

Introduction:

This analysis focuses on sales and customer data from **ShopEase**, an online retail platform. The goal is to gain useful insights that can help improve **sales performance** and **customer satisfaction**.

By examining customer behavior, product trends, and overall sales patterns, we aim to understand what drives purchases and where there may be opportunities to improve. This includes identifying top-selling products, understanding customer preferences, and finding ways to increase repeat purchases.

The insights from this analysis will support better decision-making in areas like marketing, product selection, and customer experience, helping ShopEase grow and serve its customers more effectively.

Data Cleaning:

- During the initial data inspection, it was observed that all columns, including those with numeric values, were stored as strings. To prepare the data for analysis, columns such as **Quantity** and **UnitPrice** were converted to integers, and **InvoiceDate** was converted to a date format.
- Next, a check for missing values revealed that the **CustomerID** and **Description** columns contained null values. Since both are important for analysis, the missing entries were replaced with the placeholder "**Unknown**" to retain the records without losing essential data.
- Duplicate records were then checked using the **InvoiceDate** column. Although multiple entries shared the same date and time, each represented different products purchased in the same transaction. Therefore, no records were removed at this stage.

Feature Engineering:

To support time-based analysis, new columns for **Month** and **Year** were created from the **InvoiceDate** column. Additionally, a **TotalAmount** column was calculated by multiplying **UnitPrice** and **Quantity**, giving the total value for each transaction line.

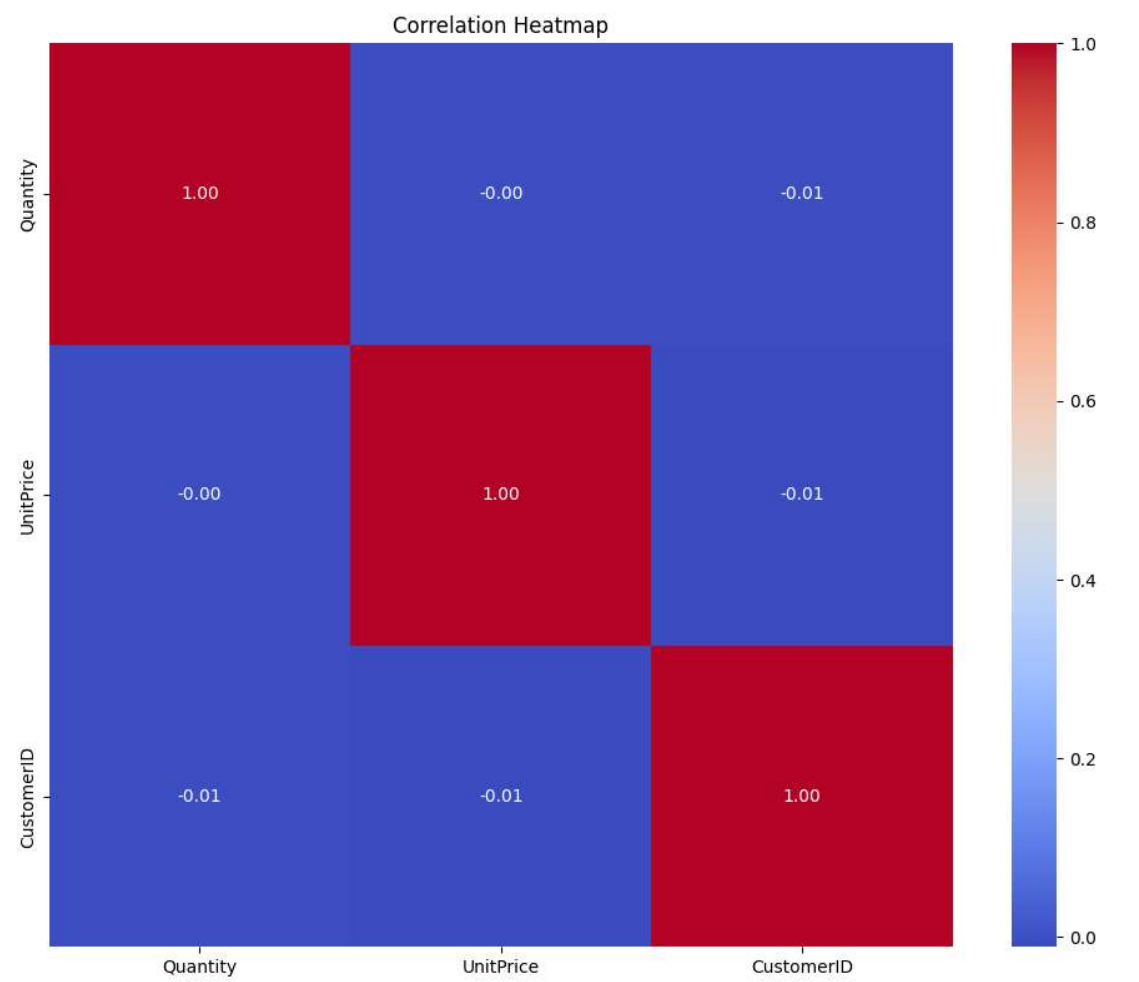
Data Exploration:

While analyzing the summary statistics, it was observed that some records had **negative values** for **Quantity** and **UnitPrice**. These were considered invalid and were removed from the dataset.

Next, a **boxplot** was used to check for outliers in the **Quantity** and **UnitPrice** columns. Outliers were detected in both cases. However, after review, these values appeared to be valid and not due to errors, so they were **retained** in the dataset.

Correlation :

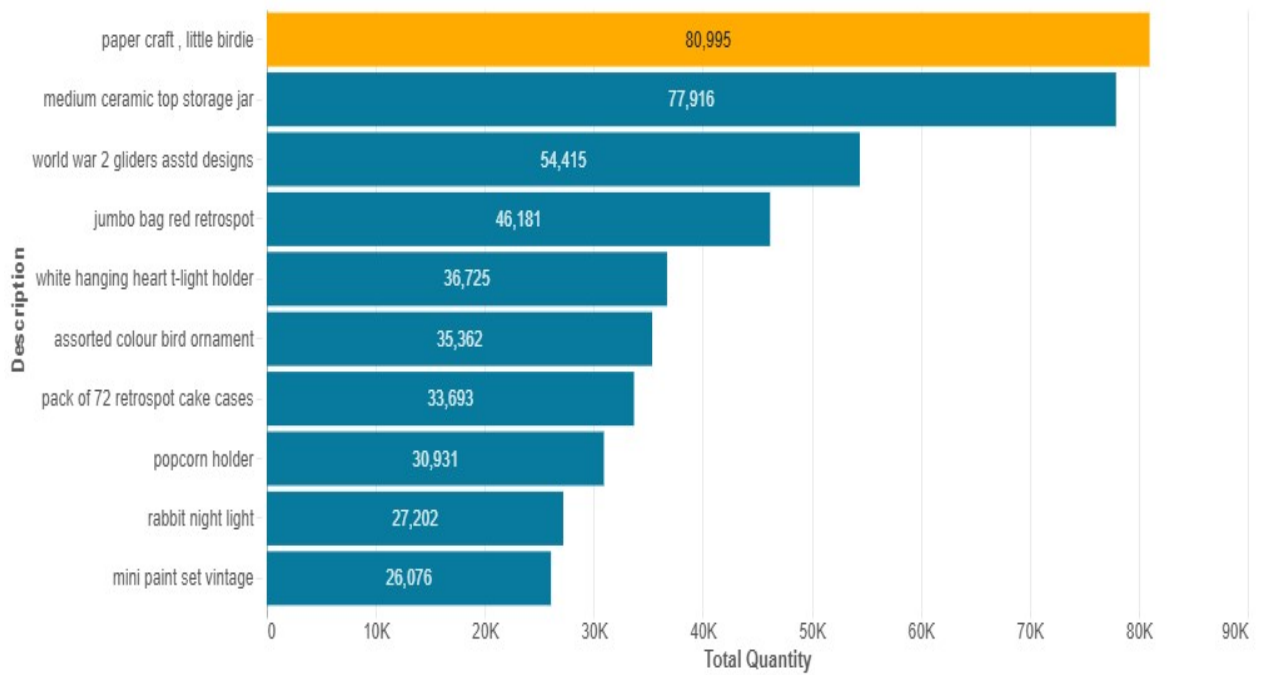
A **correlation analysis** was performed using **Quantity**, **UnitPrice**, and **CustomerID**. The results showed **no strong linear relationship** between these variables, indicating that they are not directly correlated with each other.



