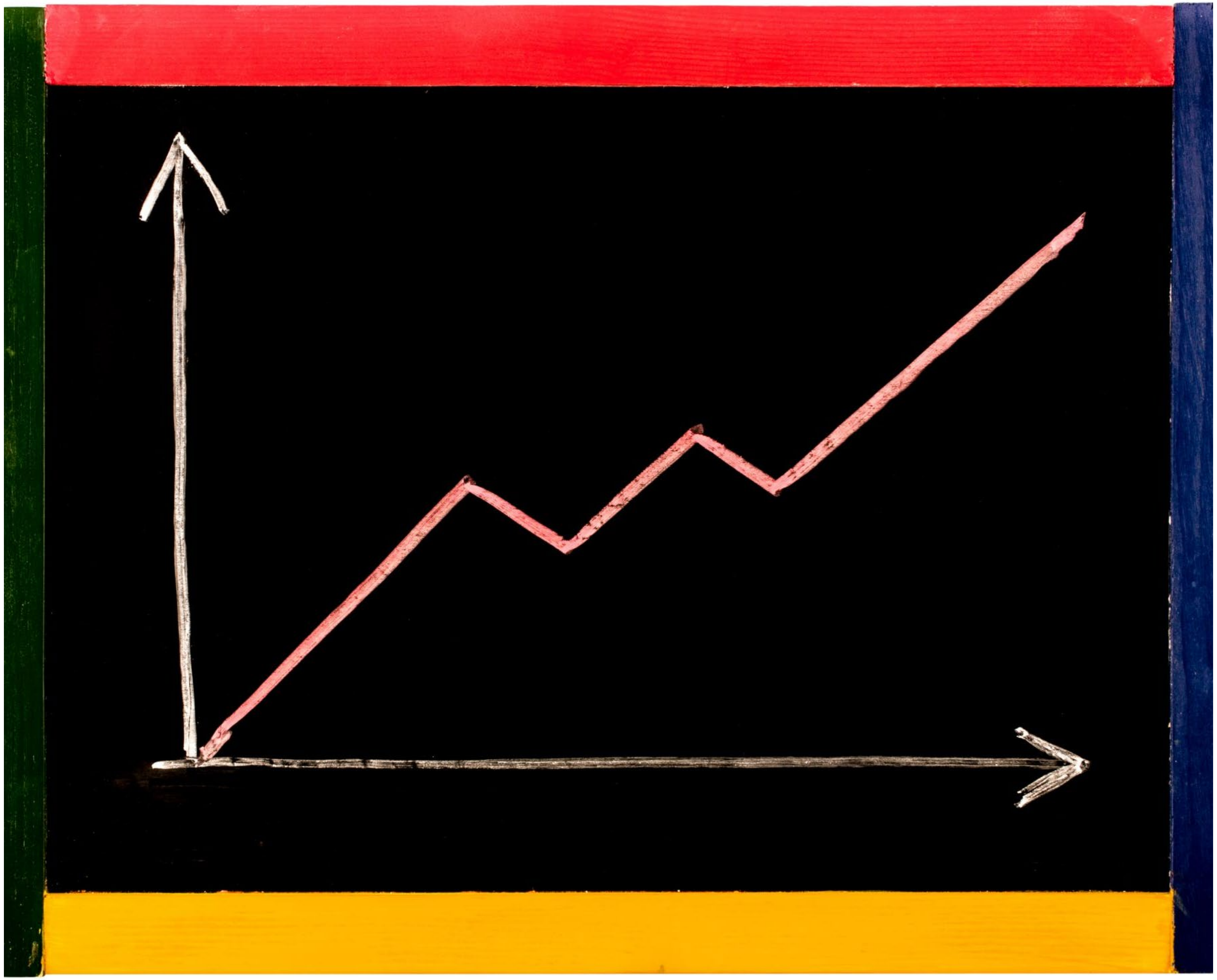
Exploratory Data Analysis of Customer Churn: Uncovering Key Insights for Retention

Analyzing patterns to enhance customer loyalty and retention

Meeting Program

- Project Overview and Dataset Description
- Data Cleaning and Preparation
- Exploring Churn Distribution and Categorical Drivers
- Numerical Feature Analysis to Identify Churn Patterns
- Key Drivers of Churn and Actionable Recommendations

Project Overview and Dataset Description



Defining the Objective and Target Variable

Objective Identification

- The main goal is to identify key factors that lead to customer churn for better retention strategies.

Target Variable:

- Exited (1 = churned, 0 = retained).

Target Variable Explanation

- The target variable indicates whether a customer has churned or remained loyal, directing the analysis approach.

Outlining Steps Taken in the Analysis

Dataset Exploration

Initial step involves examining the dataset to understand its structure and content.

Data Cleaning

Removing errors and inconsistencies from the data to ensure high quality for analysis.

Feature Analysis

Analyzing key features to identify patterns relevant to retention and churn.

Deriving Insights

Extracting meaningful conclusions to inform strategies reducing churn rates.

Initial Exploration of the Dataset

Examine Data Types

Understanding data types helps identify appropriate analysis techniques and potential preprocessing steps.

Summary Statistics

Calculating mean, median, mode, and range provides insight into data distribution and central tendencies.

Distribution Analysis

Visualizing distributions helps detect patterns, outliers, and potential data quality issues early in analysis.

Data Cleaning and Preparation

Dropping Non-Informative Identifiers

Purpose of Dropping Identifiers

- Removing unique identifiers helps eliminate irrelevant data that does not aid churn prediction.
- No Missing values found.
- Dataset Ready with 10,000 rows x 11 features.

Focus on Meaningful Features

- Streamlining the dataset allows models to concentrate on features that improve prediction accuracy.

	RowNumber	CustomerId	CreditScore	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	Exited
count	10000.00000	1.000000e+04	10000.000000	10000.000000	10000.000000	10000.000000	10000.000000	10000.00000	10000.000000	10000.000000	10000.000000
mean	5000.50000	1.569094e+07	650.528800	38.921800	5.012800	76485.889288	1.530200	0.70550	0.515100	100090.239881	0.203700
std	2886.89568	7.193619e+04	96.653299	10.487806	2.892174	62397.405202	0.581654	0.45584	0.499797	57510.492818	0.402769
min	1.00000	1.556570e+07	350.000000	18.000000	0.000000	0.000000	1.000000	0.00000	0.000000	11.580000	0.000000
25%	2500.75000	1.562853e+07	584.000000	32.000000	3.000000	0.000000	1.000000	0.00000	0.000000	51002.110000	0.000000
50%	5000.50000	1.569074e+07	652.000000	37.000000	5.000000	97198.540000	1.000000	1.00000	1.000000	100193.915000	0.000000

