CUSTOMER BANK CHURN

Trevor Moos, Ben Sullivan, Jackson Hassell, Musmin Zar

DATA DESCRIPTION

Data consists of **12 predictors:**

- Current Balance
- Tenure
- Credit card possession
- Number of products
- Membership activity
- etc.

Response:

Binary indicator if the person churned or not

PROBLEM

PROBLEM

Churn is a huge problem, how do we mitigate it?

Lots of money is lost on customers that leave, and replacing them is expensive

See if it is predictable with the data

Target people likely to churn to reduce churn rates

CHURN LOSSES

Churn over 2000 customers

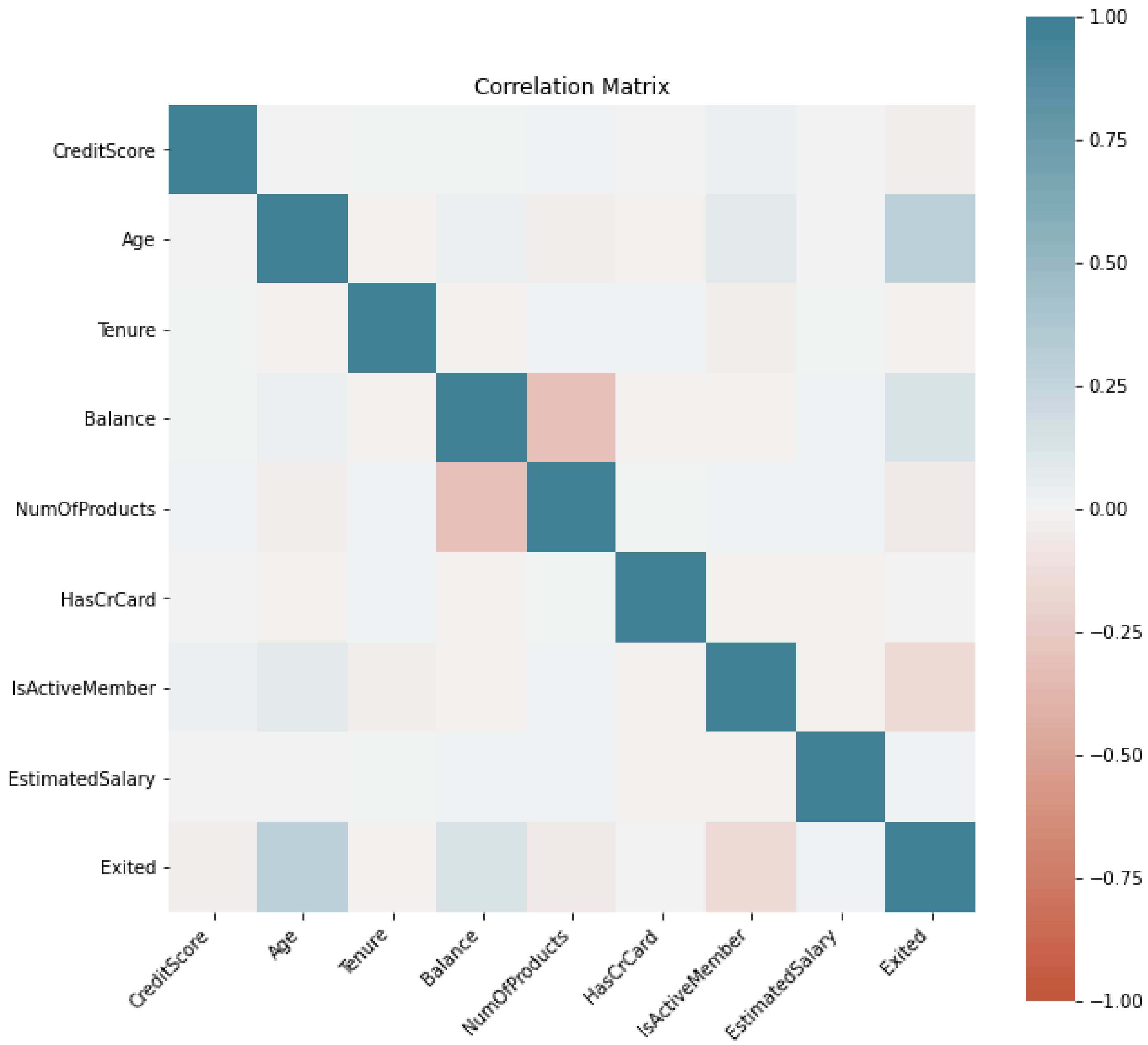
Expect to lose about $150 per customer

Churn loses banks:

>$300,000

EXPLORATORY DATA ANALYSIS

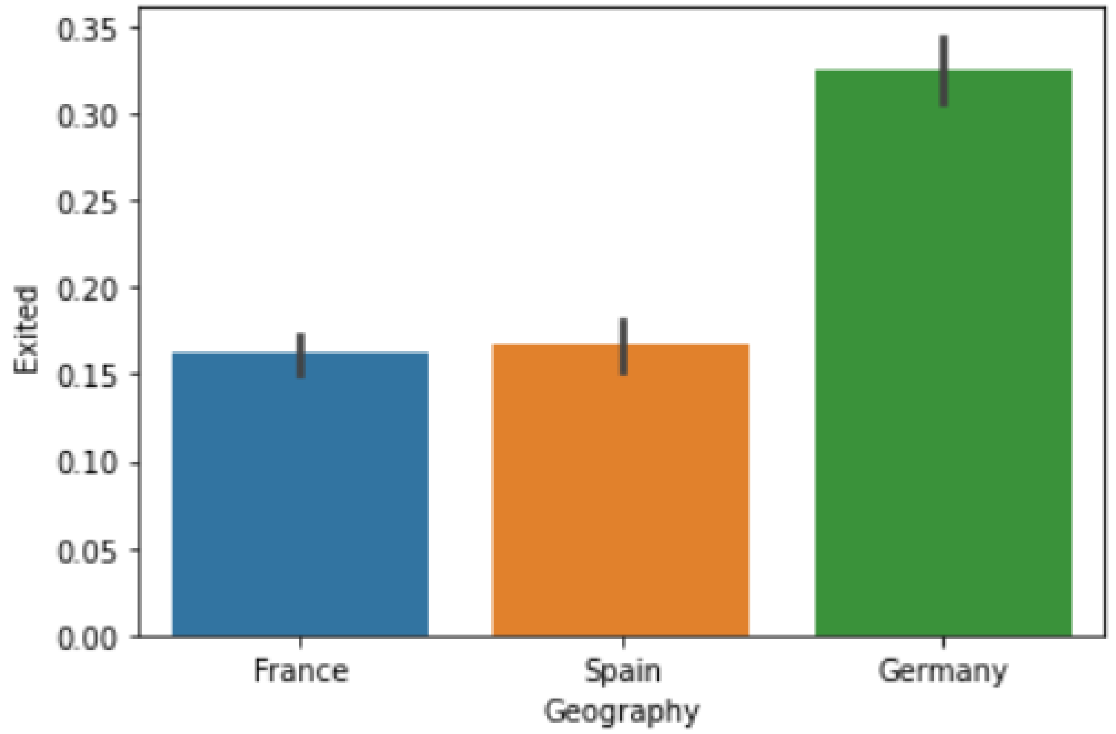
EXPLORATORY DATA ANALYSIS

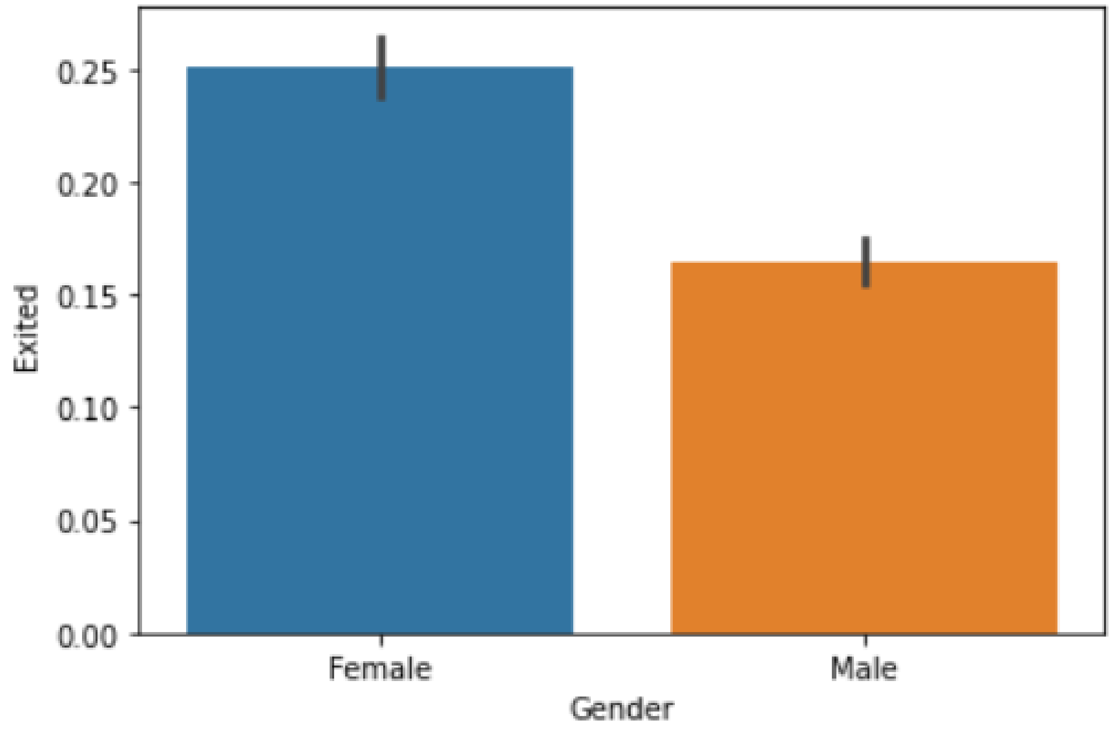


CREDITSCORE	-0.0271
AGE	0.2853
TENURE	-0.0140
BALANCE	0.1185
NUMOFPRODUCTS	-0.0478
HASCRCARD	-0.0071
ISACTIVEMEMBER	-0.1561
ESTIMATEDSALARY	0.0121
EXITED	1.0000

