Fairway Bank Churn Reduction

Photo: https://www.flickr.com/photos/anekoho/40759061335

Context

- Multinational bank (France, Spain, Germany)
- Worth €2.3B, 1M customers
- Opening branches in Belgium and Holland
- Renovating online banking platform

Problem

- Fairway's churn rate last year: **20%**

- Higher than ideal
  - 19% [1]
  - 15% [2]
  - < 5 – 10% [3]

- **Goal: reduce churn by 10% in a year**

1 https://customergauge.com/blog/average-churn-rate-by-industry
2 https://www.reviewtrackers.com/blog/bank-customer-retention
3 https://uxpressia.com/blog/how-to-approach-customer-churn-measurement-in-banking

Data Wrangling

- Random sample of Fairway's customers [1]
- 10K observations, 15 features
- Clean data, no duplicates, no missing values

1 https://www.kaggle.com/datasets/radheshyamkollipara/bank-customer-churn
Photo: https://m.bankingexchange.com/news-feed/item/9336-cfpb-requests-public-input-to-improve-bank-customer-service

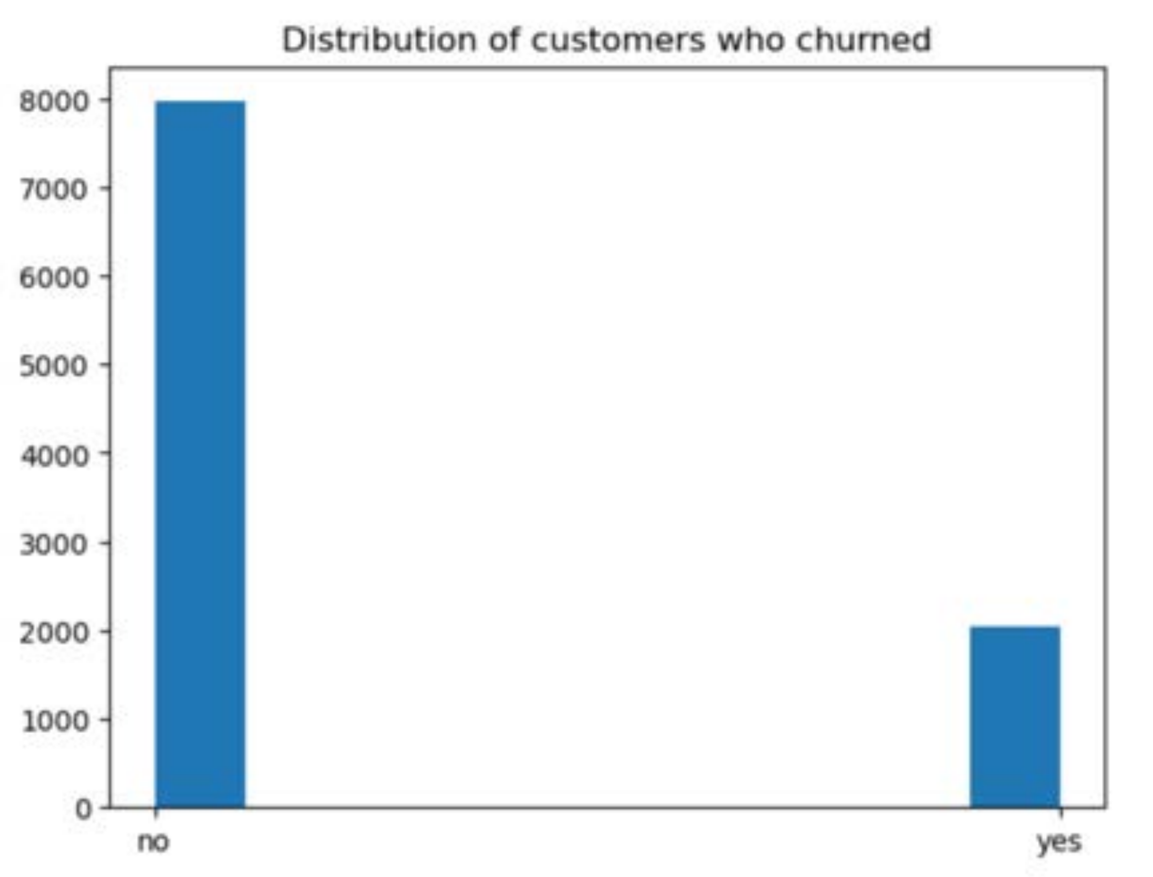
Data Wrangling

Features

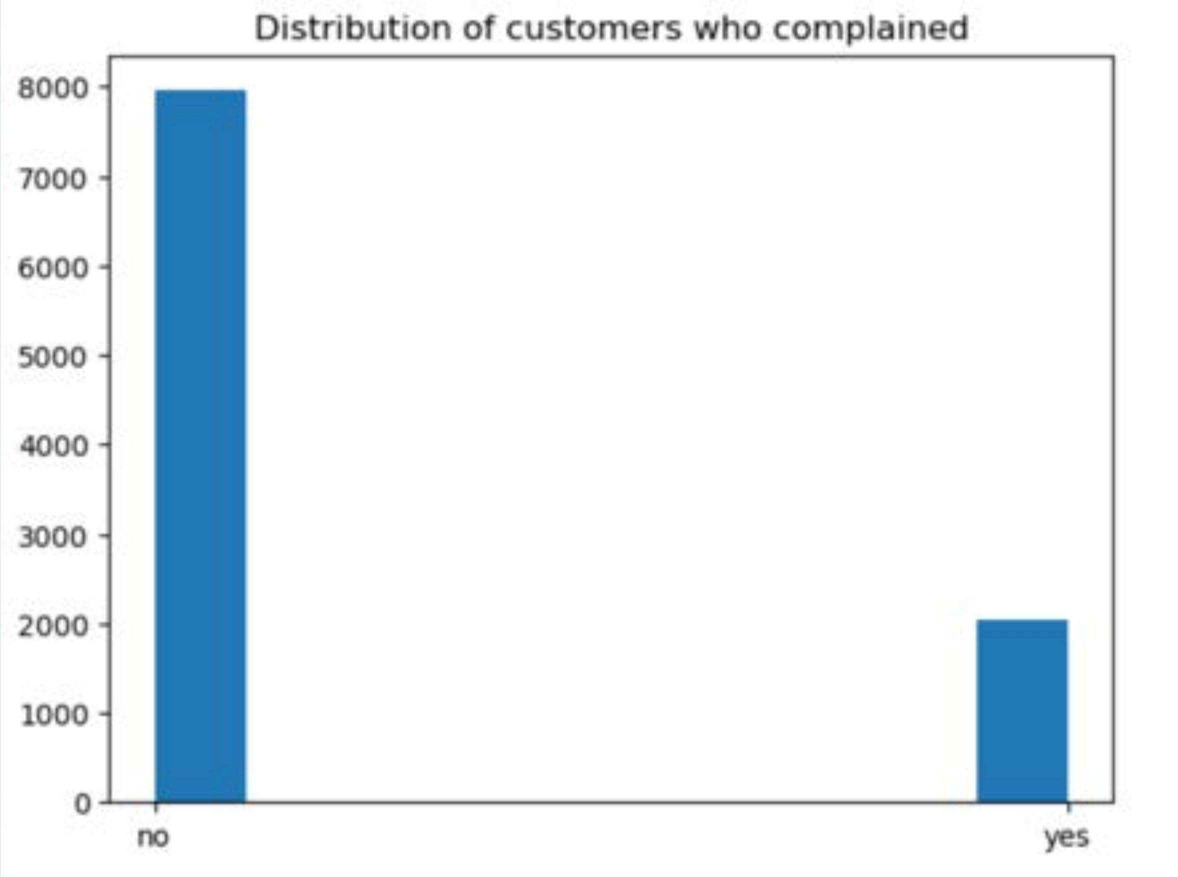
- credit score
- country
- gender
- age
- tenure
- balance
- number of products
- has credit card?
- active?
- salary
- satisfaction (1 – 5)
- card type
- points
- complained?

- **churned?**

EDA

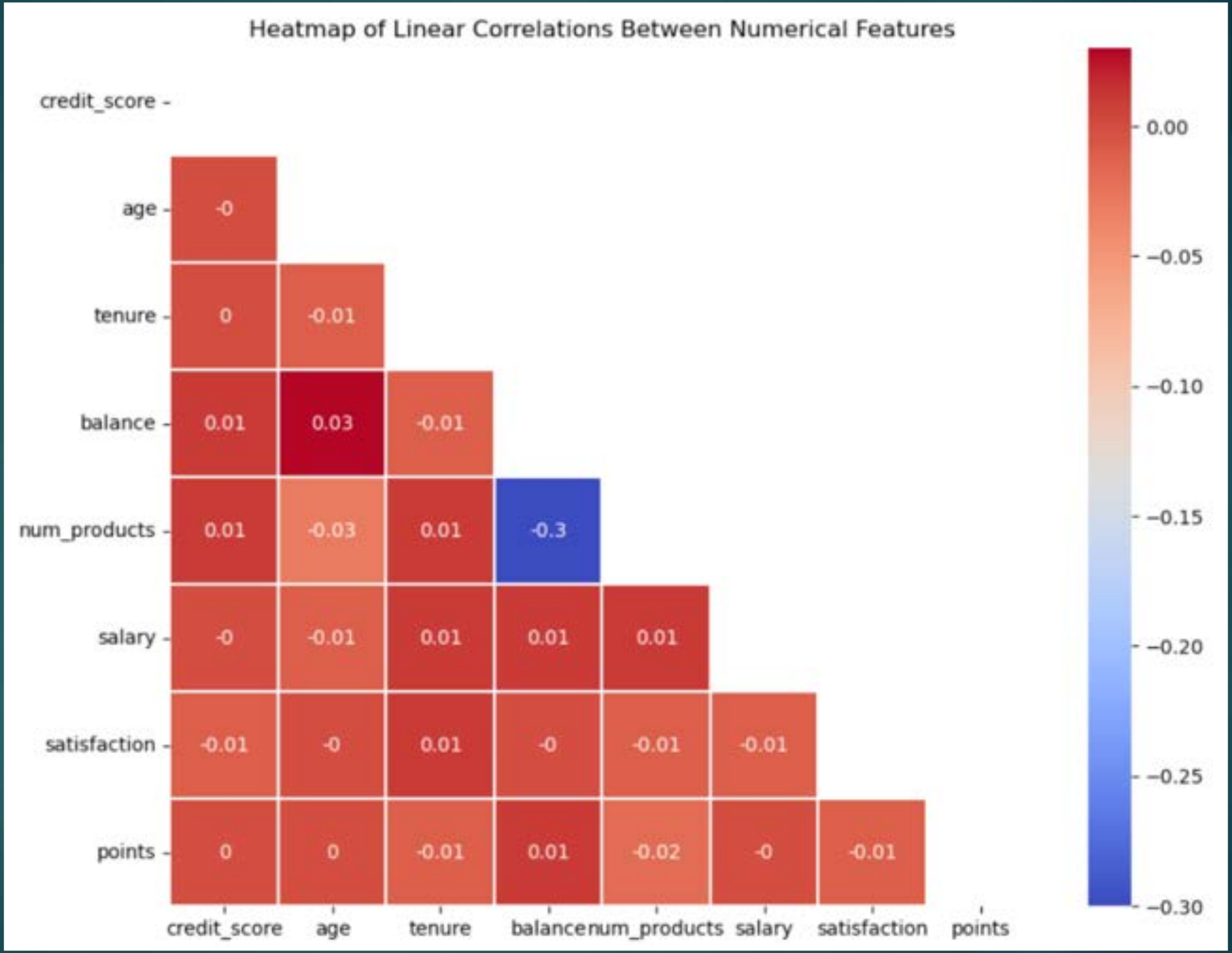


Distribution of customers who churned
8000
7000
6000
5000
4000
3000
2000
1000
0
no
yes



Distribution of customers who complained
8000
7000
6000
5000
4000
3000
2000
1000
0
no
yes