Assessing Missing Values and Data Integrity

Identifying Missing Data

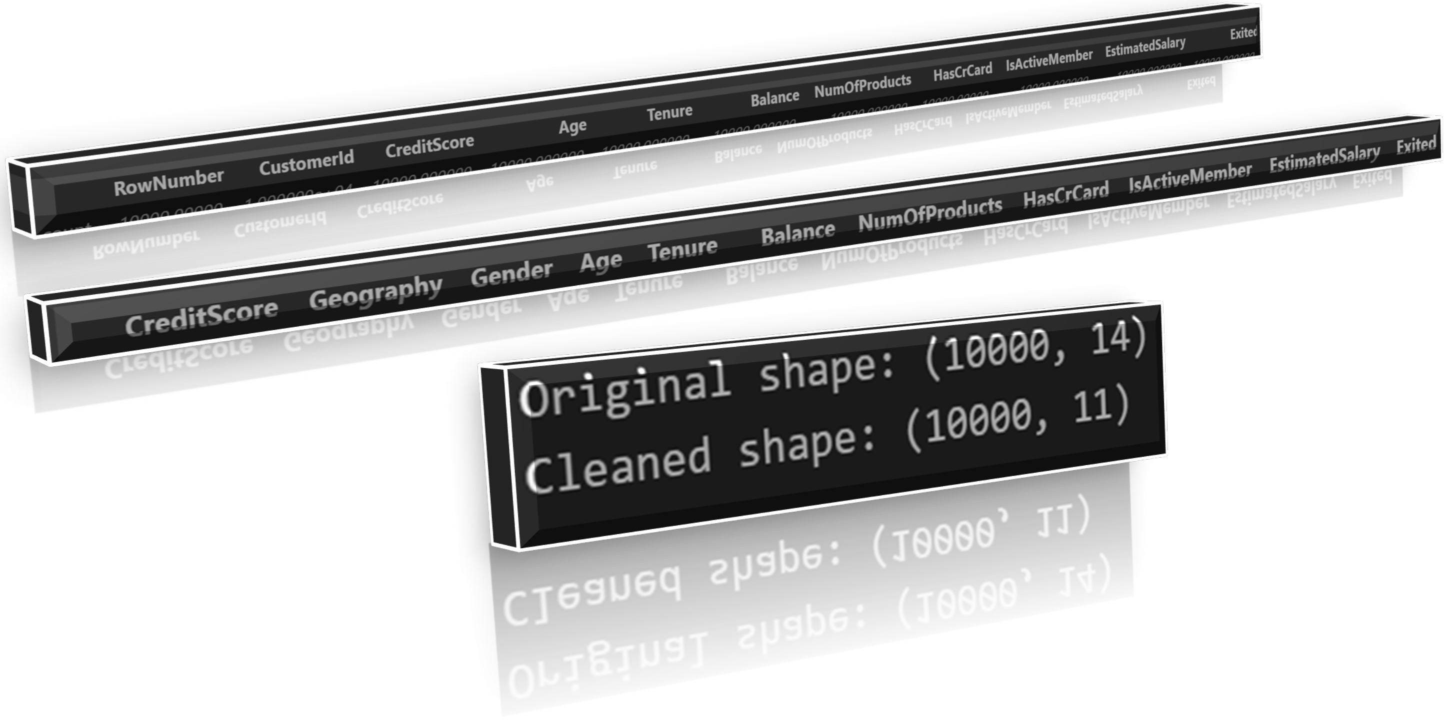
Detect and locate missing or incomplete data points in datasets to prevent analysis errors.

Ensuring Data Consistency

Check for inconsistencies or anomalies to maintain dataset integrity and increase reliability.

Improving Analysis Accuracy

Address data quality issues to ensure accurate and dependable analytical results.



Final Dataset Shape and Readiness for Analysis

Dataset Dimensions Confirmed

The dataset's rows and columns are verified to ensure accurate data structure after cleaning.

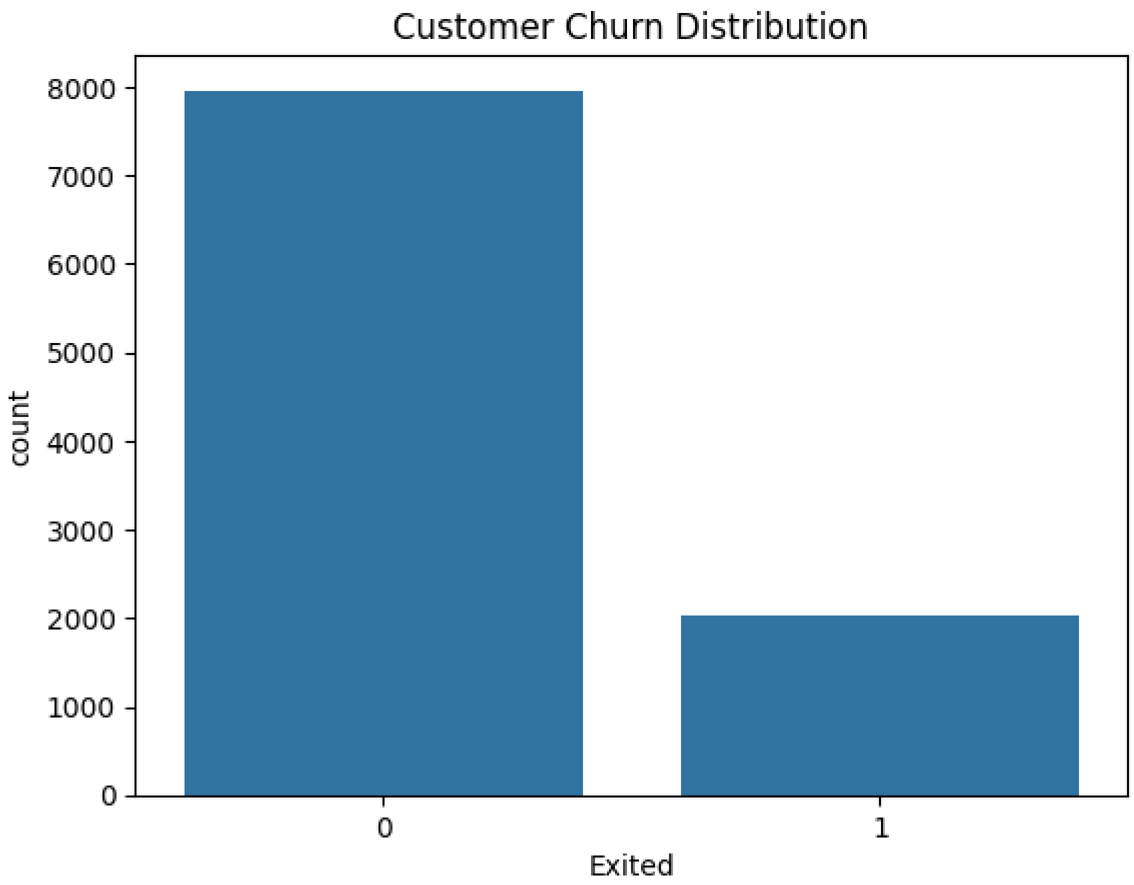
Data Cleaning Completed

Cleaning processes have refined the data, removing errors and inconsistencies for analysis readiness.

