

# **Customer Segmentation and Insights Analytics**

# AGENDA

- Overview of the dataset
- Data cleaning steps
- RFM analysis and customer segmentation
- Insights and recommendations
- Visualizations (charts and graphs)

# Overview of the dataset

## 1. **Invoice No:** Unique transaction identifier.

1. **Significance:** Differentiates individual transactions.

## 2. **Stock Code:** Product identifier

1. **Significance:** Identifies the product purchased.

## 3. **Description:** Product name/description.

1. **Significance:** Provides details about the product.

## 4. **Quantity:** Number of units purchased.

1. **Significance:** Indicates purchase volume.

## 5. **Invoice Date:** Date and time of transaction.

1. **Significance:** Useful for time-based trends (e.g., monthly or seasonal).

## 6. **Unit Price:** Price per unit of product.

1. **Significance:** Determines sales value of items purchased.

## 7. **Customer ID:** Unique customer identifier.

1. **Significance:** Tracks customer behavior and purchase patterns.

## 8. **Country:** Customer's location.

1. **Significance:** Analyzes sales by geographic region.

# Data Cleaning Steps

- **Handle Missing Data**

Use methods such as `isnull()` to detect missing data

- **Remove Duplicates**

: Use `drop_duplicates()` to detect and remove duplicate rows

- **Standardize Data Formats**

**Dates & Times:** Ensure consistency in date formats (e.g., YYYY-MM-DD) and convert strings to proper datetime objects.

- **Correct Data Types**

**Type Conversion:** Convert columns to appropriate data types (e.g., Quantity as Integer, InvoiceDate as Date).

- **Handle Outliers**

**Identify Outliers:** Use statistical methods (e.g., IQR) to identify outliers in numerical data.

- **Fix Inconsistent Data**

**Check for Inconsistencies:** Look for inconsistent naming (e.g., different spellings of the same country or product).

- **Handle Categorical Data**

**Encoding:** Convert categorical variables (e.g., Country, StockCode) into numerical formats using methods like one-hot encoding or label encoding.

- **Validate and Cross-Check**

**Cross-Verify:** After cleaning, validate the integrity of your dataset. Ensure no important information is lost.

# Insights from the Bivariate Analysis:

## . Purchase Volume vs. Purchase Amount Distribution

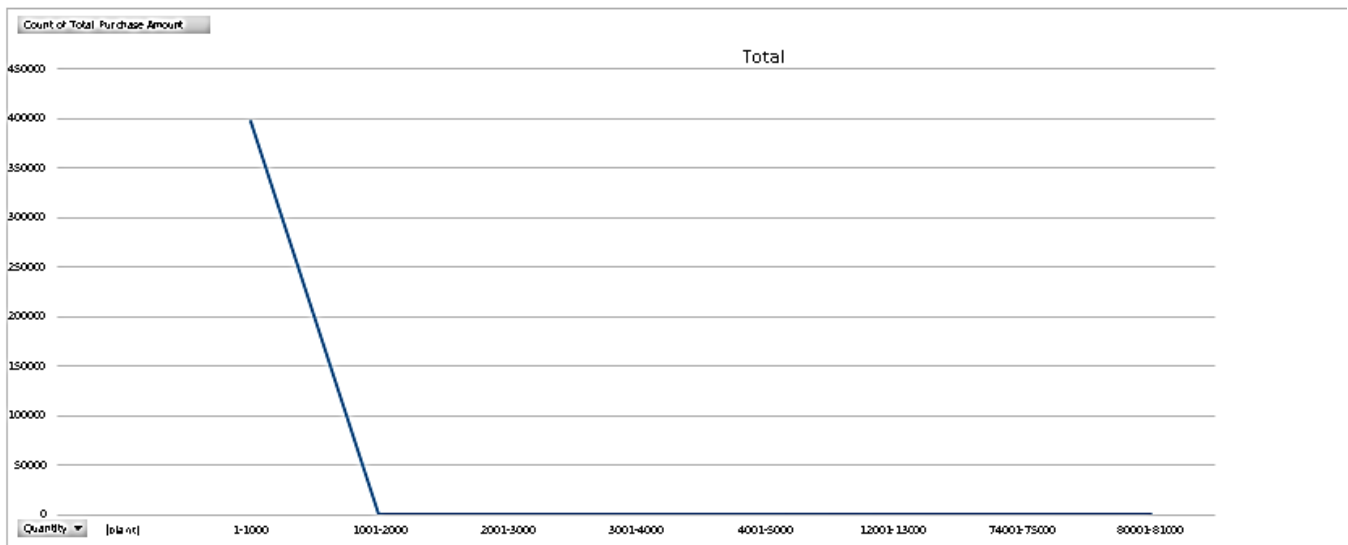
- **Most Purchases Fall in the Low Range (1-1000):**
  - **396,266 transactions** fall in the range of **1-1000 total purchase amount**.
  - This shows that the majority of purchases are relatively small, indicating that most customers tend to buy fewer items or lower-priced products.
- **Few Transactions in Higher Ranges:**
  - As the **purchase amount increases**, the number of transactions decreases significantly.
  - For example, the next significant range, **1001-2000**, only has **83 transactions**.
  - This indicates that large purchase amounts are rare and only a small fraction of customers spend significantly more.

