nuveen
A TIAA Company

Applied Analytics Capstone – Final Project

Bryant University

nuveen
A TIAA Company

Agenda

- Background
- Objectives
- Methodology
- Variables
- Findings
- Recommendations

Background

1

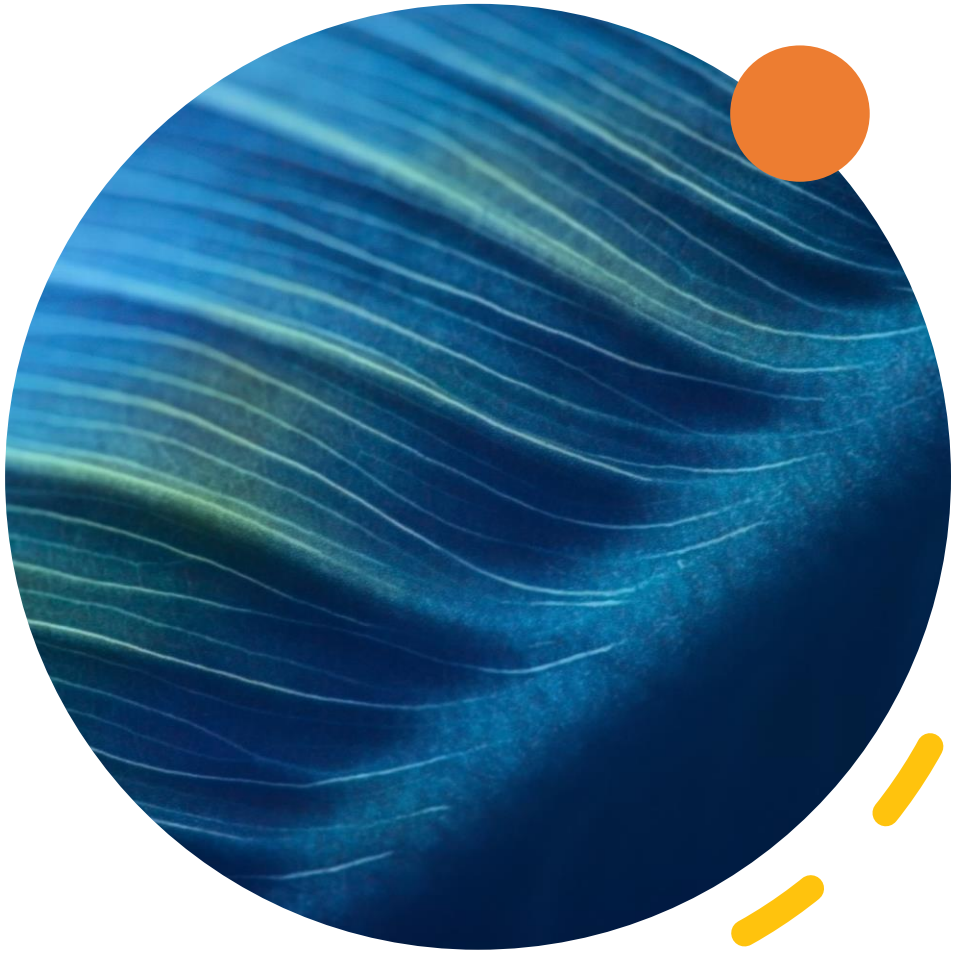
Nuveen, formerly Nuveen Investments, a subsidiary of TIAA, is an American global investment manager.

2

Nuveen is one of the largest investment managers in the world with $1.1 trillion in assets under management.

3

Nuveen offers a range of investment capabilities across income, equities, alternatives and multi-asset solutions.



nuveen
A TIAA Company

Objectives

Goal: Analyze response data set, to optimize & predict an Advisor's likelihood to increase sales.

Data Provided: An array of advisor sales activity, web activity, and 2018 transactions with clients.

50% was used to train the model
50% was used to test the model

Methodology

EDA: Data cleaning (Eliminated errors and outliers) Created new dependent and independent variables.

Analyze Data: Developed multiple machine learning models, and selected the one with the best performance.

Analysis of the Results: Validated the model and provided recommendations.

Solution:
Produced a creative solution that allows NEUVEEN to capitalize on model to identify potential targets for development.

Dependent Variables Selected

- Sales Transactions > $10K, 2018
- Redemptions > $10K ,2018
- Redemptions 2018
- Sales 2018
- 2018, Funds Sold > $10K
- Asset Rank
- Sales Rank

Target Variable = Sales Change as a Binary Variable (1 = Increase, 0= Decrease)

Positive & Negative Influencers

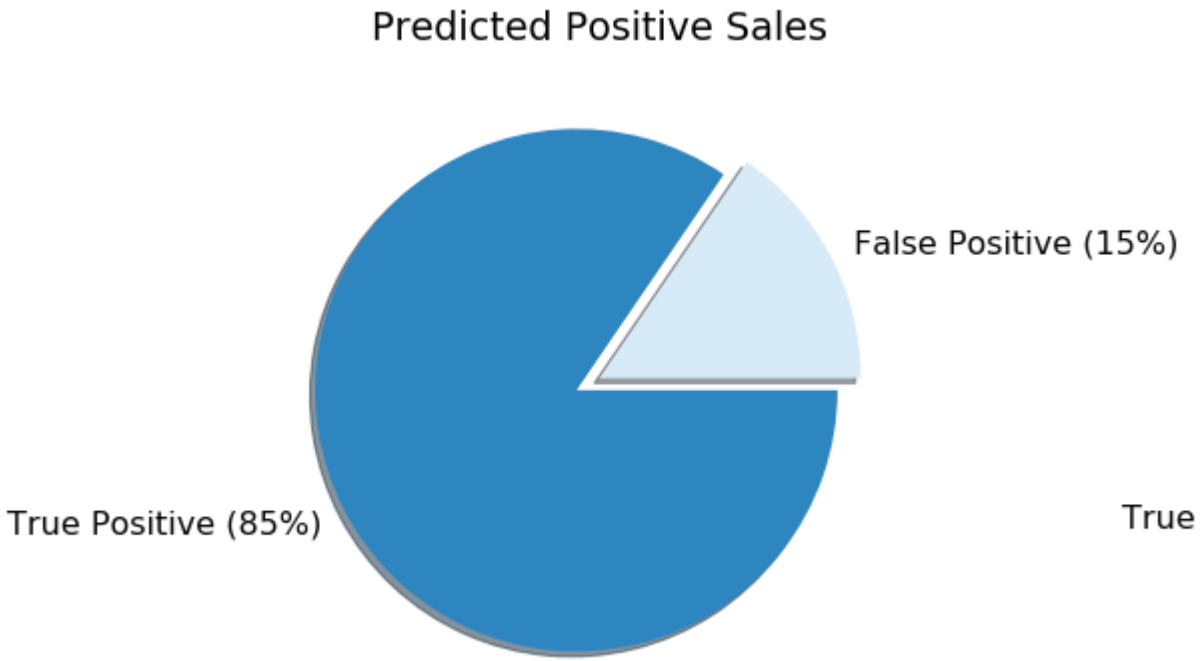
Positive
Redemptions Rank
Sales Rank
Redemptions < 10K, 2018
Sales < 10K, 2018
Sales Transactions > 10K, 2018
Open
Redemptions >10K, 2018
Clickstream
Meetings
Funds Sold > 10K, 2018
Funds Sold < 10K, 2018

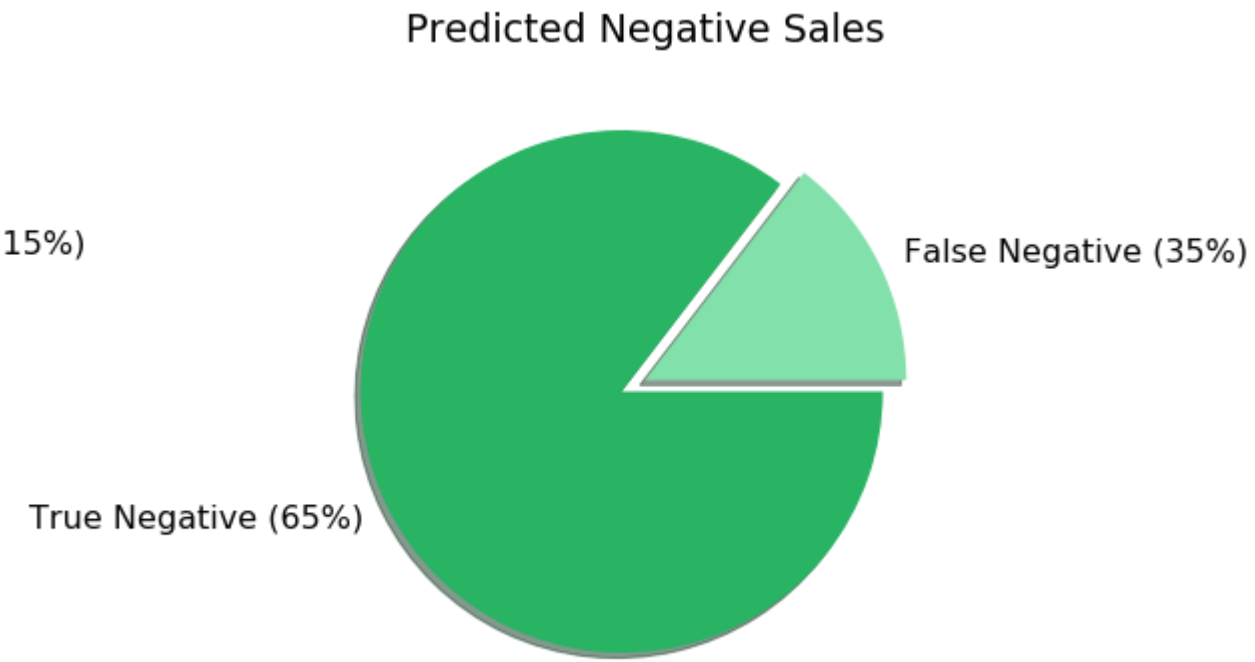
Negative
Redemptions > 10K, 2018
Sales 2018

An advisor's likelihood to generate increase in sales YOY is highly influenced by their Redemptions Rank, Sales Rank, and Number of Redemptions under 10K