Ready for Exploratory Analysis

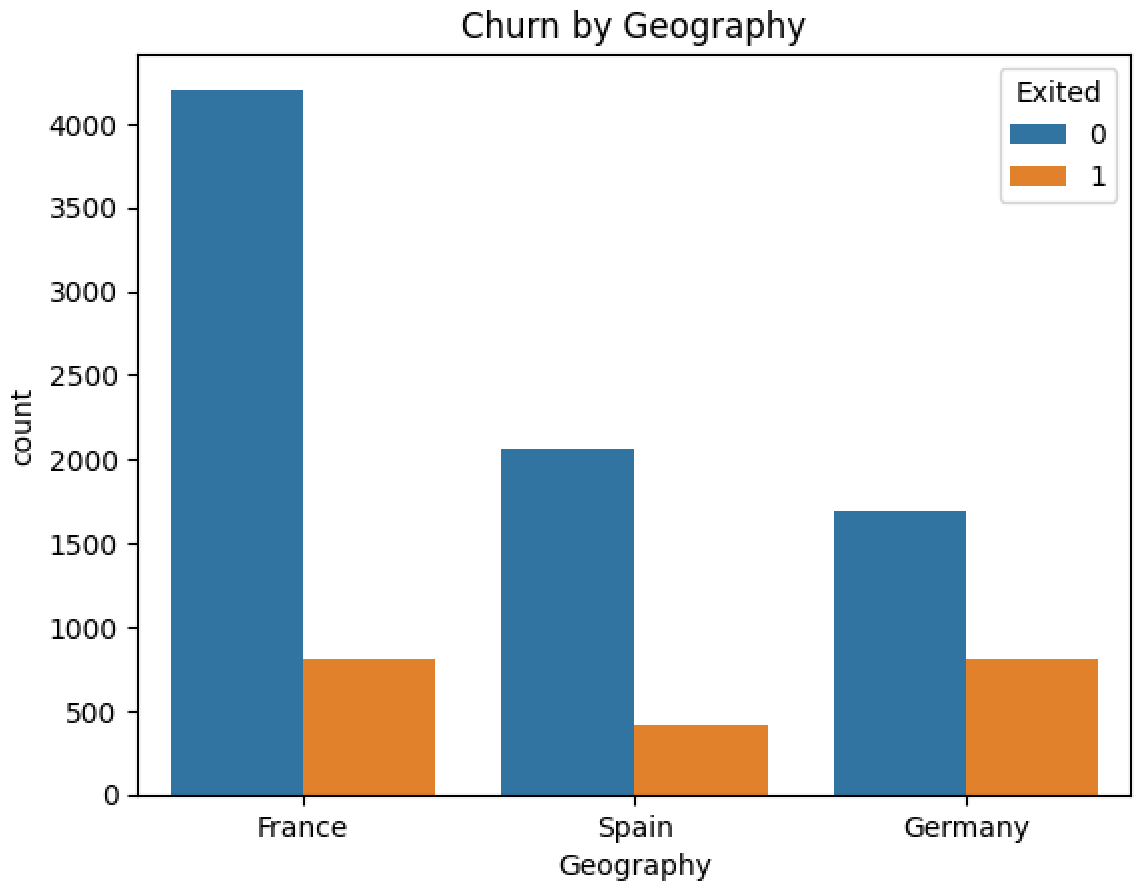
The prepared dataset is now suitable for performing exploratory data analysis effectively.

Exploring Churn Distribution and Categorical Drivers



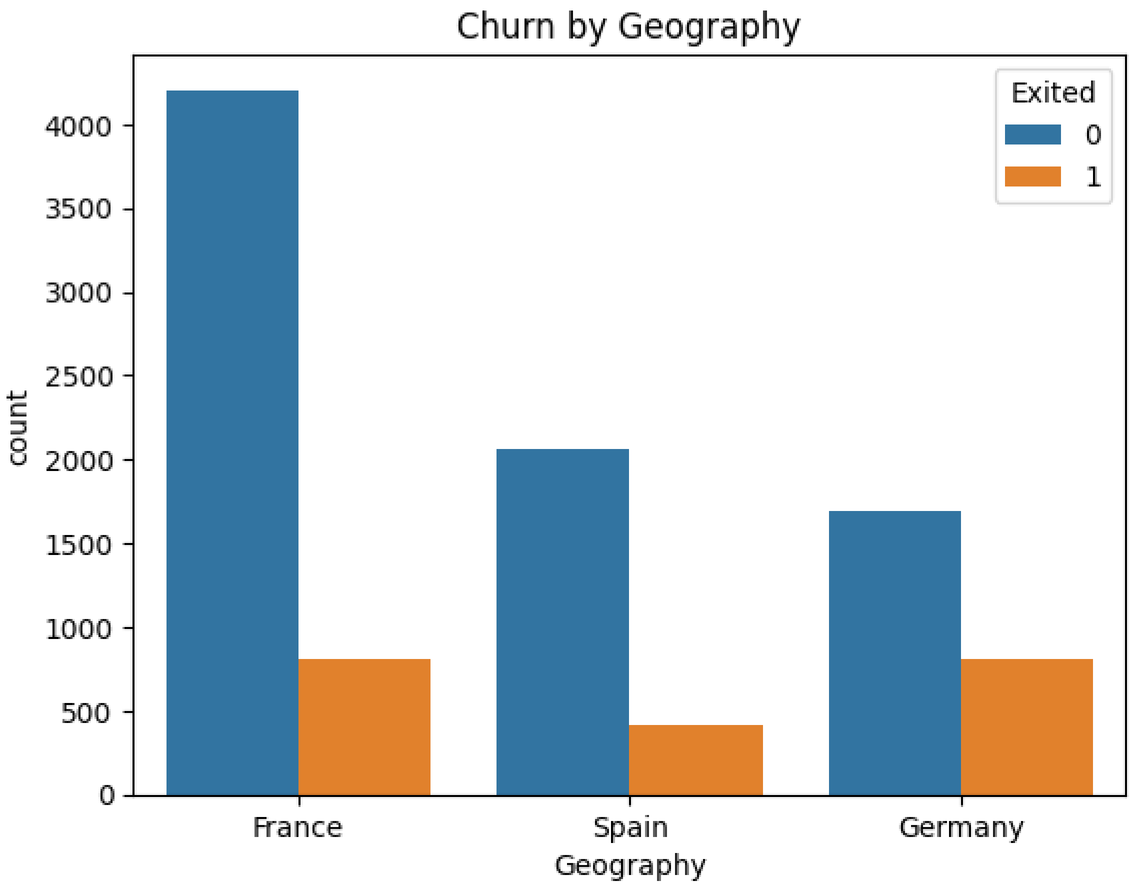
Churn Distribution

- Insight: ~20% of customers churn, ~80% retained.
- Sets up the challenge → churn is a minority class.



