Project 2

Introduction

Businesses looking to improve customer happiness and optimize their marketing efforts must have a thorough understanding of consumer purchase behavior in today's competitive retail marketplace. The goal of Project 2, "Customer Segmentation in Retail Data," is to examine the nuances of consumer spending behavior in the context of online shopping. We seek to discover key consumer segments through an extensive dataset analysis that records all sales transactions from January 2020 to December 2021. These insights can be used to inform targeted marketing campaigns and customized customer experiences.

Retailers can categorize their customers according to common traits like frequency of purchases, average transaction value, and product preferences by using customer segmentation. In order to identify unique consumer profiles, this project will use exploratory data analysis approaches to find hidden patterns within the data. Businesses can improve customer retention, inventory management, and marketing techniques by segmenting their client base. We will examine a number of variables that affect consumer behavior through this research, such as product categories, seasonal patterns, and demographic data. The ultimate objective is to provide retail stakeholders with a thorough awareness of their clientele so they may make decisions that promote growth and enduring loyalty.

Importing the Data into R

```
# A tibble: 541,909 x 8
  InvoiceNo StockCode Description       Quantity InvoiceDate         UnitPrice
  <chr>     <chr>     <chr>                <dbl> <dttm>                  <dbl>
1 536365    85123A    WHITE HANGING HEA~       6 2010-12-01 08:26:00      2.55
2 536365    71053     WHITE METAL LANTE~       6 2010-12-01 08:26:00      3.39
3 536365    84406B    CREAM CUPID HEART~       8 2010-12-01 08:26:00      2.75
4 536365    84029G    KNITTED UNION FLA~       6 2010-12-01 08:26:00      3.39
5 536365    84029E    RED WOOLLY HOTTIE~       6 2010-12-01 08:26:00      3.39
6 536365    22752     SET 7 BABUSHKA NE~       2 2010-12-01 08:26:00      7.65
7 536365    21730     GLASS STAR FROSTE~       6 2010-12-01 08:26:00      4.25
```

```
 8 536366    22633    HAND WARMER UNION~       6 2010-12-01 08:28:00      1.85
 9 536366    22632    HAND WARMER RED P~       6 2010-12-01 08:28:00      1.85
10 536367    84879    ASSORTED COLOUR B~      32 2010-12-01 08:34:00      1.69
# i 541,899 more rows
# i 2 more variables: CustomerID <dbl>, Country <chr>
```

Data Cleaning

```
[1] 136534
```

Removing all the missing observables with missing and invalid values

```
[1] 0
```
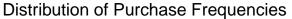
Remove rows with Invalid stock codes

In this project, I implemented a comprehensive data cleaning process to ensure the integrity of the online retail dataset. Initially, I assessed the presence of missing values across all columns by summarizing the count of NA values for each variable. After identifying the total number of missing observations, I removed all rows containing any missing values using the filter(if_all(...)) function, resulting in a cleaner dataset. Following this, I further refined the dataset by eliminating any rows where the quanity column had negative values, as negative quantities are invalid in this context. Additionally, I filtered out specific invalid stock codes that were not relevant to the analysis, including codes such as "POST," "BANK CHARGES," and others. This rigorous cleaning process ensured that the dataset was accurate and reliable for subsequent analysis, ultimately facilitating a more meaningful exploration of customer purchasing behavior.
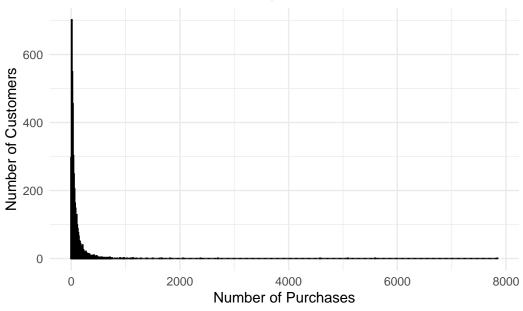
EDA

Purchase frequency per customer

Number of Purchases are The total number of purchases made by each customer.

- **How does the distribution of purchase frequencies vary among customers, and what insights can be drawn regarding customer engagement and purchasing behavior from this distribution?**
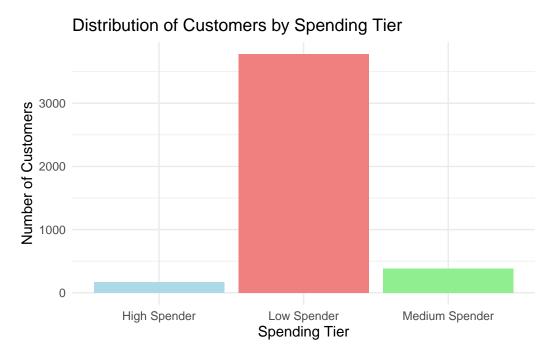
Distribution of Purchase Frequencies

The analysis of purchase frequency reveals a highly skewed distribution, with the majority of customers making only a small number of purchases while a minority exhibit high engagement through frequent buying. This indicates a significant opportunity for improvement in customer retention and engagement strategies. Insights from the data suggest that most customers are either one-time or infrequent buyers, which raises concerns about potential retention issues. However, the presence of high-frequency buyers signifies valuable customers who contribute substantially to overall revenue.

