Bank Customer Churn Prediction

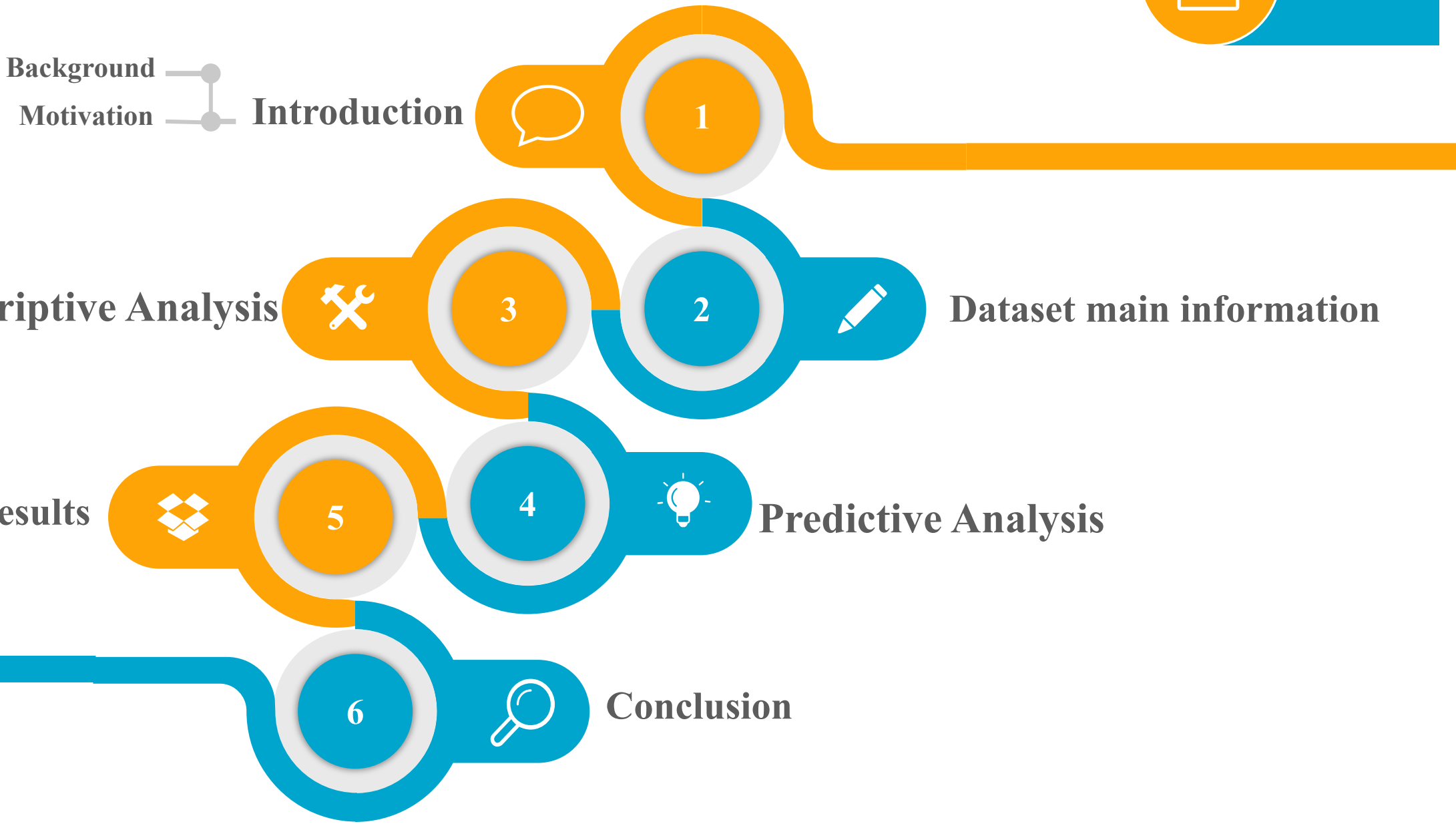
BAN 5573
Visual Analytics and Business Intelligence

Prepared by: Maryam Ahmadi , Dipendra Mainali

Advisor : Hamidreza Ahady Dolatsara

December 2022

Contents


Background
Motivation
Introduction
1
2
Dataset main information
3
Descriptive Analysis
4
Predictive Analysis
5
Results
6
Conclusion

Introduction

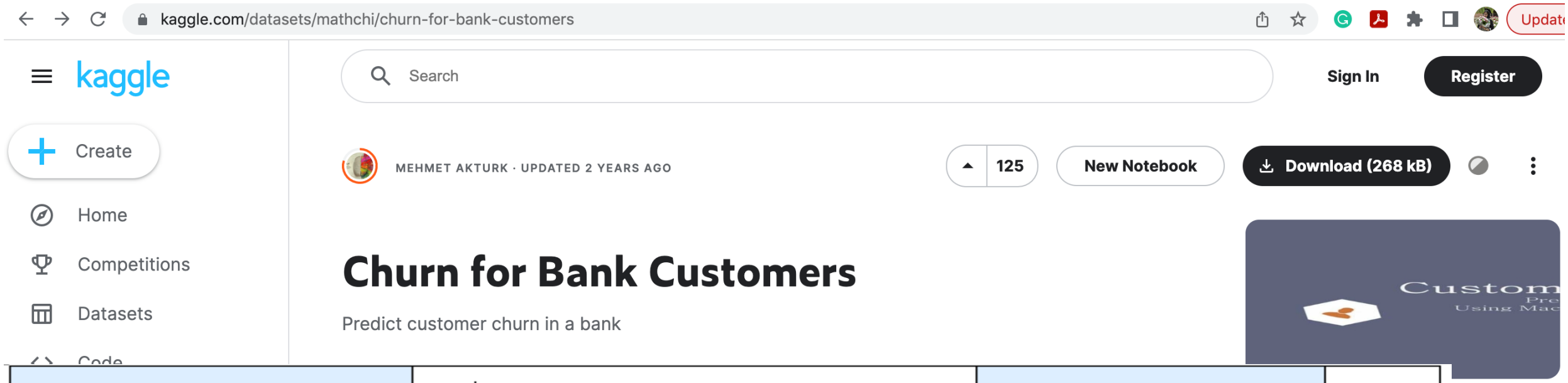
Background

- Churn customers are those who stopped being customers during certain periods
- Credit score, geography, gender, and age play a certain role in churn.

