are enough samples. According to the Tableau Forecast function, we could find out that the Forecast needs more data samples to keep accuracy. Hence, there are so many invoices in the United Kingdom, and the forecast could be very clear. Otherwise, France and Germany have few samples from the original sources, which could represent the detailed forecast.

Based on the graph below, the purchase volume would increase very fast in the United Kingdom, and it may reach 110,742 units in 2012. Regarding the forecast, it could explain the prediction needs enough samples.



2.e What does the data show at different levels of analysis

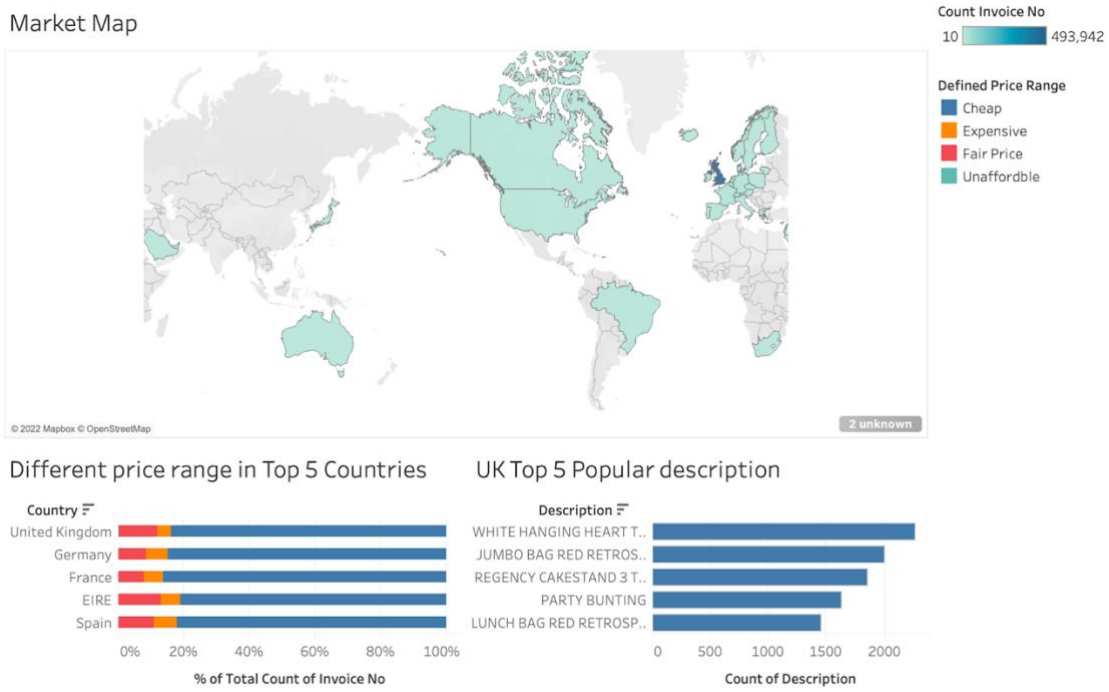
For this question, there are two points our team wants to research, which are where the most important market is, and what is the fundamental information about this market.

First, the interesting question should be how to find which country would be the most important. Then we need the details of the country and purchase volume, and use Tableau to analyse this situation. Then it could be easy to find out where the biggest market is in the below graph. Hence, the United Kingdom would be the biggest market based on the data, and it could mention there are still some countries based on the original data. But the graph below shows that all of the other countries are fewer than the United Kingdom. To deep dive into the market of the United Kingdom, it could find out the Top 5 popular descriptions in the United Kingdom. This means we could analyse what is the consumer's favourite.

2.f Insights about different customer segments

The graph under this question could define what is the different price range preference as customer segments. There are five selected countries or regions and we try to find out customer shopping preferences in different countries. According to the data, we decided to group different products into separate price ranges.

Cheap products have lower prices than 5 per unit. Besides, the fair price is between 5 to 10 per unit, the expensive price is between 10 to 100 per unit. Furthermore, Unaffordable products have higher prices than 100 per unit. Then we just found some quite interesting stuff in this research. First, France would be the most out of the five selected countries which would like to shop in the cheap price range. On the other hand, EIRE would be the most out of the five selected countries that would like to shop in a fair price range. In conclusion, it could mention customers focus on the price in France.



2.g The frequency of customer purchases

About the question of how frequently customers purchase, it could understand what we have mentioned in the first goal-specific question. Focus on the largest number of customers. It should be those who would like only to purchase once, and the number should be 1,283 customers. Then, it could find that 814 customers would purchase twice. This means most customers only purchase once and it is the largest customer group. It represents most customers who would like to purchase just once.

2.h Whether customers from certain countries purchase more often than others

Customers from the United kingdom purchase more often. According to the database, it could be understood that the largest number of samples are from the United Kingdom. Hence, there is evidence showing there are two customers purchasing 33 times from the United kingdom.

2.i High-volume weeks, months, and seasons
According to the graph of seasonal patterns in purchasing behaviour,13-20 November should be common among High-Volume weeks, and the month should be November. Furthermore, the season should be Q4, which is winter.

2.j Association between purchases and cancellations

Yes, there should exist an association between purchases and cancellations. The graph of Cancelled and fulfilled orders by time measurement shows the fact that the number of cancelled orders will increase in January. Assuming this situation, it may be a situation where consumers are always shopping at the end of the year, and they often give up shopping in January because of their financial situation.

