Feature Engineering Report

Missing values imputed (median for numerics, mode for categoricals).

Outliers clipped to 1st and 99th percentiles for numerics.

Categorical features encoded (Label/OneHot based on cardinality).

Numerical features scaled (StandardScaler).

New feature created: Spend\_per\_Year.

# Transformed Data Sample

Customer\_ID Age Gender Monthly\_Spend Tenure\_Years ... Customer\_Churn Education\_Level\_High School Education\_Level\_Master's Education\_Level\_PhD Spend\_per\_Year  
0 CUST00001 1.558431 1 -1.178436 0.858830 ... 0 0.0 0.0 1.0 -0.633966  
1 CUST00002 -1.468314 1 -0.197789 -1.477366 ... 0 0.0 1.0 0.0 0.414333  
2 CUST00003 0.361811 1 0.290450 -1.143624 ... 0 0.0 1.0 0.0 -2.022300  
3 CUST00004 0.784148 0 0.476409 0.858830 ... 0 0.0 0.0 1.0 0.256295  
4 CUST00005 1.065705 1 0.218743 -0.809881 ... 0 0.0 1.0 0.0 1.150560  
  
[5 rows x 13 columns]