BANK CUSTOMER CHURN DATASET

BANK CUSTOMER DATA FOR PREDICTING CUSTOMER CHURN

OVERVIEW

- ABC Multinational Bank is a leading global financial institution that offers a wide range of banking services, including retail, corporate and investment banking. With a diverse customer base (Spain, France and Germany), the bank's success relies on its ability to foster strong, lasting relationships with its customers. However, customer churn – a situation where customers stop using the bank's services – poses a significant challenge to the bank's growth and profitability.

- Through the research found from this dataset, it can be seen that ABC Multinational Bank has a churn rate of about 20.4%.

OBJECTIVE

- The primary goal is to develop a predictive model that helps ABC Multinational Bank identify customers at high risk of churning. By understanding the factors that influence customer churn, the bank can take proactive measures to retain these customers, improving customer satisfaction and reducing revenue.

DATA UNDERSTANDING

- The dataset is available at Kaggle.com for download through this link:

- https://www.kaggle.com/datasets/gauravtopre/bank-customer-churn-dataset

- There are 10000 rows and 12 columns in this dataset.

- We have both categorical and numerical variables in our dataset.

- Categorical variables: 'country' and 'gender'.

- Numerical variables: 'customer id', 'credit score', 'age', tenure', 'balance', 'product number', 'credit card', 'active member', 'estimated salary', 'churn'

The heatmap shows the correlation between numerical features in the dataset. From this we can see that, with regards to Churn Correlation, Age has a positive correlation(0.29). Meaning that, older customers are more likely to churn. Balance has a weak correlation(0.12). Active_member has a negative correlation(-0.16). they are less likely to churn.

Correlation Heatmap
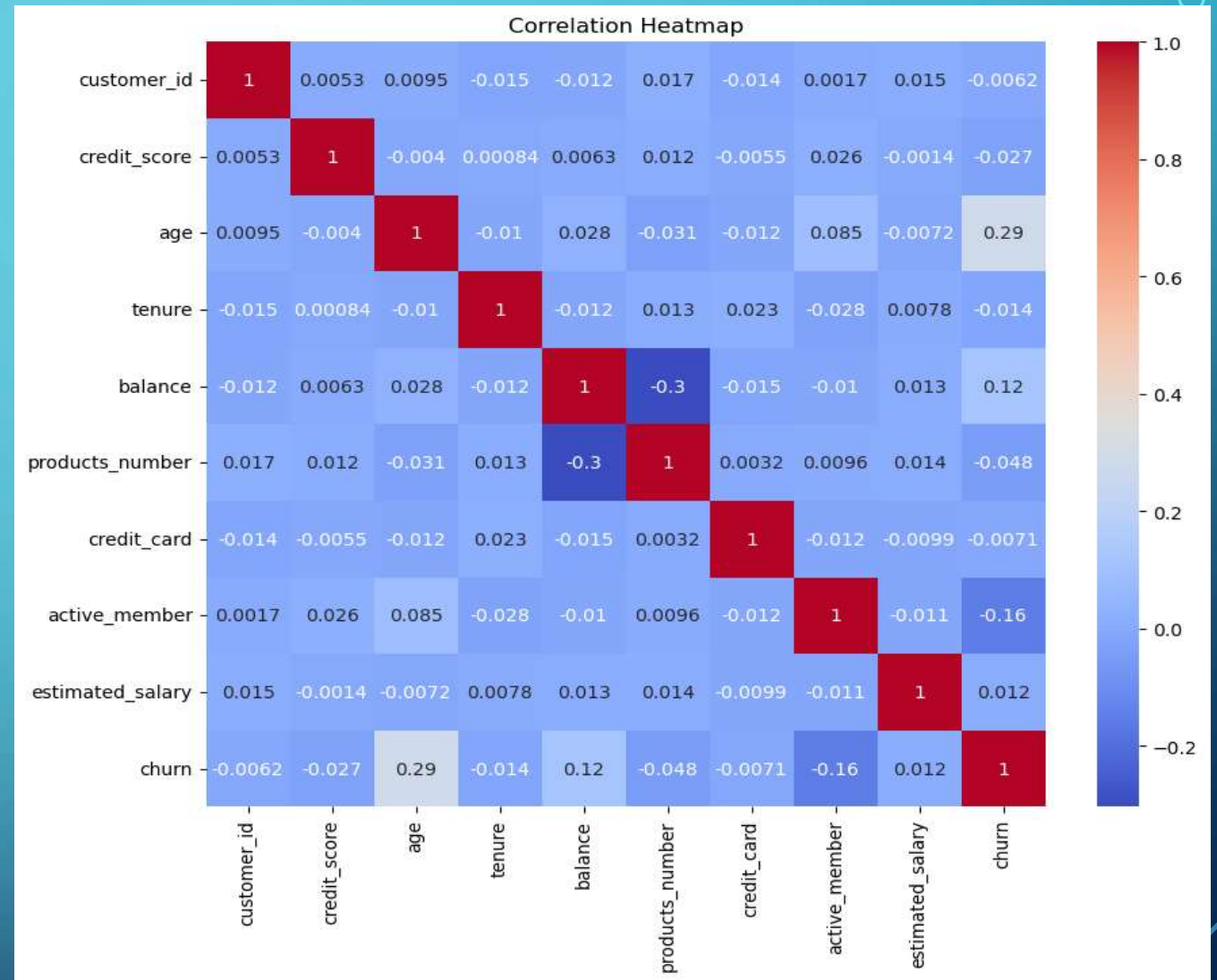
BAR PLOT AND PIE CHART SHOWING CHURN DISTRIBUTION

