

Bank Bostwana

# Data-Driven Strategies to Reduce Customer Churn

*A Comprehensive Analysis for Bank Bostwana*

- Roberto Pangihutan Situmeang
- pasitumeangroberto@gmail.com
- [LinkedIn: Roberto P Situmeang](#)



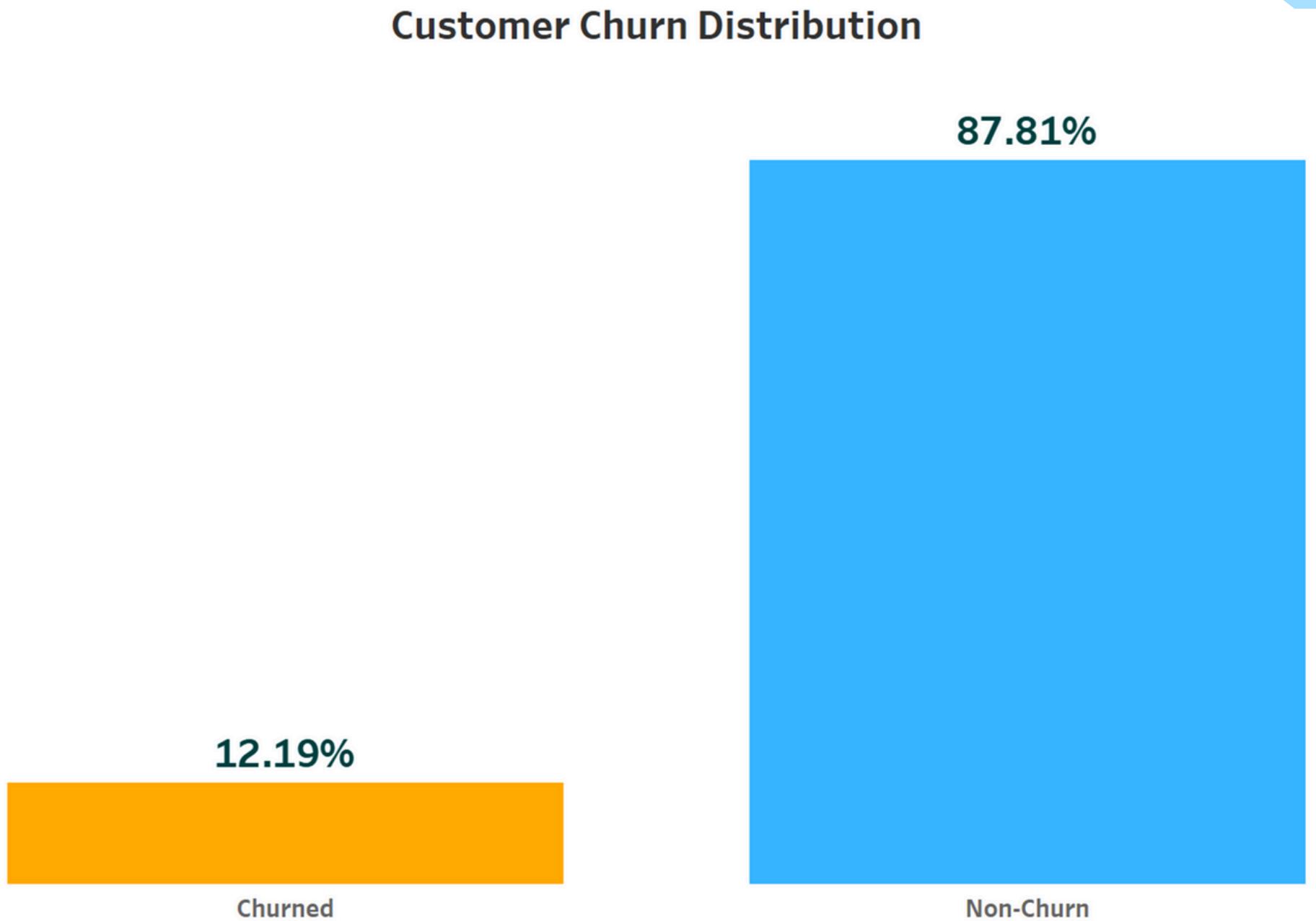
# EXECUTIVE SUMMARY

Project Description	How can we reduce customer churn from 12.19% to 10%
Objective	<ul style="list-style-type: none"><li>Understand Churn Drivers : Identify the factors contributing to customer churn, such as financial health, demographics, communication preferences, and service issues.</li><li>Segment Customers : Group customers into segments (e.g., high-risk, low-risk) based on their likelihood to churn.</li><li>Predictive Modeling : Build predictive models to forecast churn and prioritize interventions for at-risk customers.</li><li>Actionable Insights : Provide tailored strategies to reduce churn and enhance customer satisfaction.</li></ul>
Key Metrics	Retention Rate, Churn Rate, Customer Lifetime Value (CLV), Debt-to-Income Ratio, Credit Score, Customer Tenure, Outstanding Loans, Number of Complaints, Preferred Communication Channel, Reasons for Churn.
Tools	<ul style="list-style-type: none"><li>Python use for data preparation until data clean.</li><li>Tableau use for exploratory data analyst and provide data visualization</li></ul>
Analysis results	The analysis reveals that customers with high debt-to-income ratios and low credit scores are at greater risk of churning, especially divorced individuals and those with multiple dependents. Younger customers, particularly Gen Z and Millennials, are more likely to leave due to better offers from competitors. Frequent complaints and low engagement also significantly increase churn likelihood. A predictive model using Random Forest identified key churn drivers—debt-to-income ratio, credit score, number of complaints, and customer tenure—with 85% accuracy, enabling reliable forecasting.
Recommendation	To reduce churn, Bank Botswana should prioritize training staff to resolve complaints faster and implement real-time feedback tools to address dissatisfaction. Analyzing competitors' offers and introducing competitive pricing or personalized promotions can counteract churn caused by "Better Offers Elsewhere." Younger customers can be engaged through digital banking solutions and gamified savings goals, while high-risk customers should be proactively contacted with tailored solutions like fee waivers. Additionally, monitoring financial health indicators such as declining balances will help identify at-risk customers early for preemptive interventions.

# BACKGROUND OF PROJECTS



Bank Bostwana is a leading financial institution that serves a diverse customer base, ranging from individual account holders to corporate clients. Despite its strong market presence, the bank has been experiencing customer attrition, with 12.19% of customers churning, as shown in the visualization. This means that approximately 1 in every 8 customers is leaving, which can significantly impact the bank's revenue, market share, and long-term sustainability.



# PROJECT SCOPE

## 1. Data Source:

- Analyze a dataset of 115,640 customer records from Bank Botswana.
- Use internal historical data that includes demographic, financial, and behavioral attributes along with churn labels.

## 2. Objectives:

- Understand the factors that drive customer churn through detailed exploratory data analysis (EDA).
- Develop a predictive model (using Random Forest) to estimate the likelihood of churn, including cost-sensitive threshold tuning.
- Segment customers using K-Means clustering to identify groups with similar characteristics and risk profiles.
- Generate actionable insights and tailor retention strategies for different customer segments.

## 3. Processes Included:

- Data preprocessing: cleaning, handling outliers, standardizing features, and encoding categorical variables.
- Exploratory Data Analysis (EDA): analyzing distributions, correlations, and visualizing key variables to inform feature selection.
- Predictive Modeling: training, evaluating, and tuning a Random Forest classifier.
- Customer Segmentation: applying K-Means clustering and visualizing the results using PCA.
- Dashboard & Monitoring: outlining the strategy for a real-time monitoring dashboard for tracking churn and intervention outcomes.

## 4. Limitations:

- The analysis focuses solely on the internal dataset; external factors or data sources are not included.
- This is an initial phase intended to support strategic decision-making; further testing and refinements (e.g., periodic retraining or additional segmentation) might be required.

# DATA SOURCE AND OVERVIEW

This dataset about Customer Churn from Bank Botswana and contains **115,640 customer records**, including demographic, financial, geographic, and behavioral attributes. It serves as a foundation for analyzing customer churn and identifying key patterns driving customer attrition.

From those thirty one variables, RowNumber, CustomerID, Surname, Firstname are identifiers so this means these variables is unuseful for this analysis. The more important variable is “Churn Flag”, because this variable show wheter or not the customer churn.

This sample of dataset collected from [Kaggle.com](https://www.kaggle.com) .