EDA



Heatmap of Linear Correlations Between Numerical Features
credit_score
age
tenure
balance
num_products
salary
satisfaction
points
-0
0
-0.01
0.01
0.03
-0.01
0.01
-0.03
0.01
-0.3
-0
-0.01
0.01
0.01
0.01
-0.01
-0
0.01
-0
-0.01
-0.01
0
0
-0.01
0.01
-0.02
-0
-0.01
0.00
−0.05
−0.10
−0.15
−0.20
−0.25
−0.30

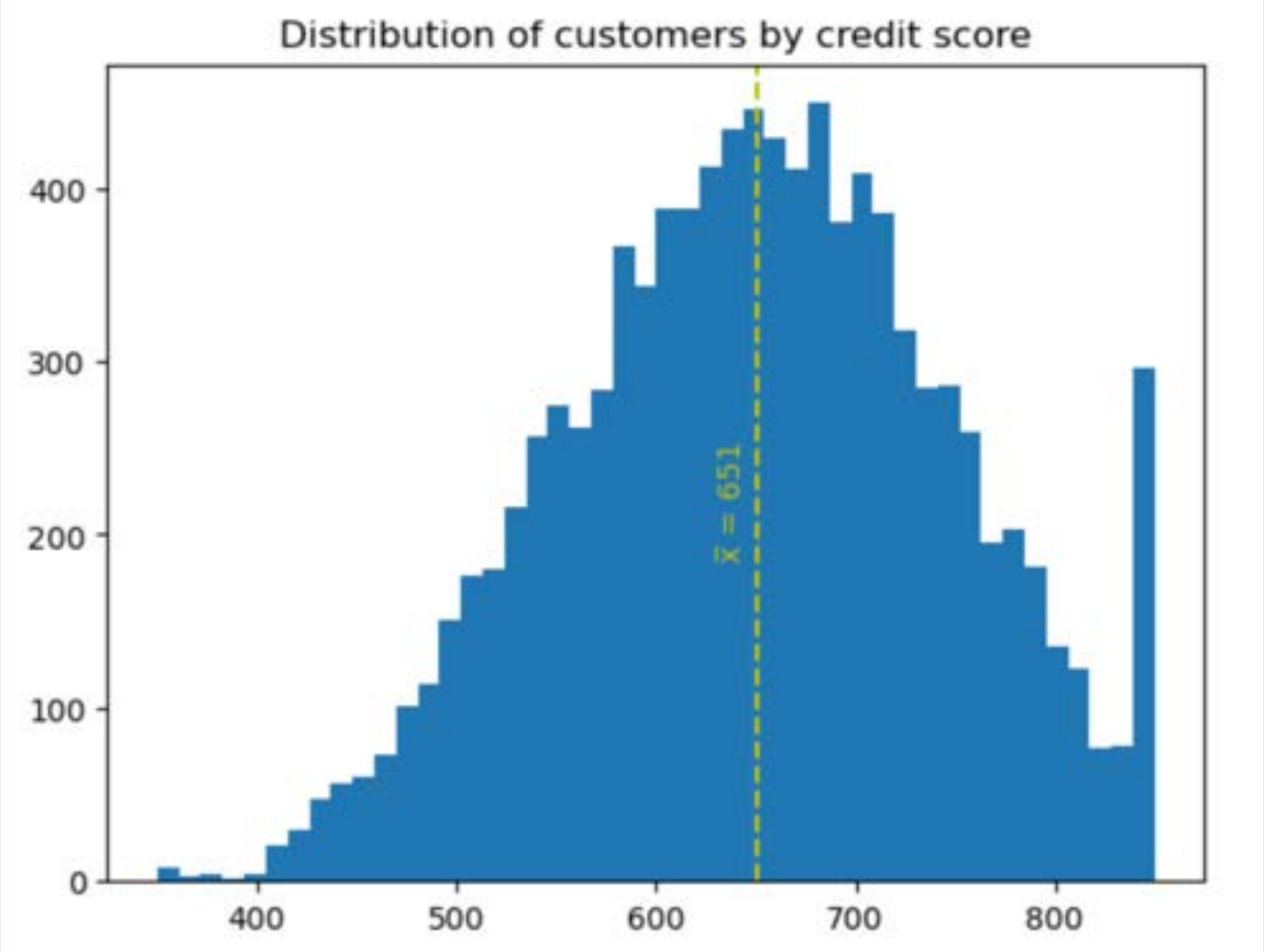
Preprocessing

- Dummy encoding (0's and 1's)
  - 19 total input features
- Bin numerical features
  - credit score
  - age
  - balance
  - salary
  - points
- Dimensionality reduction, cluster analysis

Preprocessing

credit score [1]

- 300 – 579
- 580 – 669
- 670 – 799
- 800 - 850

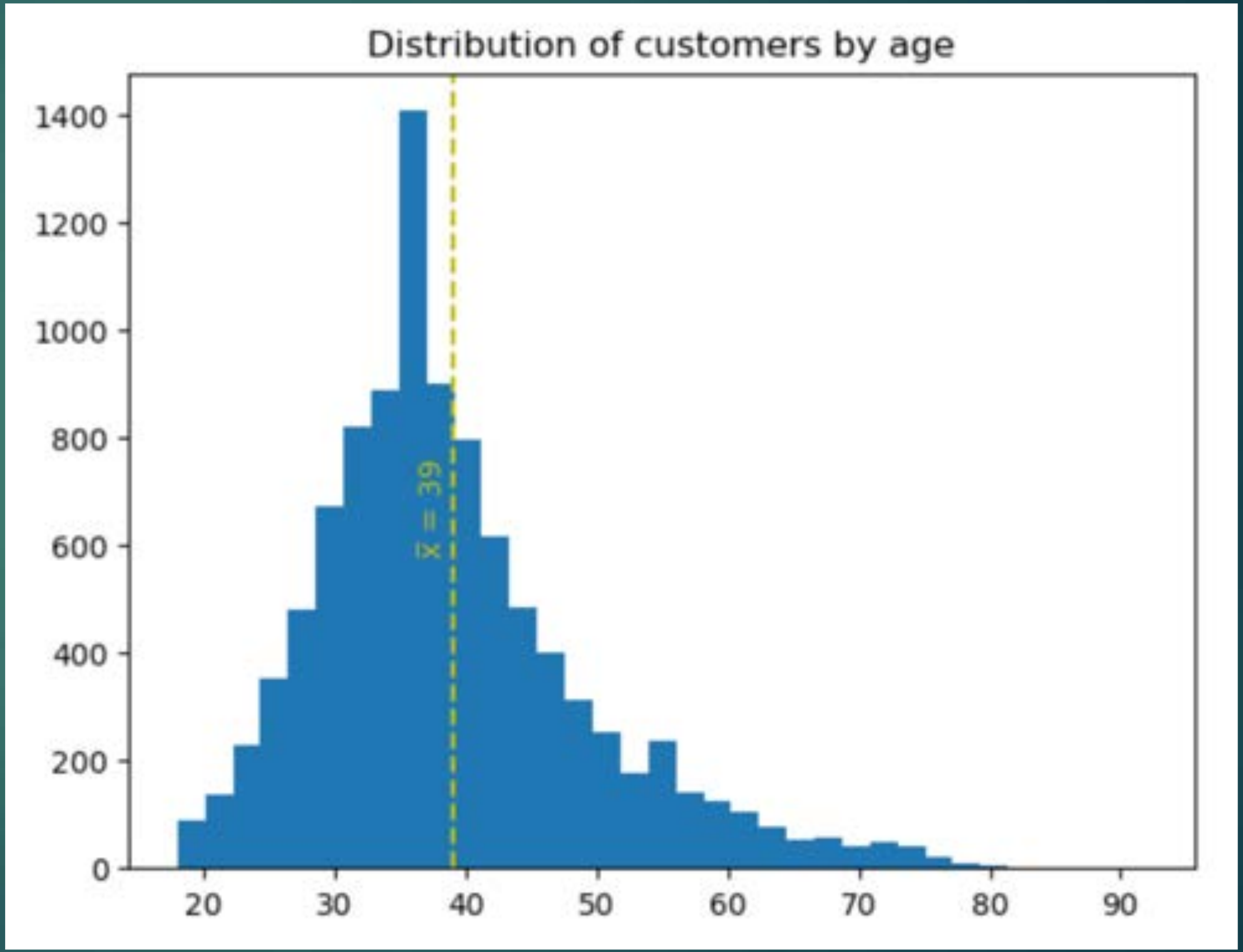


1 https://time.com/personal-finance/article/different-credit-scoring-ranges/

Preprocessing

age [1]