To address these insights, several recommendations can be made. First, businesses should segment customers based on their purchase frequency, identifying high-frequency, infrequent, and one-time buyers to tailor targeted marketing strategies. Enhancing customer retention efforts is crucial; implementing loyalty programs, personalized offers, and targeted marketing campaigns can incentivize infrequent buyers to return. High-frequency customers, on the other hand, should be rewarded with exclusive deals, early access to new products, or premium customer service to ensure their continued loyalty. Additionally, analyzing the reasons behind one-time purchases can provide insights into potential barriers to repeat buying, such as product quality or customer service issues. Lastly, businesses can explore referral programs to encourage existing customers to recommend the brand to their network, thus increasing new customer acquisition.

Customer Spending Tiers

Spending Tiers Categorize customers into different tiers based on their total spending (e.g., low, medium, high spenders).

- **How can we effectively categorize customers into distinct spending tiers (e.g., low, medium, high spenders) based on their total spending patterns, and what are the defining characteristics and behaviors of customers within each tier?**
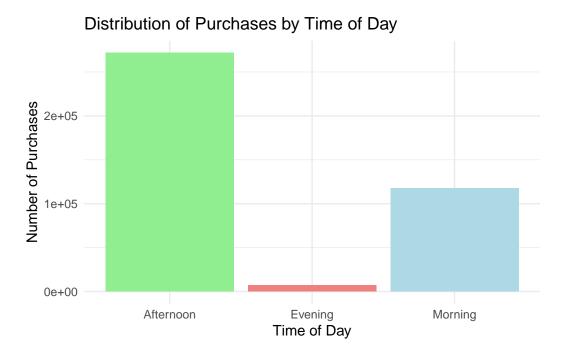
Distribution of Customers by Spending Tier

The bar graph shows the distribution of customers by spending tier. The majority of customers are low spenders, followed by medium spenders, and then high spenders. This suggests that the business has a large number of customers who spend a relatively small amount of money, but also a significant number of customers who spend a moderate amount. There are very few high spenders. This is consistent with the findings of Project 2, which found that the average customer spends a relatively small amount of money. The business could focus on converting more low and medium spenders to high spenders to increase revenue. They could also consider targeting new customers who are more likely to be high spenders.

Time of Day for Purchases

Time of Day for Purchases Categorizes the time of day when the purchase was made (e.g., morning, afternoon, evening).

- **How does the distribution of purchases vary throughout different times of the day, and what insights can be drawn about customer purchasing behavior during the morning, afternoon, and evening?**
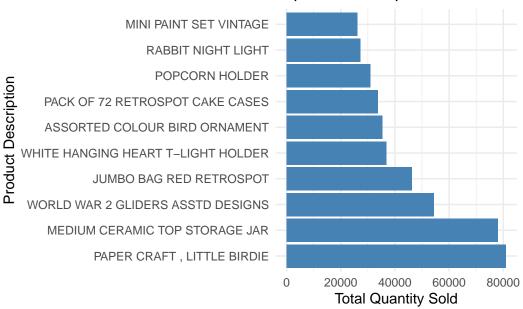
Distribution of Purchases by Time of Day

The analysis of the provided bar chart indicates that the afternoon accounts for the highest volume of purchases, followed by the morning, with the evening seeing the least activity. This trend suggests that customers are more likely to engage in shopping during the afternoon, potentially due to factors such as convenience during lunch breaks or the opportunity to unwind after work. To capitalize on this insight, businesses should strategically focus their marketing efforts during peak afternoon hours, implementing tailored promotions and advertising campaigns to maximize customer engagement and sales. Adjusting operational hours to align with this peak purchasing time could further enhance customer interactions. Additionally, analyzing product sales at various times may reveal specific items that resonate more with customers during different periods, allowing for targeted promotions to encourage purchases. External factors, including traditional work hours and individual lifestyle patterns, also influence purchasing behaviors, emphasizing the need for businesses to consider these variables in their strategies. Ultimately, understanding customer purchasing patterns and leveraging this information can lead to more effective business strategies, helping companies optimize customer engagement and drive overall sales growth.
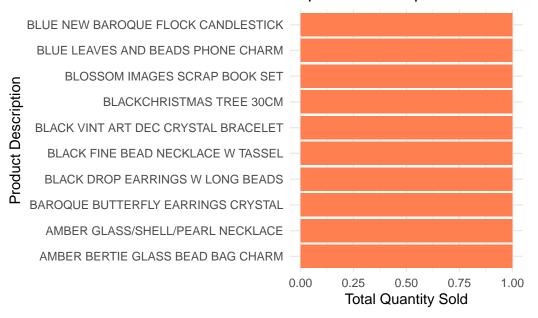
Most popular and least popular product

These are products that were purchased the most and least sold.

