IRON HACK

Data Mid-Bootcamp Classification Project

Credit Card Offer

Mario La Torre
Emanuel Panait

DAPT OCT 2022

Index

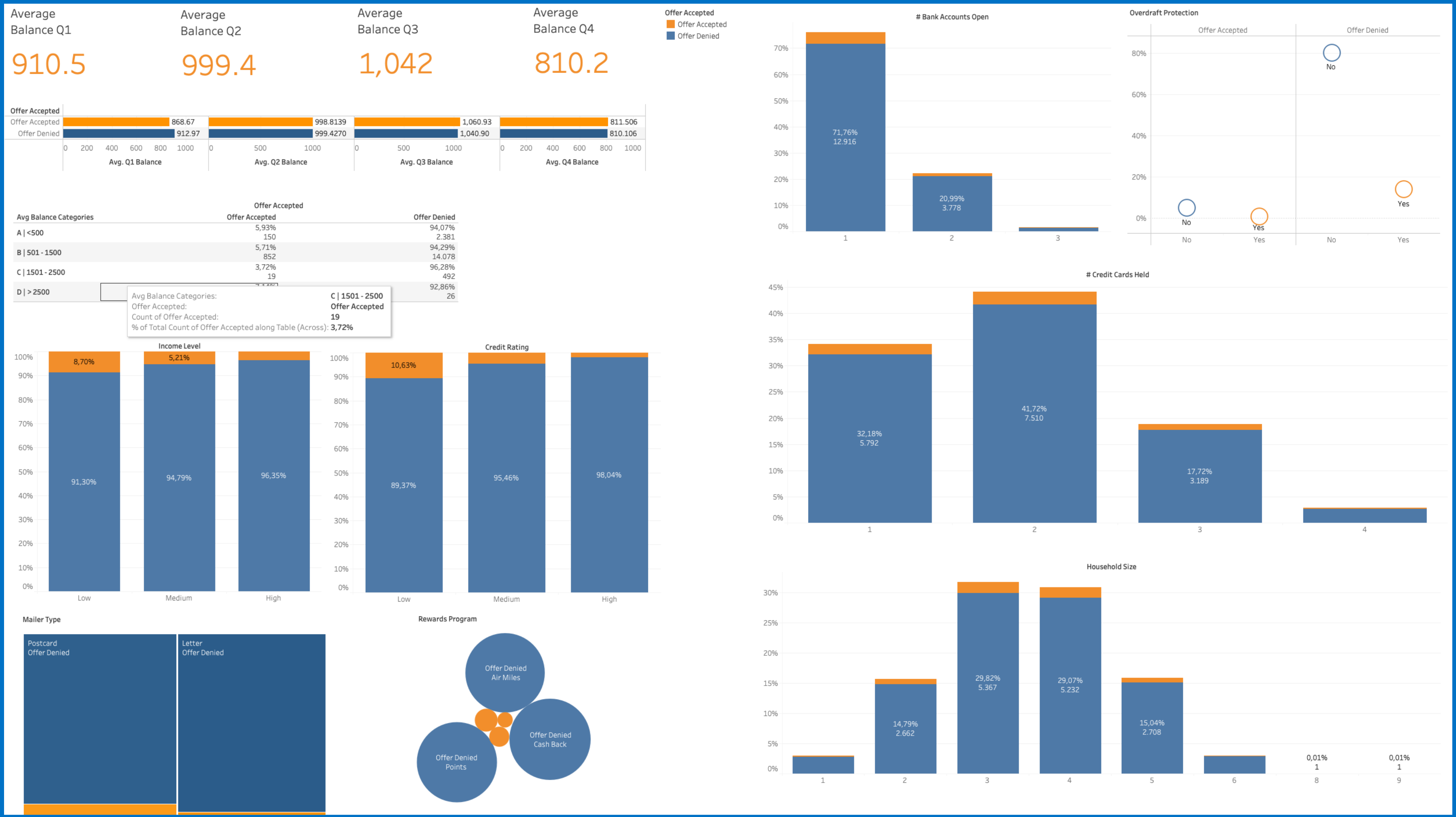
- Objectives
- Exploratory Data Analysis
- Benchmark
- Optimization
- Models Comparison
- Business Case

Objective

- **Build a model that will predict users who will accept a Credit Card offer**
- Get insights into groups of people who accepted an offer