- 18 – 24
- 25 – 34
- 35 – 44
- 45 – 54
- 55 – 64
- 65+

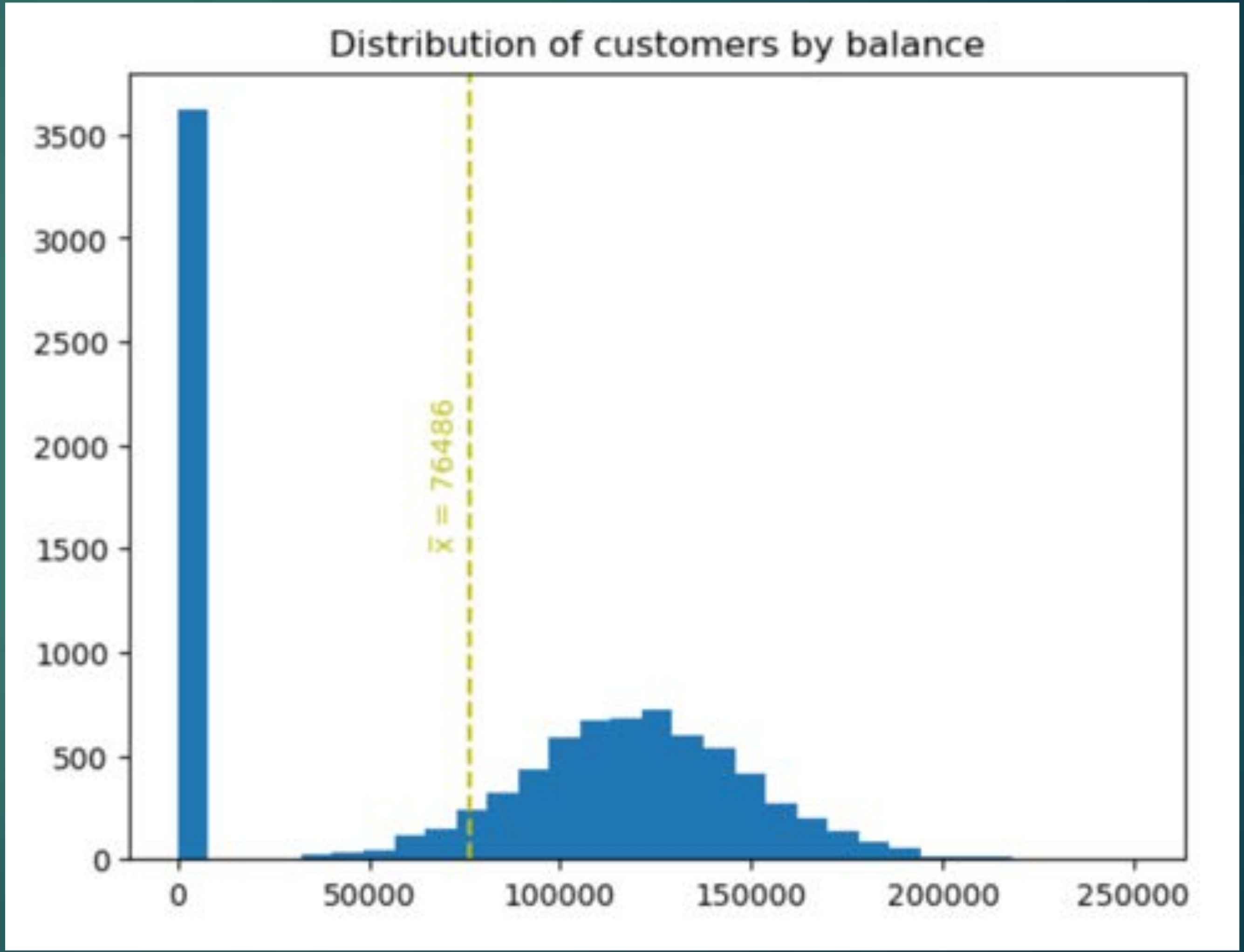


1 https://www.pickfu.com/demographic-segmentation

Preprocessing

balance

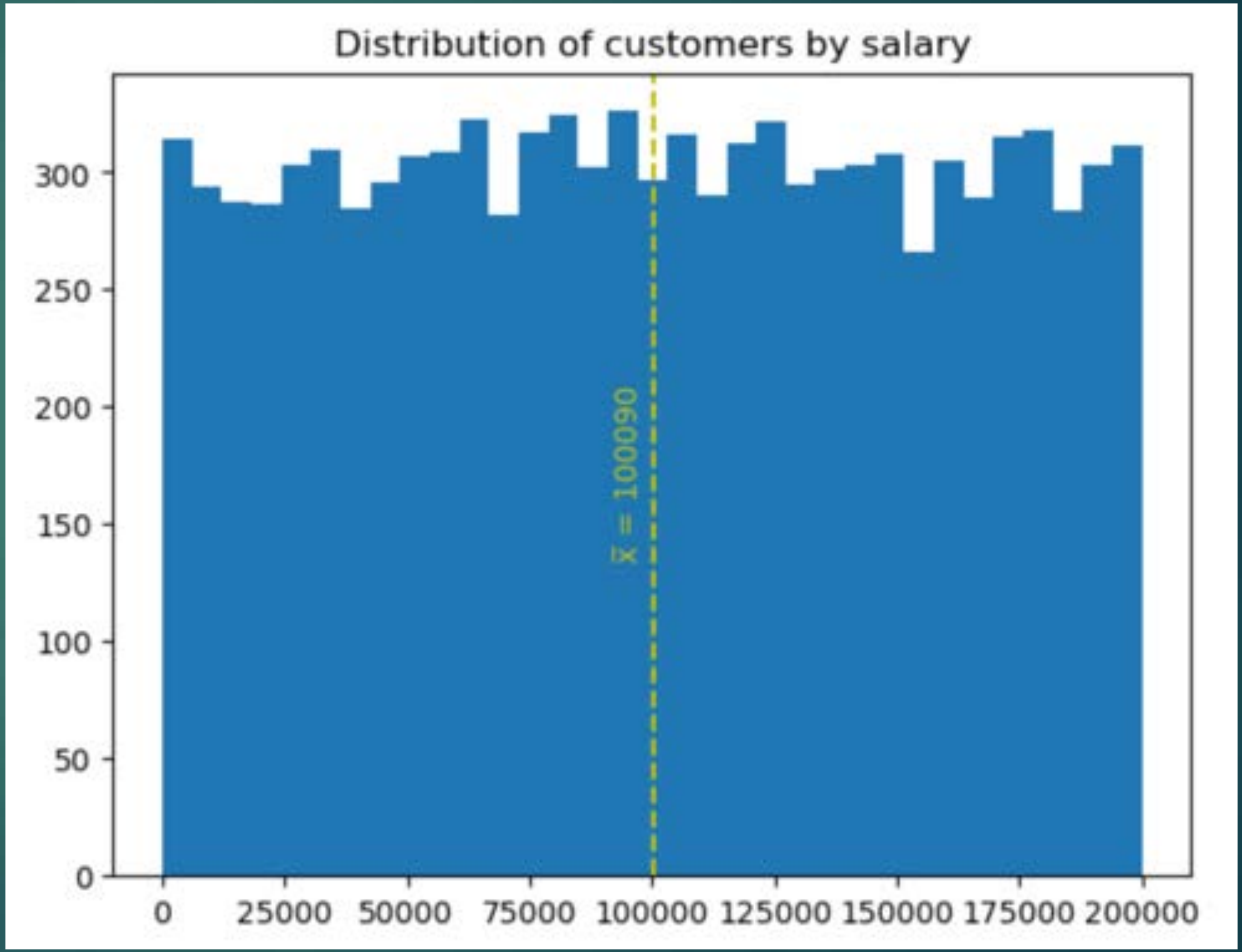
- 0
- 1 – (100K)
- 100K – (150K)
- 150K+



Preprocessing

salary

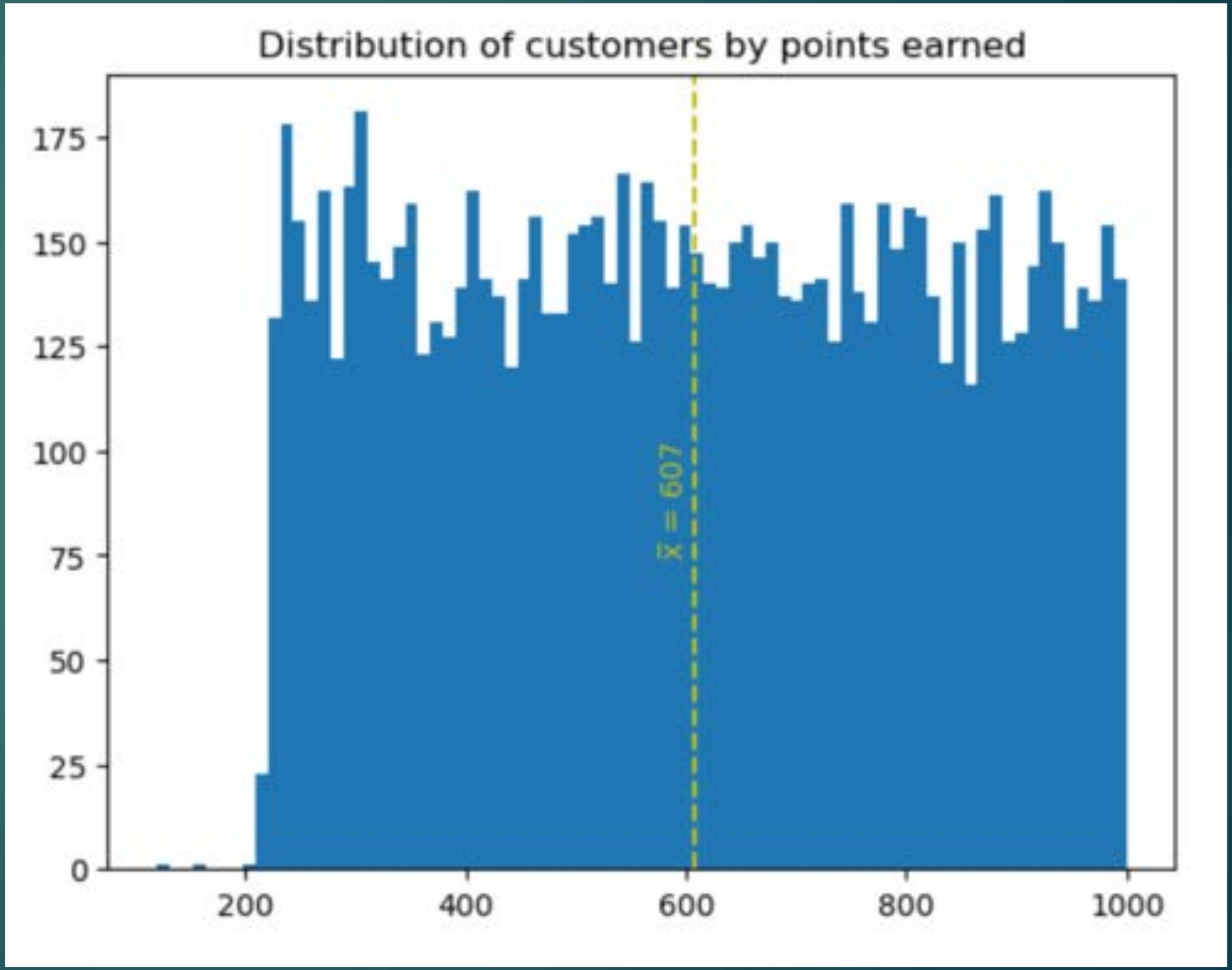
- 0 – (50K)
- 50K – (100K)
- 100K – (150K)
- 150K+



Preprocessing

points

- 0 – 299
- 300 – 399
- 400 – 499
- 500 – 599
- 600 – 699
- 700 – 799
- 800 – 899
- 900+



Modeling

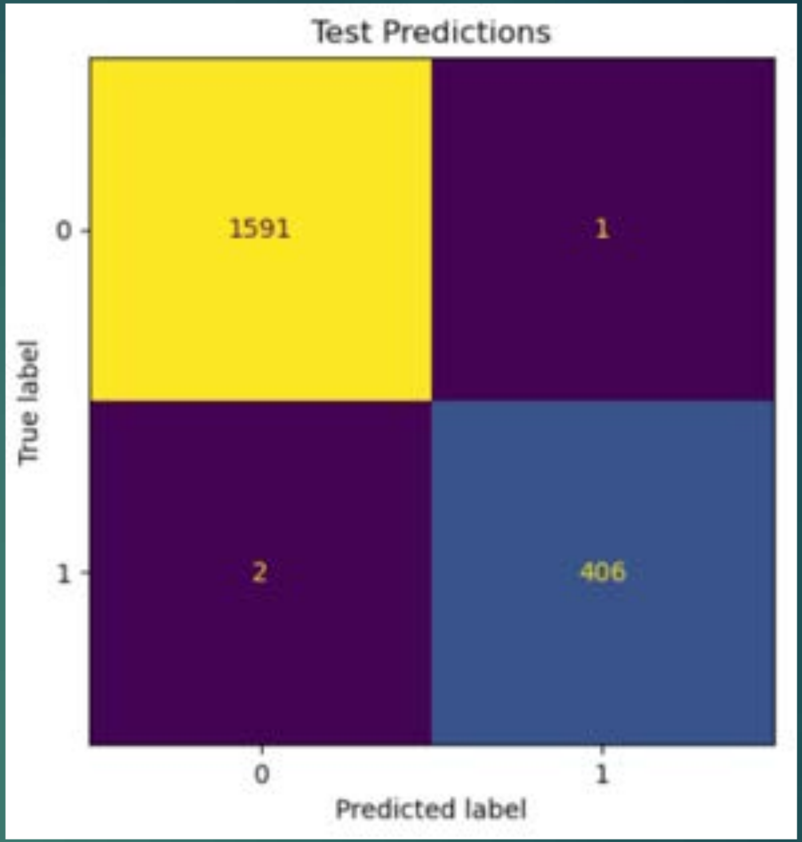
- 80 / 20 train / test split
- 5-Fold stratified K-fold on train set
- Goal: minimize false negatives
- Performance metric: **recall**

Modeling

Models

- Random Forest
- Extra Trees
- Gradient Boosting
- Hist Gradient Boosting
- AdaBoost
- XGBoost
- LightGBM
- CatBoost
- Logistic Regression
- Support Vector

Modeling



- Initial 99.88% train set recall
- Complain: golden feature
- Only 4 / 2038 customers who churned didn't complain
- Only 10 / 2044 customers who complained didn't churn

Modeling

- Drop ‘complain’ feature
- Initial median recall: 44.14%
- Lower probability threshold for more positive predictions
- Balance false positive tradeoff by optimizing f1-score
- Improved median recall: 60.50%