Logistic Regression

Results show that we have a model that is very accurate

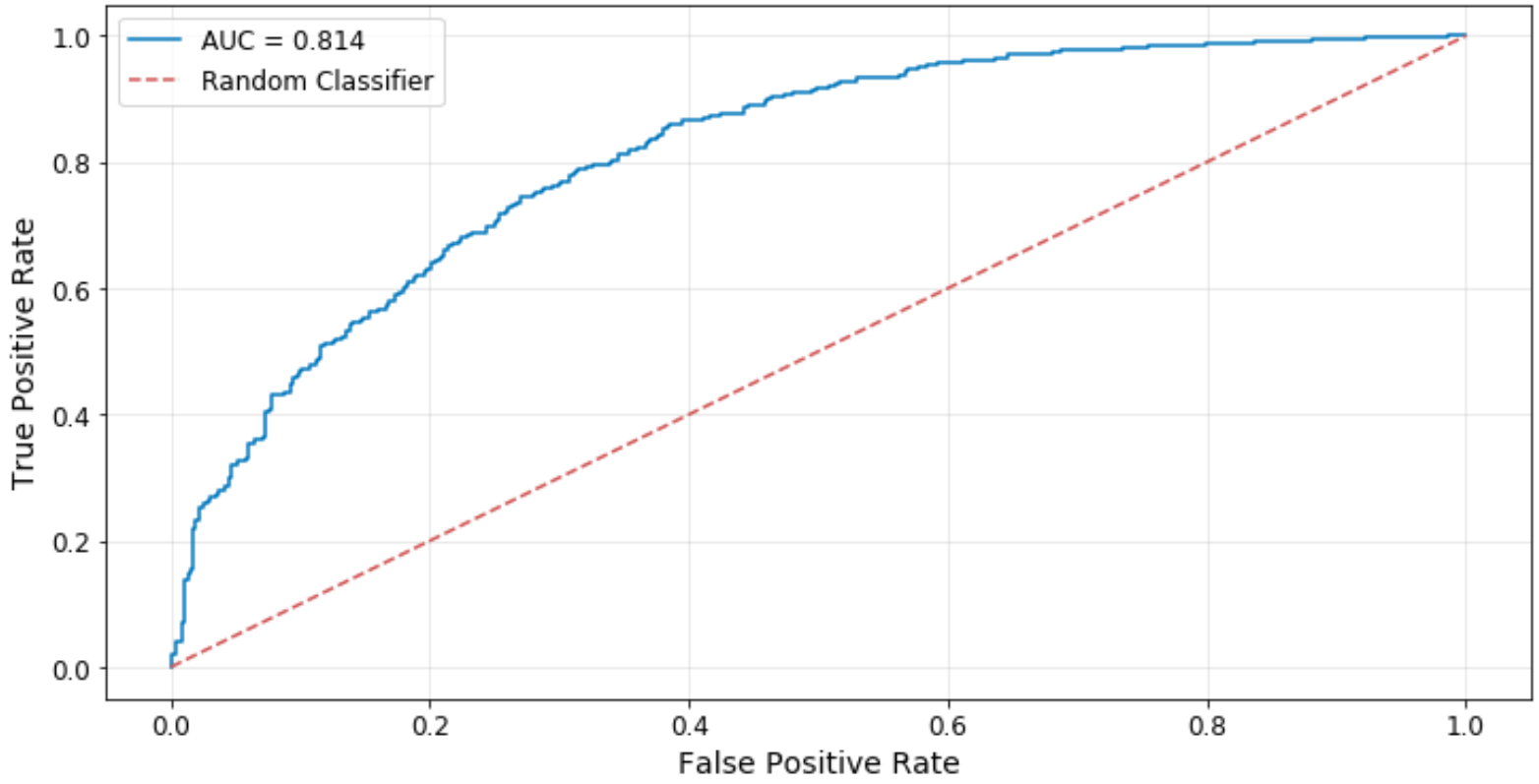




ROC Curve & AUC

Our Model works well throughout the curve and deteriorates slowly

ROC Curve shows model distinguishes well between Positive/Negative Sales

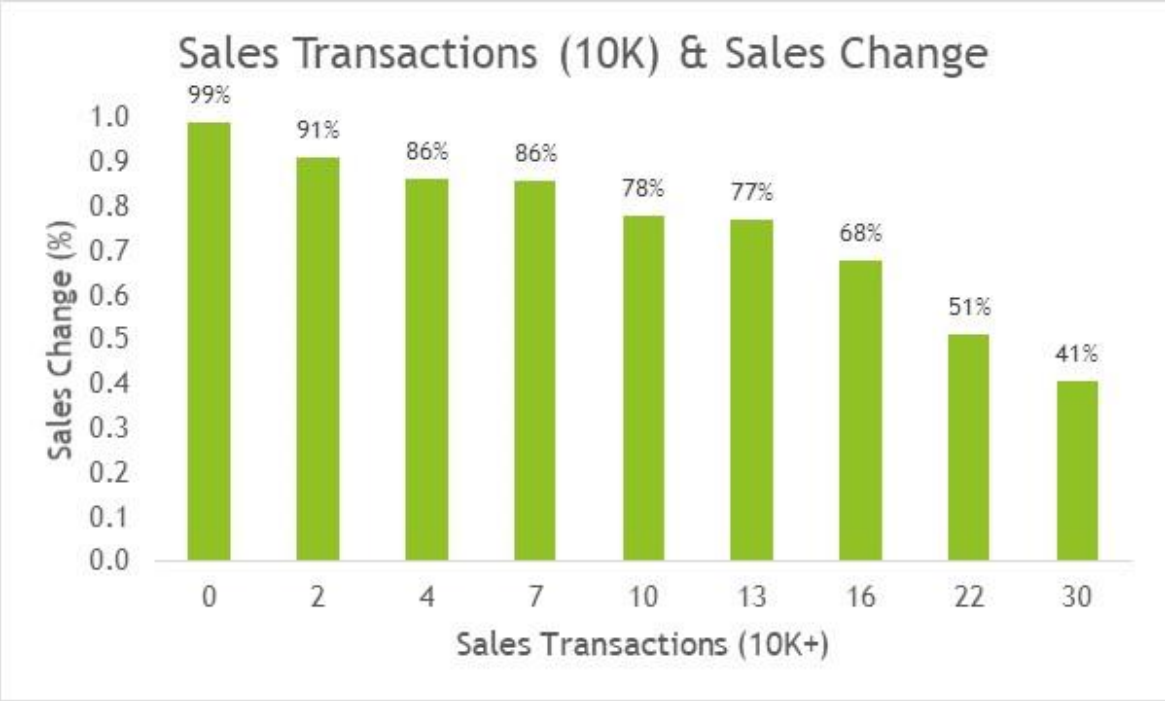


Our model performs better than just randomly selecting Advisors to focus on.

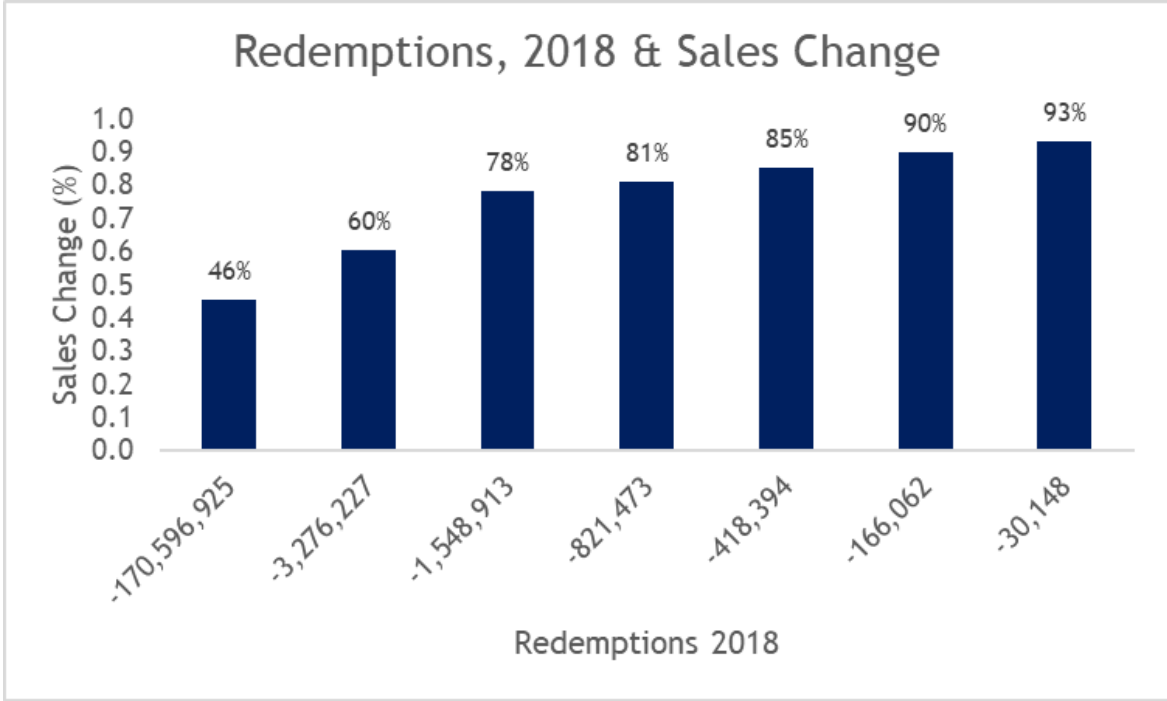
The model is most effective around the 40% point (greatest linear distance from random classifier)

nuveen
A TIAA Company

Variables

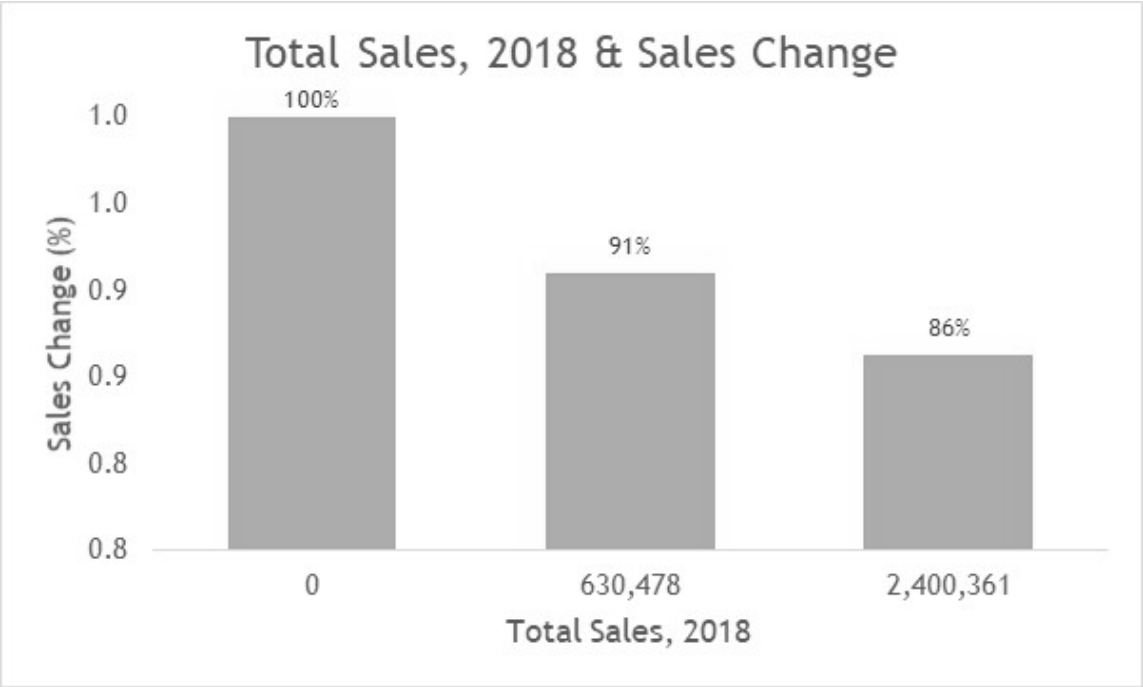


Advisors with Transactions Lower 10K+ Transactions are more likely to see an increase in sales

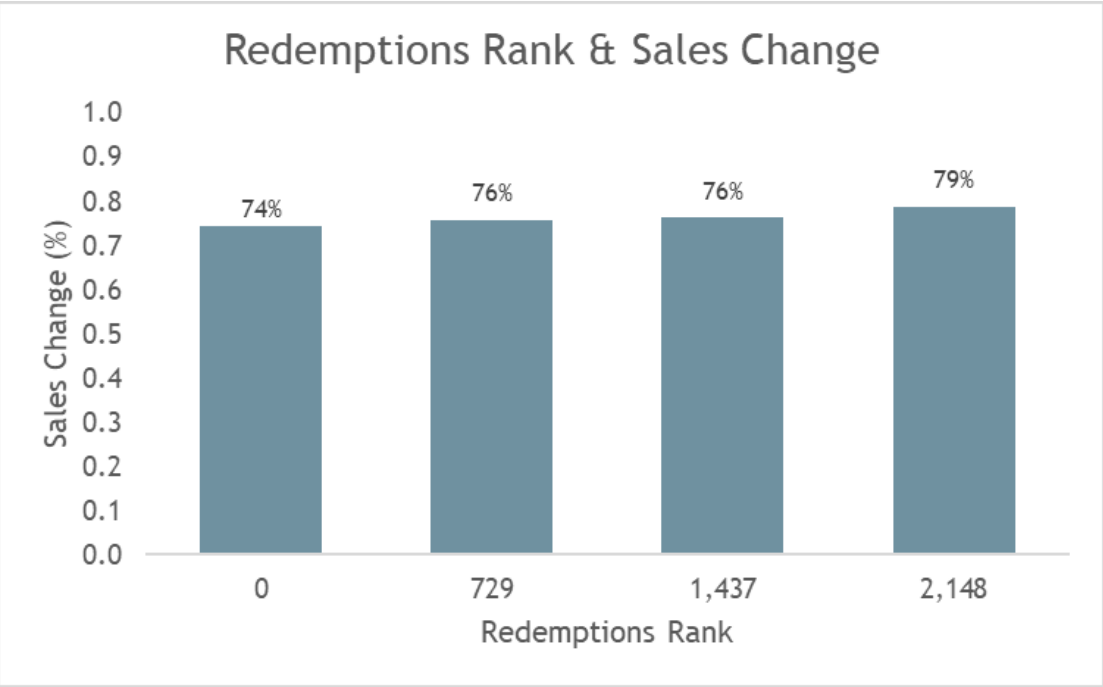


Advisors with Lower Redemptions are more likely to see an increase in sales in 2019

Variables (Cont.)



Advisors with lower total sales are more likely to achieve higher sales for 2019



Advisors with lower sales rank are more likely to see an increase in sales

nuveen
A TIAA Company

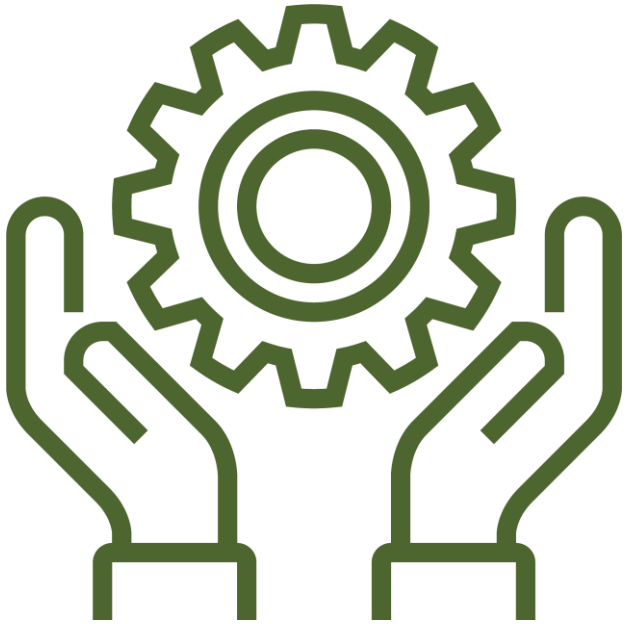
Lift Chart

	Number of	Increase in Sales	Lift Over	Cumulative	Cumulative Sales Incr.	Cumulative
Decile	Advisors	No. Advisors	Average	No. of Advisors	No. Advisors	Lift
1	199	194	27%	199	194	27%
2	199	193	26%	398	194	26%
3	199	183	20%	597	190	24%
4	199	170	11%	796	185	21%
5	199	167	9%	995	181	19%
6	199	152	-1%	1,194	177	15%
7	199	157	3%	1,393	174	14%
8	199	140	-8%	1,592	170	11%
9	199	120	-22%	1,791	164	7%
10	199	49	-68%	1,990	153	0%
Total	**1,990**	**153**	**0%**			

The advisors from deciles 1-5 are more likely to increase sales than the average.