3.0 Opportunities and Recommendations

3.a Industry benchmarks and Key Performance Indicators to Create Potential Value
In order to understand the financial and operational performance of an enterprise, industry benchmarks and key performance indicators (KPIs) are the two effective measurement methods. Although industry benchmarks and key performance indicators (KPIs) seem very similar to each other, benchmarks and key performance indicators (KPIs) are different. Benchmarks describe the process you use to compare your company's financial performance against other companies or competitors which help you compare that performance on a larger scale, while KPIs help you gather data about your internal performance (Cake, 2022).
Benchmarks can be divided into four types which are listed below (Cake, 2022).

- Internal benchmarking compares internal processes within your company.
- Competitive benchmarking refers to benchmarking that looks directly at your competitors so you can improve your internal operations.
- Functional benchmarking refers to comparing similar practices across different industries or companies.
- Generic benchmarking refers to identifying work processes that don't need to be in the same industry or job function to compare.

The key performance indicators of e-commerce usually consist of five categories as follows (Firestone, 2022).

- Sales KPIs are measures that tell you how your business is doing in terms of conversions and revenue.
- KPIs for marketing tell you how well you're doing in relation to your marketing and advertising goals.
- Customer service KPIs tell you how effective your customer service is and if you're meeting expectations.
- KPIs for manufacturing are, predictably, related to your supply chain and production processes.

- KPIs for project management give you insight into how well your teams are performing and completing specific tasks, which is critically important when running an e-commerce store.

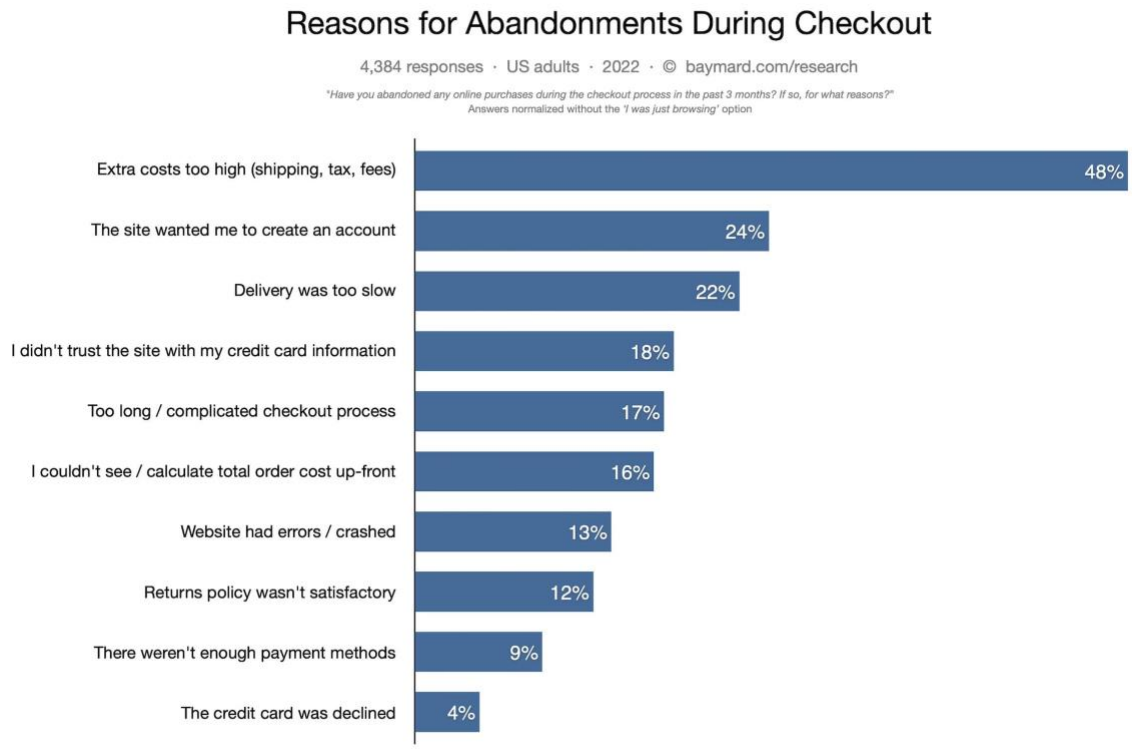
3.b Quantified Value Creation Opportunities and Areas of Improvement

The quantified value-creation opportunities can be discussed in two aspects. The first one is that total cost can be deducted up to 20 percent with the development of the delivery system (Briest et al., 2020). It is shown below. In this situation, the total revenue stays the same, however, the total cost has been reduced by around 20 percent. Then the net profit will be increased by around 20 percent. It provides a better position among the competitors.

(Briest et al., 2020)

The second one is that Using stores as micro fulfillment centers can help retailers meet increasing customer expectations for fast delivery while avoiding skyrocketing costs (Solving the Paradox of Growth and Profitability in E-commerce, 2022). With more than 90% of consumers considering two-to-three-day delivery as standard, retailers with an existing physical footprint have a unique advantage in meeting customer demand while controlling costs (Barbee et al., 2022). It not only improves work efficiency and reduces costs but also provides customers with good service.

Although there are some opportunities, the areas for improvement cannot be ignored. Customers sometimes would rather choose to abandon the cart before they make a payment. The reasons can be a variety of them. The two main reasons are the extra cost too high such as shipping, taxes, and fees, and the requirement for creating an account (48 Cart Abandonment Rate Statistics 2022 – Cart & Checkout –, n.d.). The table is shown below.



(48 Cart Abandonment Rate Statistics 2022 – Cart & Checkout –, n.d.)