Modeling

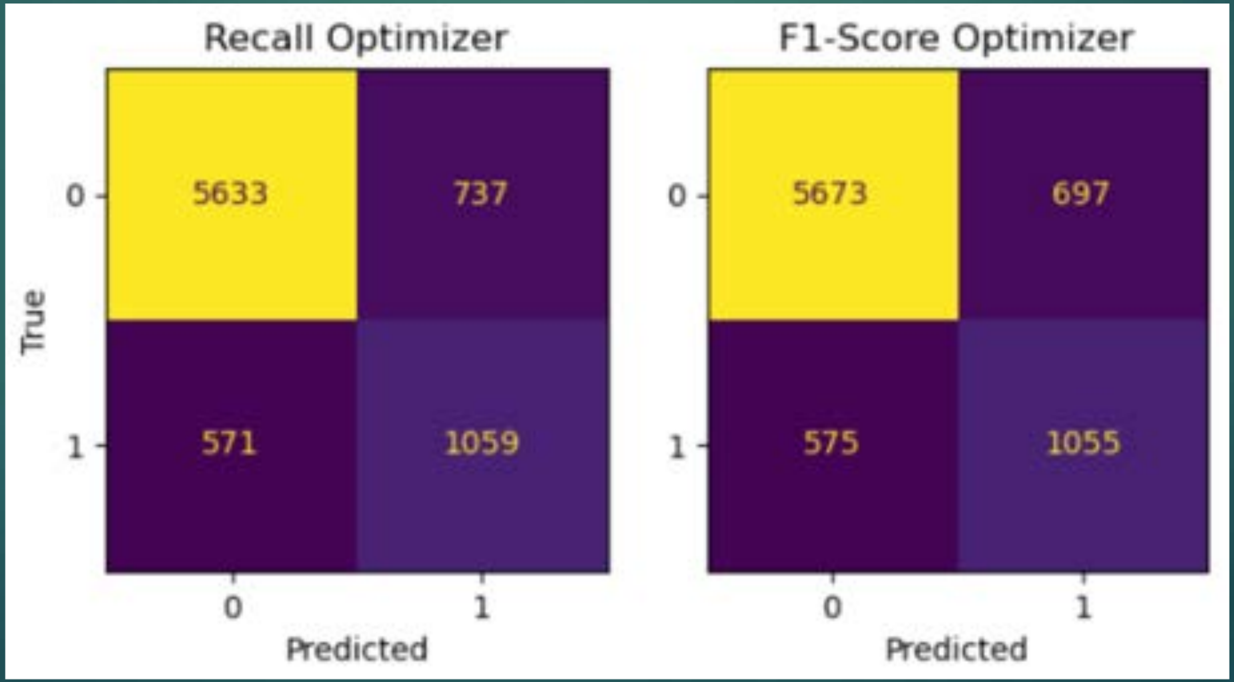
Optimum Probability Thresholds

- 0.25 Logistic Regression
- 0.27 Support Vector
- 0.29 CatBoost
- 0.30 XGBoost
- 0.31 Hist Gradient Boosting
- 0.32 Extra Trees
- 0.33 Gradient Boosting
- 0.34 Random Forest, LightGBM
- 0.50 AdaBoost

Modeling

- Hyperparameter tuning - Optuna

- F1-score best optimizer metric

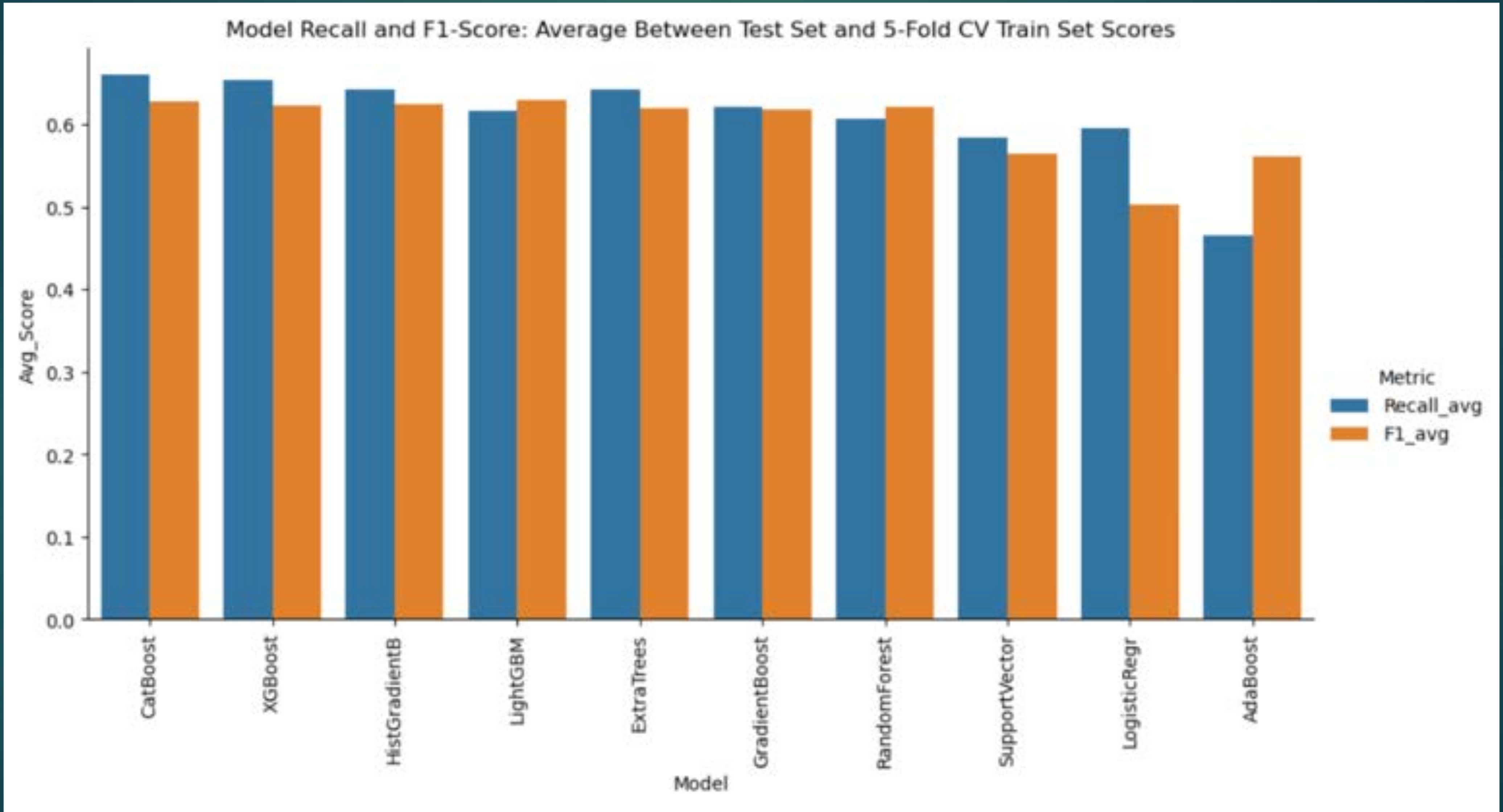


Modeling

- Aggregate score (per model, 5 metrics)
  - Train set: mean 5-fold cross-validation score
  - Average with test set score
  - Subtract train set standard deviation

- Metric weights
  - 40% recall
  - 30% f1-score
  - 20% ROC AUC
  - 5% precision
  - 5% accuracy

Modeling


Model Recall and F1-Score: Average Between Test Set and 5-Fold CV Train Set Scores
Avg_Score
0.0
0.1
0.2
0.3
0.4
0.5
0.6
CatBoost
XGBoost
HistGradientB
LightGBM
ExtraTrees
GradientBoost
RandomForest
SupportVector
LogisticRegr
AdaBoost
Model
Metric
Recall_avg
F1_avg

Modeling

Best model: CatBoost Classifier

- iterations: 545
- learning_rate: 0.013560631530876574
- depth: 6
- l2_leaf_reg: 2.548492629309027
- random_strength: 0.3660344916841721
- bagging_temperature: 0.2722303737601264

- probability threshold: 0.29

CatBoost

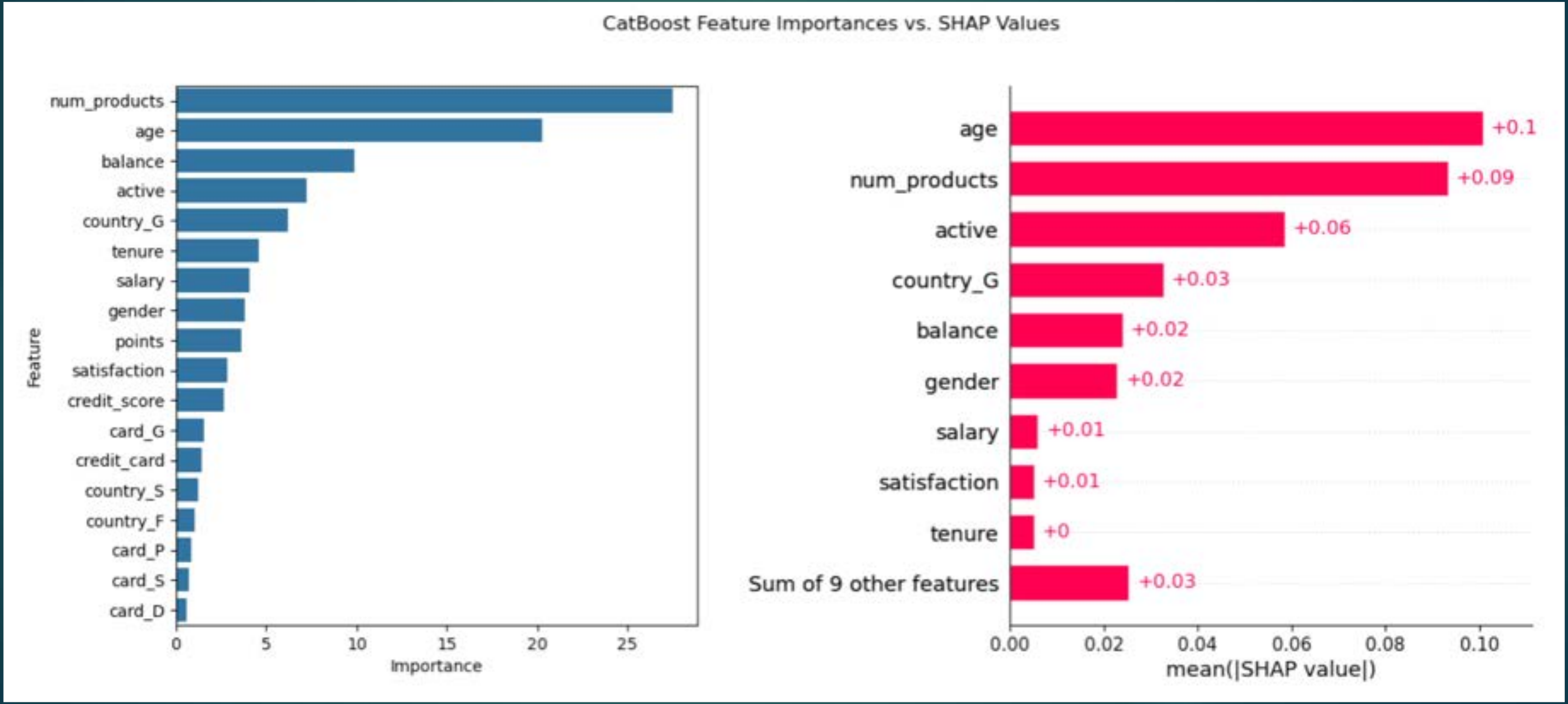
Train Report

	precision	recall	f1-score	support
0	0.91	0.89	0.90	6370
1	0.60	0.65	0.62	1630
accuracy			0.84	8000
macro avg	0.76	0.77	0.76	8000
weighted avg	0.85	0.84	0.84	8000

Test Report

	precision	recall	f1-score	support
0	0.91	0.88	0.90	1592
1	0.60	0.67	0.63	408
accuracy			0.84	2000
macro avg	0.76	0.78	0.77	2000
weighted avg	0.85	0.84	0.84	2000