Most advisors are increasing sales in 6-9, but less than the average. Decile 10 advisors are unlikely to increase sales.

Key Takeaways & Suggestions



- Introduce a Development Strategy that focuses on increasing sales at the lower deciles (Decile 6-9).
- Introduce a Retention Strategy for advisors in the top deciles (1-5), they are a fundamental part of the business.
- Give special attention to Advisors in the 10th decile, as these are less likely to increase sales – they have very valuable clients!

Thank you!

Questions?

Appendix

Additional Information

nuveen
A TIAA Company

Independent Variables

- Total Sales 2018
- Net Sales 2018 (Total Sales 2018-Total Redemptions 2018)
- Total Redemptions 2018
- Sales Transactions at $1+
- Sales Transactions at $10k+
- Redemption Transactions at $1+
- Redemption Transactions at $10k+

nuveen
A TIAA Company

Independent Variables

- Number of Funds Sold at $1+
- Number of Funds Sold at $10k+
- Meetings
- Completed Connects
- Overall Connects
- Email Opens
- Email Click Throughs
- Click Stream

nuveen
A TIAA Company

Variable Selection

1. "Voting" Variables Based on Supervised Machine Learning Attributes
2. Variables that Scored the Highest were kept in the model
3. Variables that Scored the Lowest were dropped from the model
4. Total of 23 Variables:
    a. Highest Score = 6
    b. Lowest Score = 1

Variable	IV	RF	Extratrees	Chi Square	RFE	L1	Final Score
Sales Transactions > 10K, 2018	1	1	1	1	1	1	6
Redemptions > 10K, 2018	1	1	1	1	1	0	5
Redemtions 2018	1	1	1	1	1	0	5
Sales 2018	1	1	1	1	1	0	5
2018, Funds Sold >10K	1	0	0	0	1	1	3
Asset Rank	0	0	0	1	1	1	3
Sales Rank	0	1	0	0	1	1	3

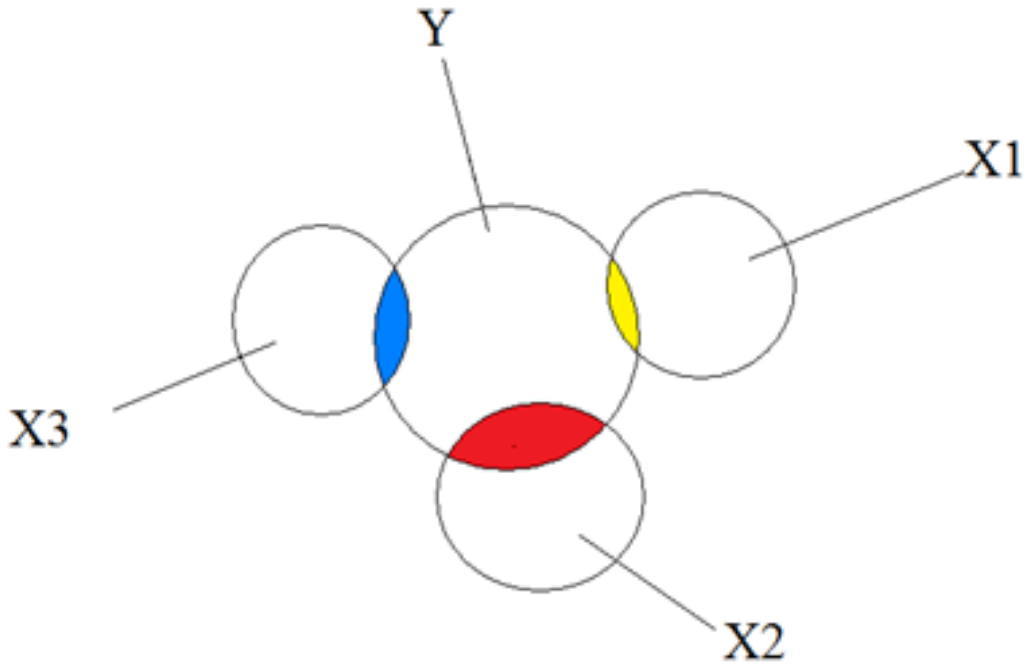
nuveen
A TIAA Company

Full - Score Table

Variable	IV	RF	Extratrees	Chi Square	RFE	L1	Final Score
Sales Transactions > 10K, 2018	1	1	1	1	1	1	6
Redemptions > 10K, 2018	1	1	1	1	1	0	5
Redemtions 2018	1	1	1	1	1	0	5
Sales 2018	1	1	1	1	1	0	5
2018, Funds Sold >10K	1	0	0	0	1	1	3
Asset Rank	0	0	0	1	1	1	3
Sales Rank	0	1	0	0	1	1	3
2018, Funds Redeemed >10K	0	0	1	0	1	0	2
2018, Funds Sold <10K	0	0	0	0	1	1	2
Open	0	0	0	0	1	1	2
Clickstream	0	0	0	0	1	1	2
Meetings	0	0	0	0	1	1	2
Redemptions <10K, 2018	0	0	0	0	1	1	2
Sales <10K, 2018	0	0	0	0	1	1	2
Clickthrough & ClickStream Rate	0	0	0	0	1	0	1
2018, Funds Redeemed <10K	0	0	0	0	1	0	1
Clickthrough	0	0	0	0	1	0	1
Completed Connects	0	0	0	0	1	0	1
Ratio of Completed Connects to Overall Connects	0	0	0	0	1	0	1
Ratio of Meetings to Overall Connects	0	0	0	0	1	0	1
Open Meetings	0	0	0	0	1	0	1
Overall Connects	0	0	0	0	1	0	1
Redemtions Rank	0	0	0	0	1	0	1

Multicollinearity

- Dropped All Variables that Had VIF Scores > 10
- A total of 13 Variables were Selected for Predictive Modeling



Features	VIF
Funds Sold >10K, 2018	9.34
Sales Transactions >10K, 2018	7.82
Redeptions >10K, 2018	4.98
Sales <10K, 2018	4.67
Redemptions <10K, 2018	4.31
Sales Rank	2.86
Funds Sold <10K, 2018	2.66
Assets Rank	2.59
Sales 2018	1.57
Clickstream	1.49
Redemptions 2018	1.49
Open	1.22
Meetings	1.13

Logistic Regression - Equation

Linear Model

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \ldots + \beta_n X_n$$

Logistic Regression Formula

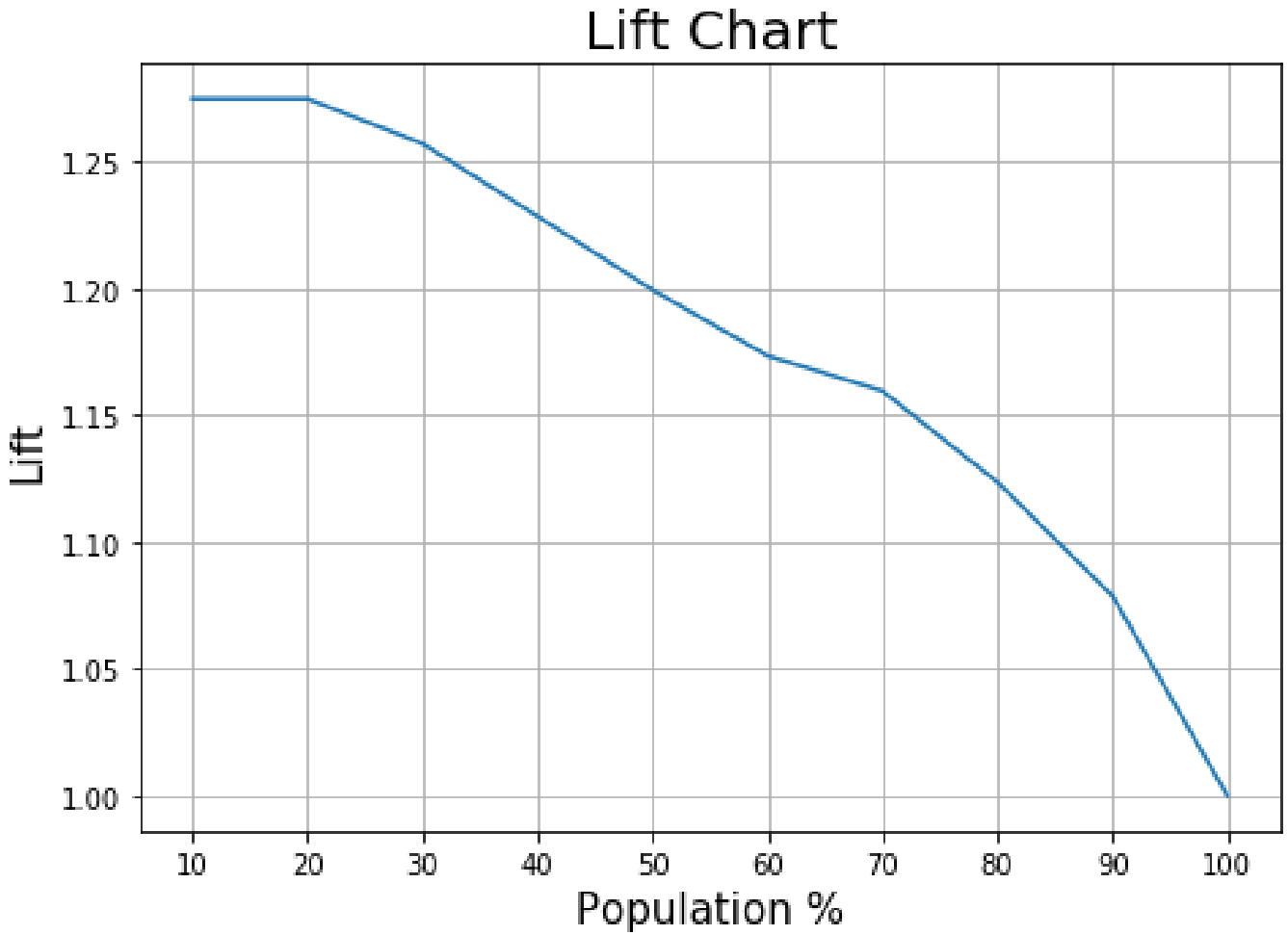
$$P = \frac{e^{Y_i}}{1 + e^{Y_i}}$$

β_0 = 0.0000004374 (Intercept)