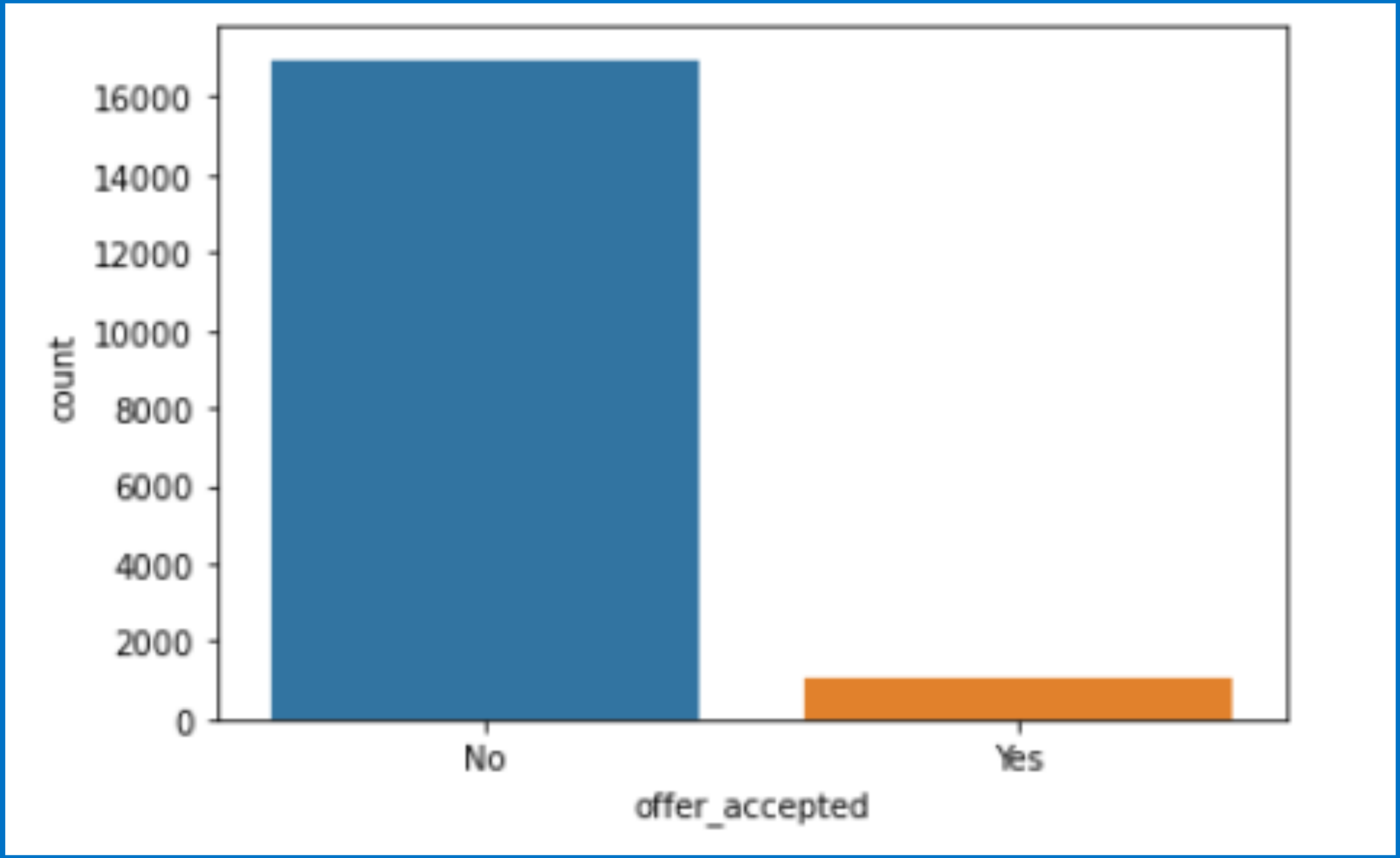
The database comes from a focused marketing study organized by the bank with participation of 18,000 current customers.

