



**Mitigating Customer Churn:
Leveraging Historical Data Analysis and Predictive Modelling to
Formulate Effective Customer Retention Strategies**

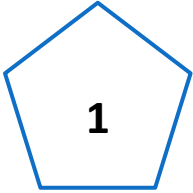
Alester D Costa

Background

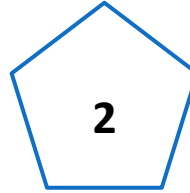
- Methods of customer retention:
 - Creating a solid relationship with customers
 - Offering exceptional services to customers
 - Providing value for customers according to their behaviors and preferences
- In 2020, the financial industry in the United States had a 25% churn rate, the joint-highest churn rate, along with the cable industry, in all industries
- Nova Apex Bank is facing customer churn within its European operations in France, Germany and Spain
 - Revenue loss
 - Higher customer acquisition costs
 - Damage to corporate reputation



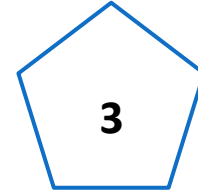
Topics of Discussion



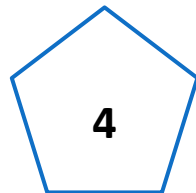
Project Objectives



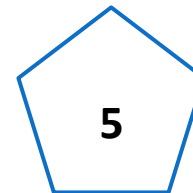
Data Source



Data Exploration



Model Exploration



Conclusion



Nova Apex Bank

Project Objectives



Project Objectives



Interactive Dashboard

Develop and implement a customer churn analysis dashboard to aggregate and visualize customer behavior through historical data analysis

Predictive Modelling

Build multiple models to predict customer churn using key customer attributes from historical customer data.

Model Selection

Determine the most reliable predictive model that predicts customer churn using different evaluation metrics.

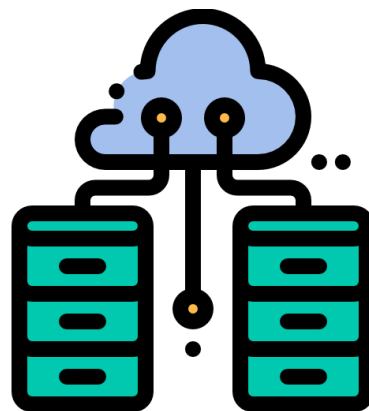
Actionable Insights

Identify key customer churn factors from the most effective predictive model, supported by historical data analysis, to develop targeted customer retention strategies.



Nova Apex Bank

Data Source



Historical Customer Data

10,000

Customer Records

12

Predictors

Exited

Target Variable

Demographics Predictors

Gender	Age
Categorized Age	Geography

Financial Information Predictors

Balance	Estimated Salary
Credit Score	Categorized Credit Score
Number of Products Held	Credit Card Holder
Tenure	Activeness as a Member