Top N Products by Quantity:

An analysis of products based on **total quantity sold** revealed that **'Paper Craft Little Birdie'** is the **top-selling product** in the dataset.

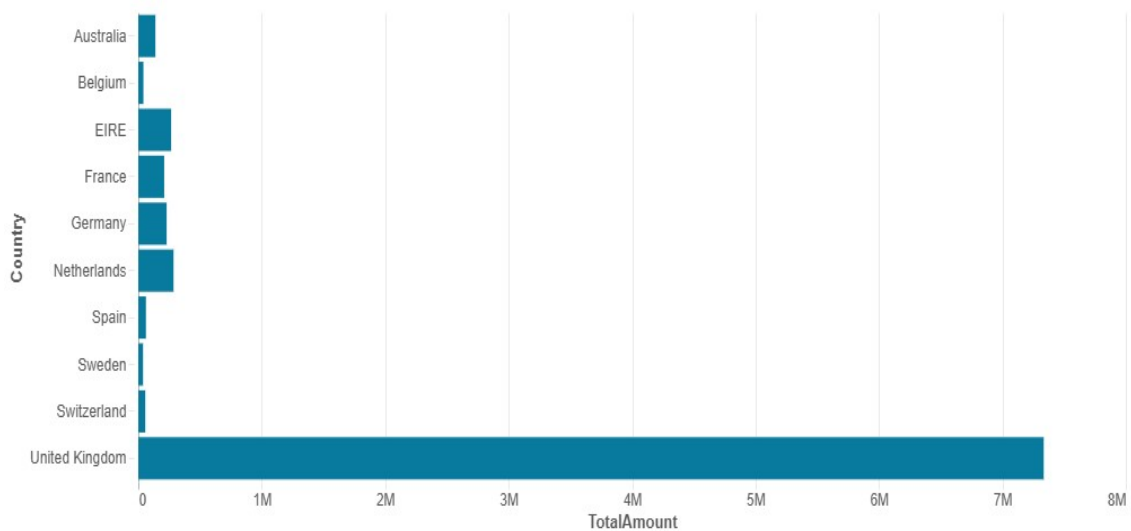
The analysis also identified **'Mini Paint Set Vintage'** as one of the **lowest-selling products** based on total quantity sold.



Top N Products by Quantity

Total Sales by Country:

An analysis of **total sales by country** showed that the **United Kingdom** has the **highest overall sales**, accounting for the majority of transactions in the dataset.