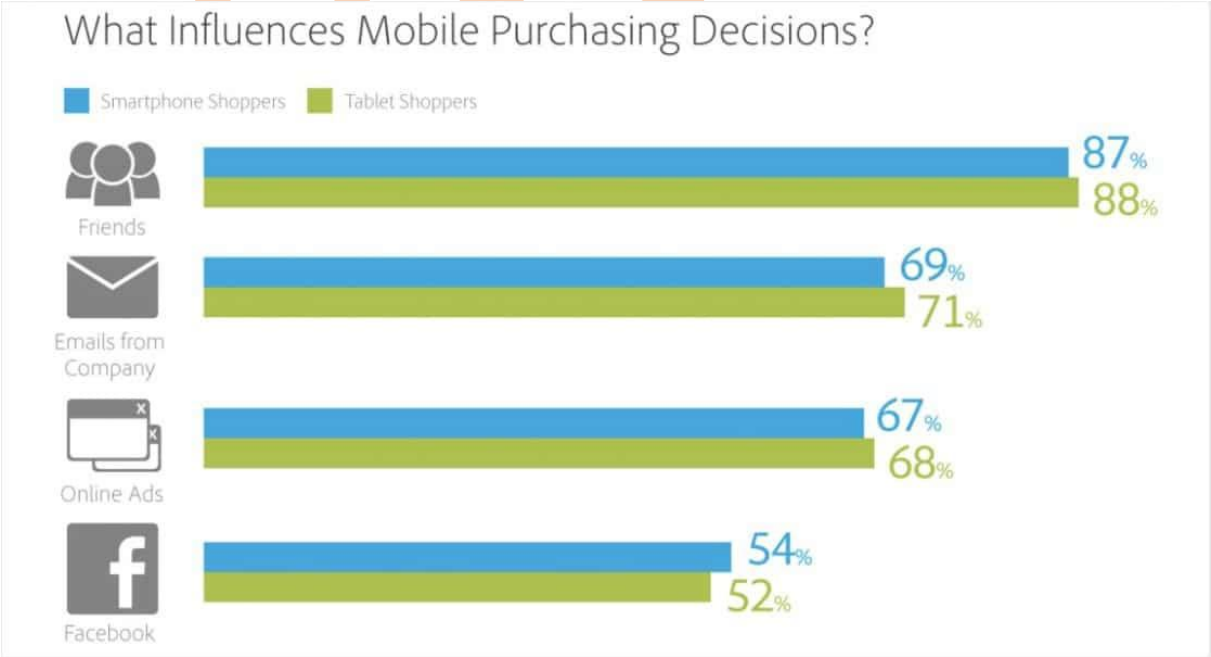
Therefore, the e-commerce business should make some adjustments accordingly based on the percentage of reasons that take place.

3.c High Impact Strategies for Increasing Revenue

There are a lot of strategies to increase revenue. The listed three are the most effective ones. The first one is to encourage customers to post product reviews, as you can see, 97% of shoppers say product reviews influence their purchase decisions. Additionally, 94% of online customers read reviews written by others, and 35% claim one bad product review causes them to reconsider their purchase (Berson, 2022). Therefore, product reviews are one of the most important factors when customers decide to make a purchase.

(Berson, 2022)

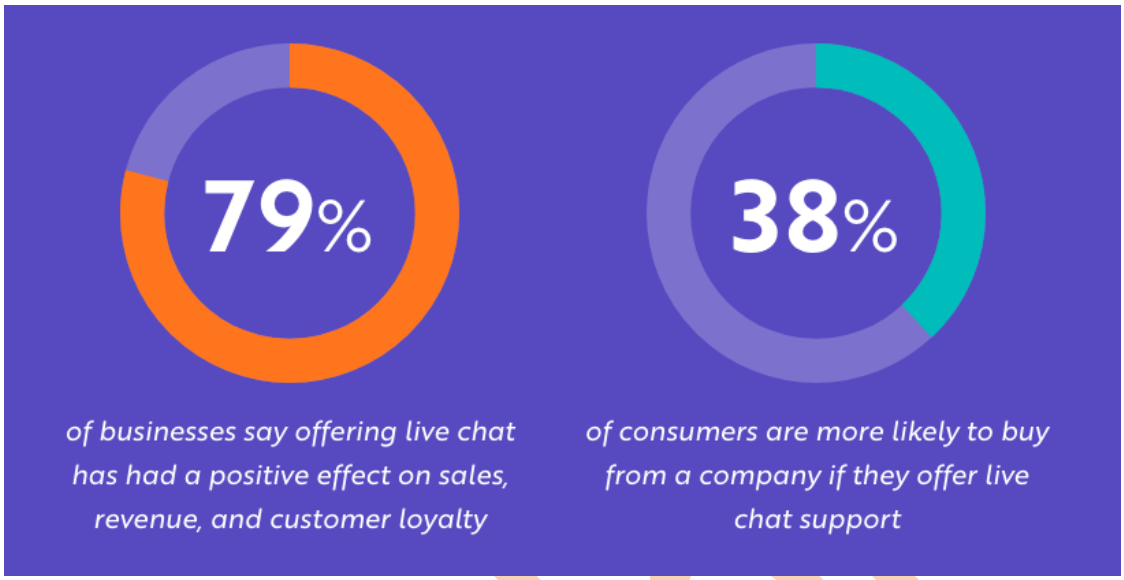
The second one is that encourage customers to recommend products to their friends. It shows below that people's friends have the most influence on mobile purchasing decisions over email marketing, online ads, or Facebook whether purchase items from cell phones or tablets (Gennaro, 2022). Therefore, the management team of the e-commerce business should come up with some ideas that can enhance the recommendation rate among friends in order to boost profits.



(Gennaro, 2022)

The third one is that Implement live chat to improve sales conversions, as you can see, 79% of consumers prefer real-time chat because they don't have to wait on hold for a customer support team to respond which leads to a 48% increase in revenue per chat

hour and a 40% increase in conversion rate (Patel, 2022). Live chat can help customers to reduce the time cost to wait for a salesperson when they need product information.