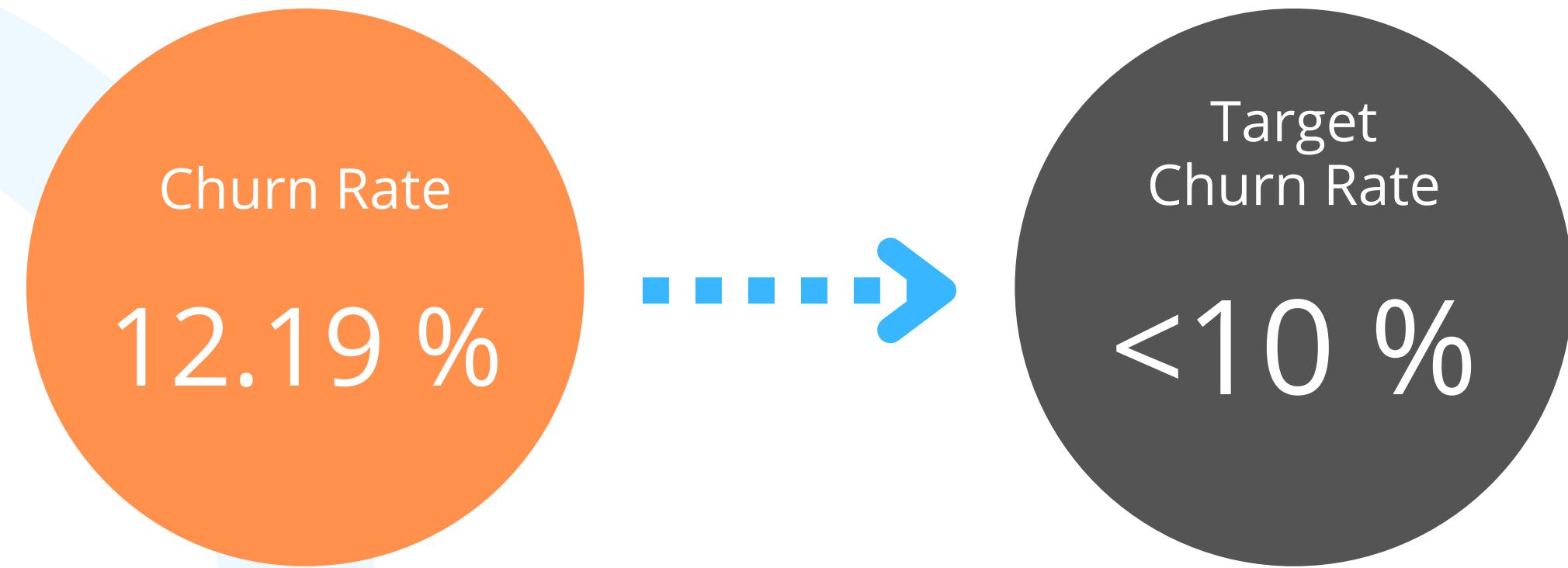
# PROBLEMS

## Problem Statements

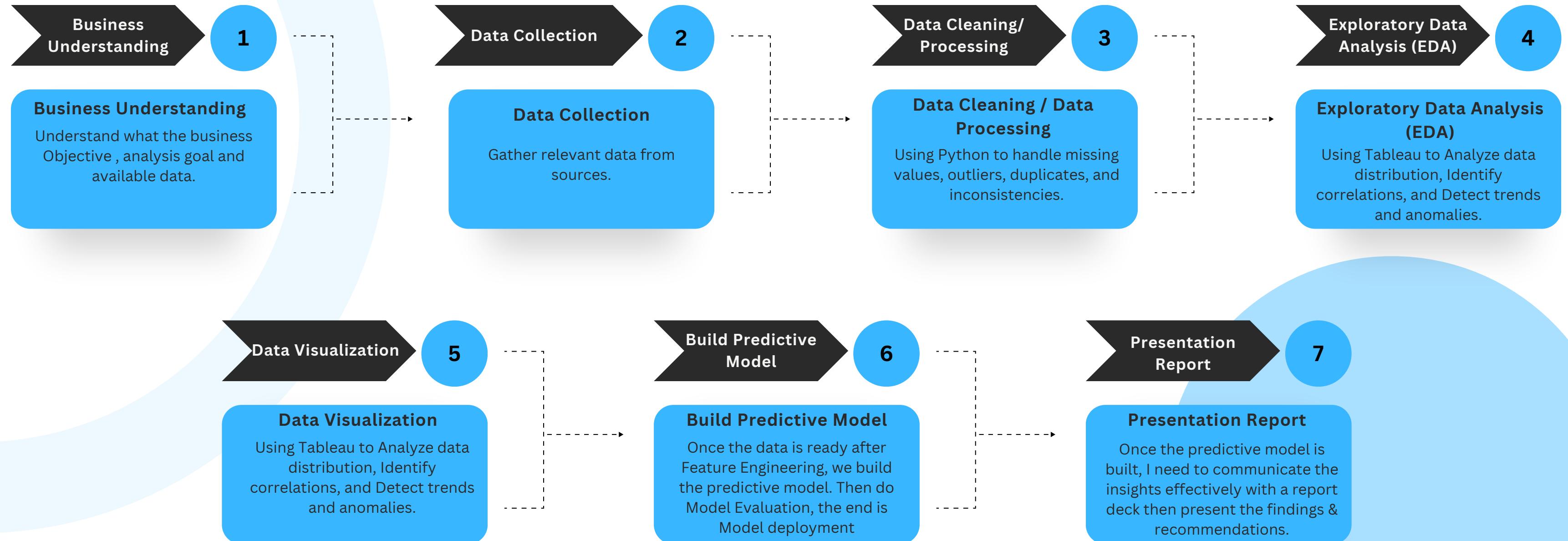
How to reduce customer churn by improving service quality, enhancing product engagement, addressing financial stress, and optimizing communication strategies, particularly among high-risk demographics, business segments, and regions with higher churn rates?

# OBJECTIVE

**"To enhance customer retention by reducing churn from 12.19% to below 10% within the next 12 months through targeted financial interventions, improved service quality, and personalized product engagement strategies."**



# Research & Methodology



# WHY DOES CUSTOMER CHURN MATTER?

- High churn leads to lost revenue and acquisition costs.
- Identifying early disengagement patterns is critical for intervention.
- This analysis connects multiple data visualizations to reveal actionable insights.