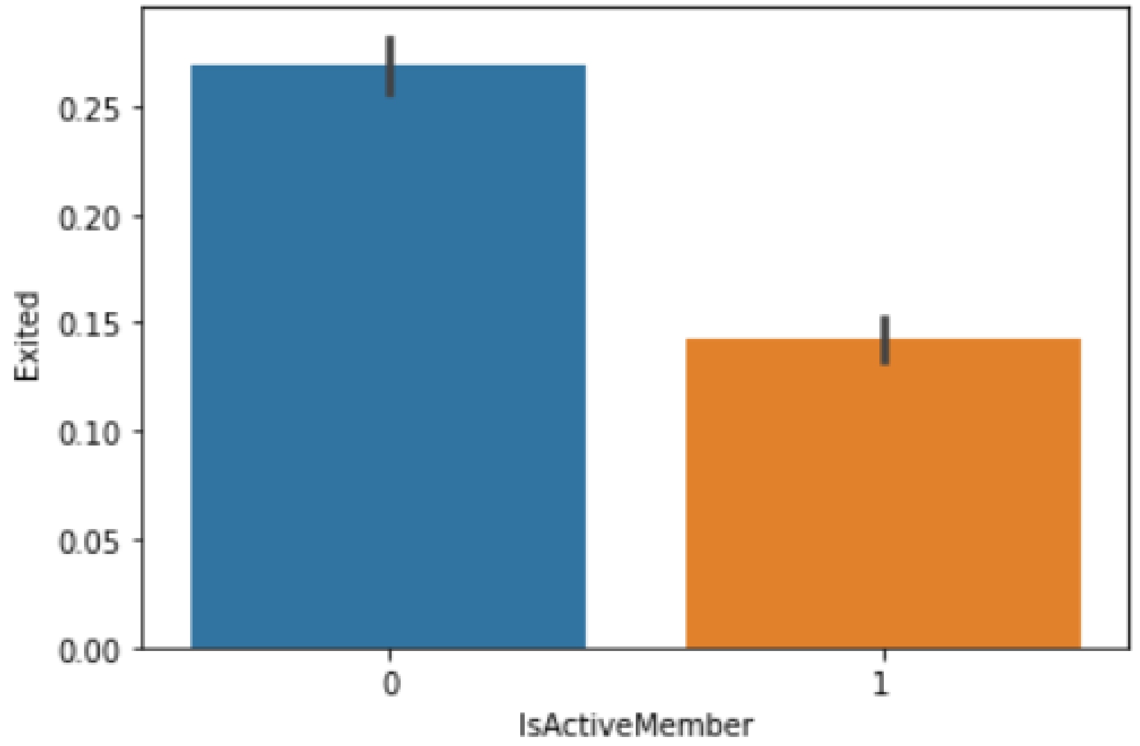
EXPLORATORY DATA ANALYSIS

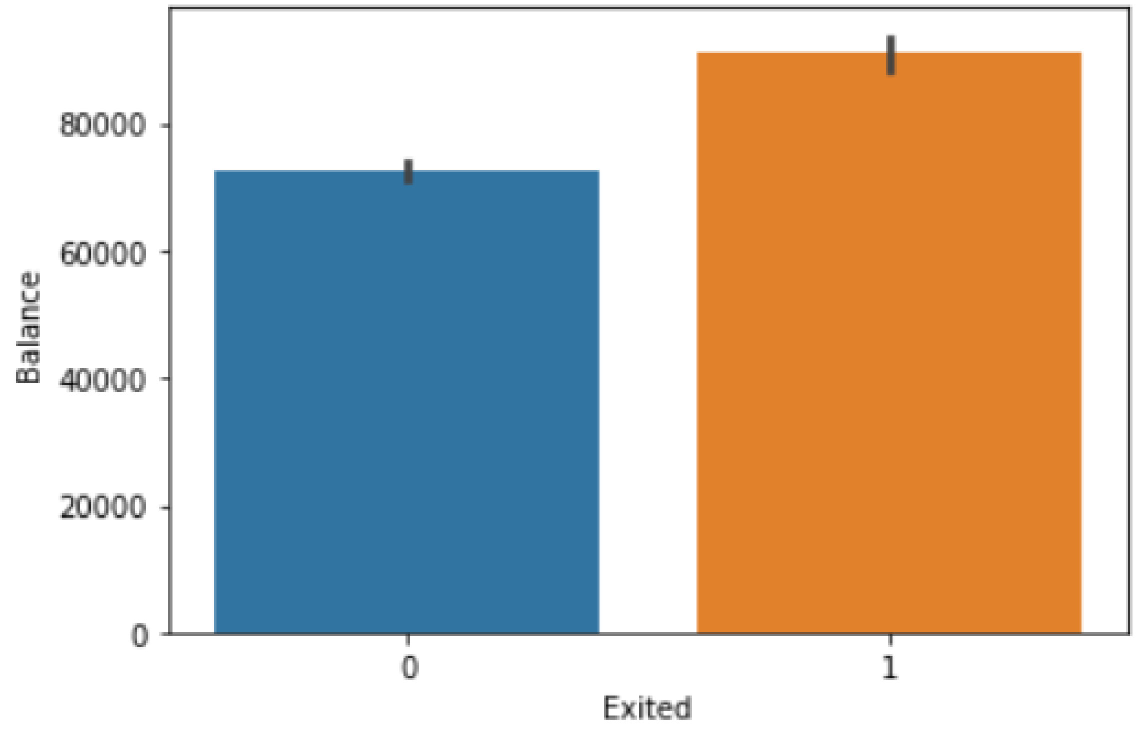
Demographic





Member Info

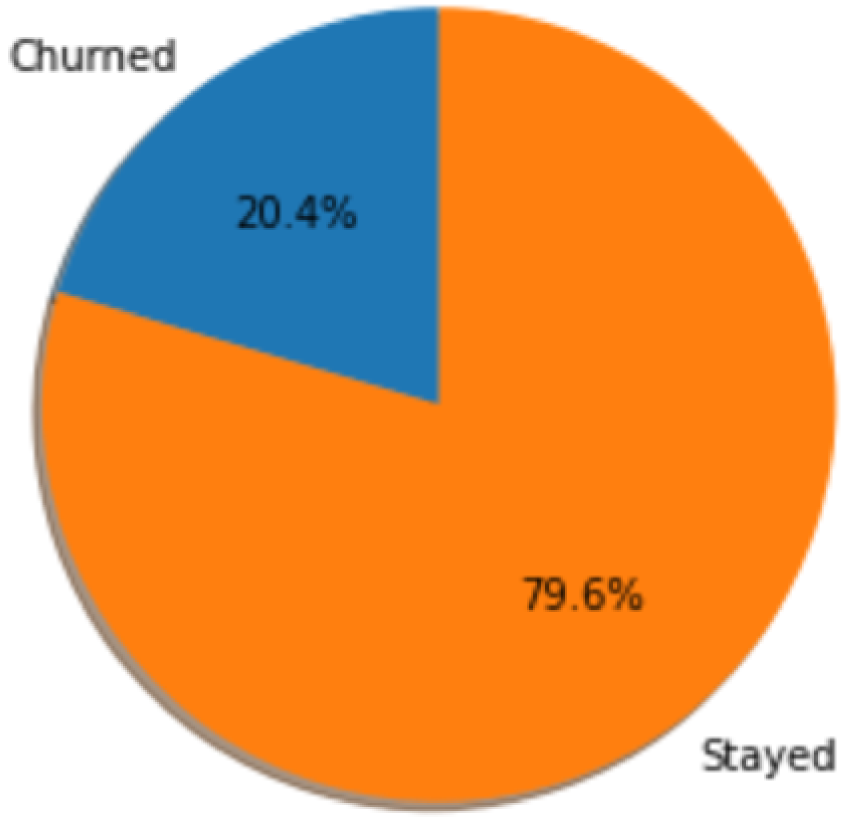