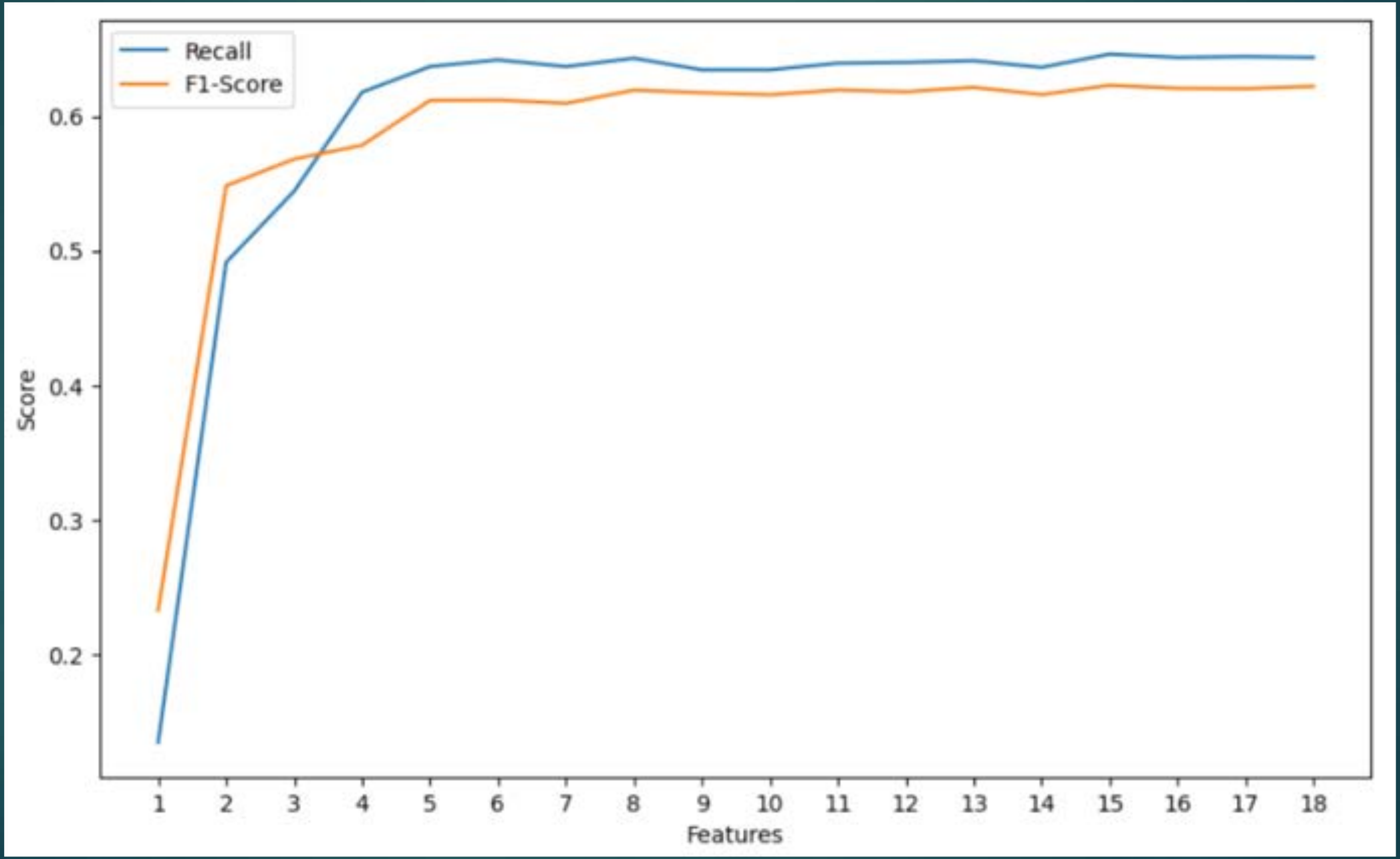
Modeling



CatBoost Feature Importances vs. SHAP Values
num_products
age
balance
active
country_G
tenure
salary
gender
points
satisfaction
credit_score
card_G
credit_card
country_S
country_F
card_P
card_S
card_D
Feature
0
5
10
15
20
25
Importance
age
num_products
active
country_G
balance
gender
salary
satisfaction
tenure
Sum of 9 other features
+0.1
+0.09
+0.06
+0.03
+0.02
+0.02
+0.01
+0.01
+0
+0.03
0.00
0.02
0.04
0.06
0.08
0.10
mean(|SHAP value|)

Modeling

CatBoost Feature Reduction: Recall and F1-Scores on Cross-Validated Train Set



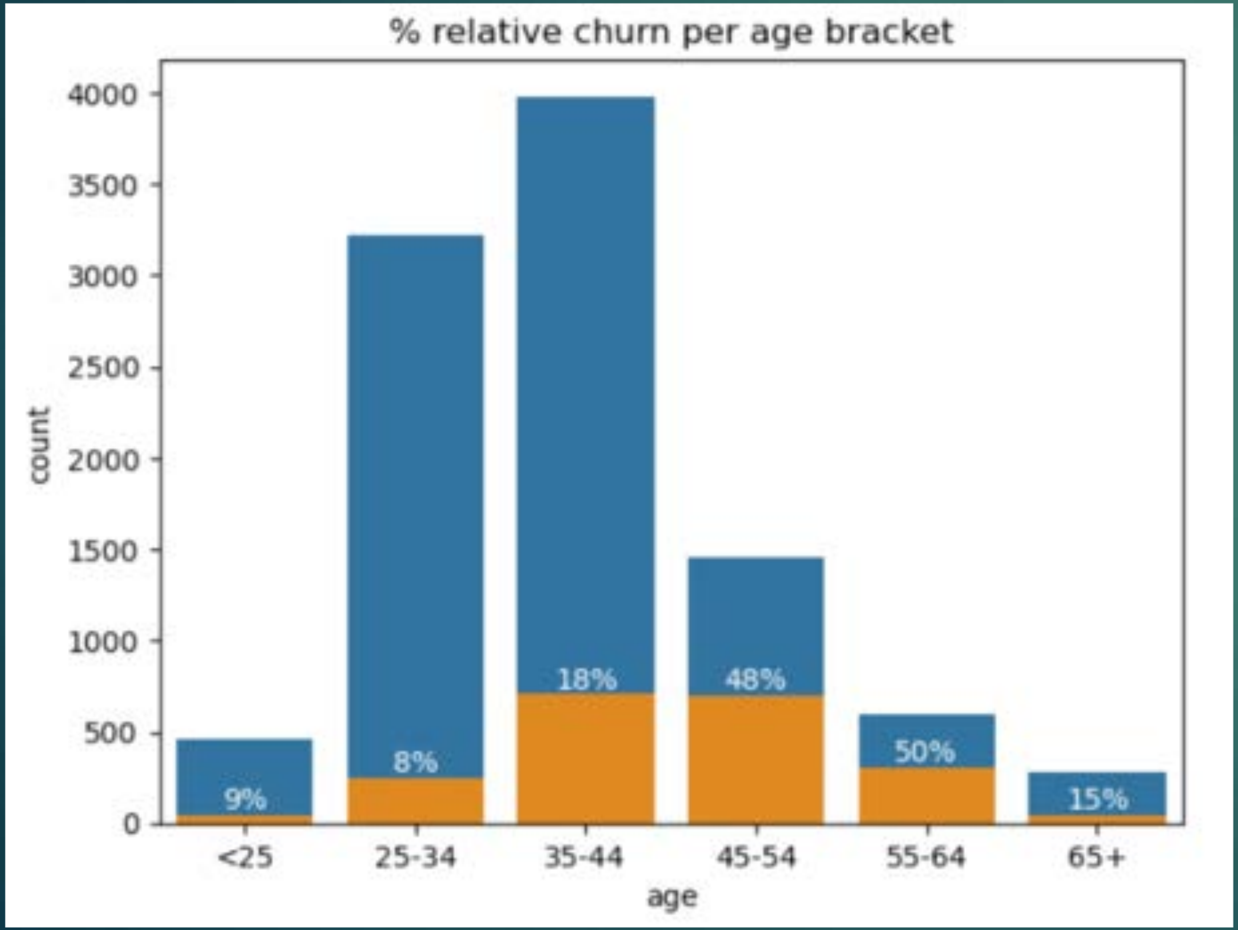
Additional EDA

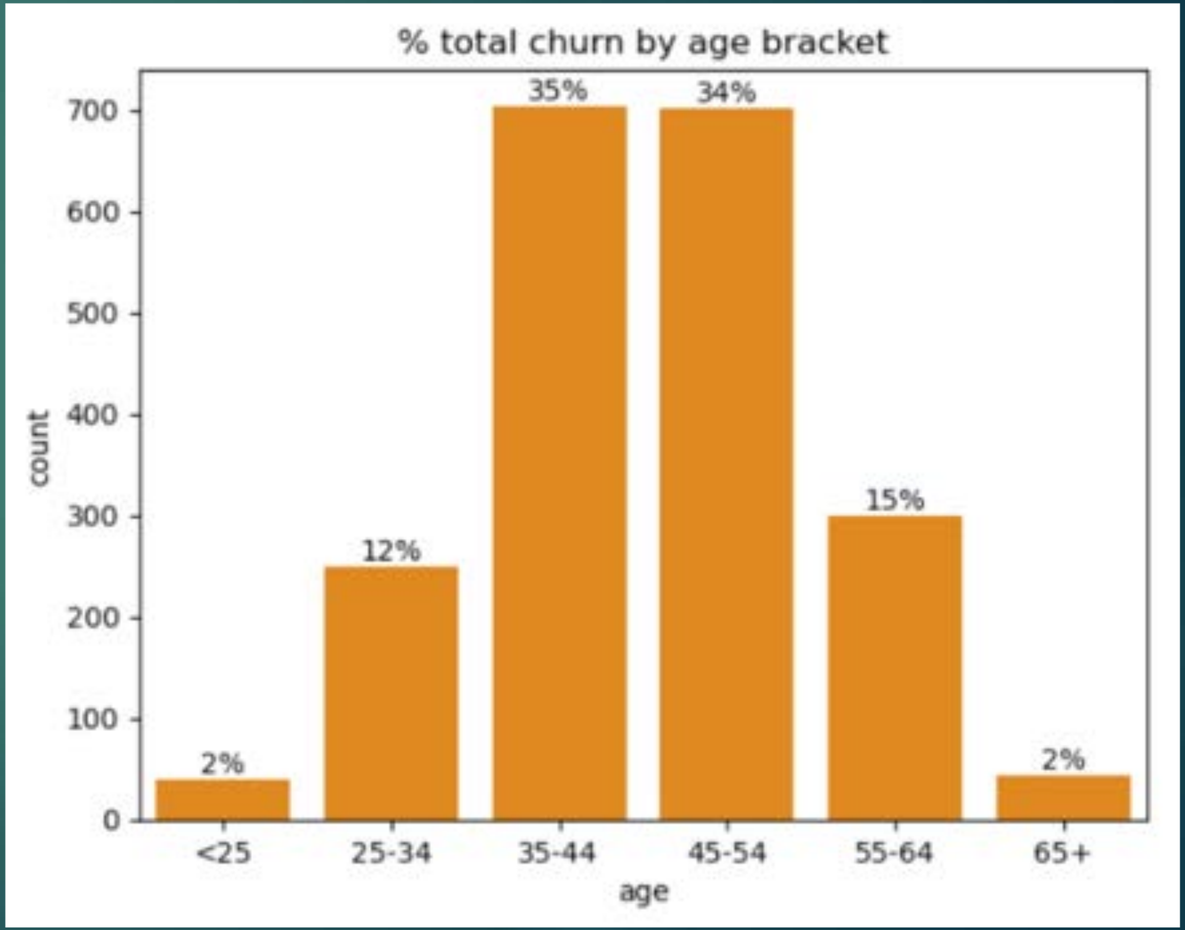
Most relevant features across top 5 models

