Predicting Churn Rate for Banks

December 7, 2021

Wharton
UNIVERSITY of PENNSYLVANIA

Agenda: Four Key Components

01 Objectives

What is the Business Question

02 Descriptive Stats

Summary statistics of the dataset

03 Two Models

Logistic Regression Model and Random Forest

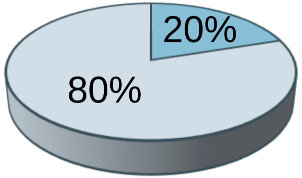
04 Implications

Results Analysis and how they can be implemented

Objectives

CHURN = ADDITION - ATTRITION

Annual Churn Rate of US Credit Provider



Churn Rate

Retention Rate

Earning Back with Reduced Churn Rate

9.9%

For Wireless Carriers [1]

Who should be Targeted?

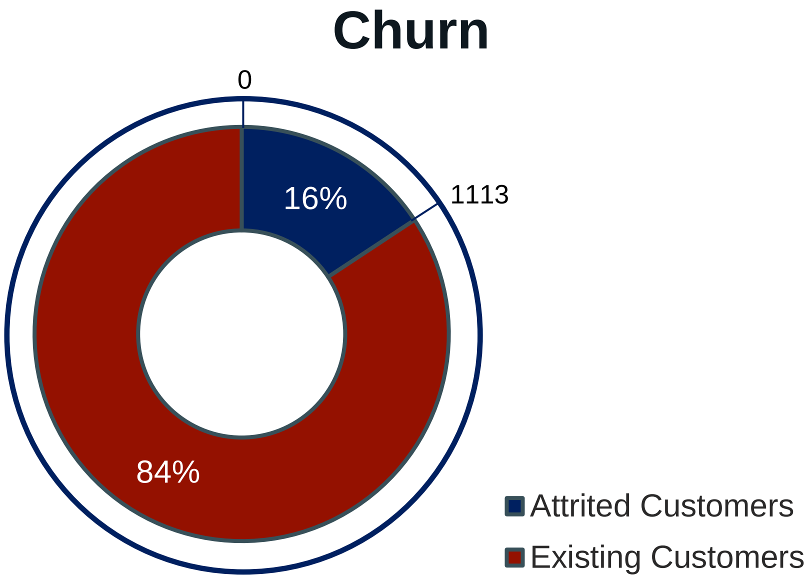
What should the company do?

Churn Rate Prediction:

1. Consumers' Characteristics
2. Churn Rate
3. Classification of customers and their reaction to incentives
4. Target higher-value customers who are most likely to defect and and respond to incentives

Wharton
UNIVERSITY of PENNSYLVANIA

Descriptive Statistics of the Dataset



Source: Kaggle (https://www.kaggle.com/sakshigoyal7/credit-card-customers) [2]

7051
Cases

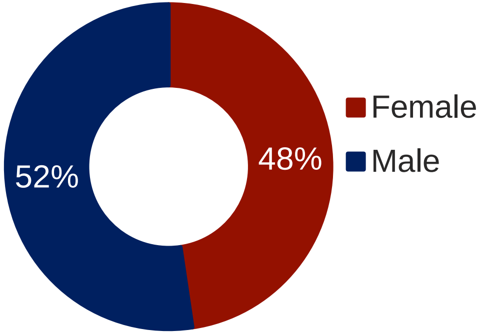
12
Features

- **Age**
- **Salary**
- **Marital Status**
- **Credit Card Limit**
- **Credit Card Category**
- **...**

Wharton
UNIVERSITY of PENNSYLVANIA

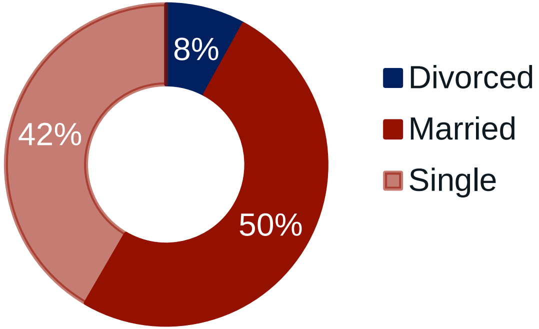
Exploratory Data Analysis—Demographic Features