Exploratory Data Analysis



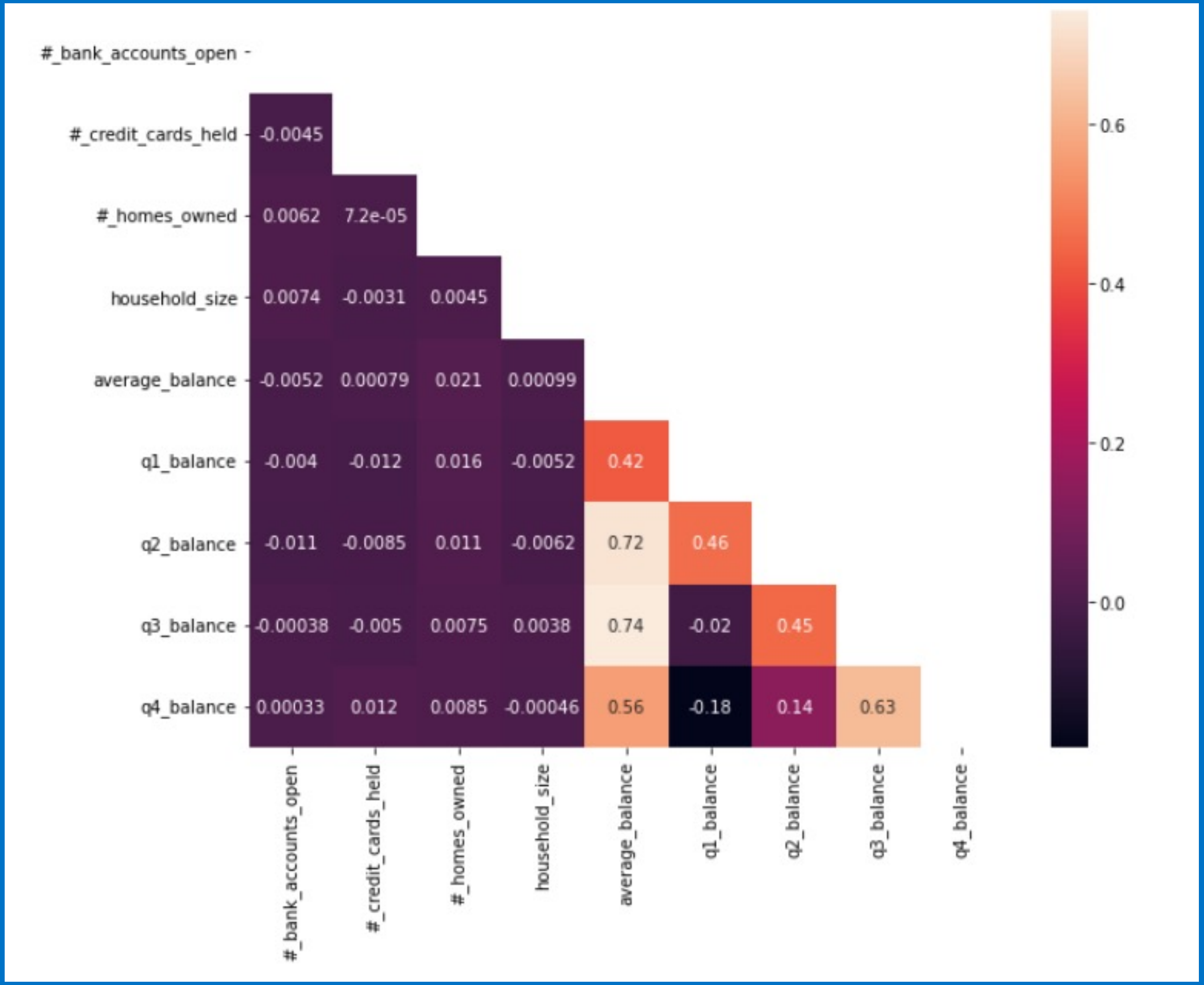
Exploratory Data Analysis

High data imbalance



Exploratory Data Analysis

No high multicollinearity



Benchmark

Because our objective is to identify those who will accept an offer, we want to get most positives as possible, which translate into high recall.

Precision

Specificity

Recall (Sensitivity)

	precision	recall	f1-score	support
No	0.97	0.68	0.80	3392
Yes	0.12	0.70	0.20	204
accuracy			0.68	3596
macro avg	0.54	0.69	0.50	3596
weighted avg	0.93	0.68	0.77	3596

Benchmark

Precision is the percentage of clients that accepted the offer were and were correctly identified.

Precision (P) = TP/(TP+FP)

Recall is the percentage of the relevant clients (who accepted an offer) that could be successfully identified.