MODELING

UNBALANCED CLASSIFICATION PROBLEM

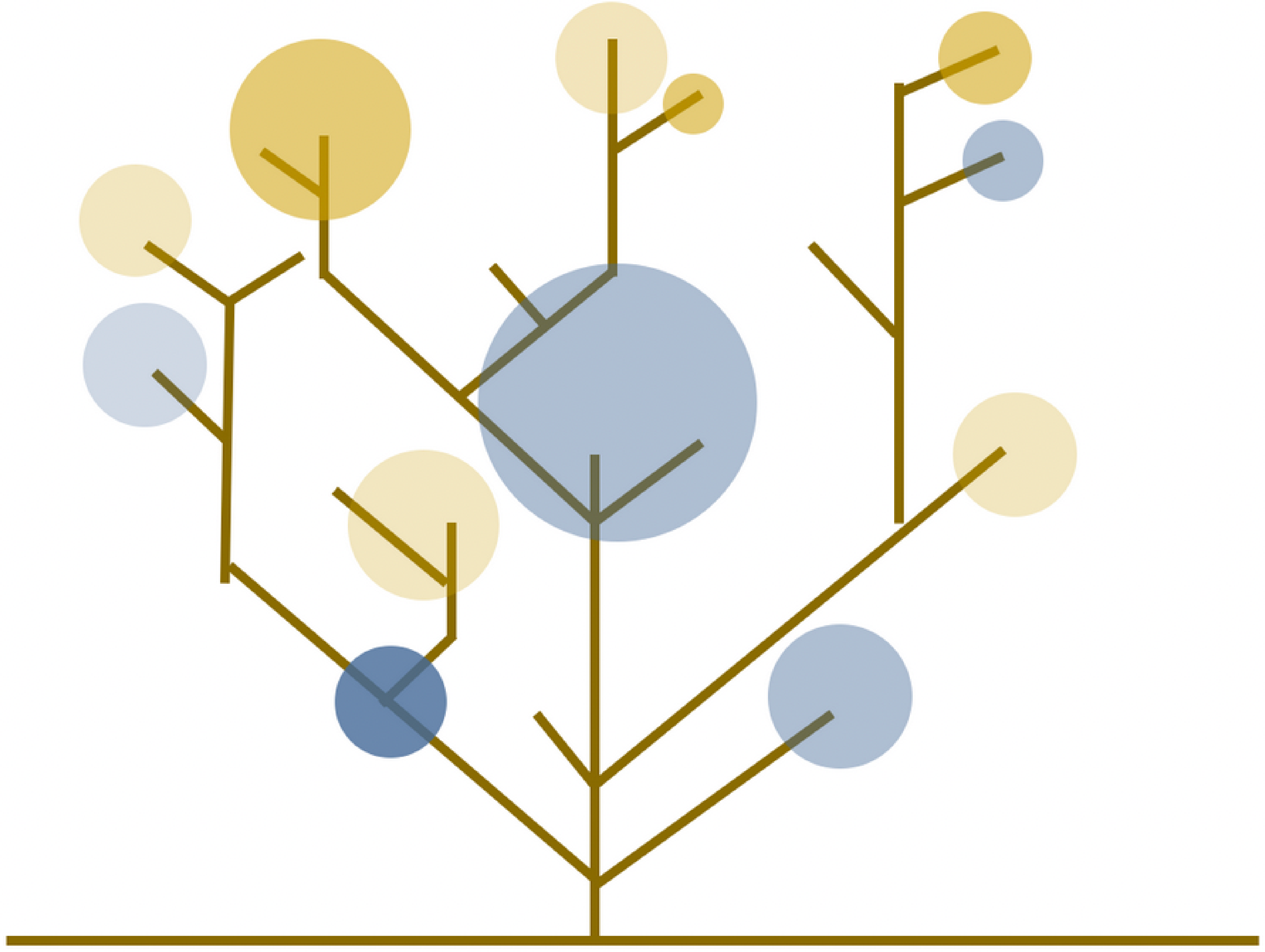
- Unbalanced data is a problem!
- Many models are built on the assumption of equal classes.
- Oversampling or undersampling?
- Balancing the training data improves predictive ability even on an unbalanced holdout set.