- This analysis of churn variable shows that 76.9% of customers do not churn, while 20.4% of the remaining customers churn from the company.
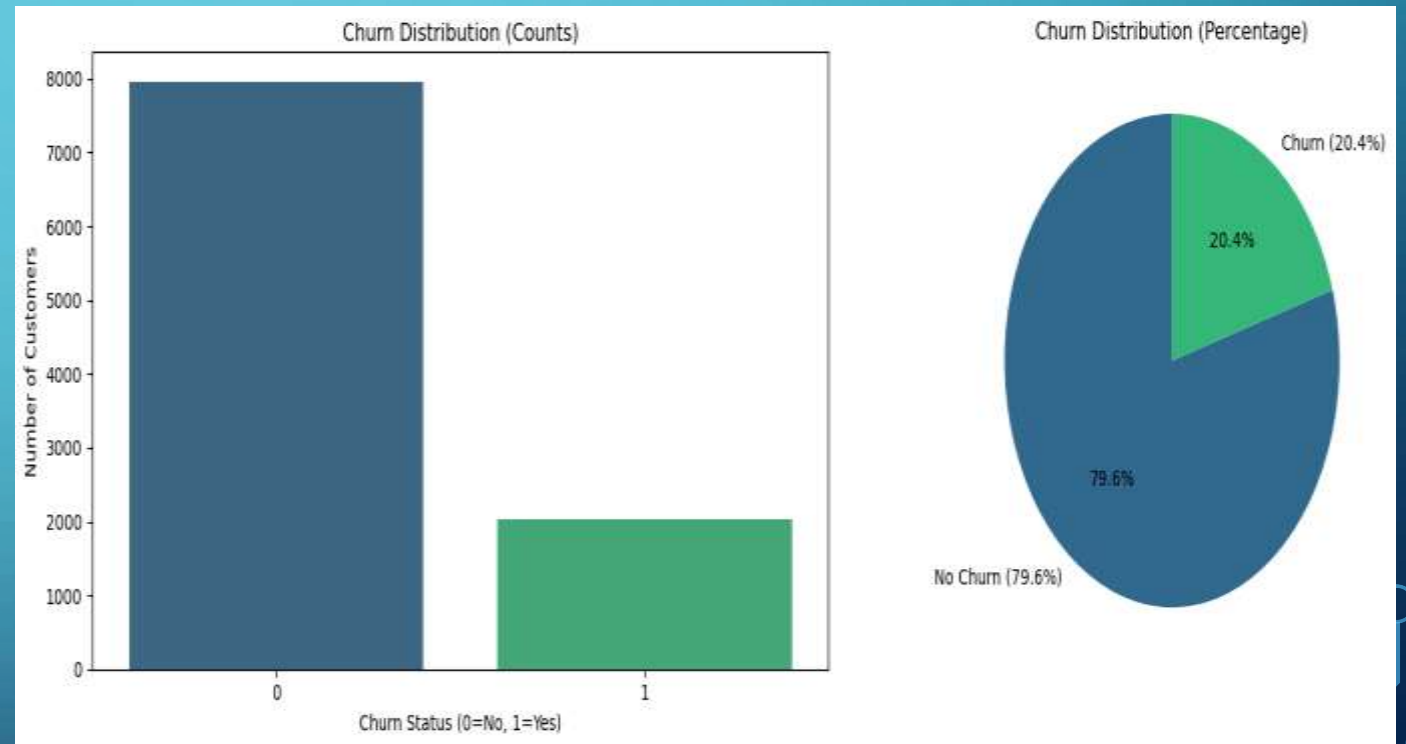
CHURN DISTRIBUTION BY GENDER

- From this visualization, we can see that there are more male customers who have not churned than female customers.

- But female customers have churned from the bank more than the male customers.
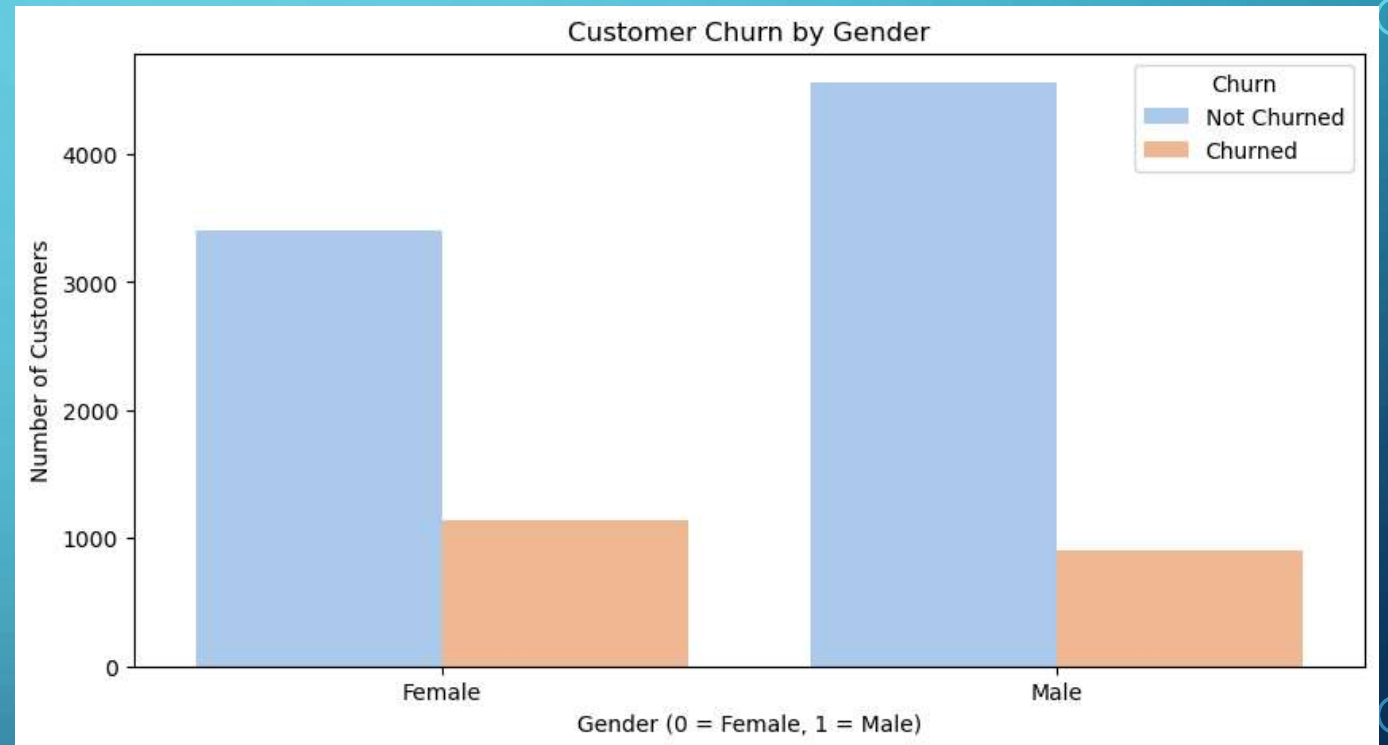
DISTRIBUTION OF DISCRETE VARIABLES

- Majority of the customers are French, and most have been with the bank for a balanced range of years.

- A significant number of customers use 1 or 2 products, which could imply limited cross-selling opportunities.