(Patel, 2022)

3.d ShopOnline Recommendations according to High Impact Strategies

The recommendations are designed for ShopOnline to deliver high-impact strategies. First of all, the specific measure that ShopOnline can take to encourage customers to post their product reviews is to give customers cash back or coupons based on their preference after their post the product reviews. It cannot only encourage the customers who like to make a comment after shopping online to continue doing more comments but also encourage the customers who dislike making a comment after shopping to make a comment driven by profit. The second one is that ShopOnline can offer discount opportunities after customers recommend the products to their friends. It is a win-win situation for both customers and their friends. Customers can gain discount opportunities and their friends can get opinions about the products from their friends which could reduce their trial-and-error cost of purchasing unwanted products. Last but not least, ShopOnline should recruit some people and develop or purchase an online live chat online system to answer the customers' questions when they are shopping online. The online live chat system can help to answer basic questions such as the delivery time. The sales representatives can help to answer difficult questions such as complaints.

4.0 Data Preparation and Analysis

4.a Choice of features used in the analysis and reasons

In the provided dataset called 'ShopOnlineData', all other features excluding 'StockCode' are selected for the analysis and visualisation in this report. This is because "StockCode" and "Description" have common traits, product name and item number

both of which indicate product information. Since 'StockCode' is fabricated from numbers, it is not a desirable option to use these numbers straightforwardly in the analysis report to demonstrate the product, therefore, the feature 'Description' is retained because it allows for a more direct presentation of the product name in the analysis.
The following table contains the reasons for using other attributes.

Name of the feature	**Arguments for choosing to use it for follow-up analysis**
InvoiceNo	It is useful when analysing the total purchase price of each order and the total number of products ordered. It can be used to analyse the number of cancelled orders. It can be used to calculate the overall transaction volume and annual turnover.
Description	Using this Attribute enables a straightforward display of product information in the analysis and visualisation of graphics, allowing the viewer to have a clear image in their mind.
Quantity	This can be used to examine the aggregate, for example, to measure the total number of units sold of an item throughout the year. This can be used to analyse aggregate, for example, to calculate the total quantity of an item sold. This allows analysts to do an analysis of the most preferred top ten products and the least favoured products.
Invoice Date	This measure supports the analysis of transaction volume and sales by month. In addition, it also allows analysts to observe if sales increase in months with holidays, for example, Thanksgiving Day, and Christmas Day.
Unit Price	This feature needs to be retained in order to calculate total sales prices.
Customer ID	It helps to analyse whether there are any returning customers and loyal customers, and who spent the most on this website.
Country	This is important when analysing the geographical distribution of sales, for instance, which region's clients are the main consumer group.

4.b Data quality issues and assumptions about the dataset

With the basic Microsoft Excel function 'Sort & Filter', a preliminary preview of the contents of each feature's column in the ShopOnline dataset can be performed. Furthermore, by inspecting the filter of each column, it is possible to determine whether there exist data quality issues in each feature, and what type of problem they belong to respectively.

The following table contains analysis on whether there are data quality issues in each feature.

Name of the feature	Existence of data quality problems (Yes/No)	Types of data quality problems	Evidence and description
Invoice No	No	/	/
StockCode	Yes	Noise	Existence of inconsistent or unexpected data. Stock Code should be a 5-digit number as stated in ‘ShopOnline Case’, however, there is inconsistent data in this column in the dataset, such as "gift_0001_10', 'DCGS0004' and 'DOT'.
Description	Yes	Missing Data	This column has a lot of elements displayed as '????' or 'Blank', which means the absence of data.
Quantity	No	/	/
Invoice Date	No	/	/
Unit Price	Yes	Outliers	‘-11062.06’and ‘38970’ is significantly different from other data, it is also unexpected data.
Customer ID	Yes	Missing Data	Existence of blank data.
Country	No	/	/

