# Predicting Customer Churn for Improved Retention and Business Growth

- Hamza Ali

# Business Goals

The goal is to help businesses predict and prevent customer churn by using data insights. By accurately identifying customers at risk of leaving, businesses can implement targeted retention strategies, improving customer loyalty and reducing churn rates.

The business goals are to:

- **Predict customer churn with high accuracy**
- **Identify at-risk customers in a timely manner to enable proactive retention strategies**
- **Increase customer retention and lifetime value**

By using data analysis and machine learning modeling techniques, the aim is to provide businesses with actionable insights to improve customer engagement, reduce churn, and enhance overall customer satisfaction.

# Project Overview

**Problem Statement:**

To meet the business goals stated earlier, the problems to be addressed are:

- Predict which customers are at risk of churning
- Accurately predict customer churn with at least 80% accuracy

**Approach**:

The following approach was undertaken to address the problems:

- *Data Analysis*: Understand key features impacting customer churn.
- *Model Development*: Build and compare predictive models.
- *Evaluation*: Use metrics like Recall and Accuracy to assess model performance.

# Data Analysis

## Dataset Overview:

The dataset contained the features (variables):

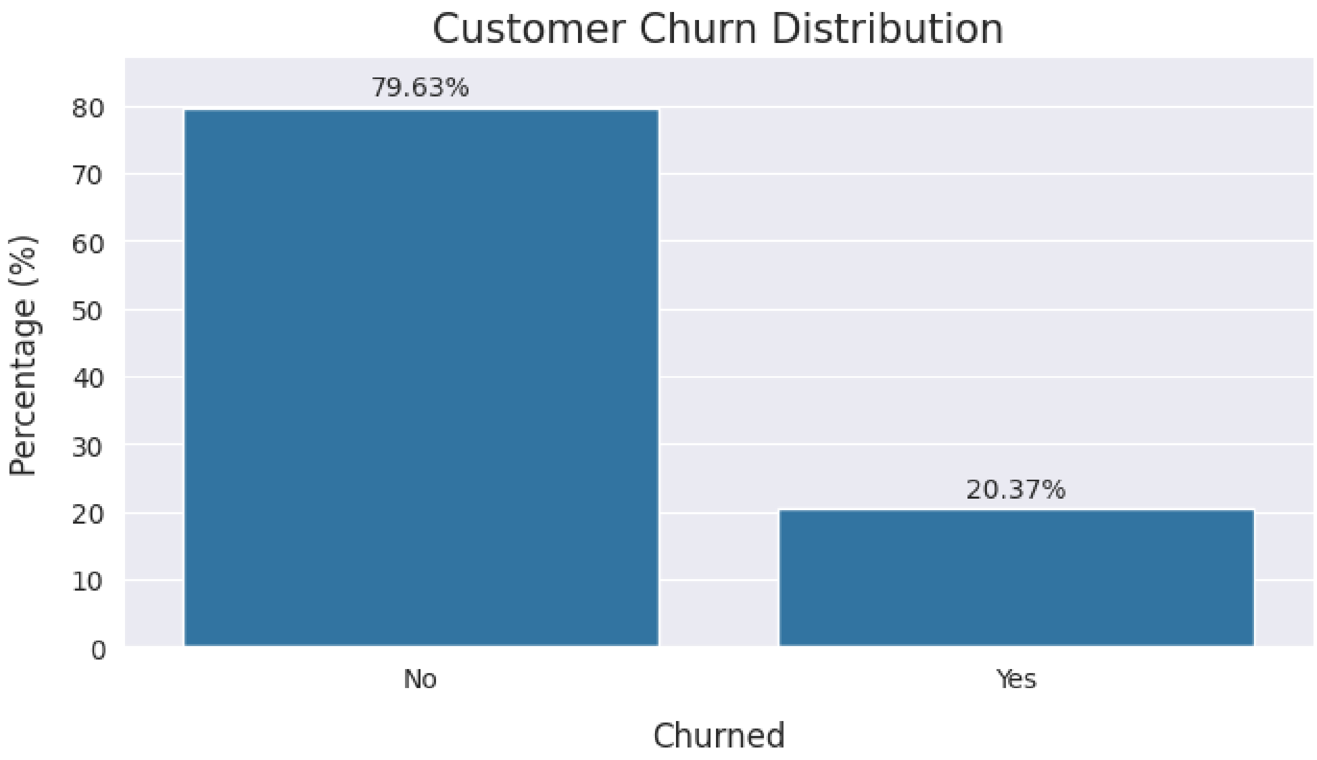
- **CreditScore**: A customer's credit score.
- **Geography**: The country where the customer belongs to.
- **Gender**: The gender of the customer.
- **Age**: The customer's age.
- **Tenure**: The time of bond with company.
- **Balance**: The account balance of the customer (the amount left with them).
- **NumOfProducts**: The products they own
- **HasCrCard**: Do they have a credit card or not? (1 = Yes, 0 = No)
- **IsActiveMember**: Whether the customer is an active member of the company.
- **EstimatedSalary**: The estimated salary of the customer.
- **Exited**: The target variable indicating whether the customer has churned (1 = Churned, 0 = Stayed).