- age
- number of products
- active
- Germany
- balance
- gender

Additional EDA

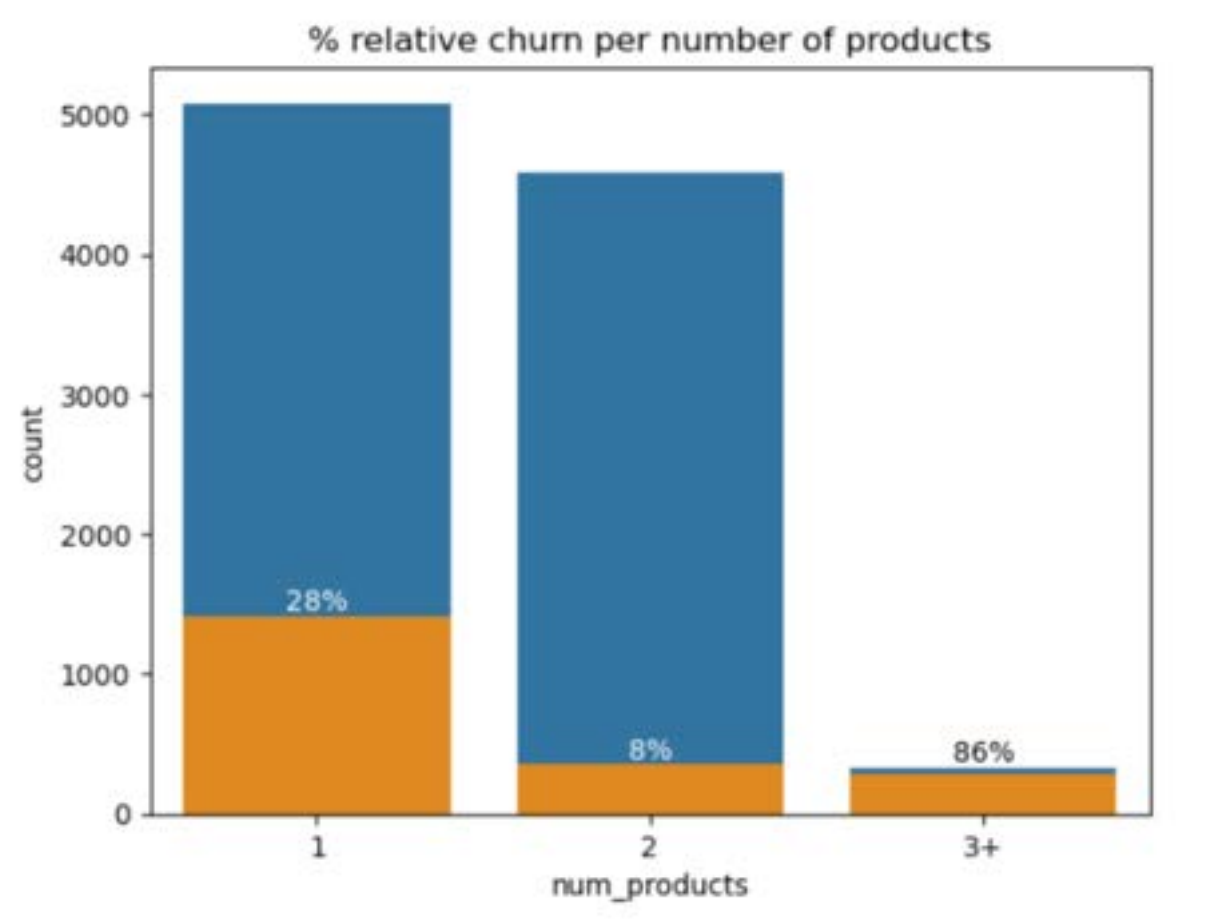
Countplots of Age Bracket and Churn Cases

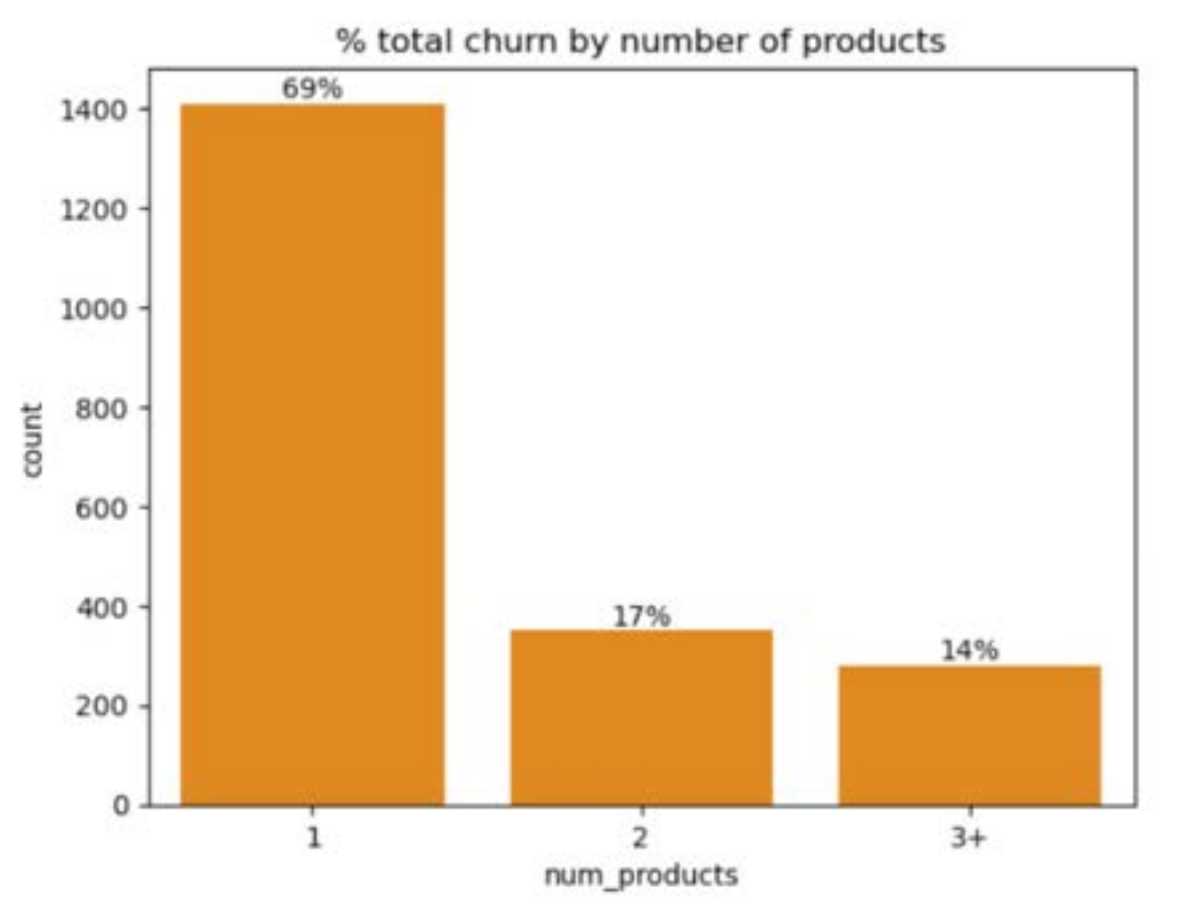




Additional EDA

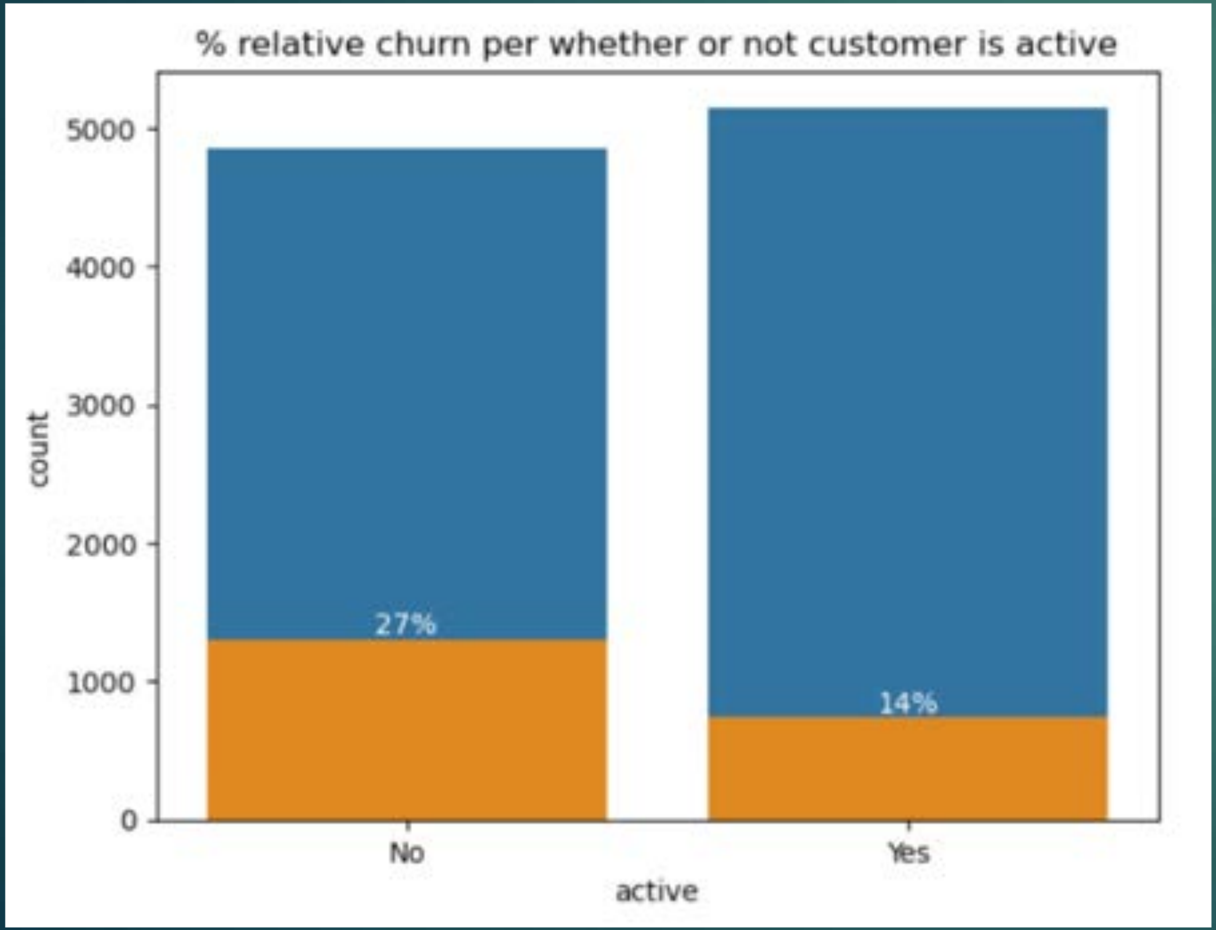
Countplots of Number of Products and Churn Cases