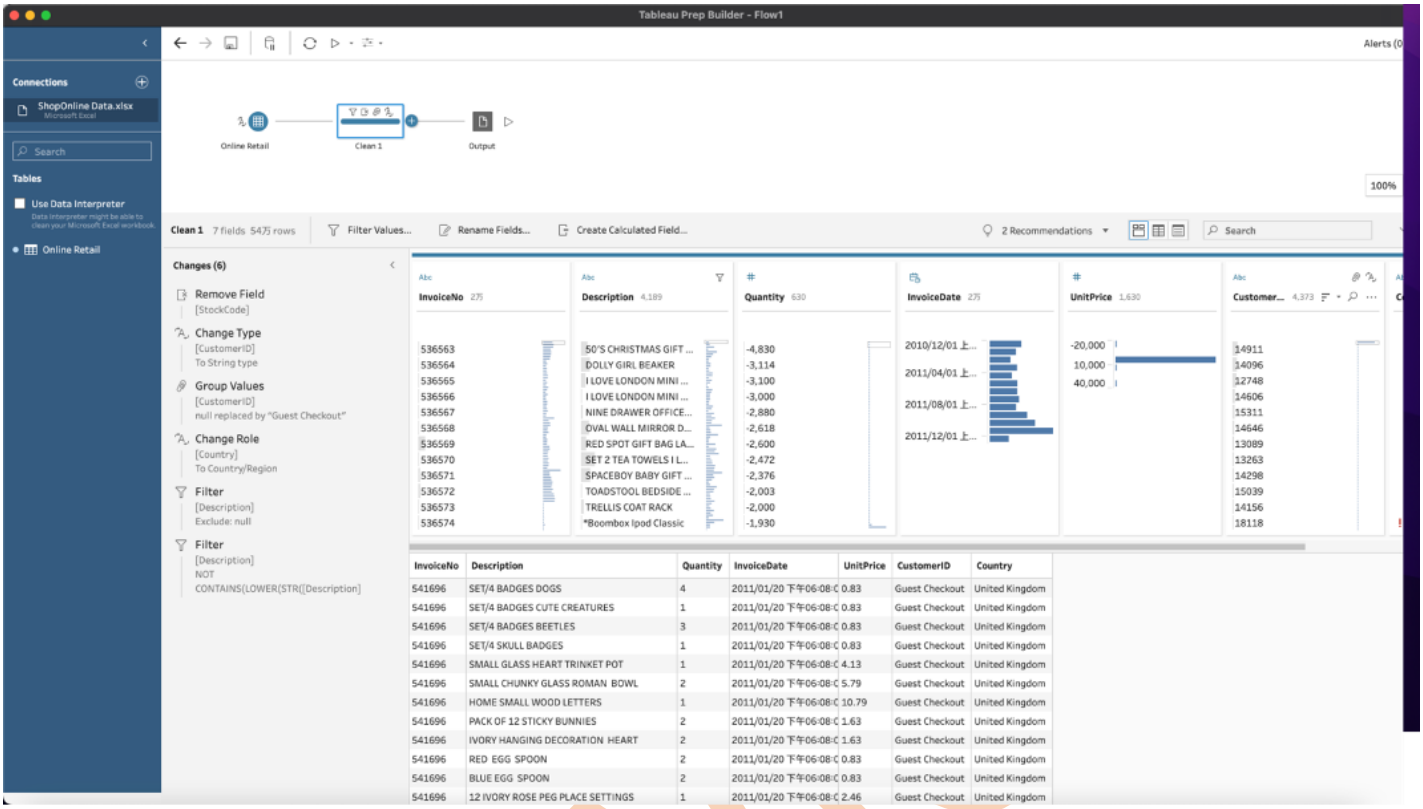
Besides, before addressing the data quality issue, we made the following underlying assumptions:

1. Suppose that the metric "StockCode" is not considered in this analysis report, which is also supported by the reasons discussed in section a4a, then this column will be deleted.
2. Assuming this report does not consider sold items where the item name is unknown, so eliminate missing data and incomplete records.
3. Suppose all the data in "Unit Price" are crucial transactions to the company, so it is necessary to keep all the data even if there are outliers.
4. Assume that all customers recorded as <blank> in the dataset are purchasers who did not register for a website account. So missing values in feature ‘Customer ID’ will be replaced by a name called “Guest Checkout".

4.c Solution selection and justification for using this proposal

Tableau Prep is the software chosen to be used for data processing and data cleaning. The following figure shows how I clean data using Tableau Prep. Our group implements data-cleaning approaches for all the features that have data quality issues. The left side of the image demonstrates the data cleaning procedure for the "ShopOnline data" file, which is:

1. Remove the ‘StockCode’ feature from the dataset. Because it has the same meaning as 'Description', and the item name is more straightforward than the item number.
2. Due to Assumption 2, remove all missing data on ‘Description’.
3. After checking the transaction of the two outliers in 'Unit Price', I found that '-11,062.06' is for adjusting bad debt, and '15098' is for manual reasons. For this reason, these two outliers should be retained.
4. By looking at the missing values in the feature ‘Customer ID’, there are 133,533 rows with missing values in this feature, which hold 25% of the total dataset. Furthermore, referring to Assumption 4, therefore decided to substitute all missing values in 'Customer ID' with a given name 'Guest Checkout.



4.d Other data considered useful for business analysis in this Case

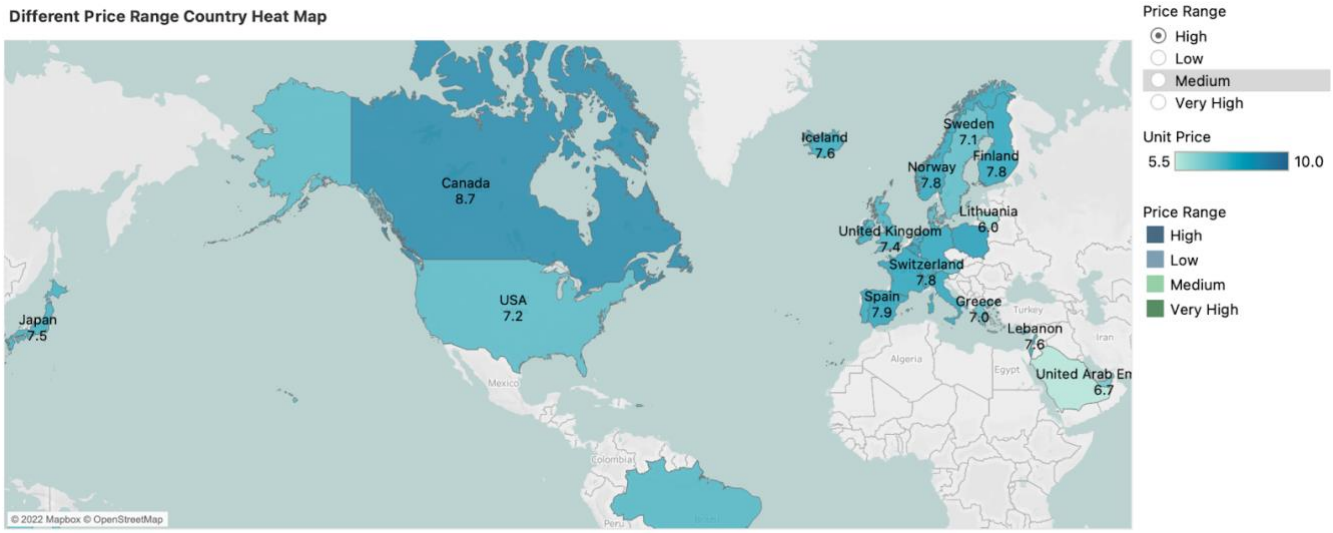
More demographic data, such as customer age, gender, marital status, education, and occupation, can be collected if circumstances permit, as this information helps companies to undertake customer segmentation analysis (Ehrens, 2019). This will enable the company to implement targeted strategies to organise marketing campaigns for specific customer segments in the future, for instance, offering college students a rational discount.

