Content

Operation Strategies Based on Data Analysis for an Online retail store

1. Abstract

This report extensively analyses an online retail dataset to optimize the online retail store's operation strategies (Panda-monium, 2024). We have performed sales and product analysis to identify peak hours and sales trends over the months. Additionally, we observed that the United Kingdom has more sales than other countries, especially on weekdays. We also applied RFM (Recency, Frequency and Monetary) analysis to segment customers and recommend specific operation strategies for some segments of customers.

2. Introduction

In recent years, the number of online retail stores has increased by 26.5% in the United Kingdom (Centre for Retail Research, n.d.) due to the pandemic. This number is expected to remain the same since most customers have become accustomed to buying products online. Online businesses often need help meeting varying customer demands and high competition.

This project comprehensively analyzes an online retail store using a computer programming language (Python). We Analyse the raw retail data (Panda-monium, 2024) to assist in enhanced decision-making, uncover hidden insights, and study critical metrics such as sales trends, customer behavior, and product performance. For example, understanding which products contribute to the highest revenue or identifying peak purchasing time can help online retailers manage their inventory. The customer behavior analysis helps us understand the customer's frequency and interest in particular products to tailor discounts and target promotions to improve customer satisfaction and drive repeat purchases. This analysis can improve the business considerably by providing data-driven strategic recommendations to identify underperforming products, understand the customer segments, and personalize marketing strategies. This project employs analytics and programming techniques to identify challenges faced by retail stores and provide actionable strategies that increase revenue and competitiveness within the market.

3. Data Processing

The data cleaning process was done to improve the accuracy and consistency of the dataset for further use in visualization and analysis:

1. Since ‘CustomerID’ column has many null values, the rows with null values in ‘CustomerID’ column were deleted.
2. ‘Description’ column also has many null values. Part of these null values was found and filled by the corresponding 'Description' values from rows with the same 'StockCode'. Since 'description' column has low impact on data analysis, the remaining null values of 'description' column are simply filled in according to the corresponding 'Stockcode' value.
3. 5268 duplicate rows were found and eliminated to help reduce the possibility of bias in data analysis.
4. To make data easier to analyze, some characteristics of the data (e.g. max value, average value, etc.) were found and shown.

4. Data Analysis and Strategies Recommendations

4.1 Sales Trend Analysis

4.1.1 Monthly Sales Trend

We determined the monthly sales trends and found out sales increased tremendously after August, peaked in November, and then came down again, as shown in Figure 1.