Coefficients For Logistic Model

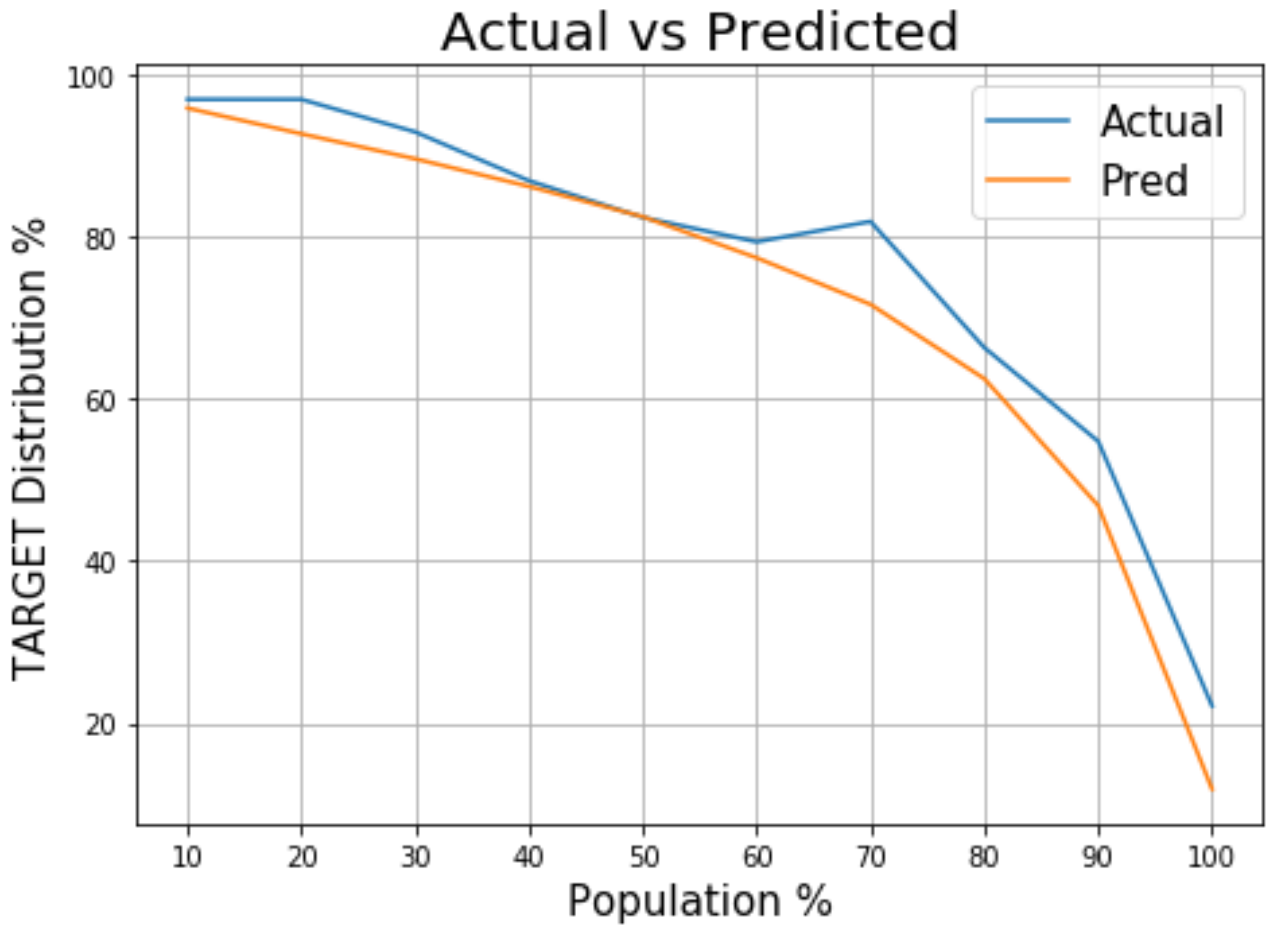
β1	0.000785	**redemtions_rank**
β2	0.0007219	**sales_rank**
β3	0.000008	**redemptions_less_10K_2018**
β4	0.0000076	sales_less_10K_2018
β5	0.0000032	sales_transactions_10K_2018
β6	0.000003	OPEN
β7	0.0000027	redemption_10K_2018
β8	0.0000022	clickstream
β9	0.0000022	meetings
β10	0.0000015	10K_funds_sold_2018
β11	0.0000007	Less_than_10K_funds_sold_2018
β12	-0.0000001	redemptions_2018
β13	-0.0000004	sales_2018

Lift Chart



A lift chart graphically represents the improvement that a mining model provides when compared against a random guess and measures the change in terms of a lift score.

Lift Charts



Model can closely, relatively, and accurately predict the outcome, based on the test data.

nuveen
A TIAA Company

Application of Model

Advisor Example:	Features
Redmption Rank	800
Sales Rank	1,607
Redemptions < 10K	8
Sales Transactions <10 K	28
Sales Transactions >10 K	20
Open	2
Redemption > 10 K	29
Clickstream	24
Meetings	0
Funds Sold > 10 K	6
Funds Sold < 10 K	0
Redemptions	-2,668,320
Sales	3,125,265

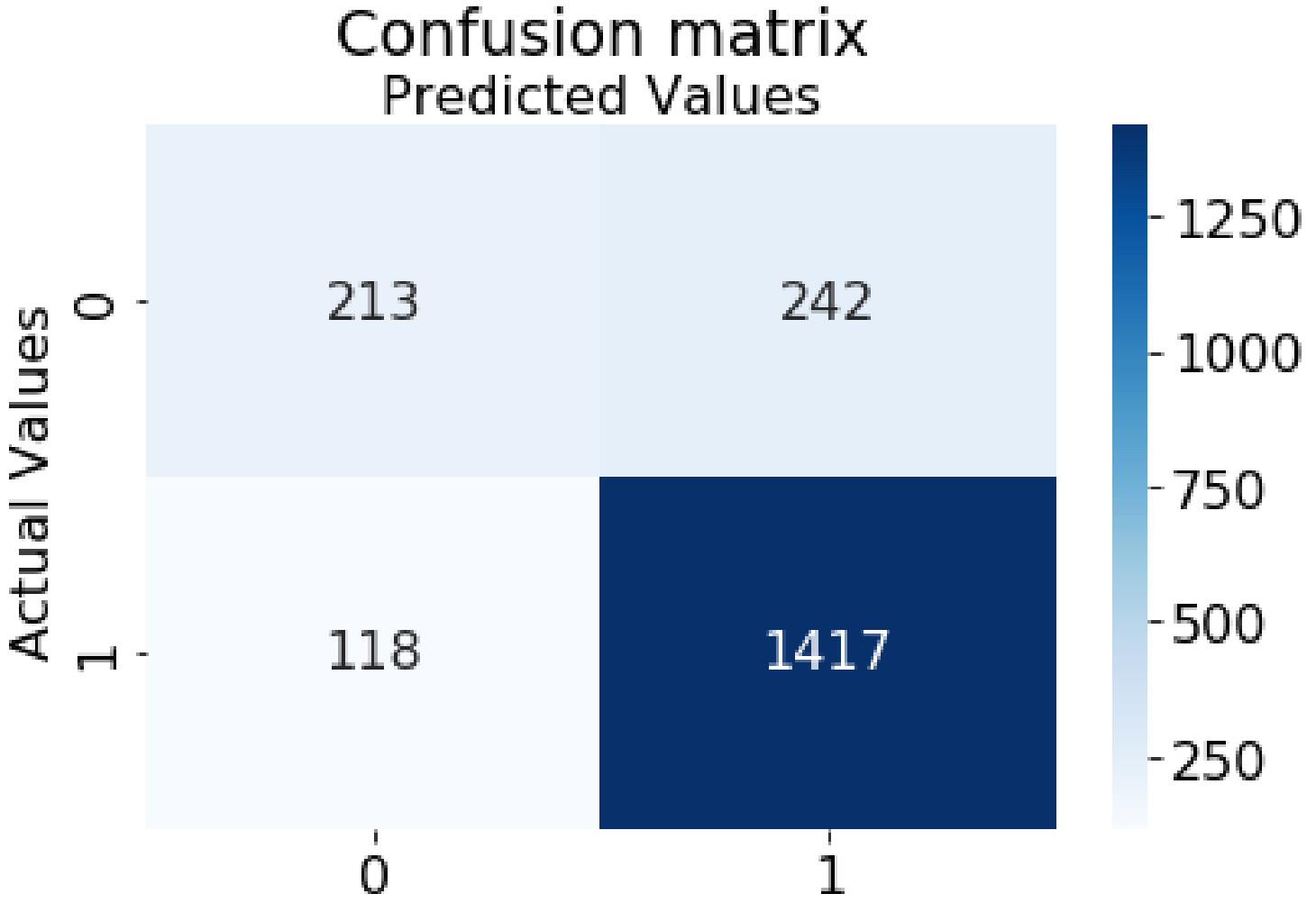
The lower the Sales & Redemption Rank - The greater the probability