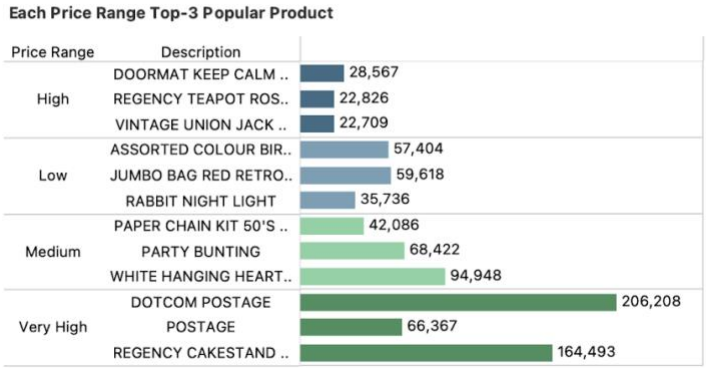
A customer feedback survey is also critically important, the company could analyse the distribution of the number of satisfied customers by collecting customer feedback. For customers who are not satisfied with the shopping experience, after-sale service should be conducted to solicit the reasons for customer dissatisfaction and then make service improvements or product development according to their opinions (Gupta et al., 2017).

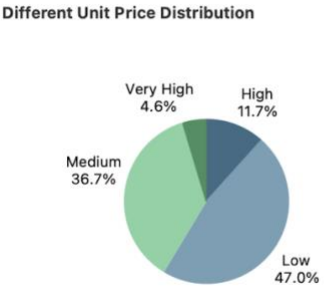
5.0 Visualization

5.a Interactive Dashboard

As the graph displayed below, the dashboard contains the distribution of sales items in different countries(classified by distinct price ranges), the top three popular products (classified by distinct price ranges), and different unit price distributions.