## 2. Increasing Purchase Amount with Fewer Transactions

- The numbers quickly drop off as the purchase amount increases, suggesting that higher purchase amounts are outliers.
  - **1 transaction in the 12001-13000** range.
  - **1 transaction in the 74001-75000** and **80001-81000** ranges.
  - This indicates that only a very small number of customers are making large purchases (likely bulk or premium items).

## 3. Patterns in Customer Behavior:

- **Majority of Customers Have Low Purchase Amounts:**
  - The bulk of customers are making purchases in the lower value range, which could indicate more frequent, smaller purchases (perhaps day-to-day essentials or low-cost items).
- **High-Spending Customers are Outliers:**
  - A very small number of customers fall into the higher purchase ranges, likely indicating bulk buying, or special offers, or high-ticket items purchased in one go. These customers might represent a smaller, high-value segment of your customer base.



**Line Graph Insights:** If you plot this data on a line graph:

- **Sharp Decline:** You'll observe a steep decline in the number of transactions as the purchase amount increases. The line will show a sharp drop after the 1-1000 range.
- **Flat or Slow Decline in Early Stages:** In the lower ranges (e.g., 1-1000), the number of transactions stays relatively high, indicating stable purchase behavior in this segment.

## Customer ID vs. Total Purchase Amount

### 1. Top Customers by Total Purchase Amount

- **Top 3 Customers with Highest Spending:**
  - **Customer 14646:** \$279,138.02
  - **Customer 18102:** \$259,657.30
  - **Customer 17450:** \$194,550.79

These top three customers account for a significant portion of the total sales, indicating they are high-value customers.

### 2. High-Spending Customer Segment:

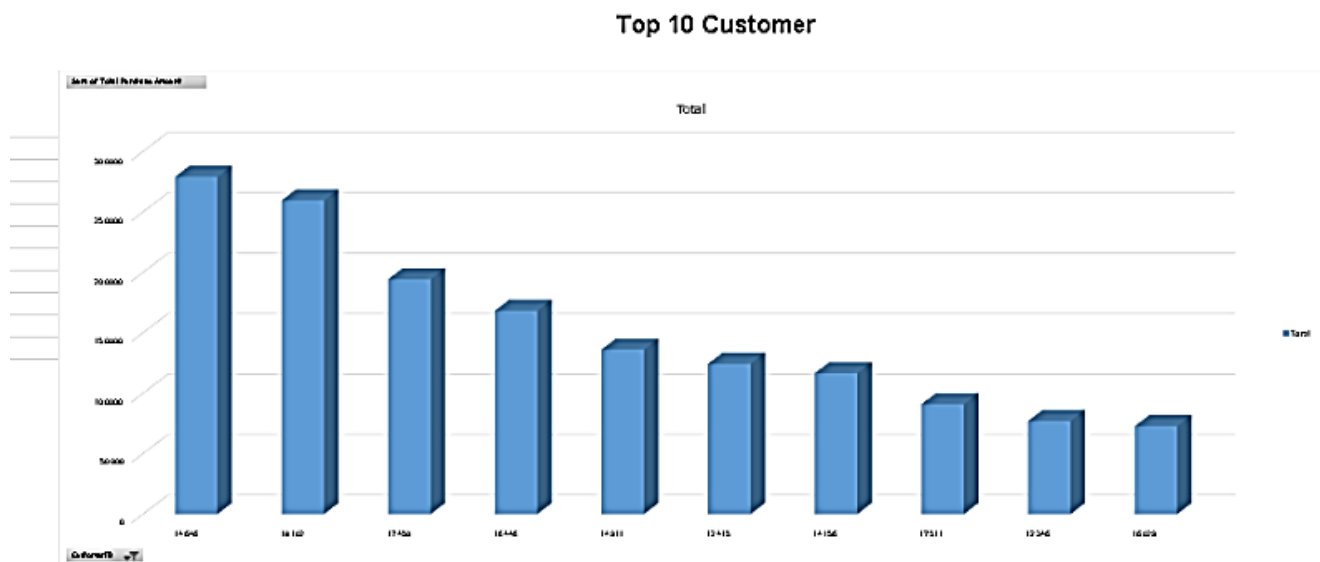
- The **first five customers** (14646, 18102, 17450, 16446, 14911) show relatively large purchase amounts ranging from **\$279,138.02 to \$136,275.72**. These customers represent a high-spending segment, and they likely contribute a large share of overall revenue.
- These customers could be key targets for **personalized marketing, exclusive offers, or loyalty programs** to retain them and increase their lifetime value.