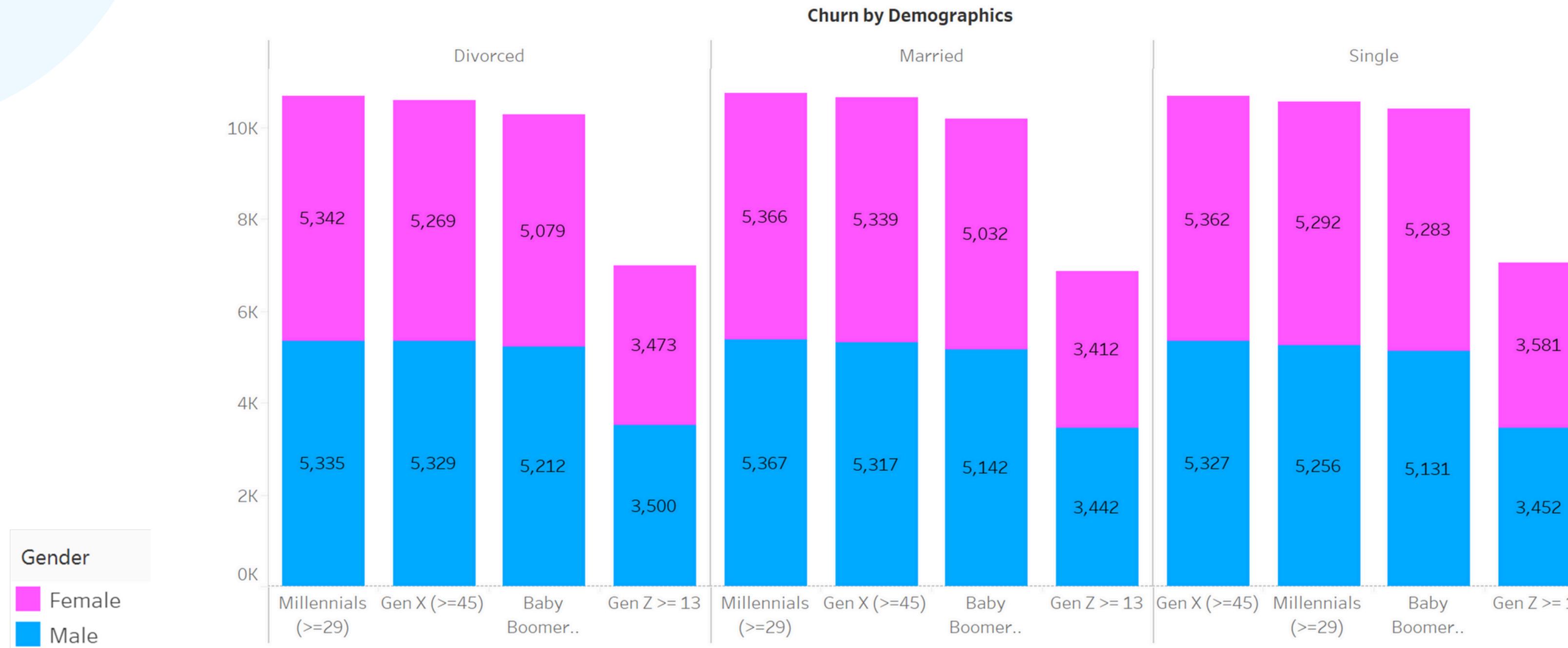




# Customer Demographics & Behavior Analysis

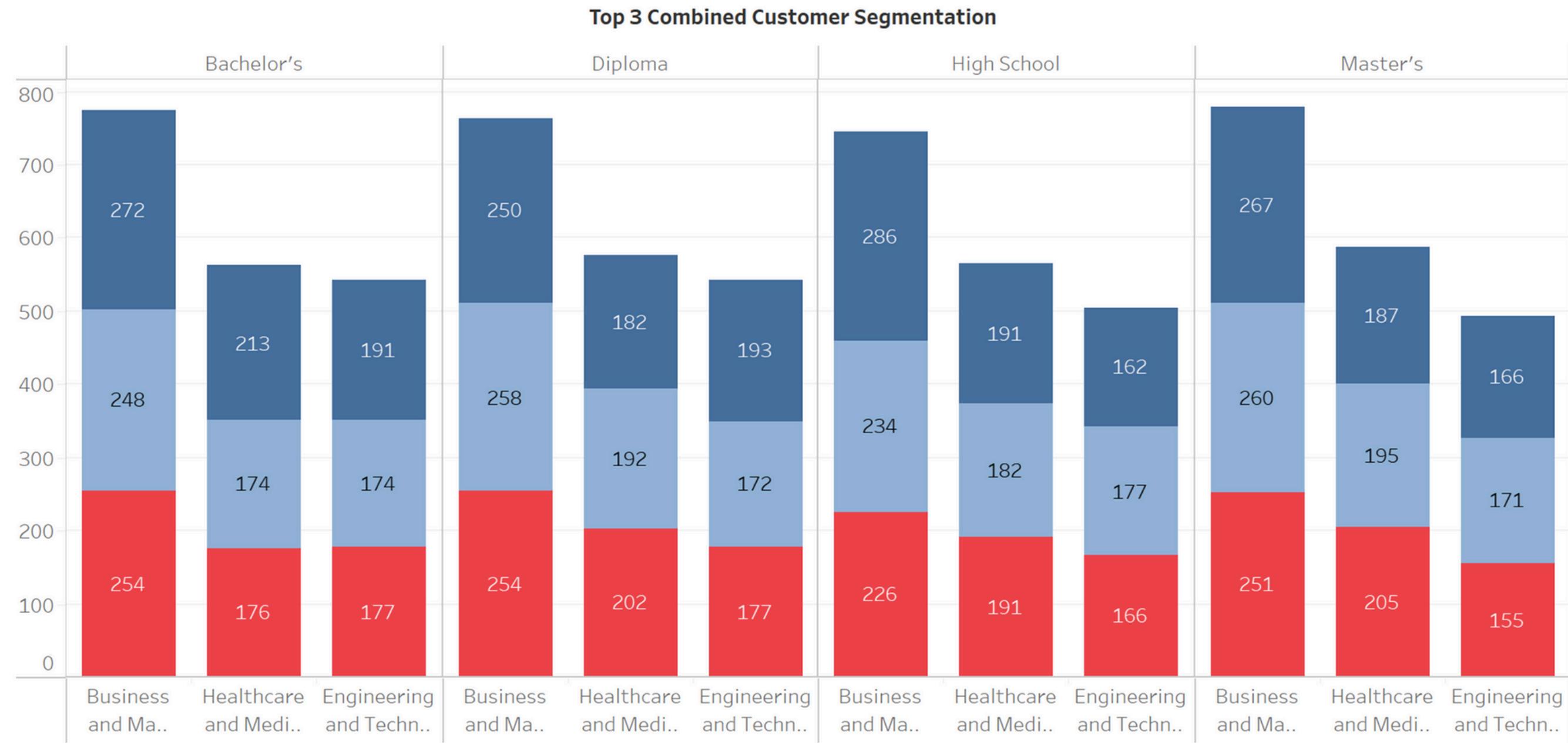
# Millennials & Divorced Customers at Highest Risk.

## Who's Most Likely to Churn?

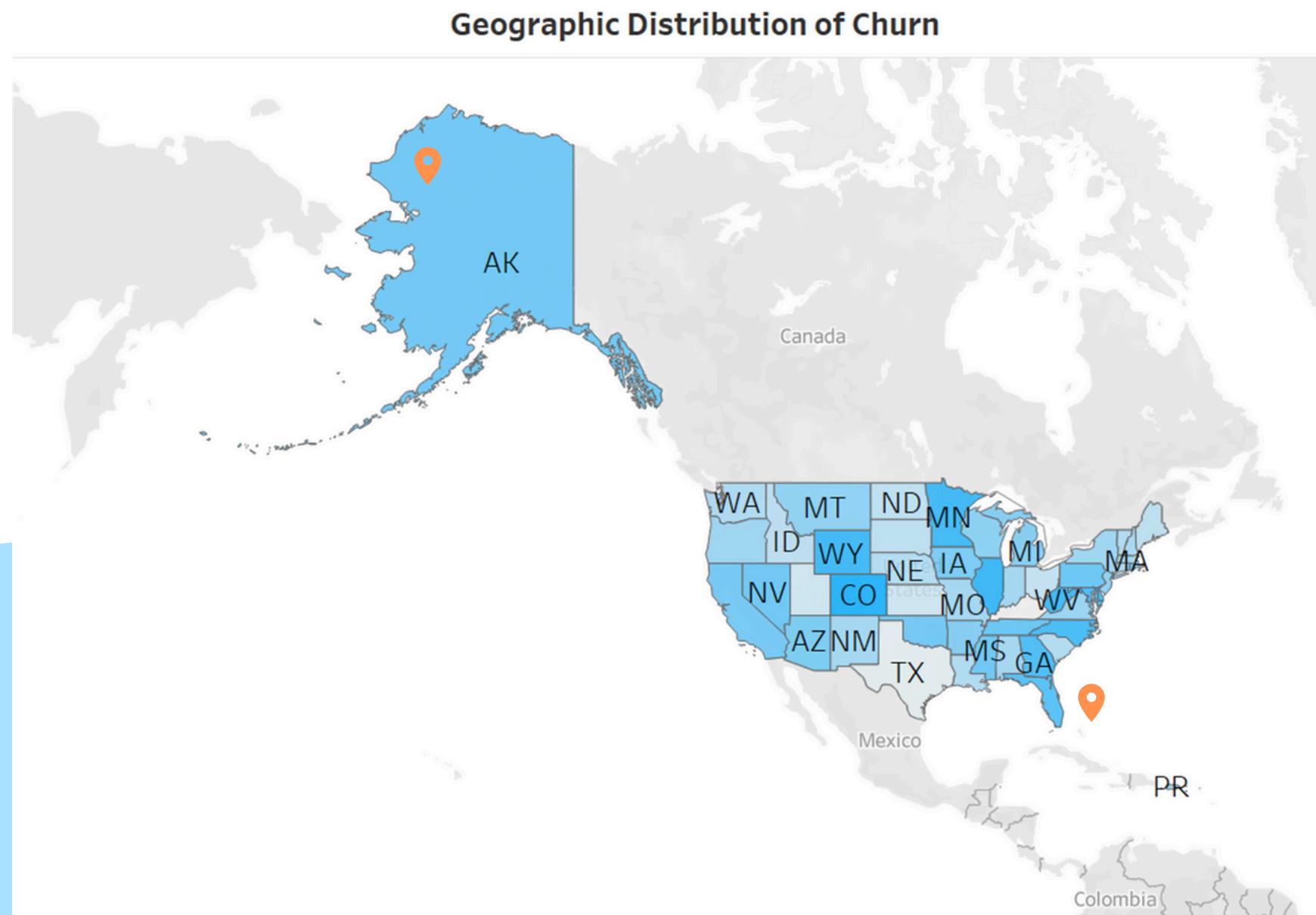


# Retail & SME Customers Drive Churn.

## How Business Type Impacts Retention?



# Alaska & Puerto Rico Are Churn Hotspots – Where We’re Losing the Most Customers



- Certain geographic regions, notably Alaska and Puerto Rico, experience significantly higher churn.
- Possible factors include limited banking access, economic downturns, or local competition.
- A regionalized strategy is needed to address the unique needs of these areas.

