Module 2

Churn Model

AI for communication and marketing project

Presented By:

MARIE ELYSE BASSIL - 503962

Objective

Customer churn is a critical issue for businesses, as losing customers directly impacts revenue and growth. Understanding and predicting which customers are likely to leave can help in developing targeted retention strategies, ultimately improving customer satisfaction and loyalty.

- Topic Analyzed: This project focuses on analyzing customer churn by leveraging historical data to identify patterns and predict future churn. The goal is to proactively address customer attrition by understanding the factors that contribute to it.
- AI Model Support: The application of a Random Forest classifier, an AI model, is used to predict churn with high accuracy. This model helps in identifying at-risk customers, enabling the business to take preventive actions and tailor retention strategies effectively.

Dataset Overview and Key Variables

Dataset Description: The dataset used in this project contains customer demographic, account, and transaction data, providing a comprehensive view of customer behavior and characteristics. It includes various attributes related to customers' engagement with the business and financial behavior.

Client Information

- customer id
- customer age
- gender
- dependent count

Account Information

- credit limit,
- months on book
- total relationship count

Behavioral Metrics

- total transaction amount
- total transaction count
- months inactive in the last 12 months
- contacts count in the last 12 month
- average utilization ratio of the credit limit.

Target Variable

- Attrition flag: indicates whether a customer has churned or not.

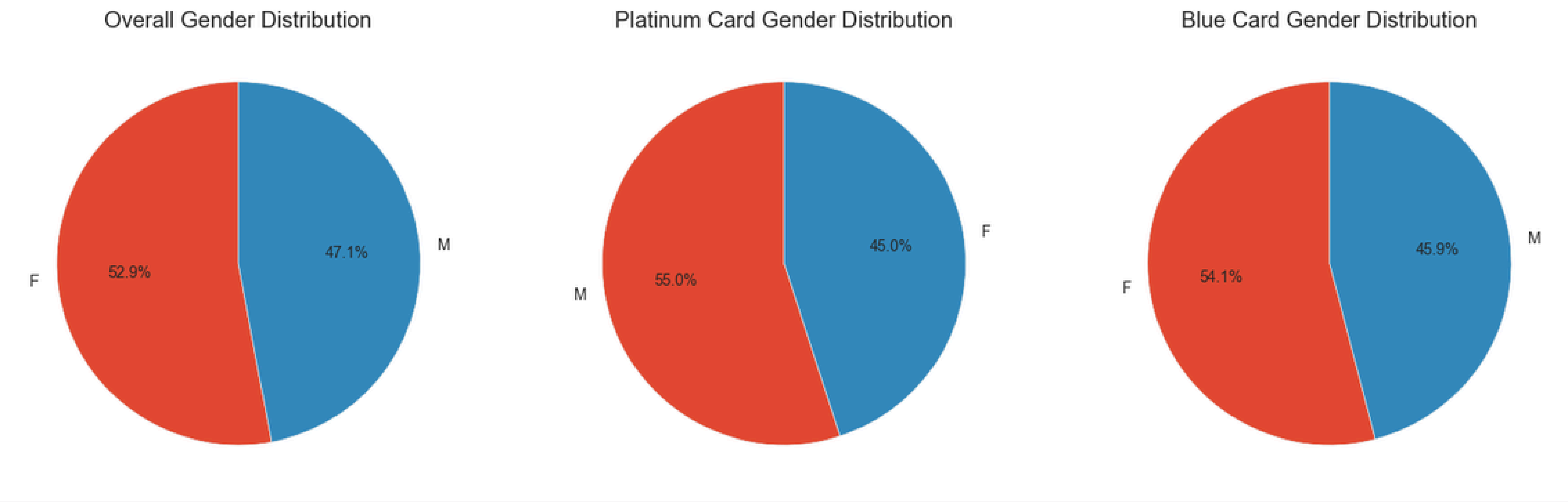
The complete dataset:

- CLIENTNUM --> Unique identifier for each client.
- Attrition_Flag --> Indicates whether the client has churned or not.
- Customer_Age --> Age of the customer.
- Gender --> Gender of the customer.
- Dependent_count --> Number of dependents the customer has.
- Education_Level --> Education level of the customer.
- Marital_Status --> Marital status of the customer.
- Income_Category --> Income category of the customer.
- Card_Category --> Category of the customer's credit card.
- Months_on_book --> Number of months the customer has been on the books.
- Total_Relationship_Count --> Total number of products the customer has with the bank.
- Months_Inactive_12_mon --> Number of months the customer was inactive in the last 12 months.
- Contacts_Count_12_mon --> Number of contacts made in the last 12 months.
- Credit_Limit --> Credit limit of the customer.
- Total_Revolving_Bal --> Total revolving balance on the customer's account.
- Avg_Open_To_Buy --> Average amount available to spend, calculated as Credit Limit minus Balance.
- Total_Amt_Chng_Q4_Q1 --> Change in transaction amount (Q4 over Q1).
- Total_Trans_Amt --> Total transaction amount in the last 12 months.
- Total_Trans_Ct --> Total transaction count in the last 12 months.
- Total_Ct_Chng_Q4_Q1 --> Change in transaction count (Q4 over Q1).
- Avg_Utilization_Ratio --> Average utilization ratio of the credit limit.