Gender



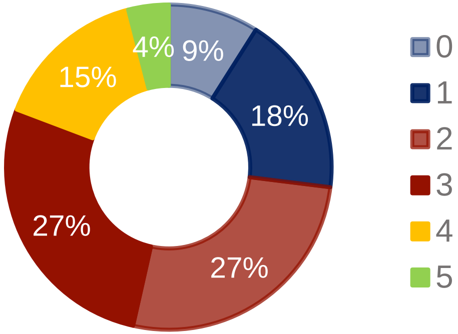
More Females

Marital Status



Married or Single

Dependent Count

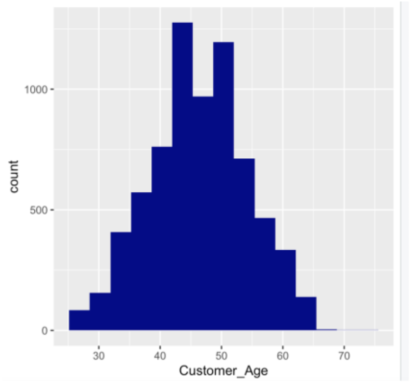


Mostly 2-3

Wharton
UNIVERSITY of PENNSYLVANIA

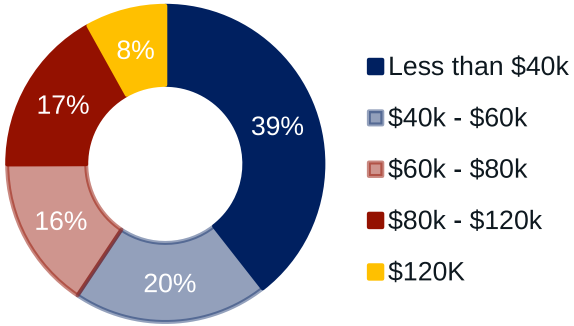
Exploratory Data Analysis—Demographic Features

Age Distribution



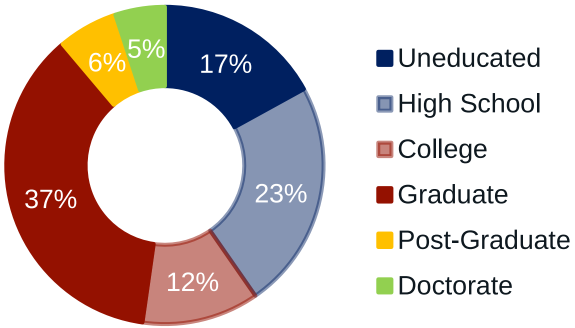
Fairly normal distribution

Income Category



>50% are below $60K

Education Level

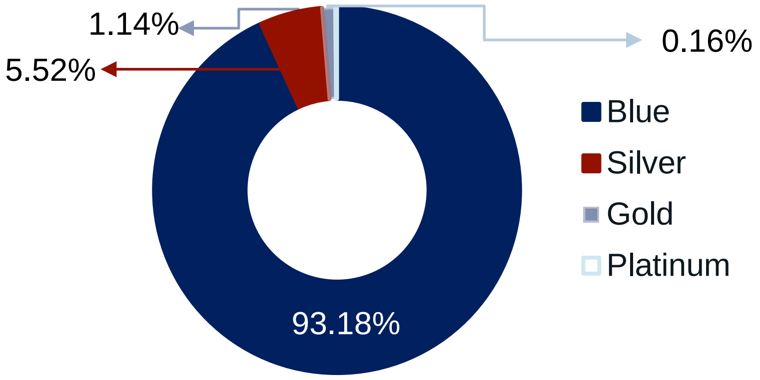


>50% with higher education

Wharton
UNIVERSITY of PENNSYLVANIA

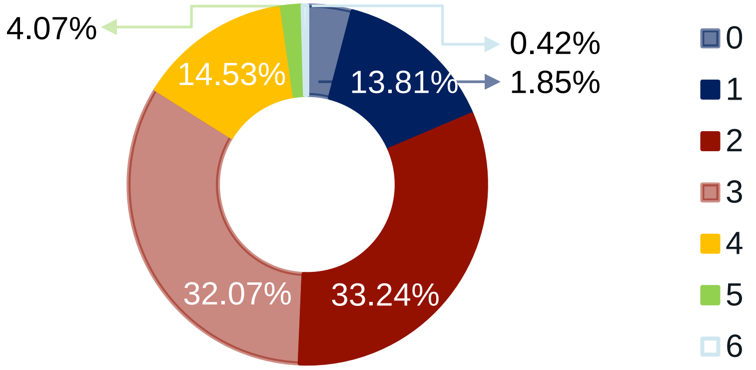
Exploratory Data Analysis—Possible Key Features