# Single-Product Users Are the Most Likely to Leave



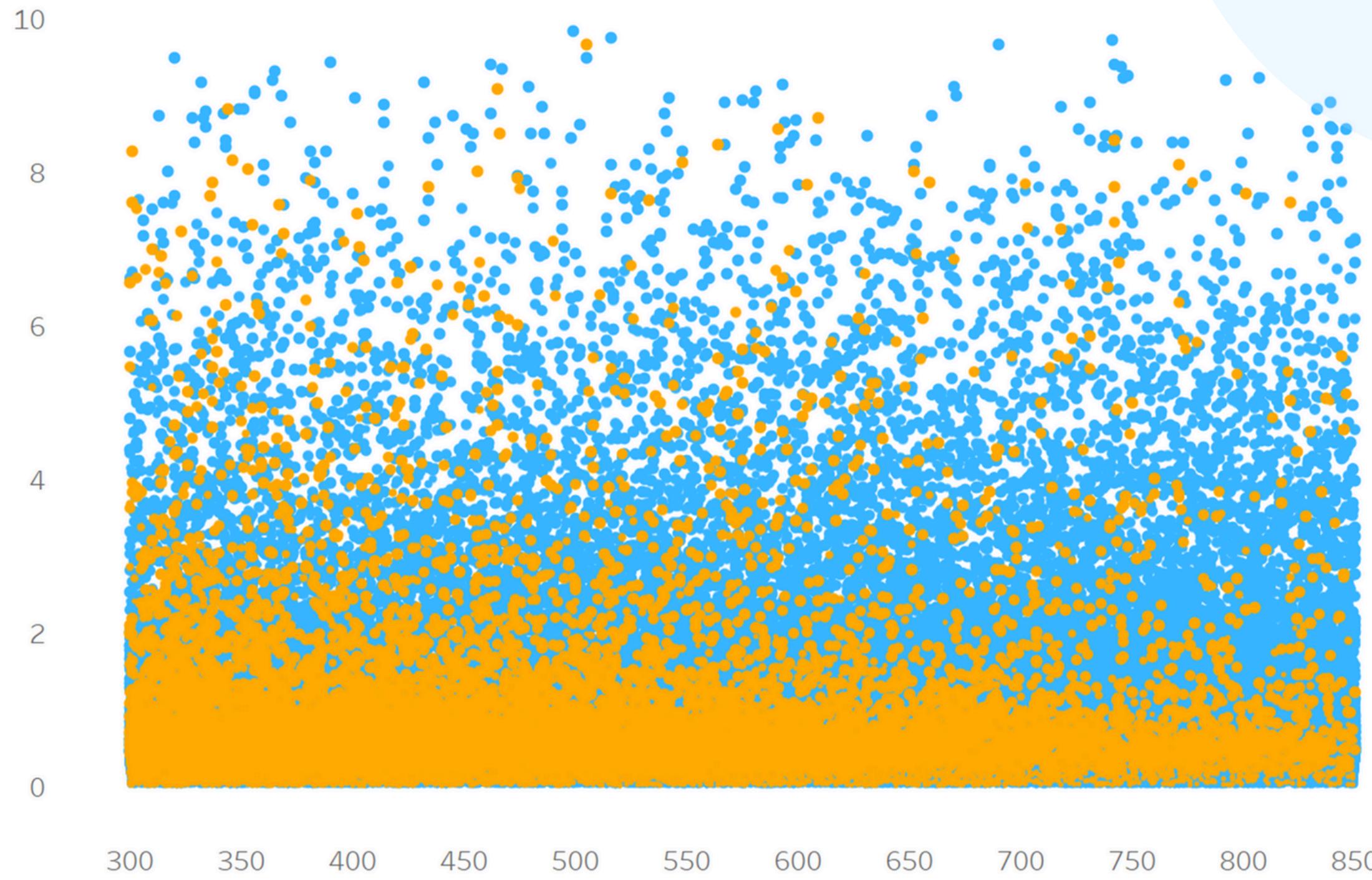
Num Of Pro..	Product and Tenure Analysis				
	Established	Lifetime	Loyal	New	Veteran
1	20.44%	20.99%	21.29%	21.73%	20.50%
2	16.14%	16.90%	15.68%	15.56%	16.58%
3	10.97%	10.93%	11.58%	11.06%	11.98%
4	8.88%	7.51%	7.73%	6.93%	8.13%
5	4.27%	4.56%	4.50%	4.89%	4.11%

- Customers with more products typically exhibit higher retention (or lower churn), since multi-product holders are often more invested in the bank's services.
- Tenure categories highlight if newer customers are more prone to leaving, or if "Veteran" customers remain loyal.
- This indicates whether you should focus on cross-selling to reduce churn or offer special onboarding programs for newer clients.

# Low Credit Score & High Debt – The Red Flags for Customer Churn

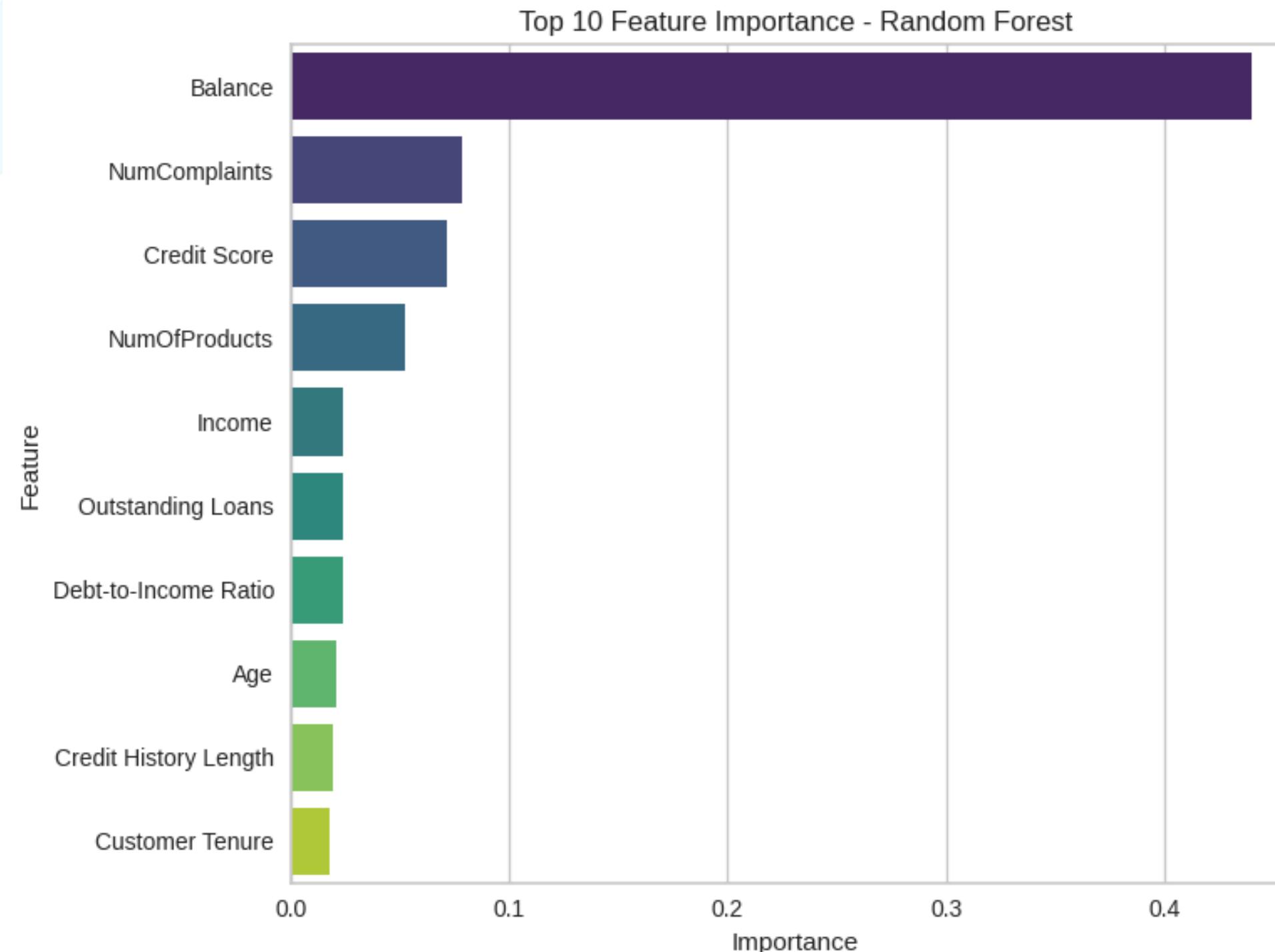
- - Customers with low credit scores and high debt-to-income ratios reduce engagement.
- - Engagement drop is an early warning sign of churn.
- - Actionable Strategy:  
Implement an AI-powered financial stress detection system.

Churn By Financial Factors



# **Linking the Pieces of the Puzzle**

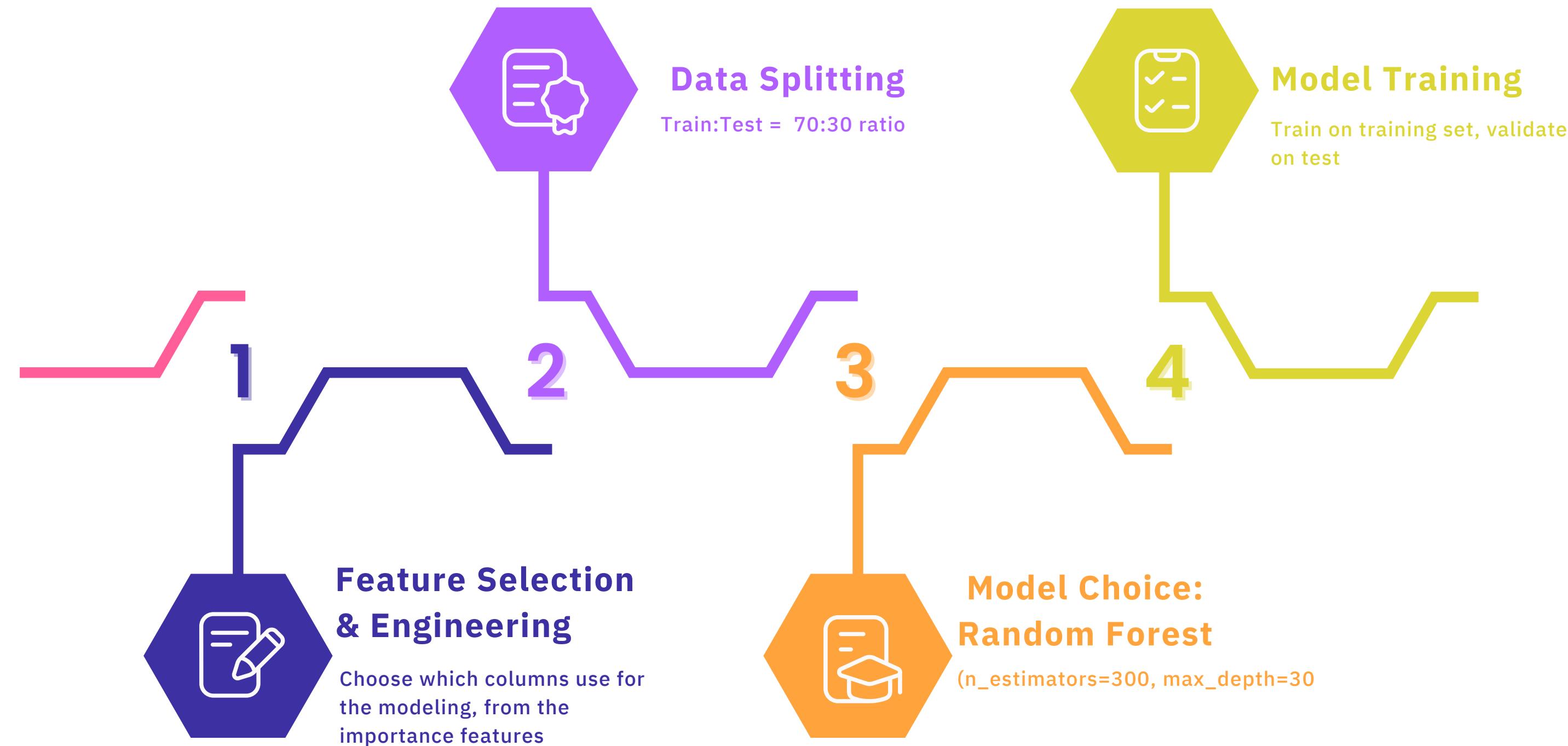
# Top 10 factors contributing to Customer Churn