Probability of Increasing Sales: **70%**

Low Sales = Higher Probability
Higher Redemption = Higher Probability

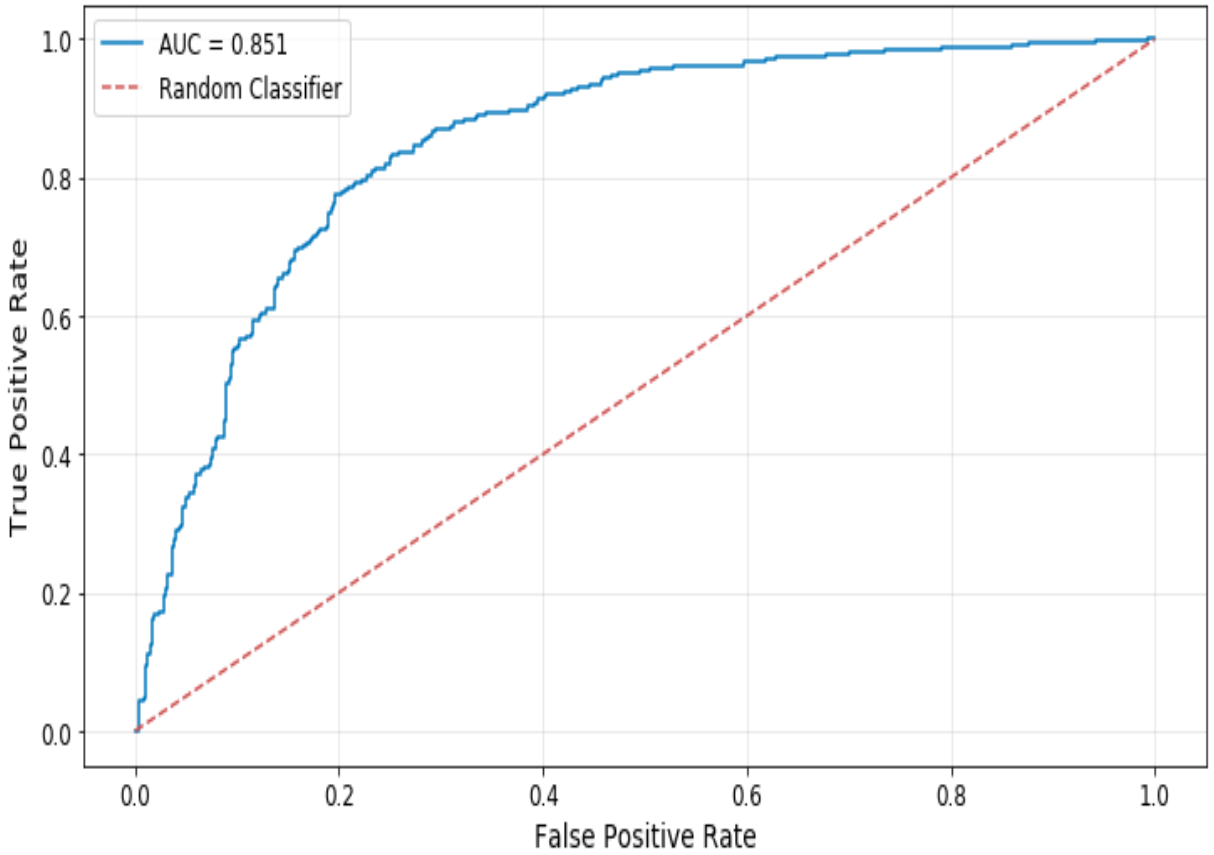
nuveen
A TIAA Company

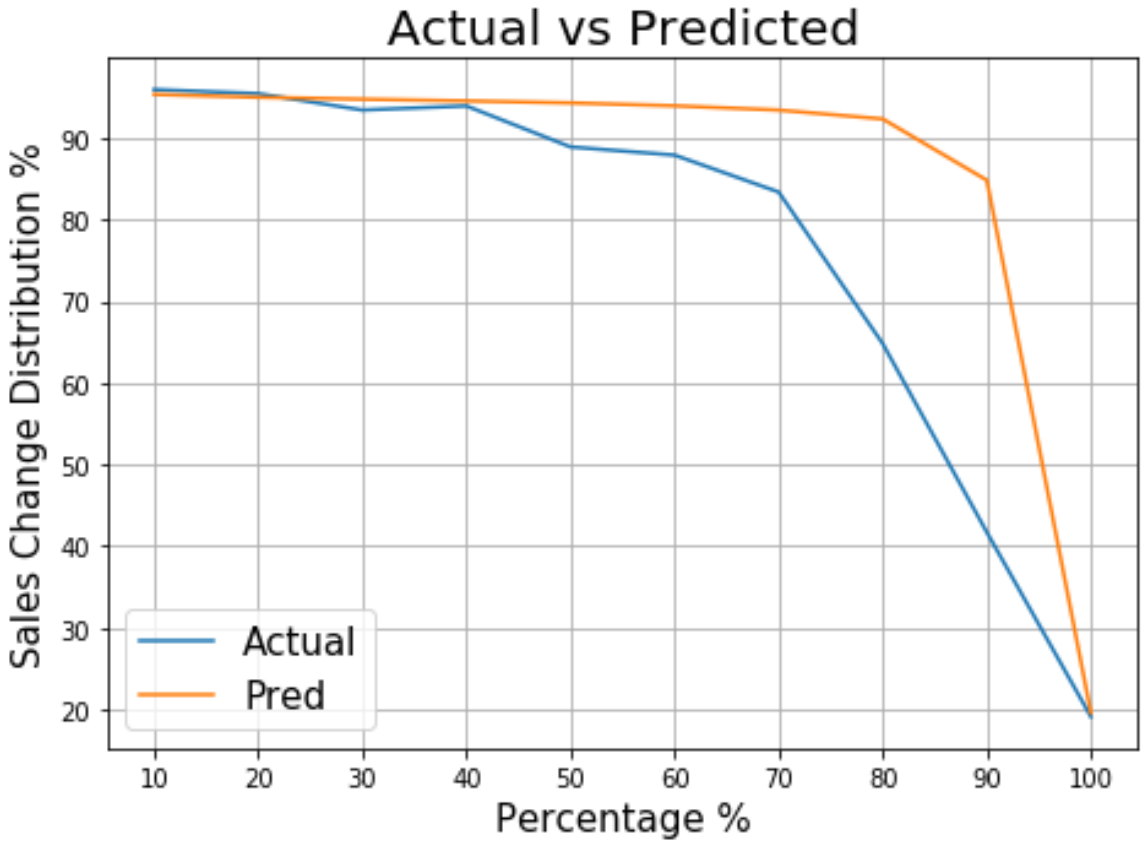
Logistic Regression – Confusion Matrix



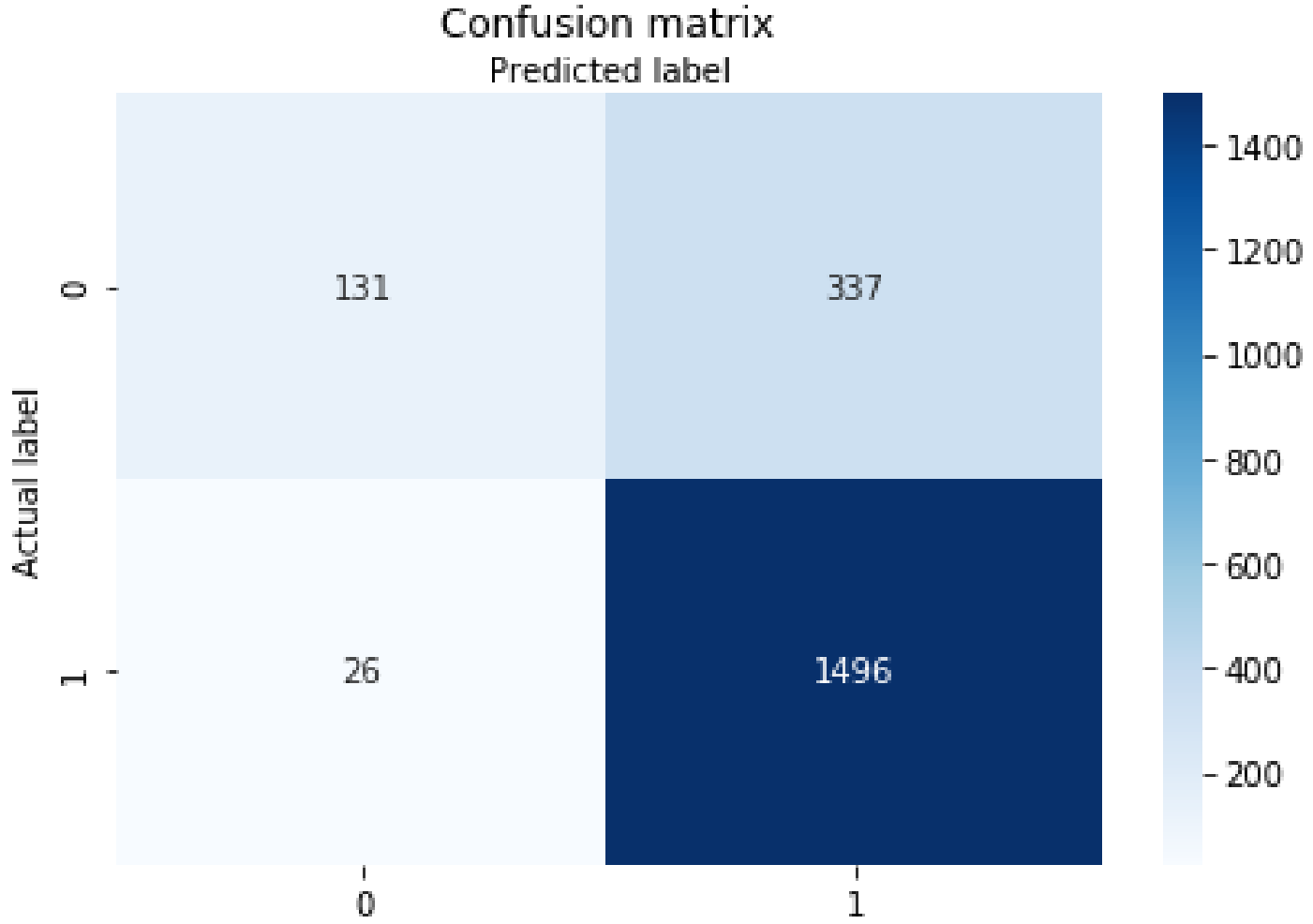
Naïve Bayes Model

Naïve Bayes Model





Naïve Bayes Confusion Matrix

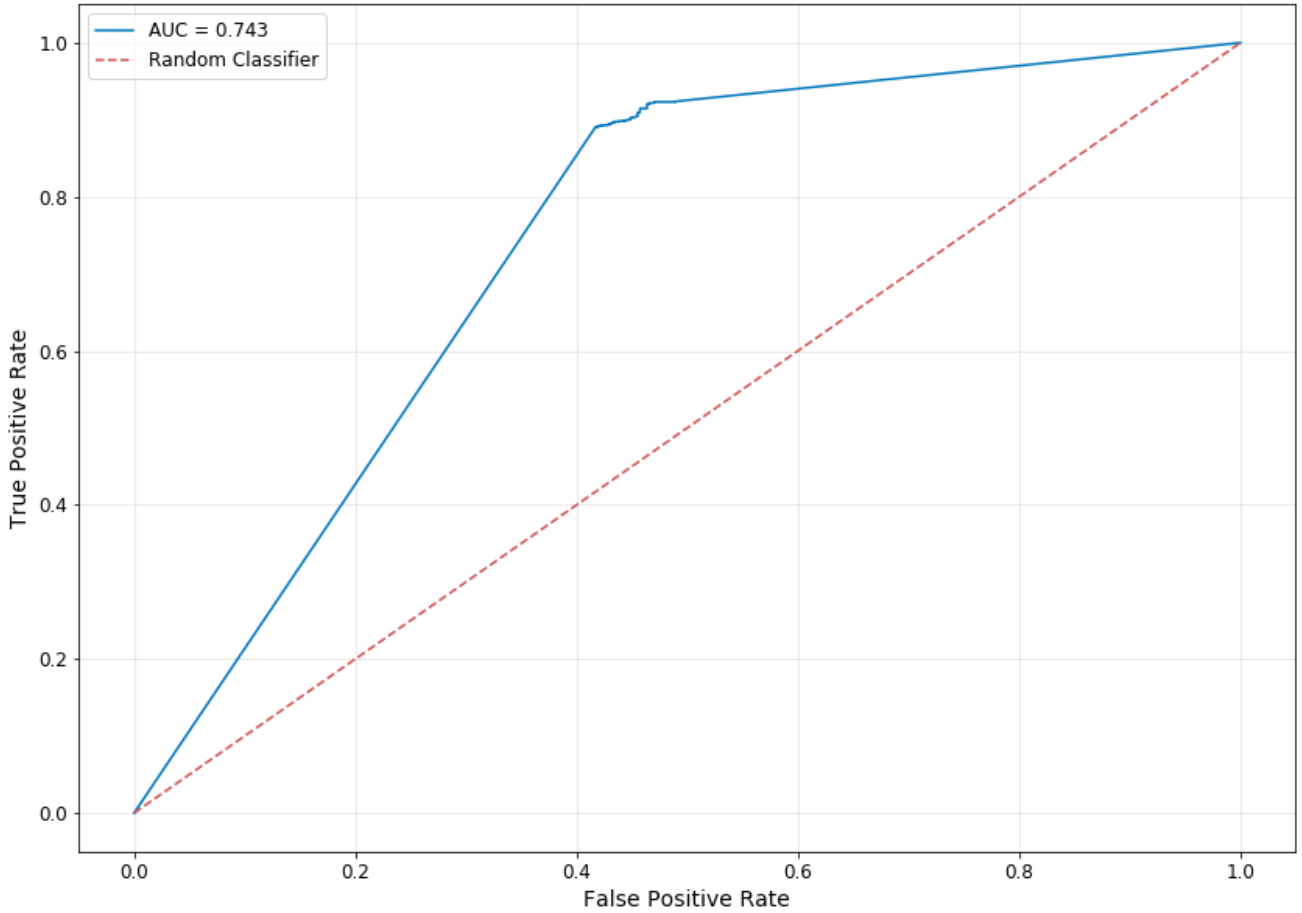


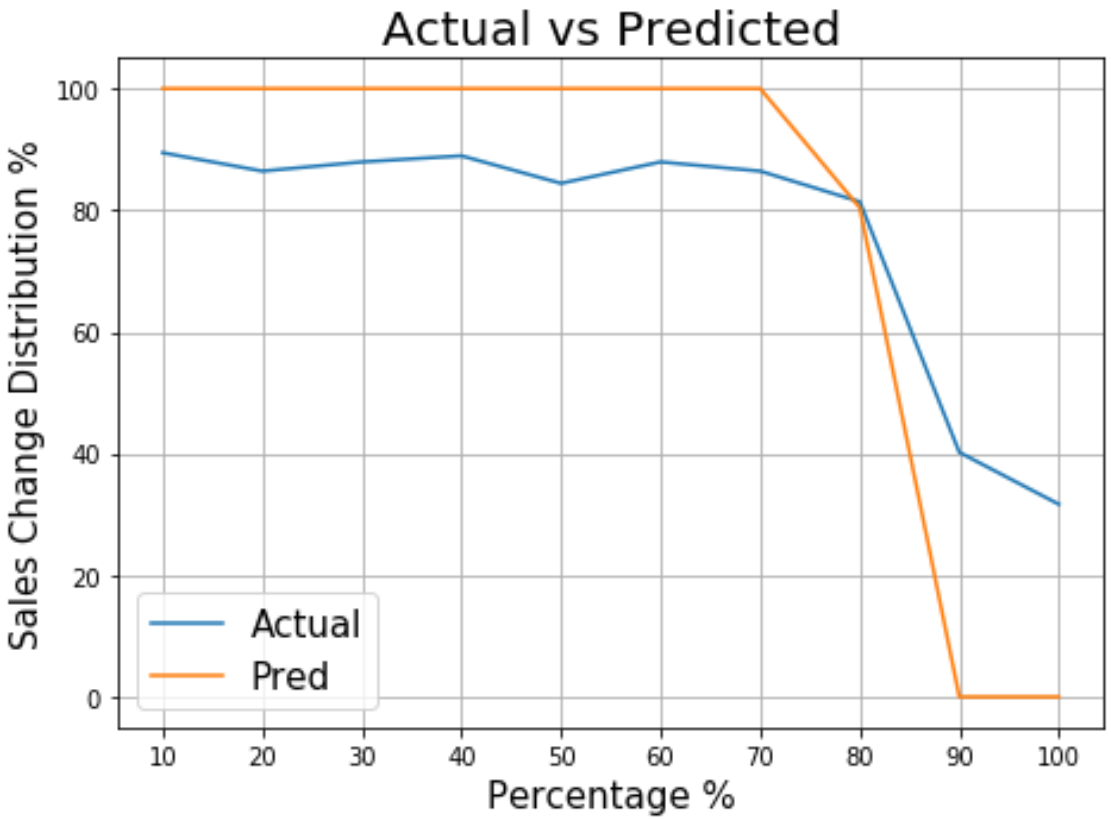
Naïve Bayes Code

```
from sklearn.naive_bayes import GaussianNB
clf = GaussianNB()

clf.fit(features_train,label_train)

pred_train = clf.predict(features_train)
pred_test = clf.predict(features_test)

from sklearn.metrics import accuracy_score
accuracy_train = accuracy_score(pred_train,label_train)
accuracy_test = accuracy_score(pred_test,label_test)

from sklearn import metrics
fpr, tpr, _ = metrics.roc_curve(np.array(label_train), clf.predict_proba(features_train)[:,1])
auc_train = metrics.auc(fpr,tpr)

fpr, tpr, _ = metrics.roc_curve(np.array(label_test), clf.predict_proba(features_test)[:,1])
auc_test = metrics.auc(fpr,tpr)

print(accuracy_train,accuracy_test,auc_train,auc_test)
```