Motivation

- What are the values of churn of customers associated with different factors?
- Predict whether a customer will exit the bank or not, based on his/her characteristics.

Dataset main information

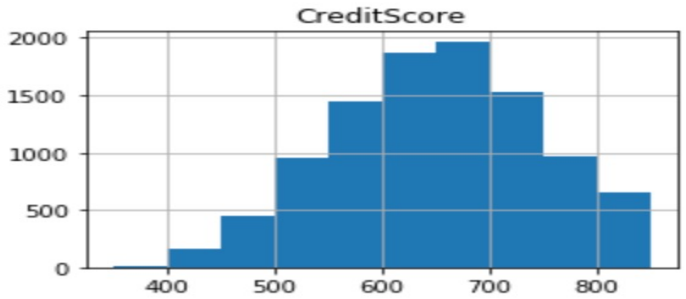


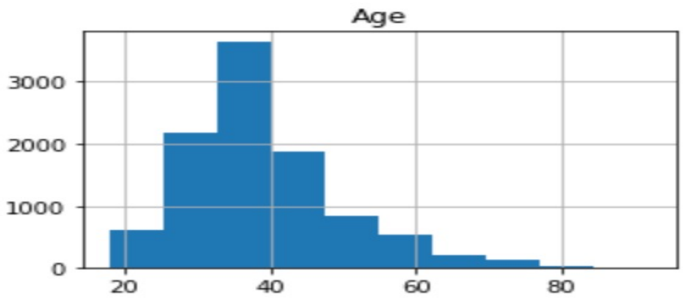
Data Set Characteristics:	Multivariate	**Number of Instances:**	10000
Attribute Characteristics:	Categorical, Numerical, Boolean, String	**Number of Attributes:**	14
Associated Tasks:	Classification	**Missing Values?**	No

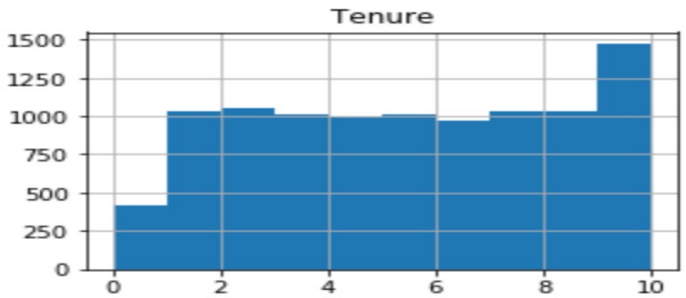
Dataset Features Description

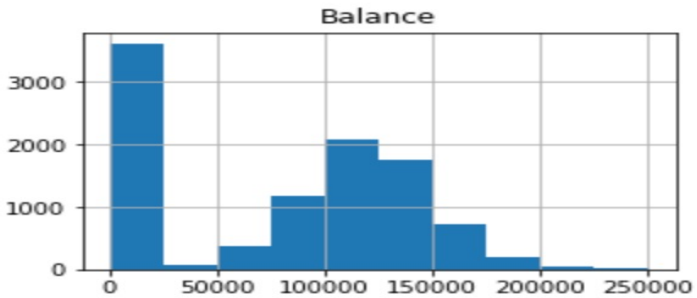
- RowNumber
- CustomerId
- Surname
- CreditScore
- Geography
- Gender
- Age
- Tenure
- Balance
- NumOfProducts
- HasCrCard
- IsActiveMember
- EstimatedSalary
- Exited