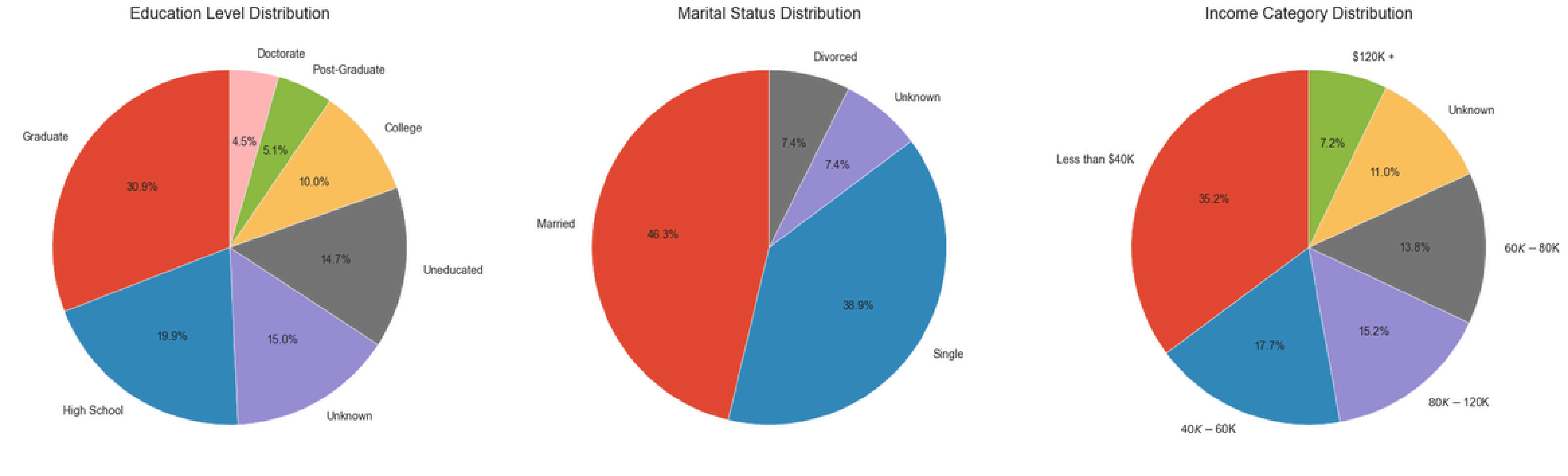
Analytical Methodology and Model Development

The Exploratory Data Analysis (EDA) phase was crucial in understanding the distribution of customer attributes and identifying key patterns that might influence customer churn. EDA also helped in detecting outliers, assessing variable distributions, and exploring relationships between variables.

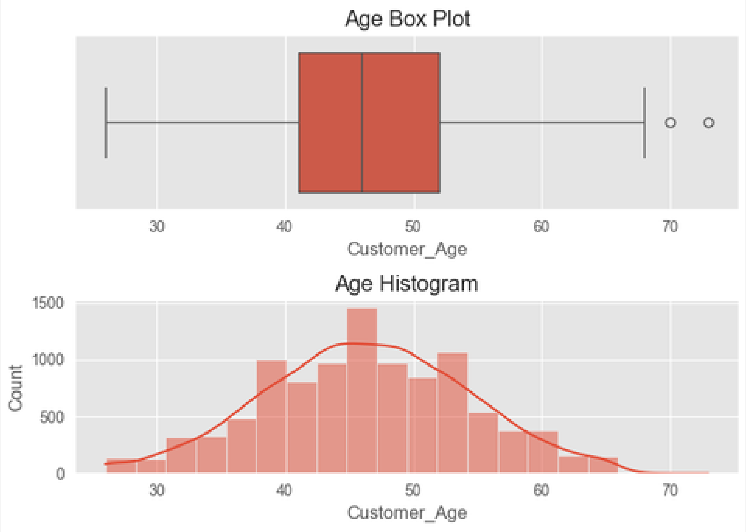
Gender and Card Type:



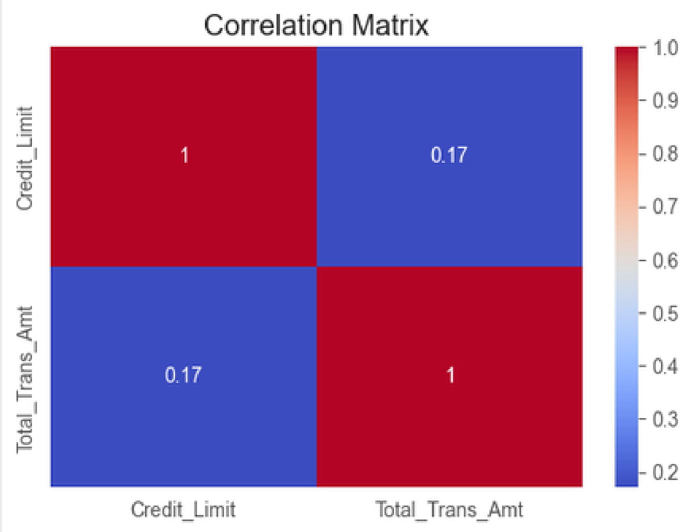
Educational and Income Levels:



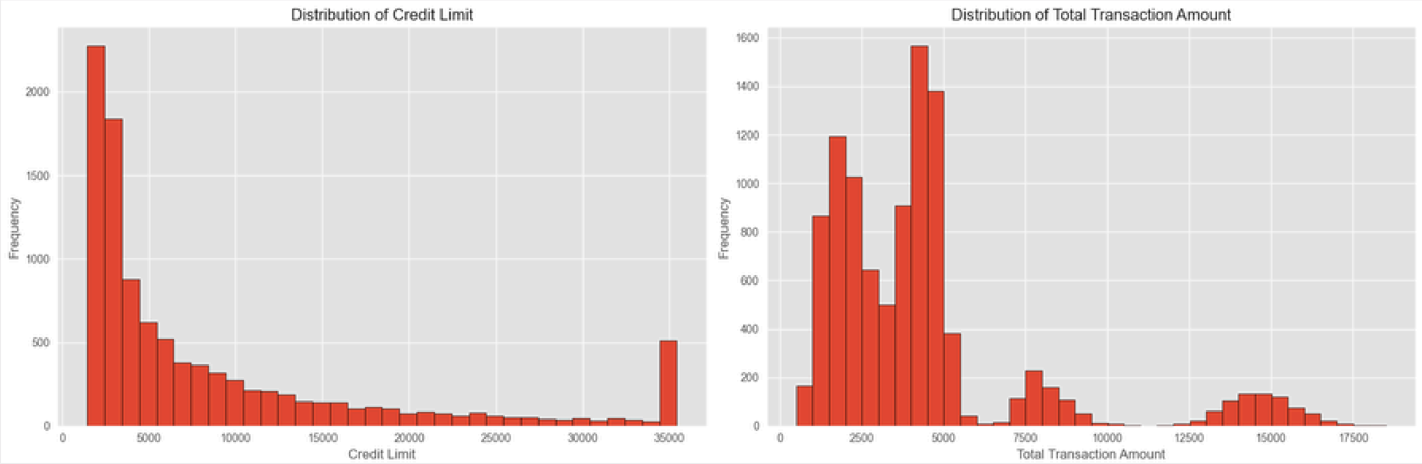
Customer Age Distribution:



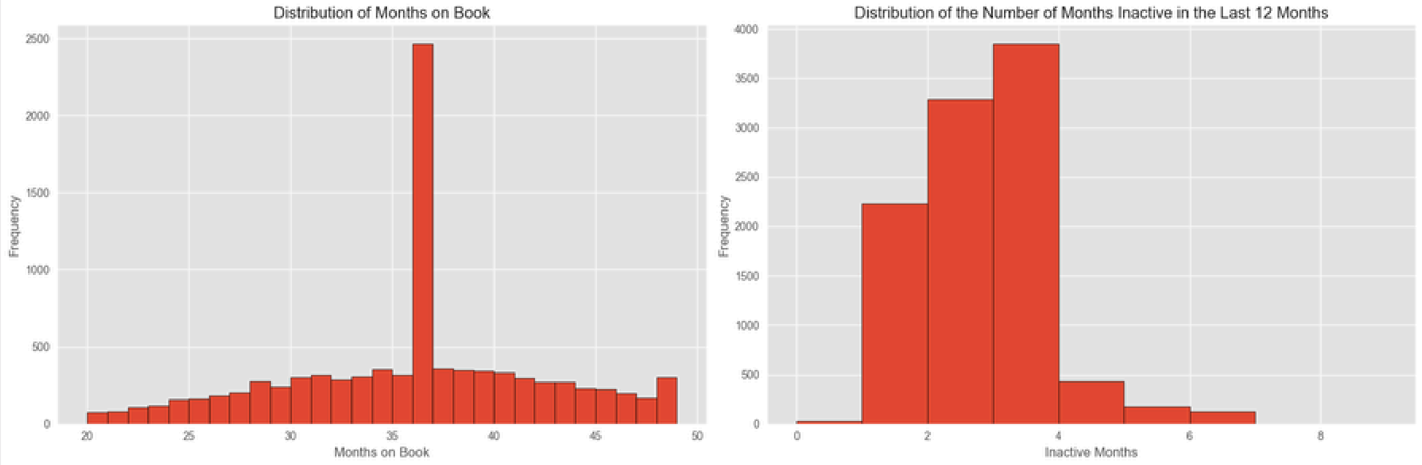
Correlation Analysis:



Credit Limit and Transactions:



Account activity



The EDA revealed that the median customer age is around 45, with most customers between 25 and 65 years old. Gender distribution shows a slight female majority overall, though males are slightly more prevalent among Platinum cardholders. Education levels are diverse, with a large portion having graduate or high school education, while income distribution is skewed towards lower income categories. Account activity shows that many customers have been on the books for about 36 months, with moderate recent inactivity. Finally, credit limits and transactions amounts display varied peaks, but there's only a weak correlation between the two.

Analytical Methodology and Model Development

Feature engineering

Categorical variables were one-hot encoded, and numerical features were standardized to improve model performance.

Feature selection

Chi-squared tests and T-tests were used to identify significant categorical and numerical features that contribute to churn, ensuring the model focuses on the most relevant data.

Model Development

A Random Forest classifier was chosen due to its robustness and ability to handle both categorical and numerical data.

Hyperparameter Tuning

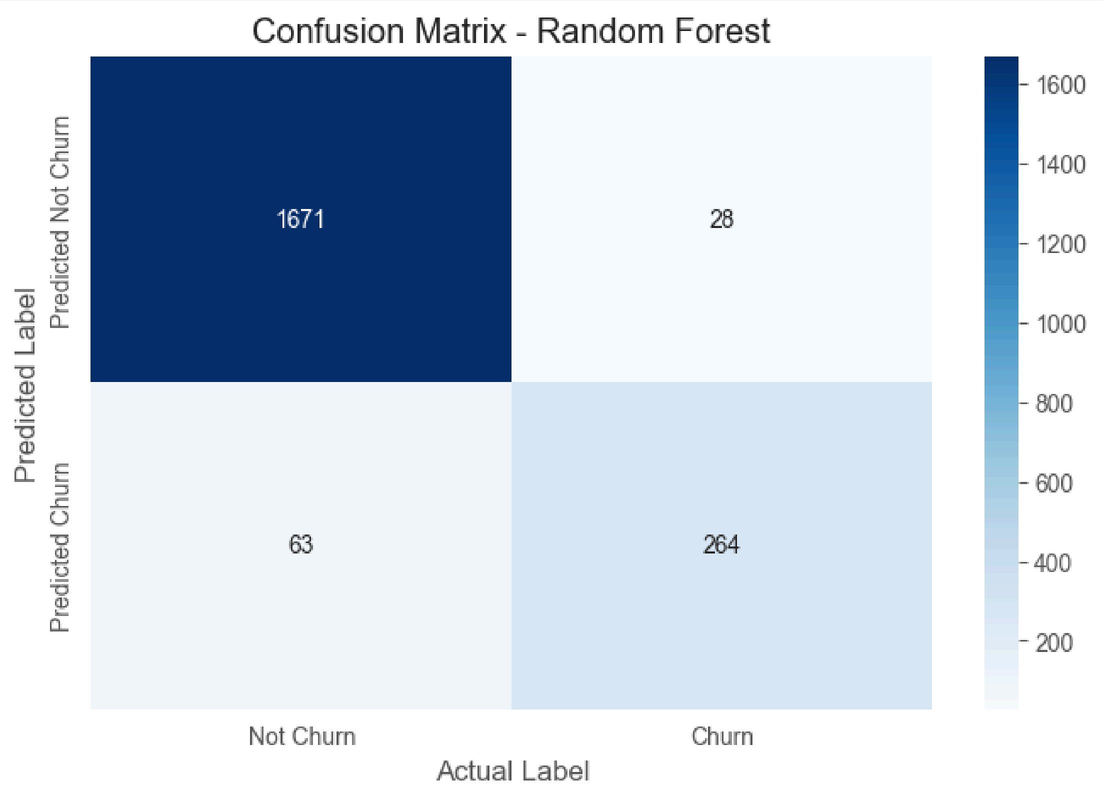
GridSearchCV was used to optimize the model's hyperparameters, ensuring the best possible performance.