Type of Card



Blue Card is Dominant

No. of Contacts in the Last 12 Months

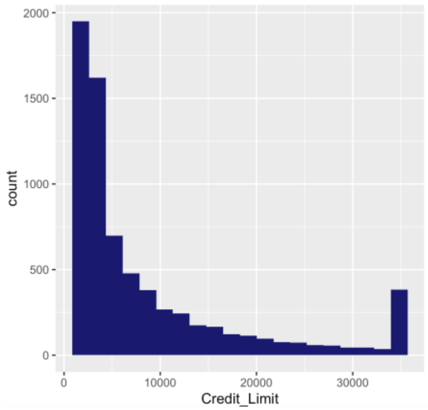


Mostly 2-3 Contacts

Wharton
UNIVERSITY of PENNSYLVANIA

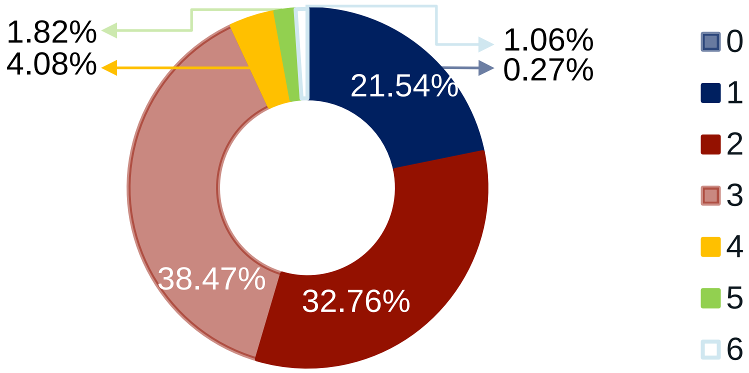
Exploratory Data Analysis—Possible Key Features

Credit Limit on the Credit Card



More on Both Ends

No. of Months Inactive in the Last 12M

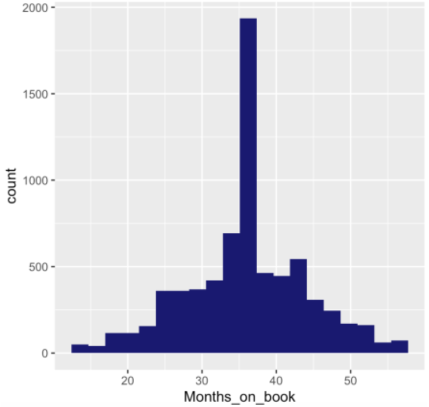


Mostly 2-3 Months

Wharton
University of Pennsylvania

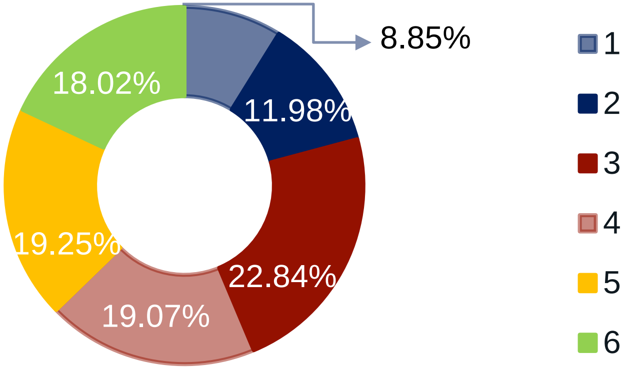
Exploratory Data Analysis—Other Features