- Gender distribution is balanced, so no strong bias exists toward male and female customers.
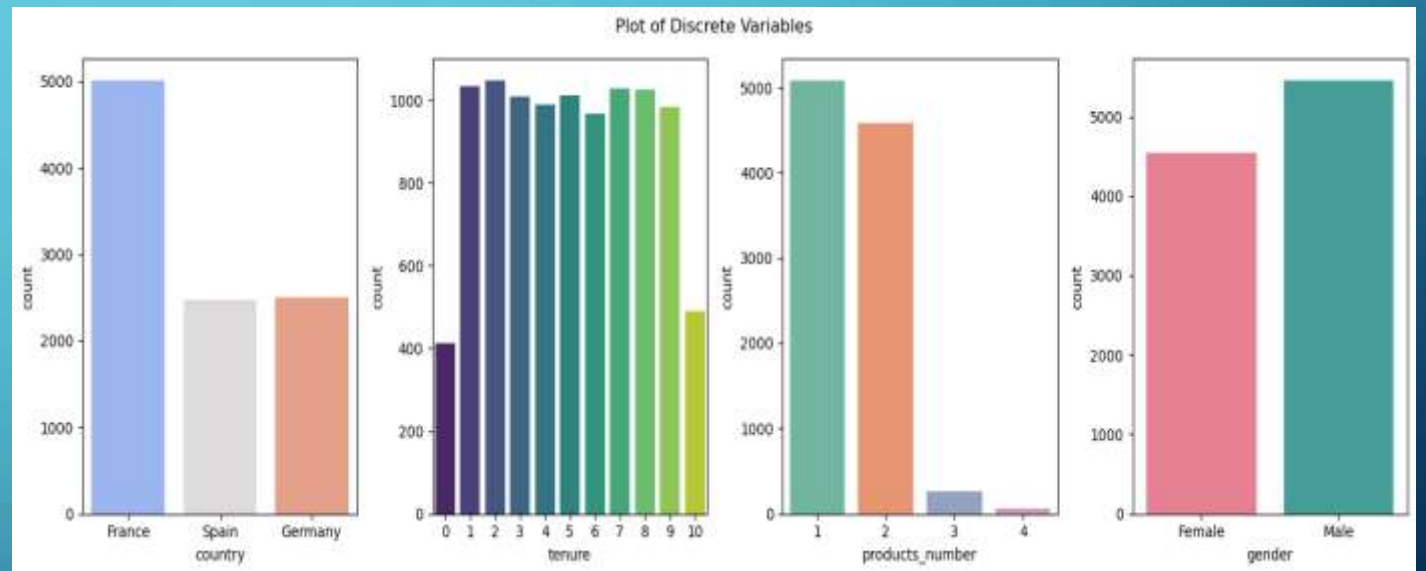
# DISTRIBUTION OF CONTINUOUS VARIABLES

- From age, the distribution is right skewed, with most customers aged 30 and 50 years.

- From balance, a significant amount have zero, which indicates customers without savings or inactive accounts.

- From estimated salary, salaries are uniformly distributed.