DATA MODELING

- Which model to use?
- What hyperparameters to use for each model?
- Test all combinations on a holdout set to evaluate models
- What metric to use? How much do we care about false negatives and false positives?

$$\text{Recall} = \frac{True\ Positive}{True\ Positive + False\ Negative}$$



MODELING RESULTS

Model	Raw Dataset	Undersample	Oversample
Logistic Regression	0.2	0.71	0.71
Naïve Bayes	0.41	0.71	0.71
SVM	0.41	0.76	0.72
K-Nearest Neighbors	0.41	0.74	0.49
Decision Tree	0.39	0.73	0.5
XGBoost	0.47	0.77	0.76
MLP	0.45	0.76	0.53

Sampling the training data helps!

APPLICATION

PROMOTION

Cost Parameters

VALUE OF CUSTOMER
$150

COST OF PROMOTION
$50

COST OF SOLICITATION
$8

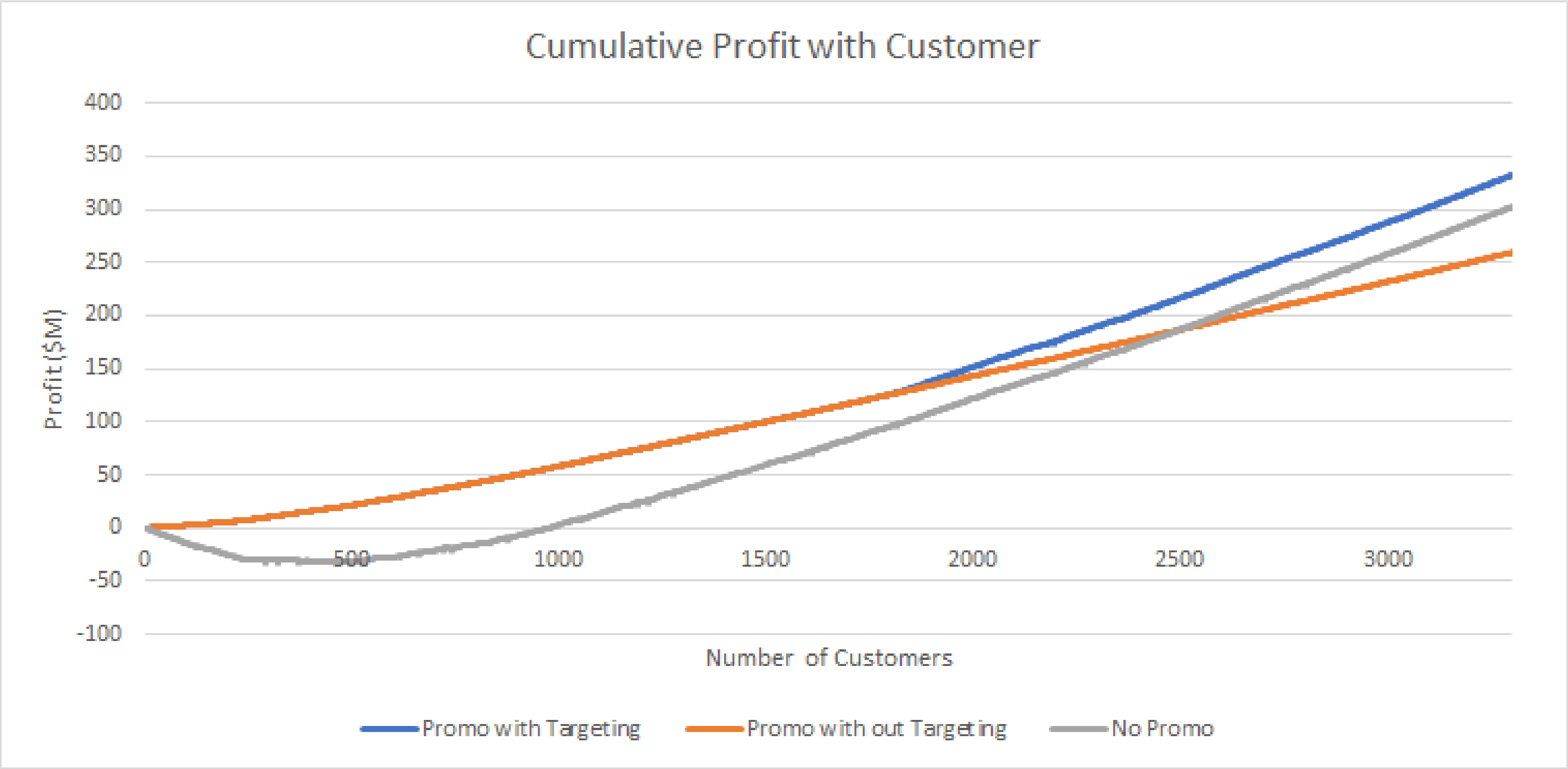
Customer	Exited	Prob Exit
0	1	99.67%
1	1	99.66%
2	1	99.65%
3	1	99.63%
4	1	99.61%
5	1	99.53%
6	1	99.51%
7	1	99.48%
8	1	99.34%
9	1	99.33%
10	1	99.28%
11	1	99.25%
12	1	99.17%
13	1	99.16%
14	1	99.12%
15	1	98.97%
16	1	98.83%

Targeting

CUT-OFF PROBABILITY
30%

TOTAL TARGETED
55%

RESULTS



Profit Margins

PROMOTION (NO TARGETING)
-$42,866

PROMOTION (TARGETING)
$29,752

CONCLUSION

CONCLUSIONS

Modeling

Recall score = 0.77

Cost of Churn

Loss of $150 per customer
20% of customers

Targeting Method

Highest-risk customers below 30% cutoff

Our Promotion

Targeting gives the bank a positive profit margin

FUTURE CONSIDERATIONS

Sources of Profit

For further analysis of profit, we wanted to consider each individual's account balance and the profit generated as a percentage of that value

Improved Targeting

Rather than simply targeting those who are at highest risk of exiting (since these customers may inherently be the most difficult to target with a promotion), we may want to find an alternative segment of customers to target

Weighting Probabilities

With more data and resources, we would have liked to determine a way to weight the probability of exit for various customers