bank customer churn

--- Final-bootcamp project

Data source:

This dataset from a U.S. multinational bank .

Links:

https://www.kaggle.com/datasets/shantanu
dhakadd/bank-customer-churn-prediction

Credit Score
Geography
Gender
Age
Tenure
Balance
Number Of Products
Has Cr Card
Is Active Member
Estimated Salary

Exited (0/1)

https://www.kaggle.com/datasets/shantanudhakadd/bank-customer-churn-prediction
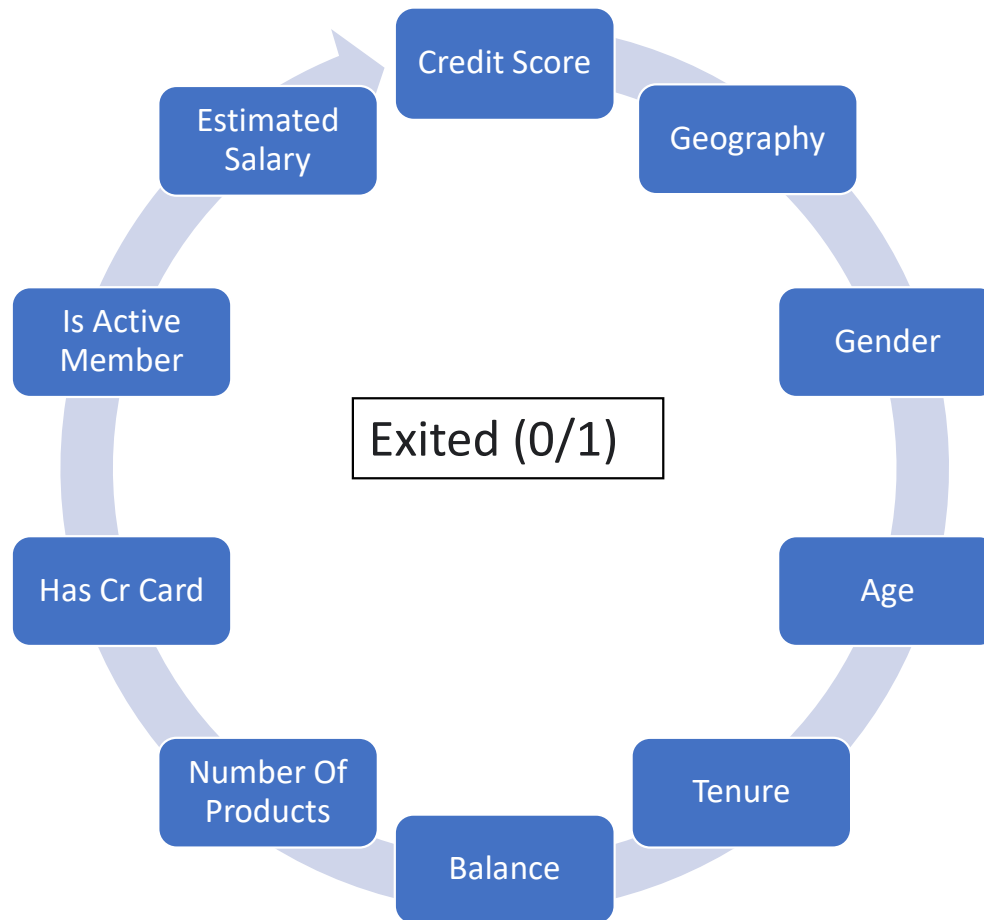
Questions

- What is the percentage of customers who leave as a percentage of all banking customers?

- In which of these countries are customers most likely to leave?

- Among bank customers, are men or women more likely to leave?

- Is it true that customers with more deposits are less likely to leave?

