CUSTOMER CHURN IN BANKS

Aishwarya Pawar

Emily Wilkins

Namita Ramesh

Thiru Vinayagam

AGENDA

Use Case

Key Factors

Data

Analysis

Findings

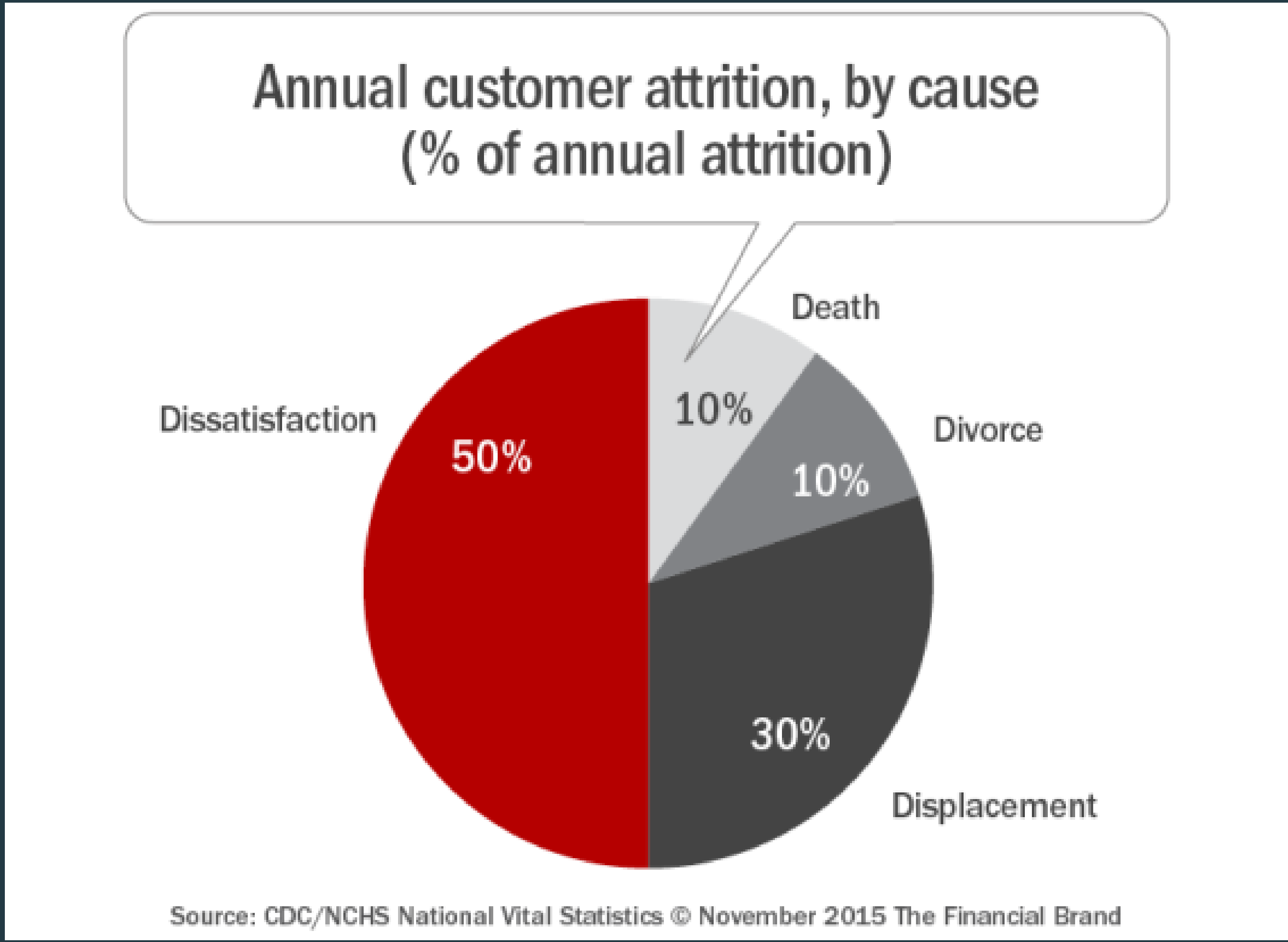
Recommendations

WHAT'S THE PROBLEM?