### 3. Long Tail of Spending:

- **Customers from 12415 to 16029** have decreasing total purchase amounts, from **\$124,564.53 to \$72,882.09**. This indicates that the top 10 customers are highly concentrated, with a few individuals contributing significantly to sales, while others contribute less in comparison.

### 4. Customer Concentration

- A small number of customers (the top 5) drive most of the revenue, while others (in the lower half of the top 10) make much smaller contributions.
- This highlights the **80/20 Rule (Pareto Principle)**, where roughly **20% of customers** generate **80% of the revenue**. These high-value customers should be prioritized.



## Recommendations Based on Insights:

### 1. Focus on Retaining High-Value Customers:

- Given that a small number of customers contribute significantly to the total purchase amount, prioritize building relationships with them. Consider offering them exclusive deals, personalized services, or loyalty rewards to encourage repeat business and customer retention.

## 2. Develop Targeted Marketing Strategies:

- **Top Spenders:** For the customers in the top 5, consider using targeted marketing campaigns or sending personalized offers to further incentivize them to make more purchases.
- **Middle Range Spenders:** Customers in the middle of the list could be incentivized to increase their spending through promotional offers like discounts or bundled deals.

## 3. Analyze Customer Behavior for Upselling:

- Since the highest spending customers are contributing the most, try to **identify patterns** in their buying behavior. Understand what kinds of products or services they purchase most often, and explore opportunities for **upselling** or offering high-margin items.

## 4. Segment and Tailor Strategies:

- You could create different segments based on **spending categories**:
  - **High-Value Segment:** (Top 5 customers) for exclusive marketing or services.
  - **Moderate Spend Segment:** (Middle 5 customers) for promotional offers.

Tailor your strategies for each segment to maximize customer retention and overall revenue.

## Country vs. Total Purchase Amount

### 1 Dominance of the United Kingdom:

- **The United Kingdom** significantly outperforms all other countries with a total purchase amount of **\$7,265,862.23**, making up the **largest share of sales**.
- This indicates that the company has a **strong presence** or more established customer base in the UK, which could be the company's primary market.

### 2 Other Significant Markets:

- **Germany** (\$205,569.89), **France** (\$183,891.68), and **Netherlands** (\$283,889.34) show relatively high total purchase amounts, signaling important markets outside of the UK.
- These markets may be good **focus areas** for targeted marketing campaigns or sales initiatives to further increase engagement and revenue.

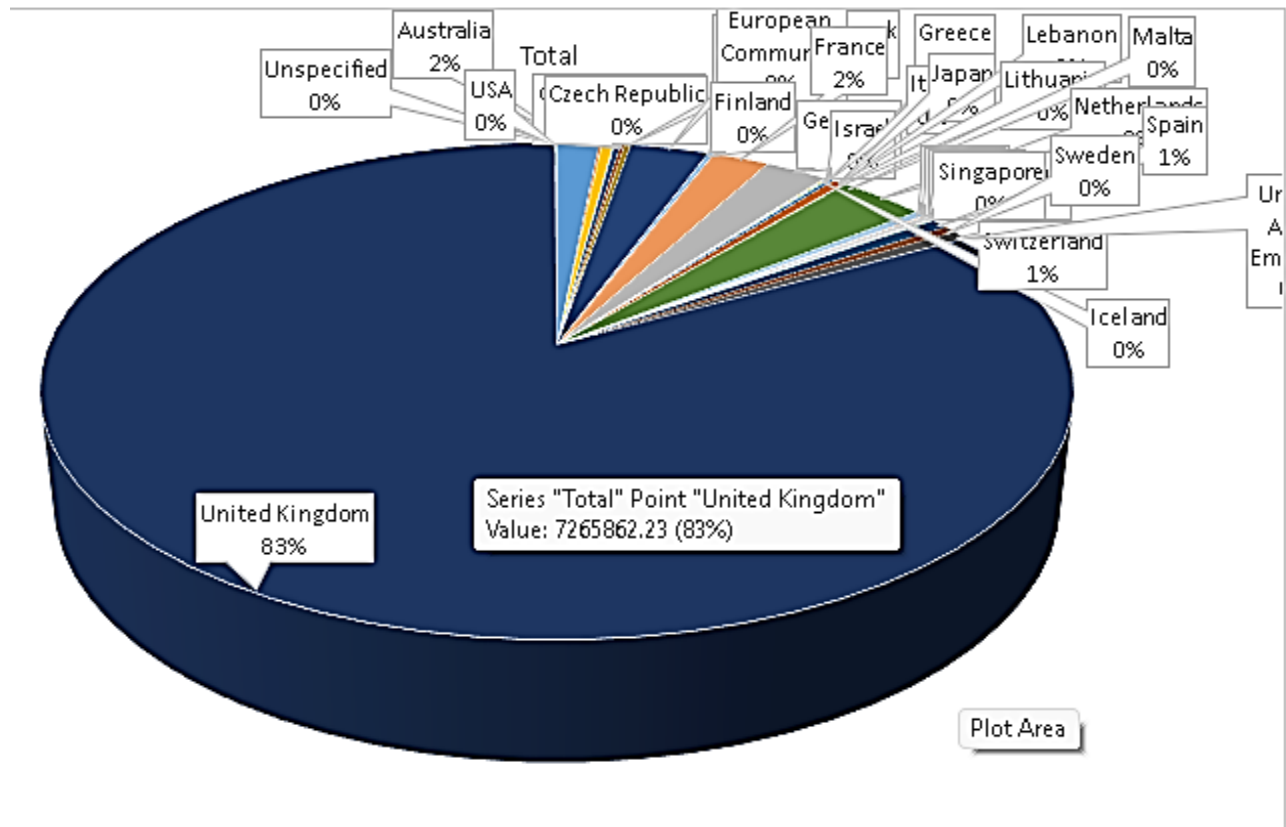
### 3 Smaller Markets:

- **Bahrain** (\$548.40), **Brazil** (\$1,143.60), and **Saudi Arabia** (\$145.92) are among the **smaller markets** contributing far less to the total purchase amount.
- While these countries contribute less, they may present **growth opportunities**. Tailored marketing strategies or localized campaigns could increase the customer base and spending in these regions.

### 4 Unspecified Country:



- The "**Unspecified**" category (\$2,667.07) suggests that some transactions are missing country information. This is an **area for improvement** in terms of data collection or system integration to better capture customer location details.



## Recommendations:

### 1. Focus on UK:

- Given the **dominant contribution** of the UK to total sales, businesses should **retain and expand** marketing efforts in this country, focusing on increasing customer lifetime value.

### 2. Growth in Smaller Markets:

- Countries with **low sales** (e.g., **Bahrain, Brazil, Saudi Arabia**) should be targeted with specific promotions or local strategies to drive higher purchase volumes.

### 3. Improve Data Collection:

- Efforts should be made to improve data accuracy, especially in terms of collecting **country information** for transactions. This could help businesses understand their global customer base more clearly.

# RFM Analysis

## RFM Analysis - Top 100 Frequency

### 1 High-Frequency Customers:

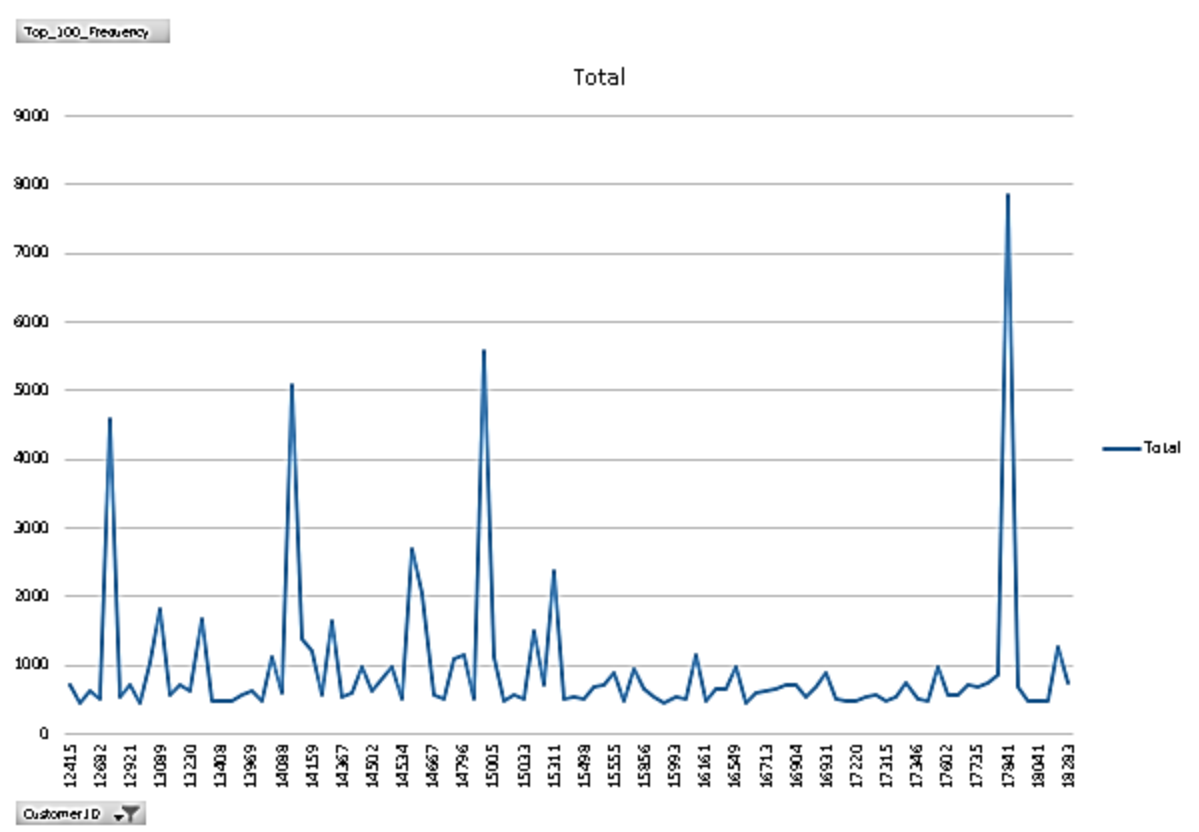
- The top customer with the highest frequency of purchases is **Customer ID 17841** with **7,838 purchases**.
- **Customer ID 14911** has the second-highest frequency with **5,591 purchases**, showing that some customers are highly engaged.