1. What is the percentage of customers who leave as a percentage of all banking customers?

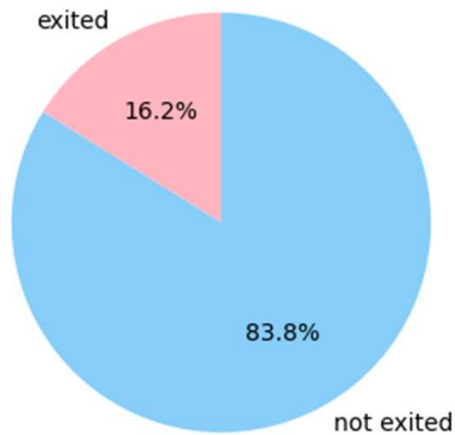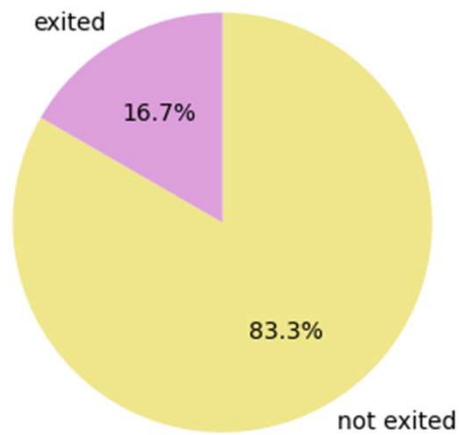**Percentage of bank customers leaving**
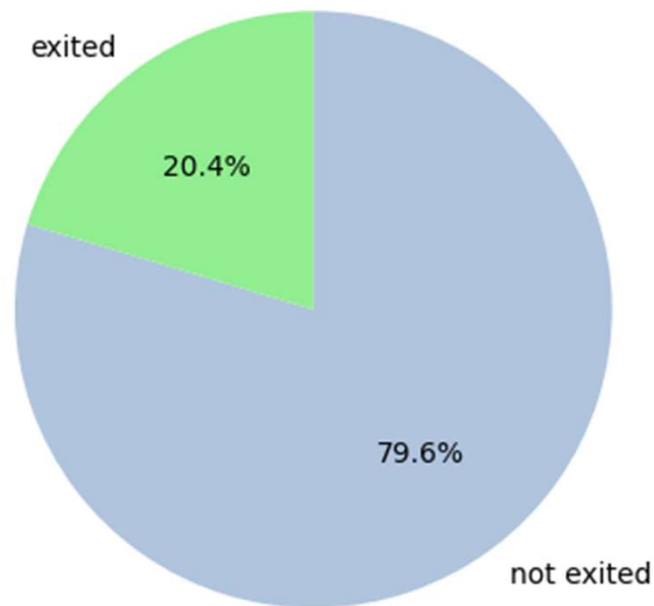
Answer 1:

about 80% of bank customers choose to stay, and only 20% of customers choose to leave.

2. In which of these countries are customers most likely to leave?

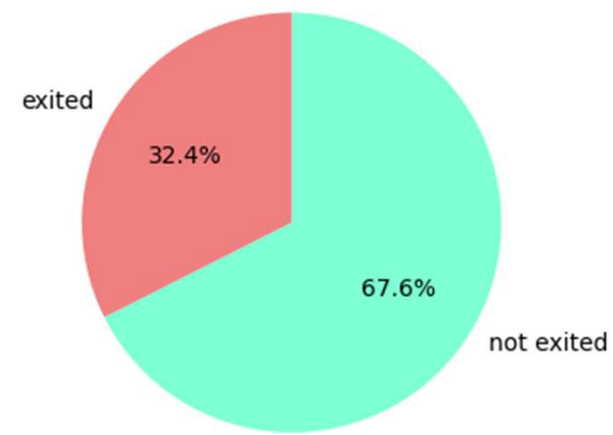## Percentage of bank customers leaving from 3 countries