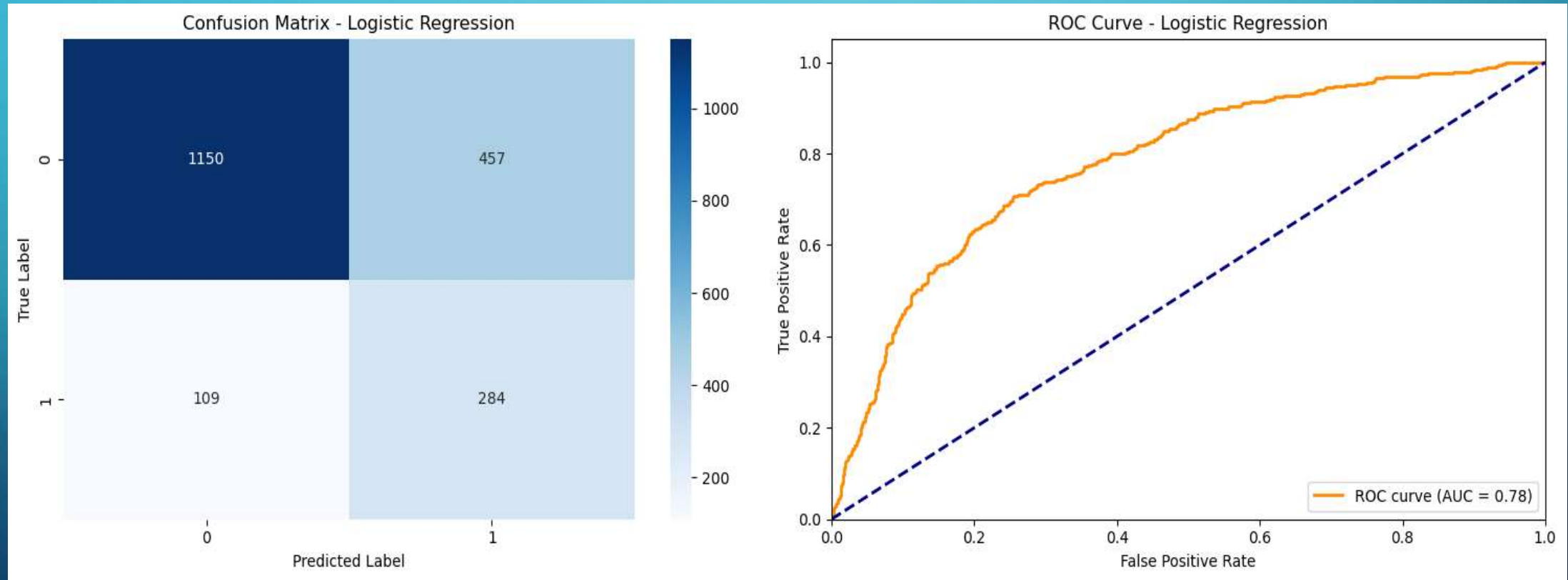- From credit score, the distribution is roughly normal.
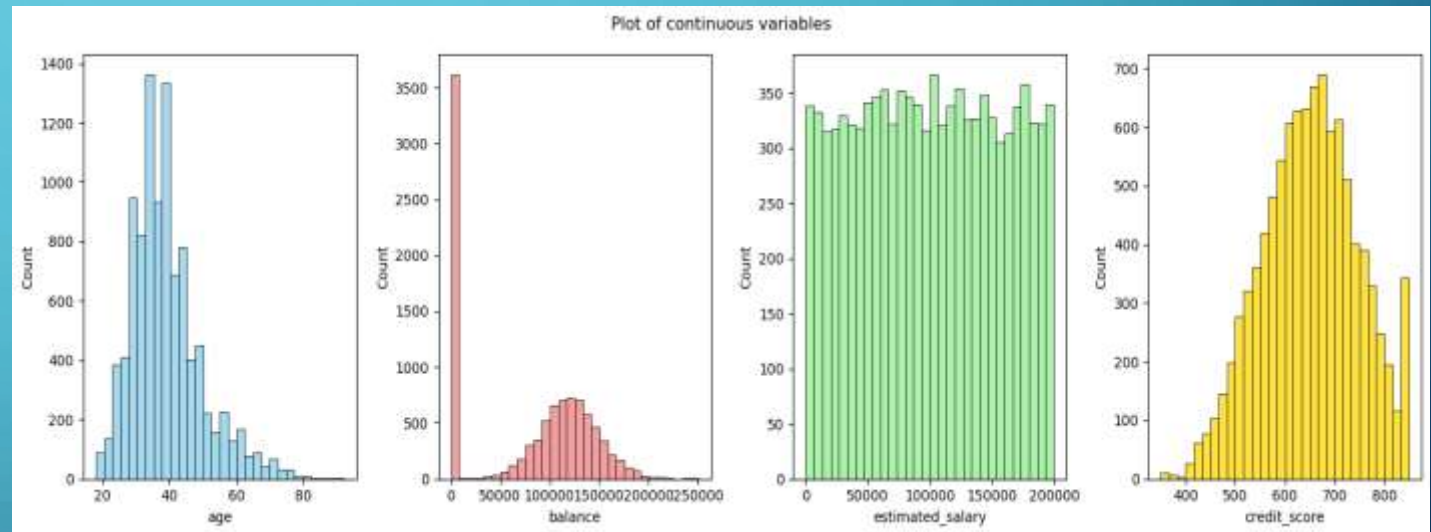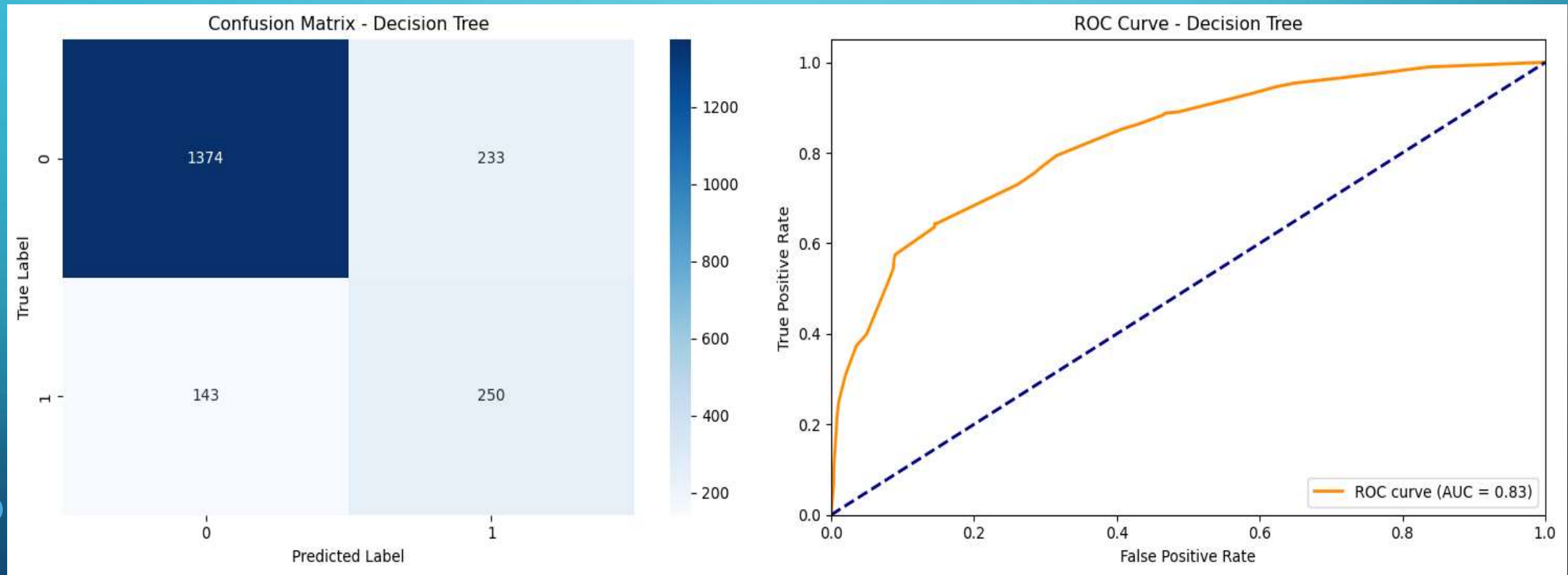
VISUALIZATION OF IMPACT BETWEEN VARIABLES
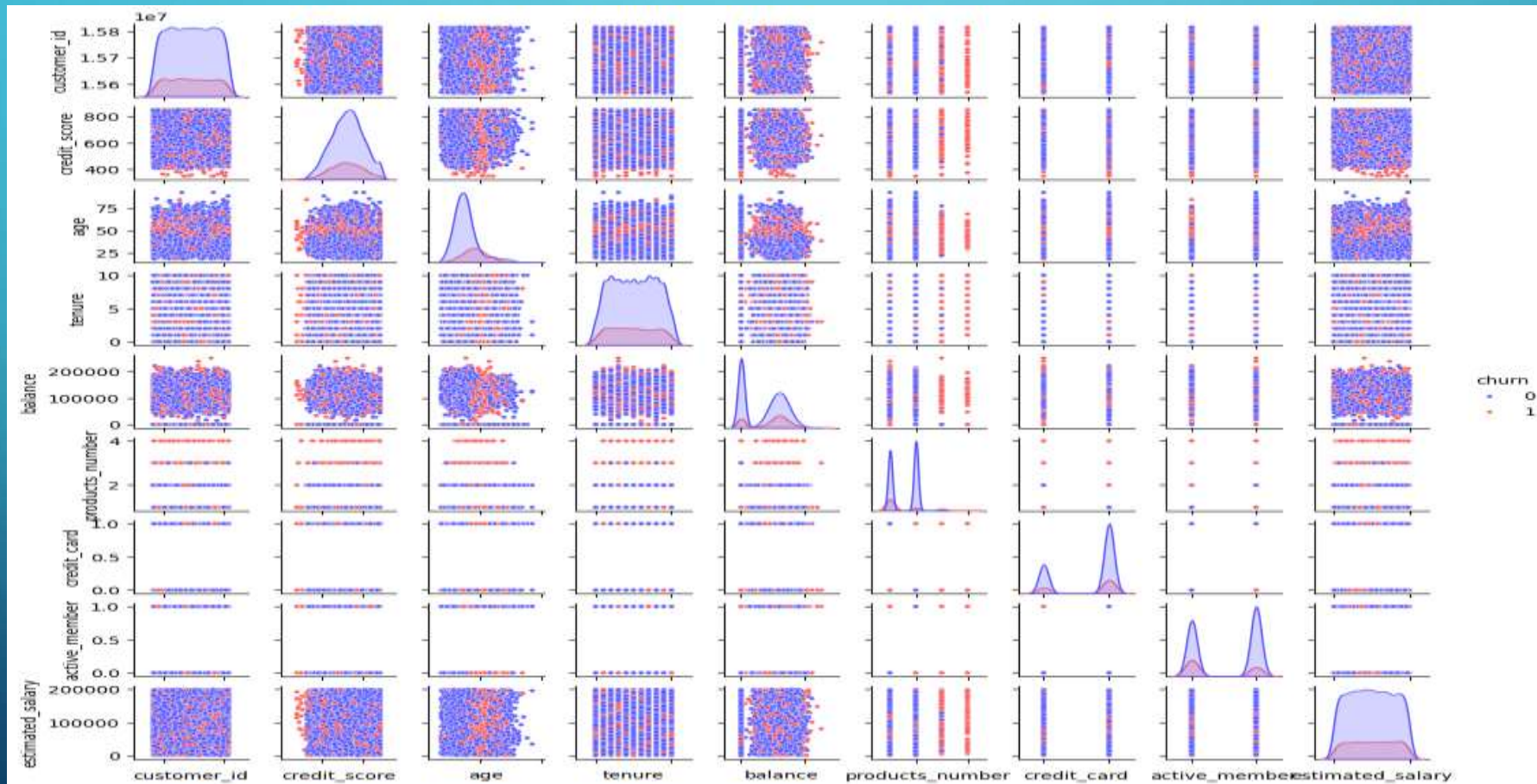
VISUALIZATION OF CORRELATION MATRIX BETWEEN FEATURES