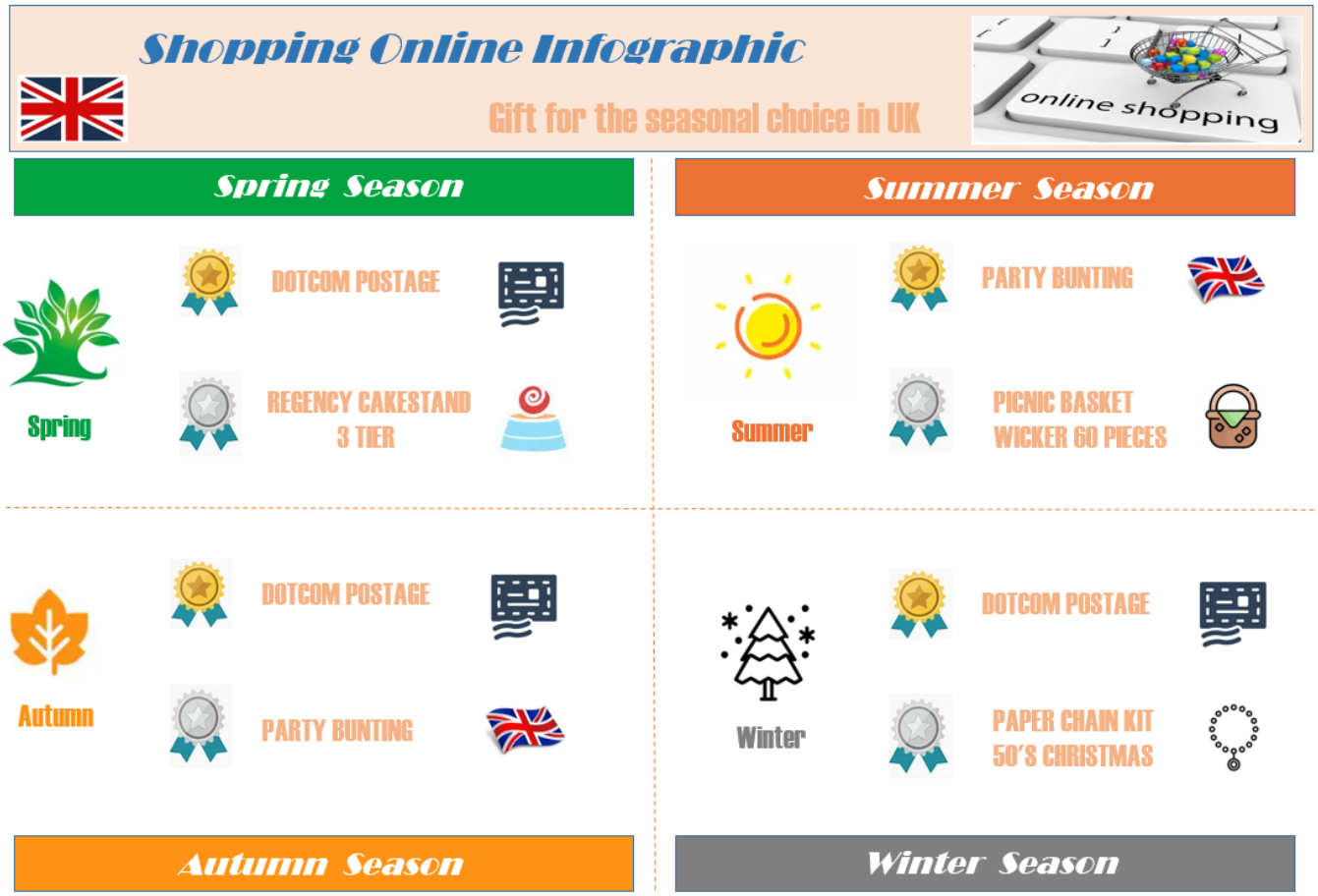






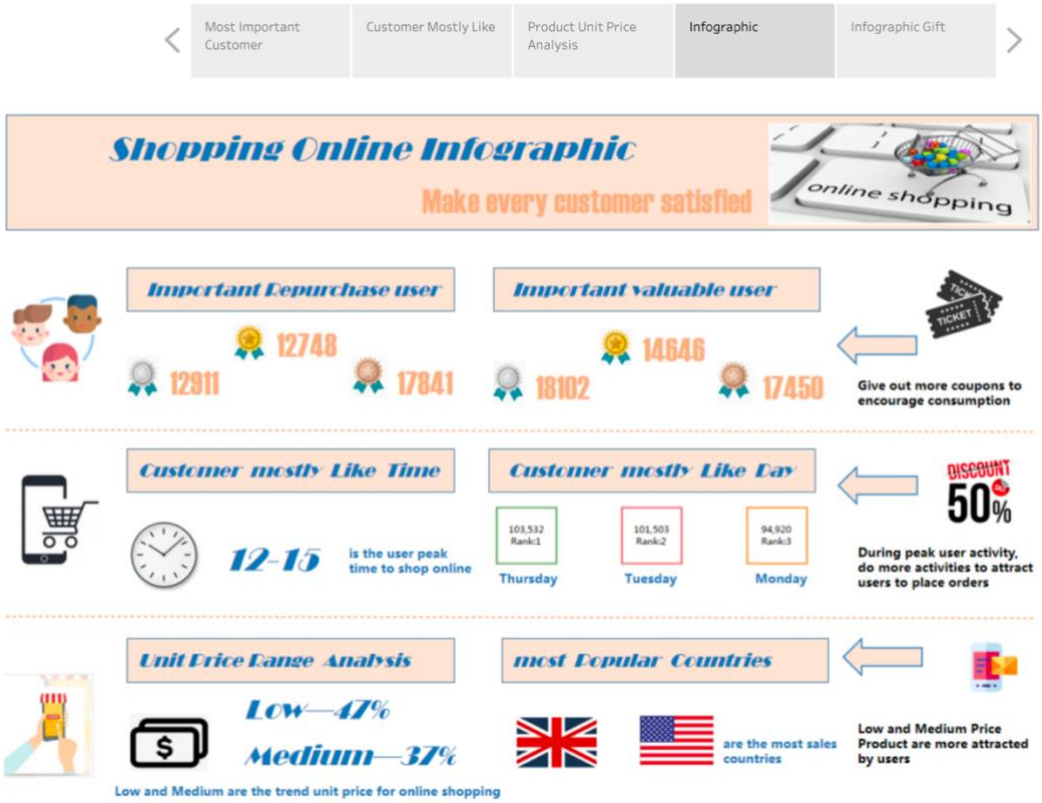
5.b Infographic

The image below indicates the two best-selling products categorized by season in the UK. In this analysis, spring refers to March 1 to May 31, summer refers to June 1 to August 31, fall refers to September 1 to November 30, and winter refers to December 1 to February 28. This graph serves as an incentive for potential UK consumers to become aware of seasonal best sellers and to purchase them if they are interested.



5.c Visual Story

This diagram is a preview of the visual story, please see the file named 'Group4-RE1-2022-S2.twbx' to look over the complete version. This visual story includes information about the most important customer, the most preferred products among the customer of the year, analysis of product unit price which include the distribution of products sold in different countries for varying price ranges and top-3 popular products in different price ranges. Also, there are two infographics images, one is a general analysis for senior management to look over the performance of the year and the other one is for motivating UK potential customers to purchase more products.



6.0 Evaluation

6.a Assessment criteria, self judgement, and evaluating process

The Shaffer 4 C's of Data VIsualization approach has been used in this analysis report as a guide to evaluate the quality of the visualization. The evaluation criteria are as follows:

Clear - The evaluation criteria in this part require clear presentation and accurate definition of the visual content. At a deeper level, this part requires visualisation: being able to understand the audience and adjust accordingly; Define what you want to convey to the audience through visualisation. In practice, visualisations should include: accurate chart titles, key labels, units of measurement, good colour selection, and avoid rotating text, cluttered, meaningless 3D effects, etc.

Clean - The evaluation criteria in this part require that the content of visualisation be as pure as possible on a comprehensive and complete basis. In practice, visualisations should not include: over-identified axes or coordinate points, dim background colours, poor colour selection, inefficient chart selection, excessive text captions, etc.

Concise - The evaluation criteria in this section require that the visual content is neither too complex nor too vague in providing information and ideas to the reader. In practice, this often requires a certain amount of experience, through a deep understanding of the research content and the audience, to accurately grasp the scale of visualisation.

Captivating - The evaluation criteria of this part require that the visual content can attract the audience's attention and keep the audience's desire to watch. In practice, visualisation should have: coordinated layout, clear point of view, strong focus, clear logic, beautiful colour selection, etc.

In addition, ethical issues is also a consideration when evaluating visualization. The data selected for visualisation should come from validated and legal sources, and the results should not mislead or deceive the audience.

