Exploratory Data Analysis on Customer Purchase Behaviour Analysis

By: k. Sai Charan

AIM:

To analyze customer churn patterns and uncover factors influencing customer retention and attrition. The goal is to identify key drivers of churn and provide actionable insights to improve customer experience, reduce churn, and optimize business decisions.

Dataset Overview:

- **Rows & Columns: 3,900 Rows,18 Columns**
- **Column Types:**
- **13 Categorical Columns:**
- **' Gender,' Item Purchased', 'Category', 'Location', 'Size', 'Color', 'Season',' Subscription Status',' Shipping Type',' Discount Applied',' Promo Code Used',' Payment Method',' Frequency of Purchases'**
- **5 Numerical Columns:**
- **'Customer ID', 'Age', 'Purchase Amount (USD)', 'Review Rating', 'Previous Purchases'**
- **Missing Data: 37 values in the Review Rating column**

Key Features:

Customer demographics (Age, Gender, Location, Subscription Status)
Purchase details (Item Purchased, Category, Purchase Amount, Season, Size, Color)
Shopping behaviour (Discount Applied, Promo Code Used, Previous Purchases, Frequency of Purchases, Review Rating, Shipping Type)

Data Types and Missing Values:

Review Rating → has missing values (because some rows had Nan values)

The dataset contains 3900 records and 18 columns.

```
#    Column                  Non-Null Count   Dtype
---  ------                  --------------   -----
0    Customer ID             3900 non-null    int64
1    Age                     3900 non-null    int64
2    Gender                  3900 non-null    object
3    Item Purchased          3900 non-null    object
4    Category                3900 non-null    object
5    Purchase Amount (USD)   3900 non-null    int64
6    Location                3900 non-null    object
7    Size                    3900 non-null    object
8    Color                   3900 non-null    object
9    Season                  3900 non-null    object
10   Review Rating           3863 non-null    float64
11   Subscription Status     3900 non-null    object
12   Shipping Type           3900 non-null    object
13   Discount Applied        3900 non-null    object
14   Promo Code Used         3900 non-null    object
15   Previous Purchases      3900 non-null    int64
16   Payment Method          3900 non-null    object
17   Frequency of Purchases  3900 non-null    object
dtypes: float64(1), int64(4), object(13)
memory usage: 548.6+ KB
```

```
Customer ID               0
Age                       0
Gender                    0
Item Purchased            0
Category                  0
Purchase Amount (USD)     0
Location                  0
Size                      0
Color                     0
Season                    0
Review Rating             37  ←
Subscription Status       0
Shipping Type             0
Discount Applied          0
Promo Code Used           0
Previous Purchases        0
Payment Method            0
Frequency of Purchases    0
```

Data Types and Missing Values:

- **Identifiers → Customer ID → string / categorical**

- **Categorical Columns:**
' Gender',' Item Purchased', 'Category', 'Location', 'Size', 'Color', 'Season',' Subscription Status',' Shipping Type',' Discount Applied',' Promo Code Used',' Payment Method',' Frequency of Purchases

- **Numerical Columns:**
'Customer ID', 'Age', 'Purchase Amount (USD)', 'Review Rating', 'Previous Purchases'

KEY FIELD DISTRIBUTIONS
• **Subscription Status → Majority "Yes" subscribers (around 70–75%), smaller group "No" → unsubscribed customers show lower purchase frequency.**
• **Age → Most customers fall between 20–40 years, with fewer older shoppers → slightly right-skewed.**
• **Gender → Fairly balanced distribution between Male and Female customers.**
• **Purchase Amount (USD) → Ranges widely from about $20 to $120 → higher purchase amounts often linked to loyal or frequent shoppers.**
• **Category → Dominated by *Clothing*, followed by *Footwear* and *Accessories* → Spending is highest in Clothing.**
• **Payment Method → Popular methods include *Credit Card*, *PayPal*, and *Venmo* → "Cash" users are fewer and typically spend less.**
• **Shipping Type → "Free Shipping" is most common, while "Next Day Air" and "Express" are used by higher-spending customers.**
• **Discount Applied → Majority of transactions have discounts → Discount users show higher frequency of purchases.**
• **Promo Code Used → Used in a significant portion of purchases, often linked with higher loyalty.**
• **Frequency of Purchases → Most customers purchase *Weekly* or *Fortnightly*, with fewer *Annual* buyers.**
• **Review Rating → Mostly between 3 and 4 stars → moderate satisfaction overall.**
• **Previous Purchases → Right-skewed, many customers with few past purchases, fewer long-term repeat buyers.**
• **Season → Spring and Winter dominate, possibly indicating seasonal trends in clothing and accessories.**