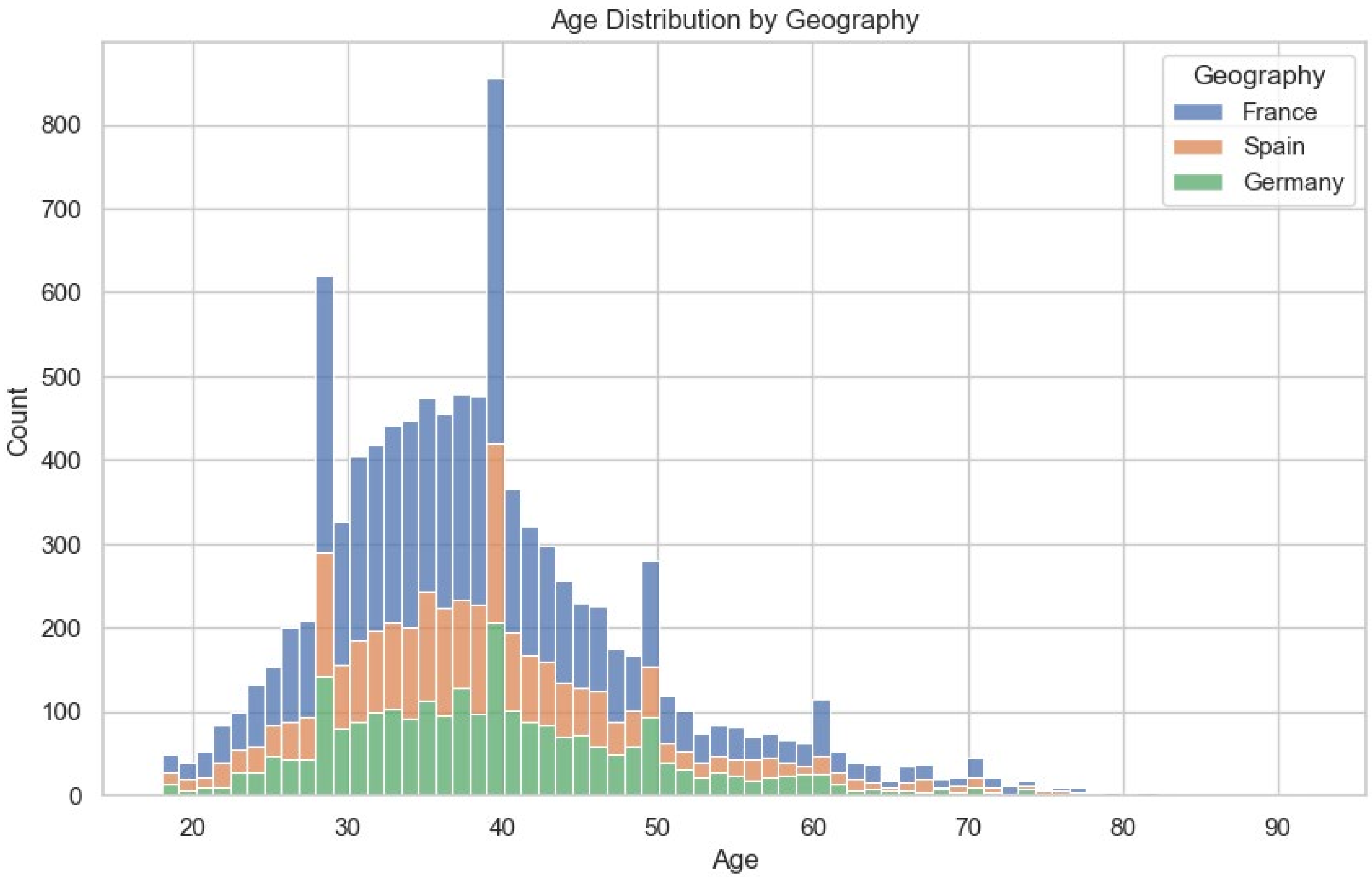
Analyzing Churn by Geography

Insight: Germany has the highest churn, France the lowest.



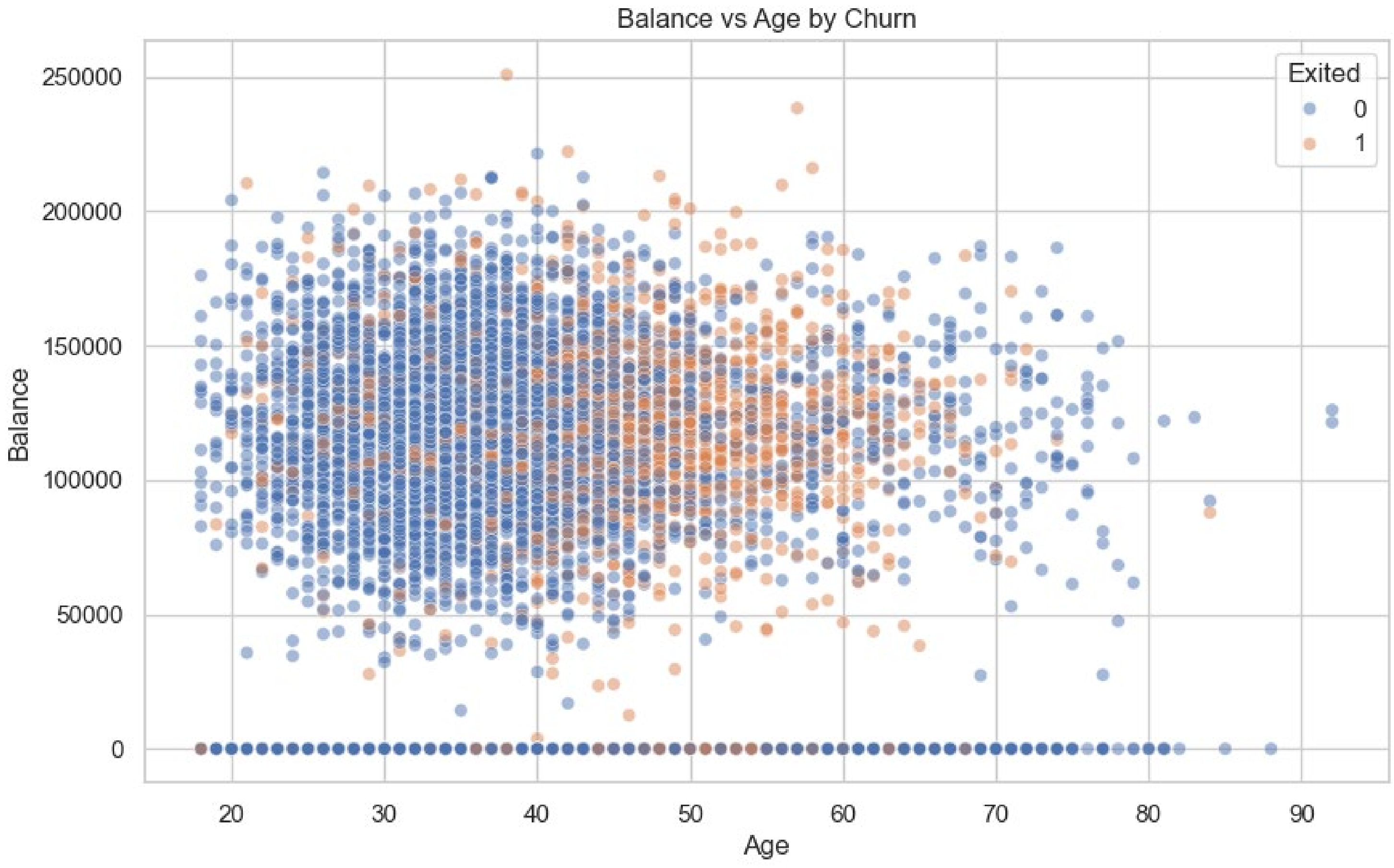
Investigating Churn by Geography

Insight: Germany has the highest churn, France the lowest.



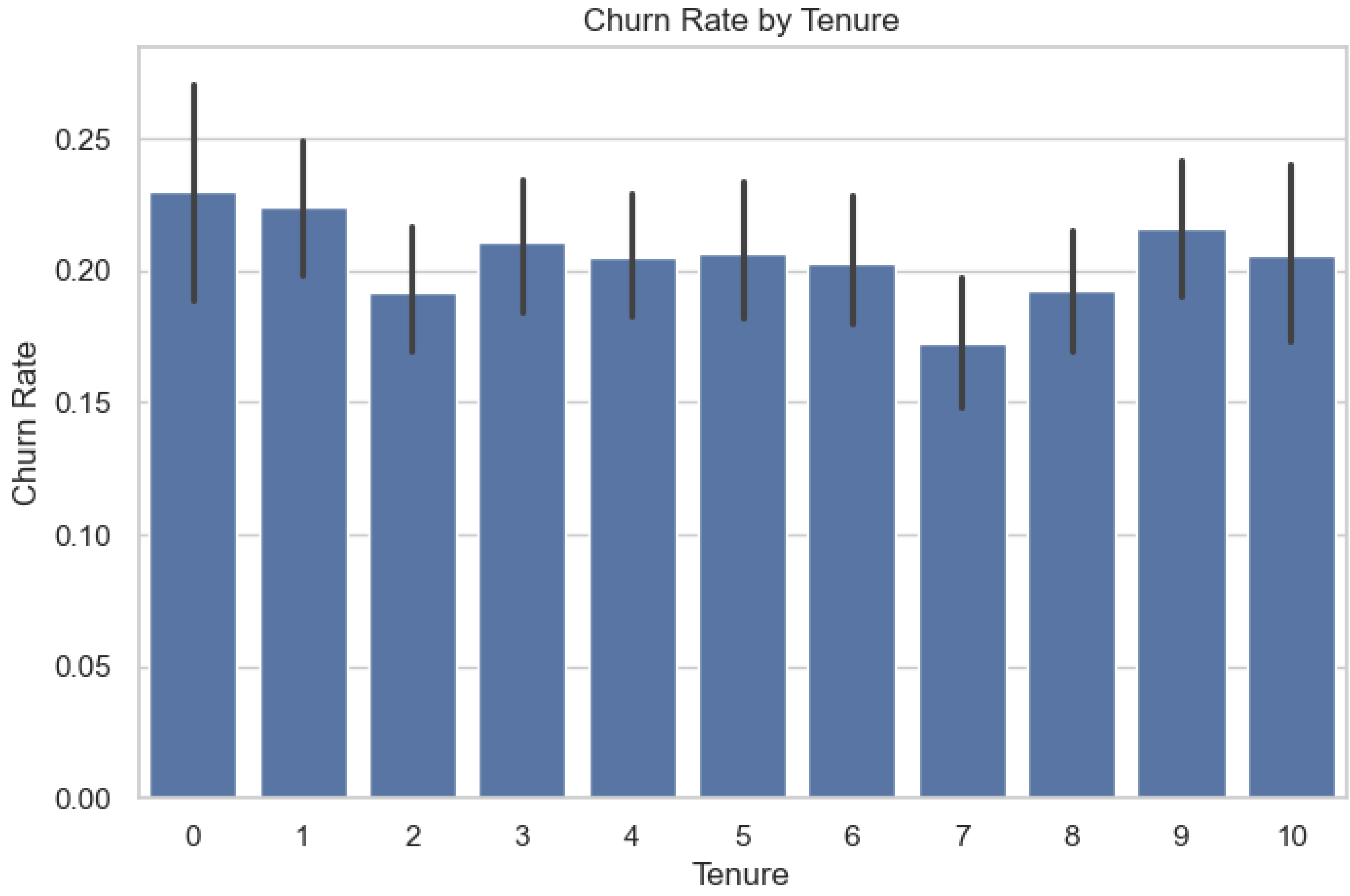
Age Distribution by Geography – Key Takeaways

