# Correlation Heatmap Analysis

This document provides a summary interpretation of the correlation heatmap generated from the dataset.  
  
Key Observations:  
  
1. Relationship with Target Variable (`exited`):  
 - age\_skewed: 0.29 → Older customers are more likely to churn.  
 - isactivemember: -0.16 → Active members are less likely to churn.  
 - geography\_germany: 0.17 → Customers from Germany show a slightly higher likelihood of churn.  
 - balancerange: 0.11 → Slight positive correlation with churn.  
 - gender\_label: -0.10 → Very weak negative correlation; may imply a small gender-based difference in churn.  
 - Other features show very weak or no correlation with churn.  
  
2. Multicollinearity Check:  
 - geography\_france vs geography\_germany: -0.58 → As expected from one-hot encoding.  
 - balancerange vs geography\_germany: 0.38 → Moderate correlation.  
 - No pairs exceed 0.8, so multicollinearity is not a serious issue.  
  
Recommendations:  
- Keep `age\_skewed`, `isactivemember`, and `geography\_\*` features (drop one geography column to avoid dummy variable trap).  
- Consider removing weak predictors such as `creditscorerange`, `gender\_label`, and `hascrcard`.  
- Proceed with modeling using logistic regression or tree-based methods.

* Recommended Columns for ML Modeling

|  |  |
| --- | --- |
| **Column** | **Reason for Inclusion** |
| age\_skewed | Moderate positive correlation with exited (0.29) |
| isactivemember | Negative correlation with exited (−0.16) — informative |
| geography\_germany | Slight positive correlation with exited (0.17) |
| geography\_spain | Use one geography feature along with germany |
| numofproducts | Though weak, still possibly useful in tree-based models |
| balancerange | Weakly correlated, but worth keeping for completeness |
| tenurerange | Not strongly correlated, but adds potential time context |

* Optional (Low Predictive Power, Keep for Testing)

|  |  |
| --- | --- |
| **Column** | **Comment** |
| gender\_label | Very low correlation (−0.10), might not help much |
| creditscorerange | Correlation ≈ 0 — usually expected to help, but in this case, might not |
| hascrcard | Almost zero correlation — likely not useful |
| estimatedsalaryrange | Very low correlation — could still be useful in tree models |

* Avoid or Drop

|  |  |
| --- | --- |
| **Column** | **Comment** |
| geography\_france | Drop to avoid dummy variable trap (keep only 2 of 3 geography columns) |