Data Quality Issues:

Missing Values
• *Review Rating* contains 37 missing values → should be imputed using the median for numerical stability.
• **Outliers**
• No significant outliers detected in numeric fields (*Age*, *Purchase Amount*, *Previous Purchases*).
• **Data Type Inconsistencies**
• No incorrect data types found, though all categorical fields (e.g., *Gender*, *Category*, *Payment Method*) are stored as objects and will require encoding before modelling.
• **Categorical Encoding**
• Columns like *Gender*, *Subscription Status*, *Shipping Type*, *Discount Applied*, and *Payment Method* require label or one-hot encoding for model training.

Data Visualization

Univariate Analysis:

Bivariate Analysis:

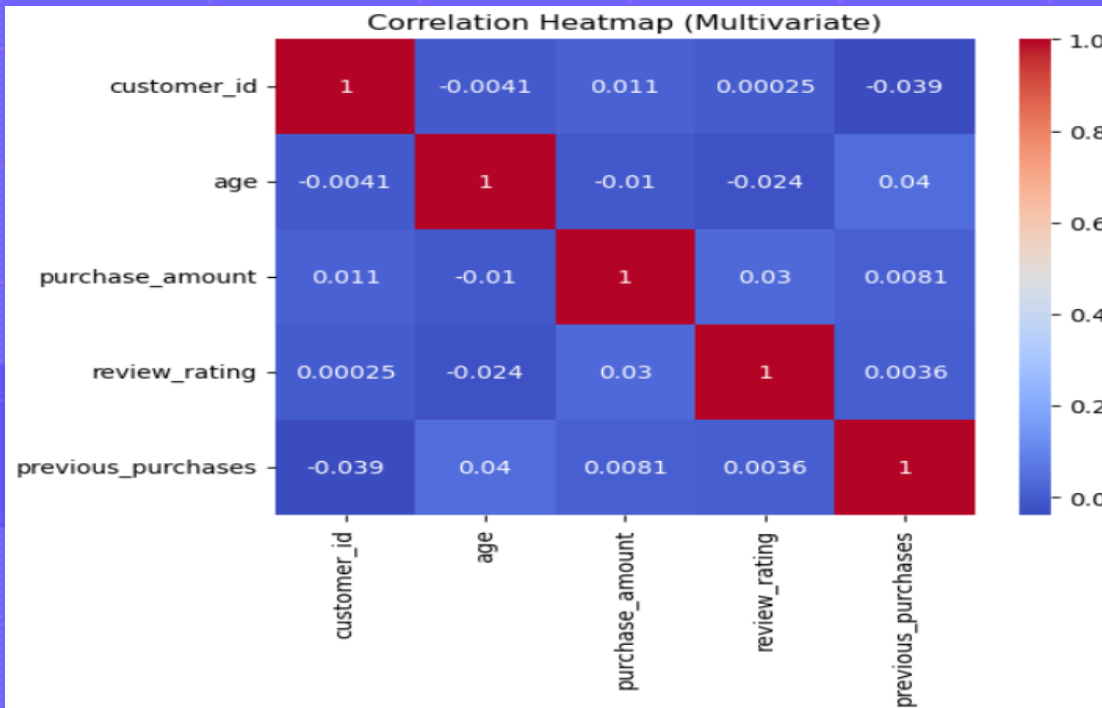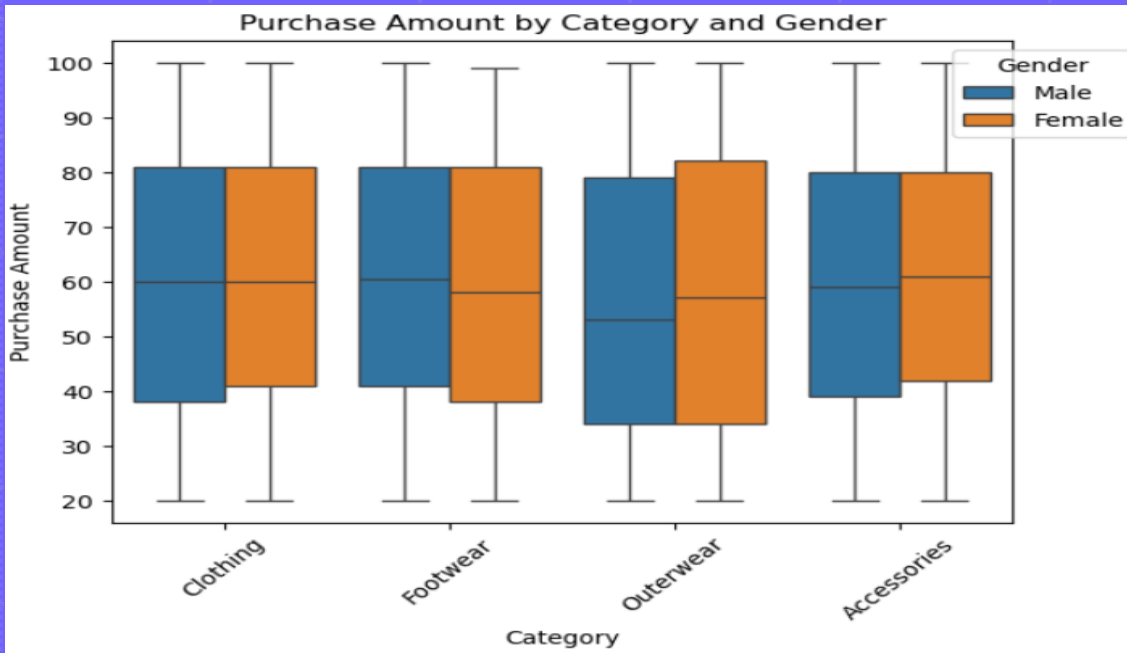

Top Performing Categories by Revenu

Seasonal Sales Trend

Purchase Amount by Category
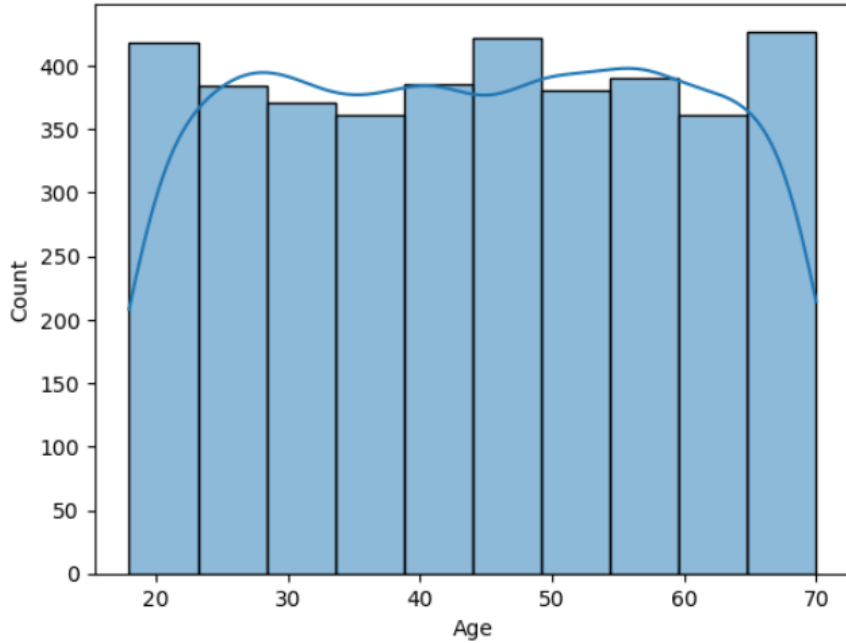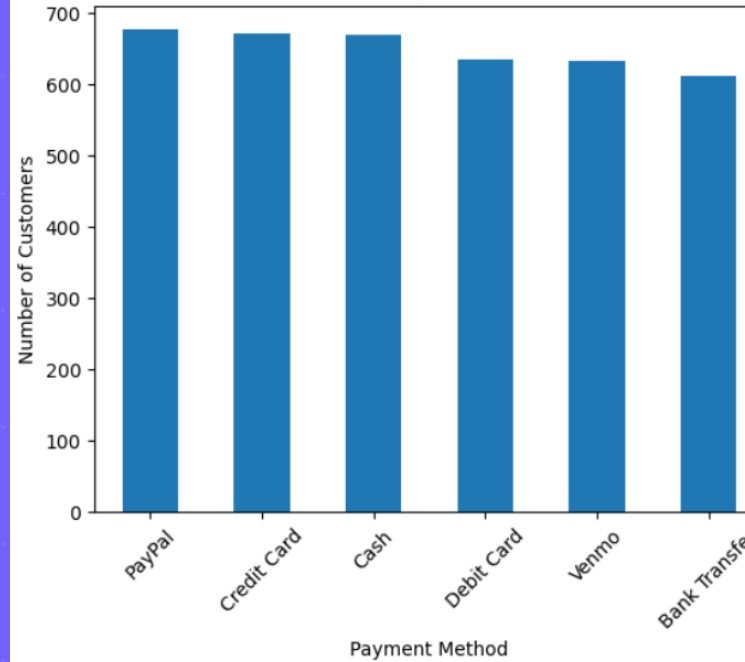