- The **age profile of customers differs significantly by geography**.
- France has the largest base of customers across most age groups, concentrated around **30–40 years old**.
- Germany shows a **heavier concentration of older customers**, which aligns with our earlier finding that **older age correlates with higher churn**.
- Spain's distribution is more balanced, sitting between the other two markets.
- This suggests that **Germany's higher churn rate is partly explained by its older customer base**, while France's younger profile may contribute to its lower churn.



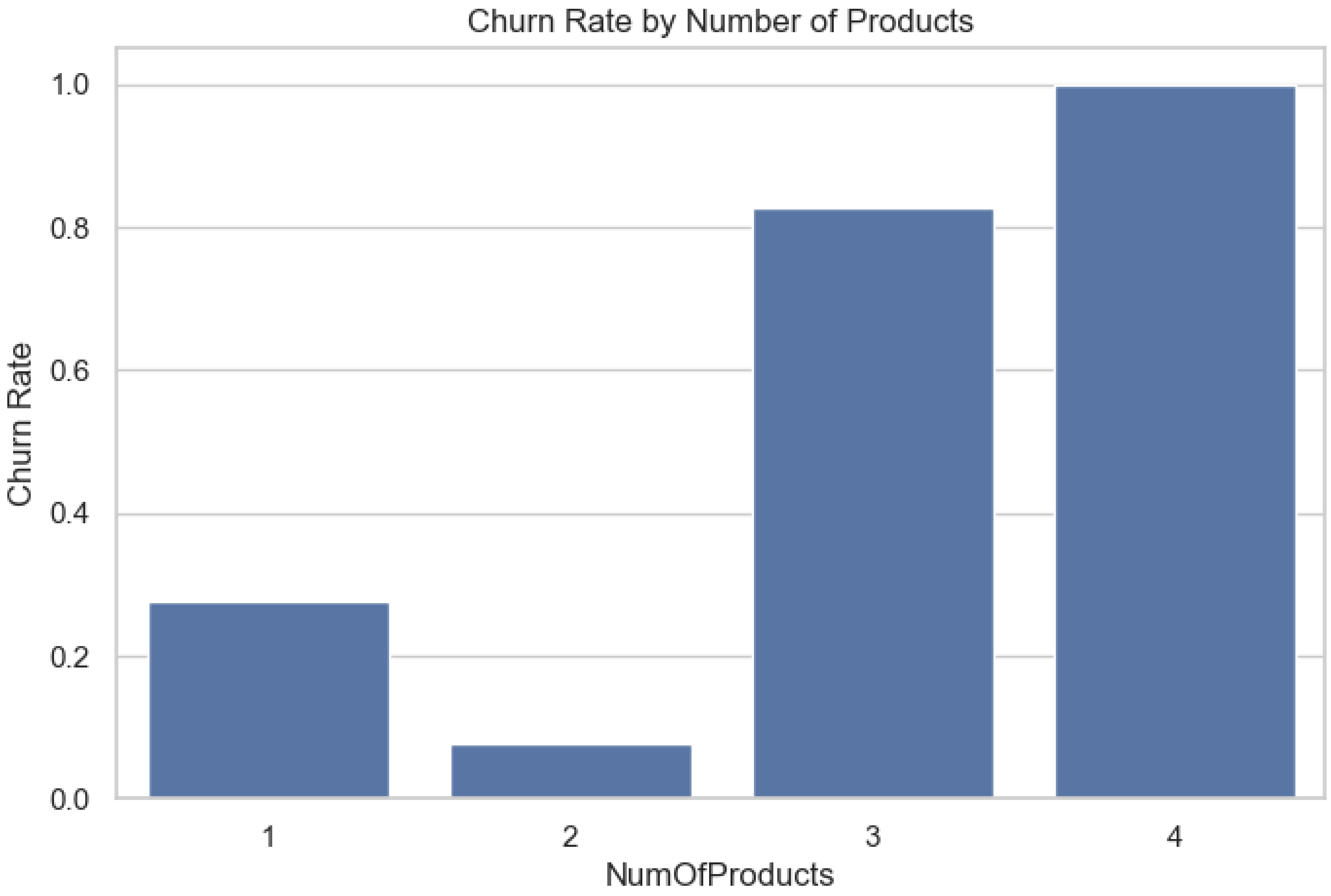
Balance vs Age by Churn - Takeaways

- **Balance vs Age by Churn – Key Takeaways**
- Churn (orange points) is **more concentrated among older customers with higher balances**.
- Younger customers with similar balances tend to remain loyal (blue points).
- A significant number of **zero-balance customers did not churn**, suggesting that inactivity doesn't always translate to leaving – possibly due to other products or stickiness factors.
- This confirms that **age and balance together explain a portion of churn risk**: older, high-value customers are more likely to exit.
- **Overall Insight:**
- Customer churn isn't explained by age or balance alone – it's their **interaction** that reveals high-risk segments, especially older customers holding larger balances.



Churn by Tenure – Key Takeaways

- Churn is relatively **consistent across tenure levels**, with rates fluctuating between ~17% and ~23%.
- Customers with **very short tenure (0–1 years)** churn more, suggesting onboarding and early experience matter.
- Churn dips slightly at **7–8 years**, indicating long-term loyalty effects, but rises again at 9–10 years.
- **Overall:** tenure alone is not a strong predictor of churn, but **new customers** represent a higher risk segment that could benefit from early retention strategies.