Bank of America

Goldman Sachs

citi CHASE



WHY DO THEY LEAVE?

WHAT ARE WE LOOKING AT?

A classification problem trying to predict the binary variable Exited

Dropped Variables

- RowNumber
- CustomerId
- Surname

Numerical Variables

- CreditScore
- Age
- Tenure
- Balance
- NumOfProducts
- EstimatedSalary

Categorical Variables

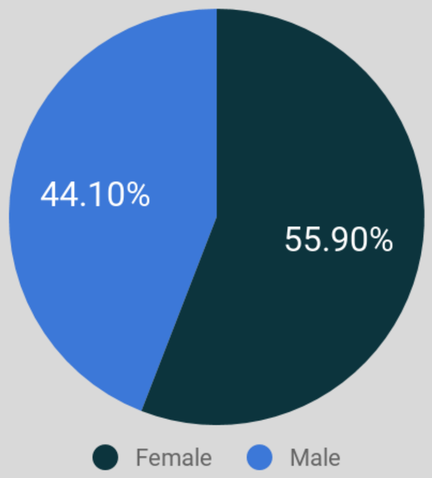
- Geography
- Gender

Binary Variables

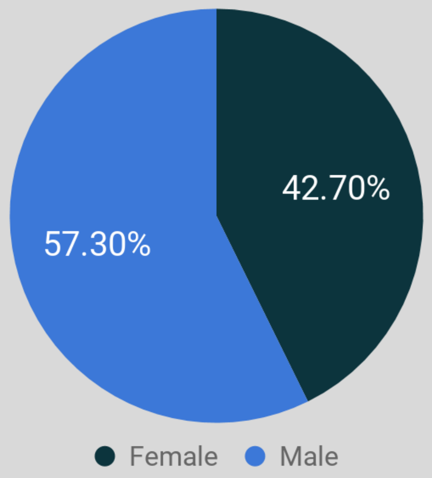
- HasCrCard
- IsActiveMember
- Exited

WHAT DID WE FIND?

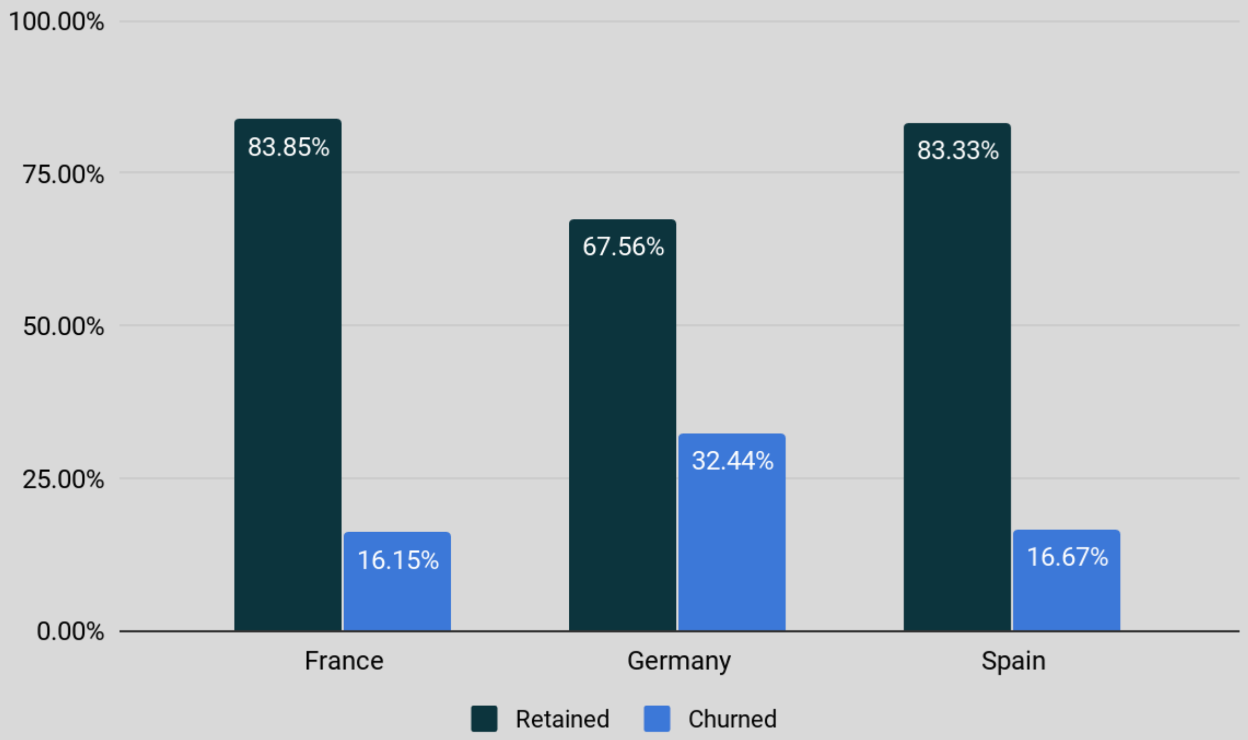
Women are more likely to churn than men. German customers are the most likely to churn of the three geographies represented in the sample.