## Insight

- The most important factor is balance, which can indicate that customers with low balances (or very high ones that suddenly decrease) tend to be more prone to churn. This could be because they feel they no longer get benefits, or they have moved their funds to another bank.
- NumComplaints: The more complaints, the higher the likelihood of churn. This means that customer service greatly influences loyalty.
- A low Credit Score usually reflects high financial risk, which can trigger churn due to customer difficulties in managing finances or not meeting certain product requirements.
- NumOfProducts: The more products you have, generally the more engaged customers are and are less likely to churn.
- Income, Outstanding Loans, Debt-to-Income Ratio: These financial factors indicate the customer's economic stability; customers with certain financial situations may be at greater risk of churn.

# Predictive Modeling – Building the Model



# Model Evaluation & Cost-Sensitive Threshold Tuning

## Evaluation Metrics

- Accuracy (~97%), ROC AUC (~0.99), Precision, and Recall.
- Present the confusion matrix and classification report.



## Threshold Tuning

- Use a cost model (e.g.,  $\text{cost}_{fp} = 100$ ,  $\text{cost}_{fn} = 1000$ ) to adjust the decision threshold.
- Optimal threshold found: 0.21 (minimizes total cost).



## Impact

Higher recall ensures more high-risk customers are flagged for intervention.



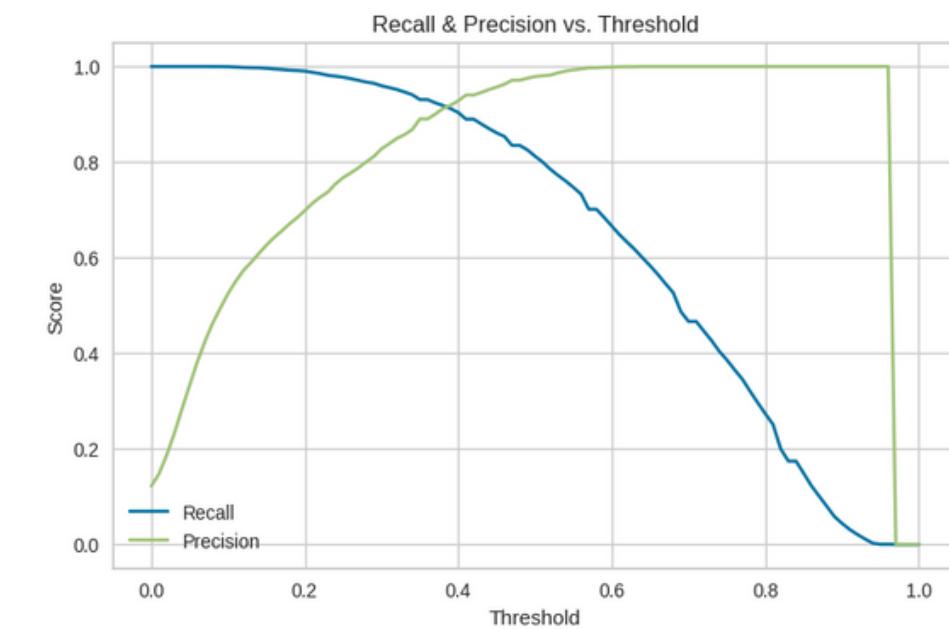
## Issue Identified

High precision but lower recall with default threshold (0.5) may miss many at-risk customers.



## Visual Examples

ROC curve, confusion matrix, and a plot of recall & precision vs. threshold.



# Customer Cluster-Based Retention Strategies for next year

| High-Risk | Moderate-Risk | Low-Risk

Cluster	Churn Risk (%)	Key Issue	Key Characteristic	Retention Strategy	How It Reduces Churn?
0	8.55% (Highest)	High Churn Risk, Financially Unstable	<ul style="list-style-type: none"> <li>Moderate balance (~\$73K)</li> <li>Low credit score (~503)</li> <li>High product adoption (avg. ~3.7 products)</li> <li>Moderate number of complaints (~2.89)</li> </ul>	<span style="color: green;">✓</span> Personalized financial coaching <span style="color: blue;">📊</span> Low-interest payment plans <span style="color: red;">🎯</span> Targeted loyalty incentives	<b>Helps struggling customers manage finances, reducing their likelihood of leaving due to financial hardship. Provides short-term relief while building long-term loyalty</b>
1	0.63% (Lowest)	Low Churn Risk, High Complaints	<ul style="list-style-type: none"> <li>Highest balance (~\$175K)</li> <li>High credit score (~644)</li> <li>Low product adoption (~2.3 products)</li> <li>High complaints (~7.1 complaints)</li> </ul>	<span style="color: green;">✓</span> Premium customer support <span style="color: blue;">💬</span> Faster response times <span style="color: orange;">🏆</span> Exclusive rewards for high-value customers	<b>Improves service experience for high-value customers, ensuring they don't churn due to frustration. Rewards reinforce loyalty, increasing retention.</b>
2	2.52% (Moderate)	Moderate Churn Risk, High Complaints	<ul style="list-style-type: none"> <li>Highest product adoption (~3.77 products)</li> <li>Moderate credit score (~463)</li> <li>High complaints (~7.19)</li> </ul>	<span style="color: green;">✓</span> Improve service touchpoints <span style="color: red;">👎</span> Proactive customer feedback mechanisms <span style="color: yellow;">💰</span> Personalized engagement offers	<b>Identifies and resolves pain points before they cause churn. Better service experiences and tailored offers keep customers satisfied.</b>
3	5.59% (High)	Lower Churn Risk, High Credit Score	<ul style="list-style-type: none"> <li>High credit score (~686)</li> <li>Low product adoption (~2.22 products)</li> <li>Low complaints (~2.79)</li> </ul>	<span style="color: green;">✓</span> Upselling premium services <span style="color: orange;">🚀</span> Personalized engagement <span style="color: blue;">✉️</span> Targeted email & app notifications	<b>Converts satisfied but inactive customers into loyal, high-value users. Keeps them engaged, reducing risk of switching to competitors.</b>

[Visit the Dashboard](#)