## Data Cleaning and Validation:

The dataset contained no missing values and only required a few data type changes. Hence the dataset underwent a cleaning procedure and was made ready for analysis.

# Key Findings

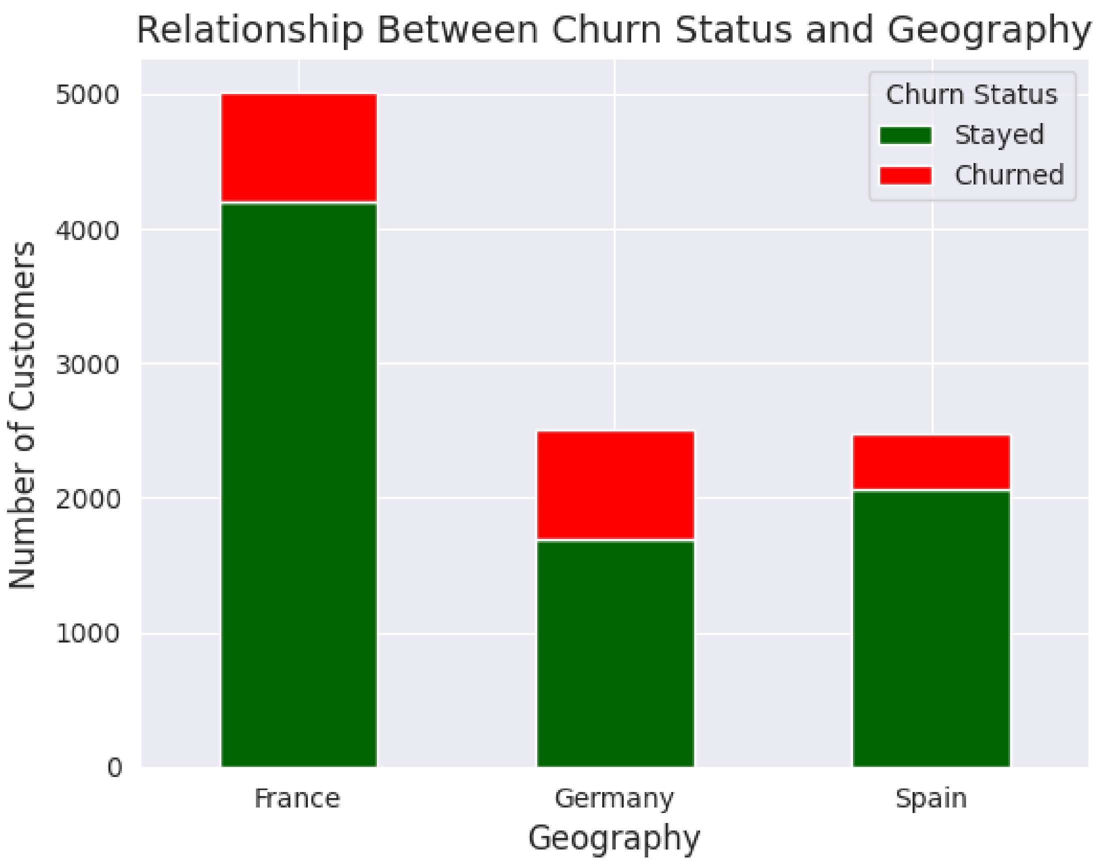
## Customer Churn Distribution




About 20% of the existing customer base has churned, and the remaining 80% haven’t done so.