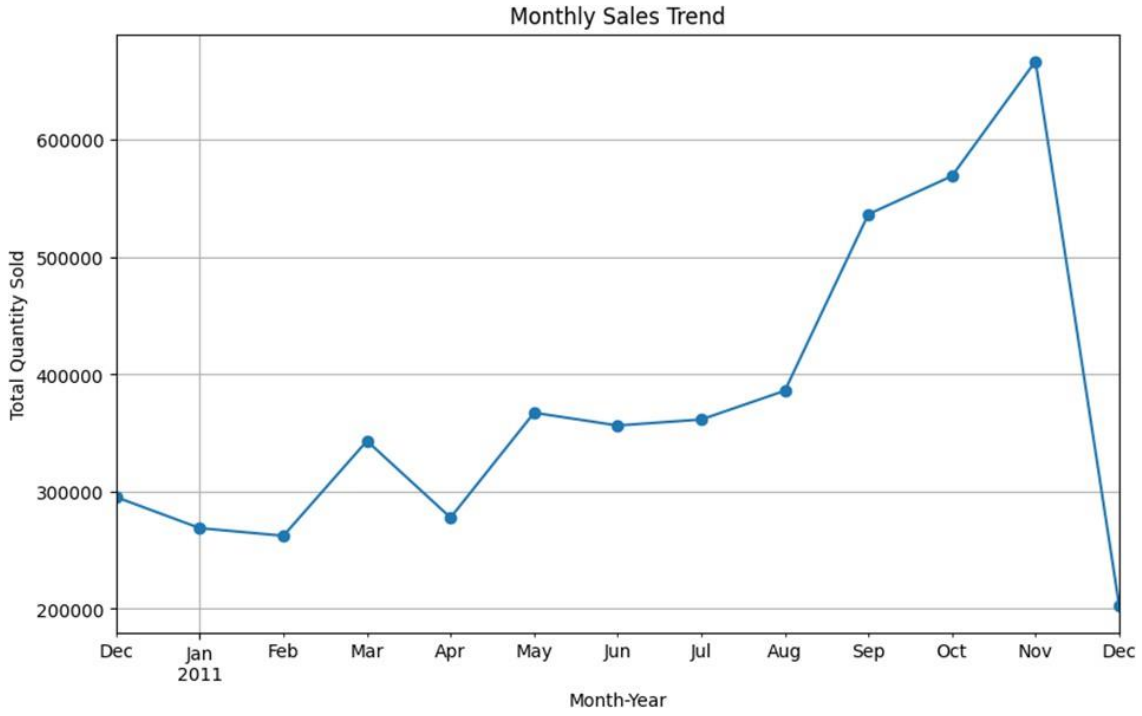

Figure 1: Monthly Sales Trend

During this period from August to November, the 20 most frequently bought items (Table 1) highlight a significant focus on gifts and decorations, indicating the preparations for the holiday season, and the top 5 countries of sales are mentioned below (Table 2).

Product	Total Quantity Sold
POPCORN HOLDER	22509
RABBIT NIGHT LIGHT	20441
JUMBO BAG RED RETROSPOT	19321
WORLD WAR 2 GLIDERS ASSTD DESIGNS	18296
ASSORTED COLOUR BIRD ORNAMENT	17194
ASSTD DESIGN 3D PAPER STICKERS	12632
60 CAKE CASES VINTAGE CHRISTMAS	11423
PAPER CHAIN KIT 50'S CHRISTMAS	11288
RED HARMONICA IN BOX	11155
WHITE HANGING HEART T-LIGHT HOLDER	10950
MINI PAINT SET VINTAGE	10791
PACK OF 72 RETROSPOT CAKE CASES	10253
JUMBO BAG VINTAGE DOILY	9413
JUMBO BAG PINK POLKADOT	8848
WOODEN HEART CHRISTMAS SCANDINAVIAN	8755
WOODEN STAR CHRISTMAS SCANDINAVIAN	8406
BUBBLEGUM RING ASSORTED	8185
DISCO BALL CHRISTMAS DECORATION	8141
PACK OF 12 LONDON TISSUES	8124
VICTORIAN GLASS HANGING T-LIGHT	7898

Table 1: Top Selling products from August to November

```
Top 5 countries by sales quantity (August–November)
 Country
United Kingdom    1773199
Netherlands         92308
EIRE                60399
France              51142
Germany             51117
Name: Quantity, dtype: int64
```

Table 2: Top 5 countries by Sales (Aug-Nov)

Based on the seasonal sales trend, we can enhance our seasonal inventory planning by stocking up on popular products in advance to minimize stockouts during peak sales periods and in the regions with high sales.

We can plan promotional campaigns around this period, give discounts on low-selling products to boost their sales, and offer bundling deals of low-selling products (Table 3) with high-selling products to further increase revenue.

Product	Total Quantity Sold
WHITE TALL PORCELAIN T-LIGHT HOLDER	-208
Discount	-36
VINTAGE BILLBOARD TEA MUG	-17
VINTAGE RED TEATIME MUG	-16
CRUK Commission	-15
PINK METAL CHICKEN HEART	-11
GLASS BEURRE DISH	-8
WHITE 3 FRAME BIRDS AND TREE	-1
BLUE EASTER EGG HUNT START POST	-1
YELLOW EASTER EGG HUNT START POST	-1

Table 3: Least Selling products from August to November

4.1.2 Geographic Trends

After calculating each region's monthly sales, we obtained a list of countries with their average monthly growth rates to get the countries to focus on for an increase in revenue as shown in Table 4.

Highest Growth Countries (Average Monthly Growth Rate):

Country	
Japan	611.061981
Netherlands	243.255870
Australia	7.435574
Italy	7.135411
Sweden	3.287162
Norway	2.143399
Finland	1.760869
Austria	1.406076
Channel Islands	0.785140
Unspecified	0.783367
Denmark	0.659959
Poland	0.644888
European Community	0.547737
Canada	0.542783
Portugal	0.496154
Switzerland	0.465699
Spain	0.439922
France	0.316128
USA	0.300462
EIRE	0.295435
Germany	0.110387
Iceland	0.088693
Belgium	0.057371
United Kingdom	0.027944
Singapore	0.009688
RSA	0.000000
Brazil	0.000000
Bahrain	0.000000
Lithuania	0.000000
Lebanon	0.000000
United Arab Emirates	-0.002016
Greece	-0.058666
Saudi Arabia	-0.106250
Israel	-0.173215
Malta	-0.272291
Czech Republic	-1.156913
Cyprus	-3.367370

Table 4: Average monthly growth rates

Even though the total quantity sold is higher in the Netherlands, Japan has a higher monthly growth rate because sales are very consistent with minor fluctuations as shown in Figure 2. On the other hand, the Netherlands has inconsistent sales with drastic sales dips in April and July, making Japan our focal point for consistent and reliable economic growth.