exited

16.2%

83.8%

not exited

France

exited

16.7%

83.3%

not exited

Spain

exited

32.4%
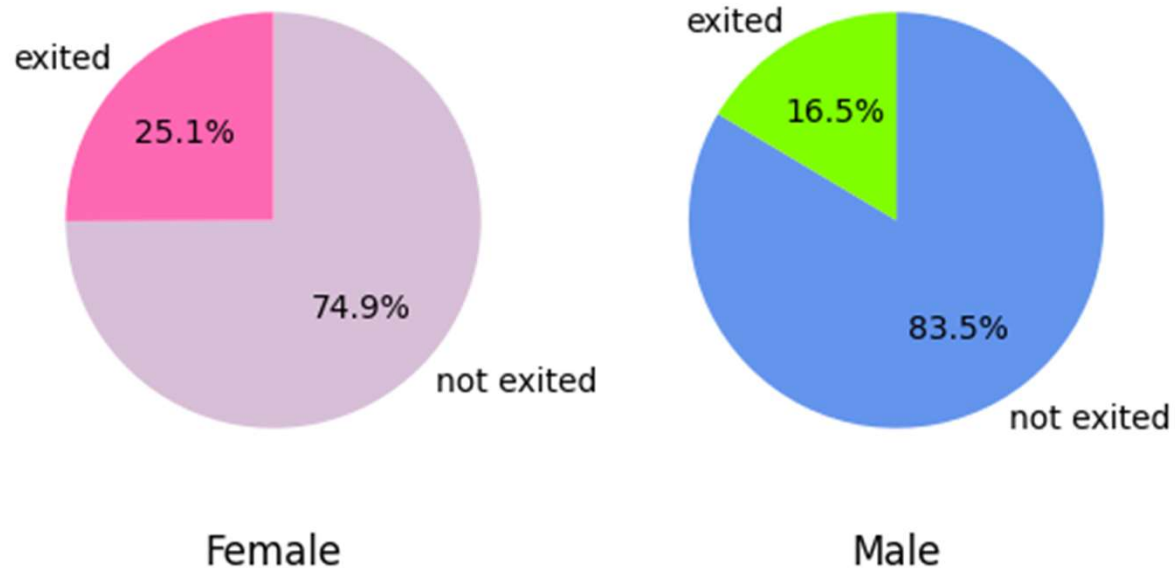
67.6%

not exited

Germany

Answer 2:

Customers from Germany are most likely to leave.

3. Among bank customers, are men or women more likely to leave?

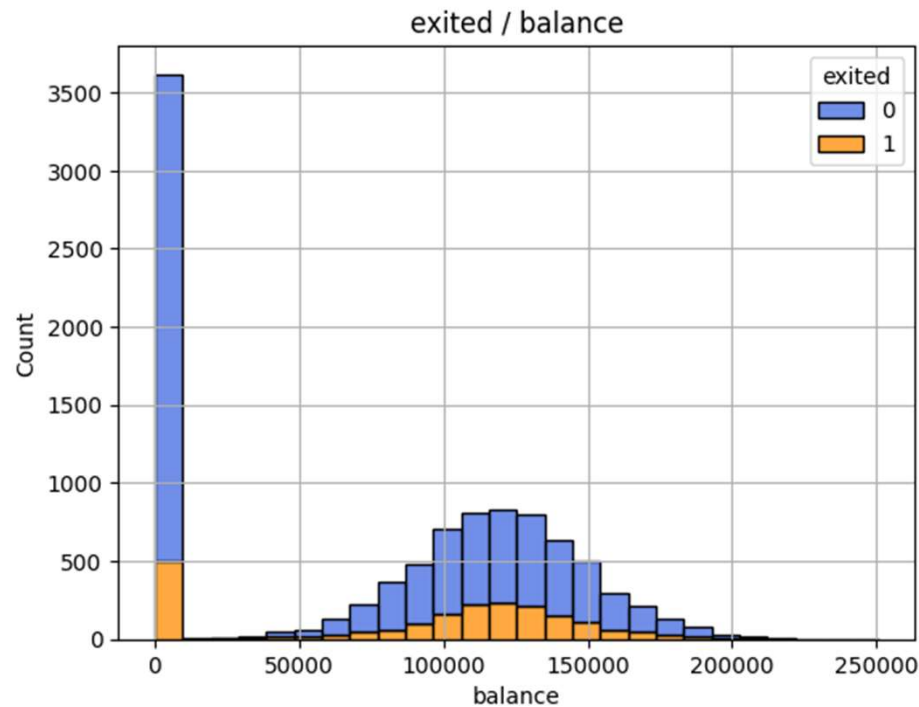### Percentage of male and female customers leaving

Female   Male

Answer 3:

More than 25% of **female customers** choose to leave, while only 16.5% of **male customers**. **Women** are more likely to leave.