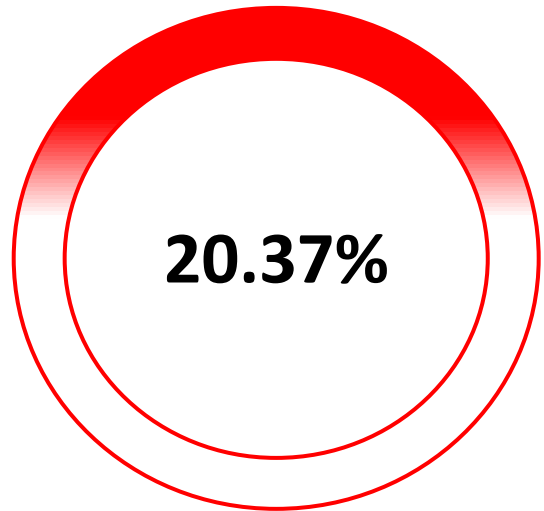


Data Exploration



Nova Apex Bank

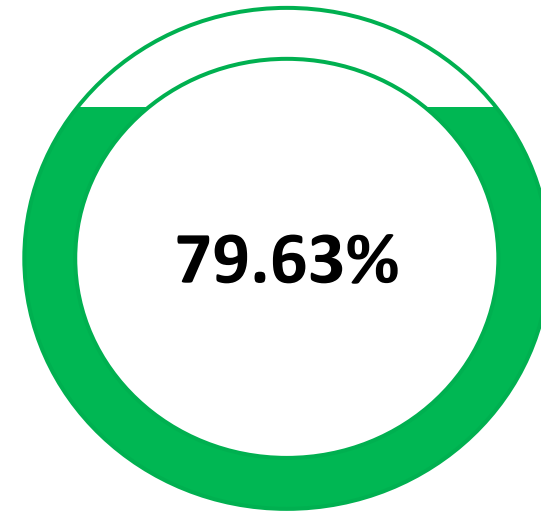
Breakdown of Customer Churn



Exited = 'Yes'

Churned Customers

10,000 Customers



Exited = 'No'

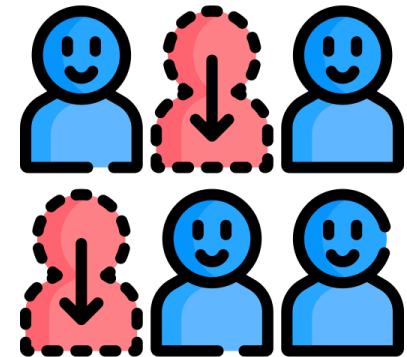
Non-Churned Customers



Nova Apex Bank

Customer Churn Analysis Dashboard

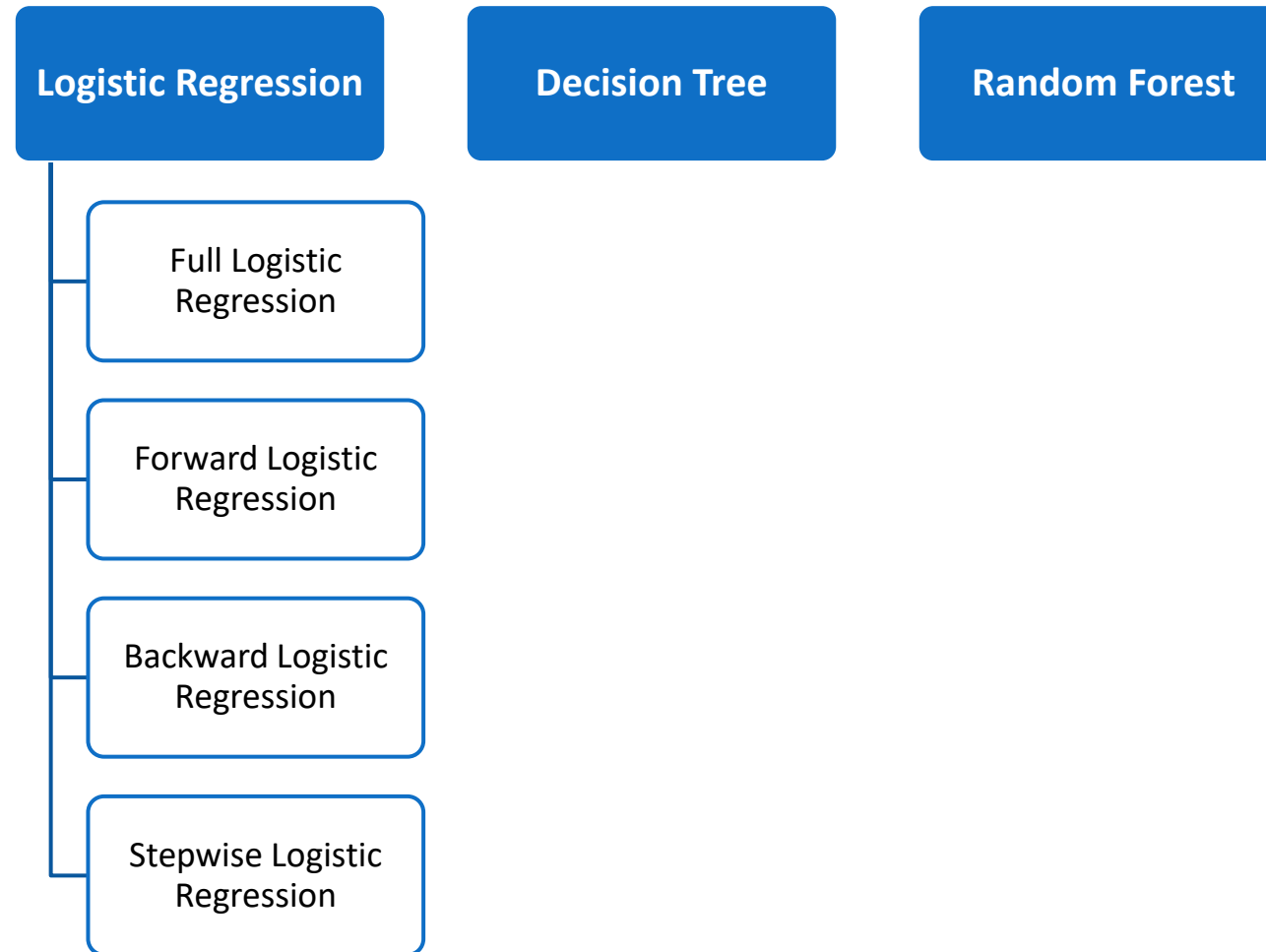
- All customer attributes on demographics and financial information aggregated and visualized at one location
- Use of churn filter on the dashboard:
 - Understand the entire customer base, or
 - Understand about a specific segment, either churned or non-churned customers
- Benefits of the dashboard:
 - Dynamic aggregations and visualizations
 - Real-time monitoring of customer churn
 - Simplification of historical data analysis