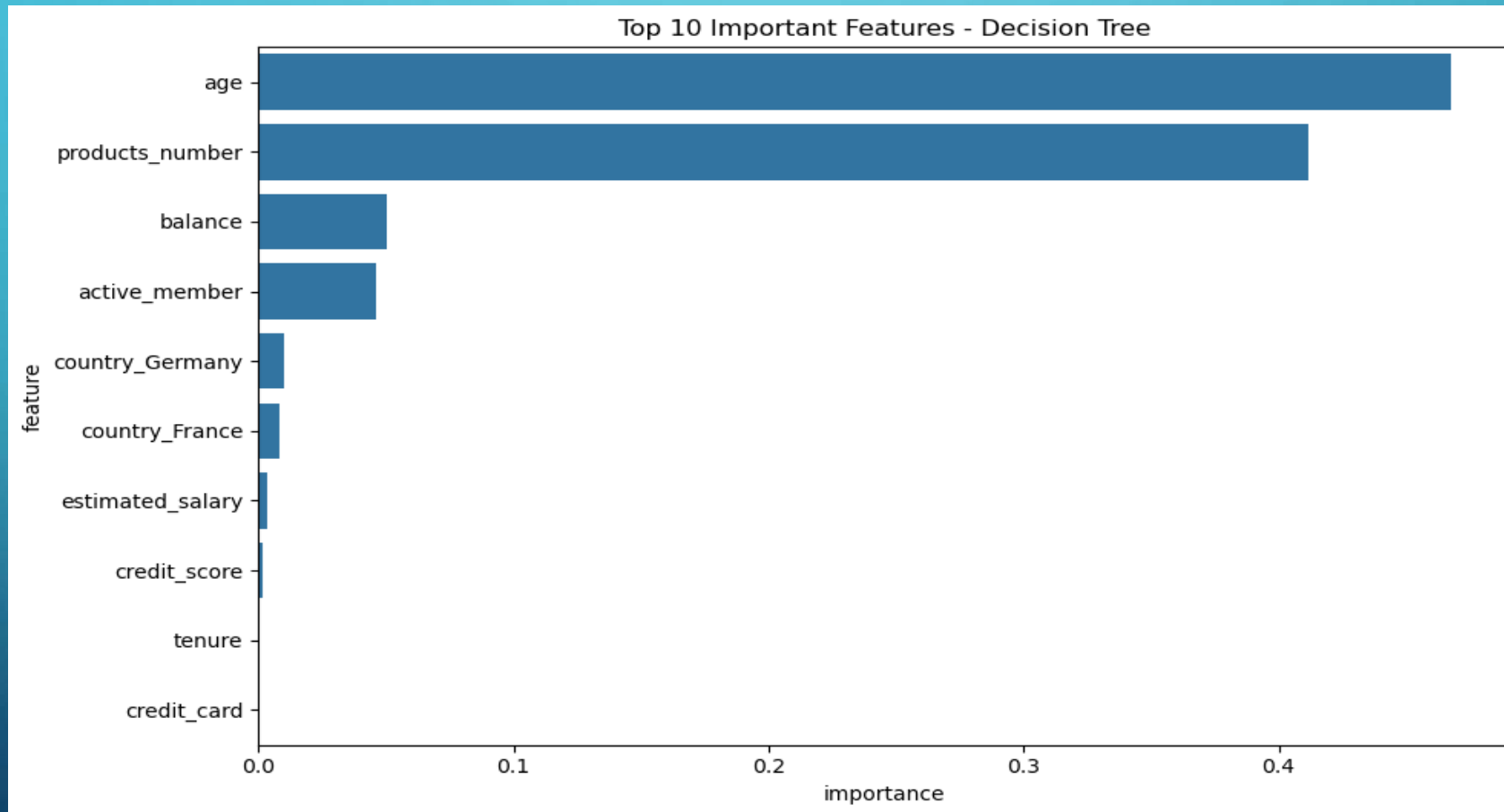Correlation between features
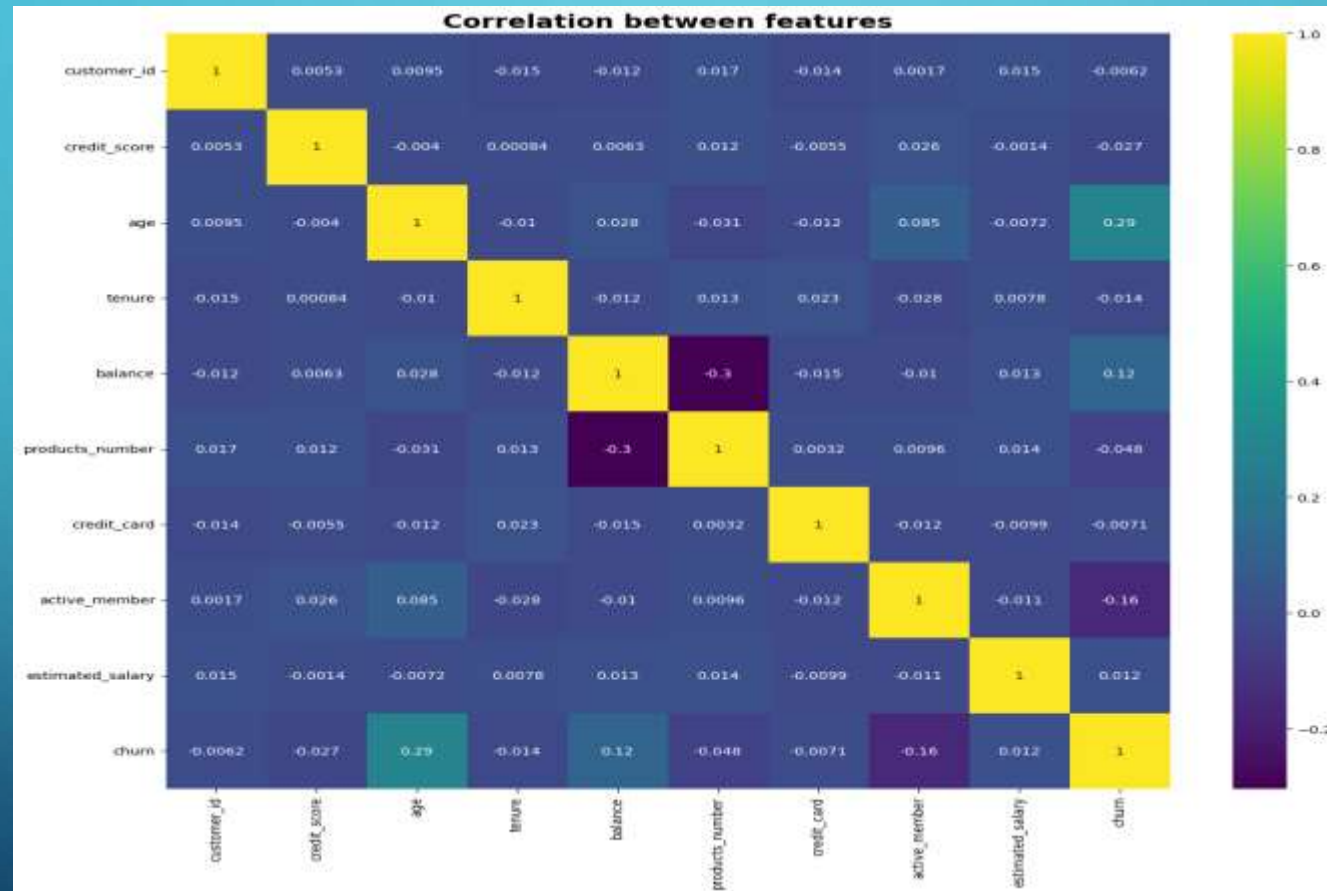
LOGISTIC REGRESSION CLASSIFIER