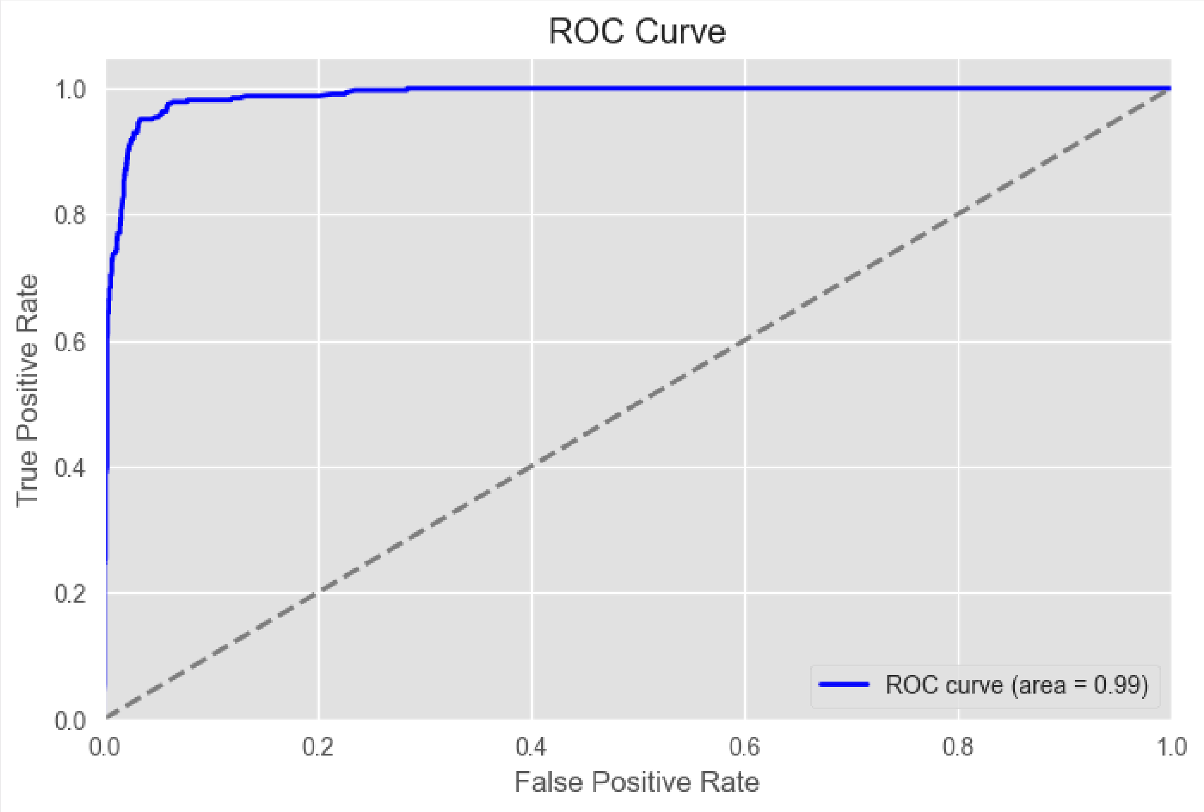
Model Evaluation:

```
Accuracy: 0.9550839091806516
F1 Score: 0.8529886914378029
Classification Report:
              precision    recall  f1-score   support

           0       0.96      0.98      0.97      1699
           1       0.90      0.81      0.85       327

    accuracy                           0.96      2026
   macro avg       0.93      0.90      0.91      2026
weighted avg       0.95      0.96      0.95      2026
```





Critical Commentary on Business Results and Practical Implications

The Random Forest model achieved an accuracy of 95.51% in predicting customer churn. It performed well in distinguishing between churned and non-churned customers, with a particularly strong performance in predicting non-churners.

Targeted Retention Strategies:
The model identifies at-risk customers, allowing the business to implement personalized retention strategies, such as targeted offers or personalized communications, to reduce churn.

Resource Allocation:
By understanding which customers are likely to churn, the business can allocate resources more effectively, focusing on the most at-risk segments.

Continuous Improvement:
Regular updates and monitoring of the model will ensure it adapts to changing customer behaviors, maintaining its effectiveness over time.

Thank you

MARIE ELYSE BASSIL - 503962