

# Data Pipeline Report

File: online\_retail\_customer\_churn.csv | Timestamp: 2025-08-23T11:51:25.030395

## 1. Data Validation Summary

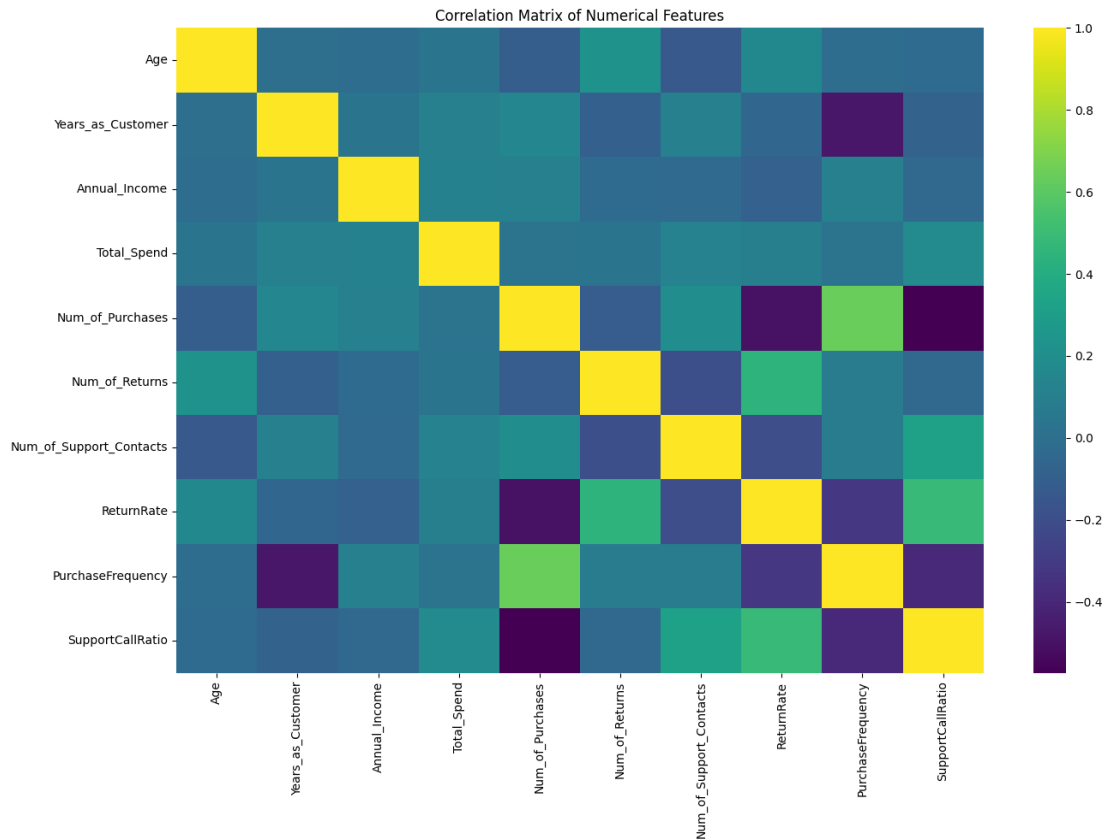
|                      |     |
|----------------------|-----|
| Total Rows           | 100 |
| Total Columns        | 9   |
| Missing Values Found | 0   |
| Duplicate Rows Found | 0   |

## 2. Feature Engineering Summary

Columns Dropped: Customer\_ID  
Columns Created: ReturnRate, PurchaseFrequency, SupportCallRatio

## 3. Exploratory Data Analysis

*Correlation Heatmap*



## 4. Final Processed Data

Shape of Final DataFrame: (100, 11)

**Processed DataFrame Head:**

| Age   | Years_as_Customer | Annual_Income | Total_Spend | Num_of_Purchases | Num_of>Returns | Num_of_Support_Contacts |
|-------|-------------------|---------------|-------------|------------------|----------------|-------------------------|
| 1.33  | 0.59              | 1.1           | 0.72        | -0.91            | 0.33           | -0.07                   |
| 1.6   | -0.83             | -0.04         | 0.35        | -1.43            | 1.73           | -0.78                   |
| -0.85 | -0.12             | -0.61         | 0.13        | -0.85            | -0.02          | 1.34                    |
| 0.94  | -1.18             | 0.75          | 0.1         | 0.76             | 0.33           | -0.78                   |
| -0.85 | -0.83             | 0.77          | -0.81       | 1.34             | 0.68           | 1.34                    |