Churned Customers

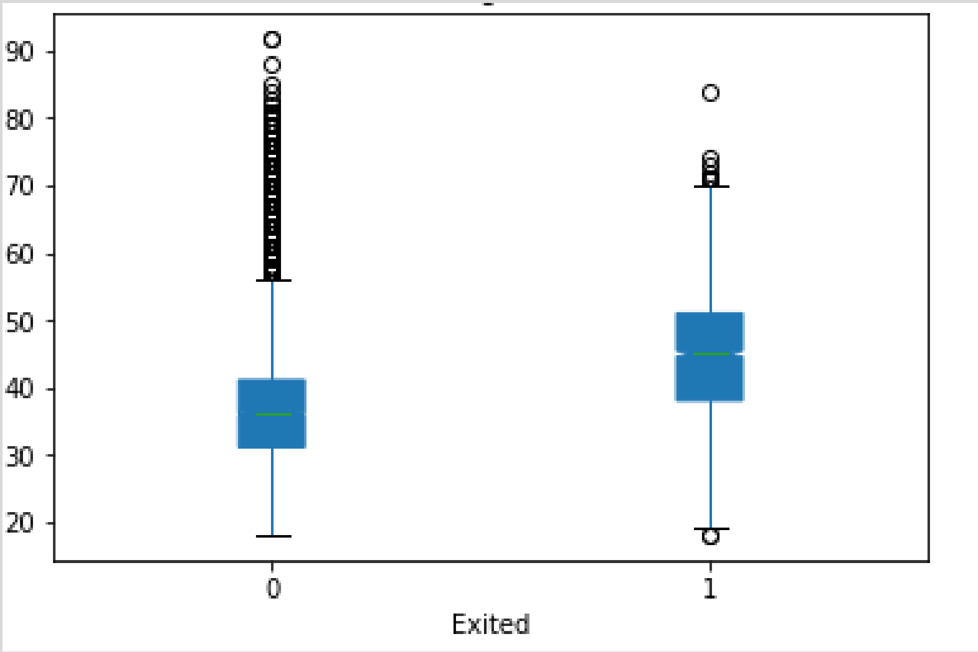


Retained Customers

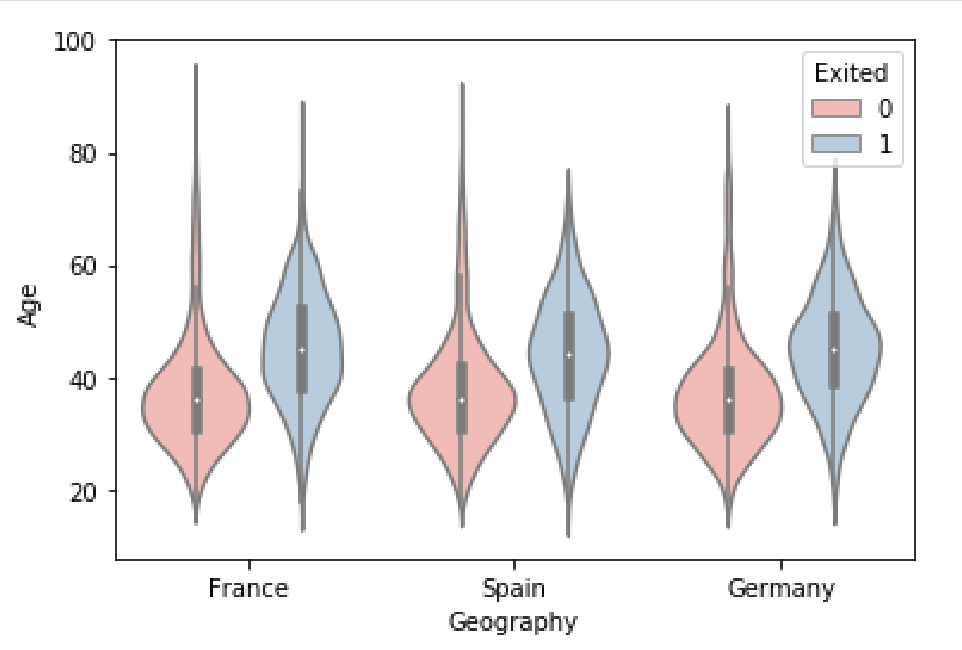


WHAT DID WE FIND?

Customers that churn are older on average. This holds true across all geographies in our sample as well.