Period of Relationship with Bank



Mostly with 3-year Relationship

Total no. of Products Held by Customers



Mostly with More than Two Products

Wharton
UNIVERSITY of PENNSYLVANIA

Using **Logistic Regression Model** to predict Churn Rate

Education Level
College

Income Category
Above $120k

Baselines for Logistic Regression

Marital Status
Divorced

Card Category
Blue

Logistic Regression Output

Important Factors

Gender (Male)
Income Category (0-80k)
Total Relationship Count

Education Level (Doctorate)
Card Category (Platinum)
Months Inactive (12 months)
Contacts count (12 months)

```
Coefficients:
                              Estimate Std. Error z value Pr(>|z|)
(Intercept)                 -2.178e+00  3.696e-01  -5.893 3.79e-09 ***
Customer_Age                 1.611e-03  7.149e-03   0.225 0.821714
GenderM                     -4.878e-01  1.331e-01  -3.666 0.000247 ***
Dependent_count              4.241e-02  2.734e-02   1.551 0.120825
Education_LevelDoctorate     3.512e-01  1.714e-01   2.049 0.040481 *
Education_LevelGraduate     -2.239e-02  1.151e-01  -0.195 0.845718
Education_LevelHigh School  -1.143e-01  1.243e-01  -0.920 0.357624
Education_LevelPost-Graduate 2.303e-01  1.670e-01   1.379 0.167797
Education_LevelUneducated   -7.338e-02  1.309e-01  -0.561 0.575079
Marital_StatusMarried       -1.203e-01  1.312e-01  -0.917 0.359130
Marital_StatusSingle         7.053e-03  1.319e-01   0.053 0.957362
Income_Category$40K - $60K  -7.652e-01  1.797e-01  -4.258 2.06e-05 ***
Income_Category$60K - $80K  -5.025e-01  1.564e-01  -3.212 0.001316 **
Income_Category$80K - $120K -1.964e-01  1.446e-01  -1.358 0.174379
Income_CategoryLess than $40K -7.465e-01 1.952e-01 -3.824 0.000132 ***
Card_CategoryGold            1.124e-01  3.413e-01   0.329 0.741927
Card_CategoryPlatinum        6.101e-01  7.136e-01   0.855 0.392587
Card_CategorySilver          2.553e-01  1.764e-01   1.447 0.147917
Months_on_book              -3.271e-03  7.125e-03  -0.459 0.646170
Total_Relationship_Count    -3.223e-01  2.327e-02 -13.853  < 2e-16 ***
Months_Inactive_12_mon       4.254e-01  3.403e-02  12.500  < 2e-16 ***
Contacts_Count_12_mon        5.862e-01  3.397e-02  17.255  < 2e-16 ***
Credit_Limit                -1.843e-05  5.864e-06  -3.144 0.001669 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Wharton
UNIVERSITY of PENNSYLVANIA

Logistic Regression Confusion Matrix

Predict \ Actual	Retain (0)	Churn(1)
0	5928	1017
1	40	96

Accuracy: 0.851

Predicted To churn (1), ended up retain (0)
40 cases

Predicted to churn (1), ended up churn (1)
96 cases

Predicted to retain (0), ended up churn (1)
1017 cases

Predicted to retain (0), ended up retain (0)
5928 cases

Logistic Regression Confusion Matrix

Predict \ Actual	Retain (0)	Churn(1)
0	5928	1017
1	40	96

Accuracy: 0.851

Among the False Predictions:

Predicted To churn (1), ended up retain (0)
40 cases

Predicted to retain (0), ended up churn (1)
1017 cases

More False-Negatives

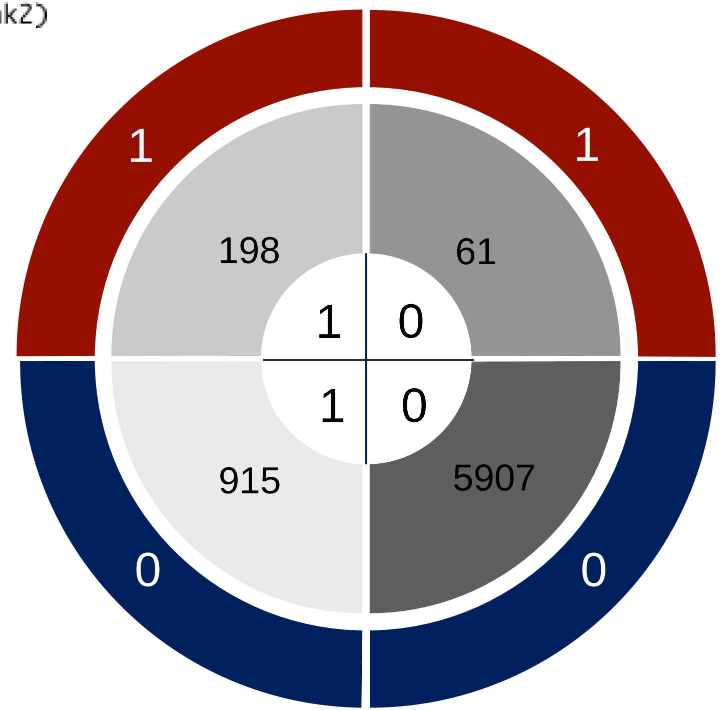
Output for **Random Forest** Model