### 2 Moderate Frequency:

- Many customers have purchase frequencies in the range of **450 to 1,000**, such as **Customer IDs 12415, 13089, and 15311**.
- These customers might represent **regular but not overly frequent buyers**.

### 3 Low Frequency:

- There are a number of customers with lower frequencies, generally below **500** purchases, such as **Customer ID 17365** with **467 purchases**.



### Observations:

- **Frequent Buyers:** A small group of customers are responsible for a significant portion of the purchases. Targeting these customers for loyalty programs or promotions can increase their lifetime value.
- **Opportunities:** Customers with low purchase frequency might benefit from **re-engagement campaigns** to increase their buying frequency.

## RFM Analysis - Spending Segmentation

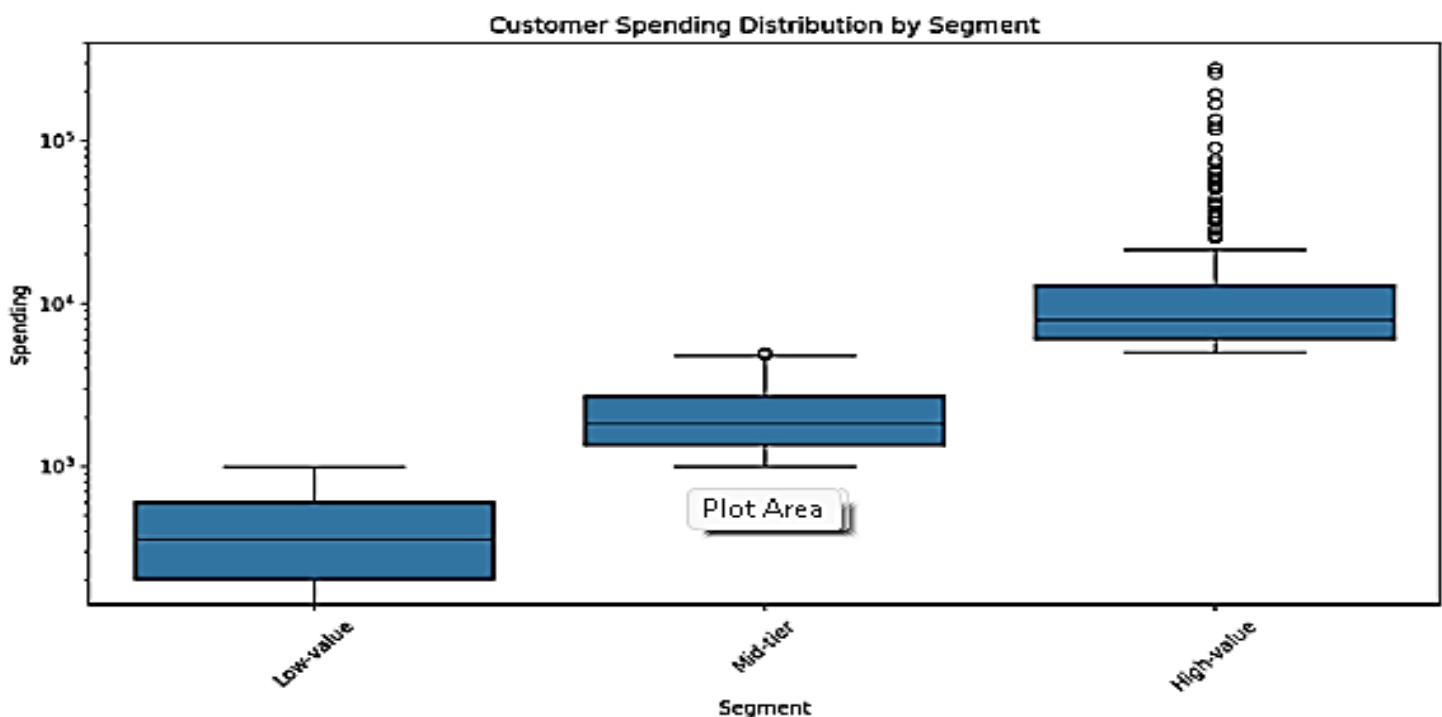
### 1 Low-Value Segment:

- **2,690 customers** in this group with a **mean spending of \$414.38**.
- Total spending: **\$1,114,671.51**.
- This group has customers with the **lowest spending**, but they make up a large portion of the customer base.

### 2 Mid-Tier Segment:

- **1,376 customers** with a **mean spending of \$2,148.53**.
- Total spending: **\$2,956,383.78**.
- These customers are **moderately engaged** with a higher average spend compared to the low-value group.

### 3 High-Value Segment:



- **269 customers** with a **mean spending of \$17,434.99**.
- Total spending: **\$4,690,011.36**.
- This segment has **highly profitable customers**, contributing significantly to overall revenue

## RFM Analysis - Recency

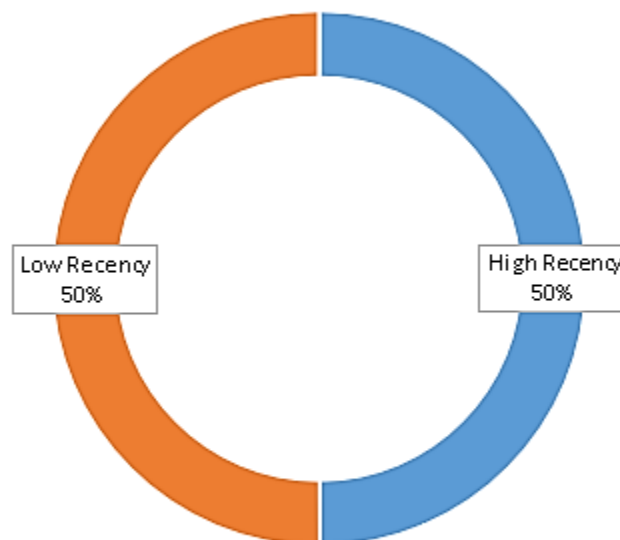
The **Recency** metric categorizes customers based on how recently they made a purchase. The dataset splits them into two groups: **High Recency** and **Low Recency**.

### 1 High Recency (197,985 customers):

- This group represents customers who have made **recent purchases**.
- These customers are likely **engaged** with the brand and may respond well to **loyalty programs** or **upsell offers**.

### 2 Low Recency (198,385 customers):

- Customers in this group have not made purchases recently.
- They might be at risk of **churning** or losing interest, requiring **re-engagement strategies** (e.g., targeted email campaigns or special offers).



# Customer Behavior Analysis

## Total Revenue per Country

### 1 United Kingdom Dominance (82.93%):

The United Kingdom contributes the largest share of total revenue, **amounting to 82.93%** of the total sales.

This indicates that the business has a strong customer base in the UK, which could be attributed to a larger market presence or brand loyalty.

**2 Other High-Contributing Countries:**

Netherlands (3.24%) and EIRE (2.94%) follow the UK, with substantial contributions to total revenue.

Countries like **France (2.10%), Germany (2.35%), and Australia (1.58%)** also make notable contributions, showing that these regions have significant market potential.

**3 Smaller Market Contributions:**

Several countries like Austria (0.10%), Canada (0.04%), and USA (0.04%) contribute only a small fraction to the total sales.

These markets may offer growth potential if targeted with focused marketing or product offerings.

**4 Low-Value Countries:**

Countries like **Bahrain (0.01%), Brazil (0.01%), and Saudi Arabia (0.00%)** contribute minimal revenue.

These regions might require a different business strategy to increase engagement and sales.