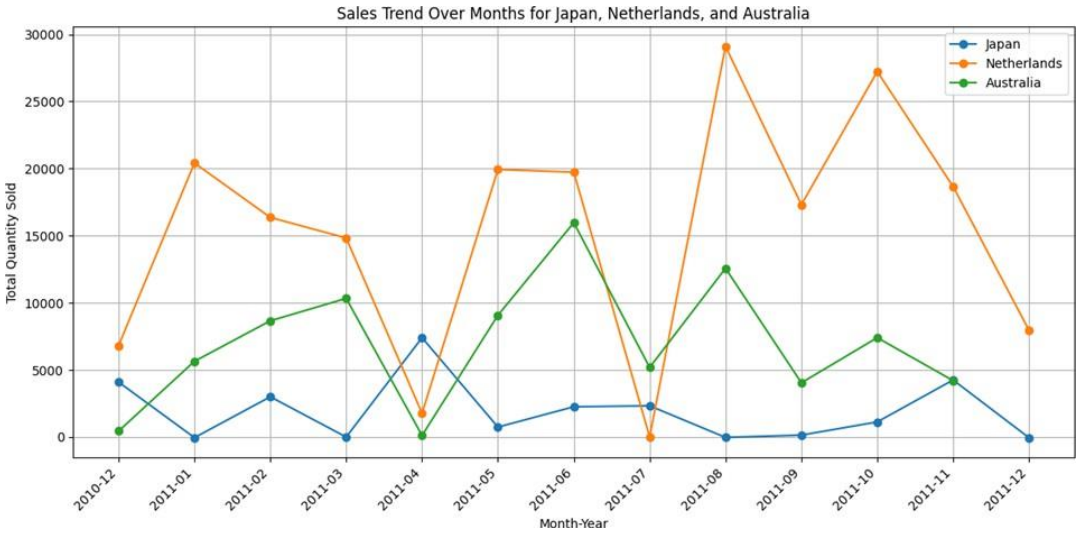

Figure 2: Sales Over Months for Japan, Netherlands and Australia

We also discovered that stationary items are the highest-selling products in Japan (Table 5), so investing in products in that category will further drive sales and increase revenue for the Business. Also, investing in warehouses to stock up on stationery products will ensure faster delivery, improve customer satisfaction, and increase customer retention while generating more revenue.

Product	Total Quantity Sold
RABBIT NIGHT LIGHT	3401
PACK OF 12 TRADITIONAL CRAYONS	1201
ROUND SNACK BOXES SET OF 4 FRUITS	870
WORLD WAR 2 GLIDERS ASSTD DESIGNS	577
MINI PAINT SET VINTAGE	577
MAGIC DRAWING SLATE BAKE A CAKE	577
MAGIC DRAWING SLATE CIRCUS PARADE	577
MAGIC DRAWING SLATE PURDEY	577
MAGIC DRAWING SLATE SPACEBOY	577
SET 12 KIDS COLOUR CHALK STICKS	576
PACK OF 12 COLOURED PENCILS	576
MINI JIGSAW CIRCUS PARADE	481
MINI JIGSAW PURDEY	481
MINI JIGSAW SPACEBOY	481
MINI JIGSAW BAKE A CAKE	481
CHARLOTTE BAG DOLLY GIRL DESIGN	398
BASKET OF TOADSTOOLS	372
WOODEN SCHOOL COLOURING SET	361
LUNCH BAG DOLLY GIRL DESIGN	359
PACK OF 12 LONDON TISSUES	336

Table 5: Highest selling products in Japan

4.2 Product Analysis

4.2.1 Top 10 Best-selling products

This analysis aims to identify high-demand products, manage inventory, understand customer behavior, and formulate marketing strategies.