DECISION TREE CLASSIFIER

TOP 10 FEATURE IMPORTANCES OF DECISION TREES

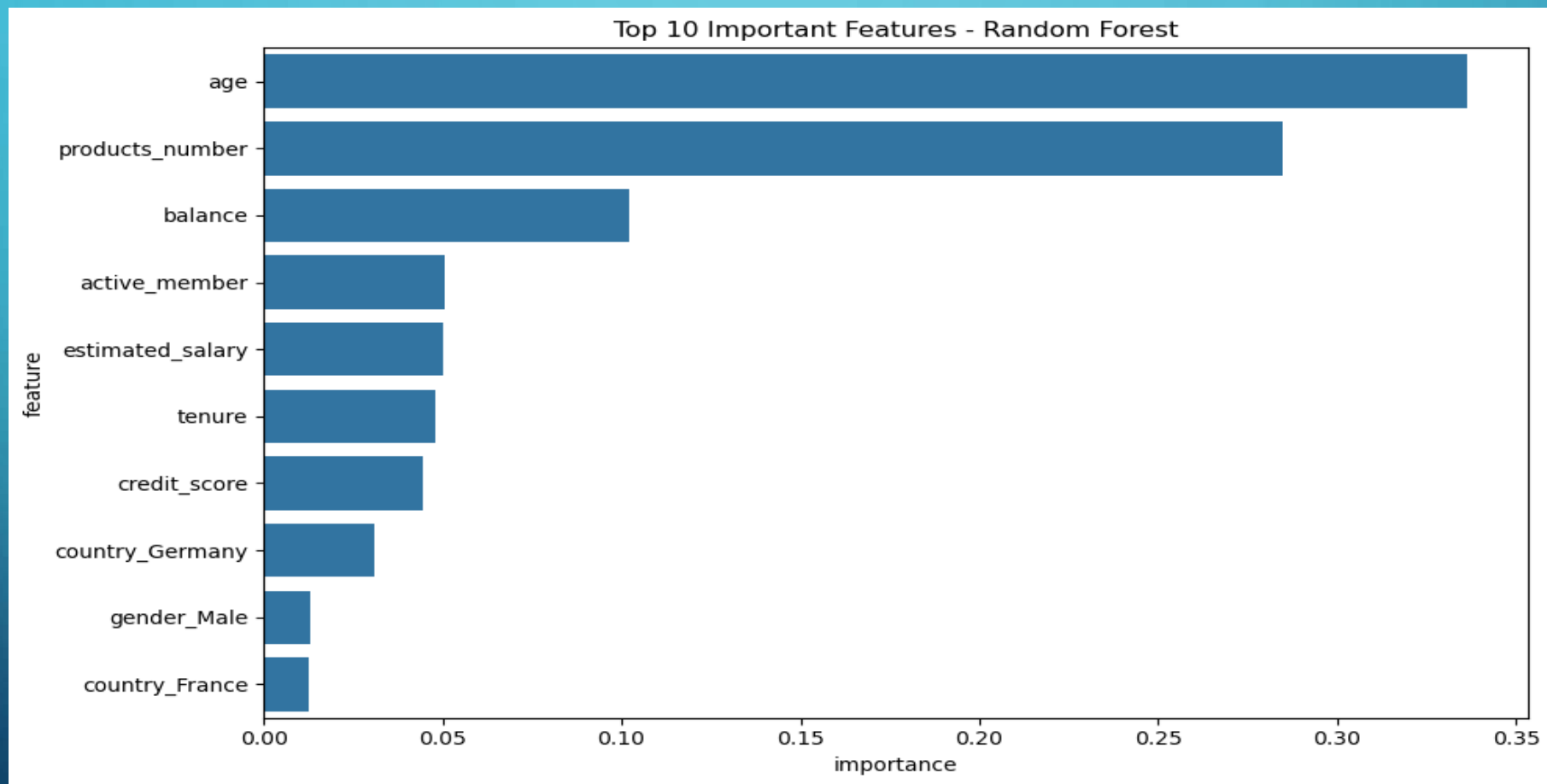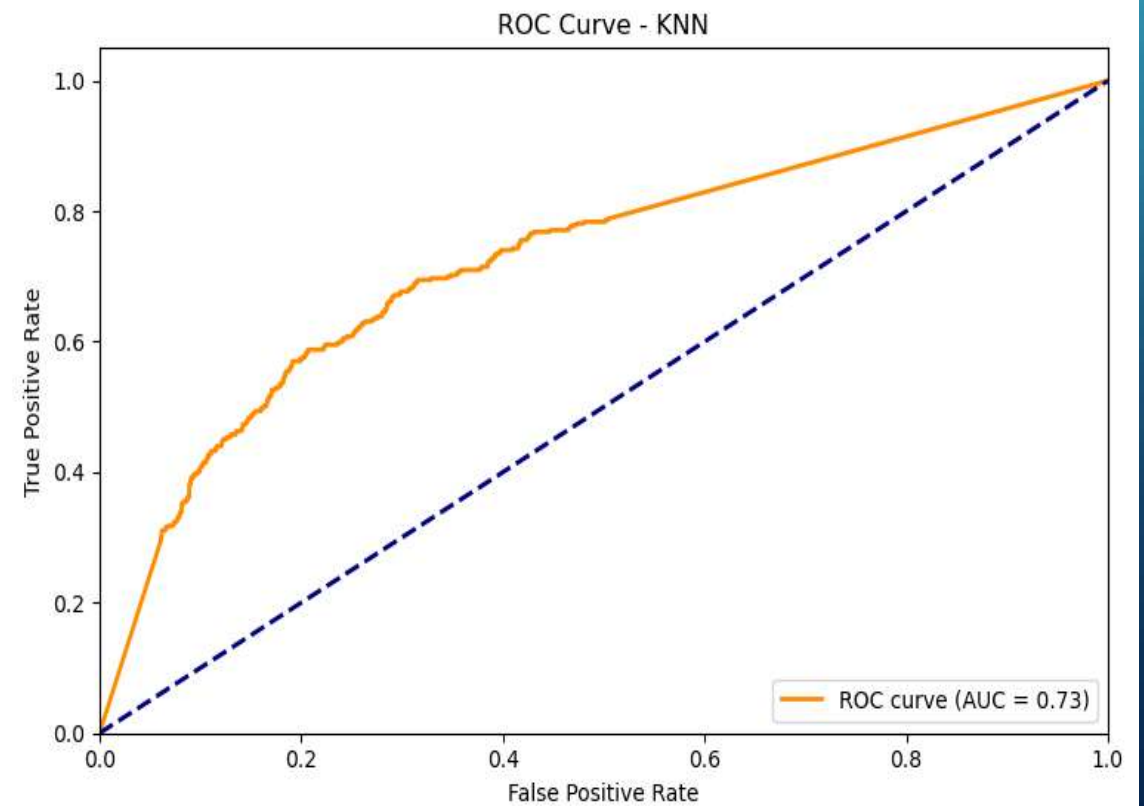