Churn by Number of Products – Key Takeaways

- Clear non-linear relationship between product ownership and churn:
  - Customers with **1 product churn moderately** (~25%).
  - Customers with **2 products churn the least** (~8%).
  - Churn **spikes dramatically** for customers with **3 (~82%) and 4 (~100%) products**.
- This suggests that **holding 2 products creates loyalty**, but customers with **too many products may feel overextended or dissatisfied**.
- **Business Implication:** Encourage **cross-sell to 2 products** for stability, while reviewing pain points for customers with 3 or more products.
- **Overall Insight:**
- Product ownership is one of the strongest churn predictors. **Two products = retention sweet spot**, while **three or more products signal high churn risk**.

Investigating Churn by Active Membership Status

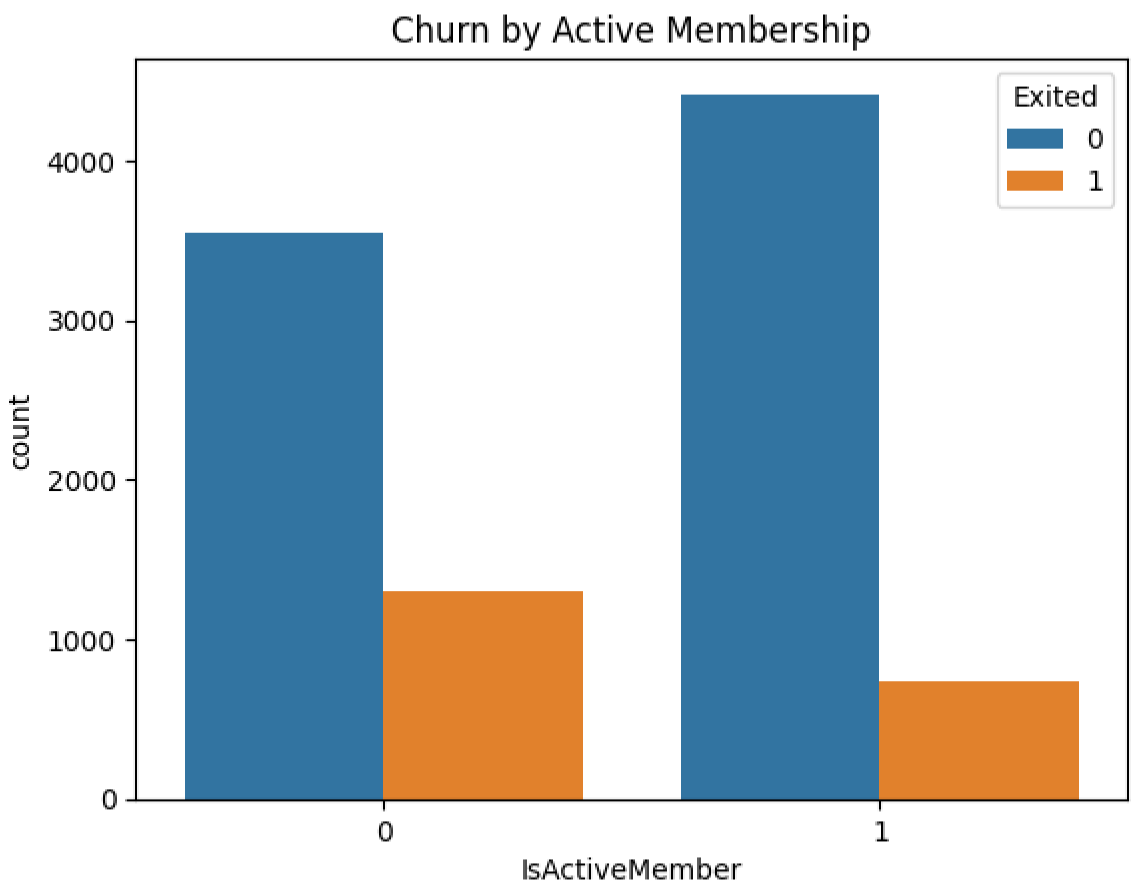
Churn Rate Analysis

Comparing churn rates reveals how member activity impacts retention rates significantly.

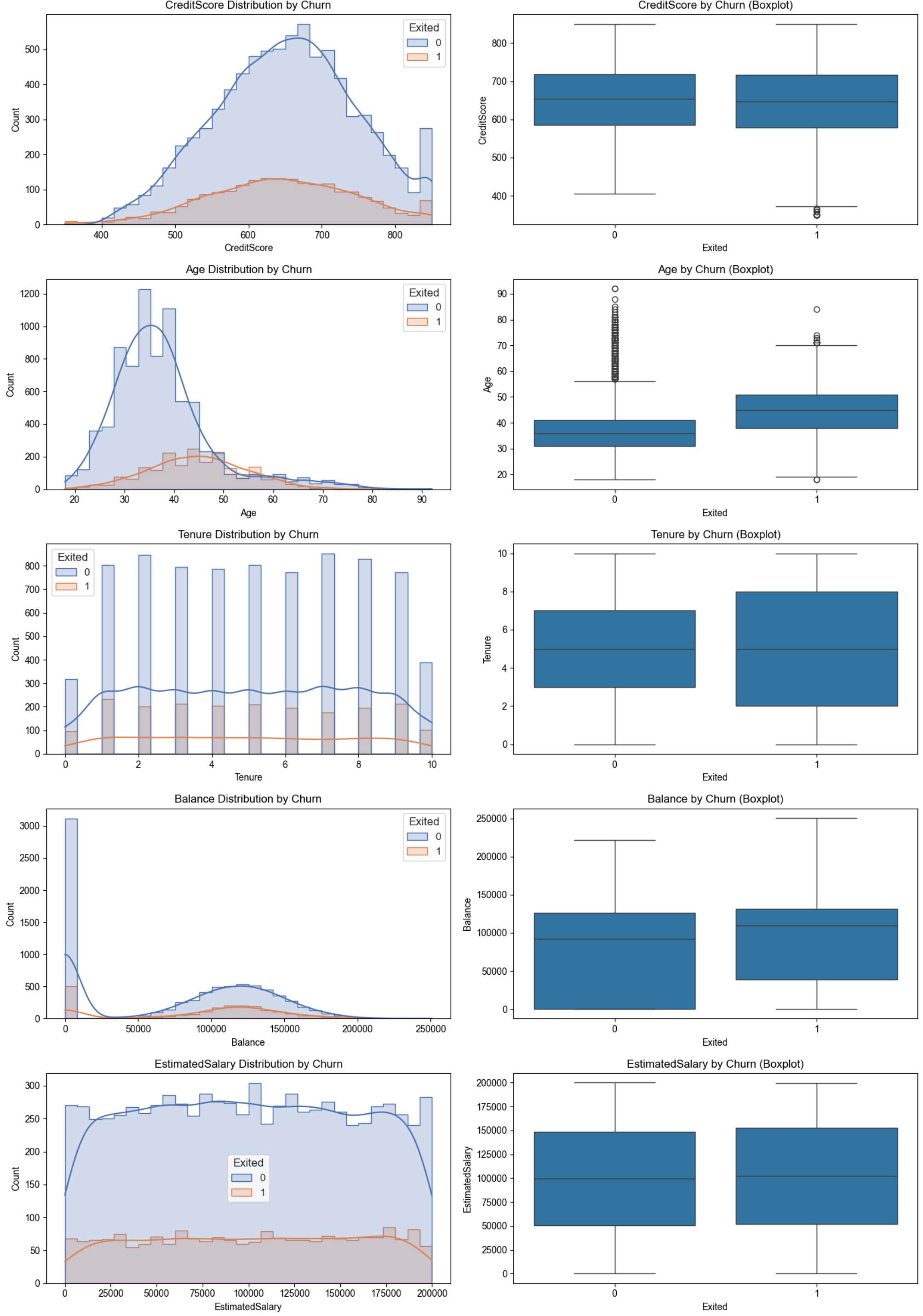
Engagement and Retention

Higher engagement levels among active members strongly correlate with improved retention.

