CUSTOMER CHURN ANALYSIS & PREDICTION

SANKA NAGA NITESH
DATA SCIENTIST, BANK XYZ

- INTRODUCTION
- BUSINESS QUESTIONs
- EXPLORATORY DATA ANALYSIS
- PREDICTIVE ANALYTICS
- BUSINESS ACTION
- BUSINESS OUTCOME
- OUTRO

BUSINESS QUESTIONS?

- What are the key factors driving customer churn in our bank?
- Can we predict which customers are likely to churn in the near future?
- How can we proactively retain customers who are at high risk of churning?
- What is the potential impact of customer churn on our bank's financial performance, and how can we mitigate this risk?

INTRODUCTION

WHAT IS CHURN IN BANKING?

Defined as movement of customer from one company to another.

For e.g. if our company gained 1000 customers and lost 32 customers in the last month, then our monthly Customer Churn Rate would be (32/1000)*100% = 3.2%.

REASONS

- Customer-friendly bank staff
- Low Interest Rates
- Location
- Services offered
- Quality
- Price
- Needs Change
- Functionality

SIGNIFICANCE

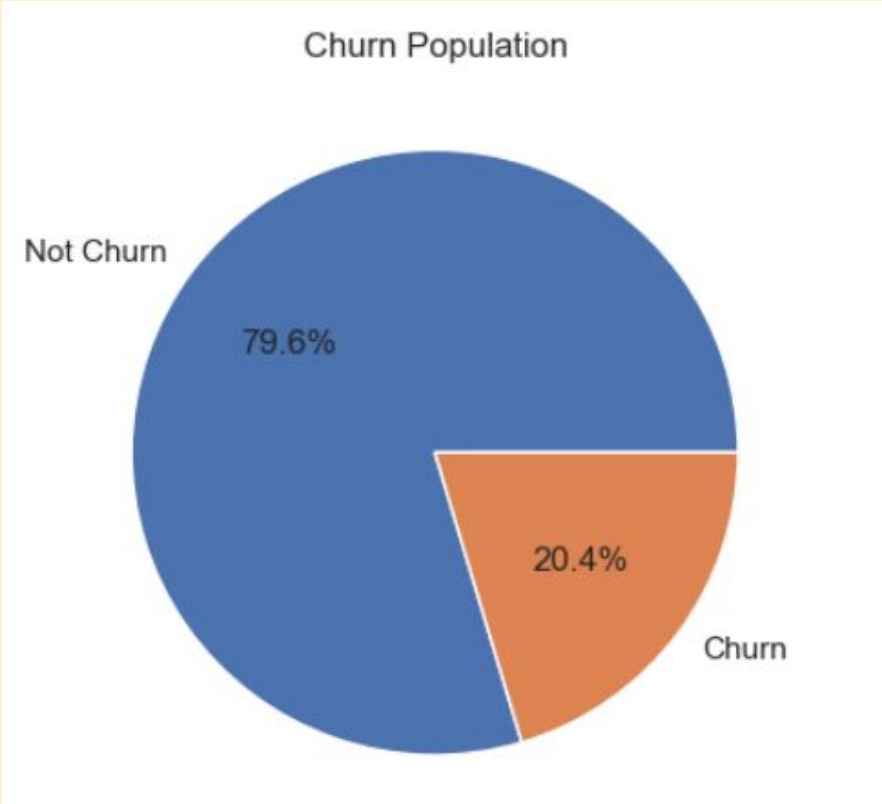
- Impacts Profitability
- Increases Customer Satisfaction
- Improves Product and services based on customer feedback
- Better understanding of Customer Behavior
- Eats into Your Bottomline
- Lowers Customer Lifetime Value
- Acquisition Costs are high