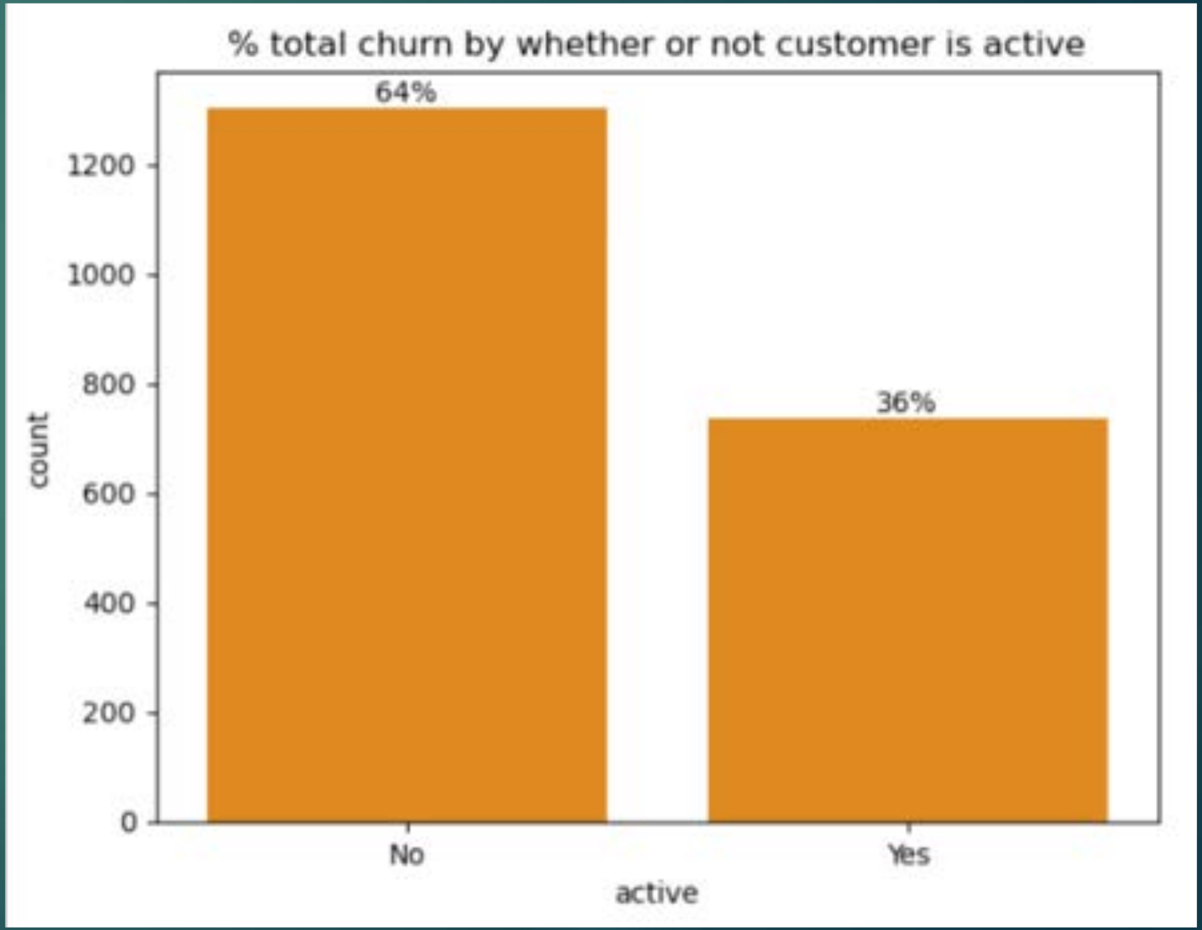




Additional EDA

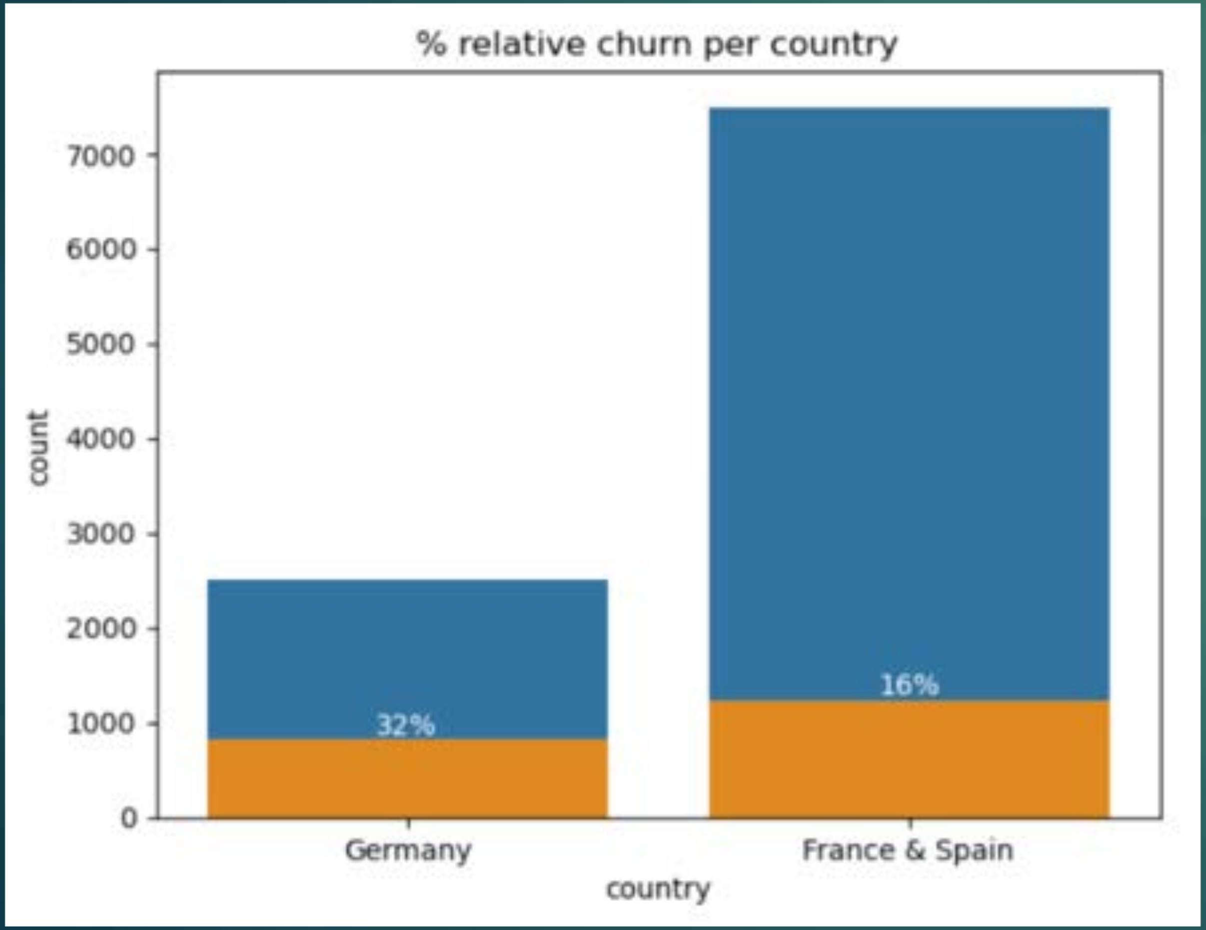
Countplots of Active and Churn Cases

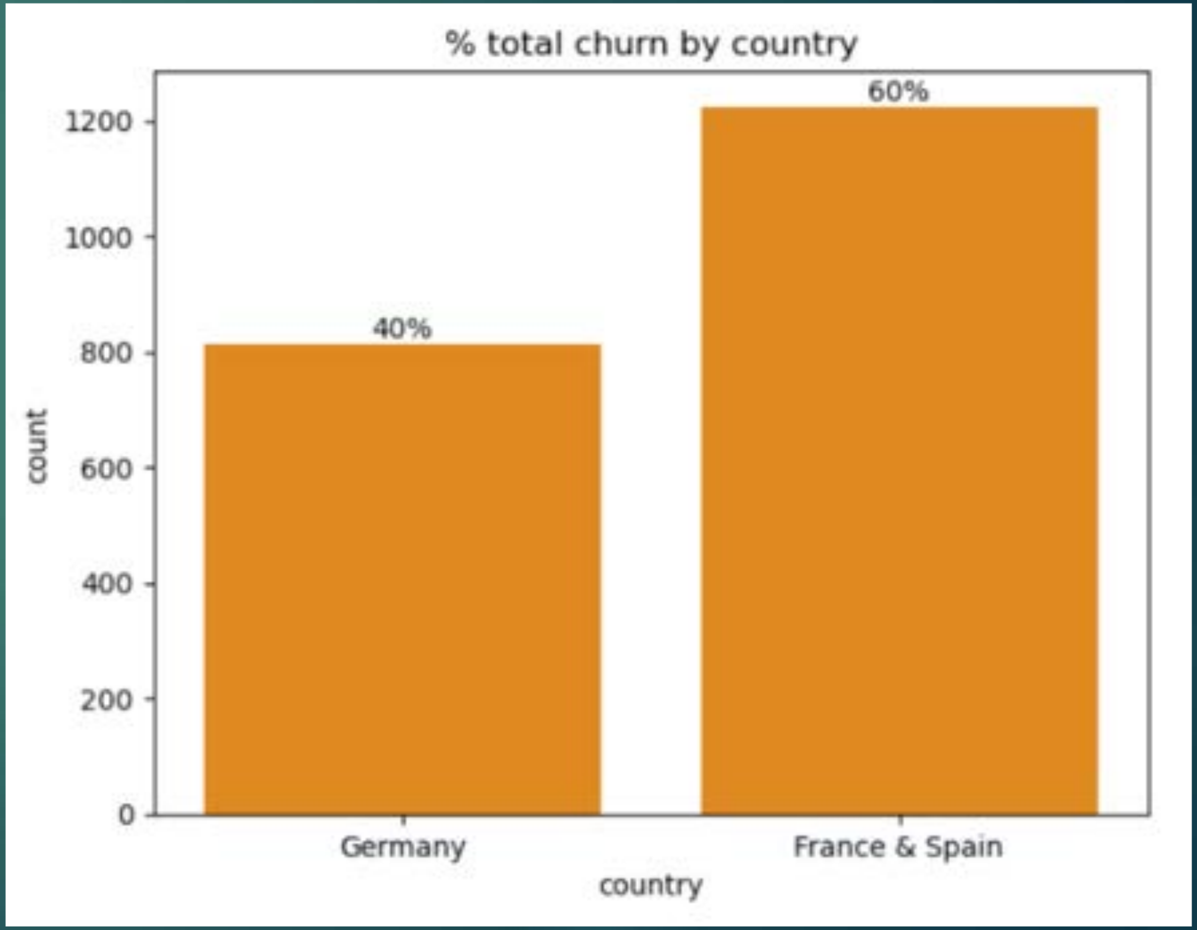




Additional EDA

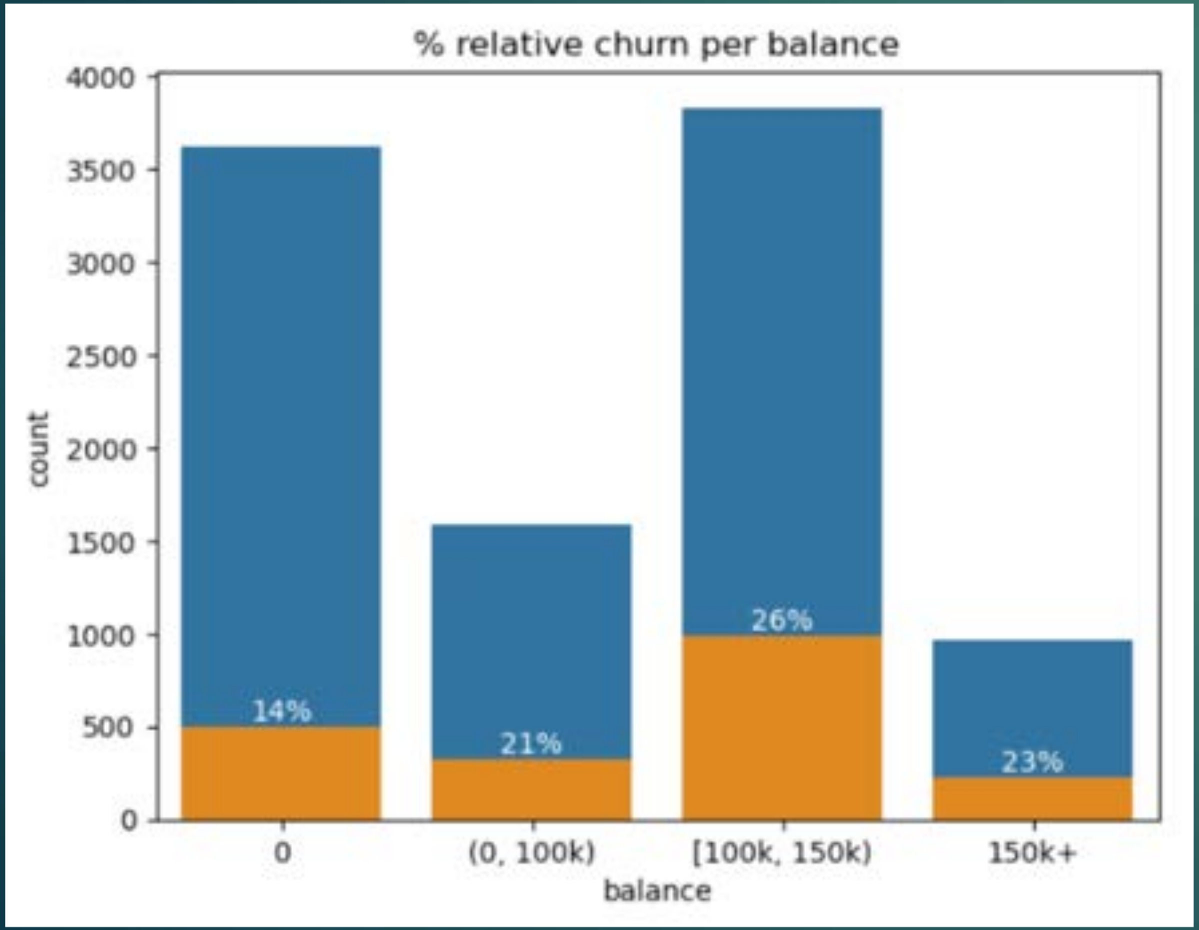
Countplots of Country and Churn Cases

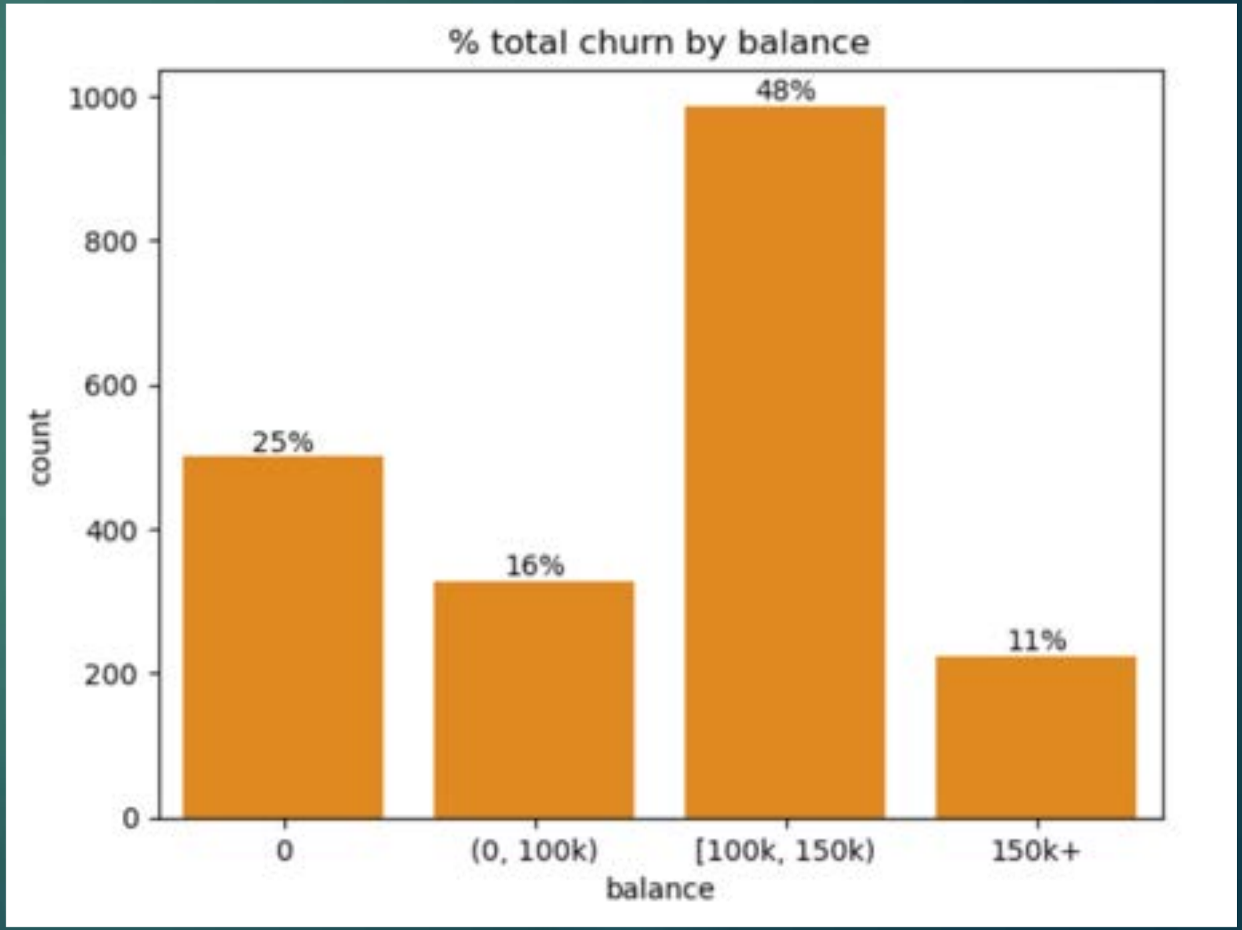




Additional EDA

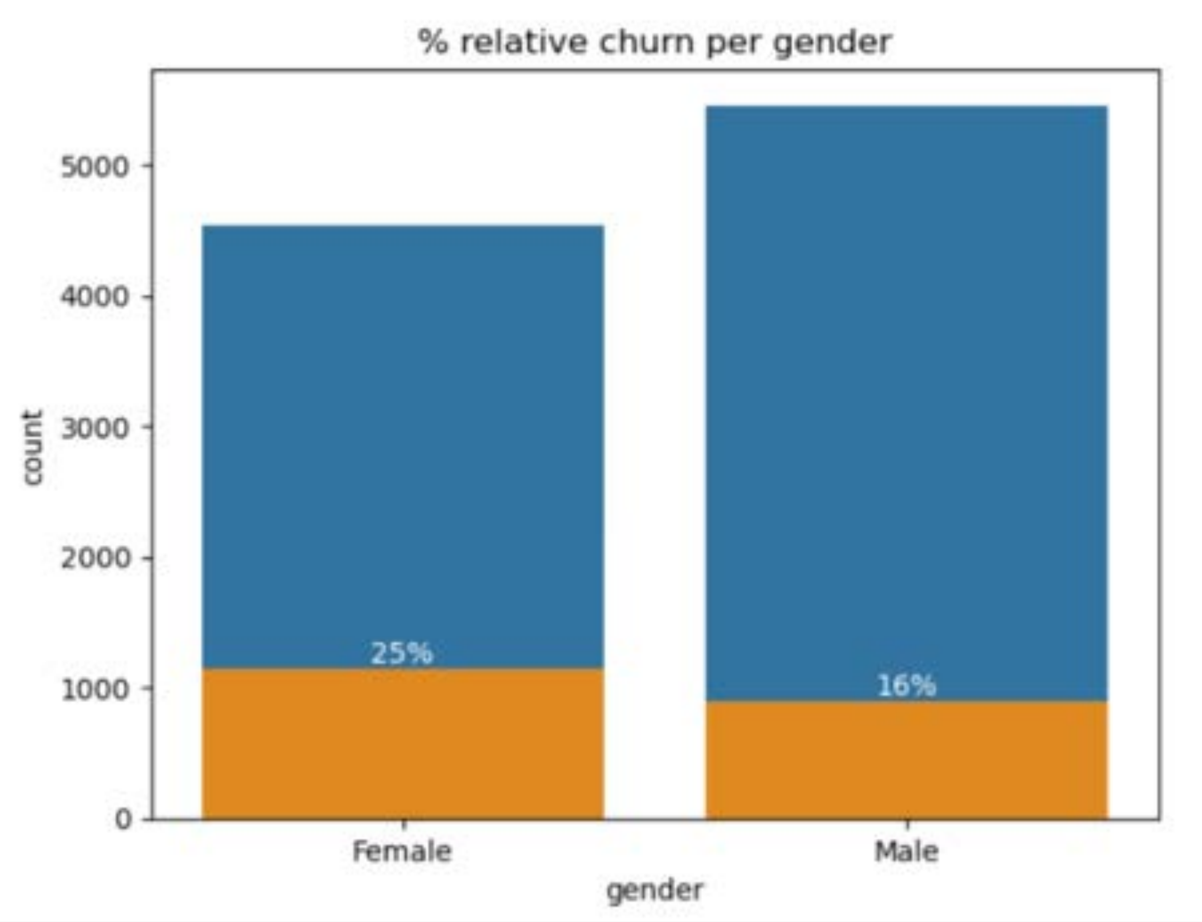
Countplots of Balance and Churn Cases