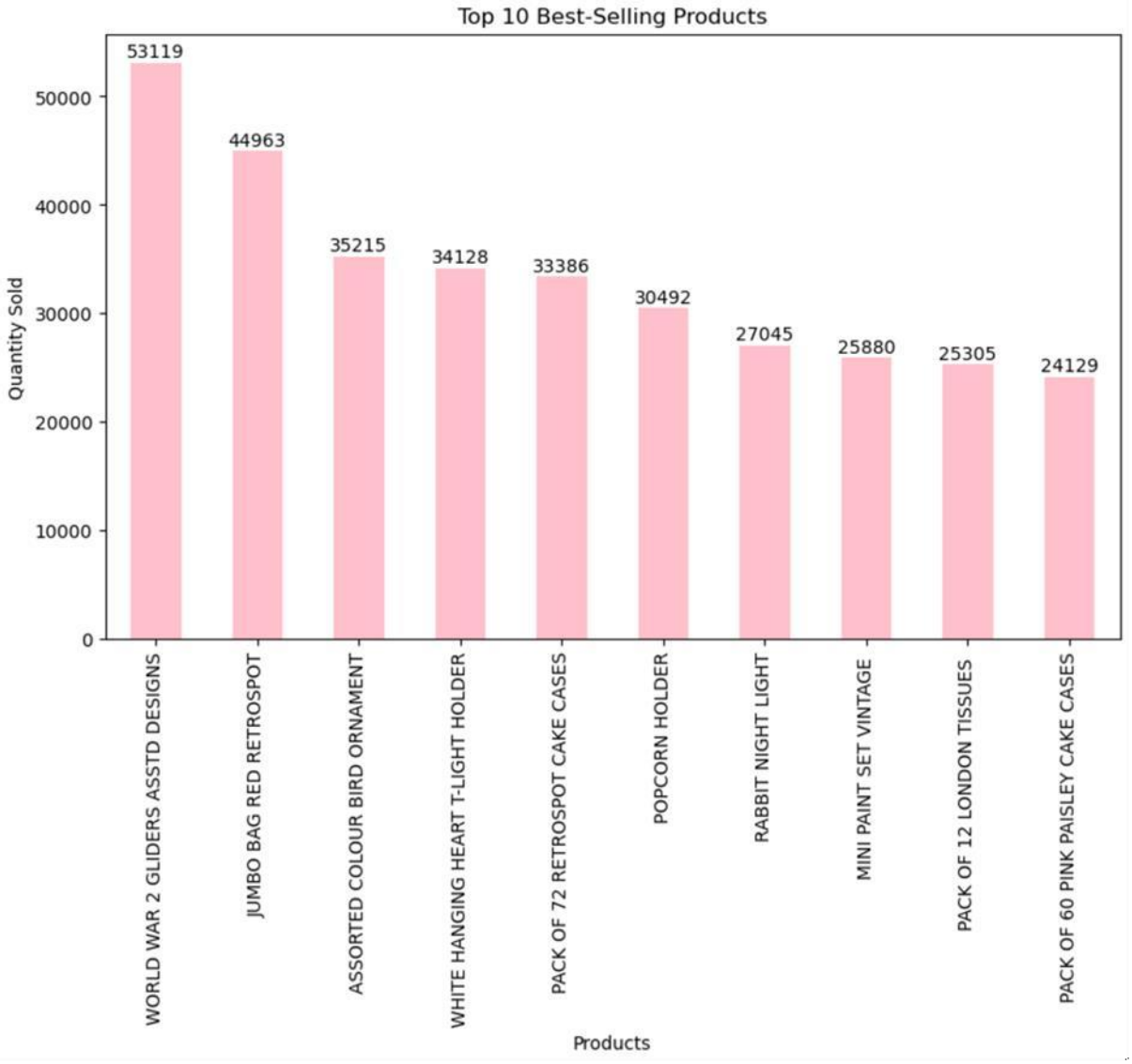

Figure 3: Top 10 Best-Selling Products-By Quantity

Figure 3 shows that 'World War 2 gliders asstd designs', 'Jumbo bag red retrospot' and 'Assorted colour bird ornament' as the top 3 highest-selling products. This shows that customers prefer buying decorative items and everyday household products. There is also a high demand for kitchen items such as 'Pack of 72 retrospot cake cases', 'Popcorn holder', 'Pack of 12 London tissues',

and ‘Pack of 60 pink paisley cake cases’. This trend suggests that customers are interested in unique decorative items and items used for baking and partying.

Always ensure that there is an adequate stock of high-demand items such as decorative products, baking accessories, and party products. Prioritize restocking these items during peak sales seasons to prevent stockouts and meet customer demand.

Suggest similar products that have similar designs, give out additional kitchen/baking products as complimentary or Bundle party accessories, and sell them at a discounted price. Implementing these strategies will lead to sales growth.

4.2.2 Analysis of the 10 Lowest Revenue Generating Products

This analysis focuses on finding the 10 lowest revenue-generating products to understand why they are underperforming and formulate suitable strategies to address the problems. Figure 4 shows that ‘The chalkboard kitchen organizer’ is the least revenue-generating item. Besides, excessive decorative lights, such as ‘White cherry lights’, ‘Pink cherry lights’, and ‘Antique lily fairy lights’ generate low revenue. The graph also shows other utility items and gift items such as a ‘Cream sweetheart magazine rack’, ‘Ream sweetheart trays’, ‘Blue padded soft mobile’, ‘Wooden box advent calendar’ and ‘Tea time cake stand’ in a gift box. These items may have negative revenue due to returns, excessive discounts, or damaged items.

