

# *Bank Churn Prediction*

*Shiyi (Ashley) Yue*

*May. 6, 2023*

# *Problem Definition*

## What we are facing

As the online third-party pay method is getting more and more popular, the traditional bank company wants to utilize their customer data to increase the customer retention rate.

## What to solve

This project is mainly about comparing the performance of different machine learning models in predicting customer churning rate based on the bank company's data

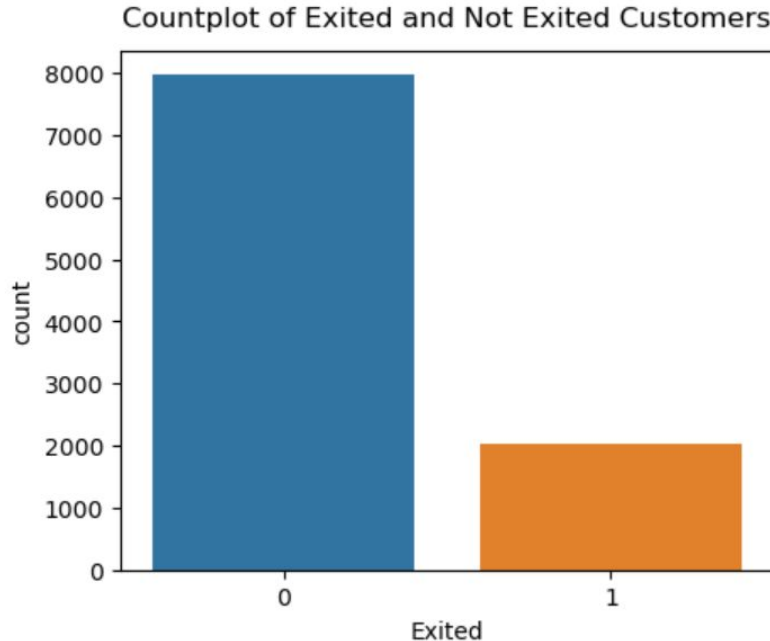
## What to expect

Good predictions and precise customer profiling help the company to segment customers better and improve the effectiveness of specific promotion or advertising for certain customer segments

## *Business Question*

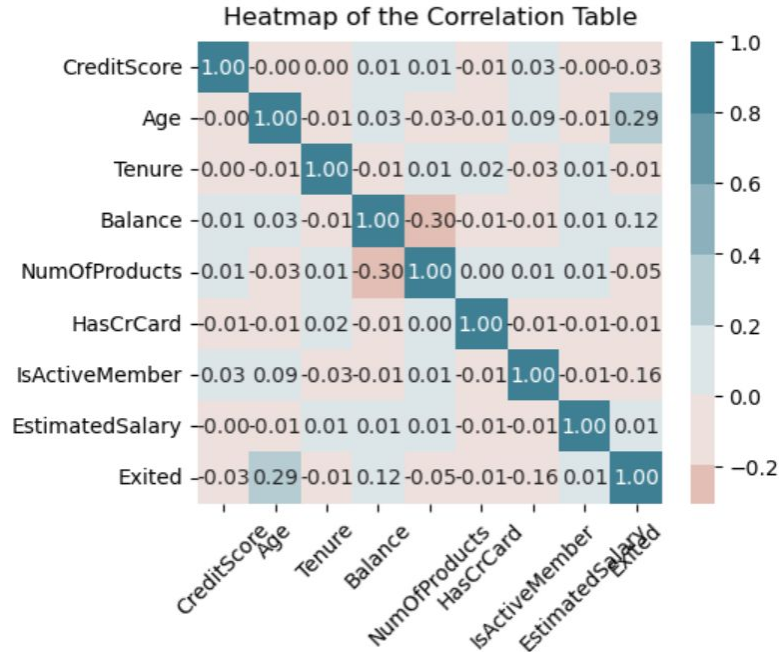
- What model would be the best fit to predict customer churning rate?
- What factors are most important to the predictive model?
- What customers tend to churn and what customers tend to stay?

# *Exploratory Data Analysis*



- There are about 4 times of negatives than positives.
- The dataset is not very balanced so accuracy score may not be the decisive metric. Other metrics should also be considered.

# Exploratory Data Analysis



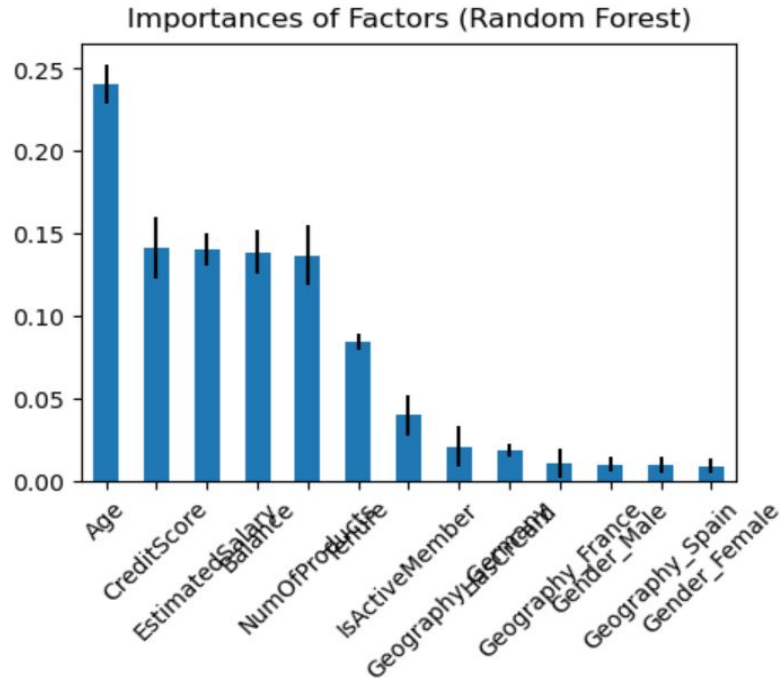
- The correlation heatmap doesn't indicate any strong linear relationship between "Exited" and any other variable.
- Desirably, there is no even moderate linear relationship between x variables, which indicates the absence of multilinearity.