## Dashboard Churn Analysis | Bank Bostwana

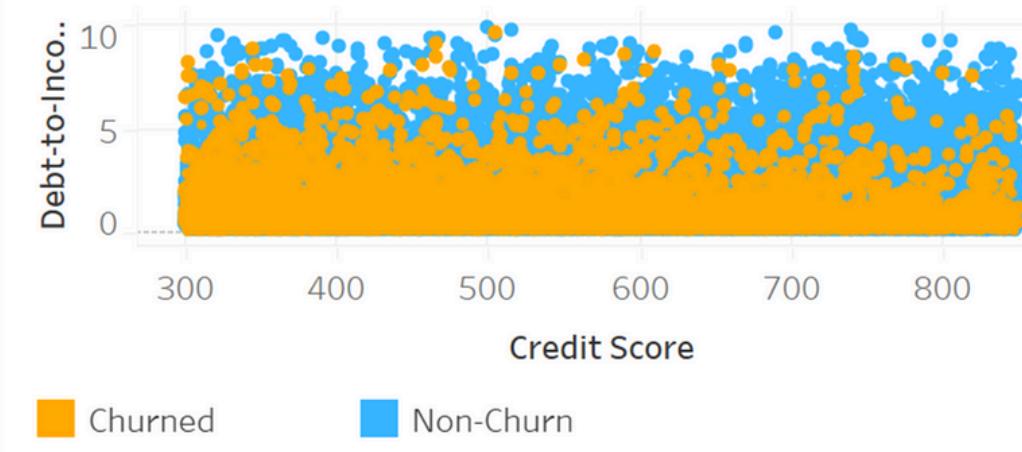
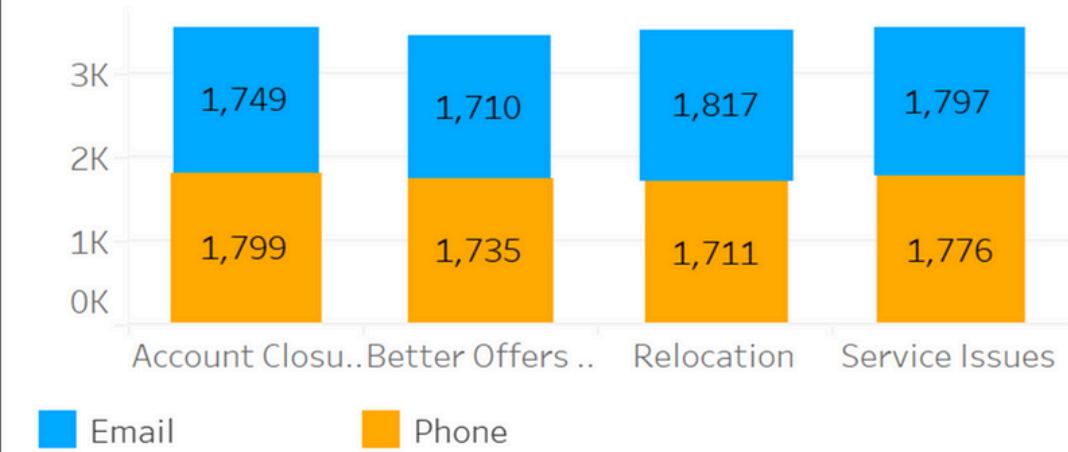
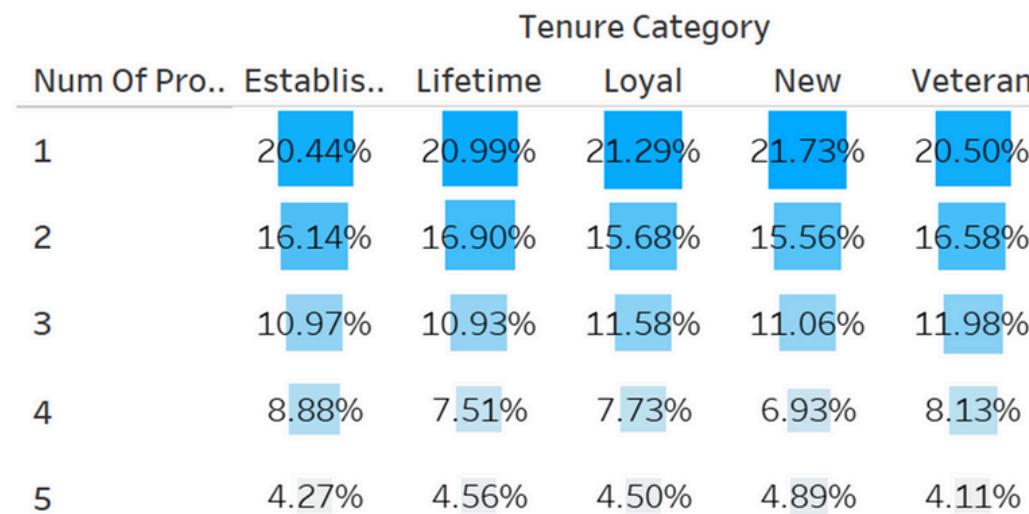
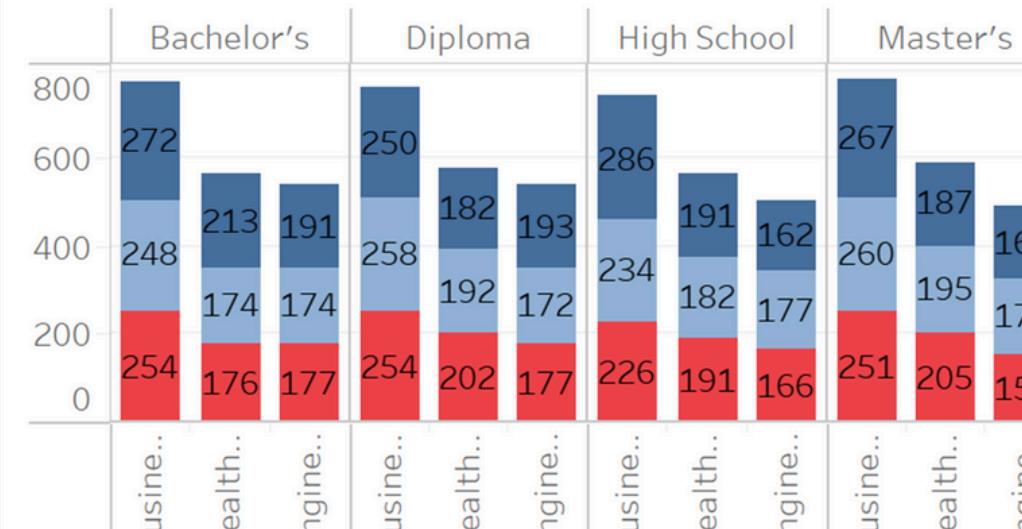
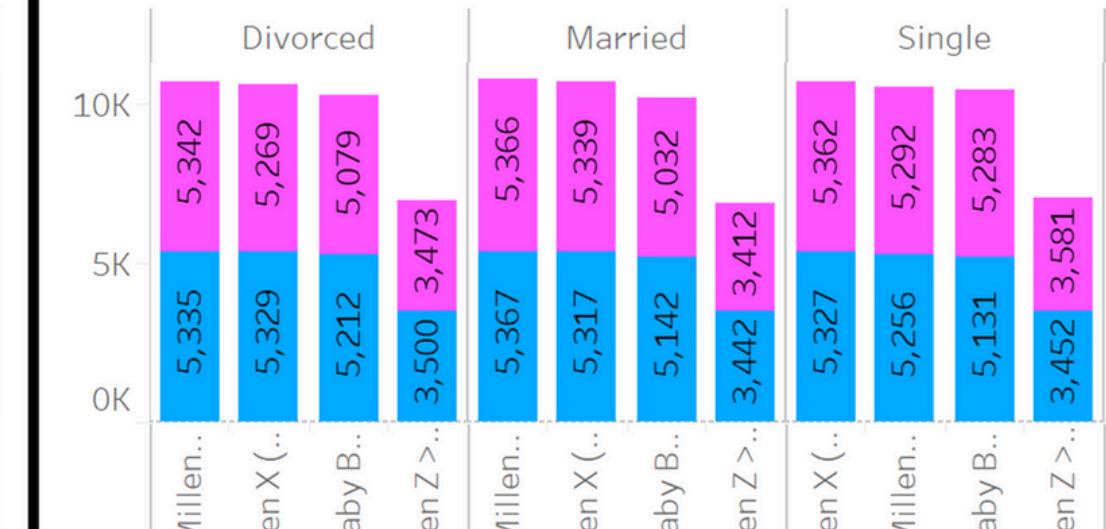
[Show Filters](#)
**Retention Rate**
**87.81%**
**Churn Rate**
**12.19%**
**Total Customers**
**115,640**
 (All)