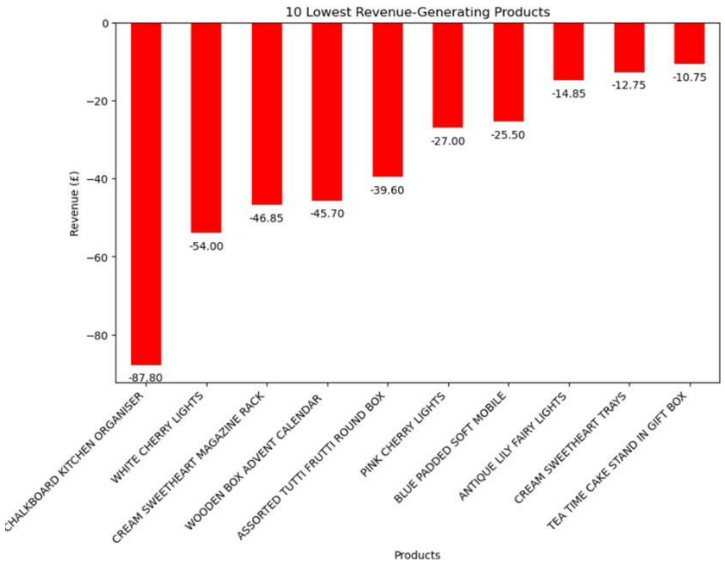

Figure 4: 10 Lowest Revenue Generating Products

We can offer special promotions and discounts for these items. For example, we could offer a buy one, get one offer for the lights during holidays and festivals when the demand for decorative items is high. We could also bundle Advent calendars with other high-performing seasonal products.

Creating gift sets for underperforming items with complementary products. For example, tea sets with ‘Teatime cake stand in gift box’.

Show low-performing products in the “Recommended for you” sections to boost sales. Encourage users to leave positive reviews on the products to attract more customers. If these strategies do not work, consider discontinuing these items or replacing them with better alternatives.

4.3 RFM Analysis

According to ‘Analytical approaches in customer relationship management’ written by Şahin et al. (2024), we realized that the RFM analysis can also be applied to our research.

Since Şahin et al. (2024) mention that RFM analysis can provide an in-depth analysis of customers in terms of three dimensions (time since the last purchase, frequency of purchase, amount of money spent) and this method is a powerful tool for customer segmentation, we select RFM analysis develop operation strategies to improve online retail store’s operating condition.

By reading Miglautsch (2000)'s ‘Thoughts on RFM scoring’, which introduces the detailed steps of RFM analysis, we successfully applied RFM analysis to develop effective operational strategies. The detailed process of our RFM analysis and the proposed marketing strategies based on this method will be presented in detail:

Initially, we calculated the time since the last purchase, the total number of orders, and the total spend amount for each customer to create ‘Recency’, ‘Frequency’, and ‘Monetary’ columns. Subsequently, we visualized the distributions of these three variables, resulting in Figures 5 and Figure 6.