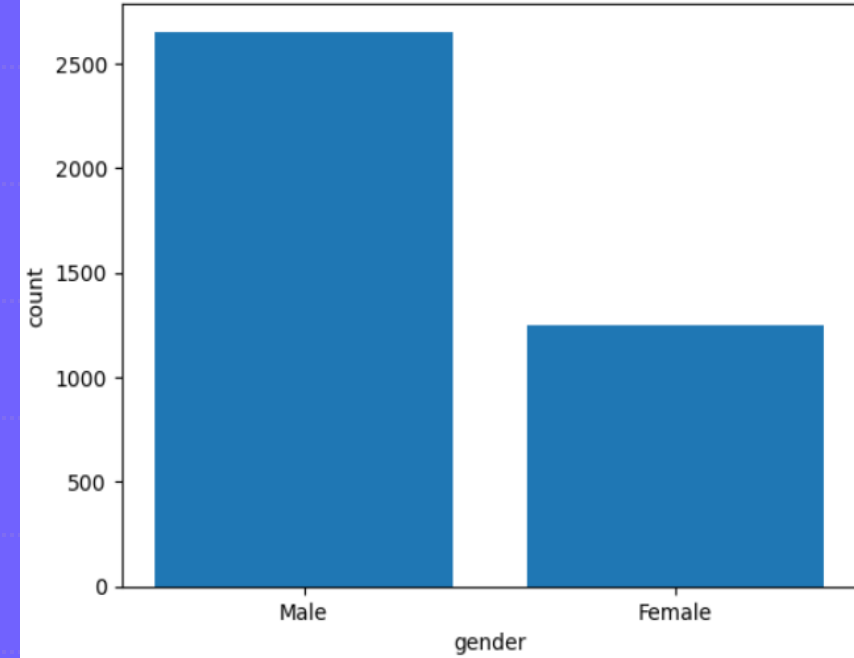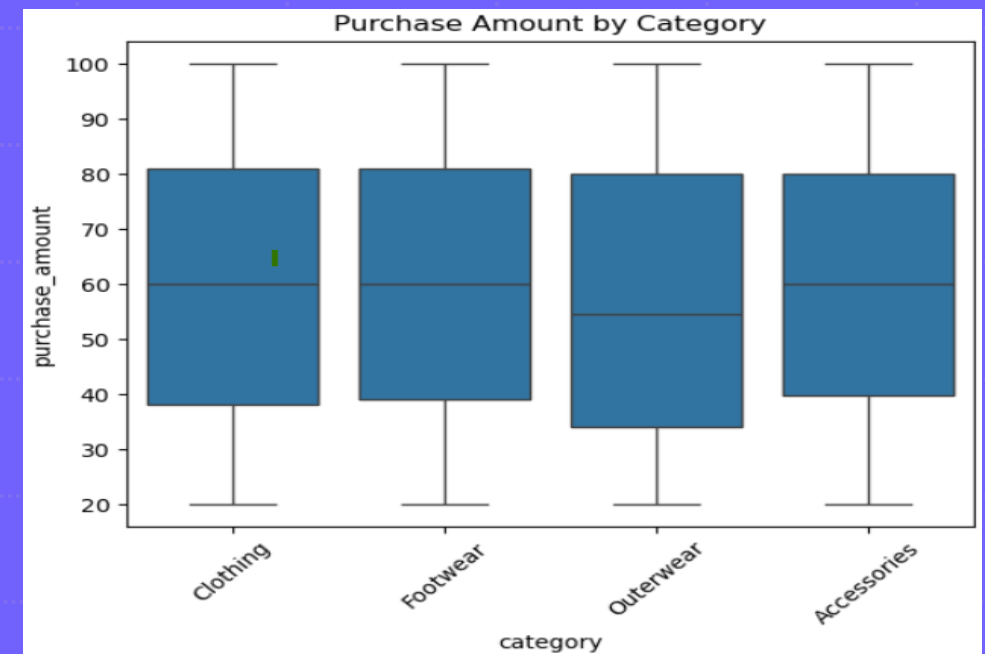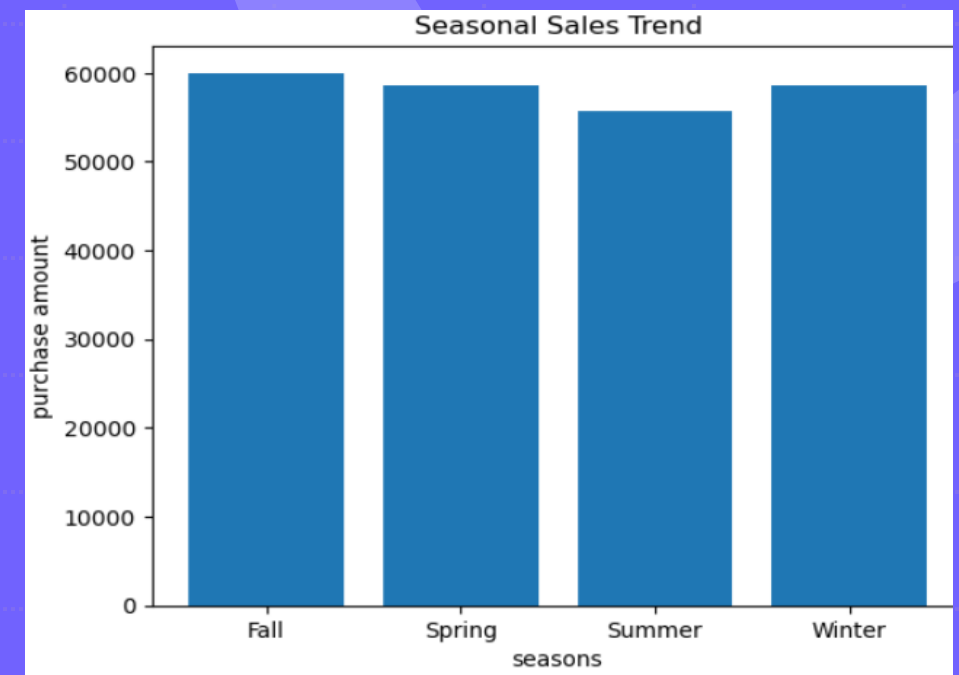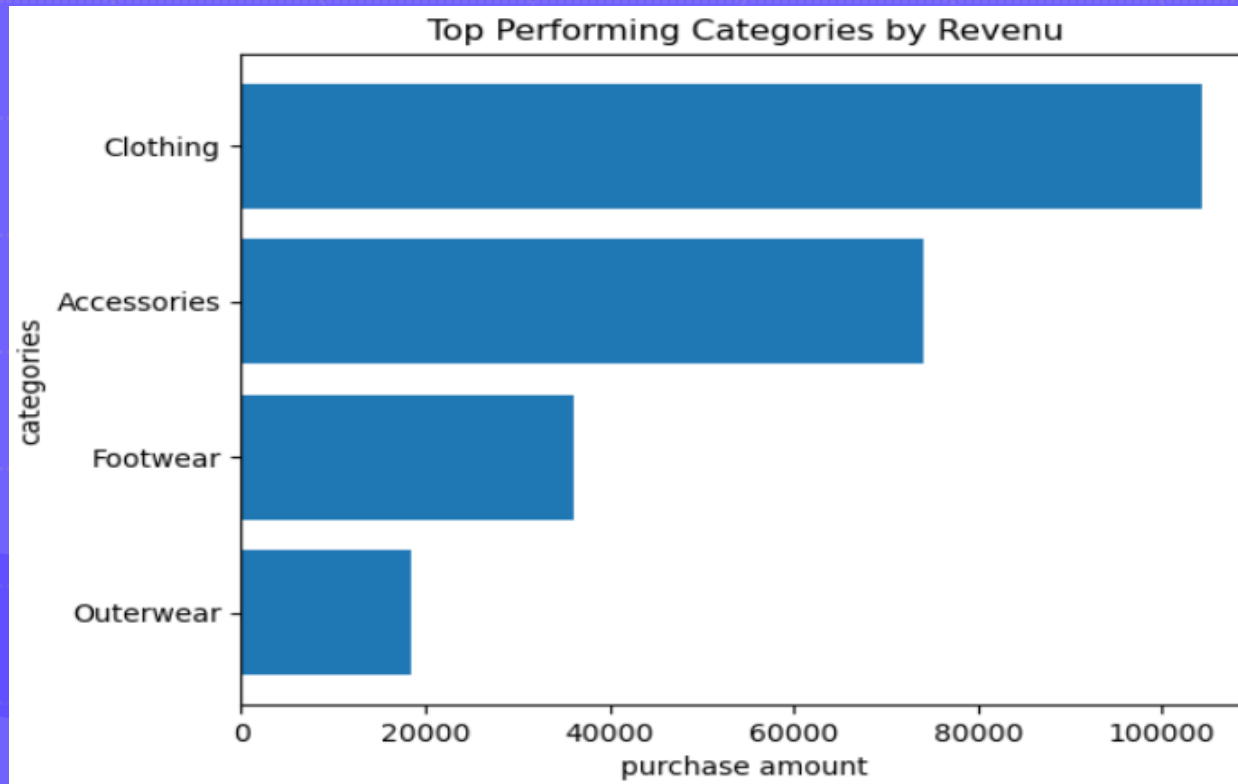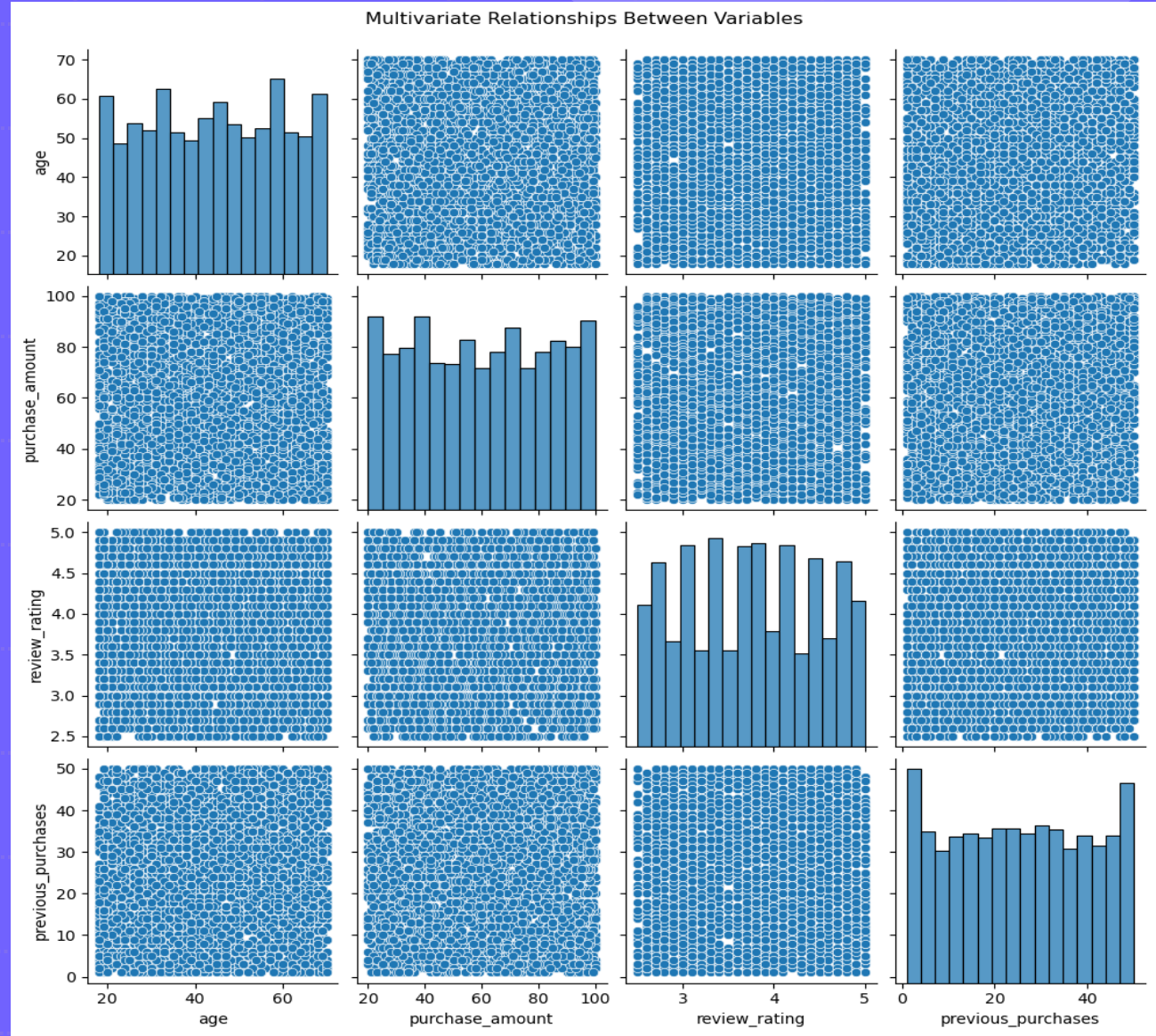
Multivariate Analysis:

Insights:

Most customers are between 20 and 40 years old, forming the core active buyer group.
• Customers with subscriptions tend to purchase more frequently and spend higher amounts.
• Weekly and fortnightly buyers show greater loyalty compared to annual shoppers.
• Discounts and promo codes strongly influence purchase frequency, encouraging repeat transactions.
• Clothing dominates sales, followed by Footwear and Accessories, showing fashion-related buying trends.
• Customers using digital payments (Credit Card, PayPal, Venmo) spend more and buy more often than cash users.
• Free Shipping is preferred by most, though customers opting for Express or Next Day Air tend to make larger purchases.
• Review ratings are mostly moderate (3–4), indicating average satisfaction and potential scope for service improvement.
• Customers with more previous purchases exhibit consistent buying behaviour, reflecting brand loyalty.
• Seasonal patterns show increased activity in Spring and Winter, suggesting seasonal demand peaks.

Recommendations:

- **Enhance Customer Loyalty Early**
  Provide personalised welcome offers and loyalty points for new customers, encouraging repeat purchases during their first few interactions.
- **Strengthen Subscription Benefits**
  Offer exclusive discounts, early access to new collections, or free shipping for subscribers to maintain engagement and boost retention.
- **Leverage Discounts Strategically**
  While discounts drive purchases, monitor their frequency to maintain profitability. Introduce targeted promotions for high-value or inactive customers.
- **Optimize Payment Experience**
  Encourage the use of secure digital payments (Credit Card, PayPal, Venmo) by offering cashback or reward points, reducing dependency on cash transactions.
- **Improve Customer Satisfaction**
  Use product reviews and feedback data to identify improvement areas. Focus on products with lower ratings to enhance customer trust and repeat buying.
- **Seasonal Marketing Campaigns**
  Increase marketing efforts during Spring and Winter when purchase activity peaks, aligning promotions with seasonal trends and top-selling categories.

Thank You!