Insight: Inactive members churn at much higher rates.



Numerical Feature Analysis to Identify Churn Patterns



Age and Balance Correlations with Churn

Age Impact on Churn

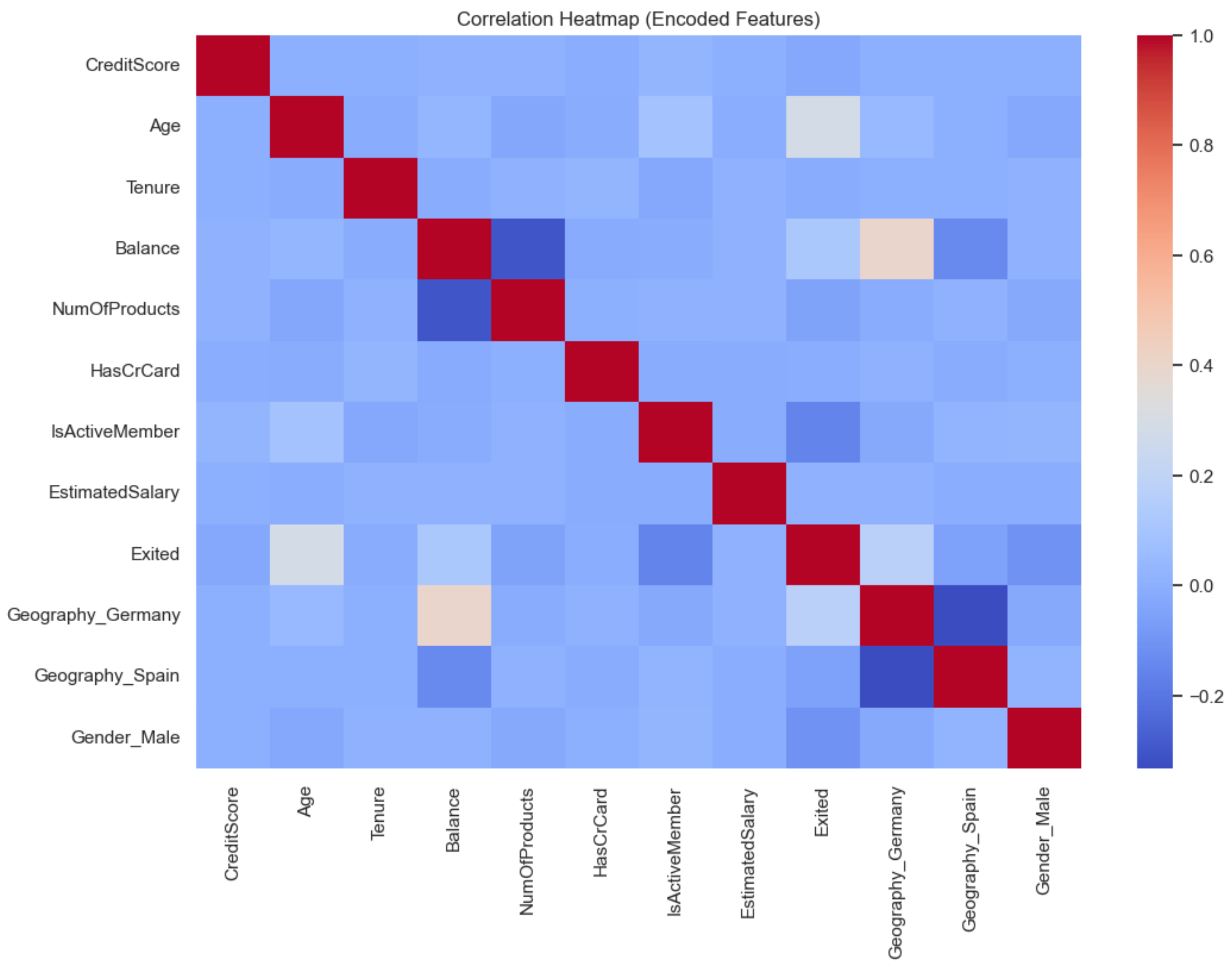
Age is a strong churn driver – older customers are more likely to leave.

Account Balance Influence

Balance shows extremes: customers with very high balances churn more, but zero-balance customers are often retained.

Combined Factors Analysis

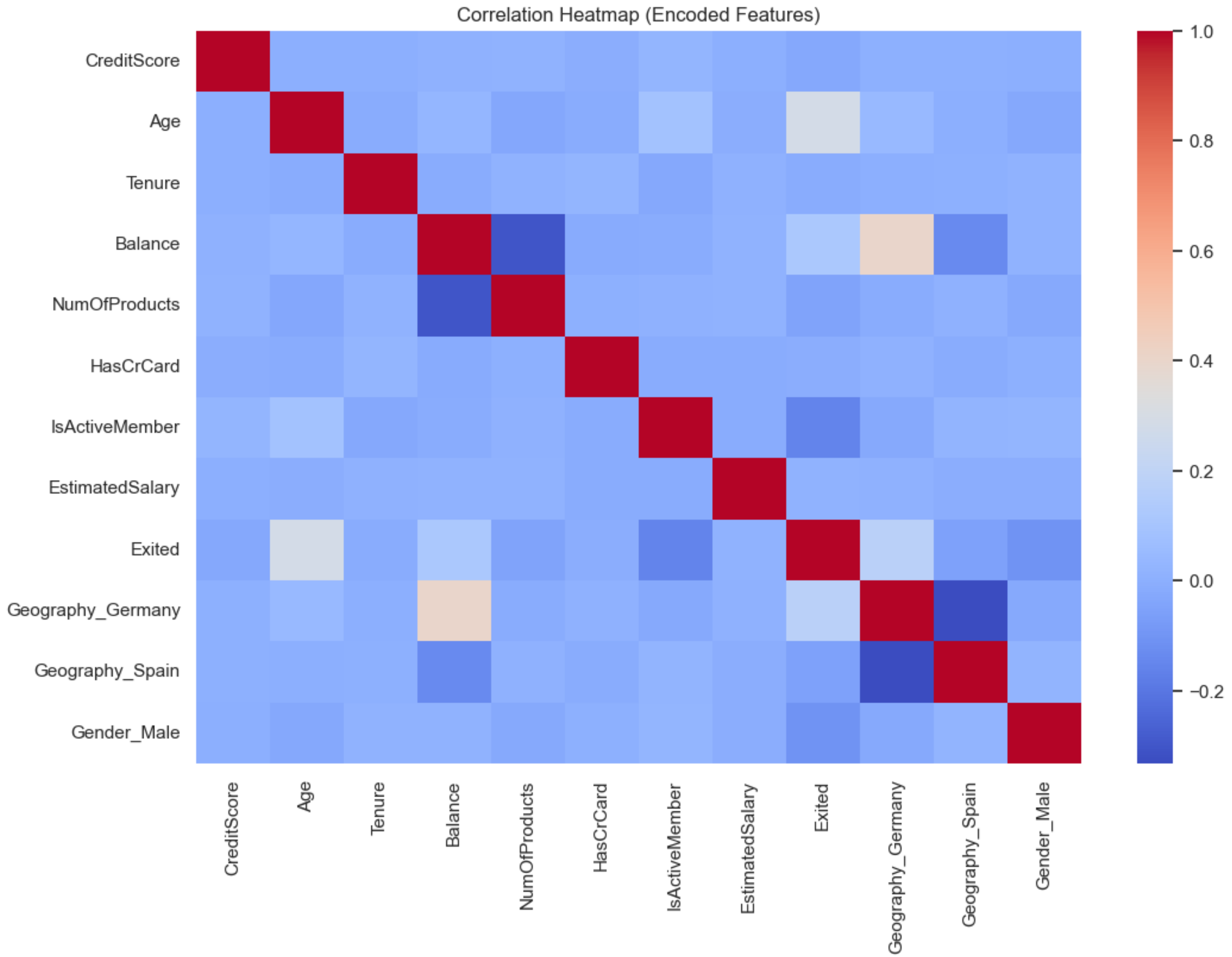
Credit score, tenure, and salary have limited impact on churn.