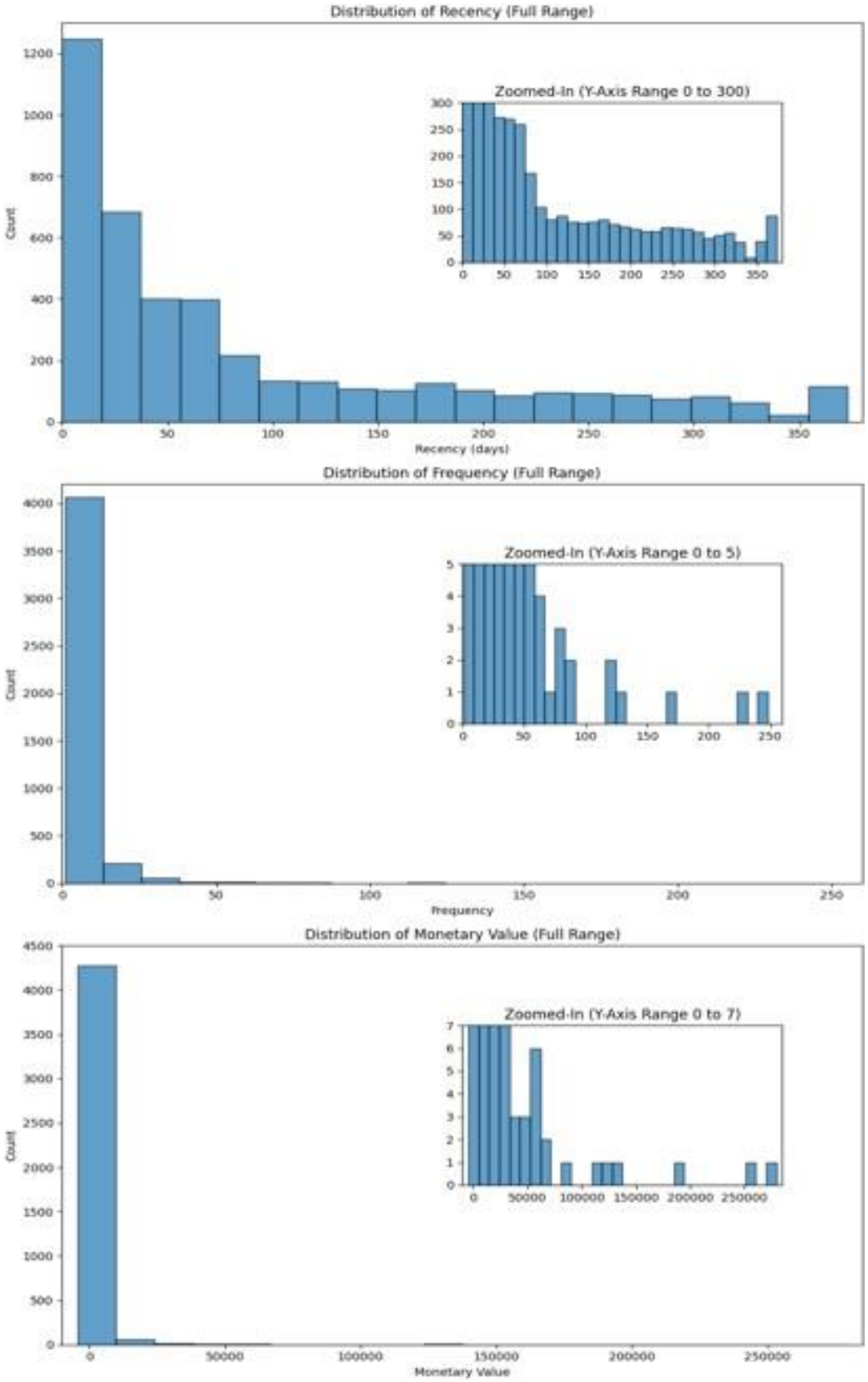

Figure 5: Distribution of Recency, Frequency, and Monetary Value (With Zoomed-In View)

Figure 5 shows most customers have made purchases within the last 30 days, majority of customers' total order numbers are less than 10. In terms of the amount of money spent, and most customers total spend in this online retail store less than 10000£.