```
Call:
 randomForest(formula = factor(Attrition_Flag) ~ ., data = bank2)
               Type of random forest: classification
                     Number of trees: 500
No. of variables tried at each split: 3

        OOB estimate of  error rate: 13.78%
Confusion matrix:
     0   1 class.error
0 5907  61  0.01022118
1  915 198  0.82210243
```

```
> mean(bank2$Attrition_Flag==rf$predicted)
[1] 0.8621664
> table(bank2$Attrition_Flag, rf$predicted)

      0    1
  0 5907   61
  1  915  198
```



Wharton UNIVERSITY of PENNSYLVANIA

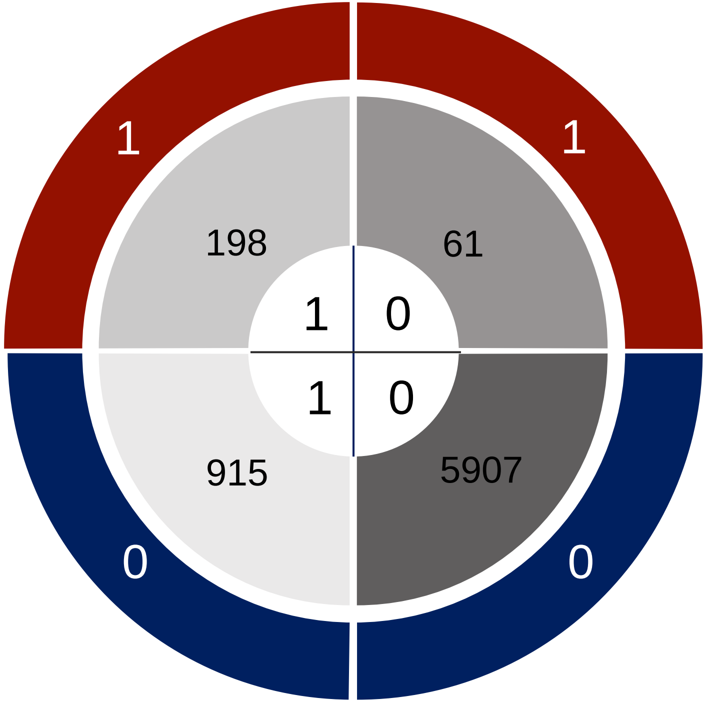
Output for **Random Forest** Model

Among the False Predictions:

Predicted To churn (1), ended up retain (0)
915 cases

Predicted to retain (0), ended up churn (1)
61 cases

More False-Positives



Implications of the Project—Comparison between Logistic Regression and Random Forest

	Accuracy Level	Able to Describe Significance of Variables	False Prediction Handling
Logistic Regression	85.1%	Yes (The Coefficients)	Generate more False-Negatives
Random Forest	86.22%	No	Generate more False-Positives

Wharton
UNIVERSITY of PENNSYLVANIA

Implementations of the Results—Suggestions

Customers with Doctorate Education → Adjust their promotion activities targeting the needs of the customer with higher education level

Customers with Income between 40k to 60k → Adjust the annual fees and other benefits to meet the need of the group

Customers with a Platinum Card → Suggesting that the benefits of the platinum card are not attractive enough to the customers or the cost is too high

NEXT STEP? →
1. Design a detailed questionnaire: reasons why these groups fairly prefer to leave
2. Create the targeting marketing strategies

Wharton
UNIVERSITY of PENNSYLVANIA

Team Members


Siyuan Xu


Fengsui Xie


Jie Xu


Yufei Qin


Xinyuan Hu

Wharton
UNIVERSITY of PENNSYLVANIA

Reference

1. Forbes: https://www.forbes.com/sites/hbsworkingknowledge/2013/11/11/a-smarter-way-to-reduce-customer-churn/?sh=22e02bab2c0a

2. Kaggle: https://www.kaggle.com/sakshigoyal7/credit-card-customers

Wharton
UNIVERSITY of PENNSYLVANIA