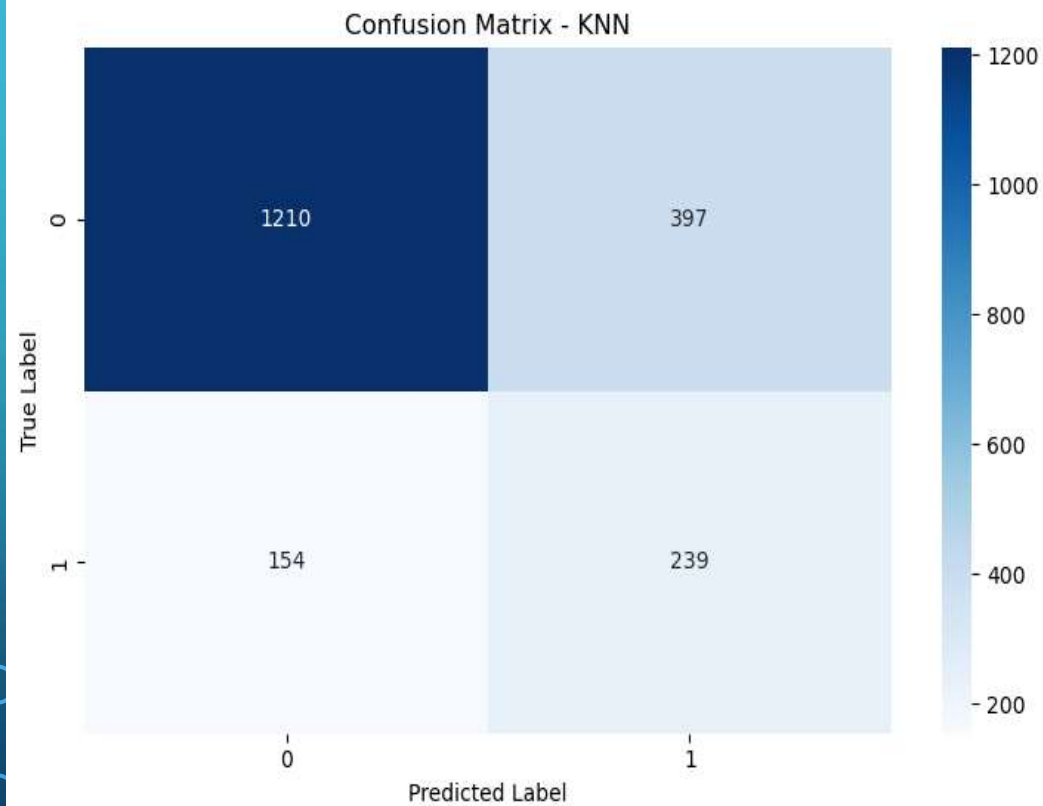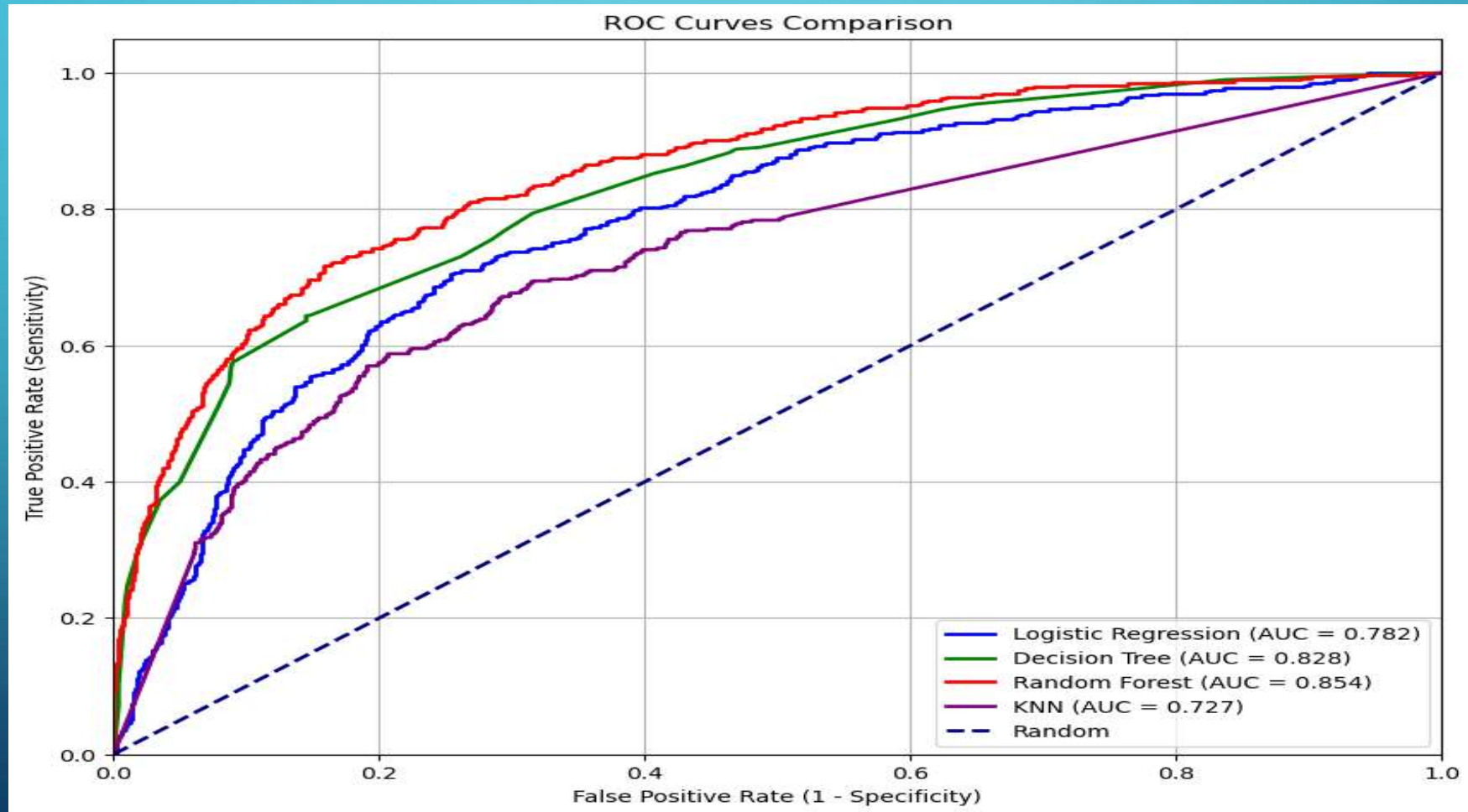Top 10 Important Features - Decision Tree