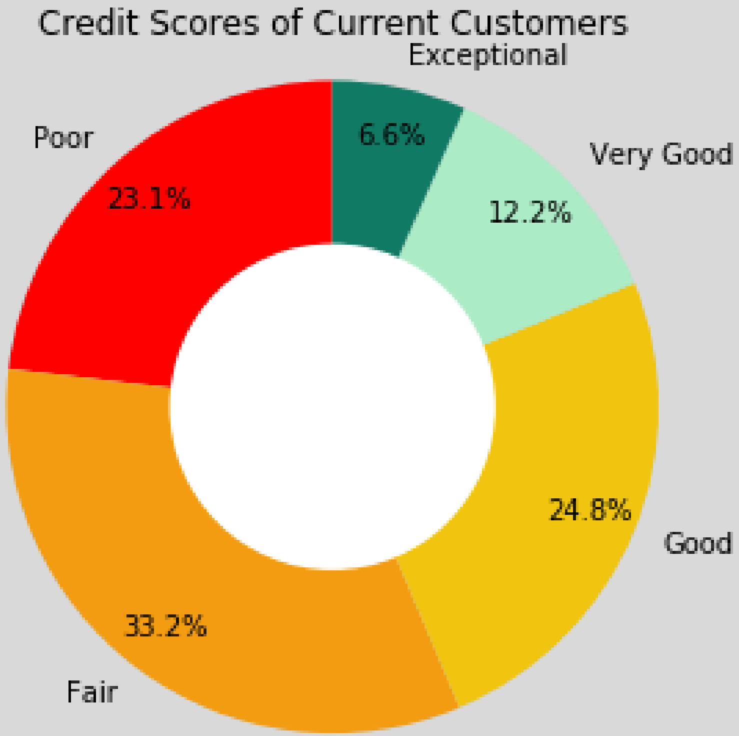
Churn vs Age

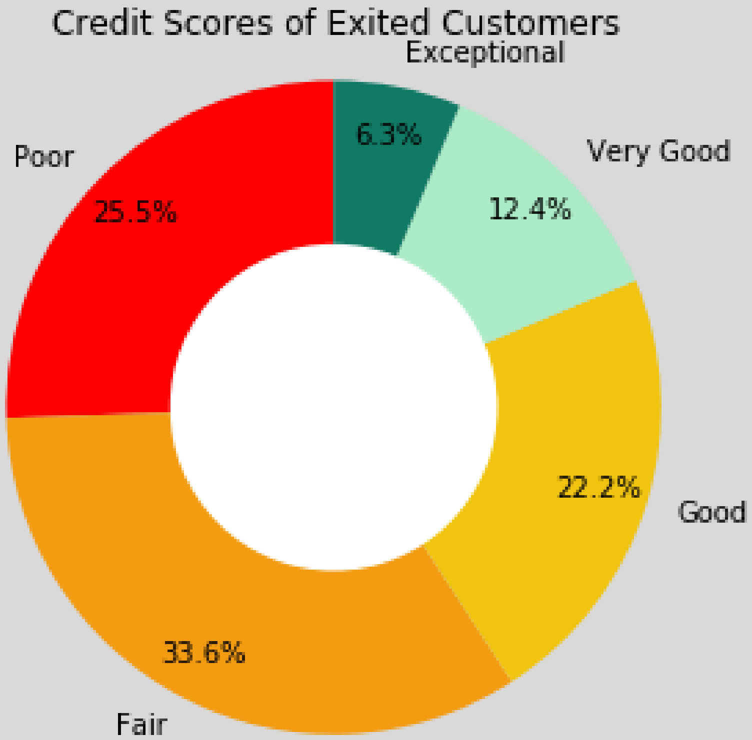


Churn vs Age by Geography

WHAT DID WE FIND?

Larger proportion of Exited customers have poor credit scores than current customers.