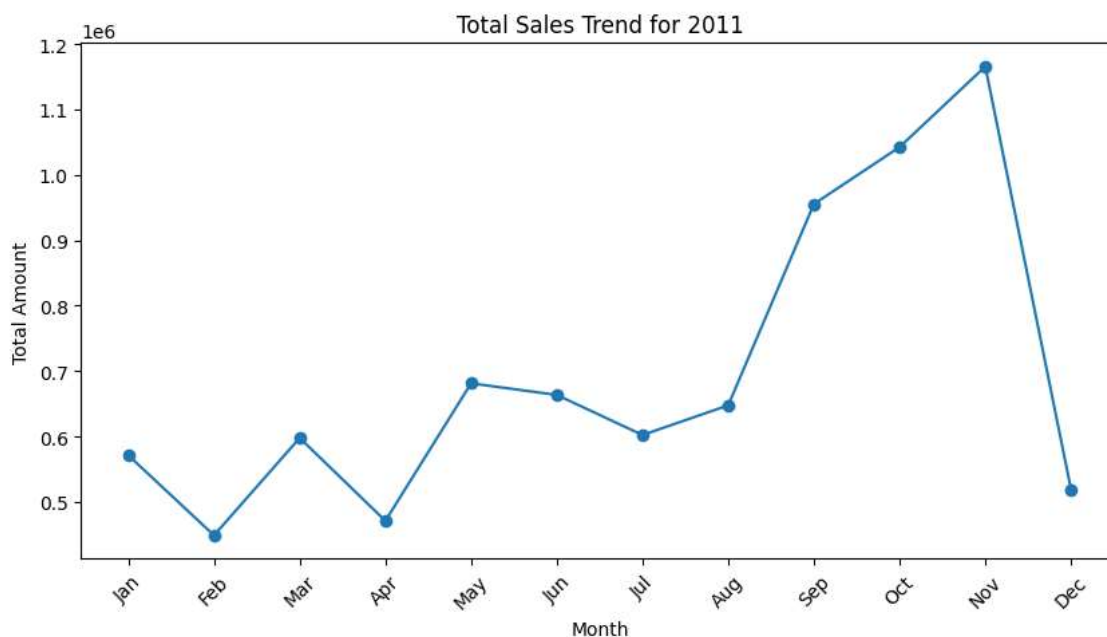
Total Sales by Country

- The UK may be a mature or high-demand market, indicating successful market strategies or stronger customer loyalty.

- The contrast with Saudi Arabia suggests potential for growth through localized marketing, promotions, or partnerships.

Trend Analysis by Month for 2011:

- The data for the years 2010 and 2011 reveals a significant increase in sales, with an overall growth of approximately 13.56%. A deeper look into the monthly breakdown indicates that sales consistently peak in December and January for both years.
- Monthly sales analysis showed that **February** recorded the **lowest sales**.
- Seasonal Pattern Observed December and January are high-performing months.
- This trend suggests a seasonal sales cycle, possibly influenced by:
Holiday shopping season (e.g., Christmas, New Year) Post-holiday promotions New Year resolutions and spending habits.



EDA Summary:

The exploratory data analysis revealed several key insights. Correlation analysis between Quantity, UnitPrice, and CustomerID showed no strong linear relationships. In terms of product performance, '**Paper Craft Little Birdie**' emerged as the top-selling product, while '**Mini Paint Set Vintage**' was among the lowest sold. Country-wise sales analysis showed that the **United Kingdom** had the highest total sales. Monthly trends indicated that **December** had the **highest sales**, likely due to the holiday season, while **February** recorded the **lowest**. Outlier detection using boxplots identified extreme values in Quantity and UnitPrice, but these were retained as they appeared to be valid and not data entry errors.

Data Visualizations:

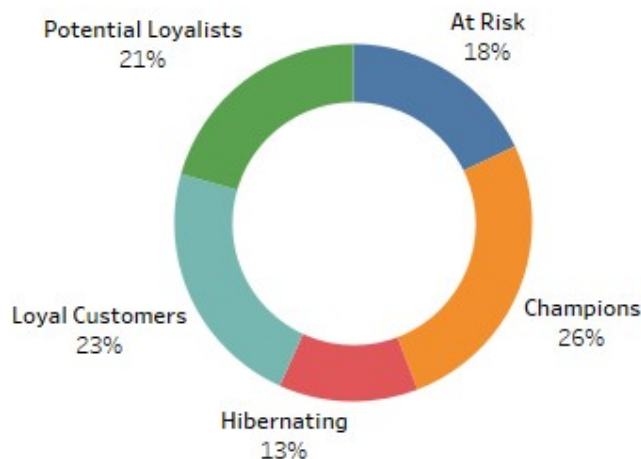
[Sales Customer churn analysis | Tableau Public](#)