 Non-Churn

 Churned

**Geographic Distribution of Churn**


10.57% 13.77%

**Churn by Financial Factors**

**Churn Reasons and Communication Channel**

**Product and Tenure Analysis**

**Top 3 Combined Customer Segmentation**

**Churn by Demographics**


**THANK YOU**  
FOR YOUR NICE ATTENTION

# APPENDIX

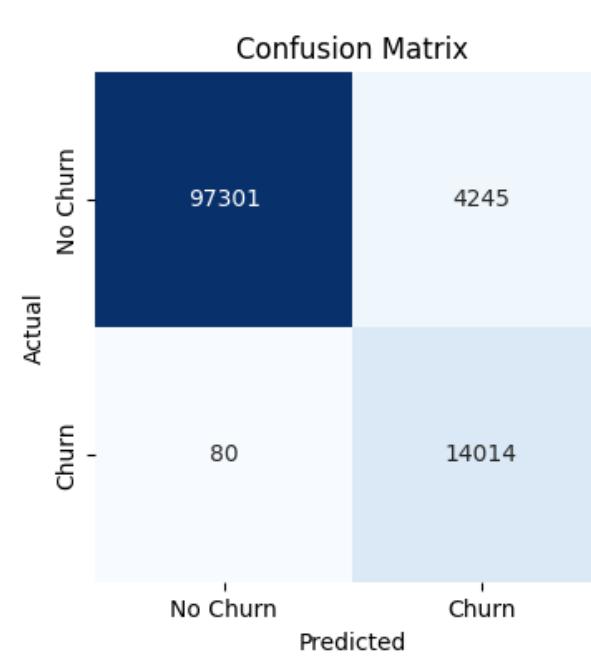
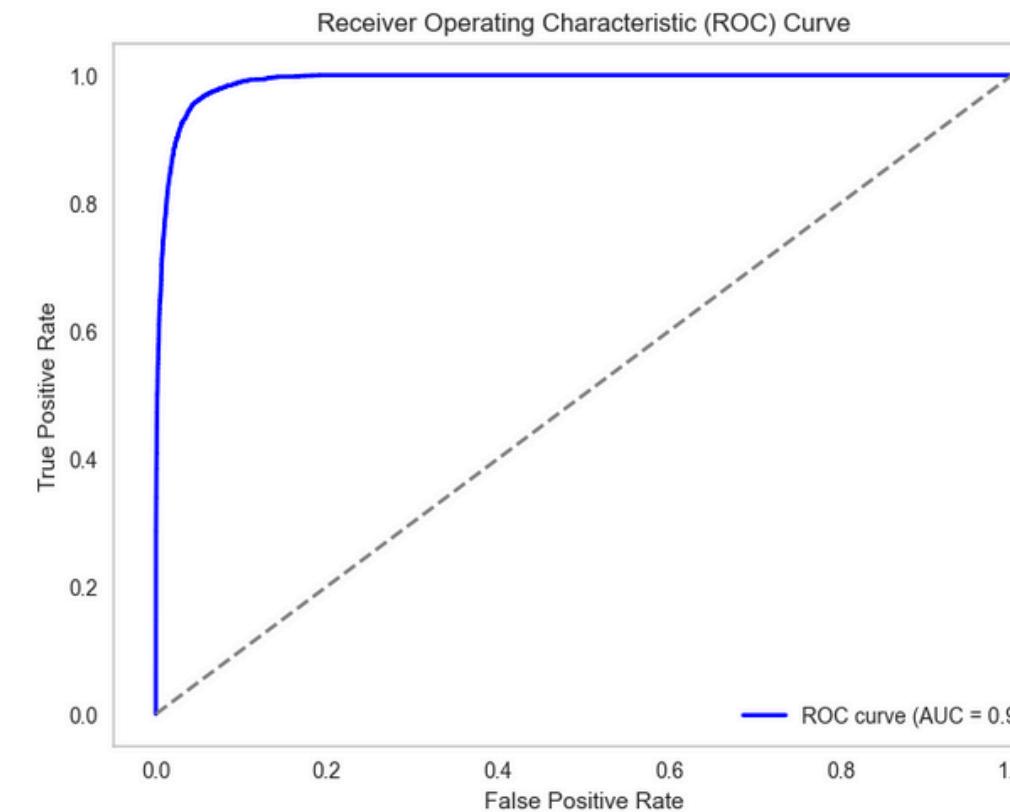
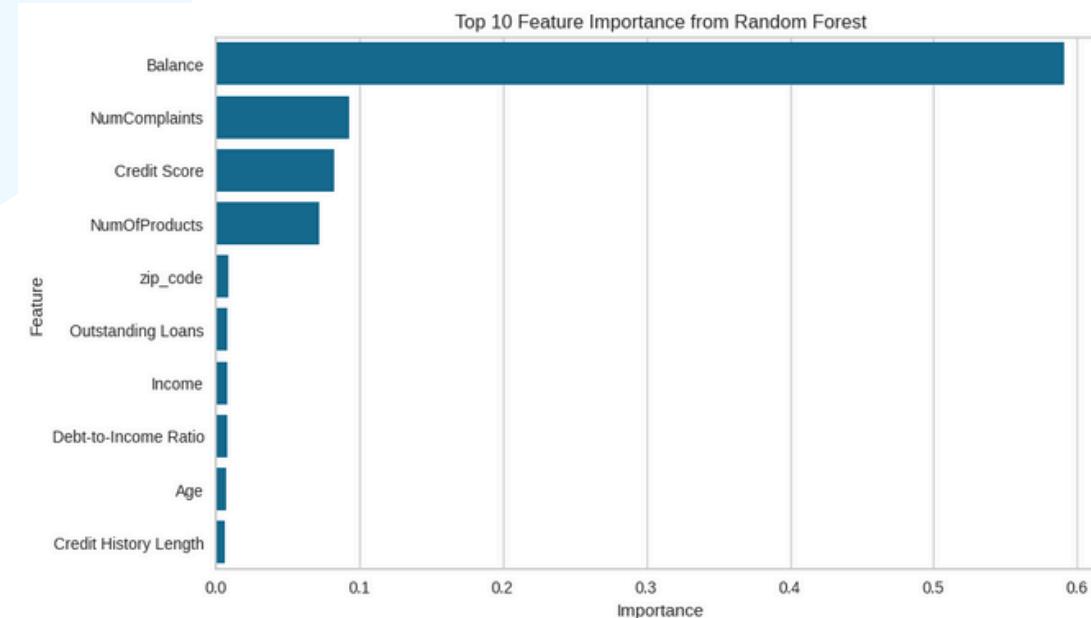
Dataset : **KAGGLE**

Python Code : **Google Colab**

Data Visualization : **Tableau**

# Random Forest Test Result - ROC AUC Curve

```
RandomForestClassifier
RandomForestClassifier(class_weight='balanced', max_depth=30, n_estimators=200,
random_state=42)
```



Predict on the test data

Optimized Model Performance (Data Test):

- Accuracy (Data Test): 0.952438602559668
- Precision (Data Test): 0.7293835068054444
- Recall (Data Test): 0.9694927279177014
- F1 Score (Data Test): 0.8324703015534572

Classification Report (Data Test):

	precision	recall	f1-score	support
0	1.00	0.95	0.97	20309
1	0.73	0.97	0.83	2819
accuracy			0.95	23128
macro avg	0.86	0.96	0.90	23128
weighted avg	0.96	0.95	0.96	23128