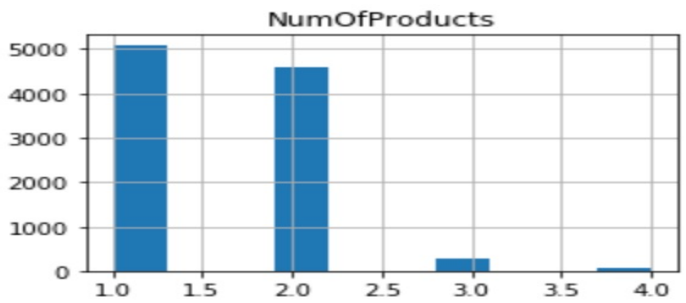
Descriptive Analysis

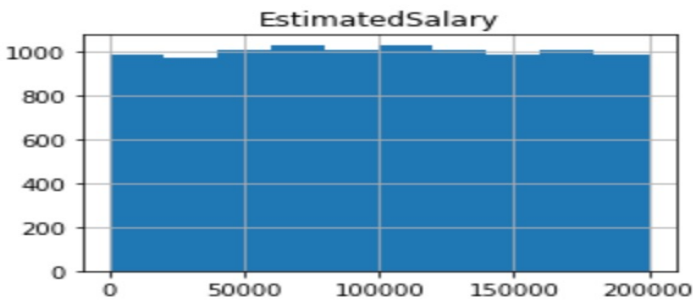


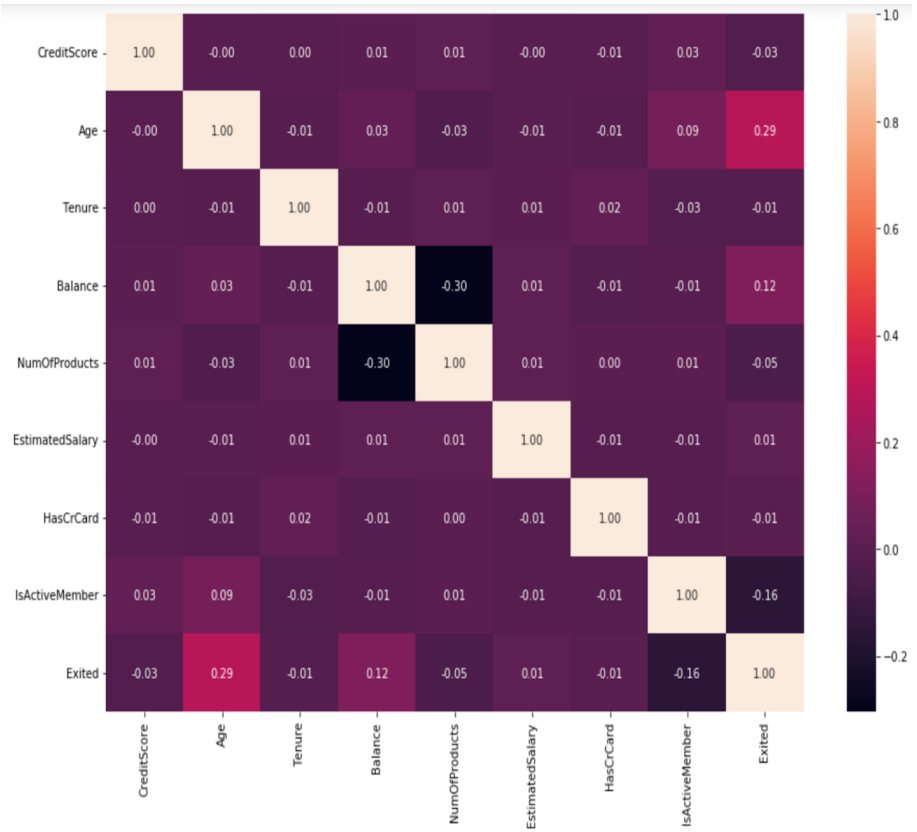




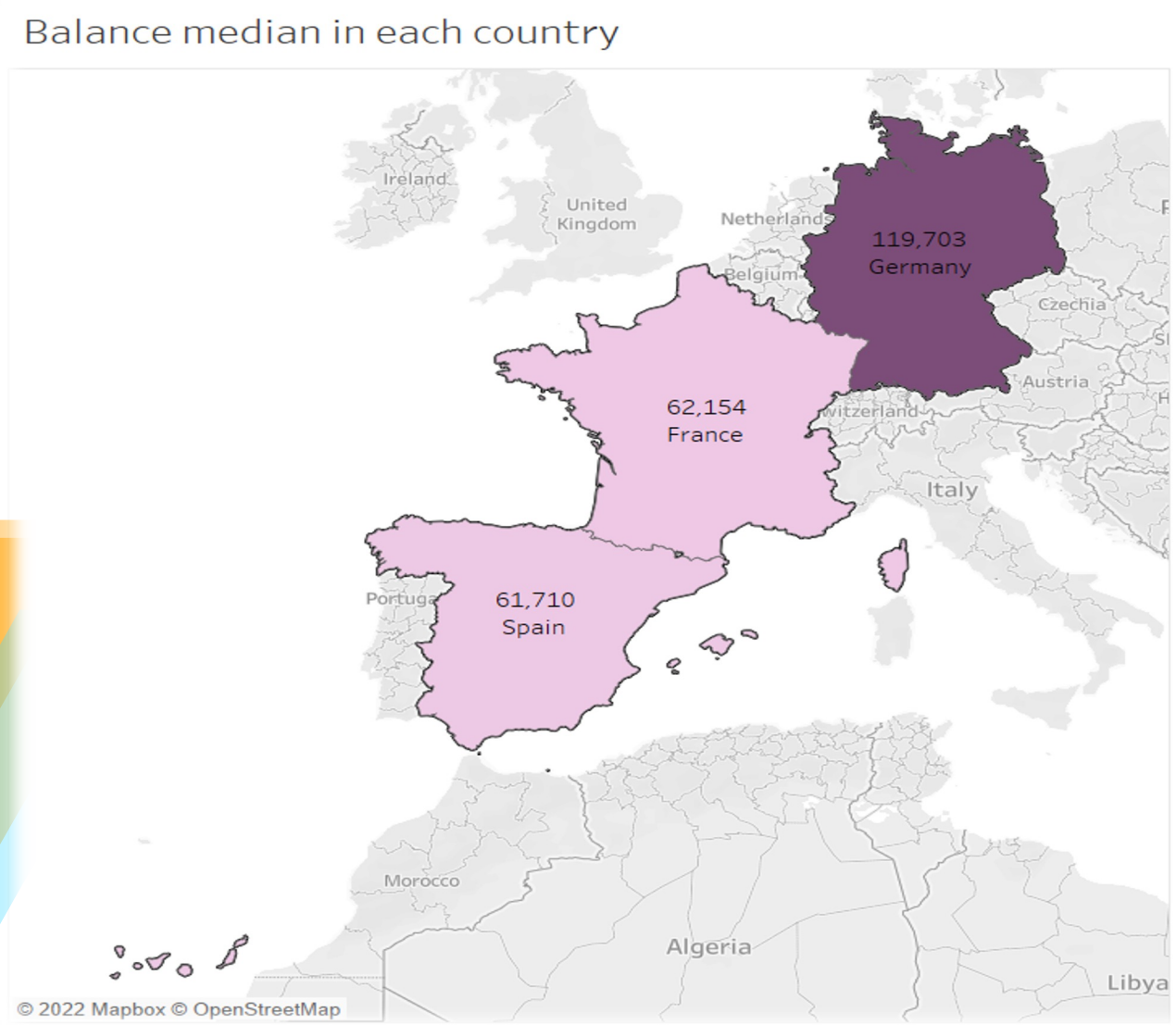