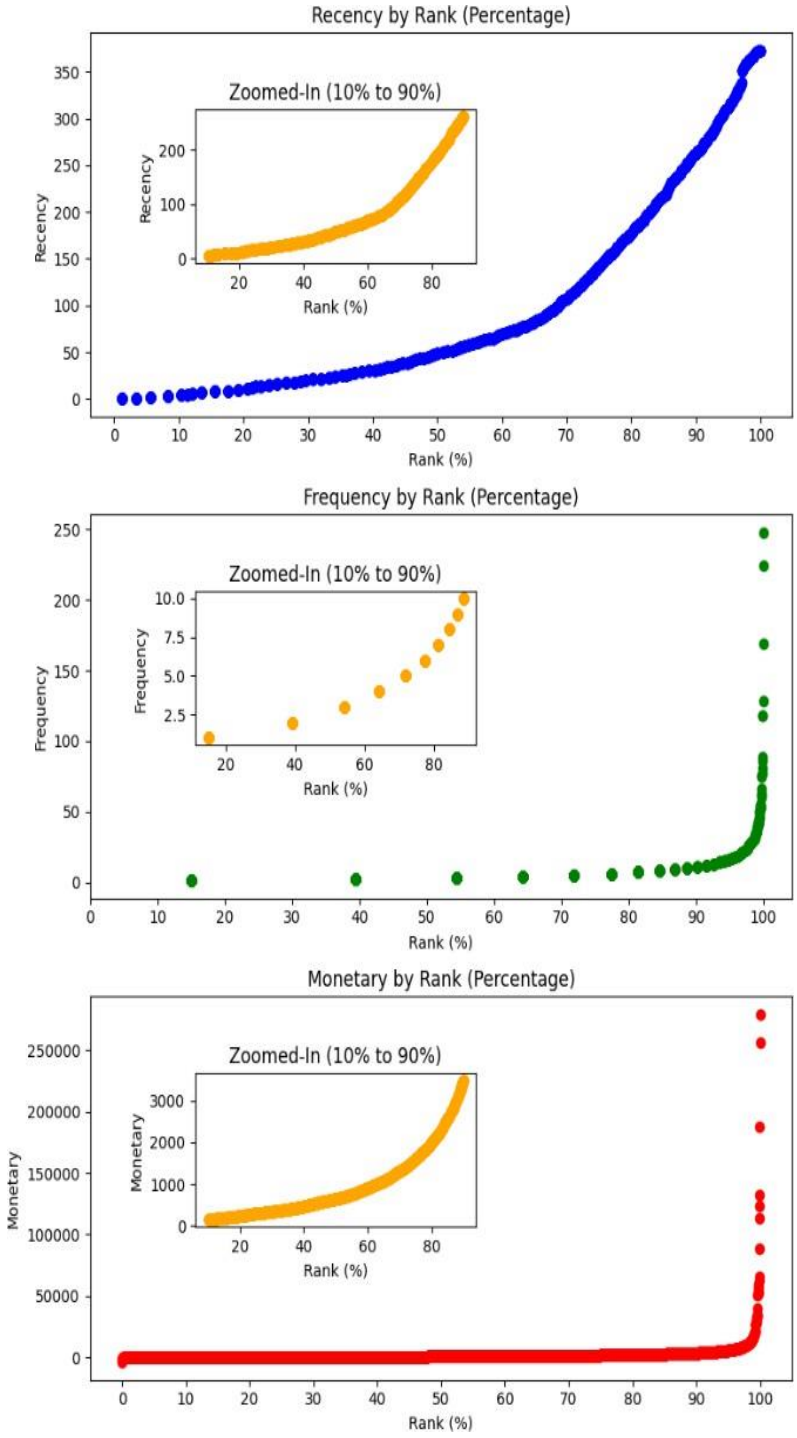

Figure 6: Recency, Frequency, and Monetary Value by Rank (With Zoomed-In View)

The overall distributions of Frequency and Monetary in Figure 6 reveal nearly 1% of customers have exceptionally high Frequency and Monetary values. However, according to the zoomed-in views, most customers show a continuous trend, with no significant change of slope at any point.

Therefore, we applied the quantile-based scoring method to divide ‘Recency’, ‘Frequency’, and ‘Monetary’ columns into three groups and created 4 new columns (Table 6):

1. Dividing ‘Recency’ column into three groups from highest to lowest and assigning values 1, 2, and 3 to generate column ‘R Score’.
2. Dividing ‘Frequency’ column into three groups from lowest to highest and assigning values 1, 2, and 3 to generate column ‘F Score’.
3. Dividing ‘Monetary’ column into three groups from lowest to highest and assigning values 1, 2, and 3 to generate column ‘M Score’.
4. Combining ‘R Score’, ‘F Score’, and ‘M Score’ columns to create column ‘RFM Score’.

CustomerID	R_Score	F_Score	M_Score	RFM_Score
12346.0	1	1	1	111
12347.0	3	3	3	333
12348.0	2	2	3	223
12349.0	3	1	3	313
12350.0	1	1	1	111

Table 6: Dataset after processing

Then, we analyzed the real meaning of each kind of RFM Score to create a customer segments table (Table 7).

RFM Score	Segment	Description
33x	Loyal	Loyal Customers (Bought recently and most f
xx3	High-Value	Big Spenders (Spent Most)
31x	New	New Customers
1xx	Lost	Customers who have churned
2xx	Losing	Customers with potential churn risk
xx1	Low-Spending	Customers who spend less
x3x	High-Frequence	Customers who purchase frequently
11x	Accidental customers	Accidental customers
x1x	Low-Frequence	Customers who purchase infrequently
322	Ordinary	Ordinary Customers

x can be 1, 2 and 3

Table 7: Customer Segments Based on RFM Score

Based on Table 7, we generated Figure 7 (RFM customer segment distribution) and CustomerID-Segment table (Table 8).