Model Exploration



Predictive Models Developed



Evaluation Metrics of Predictive Models

Receiver Operating Characteristic – Area Under the Curve (ROC-AUC)

Assesses the effectiveness of each predictive model to distinguish between customers likely to churn and those not likely to churn

Accuracy

Calculates the proportion of correct predictions (both churn and non-churn) out of the total predictions

F1-Score

Balances precision (preciseness of customer churn predictions) and recall (capacity to discover actual churn cases) into a single value



Comparison of Predictive Models

Model	ROC-AUC	Accuracy	F1-score (Exited = "Yes")
Random Forest	0.86	0.86	0.57
Decision Tree	0.85	0.86	0.58
Backward Logistic Regression	0.79	0.83	0.45
Full Logistic Regression	0.79	0.82	0.44
Stepwise Logistic Regression	0.76	0.83	0.42
Forward Logistic Regression	0.76	0.82	0.42



Confusion Matrix of the Random Forest Model

Actual Churn	Predicted Churn	
	Predicted to Stay	Predicted to Leave
Actually Stayed	2,300	73
Actually Left	346	281

 Correct Predictions
 Incorrect Predictions

Precision (Exited = "Yes")

79% of customers predicted to churn actually churned

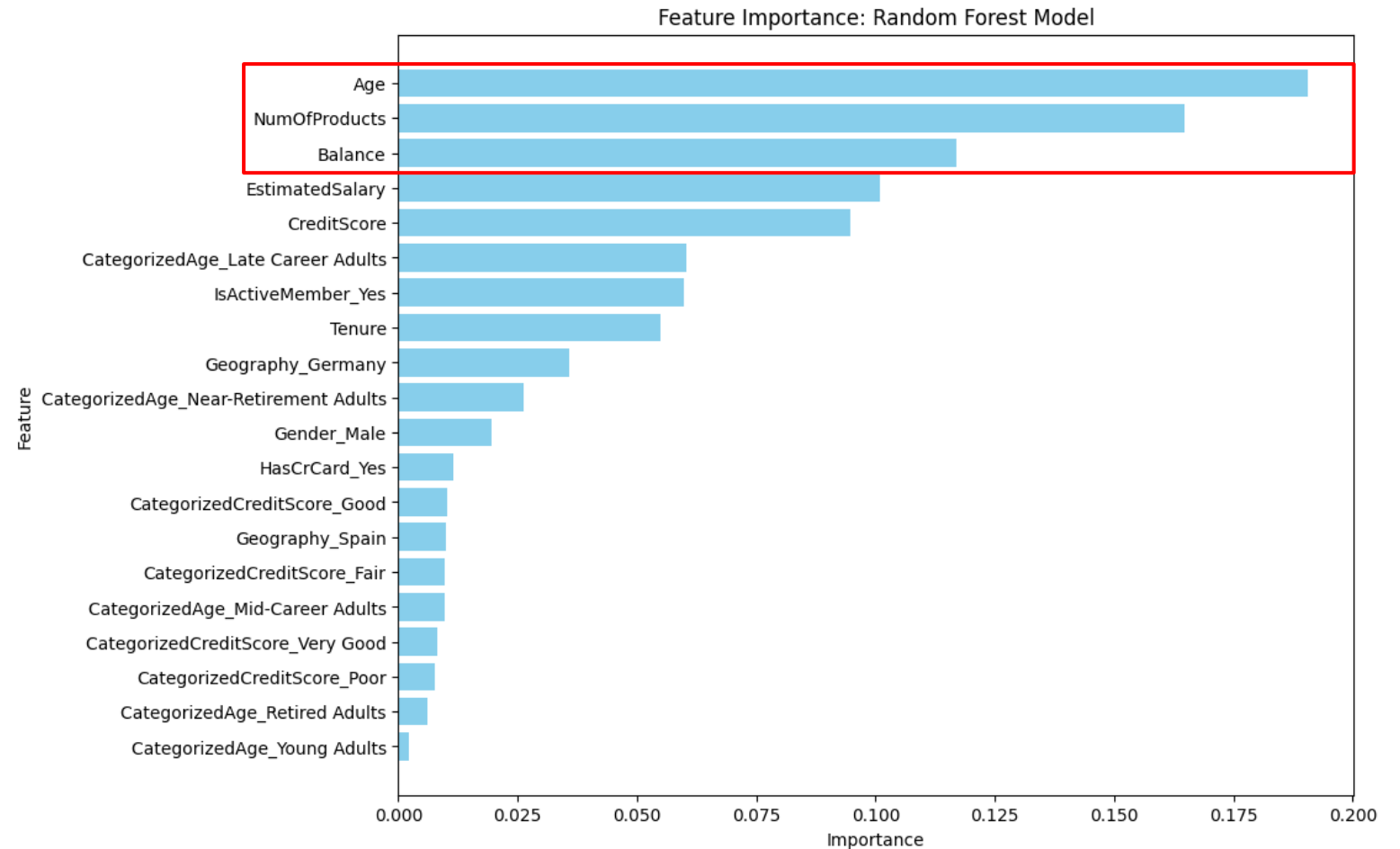
Recall (Exited = "Yes")

45% of customers who actually churned have been correctly identified



Feature Importance of the Random Forest Model

- Top three features influencing customer churn:
 - Age of customers
 - Number of products held by customers
 - Balance of customers in their accounts



Conclusion



Mitigating Customer Churn: Customer Retention Strategies

- **Tailored offerings for customers of different age groups**
 - Younger customers
 - Targeted marketing campaigns on technology-driven banking solutions
 - Older customers
 - Stability-focused services and personalized guidance
- **Cross-selling through bundling for customers with limited products**
 - Encourage customers to obtain multiple products by offering bundled packages with appealing discounts
- **Exclusive benefits for customers with higher balances**
 - Provide targeted incentive programs to enhance engagement



Thank You!