EXPLORATORY DATA ANALYSIS

TARGET VARIABLE – EXITED

80%

Retention

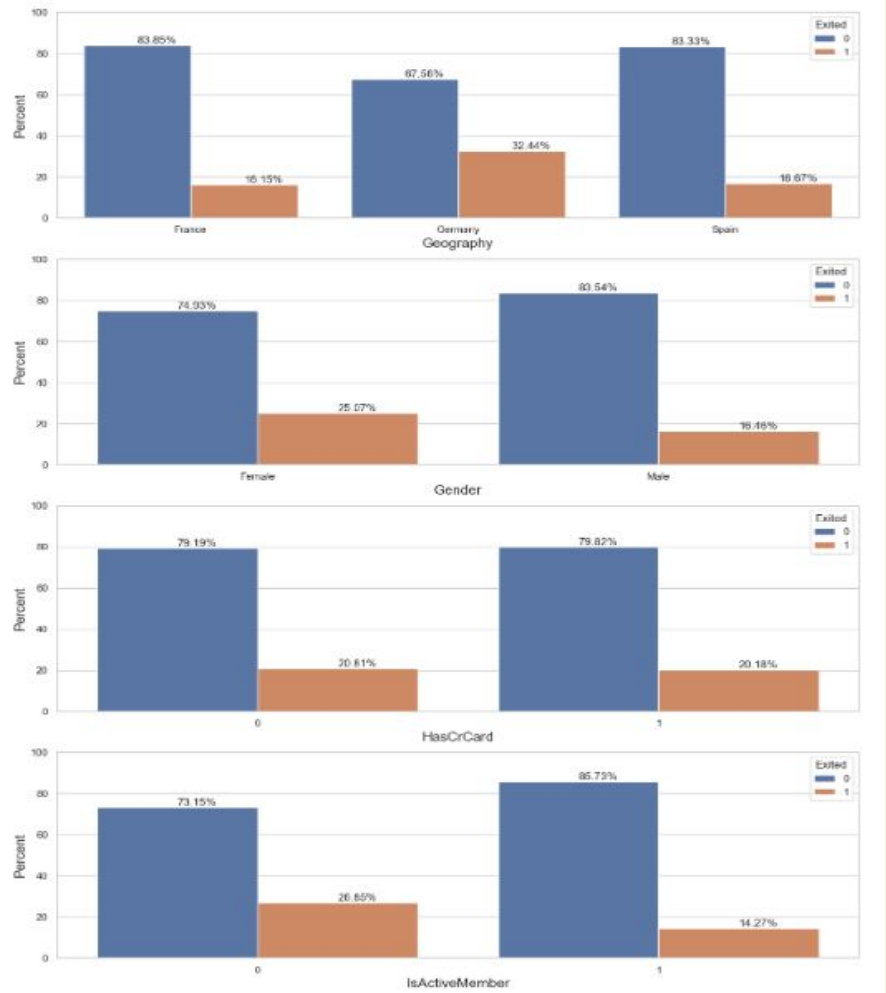


20%

Churn

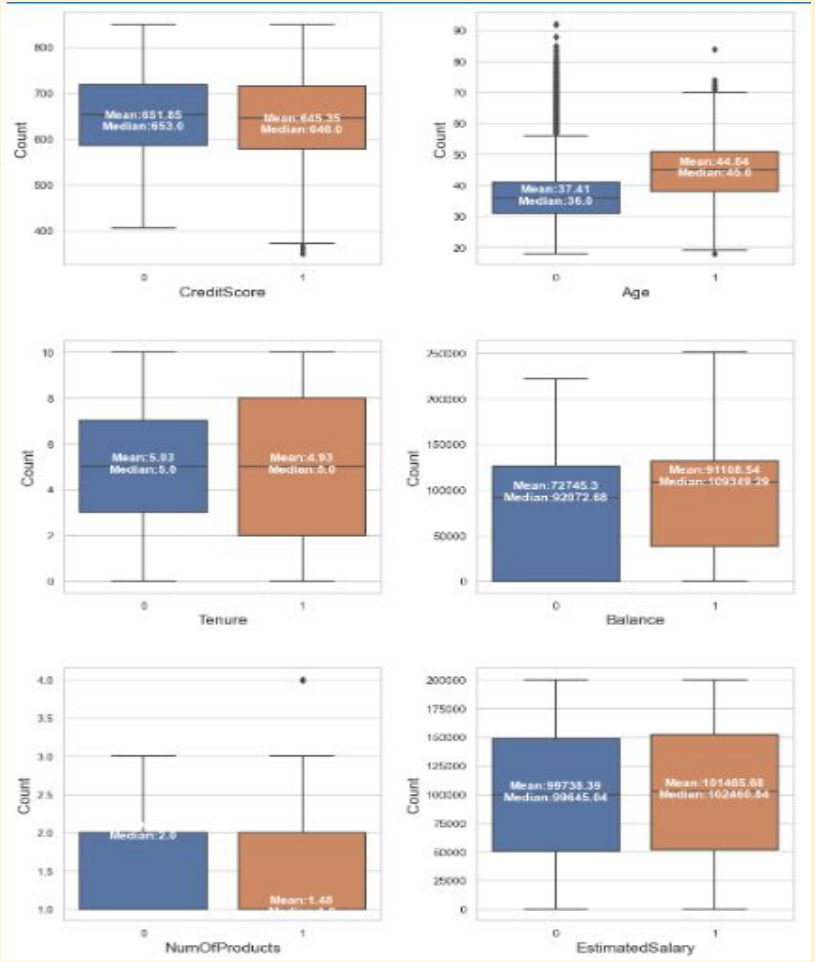
CATEGORICAL VARIABLES

- France has least churn rate and germany has higher
- Proportion of female customers churning is slightly higher than that of male customers
- Majority of customers that churned are those with credit card. Could be a coincidence
- Inactive members have greater churn
- Overall proportion of inactive members is high, bank can start programs to turn these users into active customers

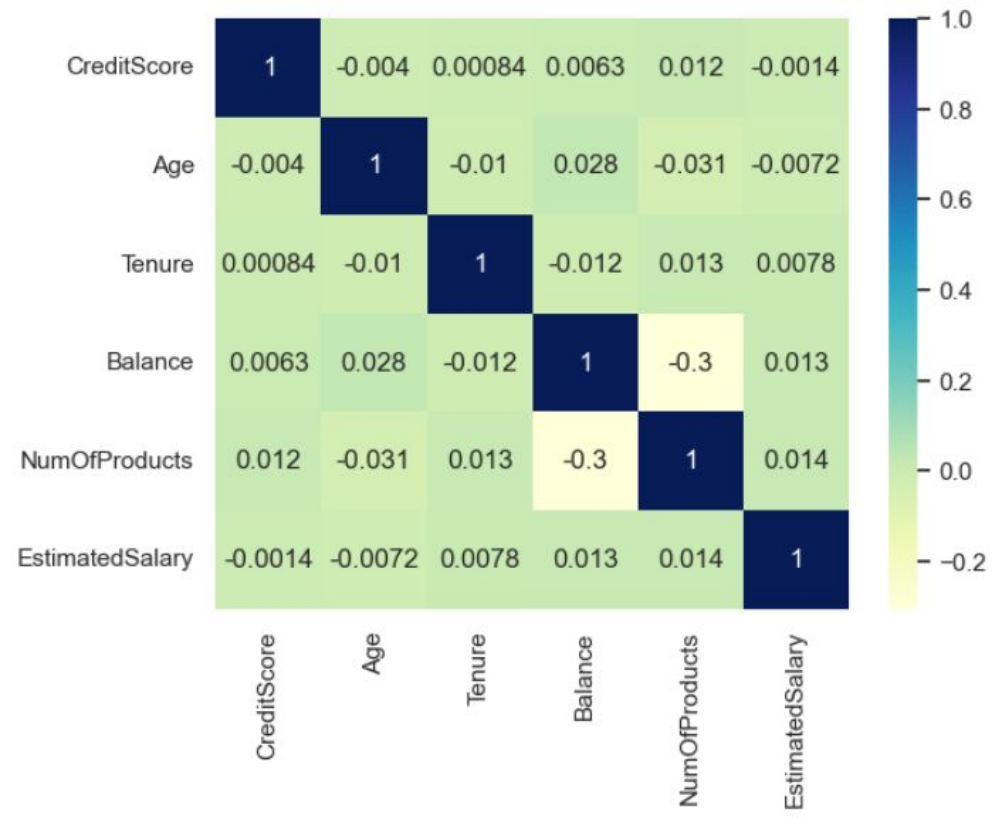


NUMERICAL VARIABLES

- Neither product nor salary has an impactful effect on the likelihood to churn
- No significant difference in the credit score distribution between retained and churned
- Tenure is an important factor in predicting churn, with customers who have spent less or more time with the bank being more likely to churn. The highest churn rates were observed in years 2, 4, and 10 out of the 11 years examined
- Churned customers have a high average bank balance of $91,000, which could impact the bank's lending capital. The overall average bank balance for churned customers is $101,465



CORRELATION MATRIX



The correlation coefficient map indicates that there are no significant correlations between the numerical variables, which means that multicollinearity is not a concern.

PREDICTIVE ANALYTICS

MODELS USED...