The visualisations in this paper are in good compliance with the above standards, but some of them have some flaws in details. The following figure is an example. The selection of the two charts above is suitable for the data type, the colour selection is clear and friendly to the colour blind, the meaning of the coordinate axes is clear and there is no meaningless sign, and the language is professional. However, the placement of the subheadings is inconsistent, the spelling of Cancel appears in two forms, and the capitalization of the storyline is not consistent.

In conducting the evaluation, the accurate delivery of the content is the prioritise. Therefore, in the case of bringing in the question, whether the author's point of view can be clearly and accurately expressed through visualisation is another consideration.

6.b Recommendations for ShopOnline to gain data quality

Problems with the company	**Bulleted list of recommendations**
Massive missing Customer ID data	Query system records, trace problems and improve data entry and storage system
Outliers appear in product price data	Set the data validity range in the database. When the input data exceeds the set range, the system will report an error message to correct the data
Inconsistent data exists in Stockcode	Strengthen the training of data entry personnel and develop an effective assessment system
Duplicate data was entered	Set the error condition of duplicate value in the database, when the same data is repeatedly entered, the system will report error prompt correction
Dirty data exists	Maintain the database regularly, find problems in time, and clean up the existing data

6.c Ethics and PAPA framework

ShopOnline protected customers' personal information. In the database used for analysis and presentation, the information relevant to the individual customer is the country area and the product price and sales volume. The analysis and presentation of such data will not violate the personal privacy of customers.

ShopOnline is potentially at risk of violating the PAPA framework in the future.

In terms of privacy, the actual decision-making power of the data that customers provide to the company through consumption rests with the company. Even though ShopOnline may have signed legally protected confidentiality agreements with customers in various forms, when the company decides to use data to gain benefits by infringing consumers' privacy, consumers often find it difficult or cannot bear the cost of safeguarding rights and interests. Companies often infringe on consumers' privacy in a wide range, and consumers, as independent individuals, can hardly bear the responsibility to fight against companies.

In terms of property rights, the transaction information between customers and the company does not have a clear definition of ownership, and in general, it is only collected and utilised by the company. Since the revenue generated is entirely attributable to the company, there is a dispute over the fairness of the data property rights. For a single transaction event, both parties get a very small amount of data, but only the company can make full use of the transaction data at a low cost and generate revenue from the data. However, if the revenue generated by analysis is averaged into a single transaction, first of all, it is difficult to estimate the revenue generated by analysis, and second, the average revenue generated by each transaction is likely to be less than 1 cent. Therefore, despite the controversy, consumers are always a vulnerable group.

Reference List

48 Cart Abandonment Rate Statistics 2022 – Cart & Checkout –. (n.d.). Baymard Institute. Retrieved October 21, 2022, from https://baymard.com/lists/cart-abandonment-rate

Barbee, J., Jayakumar, J., Touse, S., & Venkataraman, K. (2022, April 13). *Retail's need for speed: Unlocking value in omnichannel delivery*. McKinsey & Company. Retrieved October 21, 2022, from https://www.mckinsey.com/industries/retail/our-insights/retails-need-for-speed-unlocking-value-in-omnichannel-delivery

Berson, S. (2022, October 4). *40 Strategies to Increase Ecommerce Sales in 2022*. Comrade Digital Marketing Agency. Retrieved October 21, 2022, from https://comradeweb.com/blog/strategies-to-increase-ecommerce-sales/

Briest, P., Dragendorf, J., Ecker, T., Mohr, D., & Neuhaus, F. (2020, November 25). *The endgame for postal networks: How to win in the age of e-commerce*. McKinsey & Company. Retrieved October 21, 2022, from https://www.mckinsey.com/industries/travel-logistics-and-infrastructure/our-insights/the-endgame-for-postal-networks-how-to-win-in-the-age-of-e-commerce

Cake, H. (2022, May 4). *What Are the Differences Between Benchmarking and KPIs?* Fully Accountable. Retrieved October 20, 2022, from https://fullyaccountable.com/what-are-benchmark-kpis/

Ehrens, T. (2019, April 4). *What is customer experience management (CEM or CXM)?* SearchCustomerExperience. Retrieved October 15, 2022, from https://www.techtarget.com/searchcustomerexperience/definition/customer-segmentation

Firestone, E. S. A. \. C. C. (2022, October 10). *Ecommerce KPIs: 67 Key Performance Indicators for Ecommerce*. Shopify. Retrieved October 20, 2022, from https://www.shopify.com/hk-en/blog/7365564-32-key-performance-indicators-kpis-for-ecommerce

Gennaro, L. (2022, June 24). *68 Useful eCommerce Statistics You Must Know in 2022*. WPForms. Retrieved October 21, 2022, from https://wpforms.com/ecommerce-statistics/

Gupta, R. K., Belkadi, F., & Bernard, A. (2017). Evaluation and management of customer feedback to include market dynamics into product development:

Satisfaction Importance Evaluation (SIE) model. In DS 87-4 Proceedings of the 21st International Conference on Engineering Design (ICED 17) Vol 4: Design Methods and Tools, Vancouver, Canada, 21-25.08. 2017 (pp. 327-336).

Patel, S. (2022, February 18). *15 Powerful Strategies to Increase Online Sales in 2022*. REVE Chat. Retrieved October 21, 2022, from https://www.revechat.com/blog/increase-online-sales/

Solving the paradox of growth and profitability in e-commerce. (2022, January 6). McKinsey & Company. Retrieved October 21, 2022, from https://www.mckinsey.com/industries/retail/our-insights/solving-the-paradox-of-growth-and-profitability-in-e-commerce