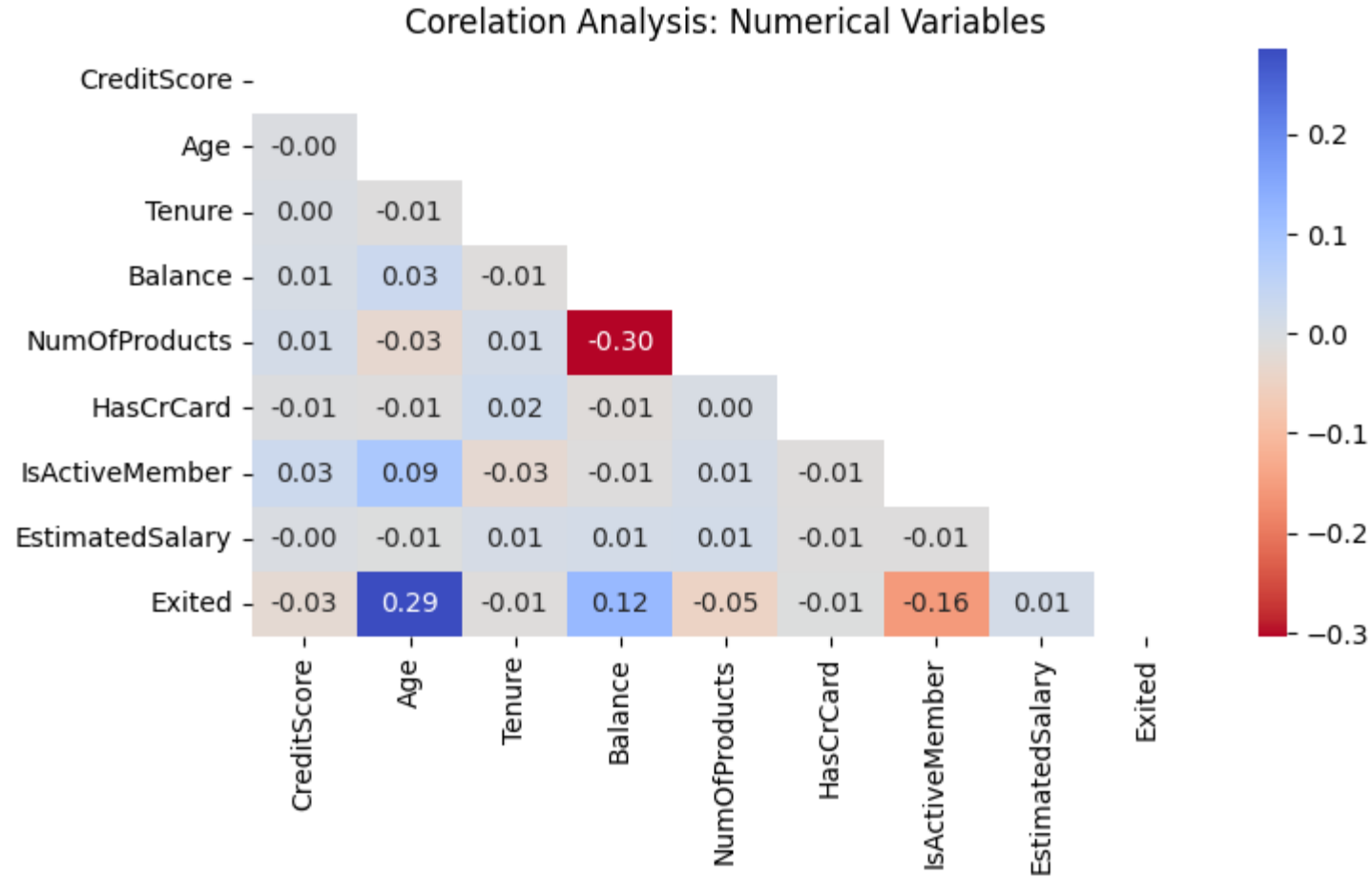
References

- Glassbox. (2023, September 27). *Customer retention for banks: 5 ways to boost loyalty and lifetime value*. <https://www.glassbox.com/blog/customer-retention-in-banking>
- Investopedia. (2024, March 21). *Churn Rate: What It Means, Examples, and Calculations*. <https://www.investopedia.com/terms/c/churnrate.asp>
- Mihup (2024, February 15). *Maximizing Retention: The Power of Bank Customer Churn Prediction*. <https://mihup.ai/maximizing-retention-bank-customer-churn-prediction>
- Rodgers, E. (2023, October 24). Banking Customer Retention Strategies to Implement in 2024. *Drive Research*. <https://www.driveresearch.com/market-research-company-blog/banking-customer-retention-strategies-that-work-in-2023>
- Statista. (2022, July 6). *Customer churn rate in the United States in 2020, by industry*. <https://www.statista.com/statistics/816735/customer-churn-rate-by-industry-us>