- **Logistic Regression**: A statistical method used to analyze the relationship between a binary dependent variable and one or more independent variables by estimating probabilities using a logistic function.
- **Support Vector Machine**: A supervised machine learning algorithm used for classification or regression analysis that finds the optimal hyperplane (decision boundary) to separate classes by maximizing the margin between them.
- **Random Forest**: A machine learning algorithm that uses multiple decision trees to make predictions by aggregating the outputs of individual trees, reducing the risk of overfitting and increasing accuracy.
- **XGBoost**: A gradient boosting machine learning algorithm that uses a combination of decision trees and gradient descent optimization to achieve high predictive accuracy in classification and regression tasks. It's known for its speed and scalability in handling large datasets.

PERFORMANCE METRICS MEASURED

- **Accuracy**: Ratio of correct predictions to the total number of predictions made by the model.
- **Precision**: Ratio of true positive predictions to the total number of positive predictions made by the model.
- **Recall**: Ratio of true positive predictions to the total number of actual positive cases in the data.
- **F1 score**: Harmonic mean of precision and recall, which gives a balance between the two metrics.
- **ROC AUC** (Area Under the Receiver Operating Characteristic Curve): Measure of how well a model can distinguish between two classes, with a value of 1 indicating perfect separation and 0.5 indicating random guessing.

MODEL PERFORMANCE

	LOGISTIC REGRESSION	SVM	RANDOM FOREST	XGBOOST
ACCURACY	0.80	0.86	0.86	0.85
PRECISION	0.59	0.83	0.77	0.70
RECALL	**0.19**	**0.40**	**0.45**	**0.48**
F1 SCORE	0.28	0.54	0.57	0.57
ROC AUC	0.57	0.69	0.71	0.71

WHICH MODEL TO PICK???

Most important metric used is **RECALL** in this case. It reveals the proportion of churn identified correctly by the total number of churns.

Based on this, we will choose **XGBoost Model** because

- High Accuracy
- Ability to handle missing values
- Feature Importance
- Regularization
- Speed
- Scalability
- Early customer intervention

WHAT ARE THE SIGNIFICANT VARIABLES AFFECTING CUSTOMER CHURN???

- CreditScore
- Age
- Balance
- NumOfProducts and
- IsActiveMember

BUSINESS ACTION

To prevent churn of bank customers, the following business actions could be taken:

- **Improve customer service:** Provide exceptional customer service, improve customer satisfaction and reduce churn.
- **Offer incentives**: Offering incentives such as loyalty programs, discounts, and personalized offers can improve customer retention.
- **Improve product offerings**: Providing better and more diverse product offerings can improve customer engagement and satisfaction, reducing the likelihood of churn.
- **Proactive customer outreach**: Proactive outreach to customers who may be at risk of churning can help identify and address concerns before they lead to customer loss.
- **Monitor customer behavior**: Monitoring customer behavior can help identify patterns of churn and develop strategies to mitigate the risk of customer loss.
- **Streamline customer experience**: Simplifying the customer experience and reducing friction points can improve satisfaction and reduce the likelihood of churn.
- **Personalize communication**: Tailoring communication and offers to individual customers based on their preferences and behavior can improve engagement and satisfaction, reducing churn.

BUSINESS OUTCOME

Retaining Customers: By identifying customers who are likely to churn, banks can take proactive measures to retain these customers, such as offering personalized incentives, improved customer service, or new product offerings.

Cost Reduction: Customer churn can be an expensive problem for banks, as it can result in lost revenue, increased marketing costs, and other expenses associated with acquiring new customers. By identifying and preventing churn, banks can save significant amounts of money.

Improved Customer Experience: By proactively addressing issues that are causing customers to consider churning, banks can improve overall customer satisfaction and loyalty. This can lead to increased engagement and higher lifetime customer value.

Competitive Advantage: Banks that are able to predict and prevent customer churn can gain a competitive advantage in the marketplace by offering superior customer service and tailored products and services that meet the unique needs of their customers.

OUTRO

Banking customer churn prediction model is a valuable tool for financial institutions like banks like ours to proactively identify and address customer churn.

By leveraging machine learning algorithms like XGBoost (in this case) and evaluating performance metrics such as accuracy, precision, recall, F1 score, and ROC AUC, businesses can make data-driven decisions to retain customers and maintain a stable customer base.

Ultimately, the implementation of a churn prediction model can have a significant impact on the financial health and stability of the organization.

THANKS