**Recommendations:**

1. **Focus on High Revenue Countries:** Prioritize market growth strategies in the **United Kingdom, Netherlands, EIRE, France, and Germany** due to their high contributions.
2. **Growth Opportunities in Low-Contributing Regions:** Explore **targeted campaigns** in low-revenue countries such as **USA, Saudi Arabia, and Bahrain** to expand their revenue share.
3. **Local Adaptation:** Tailor products, services, or marketing campaigns according to the specific demands of countries that contribute less revenue.

## Product Insights:

### 🔍 Top Products by Quantity Sold:

- The **most sold products** are the following:
  - **WHITE METAL LANTERN:** Sold 6 units.
  - **KNITTED UNION FLAG HOT WATER BOTTLE:** Sold 6 units.
  - **RED WOOLLY HOTTIE WHITE HEART:** Sold 6 units.
  - **CREAM CUPID HEARTS COAT HANGER:** Sold 8 units.
  - **WHITE HANGING HEART T-LIGHT HOLDER:** Sold 6 units.

These products are sold in relatively small quantities, with **6 or 8 units** for each listed product.

### 🔍 Total Quantity Sold:

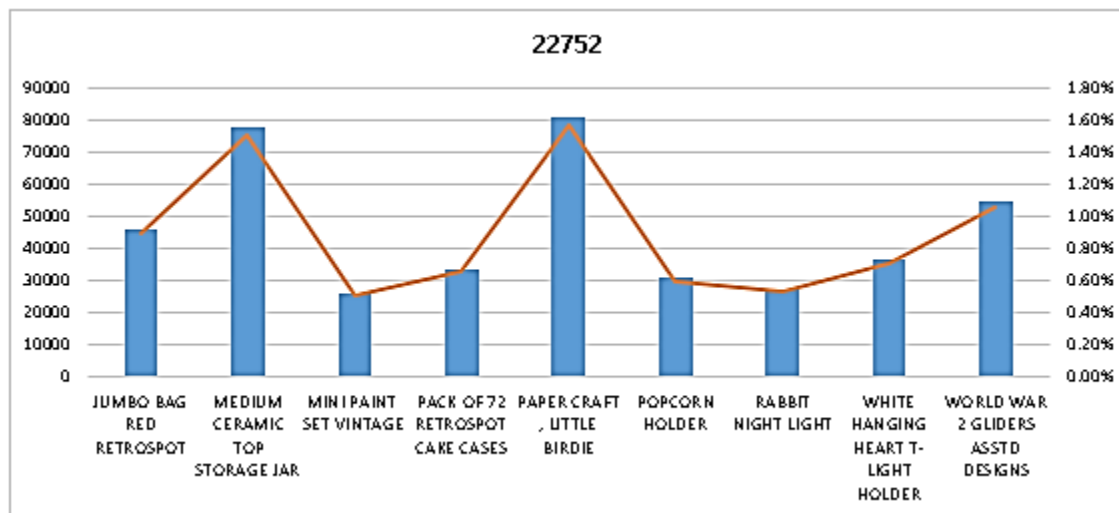
- The **grand total of all quantities** sold across products is **5,171,231** units. This reflects the overall sales volume for all products in the dataset. **Potential Takeaways:**
- **Focus on Bestsellers:** The products listed above may have **niche demand** based on the smaller quantities. It could be insightful to explore customer preferences, seasonality, or product promotions to see if these products can be marketed more aggressively to boost their sales.
- **Explore Product Categories:** While this list shows a small selection of products, there could be other items in the dataset contributing to the bulk of sales. Analyzing **categories** or **product types** could help identify major sellers and gaps in the product lineup.
- **Sales Volume Distribution:** The **grand total of 5,171,231** units sold suggests that the majority of sales are likely spread across a broader array of products, not just the ones listed here. Further analysis can reveal whether a **small group of products** contribute significantly to the total sales

| StockCode   | 22752                              | Quantity sold | Top_10_Rate |
|-------------|------------------------------------|---------------|-------------|
| Description |                                    |               | 0.00%       |
| Description | JUMBO BAG RED RETROSPOT            | 46181         | 0.89%       |
| Description | MEDIUM CERAMIC TOP STORAGE JAR     | 77916         | 1.51%       |
| Description | MINI PAINT SET VINTAGE             | 26076         | 0.50%       |
| Description | PACK OF 72 RETROSPOT CAKE CASES    | 33693         | 0.65%       |
| Description | PAPER CRAFT , LITTLE BIRDIE        | 80995         | 1.57%       |
| Description | POPCORN HOLDER                     | 30931         | 0.60%       |
| Description | RABBIT NIGHT LIGHT                 | 27202         | 0.53%       |
| Description | WHITE HANGING HEART T-LIGHT HOLDER | 36719         | 0.71%       |
| Description | WORLD WAR 2 GLIDERS ASSTD DESIGNS  | 54415         | 1.05%       |