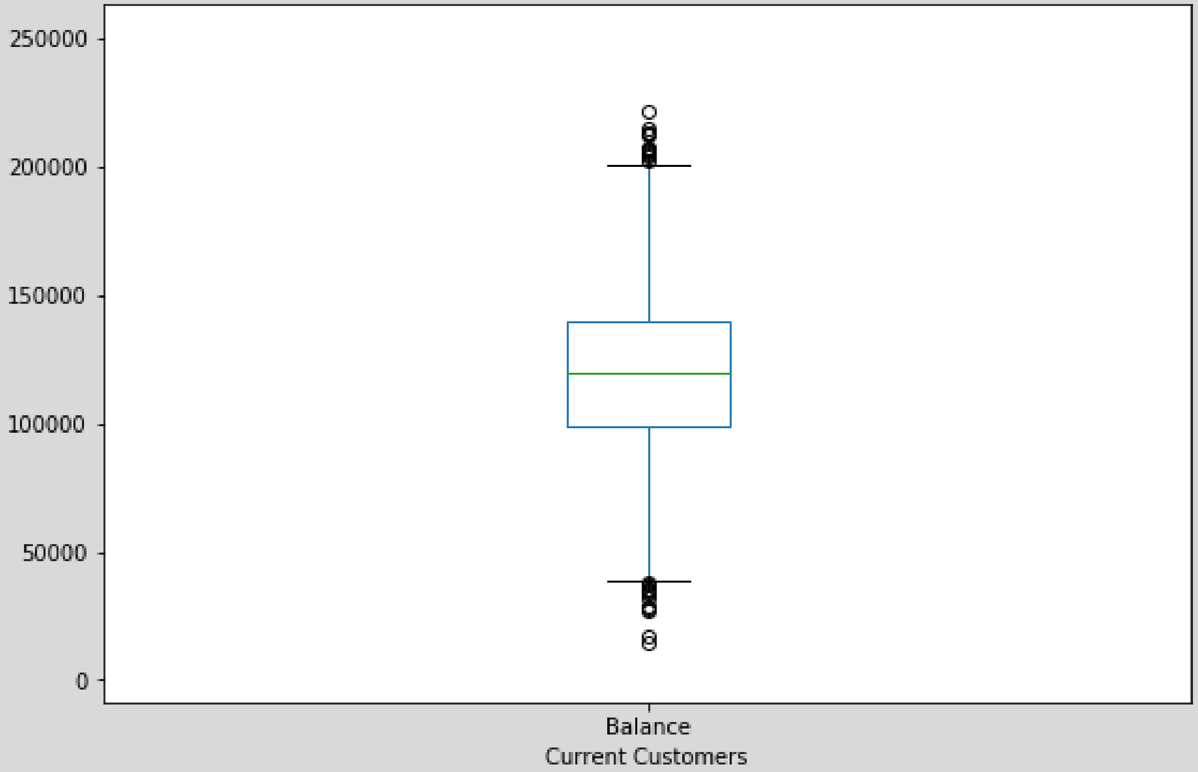


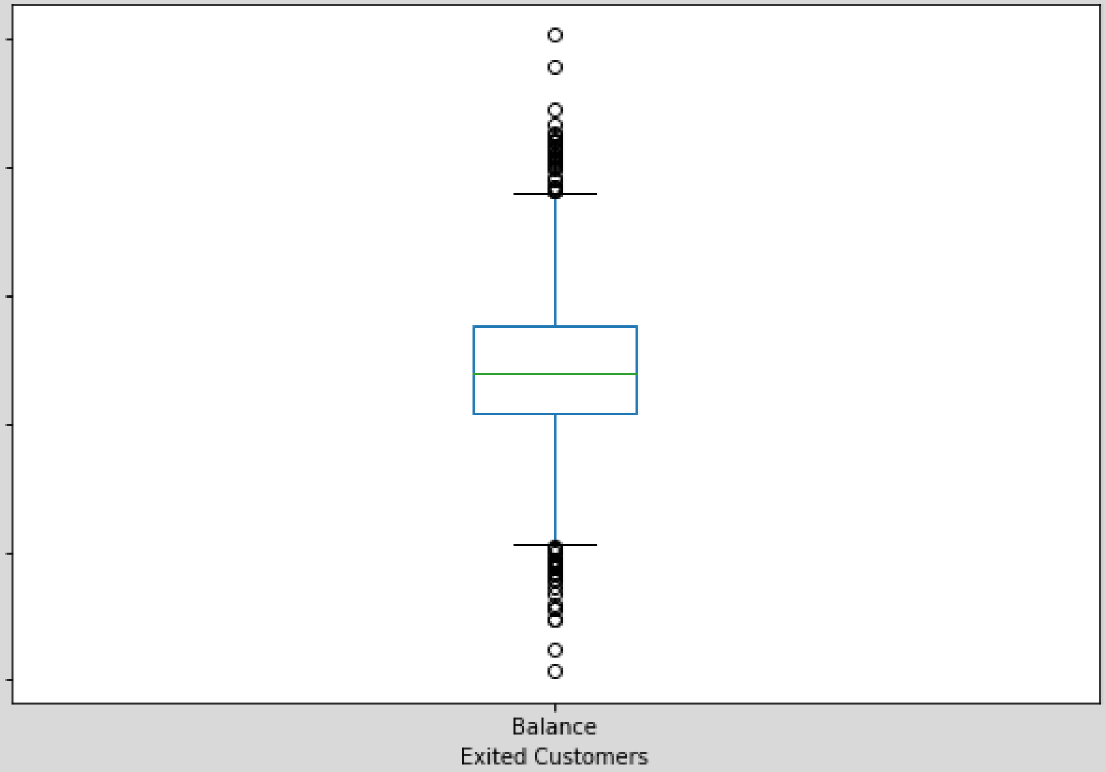


Note: Buckets for credit scores were determined based on the Expirian credit score scale.
Credit Buckets : (300-579 : Poor); (580-669 : Fair); (670-739 : Good); (740-799: Very Good) ; (800-850 : Exceptional)

WHAT DID WE FIND?

Current customers and exited customers have a similar distribution of account balances, but on average, exited customers have a higher account balances.