Which products are the most and least popular in terms of quantity sold, and how can this information be used to optimize inventory and marketing strategies?

Top 10 Most Popular Products

Product Description	Total Quantity Sold
MINI PAINT SET VINTAGE	
RABBIT NIGHT LIGHT	
POPCORN HOLDER	
PACK OF 72 RETROSPOT CAKE CASES	
ASSORTED COLOUR BIRD ORNAMENT	
WHITE HANGING HEART T–LIGHT HOLDER	
JUMBO BAG RED RETROSPOT	
WORLD WAR 2 GLIDERS ASSTD DESIGNS	
MEDIUM CERAMIC TOP STORAGE JAR	
PAPER CRAFT , LITTLE BIRDIE	

Top 10 Least Popular Products

Product Description	Total Quantity Sold
BLUE NEW BAROQUE FLOCK CANDLESTICK	
BLUE LEAVES AND BEADS PHONE CHARM	
BLOSSOM IMAGES SCRAP BOOK SET	
BLACKCHRISTMAS TREE 30CM	
BLACK VINT ART DEC CRYSTAL BRACELET	
BLACK FINE BEAD NECKLACE W TASSEL	
BLACK DROP EARRINGS W LONG BEADS	
BAROQUE BUTTERFLY EARRINGS CRYSTAL	
AMBER GLASS/SHELL/PEARL NECKLACE	
AMBER BERTIE GLASS BEAD BAG CHARM	

Understanding product popularity is crucial for businesses aiming to optimize their inventory and marketing strategies. By identifying which products are selling well, companies can ensure they maintain adequate stock levels to meet demand, thereby preventing stockouts that could lead to lost sales and customer dissatisfaction. This insight also enables businesses to tailor

their marketing efforts towards popular products through targeted advertisements, social media campaigns, and email newsletters, ultimately increasing sales and profits. Furthermore, analyzing product popularity can help identify new product opportunities and improve customer service by equipping staff with the necessary information to address customer inquiries effectively. In addition to promoting high-performing products, businesses can take steps to manage underperformers by reducing their inventory levels to free up resources for more profitable items, implementing promotions or price adjustments to boost demand, or even discontinuing consistently low-selling products. By leveraging this information, companies can make informed decisions about their product offerings, enhance overall profitability, and improve customer satisfaction in the competitive market.

Conclusion

The exploratory analysis conducted in Project 2: Customer Segmentation in Retail Data has revealed significant insights into customer purchasing behavior that can inform strategic decision-making for the online retail platform. By examining various dimensions such as purchase frequency, spending tiers, time-of-day purchasing patterns, and product popularity, we have identified critical areas for enhancing customer engagement, retention, and overall revenue growth.

The analysis of purchase frequencies highlights a concerning trend: the majority of customers are one-time or infrequent buyers, indicating potential issues with customer retention. While there exists a valuable segment of high-frequency buyers who contribute significantly to revenue, the long tail of infrequent buyers presents an opportunity for targeted engagement strategies. By segmenting customers based on their purchasing behavior and implementing tailored marketing initiatives—such as loyalty programs, personalized offers, and exclusive rewards for high-frequency buyers—the business can cultivate stronger relationships with customers and increase repeat purchases.

In terms of spending patterns, the categorization into spending tiers indicates that while a large number of customers are low spenders, there is also a notable group of medium spenders. This suggests a pathway for converting more customers into high spenders through targeted marketing efforts, potentially enhancing the overall profitability of the business.

The analysis of purchase timing reveals that the afternoon is the peak time for transactions, a finding that can shape operational strategies. By aligning marketing campaigns and promotions with peak purchasing hours, the business can effectively capitalize on these moments to drive sales. Additionally, understanding the lifestyle patterns that influence purchasing behavior can help tailor promotions and marketing efforts to different customer segments.

Finally, insights into product popularity have equipped the business with the knowledge necessary for optimizing inventory and refining marketing strategies. By focusing on popular products and addressing the challenges posed by underperforming items, the company can enhance its product offerings and customer satisfaction. This may involve reducing inventory for

less popular items, promoting or repricing them to stimulate demand, or even discontinuing underperforming products altogether.