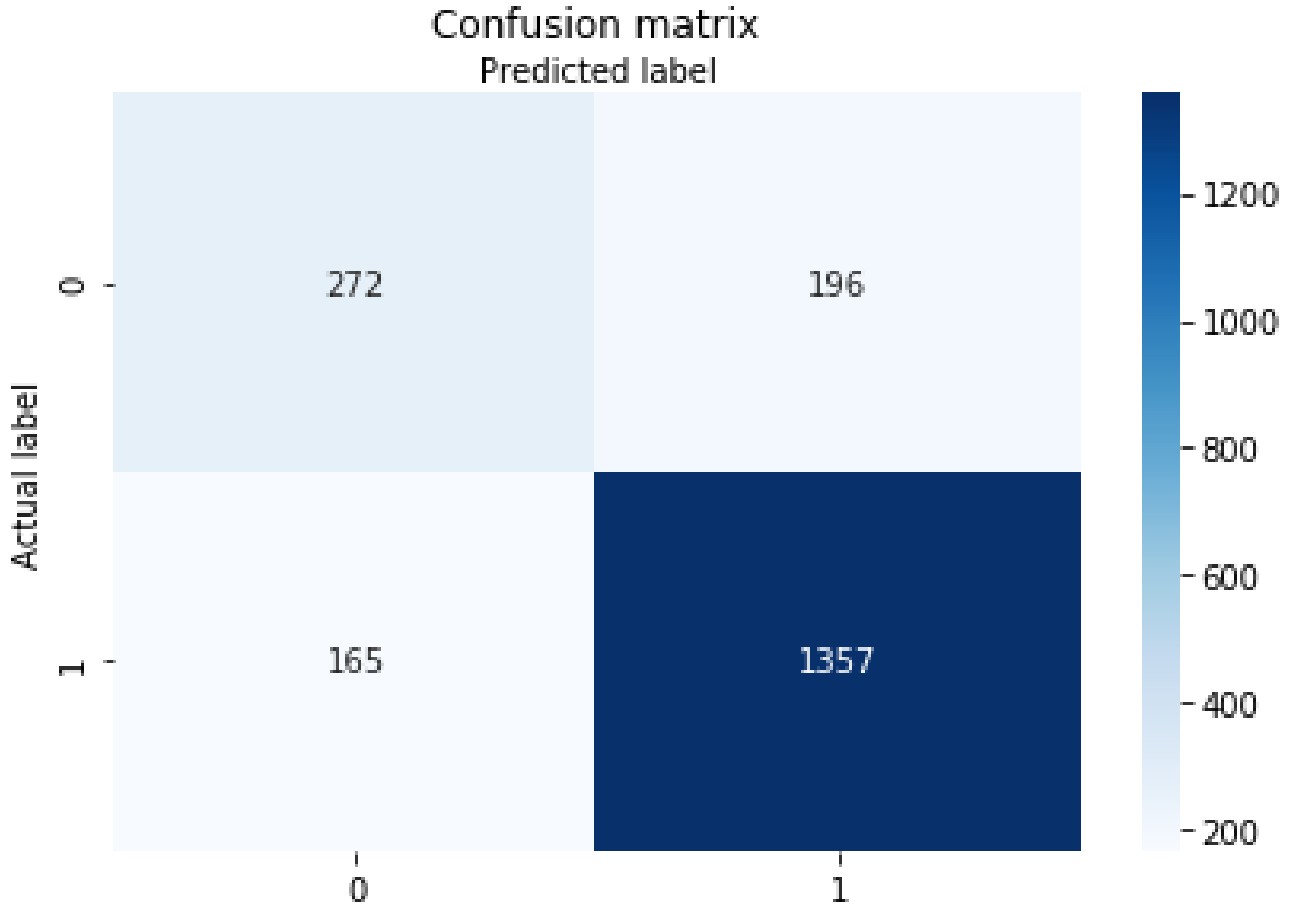
Neural Network Model

Neural Network Model





Neural Network Model



Neural Network - Code

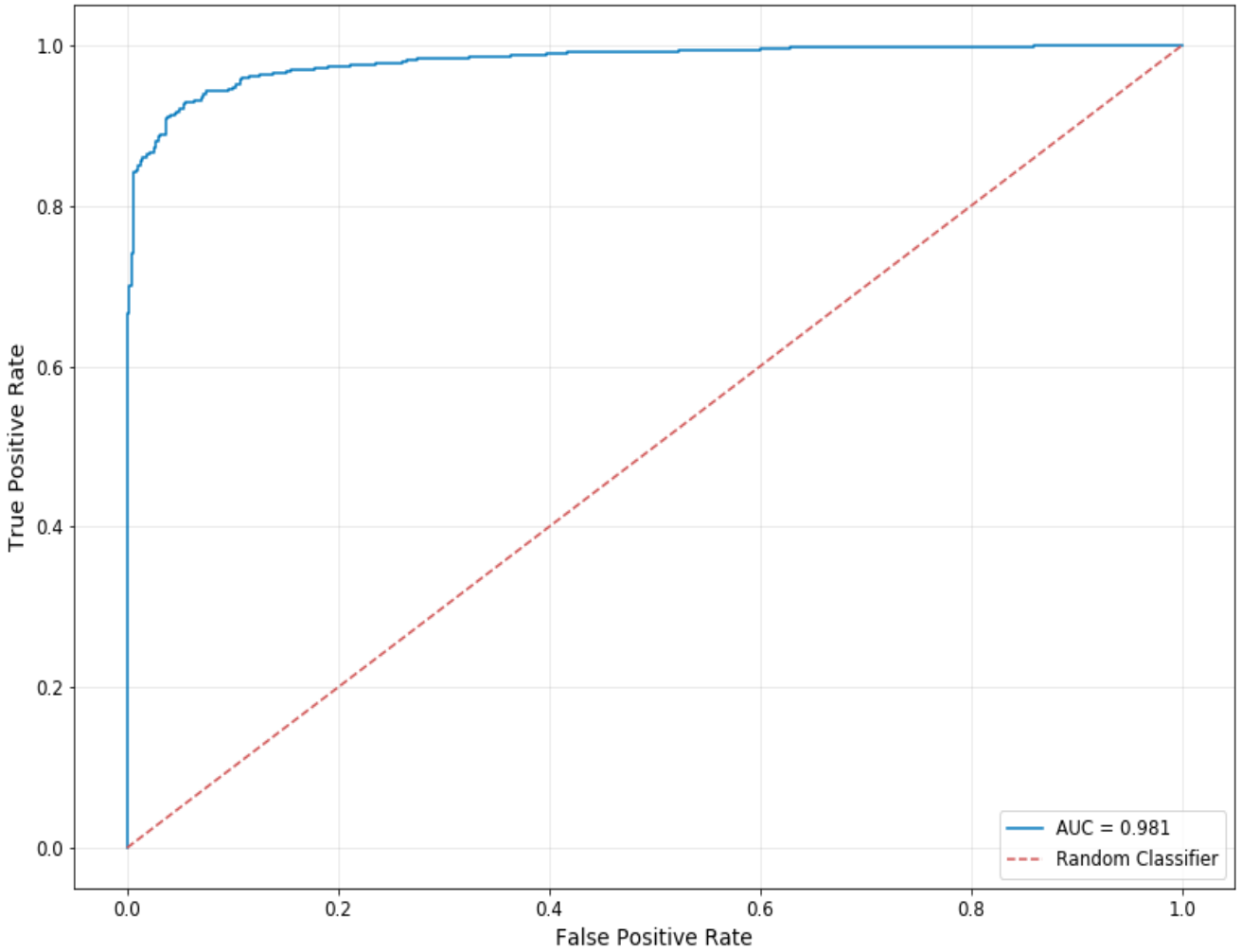
```
1  from sklearn.neural_network import MLPClassifier
2  clf = MLPClassifier()
3
4  clf.fit(features_train,label_train)
5
6  pred_train = clf.predict(features_train)
7  pred_test = clf.predict(features_test)
8
9  from sklearn.metrics import accuracy_score
10 accuracy_train = accuracy_score(pred_train,label_train)
11 accuracy_test = accuracy_score(pred_test,label_test)
12
13 from sklearn import metrics
14 fpr, tpr, _ = metrics.roc_curve(np.array(label_train), clf.predict_proba(features_train)[:,1])
15 auc_train = metrics.auc(fpr,tpr)
16
17 fpr, tpr, _ = metrics.roc_curve(np.array(label_test), clf.predict_proba(features_test)[:,1])
18 auc_test = metrics.auc(fpr,tpr)
19
20 print(accuracy_train,accuracy_test,auc_train,auc_test)
```

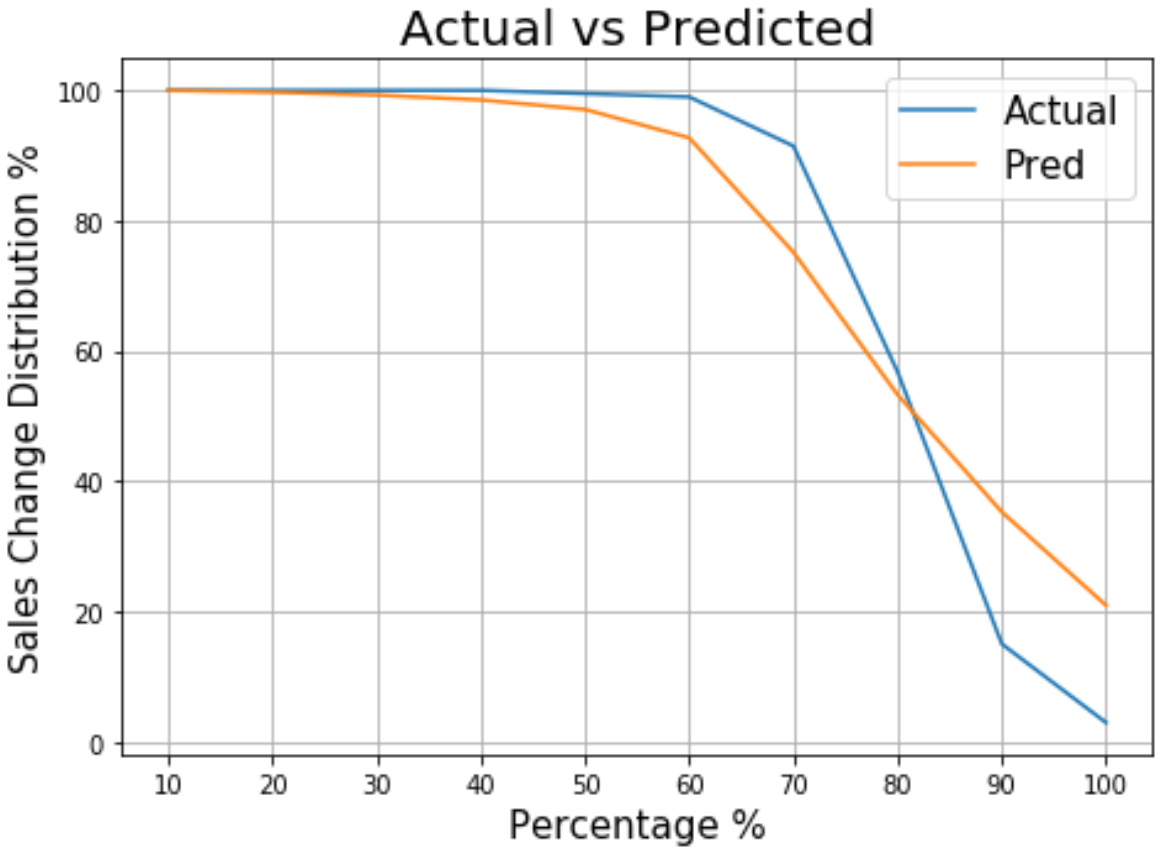
```
0.8095477386934673 0.8185929648241206 0.7436197728761254 0.7425213675213674
```