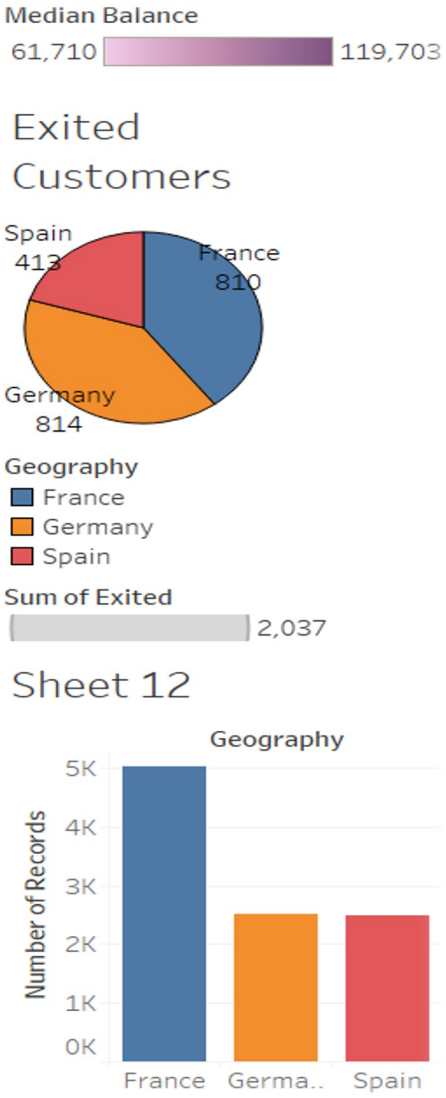




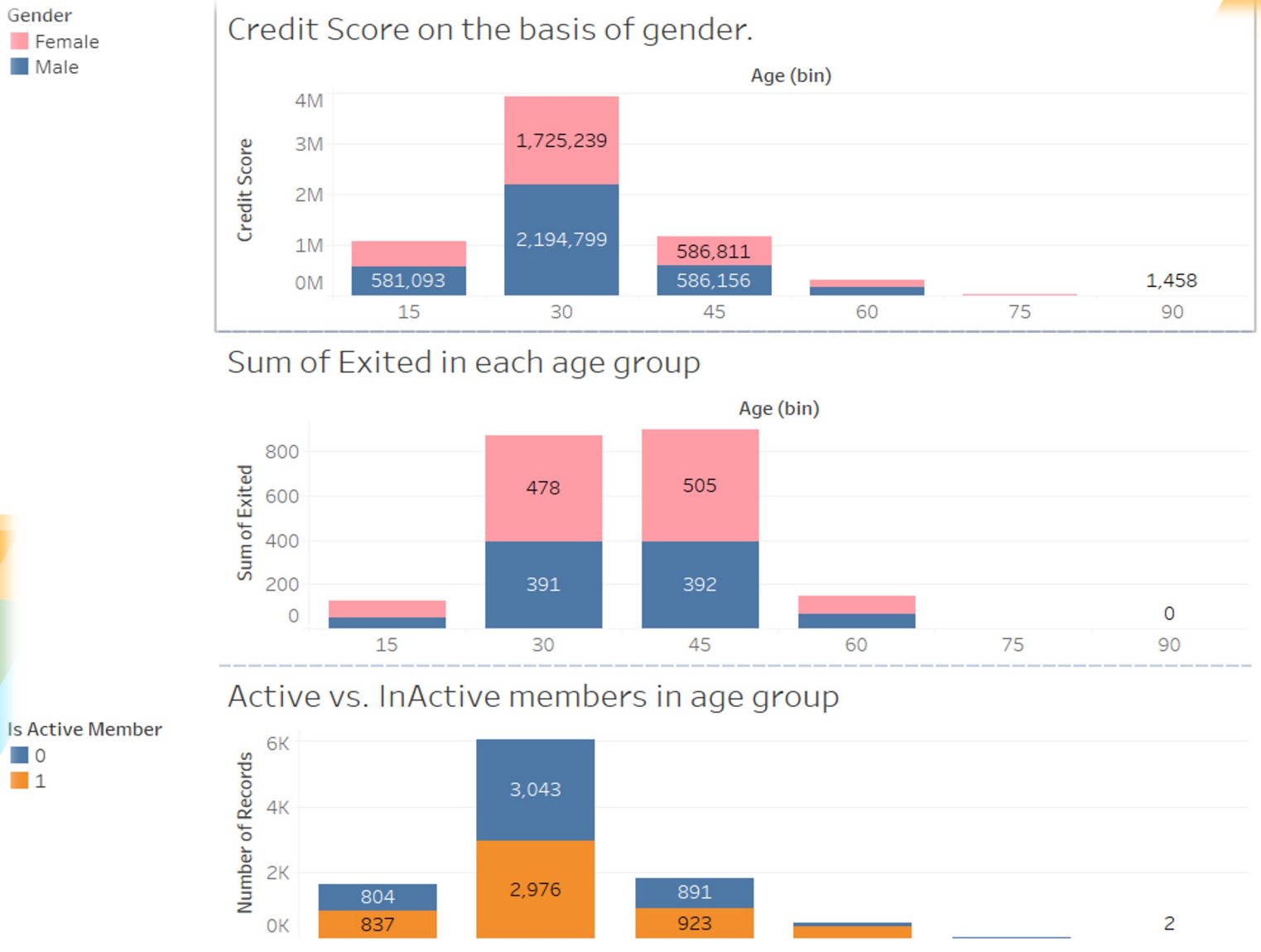


Visualization of Customers' Information in Different Countries

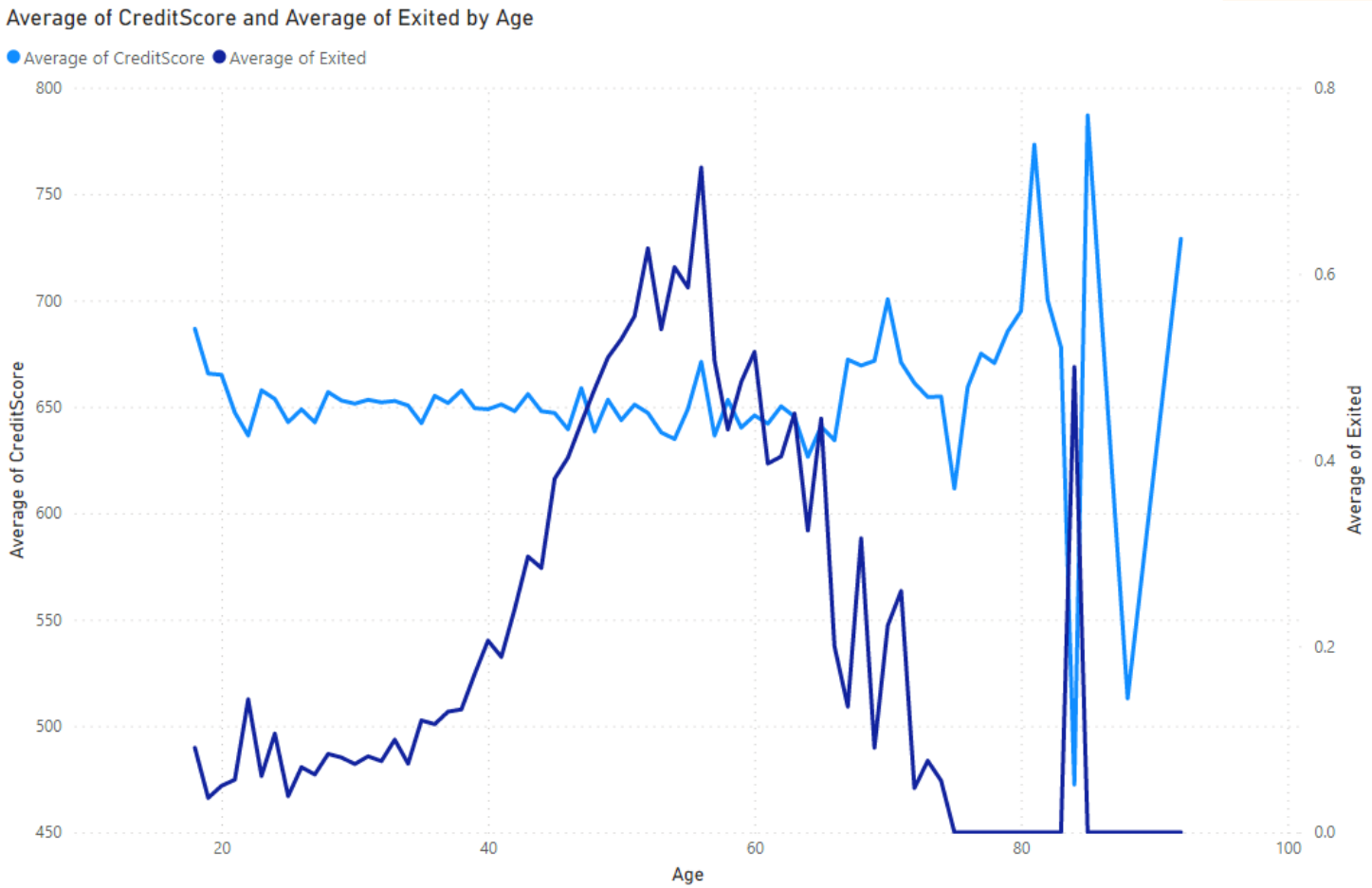




Customer Behavioral Pattern in Different Age Groups



