

# Data Pipeline Report

File: churn.csv | Timestamp: 2025-08-23T22:51:12.582832

## 1. Data Validation Summary

|                      |     |
|----------------------|-----|
| Total Rows           | 100 |
| Total Columns        | 14  |
| Missing Values Found | 0   |
| Duplicate Rows Found | 0   |

## 2. Feature Engineering Summary

**Columns Dropped:** RowNumber, CustomerId, Surname

**Columns Created:** BalanceSalaryRatio, TenureByAge, CreditScoreGivenAge

## 3. Exploratory Data Analysis

*Correlation Heatmap*



## 4. Final Processed Data

Shape of Final DataFrame: (100, 15)

**Processed DataFrame Head:**

| CreditScore | Age   | Tenure | Balance | NumOfProducts | HasCrCard | IsActiveMember |
|-------------|-------|--------|---------|---------------|-----------|----------------|
| -0.49       | -0.43 | 1.36   | 0.06    | -0.15         | 0.87      | 1.02           |
| 1.29        | -0.91 | 0.25   | 0.55    | -1.37         | 0.87      | -0.98          |
| 0.54        | -1.29 | -0.86  | 0.61    | -1.37         | -1.15     | -0.98          |
| -0.22       | -0.24 | -1.6   | -1.19   | -1.37         | 0.87      | 1.02           |
| -0.33       | -0.15 | 1.36   | 1.2     | 1.08          | 0.87      | 1.02           |