Random Forrest – Hyper Tuning Model

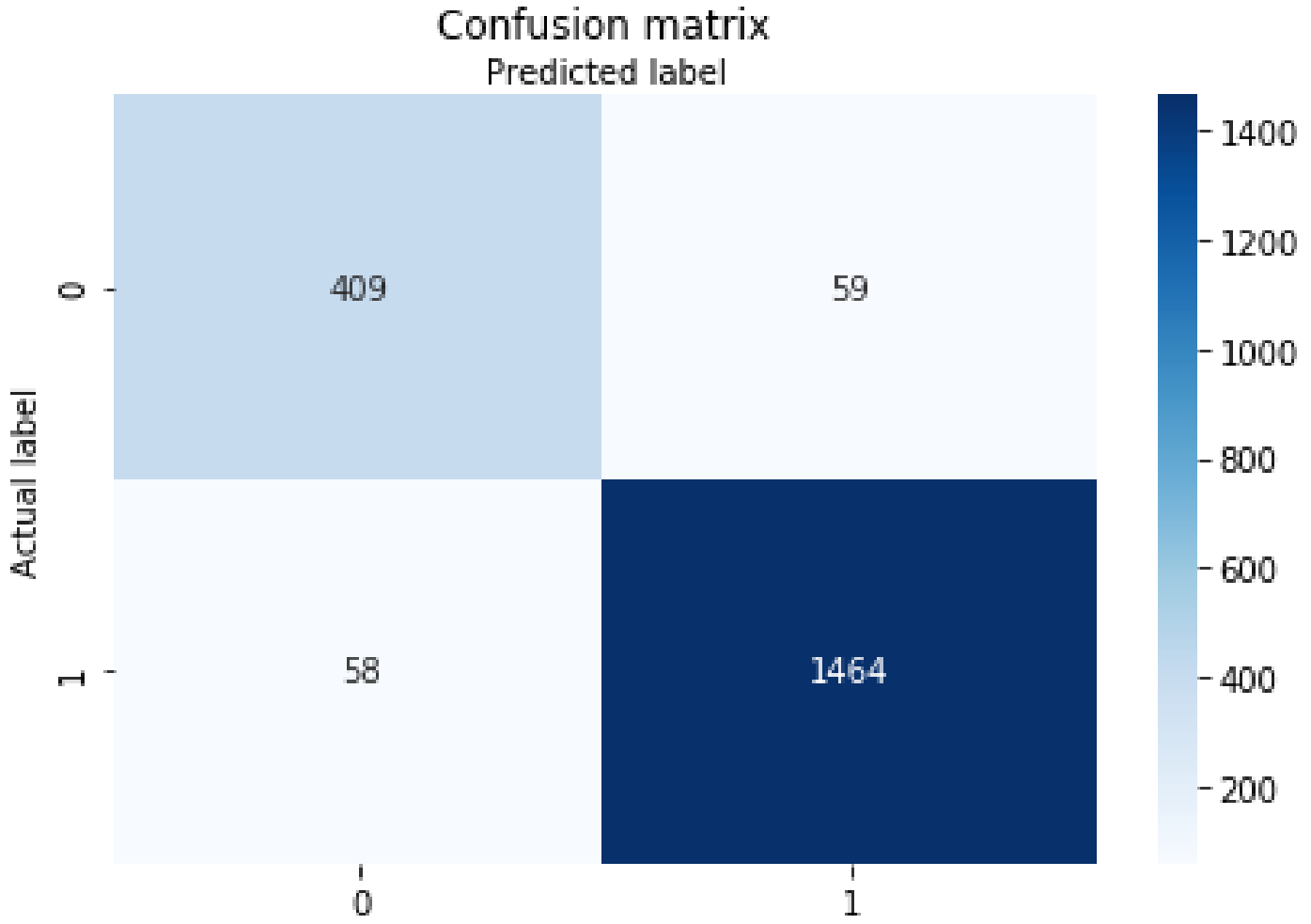
nuveen
A TIAA Company

Hyper Tuning – Random Forrest Model





Confusion Matrix – Hyper Tuning RF



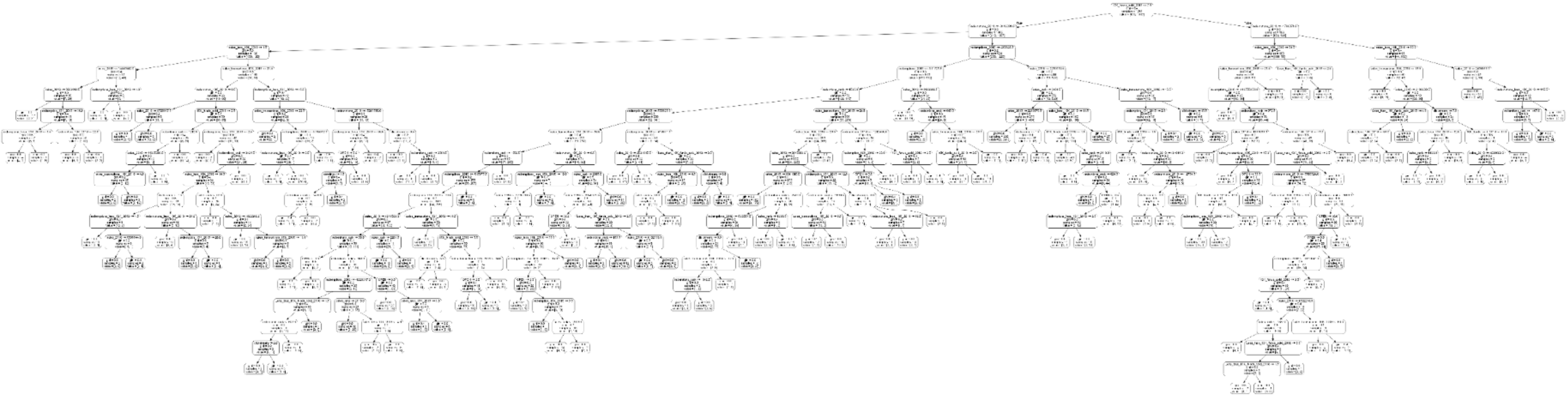
Hyper Tuning Random Forrest - Code

In [96]:

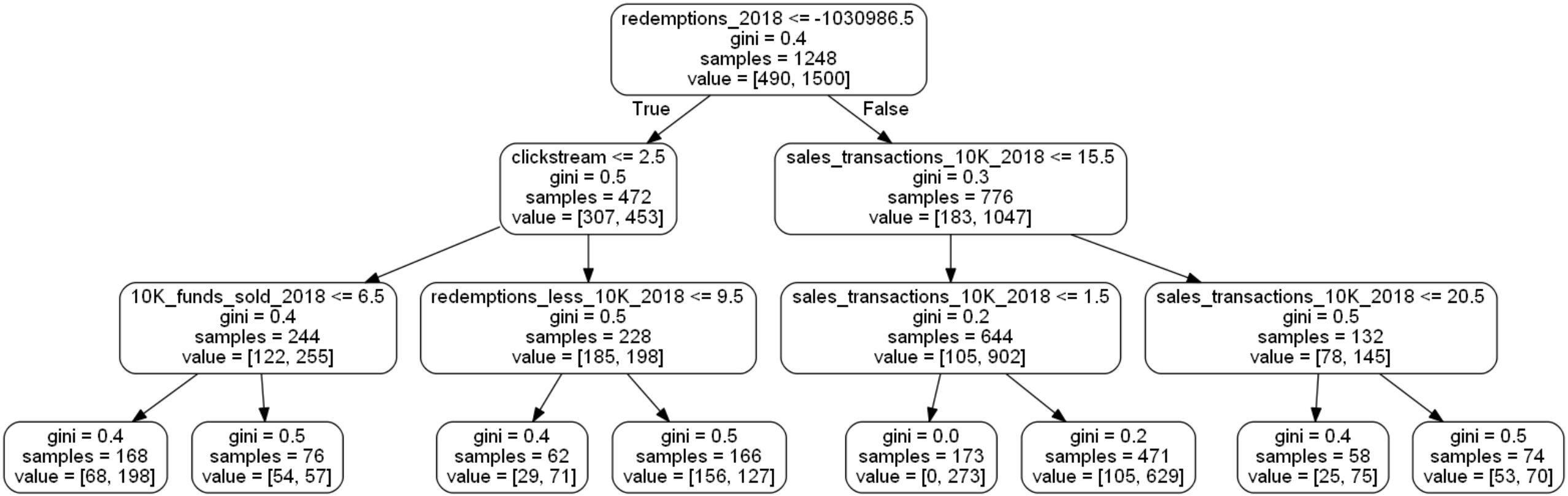
```
from sklearn.model_selection import RandomizedSearchCV
from sklearn.ensemble import RandomForestClassifier

n_estimators = [int(x) for x in np.linspace(start = 10, stop = 500, num = 10)]
max_features = ['auto', 'sqrt']
max_depth = [int(x) for x in np.linspace(3, 10, num = 1)]
max_depth.append(None)
min_samples_split = [2, 5, 10]
min_samples_leaf = [1, 2, 4]
bootstrap = [True, False]

random_grid = {'n_estimators': n_estimators,
               'max_features': max_features,
               'max_depth': max_depth,
               'min_samples_split': min_samples_split,
               'min_samples_leaf': min_samples_leaf,
               'bootstrap': bootstrap}

rf = RandomForestClassifier()

rf_random = RandomizedSearchCV(estimator = rf, param_distributions = random_grid, n_iter = 10, cv = 2, verbose=2, random
rf_random.fit(features_train, label_train)

print(rf_random.best_params_)

#-------------------------------------------------------------------------------------------------------------------
from sklearn.ensemble import RandomForestClassifier
clf = RandomForestClassifier(**rf_random.best_params_)

clf.fit(features_train,label_train)

pred_train = clf.predict(features_train)
pred_test = clf.predict(features_test)

from sklearn.metrics import accuracy_score
accuracy_train = accuracy_score(pred_train,label_train)
accuracy_test = accuracy_score(pred_test,label_test)

from sklearn import metrics
fpr, tpr, _ = metrics.roc_curve(np.array(label_train), clf.predict_proba(features_train)[:,1])
auc_train = metrics.auc(fpr,tpr)

fpr, tpr, _ = metrics.roc_curve(np.array(label_test), clf.predict_proba(features_test)[:,1])
auc_test = metrics.auc(fpr,tpr)

print(accuracy_train,accuracy_test,auc_train,auc_test)
```

nuveen
A TIAA Company

Full Tree Random Forrest – Hyper Tuning

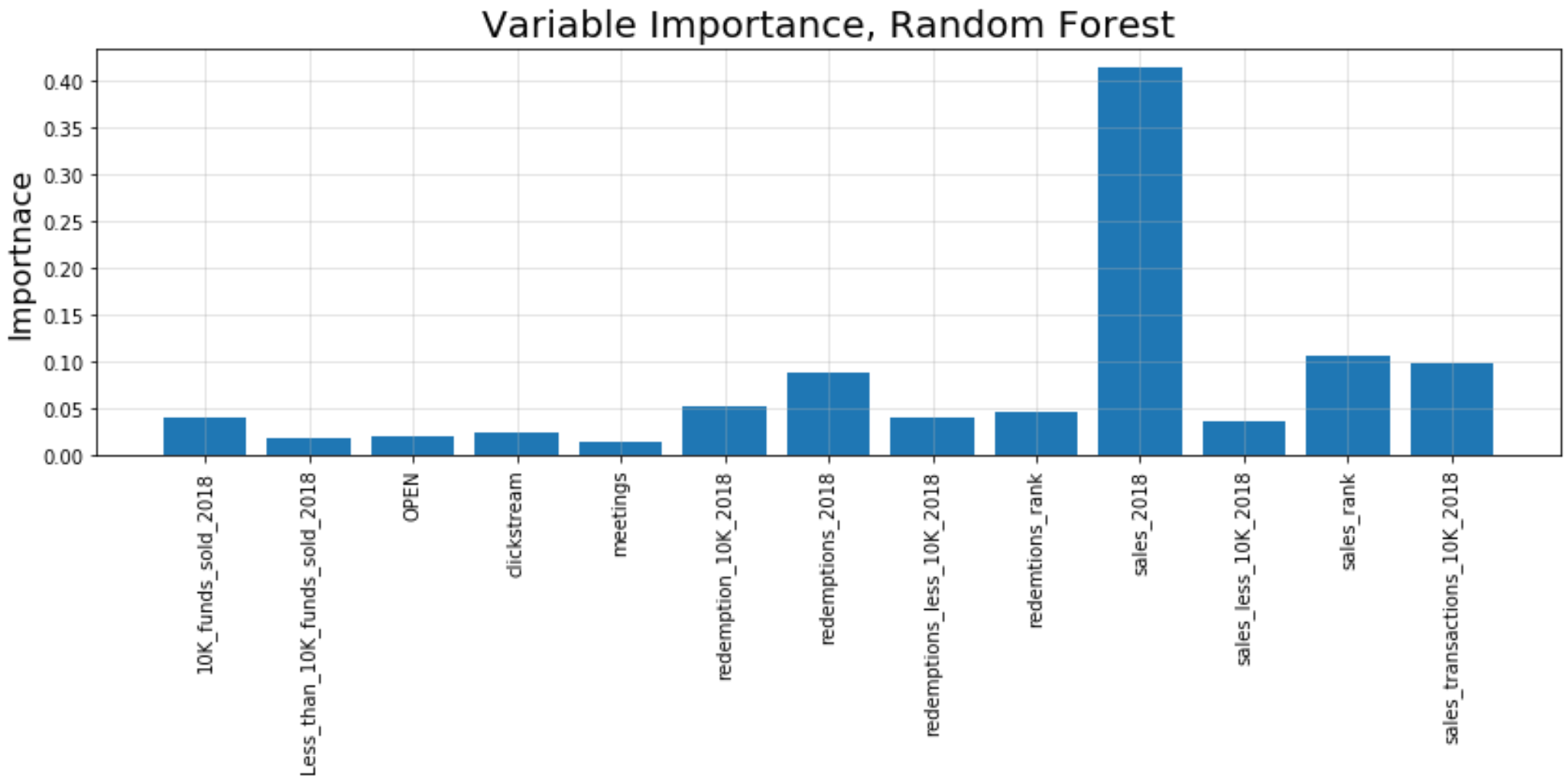


Small Tree


redemptions_2018 <= -1030986.5
gini = 0.4
samples = 1248
value = [490, 1500]
True
False
clickstream <= 2.5
gini = 0.5
samples = 472
value = [307, 453]
sales_transactions_10K_2018 <= 15.5
gini = 0.3
samples = 776
value = [183, 1047]
10K_funds_sold_2018 <= 6.5
gini = 0.4
samples = 244
value = [122, 255]
redemptions_less_10K_2018 <= 9.5
gini = 0.5
samples = 228
value = [185, 198]
sales_transactions_10K_2018 <= 1.5
gini = 0.2
samples = 644
value = [105, 902]
sales_transactions_10K_2018 <= 20.5
gini = 0.5
samples = 132
value = [78, 145]
gini = 0.4
samples = 168
value = [68, 198]
gini = 0.5
samples = 76
value = [54, 57]
gini = 0.4
samples = 62
value = [29, 71]
gini = 0.5
samples = 166
value = [156, 127]
gini = 0.0
samples = 173
value = [0, 273]
gini = 0.2
samples = 471
value = [105, 629]
gini = 0.4
samples = 58
value = [25, 75]
gini = 0.5
samples = 74
value = [53, 70]