RANDOM FOREST CLASSIFIER
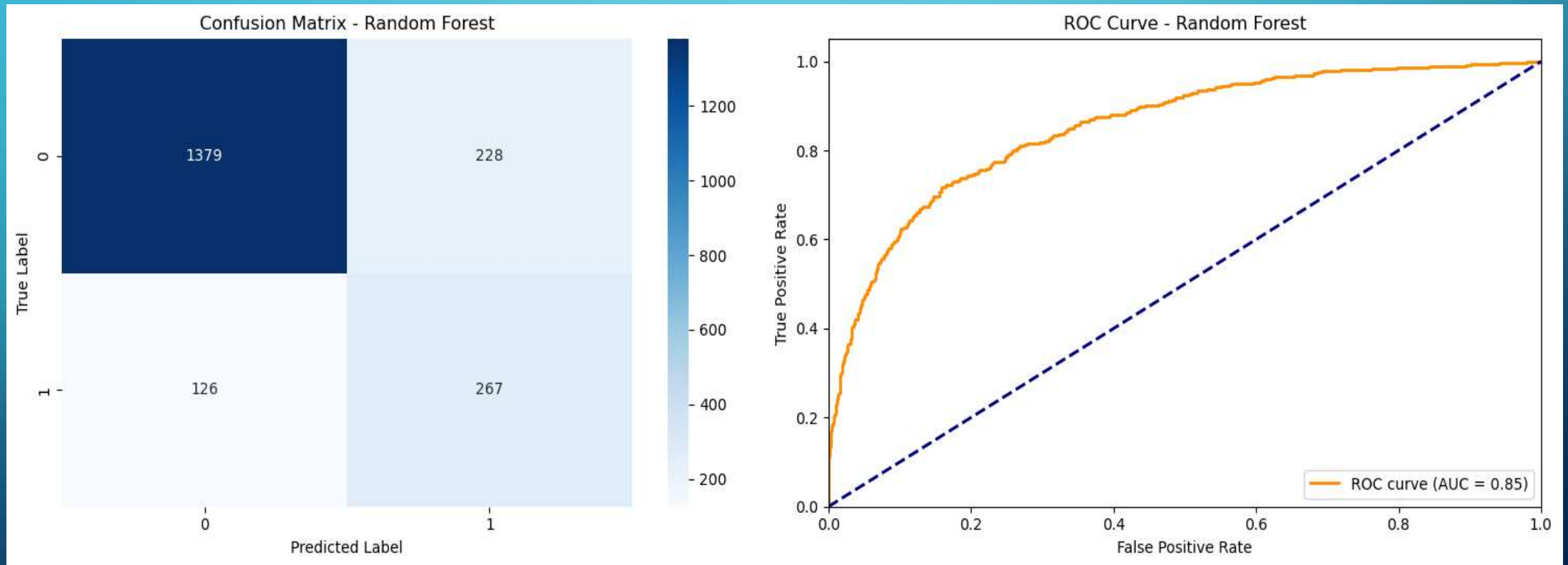
TOP 10 FEATURE IMPORTANCES OF RANDOM FOREST

KNN CLASSIFIER

ROC CURVES COMPARISON

RECOMMENDATIONS

1. Model Implementation-

   - Deploy Random Forest model (82.30% accuracy, 0.85 AUC score)
   - Set up automated alerts for high-risk customers

2. Key Business Actions-

- Age-based strategies: - Targeted programs for older customers - Age-specific products/services

- Product engagement: - Promote multi-product relationships   - Optimize product bundles

- Customer activation: - Re-engagement campaigns   - Rewards for active usage

RECOMMENDATIONS (CONTINUATION)

3. Retention Strategy:- Early warning system, Personalized retention offers, Focus on high-value customers, Track retention program ROI

4. Data & Monitoring:- Regular performance tracking, Customer feedback collection, Measure intervention success

Monitor key metrics: Churn Rates, Customer satisfaction, Retention success rates