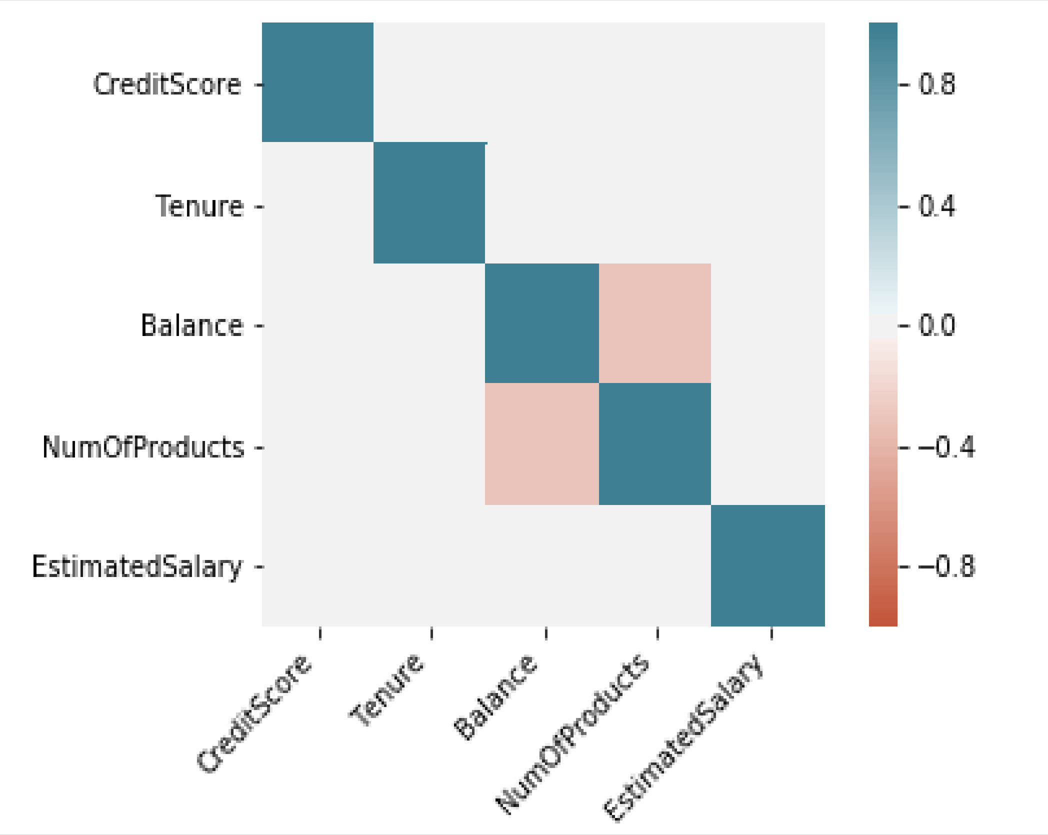




DATA PROCESSING

- Missing Value Treatment for Estimated Salary
- One-hot encoding for categorical variables:
    - Geography
    - Gender
- Removed customer specific details like name and IDs
- Correlation Matrix to check multicollinearity problem
- Training and Test Data Split

Correlation Matrix



MODELS

86% Random Forest

- Aggregated 1000 trees to get the average accuracy
- 83% accuracy with single decision tree
- Age, Estimated Salary, Credit Scores, Number of products turned out to be significant variables