Feature Relationships with Churn.

- **Correlation Heatmap – Key Takeaways**
- The heatmap highlights relationships between **customer attributes** and the **churn outcome (Exited)**.
- **Age** shows a positive correlation with churn, confirming that **older customers are more likely to leave**.
- **Balance** has a modest positive correlation with churn, suggesting that customers with higher account balances are more prone to exit.
- **IsActiveMember** is negatively correlated with churn, meaning **active members are less likely to churn**, supporting earlier categorical insights.
- **Number of Products** shows a negative correlation, but recall our earlier finding that churn rises again at 3–4 products, showing a **non-linear effect** not fully captured by correlation alone.
- Features like **CreditScore**, **Tenure**, and **EstimatedSalary** show very weak correlations with churn, indicating limited predictive value.
- Geography dummies (Germany, Spain) show clear relationships: **German customers are more likely to churn**, while French customers (the reference group) are least likely.
- **Overall Insight:**
  The correlation heatmap reinforces that **age, activity status, geography, balance, and product ownership** are the most influential features, while others like credit score and salary contribute very little. This helps prioritize which factors to focus on in predictive modeling and business strategies.

Key Drivers of Churn and Actionable Recommendations



Top Predictive Features: Geography, Membership, Products, Age, Balance

Key Drivers Identified in EDA

- **Geography** → German customers churn at much higher rates; French customers are least likely to churn.
- **Membership** → Inactive members are nearly twice as likely to leave compared to active members.
- **Number of Products** → Customers with **2 products churn the least**; churn spikes dramatically for those with **3+ products**.
- **Age** → Older customers show a clear trend toward higher churn.
- **Balance** → Higher account balances are modestly linked with churn; customers with **zero balances often remain loyal**.

Overall Insight:

- Churn is most influenced by **customer demographics (age, geography)** and **behavioral factors (products, membership activity)**, while variables like credit score, tenure, and salary play a minimal role.

Retention Strategies to Reduce Churn

Personalized Offers

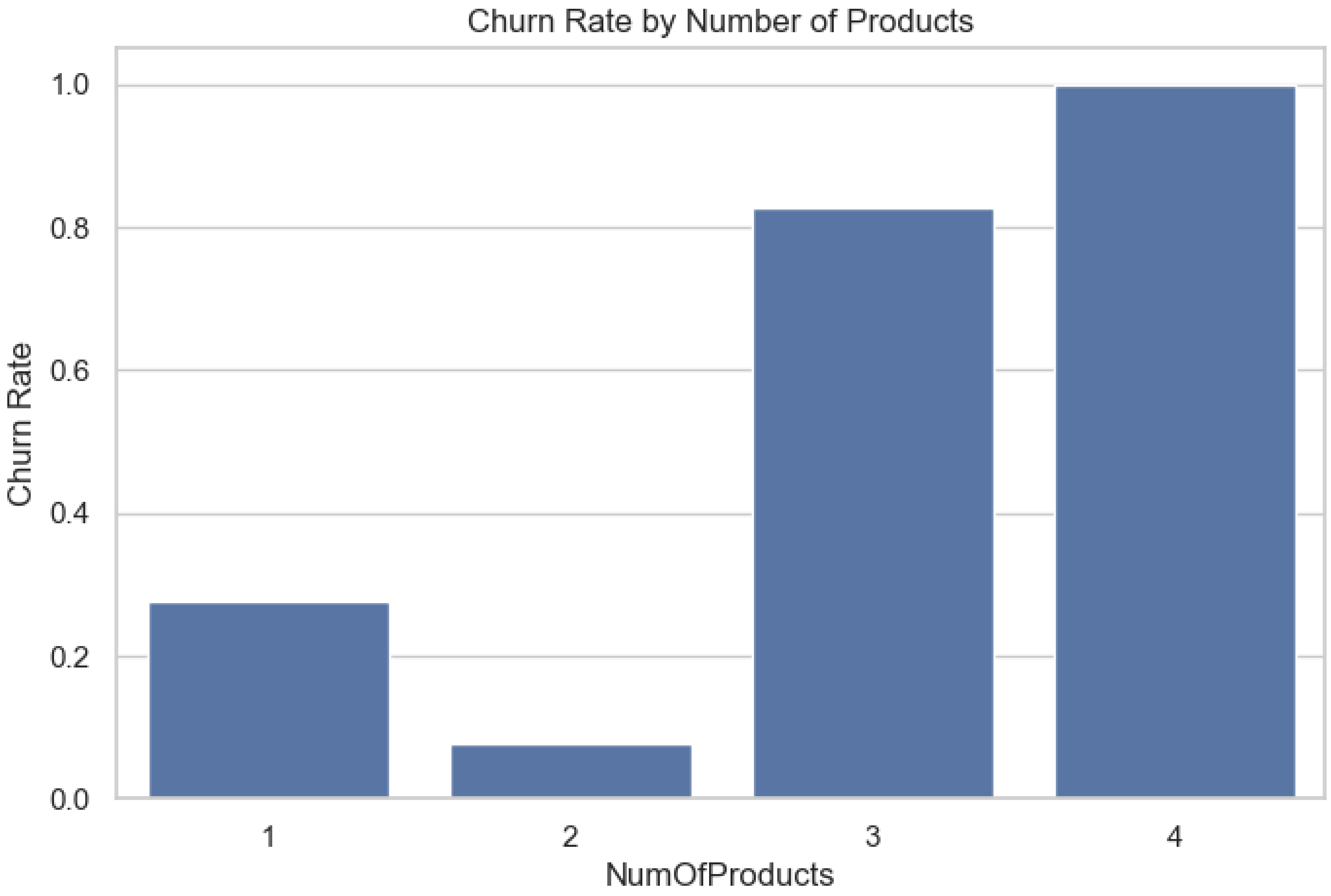
Tailoring offers to individual customer preferences increases satisfaction and reduces churn effectively.

Improved Customer Service

High-quality and responsive customer support fosters loyalty and decreases customer attrition.

Enhanced Engagement

Engaging customers through interactive communication channels strengthens relationships and lowers churn rates.



Targeted Engagement for High-Risk Segments

High-Risk Profiles to Address

- **Single-product customers**: More likely to churn; should be encouraged to adopt a second product through cross-sell offers.
- **3+ product customers**: Show extremely high churn; investigate dissatisfaction or product complexity driving exits.
- **Inactive members**: Engagement campaigns (loyalty rewards, reminders) can reduce churn.
- **Older, high-balance customers**: Proactive relationship management needed, as they represent valuable but vulnerable clients.

Proactive Retention Strategy:

- By focusing on these **specific high-risk segments**, organizations can reduce churn more effectively than with one-size-fits-all strategies.

Conclusion

Value of Data Insights

Exploratory data analysis revealed clear churn drivers: geography, membership activity, product ownership, age, and balance. These insights provide a fact-based understanding of why customers leave.

Reducing Customer Churn

By targeting inactive members, German customers, single-product holders, and older high-balance clients, organizations can design retention programs that directly address high-risk segments.

Driving Business Growth

Focusing on churn reduction not only improves retention but also creates opportunities for cross-selling, stronger customer relationships, and long-term profitability. Data-driven decisions ensure strategies are both effective and scalable.