# Key Findings

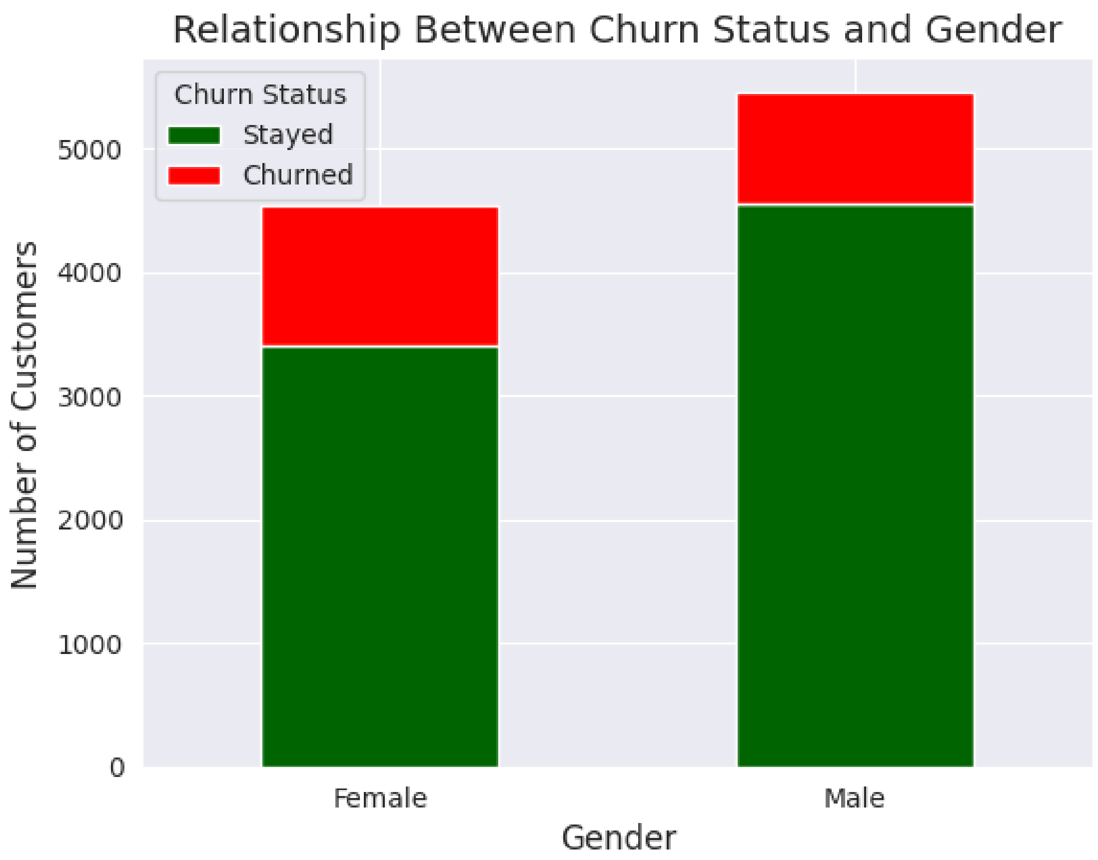
## Churn Status by Geography




France and Spain both have a small portion of their customer base who churned, however, customers from Germany are an exception, as more than one-third of German customers have churned. Customers from Germany seem particularly dissatisfied.