4. Is it true that customers with more balance are less likely to leave?

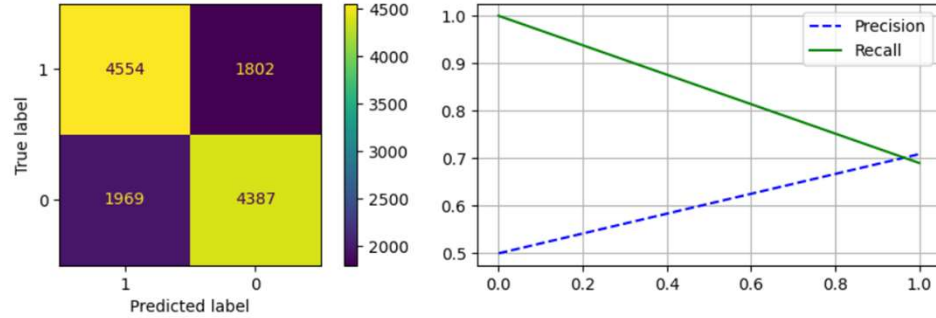exited / balance

Answer 4:

Many bank customers have no balance in their accounts.

As the balance increases, the ratio of customers who stay to customers who leave does not change.

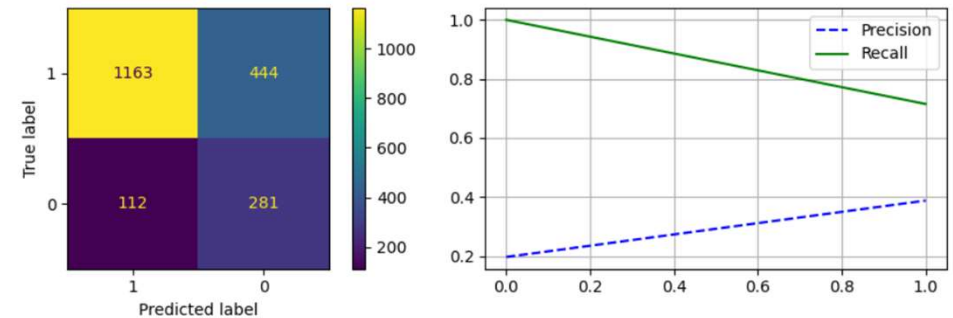So the balance has basically no effect on the customer leaving

LogisticRegression

LogisticRegression / Train set

	precision	recall	f1-score	support
0	0.70	0.72	0.71	6356
1	0.71	0.69	0.70	6356
accuracy			0.70	12712
macro avg	0.70	0.70	0.70	12712
weighted avg	0.70	0.70	0.70	12712

cohen-kappa-score: 0.40670232850849586

LogisticRegression / Test set

	precision	recall	f1-score	support
0	0.91	0.72	0.81	1607
1	0.39	0.72	0.50	393
accuracy			0.72	2000
macro avg	0.65	0.72	0.65	2000
weighted avg	0.81	0.72	0.75	2000

cohen-kappa-score: 0.3325931038621973

KNeighborsClassifier

KNeighborsClassifier / Train set
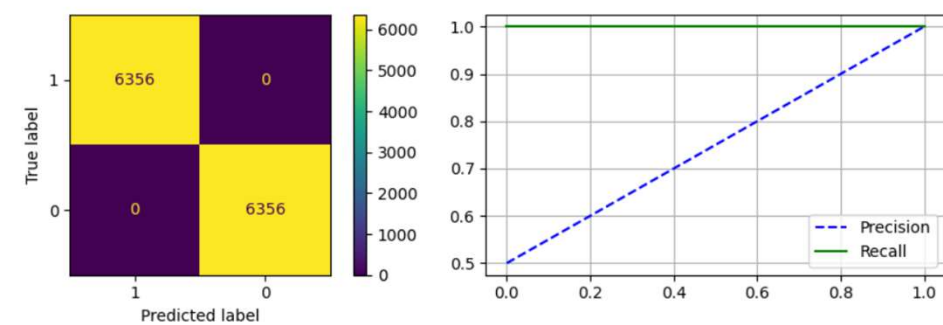
	precision	recall	f1-score	support
0	1.00	1.00	1.00	6356
1	1.00	1.00	1.00	6356
accuracy			1.00	12712
macro avg	1.00	1.00	1.00	12712
weighted avg	1.00	1.00	1.00	12712