Average credit score and average exited age

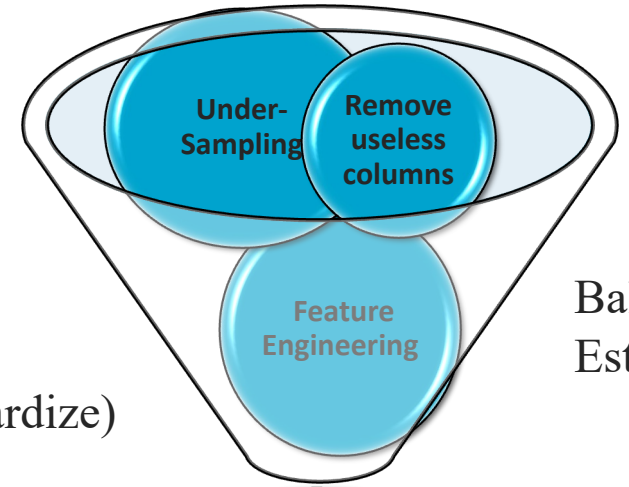


Data Preprocessing

0	7963
1	2037

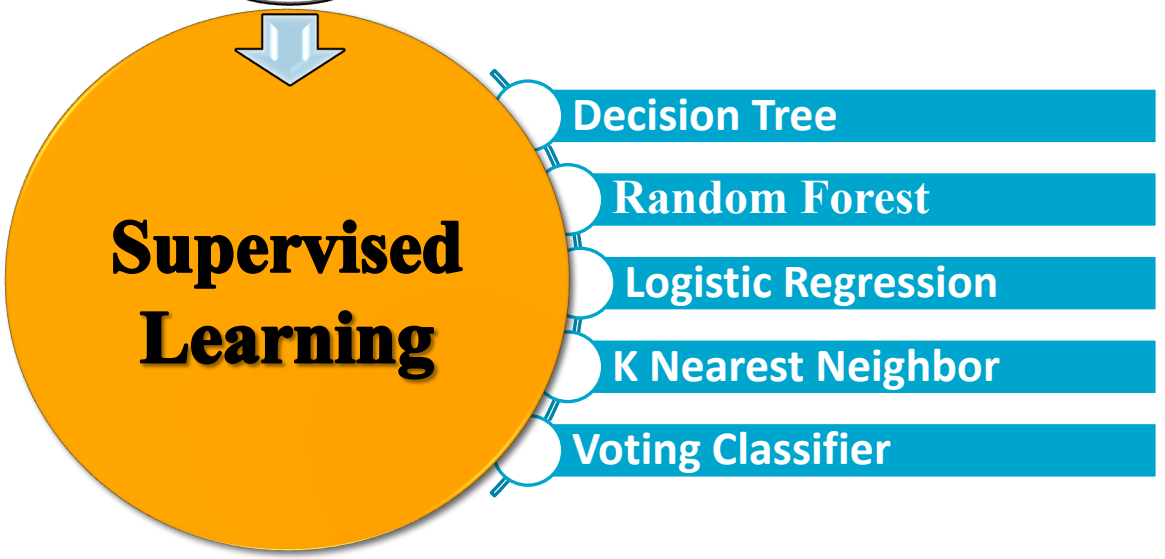
0	2037
1	2037

Scale (Standardize)