Appendix



Correlation Analysis: Numerical Variables



Customer Churn Analysis Dashboard



Customer Churn Analysis Dashboard: Aggregations



10,000

Total Customers



2,037

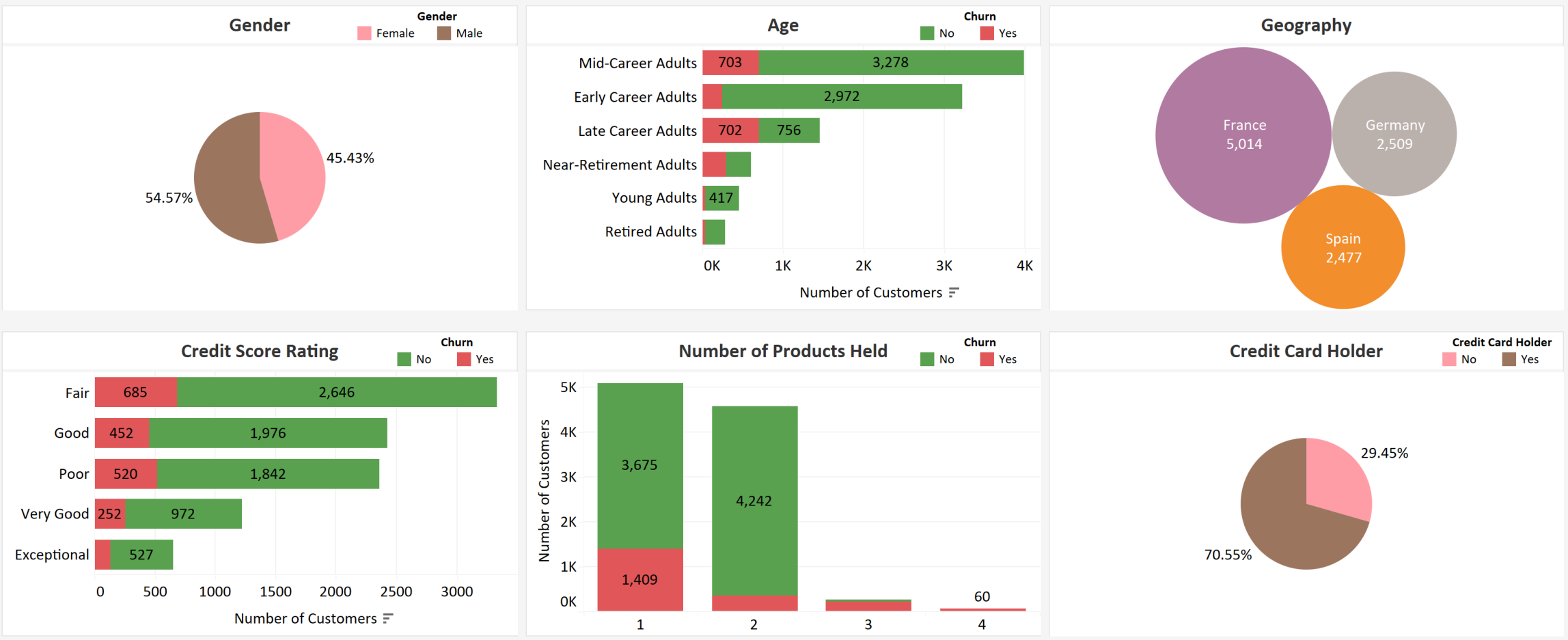
Churned Customers



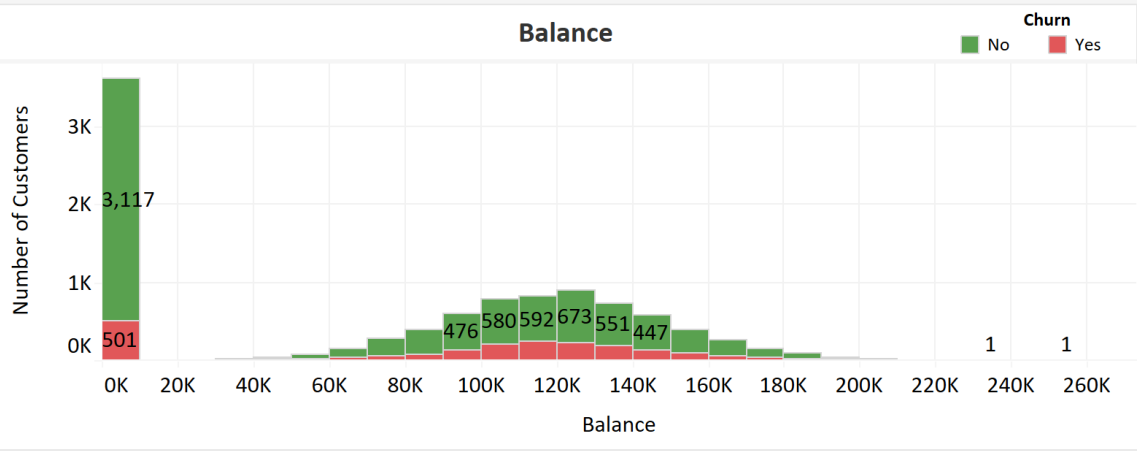
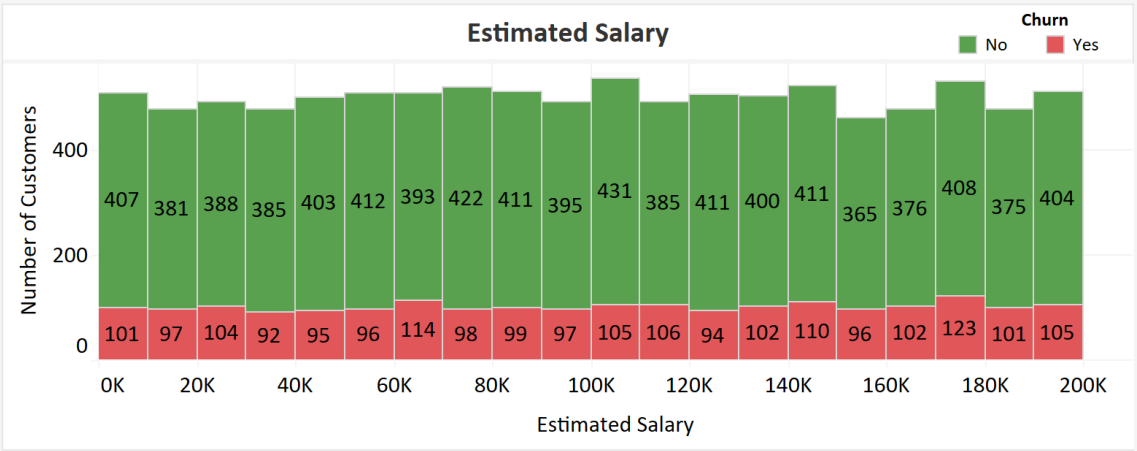
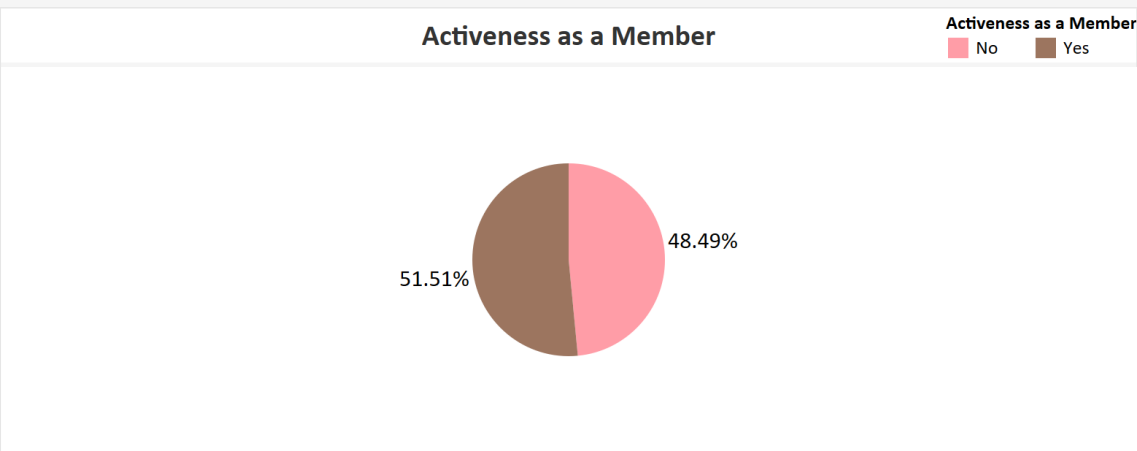
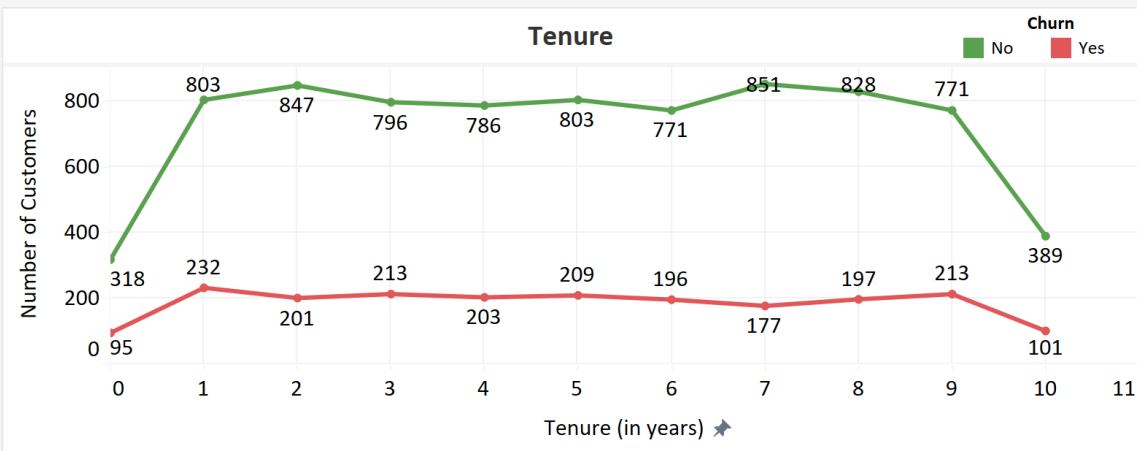
20.37%

Churn Rate

Customer Churn Analysis Dashboard: Visualizations



Customer Churn Analysis Dashboard: Visualizations



Feature Importance of the Random Forest Model: Gini Impurity Values

	Feature	Importance (Gini Impurity)
0	Age	0.190698
4	NumOfProducts	0.164777
1	Balance	0.117053
2	EstimatedSalary	0.100888
3	CreditScore	0.094838
9	CategorizedAge_Late Career Adults	0.060458
19	IsActiveMember_Yes	0.059714
5	Tenure	0.054995
7	Geography_Germany	0.035899
11	CategorizedAge_Near-Retirement Adults	0.026268
6	Gender_Male	0.019433
18	HasCrCard_Yes	0.011651
15	CategorizedCreditScore_Good	0.010125
8	Geography_Spain	0.010078
14	CategorizedCreditScore_Fair	0.009665
10	CategorizedAge_Mid-Career Adults	0.009597
17	CategorizedCreditScore_Very Good	0.008060
16	CategorizedCreditScore_Poor	0.007543
12	CategorizedAge_Retired Adults	0.006026
13	CategorizedAge_Young Adults	0.002232