Predict on the train data

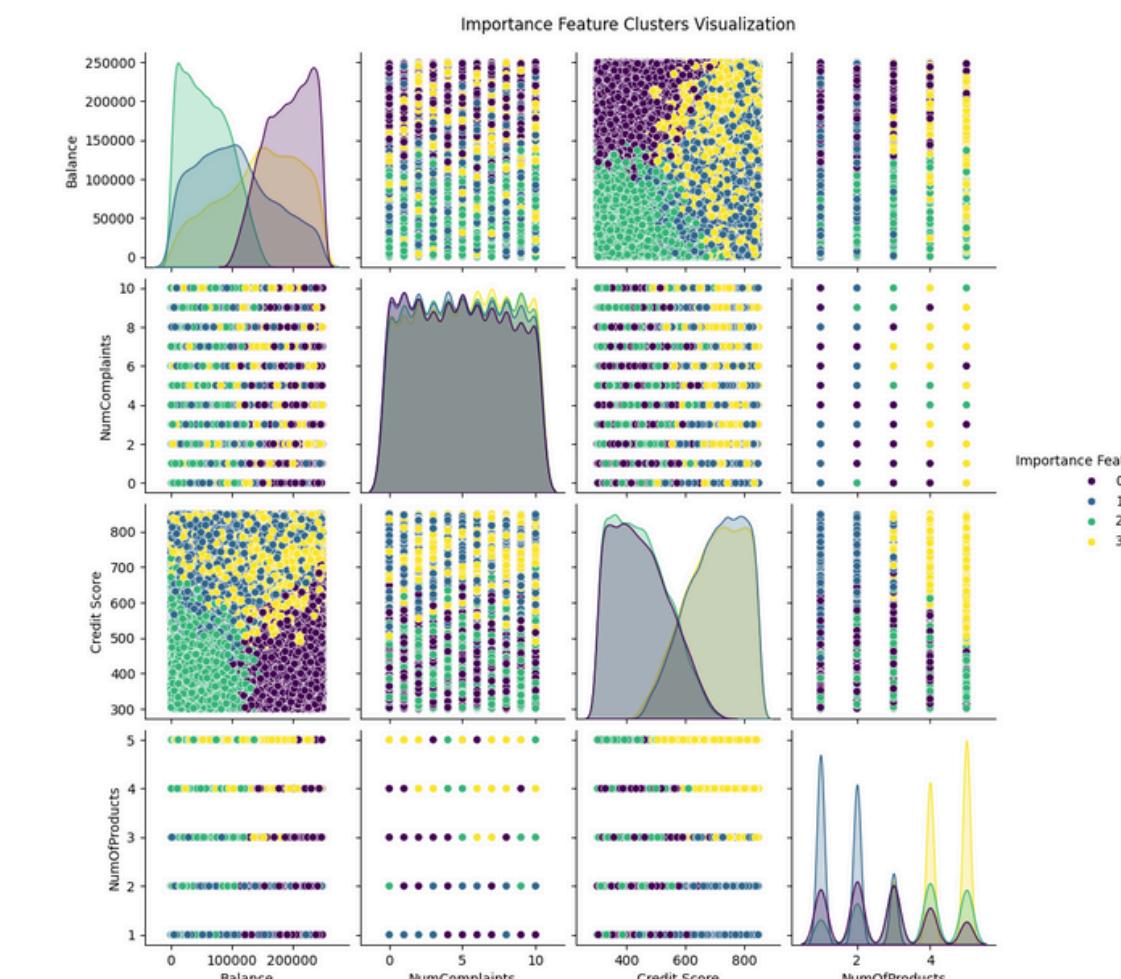
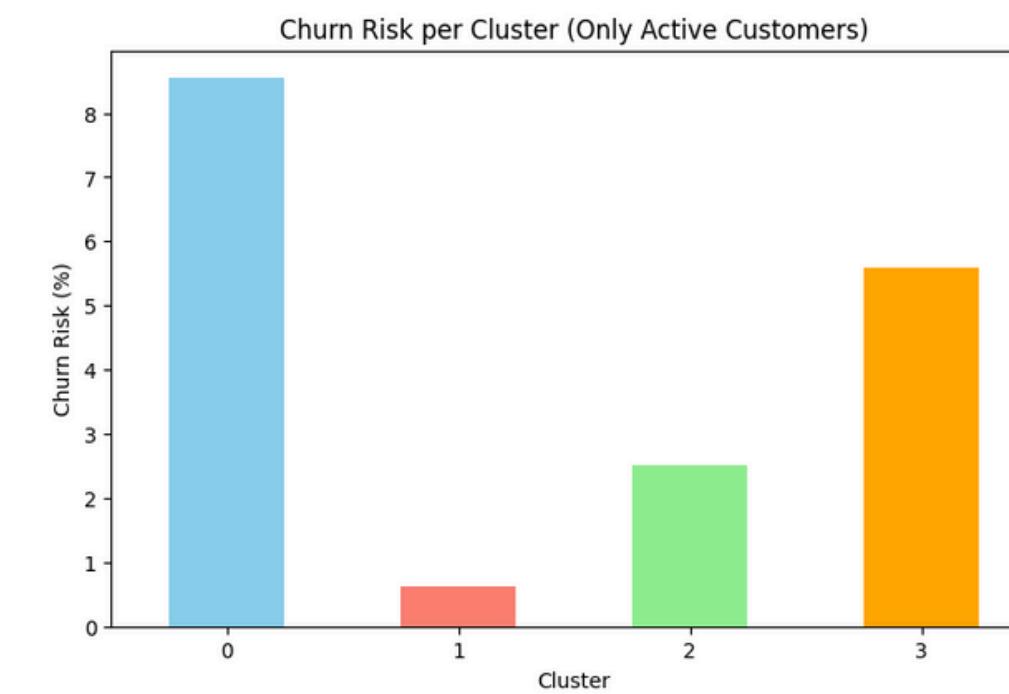
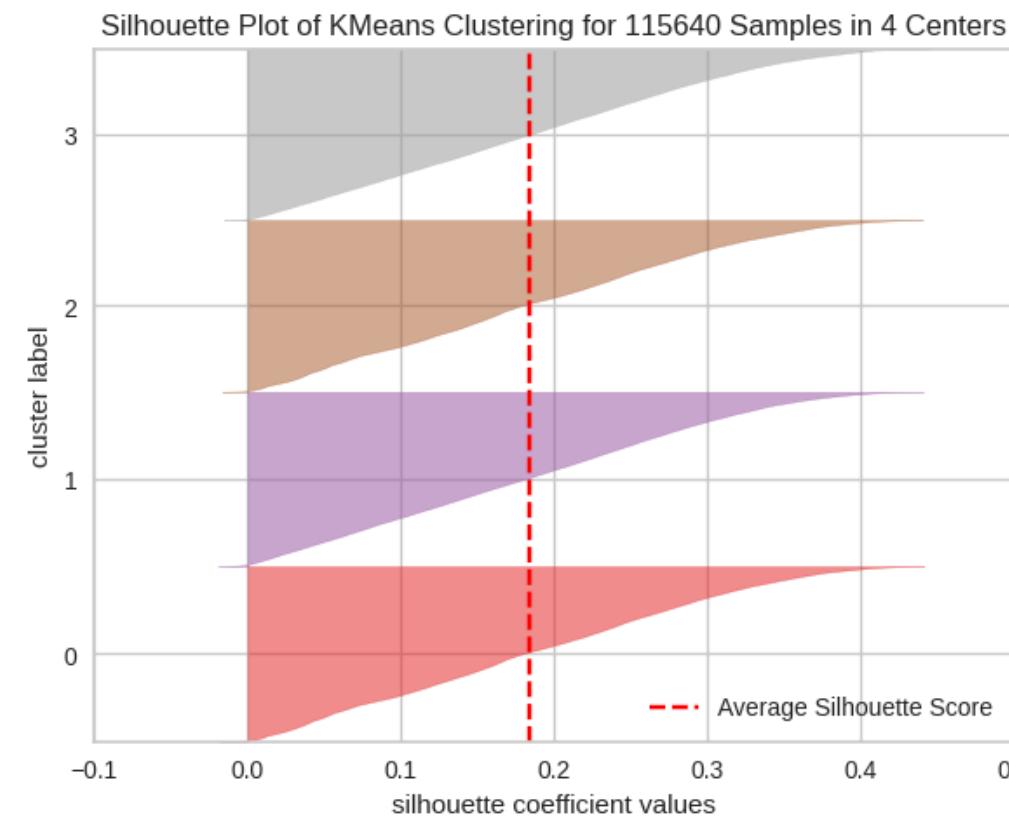
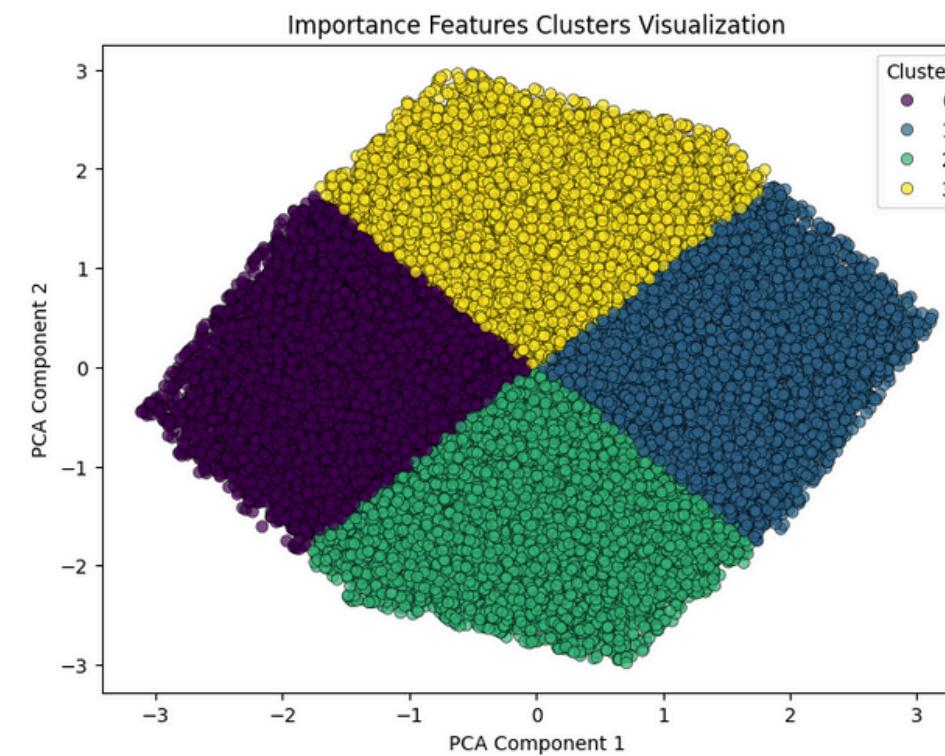
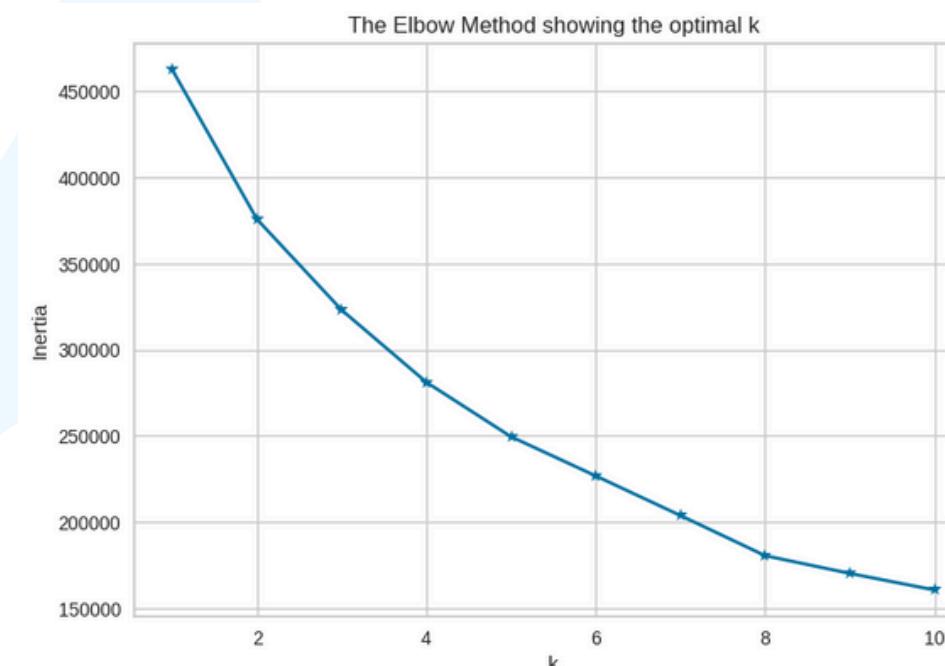
Optimized Model Performance (Data Train):

- Accuracy (Data Train): 0.9642316672431684
- Precision (Data Train): 0.7735952740761094
- Recall (Data Train): 0.9988470066518848
- F1 Score (Data Train): 0.8719080246196725

Classification Report (Data Train):

	precision	recall	f1-score	support
0	1.00	0.96	0.98	81237
1	0.77	1.00	0.87	11275
accuracy			0.96	92512
macro avg	0.89	0.98	0.93	92512
weighted avg	0.97	0.96	0.97	92512

# K-Means Clustering



Predicted Churn		0	1
Importance Features Cluster PCA			
0	21066	1969	
1	26909	170	
2	24389	630	
3	24937	1476	