Recall(R) (Sensitivity) = TP/(TP+FN)

When we want accurately identify people who rejected an offer, then the number of true negatives should be high, which would require a high specificity.

Specificity = TN/(TN+FP)

Optimization

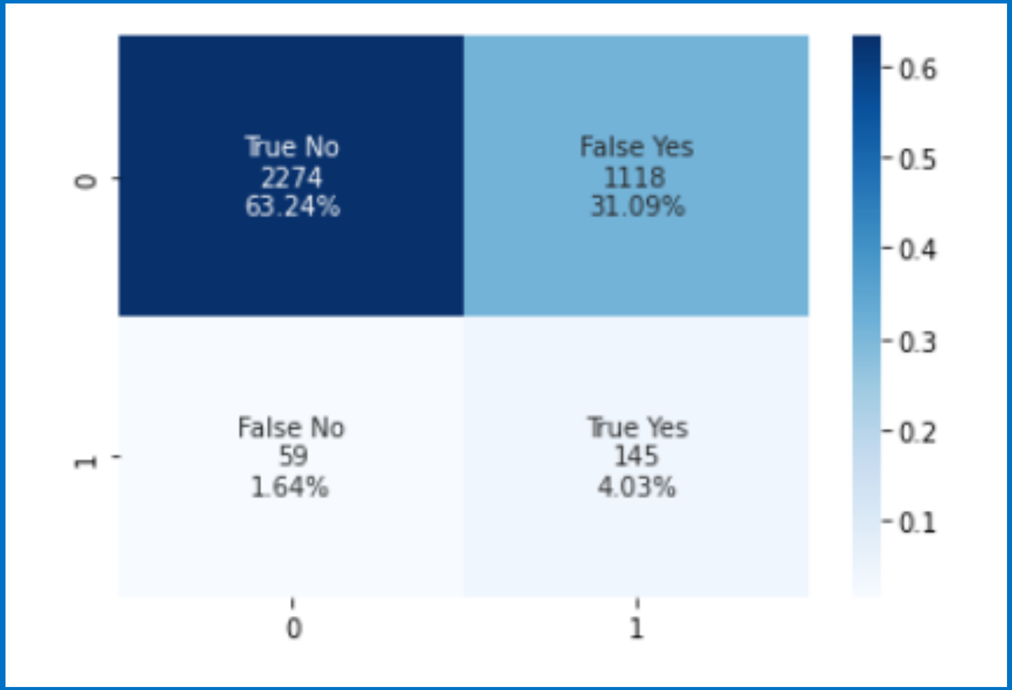
To optimize our model, we applied different methods and compared how much they would improve or not the recall of the model

1. Encoders (OneHot Encoder)
2. Scalers (StandardScaler, Normalizer, BoxCox)
3. Outliers (Drop for Average, Q2, Q3 and Q4 balances)
4. Balancers (SMOTE, Over and Undersampling, TomekLinks)

Optimization

Final Model

	precision	recall	f1-score	support
No	0.97	0.67	0.79	3392
Yes	0.11	0.71	0.20	204
accuracy			0.67	3596
macro avg	0.54	0.69	0.50	3596
weighted avg	0.93	0.67	0.76	3596



Our final model managed to optimize both specificity and recall, so we could correctly identify customers that will accept the offer and also those who will reject it

Models Comparison

Models	Recall Score
1. LogisticRegression	0,71
2. MultinomialNB	0,60
3. GaussianNB	0,58
4. SVC	0,06
5. KNeighborsClassifier	0,14
6. DecisionTreeClassifier	0,15

Business Case

Assuming that the bank provides us with the following information:

- 1M active clients in the last fiscal year;
- \$5 is the cost to print a new card and send it via mail to a client;
- \$150 is the average spent on a newly printed card that was accepted by the client.

Then:

- Every true positive: \$150 - \$5 = +\$145 (card sent, offer accepted)
- Every true negative: \$0 (card not sent)
- Every false negative: - \$150 (card not sent, lost opportunity for an offer acceptance)
- Every false positive: - \$5 (card sent, offer not accepted)

Business Case

With the active customers from last fiscal year, according to our model:

- If the bank sent a card to every client, it would have - $1,2M in revenue
- Sending a card to clients identified as potential accepters would generate $1,8M

- Savings: $ 3,1M

Thank You!!