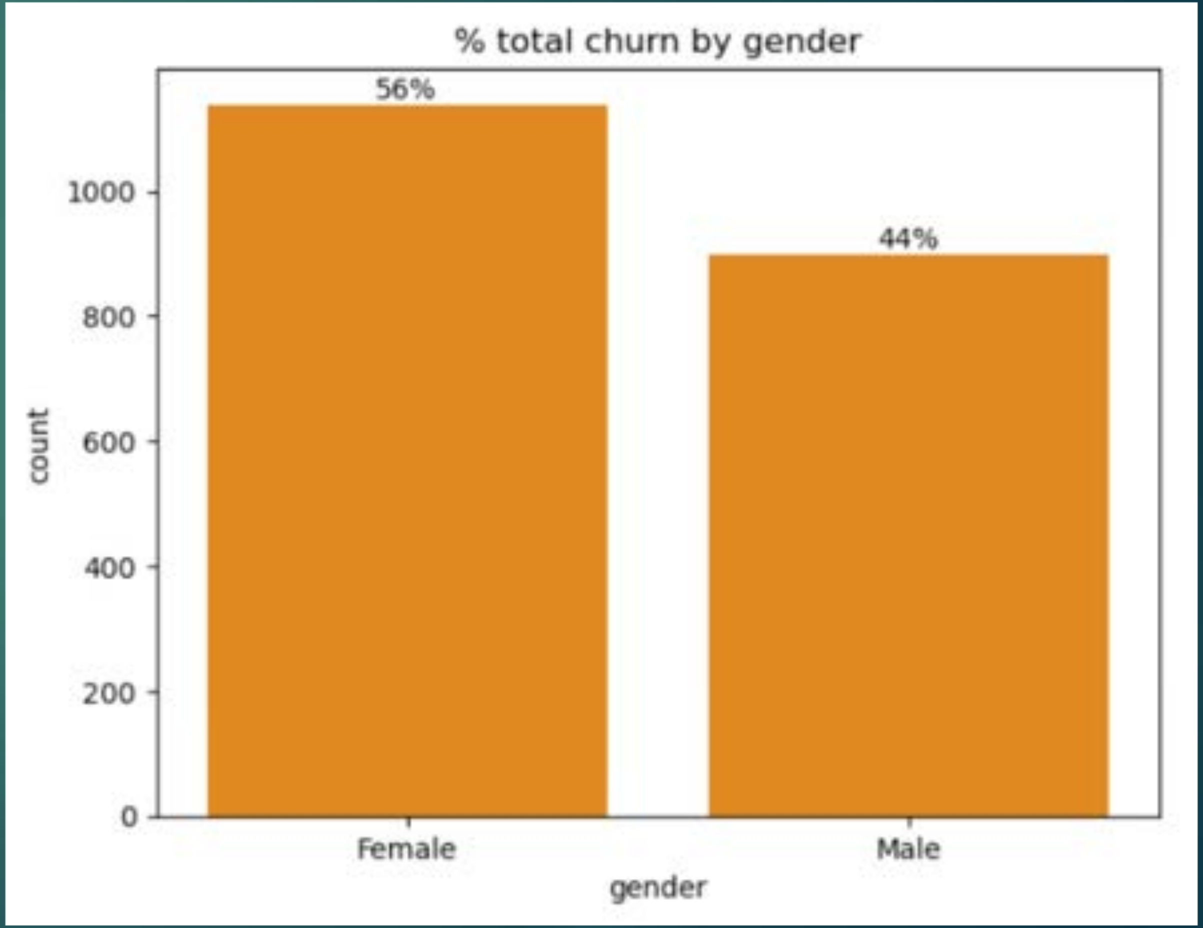




Additional EDA

Countplots of Gender and Churn Cases





Future Implications

- Better complaint resolution
- CatBoost Classifier: 67.4% recall (test set)
- Further investigate key features
- Additional features, larger sample, other models
- Customer segmentation – cluster analysis

Thanks

Photo: https://en.wikipedia.org/wiki/Soci%C3%A9t%C3%A9_G%C3%A9n%C3%A9rale