| StockCode   | 71053               | Quantity sold | Rate |
|-------------|---------------------|---------------|------|
| Description | WHITE METAL LANTERN | 6             | 100% |

| StockCode   | 84029                               | Quantity sold | Rate |
|-------------|-------------------------------------|---------------|------|
| Description | KNITTED UNION FLAG HOT WATER BOTTLE | 6             | 50%  |
| Description | RED WOOLLY HOTTIE WHITE HEART.      | 6             | 50%  |

| StockCode   | 85123                              | Quantity sold | Rate |
|-------------|------------------------------------|---------------|------|
| Description | WHITE HANGING HEART T-LIGHT HOLDER | 6             | 100% |



### Recommendation:

- **Further Product Segmentation:** Investigate product categories to understand the relationship between sales volume and product types.

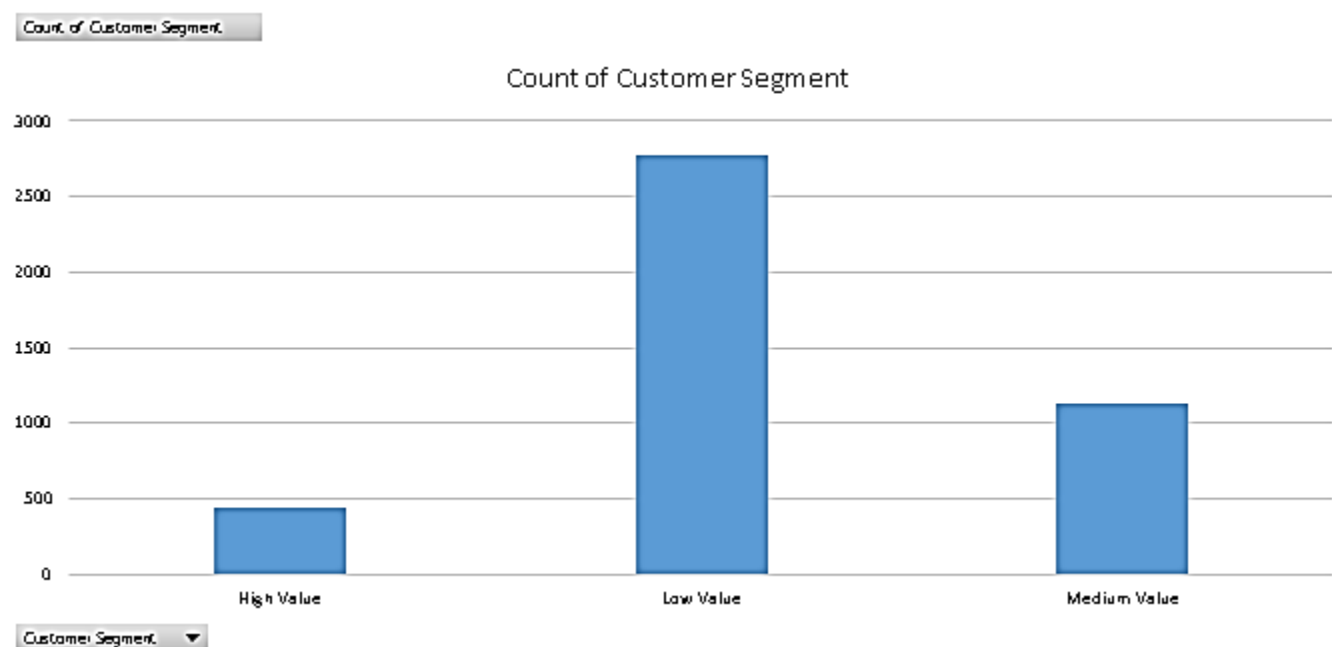
- **Identify High-Demand Products:** Explore whether the top products could be promoted more heavily or bundled to **increase visibility** and sales.

## Customer Purchase Behavior

**1 High-Value Customers:** Only **437 customers** fall into the high-value segment, representing a small but highly valuable portion of the customer base. This segment contributes the most to revenue and should be prioritized for retention strategies like loyalty programs or exclusive offers.

**2 Low-Value Customers:** The largest segment with **2,770 customers**, these customers contribute the least in terms of revenue. Efforts like targeted promotions or upselling can help increase their value.

**3 Medium-Value Customers:** Comprising **1,128 customers**, this segment has potential to be moved into the high-value category with strategic engagement and personalized offers.



### Key Takeaway:

The data highlights the **classic Pareto Principle (80/20 rule)**—a small fraction of customers (high-value) drives a significant portion of revenue. Effective customer relationship management can further maximize business outcomes