# Key Findings

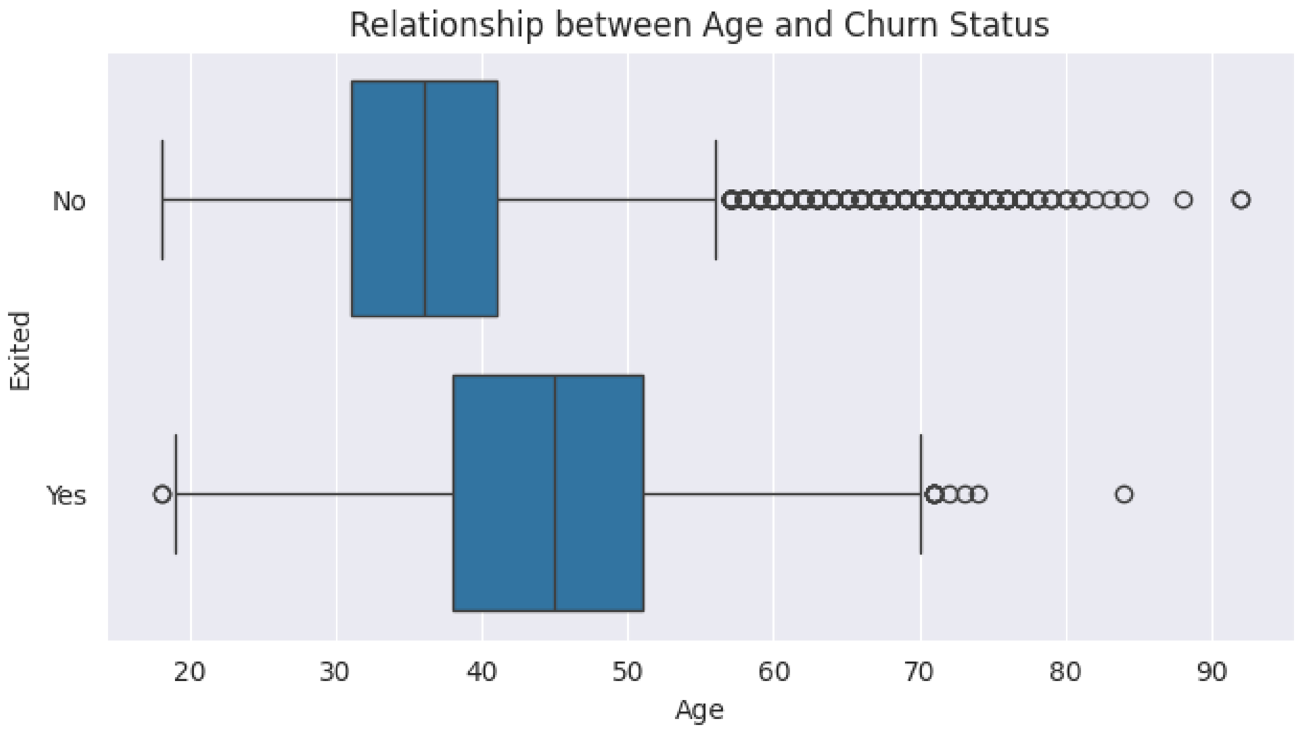
## Churn Status by Gender




Even though the customers who have churned have a higher proportion of females, it is worth noting that more than a quarter of female customers have churned, despite there being considerably more male customers than female customers, which means females have a higher probability of churning than males.

# Key Findings

## Relationship between Age and Churn Status




The median age of customers who have churned is around 45, while for those who haven't, it’s about 35. For non-churning customers, anyone over 56 is considered an outlier, but this is not the case for churned customers. This suggests that customer churn is influenced by age, with the probability of churn increasing as customers age.

# Model Development

**Data Preprocessing:**

- Standardized numeric features
- Encoded categorical features using label and one-hot encoding.

**Model Development:**

Three models were created and evaluated:

- Logistic Regression
- Random Forest
- XGBoost

**Evaluation Metrics:**

- **Recall**: Ensures that most customers at risk of churning are identified.
- **Accuracy**: Measures the overall performance of the model in classifying churned and non-churned customers.

# Predictive Modeling Results

- The **XGBoost model** performs best with a recall of 0.52 and an accuracy of 0.87.
- The **Random Forest model** shows good performance after hyperparameter tuning, with a recall of 0.45 and accuracy of 0.86.
- The **Logistic Regression model** underperformed with a recall of 0.15 and accuracy of 0.81.
- The **XGBoost model** is chosen for its superior recall and overall model quality.

# Business Metrics

**Key Metrics to Monitor:**

- **Recall (0.52)**: Ensures that 52% of at-risk customers are correctly identified, maximizing retention efforts and reducing churn.
- **Accuracy (0.87)**: Measures the overall performance of the model, ensuring that both churned and non-churned customers are correctly predicted.

**Impact on Decision-Making:**

- **Customer Retention**: High recall helps ensure that most at-risk customers are targeted with retention strategies, improving customer loyalty.
- **Performance Monitoring**: Regularly tracking these metrics will help fine-tune the model to improve both recall and accuracy, optimizing retention efforts.

**Monitoring Strategy:**

- **Regular Tracking**: Monthly evaluation of recall and accuracy to assess the model's performance and adjust.
- **Model Adjustment**: Fine-tune decision thresholds and retrain the model as needed to maintain high performance.

# Recommendations

To maximize customer retention, the business should:

- Offer credit score improvement programs to customers with low and highly variable credit scores.
- Investigate reasons for high churn among German customers and tailor retention strategies for this group.
- Create targeted retention campaigns for female customers, addressing their higher churn rate.
- Provide loyalty programs and premium services to high-balance customers to encourage retention.
- Engage credit card holders with rewards programs and personalized offers to reduce churn.
- Offer age-specific retention strategies, especially for older customers, to reduce churn.
- Regularly monitor and improve the churn prediction model, focusing on recall and accuracy.
- Implement continuous customer feedback loops to identify and address reasons for churn.

These actions will help sustain high user engagement and increase subscription rates.

# Thank You!