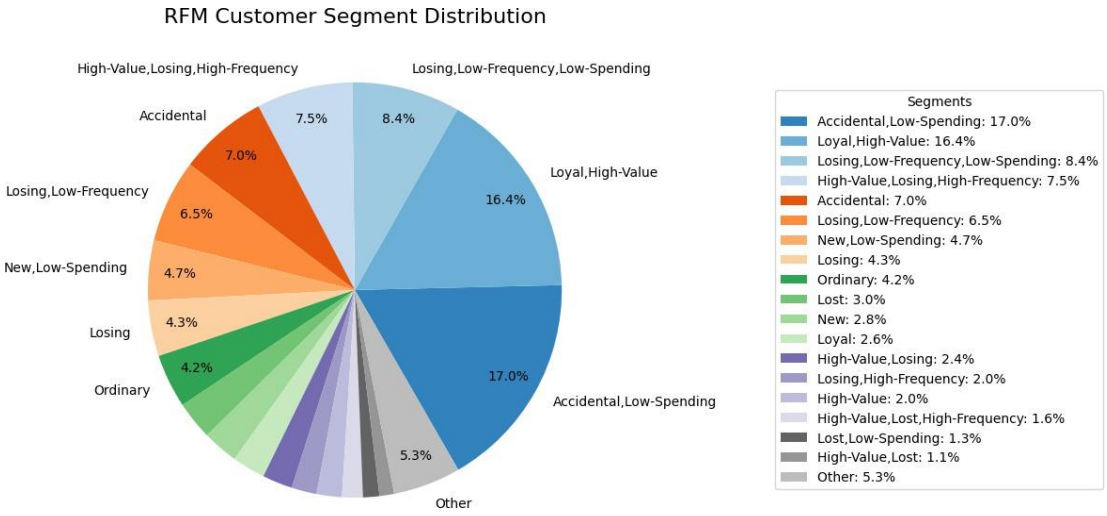

Figure 7: RFM Customer Segment Distribution

Through Figure 7, the online retail store can gain a clearer understanding of its overall customer situation in time, such as when a lot of customers in a situation of near churn, the online retail store can find this situation in time with the help of Figure 6 and use effective operational strategies to solve it.

CustomerID	Segment
12346.0	Accidental,Low-Spending
12347.0	Loyal,High-Value
12348.0	High-Value,Losing
12349.0	High-Value,New
12350.0	Accidental,Low-Spending
12352.0	High-Value,Losing,High-Frequency
12353.0	Accidental,Low-Spending
12354.0	Accidental
12355.0	Accidental
12356.0	High-Value

Table 8: CustomerID-Segment Table

With CustomerID-Segment Table (Table 8), the online retail store can master the detailed segment of each customer.

Based on customer segments (Table 7), we suggest the online retail store use these operation strategies:

1. For high-valuable (xx3) and Loyal (33x) customers, online retail stores can provide VIP services to them, which includes a personal assistant to improve the experience of shopping. Subsequently, online retail stores can collect suggestions from this kind of customer to maintain them.
2. For low-frequency (x1x) customers, online retail stores can provide time-limit coupons to increase their desire to purchase.
3. For losing (2xx) customers, online retail stores can provide discounts to enhance their willingness to purchase and feedback questionnaires to find the reason why they are losing.
4. For lost (1xx) but high-value (xx3) customers, online retail stores can provide them discounts and notify them with emails and messages.

5. For loyal (33x) but low-spending (xx1) customers, it’s useful to provide some discount if customers’ consumption amount reaches a certain amount, to enhance their consumption amount per purchase.
6. For new (31x) customers, online retail stores can give them some accumulated points that can be exchanged for coupons for their every consumption

5. Conclusion

In this report, we analyzed an online retail dataset and performed data cleaning and exploratory analysis to get actionable insights using Python programming. We identify that the United Kingdom is the leading region in sales and Japan has the highest growth rate, while "World War 2 gliders asst designs" are the most selling products. Through RFM analysis, we create effective methods to help the online retail store determine segments of each customer, grasp the overall sales situation based on sales data, and get some effective operation strategies. By implementing these insights and recommendations the business can optimize their processes, adopt data-driven operational strategies, and improve customer engagement and retention, which can further improve the growth of the online retail business.

References

Centre for Retail Research. (n.d.) Online Trends & Statistics for UK, Europe & N. America 2022-2023. *The Centre for Retail Research*. Available at: https://www.retailresearch.org/online-retail.html (Accessed: 15 November 2024).

Miglautsch, J. R., (2000). 'Thoughts on RFM scoring', *Journal of Decision Models*, 7(1), pp. 45-53. Available at: https://link.springer.com/article/10.1057/palgrave.jdm.3240019 (Accessed: 18 November 2024).

Panda-monium, (2024). *Online Retail Dataset*. Kaggle. Available at: https://www.kaggle.com/datasets/divanshu22/online-retail-dataset/data (Accessed: 16 November 2024).

Şahin, Z. et al., (2024). Analytical approaches in customer relationship management. *IEEE Transactions on Industrial Informatics*, 20(5), pp. 1234-1245. Available at: https://doi.org/10.1109/TII.2024.10391021 (Accessed: 19 November 2024).

Appendix

Project Code:

https://colab.research.google.com/drive/1poYM4iC3z-l1iXa5ZahcdGelZTpGapwj