cohen-kappa-score: 1.0

KNeighborsClassifier / Test set
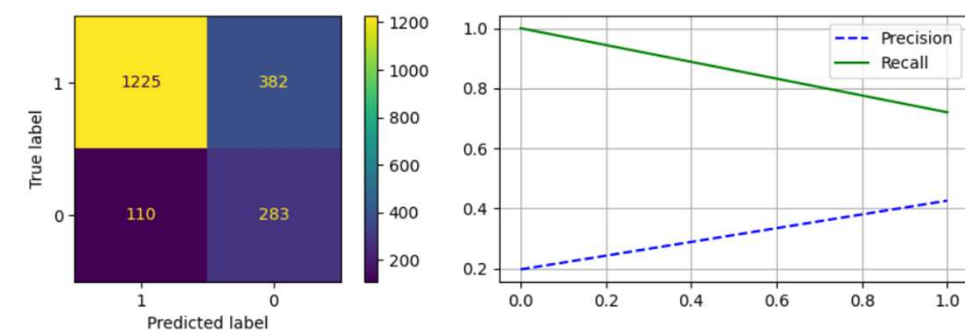
	precision	recall	f1-score	support
0	0.92	0.76	0.83	1607
1	0.43	0.72	0.53	393
accuracy			0.75	2000
macro avg	0.67	0.74	0.68	2000
weighted avg	0.82	0.75	0.77	2000

cohen-kappa-score: 0.3824177341509185

RandomForestClassifier

RandomForestClassifier / Train set

	precision	recall	f1-score	support
0	1.00	0.99	0.99	6356
1	0.99	1.00	0.99	6356
accuracy			0.99	12712
macro avg	0.99	0.99	0.99	12712
weighted avg	0.99	0.99	0.99	12712

cohen-kappa-score: 0.9896161107614853

\RandomForestClassifier / Test set

	precision	recall	f1-score	support
0	0.90	0.92	0.91	1607
1	0.63	0.59	0.61	393
accuracy			0.85	2000
macro avg	0.77	0.75	0.76	2000
weighted avg	0.85	0.85	0.85	2000

cohen-kappa-score: 0.5202461007839763
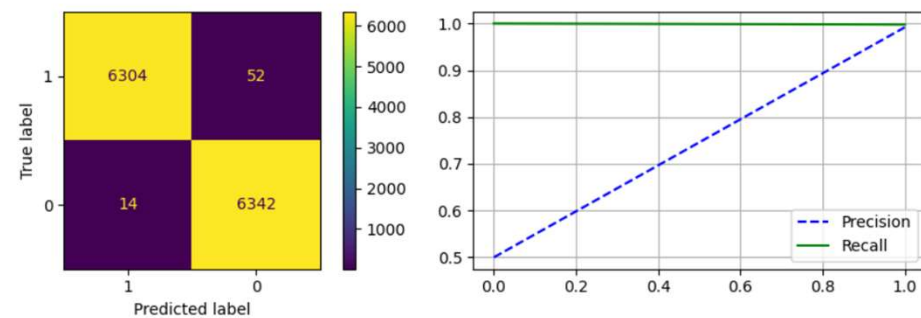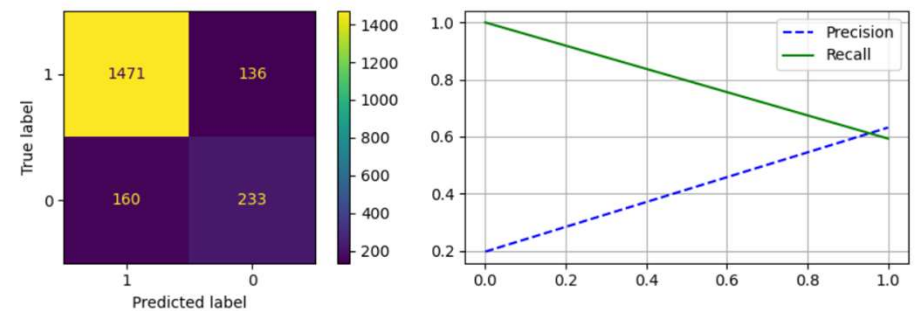
Conclusion: Looking at the scores, the **RandomForestClassifier** is the **best** model.

Thanks for your listening

--- Ironhack student Jian Lang