Columns:
1- RowNumber, 2- CustomerId,
3- Surname

Balance: Poor, Good, Excelent, Top
EstimatedSalary: Poor, Good, Excelent, Top

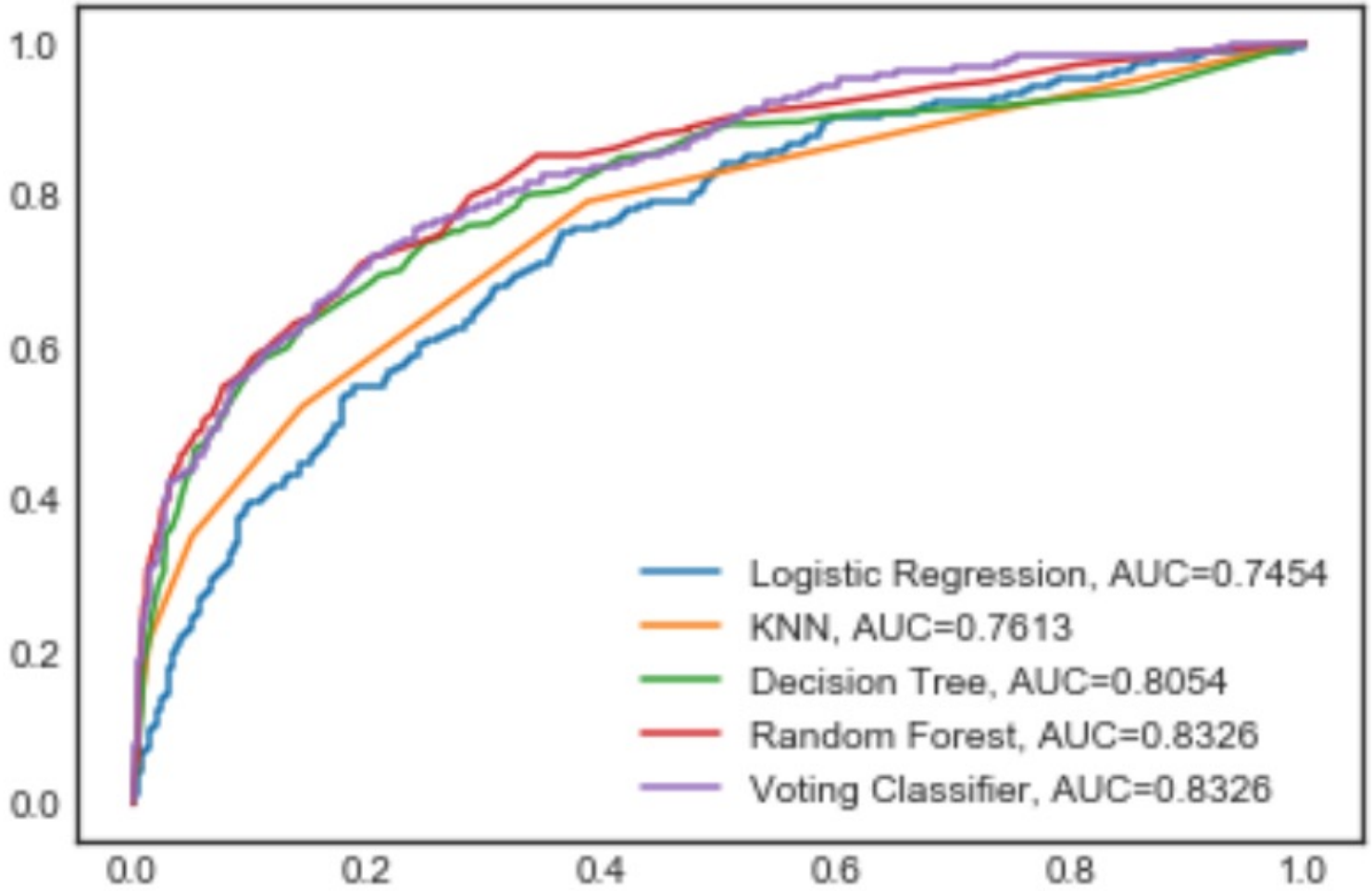


Predictive Analysis

Model	Accuracy	Precision	Recall	F1-score
Decision Tree	0.85	0.79	0.69	0.72
Random Forest	**0.86**	0.81	**0.70**	**0.73**
Logistic Regression	0.80	0.69	0.56	0.56
KNN	0.83	0.75	0.65	0.68
Voting Classifier	0.85	**0.82**	0.67	0.71