RFM Analysis:

Customers were segmented based on their **RFM Aggregate Score (RFM_ag)**, which combines Recency, Frequency, and Monetary values to evaluate customer engagement and value. The segmentation ranges from highly engaged **Champions (RFM_ag ≥ 11)**—our most valuable customers—to **Hibernating (RFM_ag < 5)** customers who show minimal activity and low value. In between, **Loyal Customers (9–10.9)** and **Potential Loyalists (7–8.9)** are seen as key segments for retention and growth through targeted offers and loyalty strategies. **At Risk** customers (5–6.9) require re-engagement efforts, as they show signs of churn. This segmentation helps prioritize marketing actions and customer relationship strategies based on engagement levels.

The analysis revealed a **high number of Champions (RFM_ag ≥ 11)**—our most valuable and loyal customers who purchase frequently and spend the most. At the same time, around **18% of customers fall into the At Risk segment ($5 \leq \text{RFM_ag} < 7$)**. These customers have not purchased recently and show signs of disengagement, making them key targets for reactivation efforts. This segmentation allows for focused strategies to **retain top customers** and **re-engage those at risk of churning**.

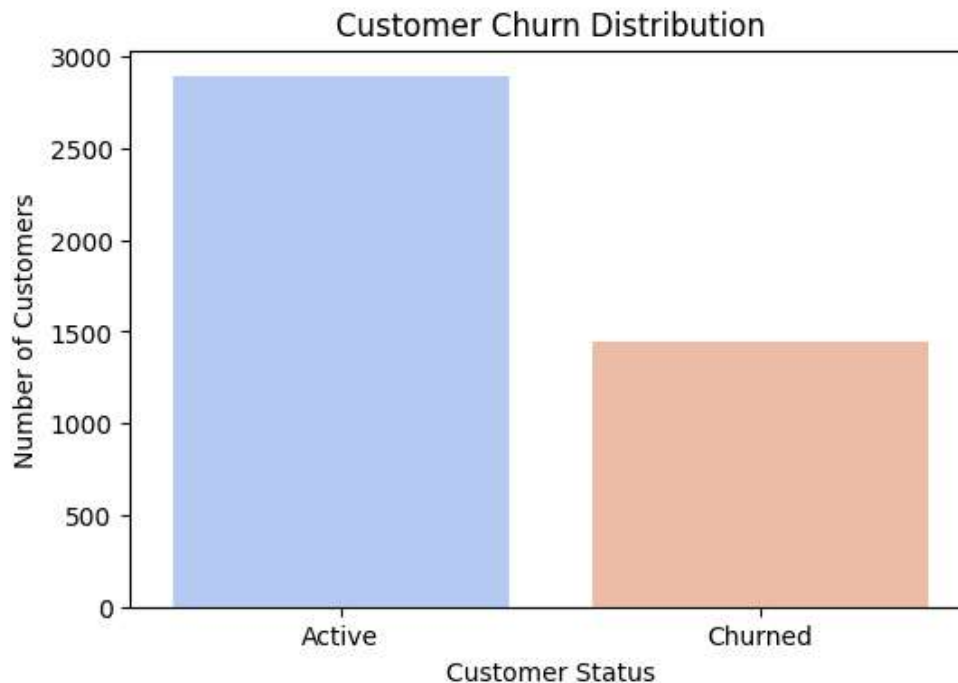
Loyalty Customers



Churn Analysis:

Customer churn analysis is key to understanding why customers stop buying from you and more importantly, how to reduce churn and improve retention.

A bar graph comparing **active** and **churned customers** shows that the number of **active customers** is significantly **higher** than churned customers. This indicates a strong base of engaged customers who continue to make purchases. However, the presence of churned customers also highlights the need to monitor and address drop-off behavior before it grows.



