Feature Analysis Report

RowNumber

- Represents: Sequential record identifier
- Datatype: Numerical (integer)
- Importance: Non-predictive index
- Range: 1-10,000 (μ=5000.5, σ=2886.9)
- Transformation: Exclude

CustomerId

- Represents: Unique customer identifier
- Datatype: Numerical (integer)
- Importance: Non-predictive identifier
- Range: 15.57M-15.82M (all unique)
- Transformation: Remove

Surname

- Represents: Customer's last name
- Datatype: Categorical
- Importance: Low predictive value
- Levels: 2932 unique surnames
- Transformation: Exclude (high cardinality)

CreditScore

- Represents: Financial credibility metric
- Datatype: Numerical (integer)
- Importance: High predictor potential
- Range: 350-850 (μ=650.5, σ=96.7)
- Transformation: Standard scaling

Geography

- Represents: Customer country
- Datatype: Categorical
- Importance: Significant regional patterns
- Levels: Categorical
- Transformation: One-hot encoding

Gender

- Represents: Biological sex
- Datatype: Categorical
- Importance: Moderate demographic factor
- Levels: Male/Female
- Transformation: Binary encoding

Age

- Represents: Customer age
- Datatype: Numerical (integer)
- Importance: Key demographic predictor
- Range: 18-92 (μ=38.9, σ=10.5)
- Transformation: Binning for non-linear effects

Tenure

- Represents: Account longevity (years)
- Datatype: Numerical (integer)
- Importance: Moderate retention signal
- Range: 0-10 (μ=5.0, σ=2.9)
- Transformation: Ordinal encoding

Balance

- Represents: Account funds
- Datatype: Numerical (float)
- Importance: Critical financial indicator
- Range: $0-$250K (μ=$76,486, σ=$62,397)
- Transformation: Log transform for skew

NumOfProducts

- Represents: Banking products used
- Datatype: Numerical (integer)
- Importance: Strong retention signal
- Range: 1-4 (μ=1.5, σ=0.6)
- Transformation: Ordinal encoding

HasCrCard

- Represents: Credit card ownership
- Datatype: Binary
- Importance: Moderate spending indicator
- Levels: 0=No, 1=Yes
- Transformation: Keep as binary

IsActiveMember

- Represents: Account activity status
- Datatype: Binary
- Importance: Cructive retention indicator
- Levels: 0=Inactive, 1=Active
- Transformation: Binary flag

EstimatedSalary

- Represents: Annual income
- Datatype: Numerical (float)
- Importance: Moderate financial factor
- Range: $11.58-$199,992 (μ=$100,090, σ=$57,510)
- Transformation: Robust scaling

Exited (Target)

- Represents: Churn status
- Datatype: Binary
- Levels: 0=Retained, 1=Churned
- Class Balance: 20.4% churn rate

Complain

- Represents: Customer complaints
- Datatype: Binary
- Importance: Critical churn predictor
- Levels: 0=No, 1=Yes
- Transformation: Binary flag

Satisfaction Score

- Represents: Service rating
- Datatype: Ordinal
- Importance: Key experience metric
- Range: 1-5 (μ=3.0, σ=1.4)
- Transformation: Ordinal encoding

Card Type

- Represents: Credit card tier
- Datatype: Categorical
- Importance: Moderate spending signal
- Levels: DIAMOND/GOLD/SILVER/PLATINUM
- Transformation: One-hot encoding

Point Earned

- Represents: Loyalty points
- Datatype: Numerical (integer)
- Importance: Engagement metric
- Range: 119-1000 (μ=606.5, σ=225.9)
- Transformation: Min-Max scaling

1. Priority Predictors: CreditScore, Geography, Age, Balance, Complain, Satisfaction Score
2. Encoding Strategy: One-hot for Geography/CardType, ordinal for Satisfaction Score
3. Feature Removal: Exclude RowNumber, CustomerId, Surname
4. Handling Skew: Log transform Balance, robust scale EstimatedSalary
5. Class Imbalance: Address 20.4% churn rate with sampling techniques