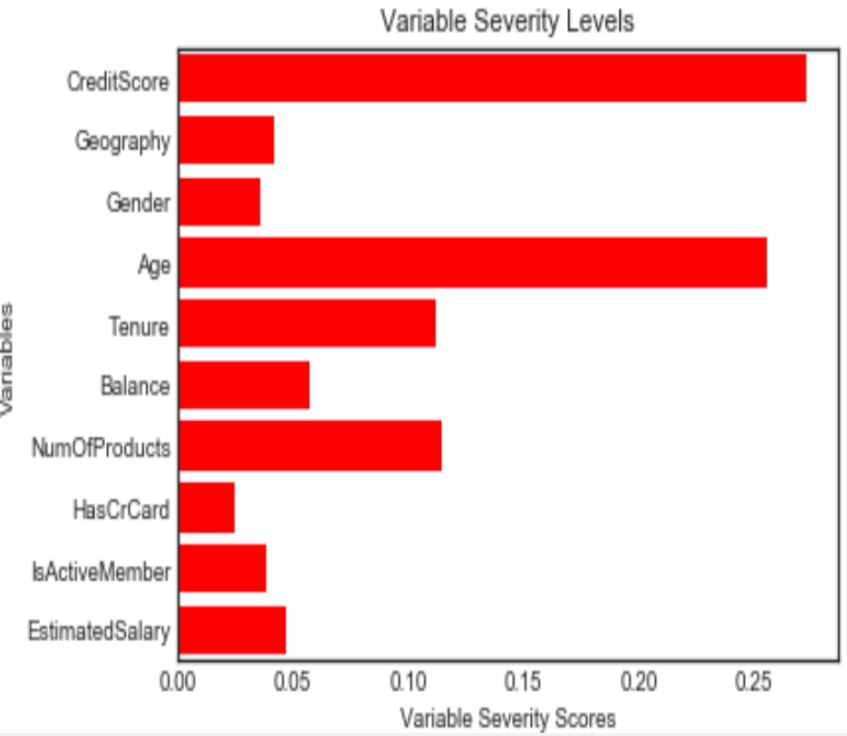
Predictive Analysis

Compare Models Using the area under the ROC curve

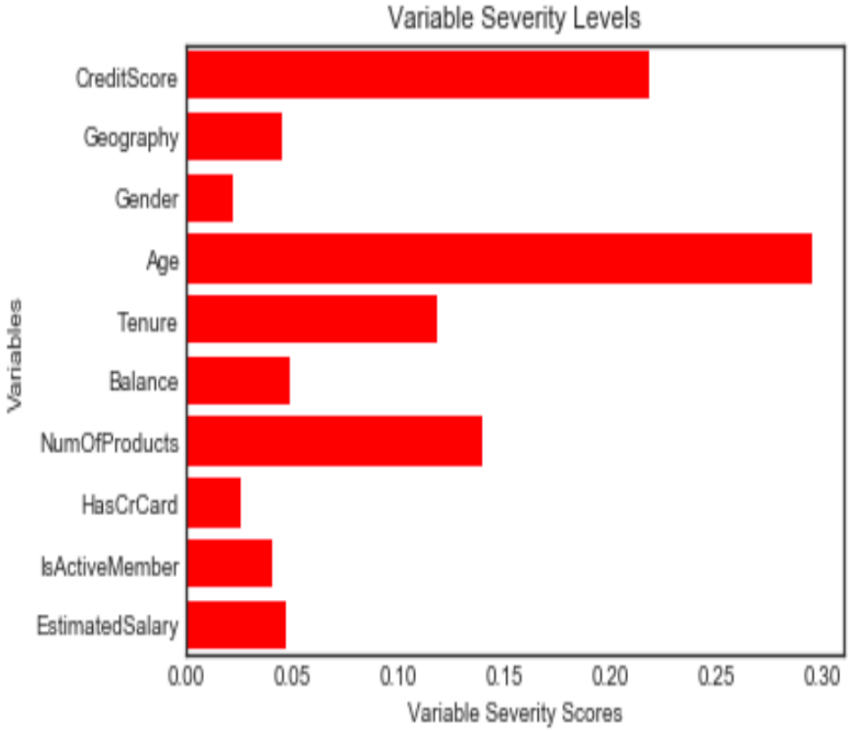


Feature Importance

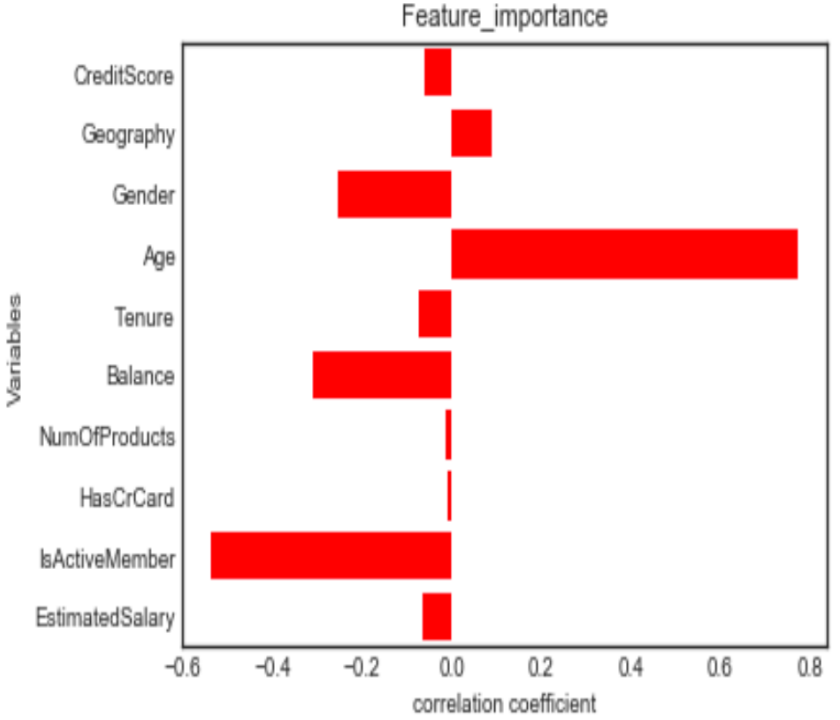
Decision Tree



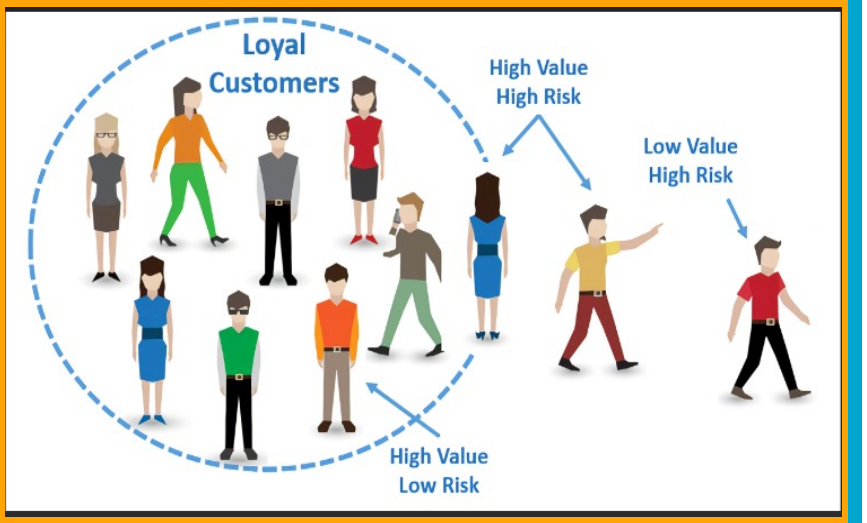
Random Forest



Logistic Regression



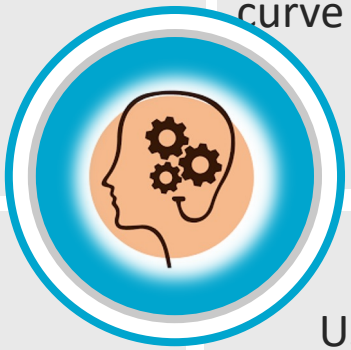
Focus on High Value - High Risk Customers



Result/Conclusion

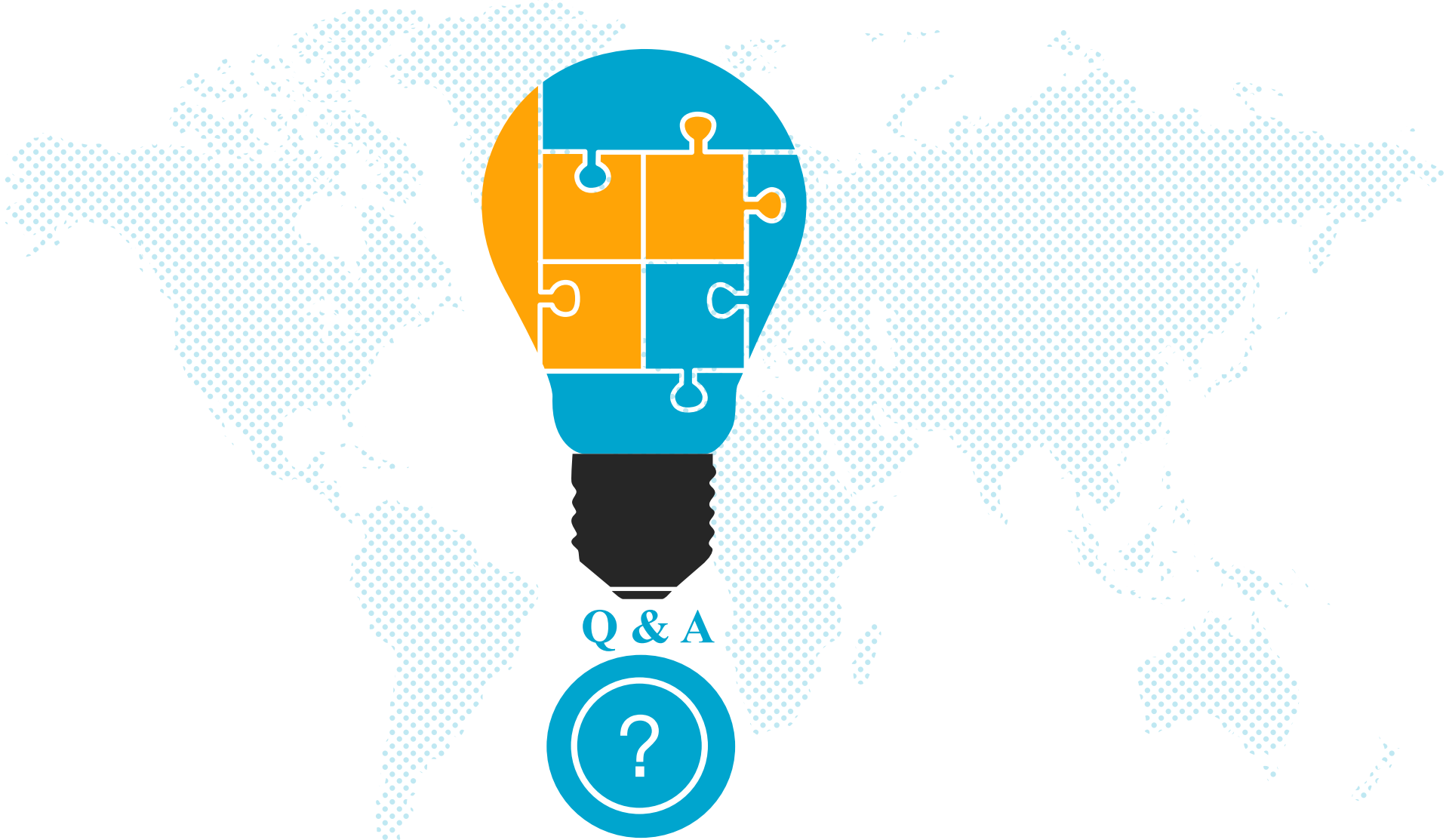
Customers' age, credit scores, and the number of products they use are the most important features for customer churn prediction.

Among models, Random Forest and voting model have the best performance based on accuracy, precision, recall, f-1 score, and area under the curve measurements.



Based on visualization, Almost half of the customers from Germany leave the bank. Therefore, we should investigate these two groups and find the reason so we can design proper strategies to prevent them from exiting

Using machine learning algorithms, we were able to predict 70% of exited customers. Offering promotions to them makes it possible to encourage them to stay.


Q & A

THANK YOU