Summary of insights:

The data analysis of ShopEase's online retail data uncovered several valuable insights:

- **Best-Selling Month:** **December** recorded the **highest sales**, likely driven by holiday shopping activity, while **February** saw the **lowest sales**.
- **Top Product:** **"Paper Craft Little Birdie"** emerged as the **best-selling product** based on quantity sold, indicating high customer interest in this item. In contrast, **"Mini Paint Set Vintage"** was among the lowest-selling products.
- **Country with Highest Sales:** The **United Kingdom** accounted for the majority of total sales, indicating a strong customer base and market presence.
- **Customer Behavior & Trends:**
 - A high number of **Champion customers** were identified—these are loyal, high-value customers who purchase frequently and recently.
 - Around **18% of customers** were flagged as **At Risk**, showing signs of churn due to a drop in activity.
 - Monthly sales trends and product performance suggest that **seasonality** plays a role in customer purchasing patterns.
- **Sales Correlation:** No strong **linear correlation** was found between **Quantity**, **UnitPrice**, and **CustomerID**, suggesting customer behavior is influenced by a mix of non-linear or external factors not captured in these features alone.

Recommendations:

Based on the key insights from the analysis, the following recommendations can help improve **sales performance** and **customer retention**:

- **Strengthen Year-End Campaigns:**
Leverage the high sales volume in **December** by launching targeted promotions, bundle offers, and email marketing campaigns ahead of the holiday season to maximize revenue.
- **Boost February Sales with Promotions:**
Since **February** has the lowest sales, introduce limited-time discounts, Valentine’s-themed offers, or loyalty point boosters to drive engagement during this slow period.
- **Retain High-Value Customers (Champions):**
Implement **exclusive perks** such as early access to new products, personalized emails, or VIP loyalty programs to keep Champions engaged and loyal.
- **Re-Engage At-Risk Customers:**
Develop **reactivation campaigns** (e.g., personalized discounts, reminder emails, or surveys) targeted at the **18% of customers at risk** of churning to win back their interest.
- **Promote Top-Selling Products:**
Highlight “**Paper Craft Little Birdie**” and similar high-performing items through homepage features, social media, and cross-selling to increase sales of popular products.
- **Optimize Low-Selling Inventory:**
Consider bundling, discounting, or phasing out low-performing products like “**Mini Paint Set Vintage**” to clear inventory and focus on more profitable items.
- **Expand UK Market Strategies:**
With the **United Kingdom** being the strongest market, explore further growth opportunities there—such as local promotions, faster shipping, or localized content.
- **Implement Customer Segmentation in Marketing:**
Use RFM segments to tailor communication. For example, send loyalty rewards to **Loyal Customers**, and educational content or first-time purchase offers to **Potential Loyalists**.

Effective Win-Back Approaches:

- **Personalized Email Campaigns**
 1. Send customized emails reminding customers of products they liked or bought before.
 2. Include special discounts or offers like “We miss you – here’s 20% off your next purchase.”
- **Exclusive Limited-Time Offers**
 1. Create urgency by offering a time-sensitive promotion (e.g., “72-hour flash sale just for you”).
- **Re-Engagement Surveys**
 1. Ask churned customers why they stopped purchasing and what would bring them back. Use this feedback to adjust offerings or service.
- **Loyalty Point Bonuses**
 1. Offer bonus loyalty points or rewards if they return and make a purchase within a specific period.
- **Product Recommendations**
 1. Recommend new or trending products based on their previous purchases or browsing history.
- **Abandoned Cart Reminders**

1. If applicable, remind them about items left in their cart or wishlist with an added incentive.
- **Multi-Channel Outreach**
 1. In addition to email, use SMS, push notifications, or retargeting ads to reach them where they're most likely to respond.

Ethical Considerations:

The dataset used for this analysis does **not contain any personally identifiable information (PII)** such as names, email addresses, or contact details. As a result, **no data masking or anonymization** was required. All visualizations and insights were derived from aggregated or non-sensitive attributes such as customer ID, product details, and transaction data. This ensures that the analysis is conducted in line with ethical data practices and privacy standards.