## Insights

	Logistic Regression	K Nearest Neighbor	Random Forest
Accuracy	0.81	0.83	0.86
Recall	0.29	0.41	0.45
Precision	0.56	0.62	0.75
AUC	0.62	0.67	0.70

- Random Forest outstrips Logistic Regression and KNN in terms of accuracy score, recall score, precision score, and AUC. Therefore, undoubtedly, Random Forest should be the best fit to predict customer churn.
- We want to identify customers who tend to churn so we care more about the true positive rate, i.e. recall.

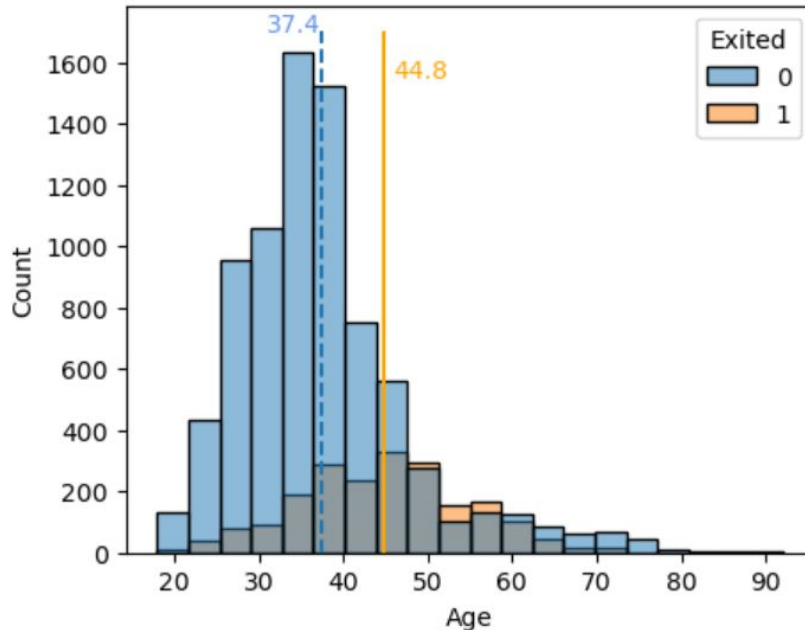
# Insights



- Customer age is the most important factor when predicting customer's churning rate.
- Estimated salary, credit score, balance amount, and the number of products the customer purchased are also important in predicting the churning rate.

# Business Action

Distributions of Age separated by Exited or not

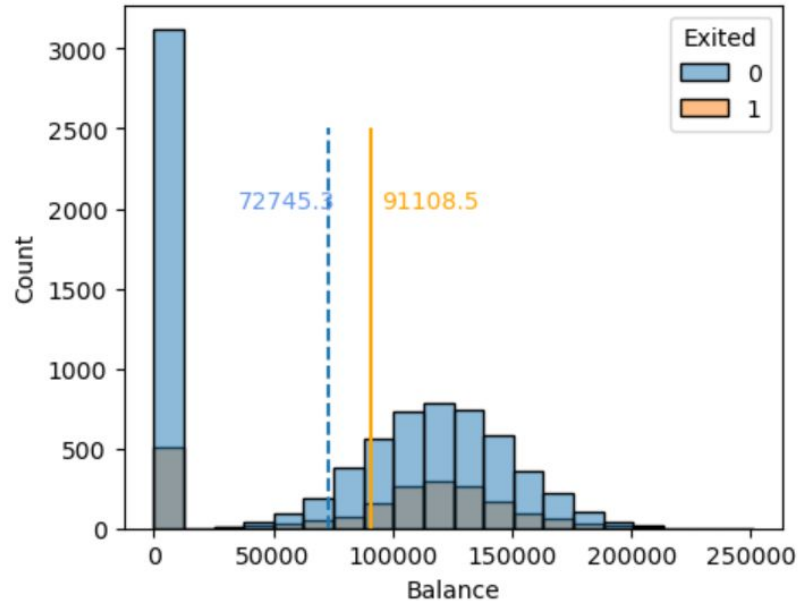


- The mean age of customers who churned is 44.8 while the number is 37.4 for customers who did not churn.
- Customers whose age ranges from 30-40 are less likely to churn while customers whose age is higher than 50 are more likely to churn.
- The bank may target on customers who aged 30-40 as they are in the most loyal customer segment.



# Business Action

Distributions of Balance separated by Exited or not



- Customers whose balance amount is zero are more likely to churn.
- The bank may incentivize customers to deposit money in their account to increase customer retention

## *Business Outcome*

- Random Forest would be the best predictive model in this scenario.
- Customer age is the most important factor when predicting the churning rate. Estimated salary, credit score, balance amount, and the number of products the customer purchased are also important in predicting the churning rate.
- The bank may target on customers who aged 30-40 as they are in the most loyal customer segment. Besides, The bank may incentivize customers to deposit money in their account to increase customer retention



*Thank You for Listening*