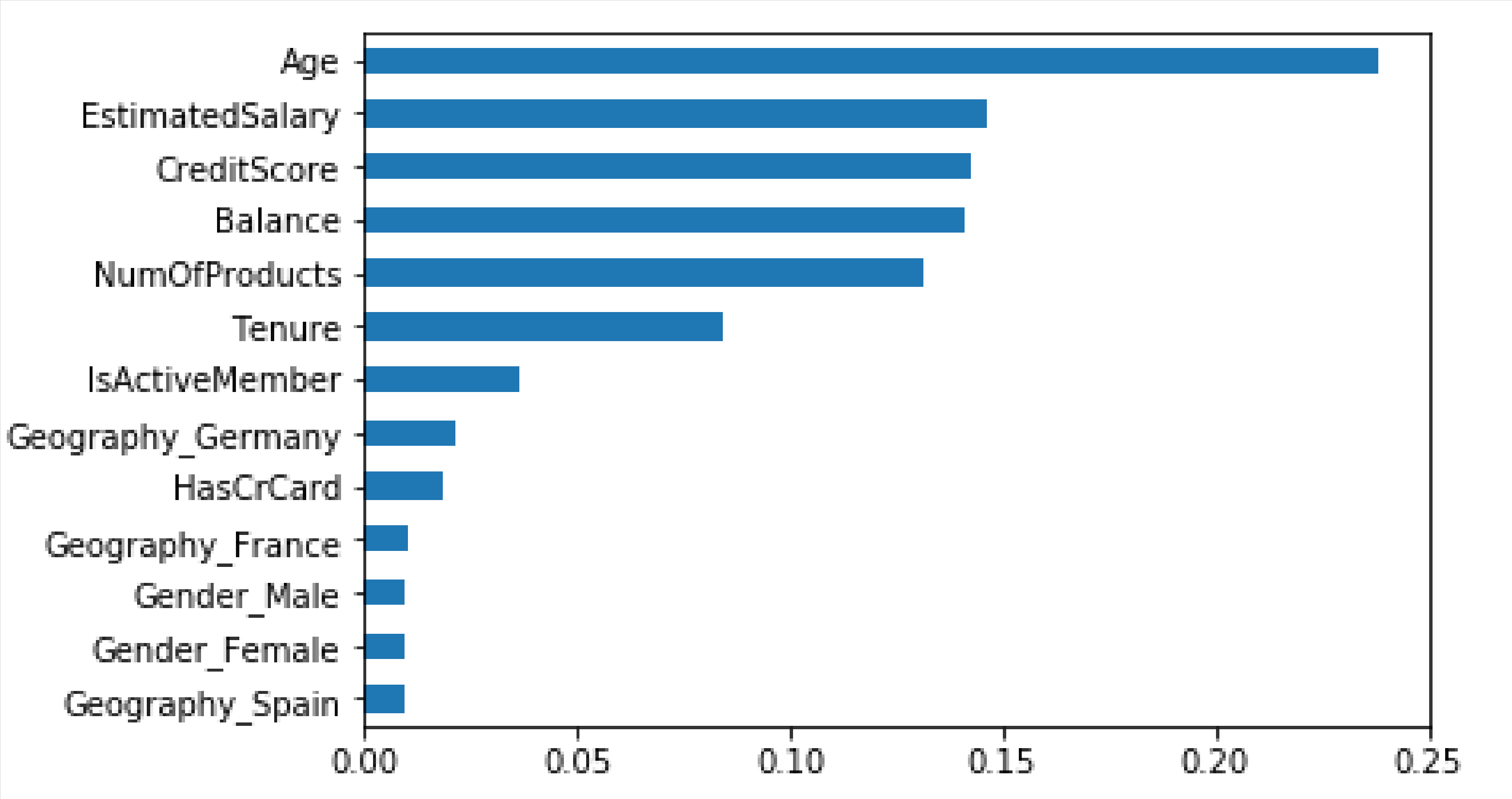
79% Logistic Regression

- Age, Estimated Salary, Credit Score turned out to be significant variables

70% K-Nearest Neighbours

- Found 20 nearest neighbours to be most optimal after plotting N values versus cross-validated accuracy scores

FEATURE IMPORTANCE



RECOMMENDATIONS

Key factors		Suggested implementations
01	Credit Score: Inversely proportional to Churn Rate	Increase awareness amongst customers on importance of maintaining a healthy credit score
02	Balance: Directly proportional to Churn Rate	Customized financial advisory and investment offerings to different segments of customers
03	Age: Directly proportional to Churn Rate	Launch Digital products and diverse schemes to woo customers from different age brackets

Questions?
Thank you!