nuveen
A TIAA Company

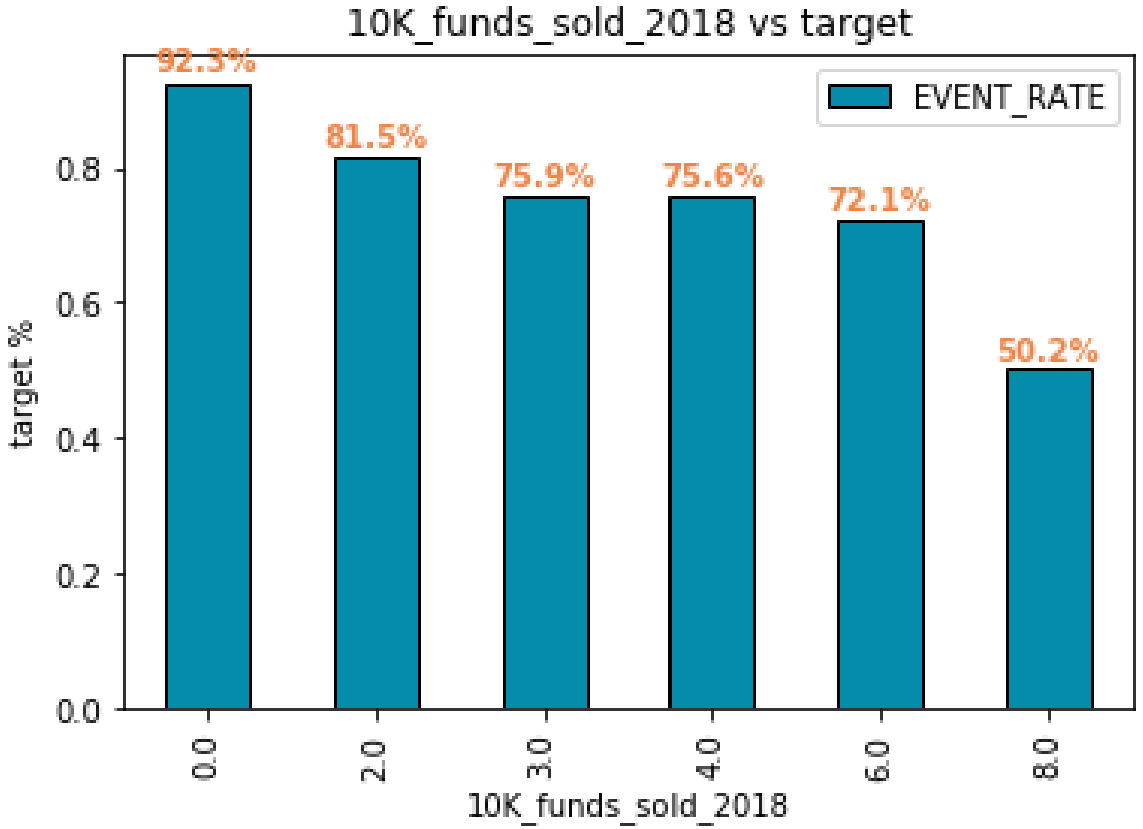
Random Forest Variable Importance

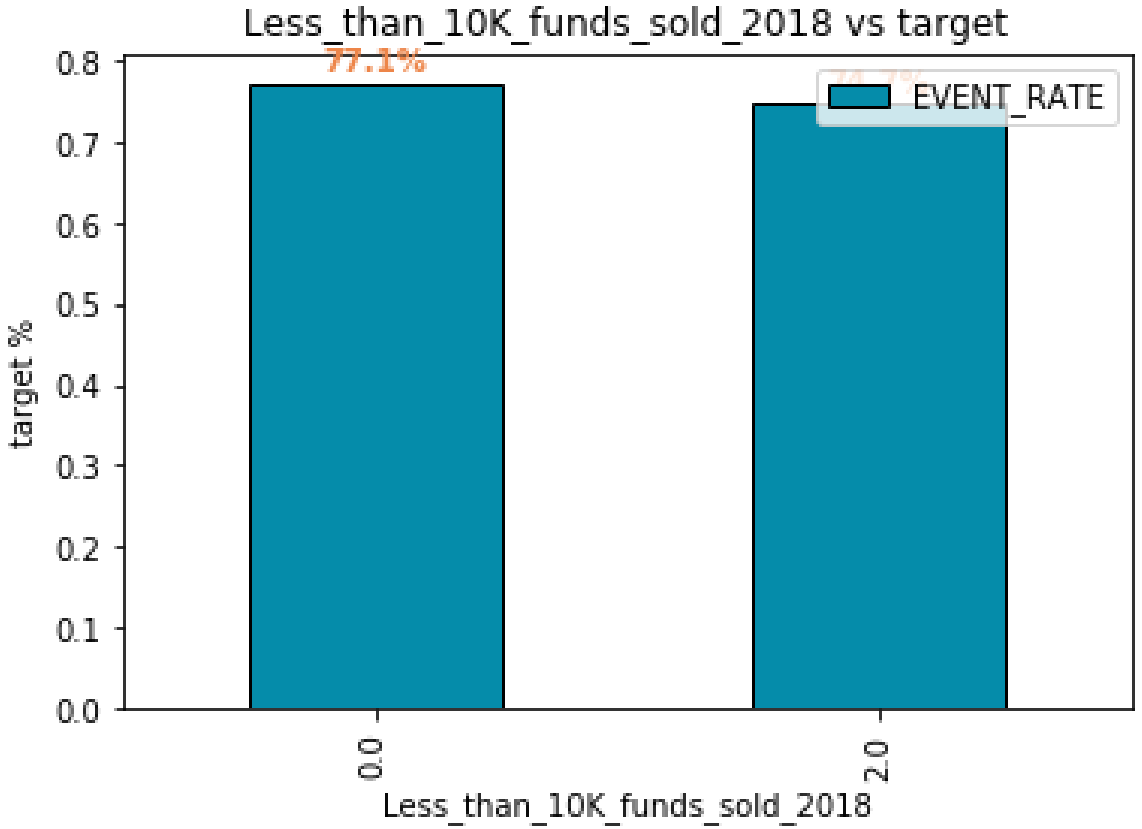


Bivariate Bins

nuveen
A TIAA Company

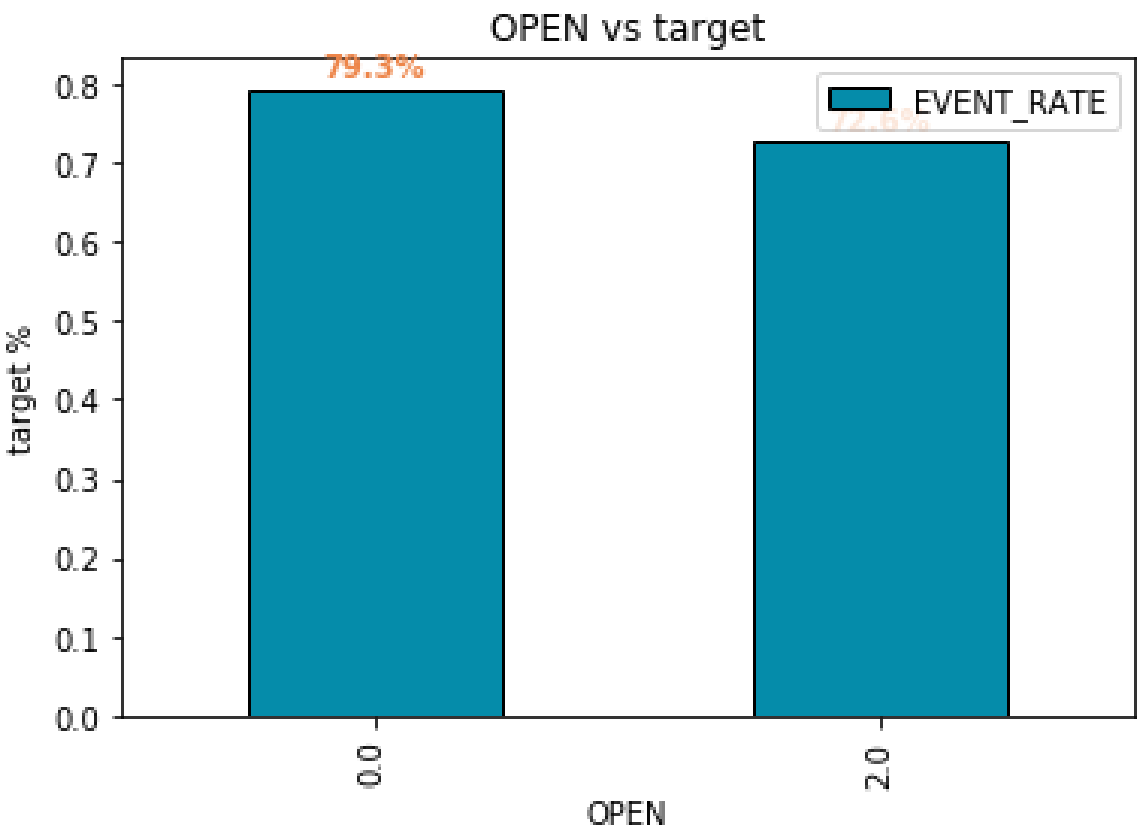
Bivariate Histograms

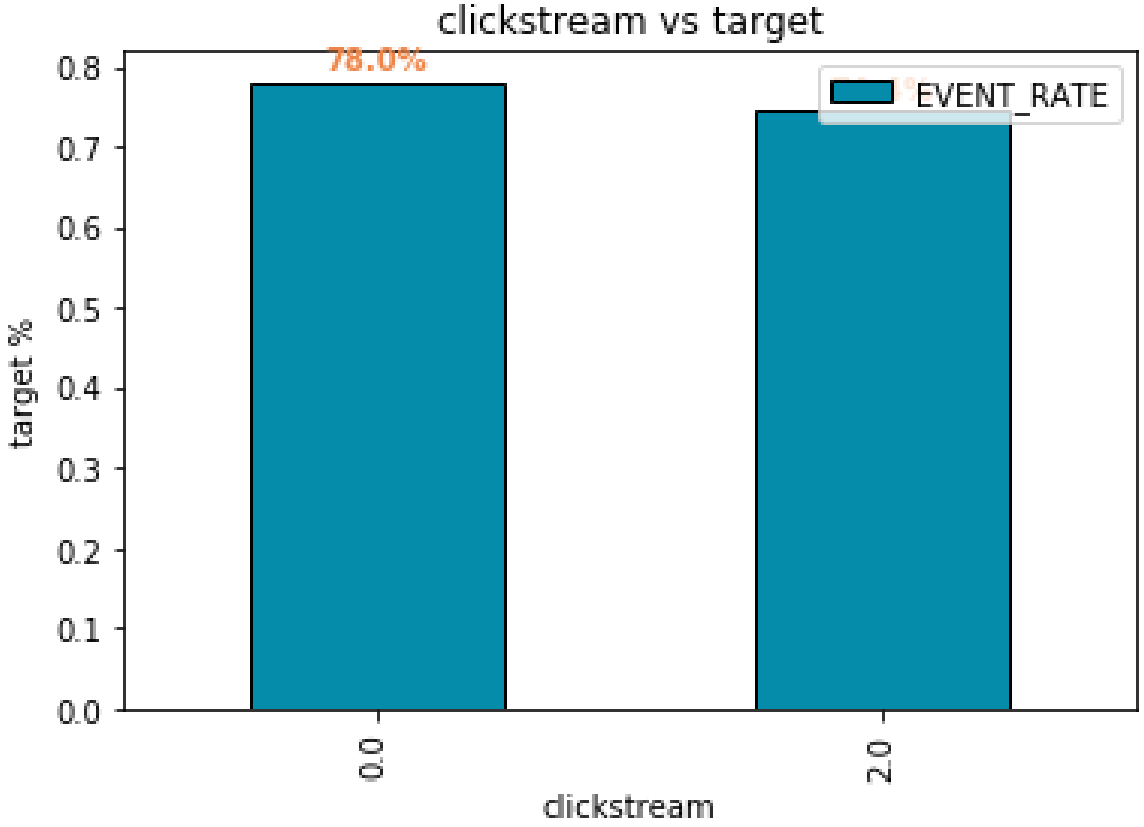




nuveen
A TIAA Company

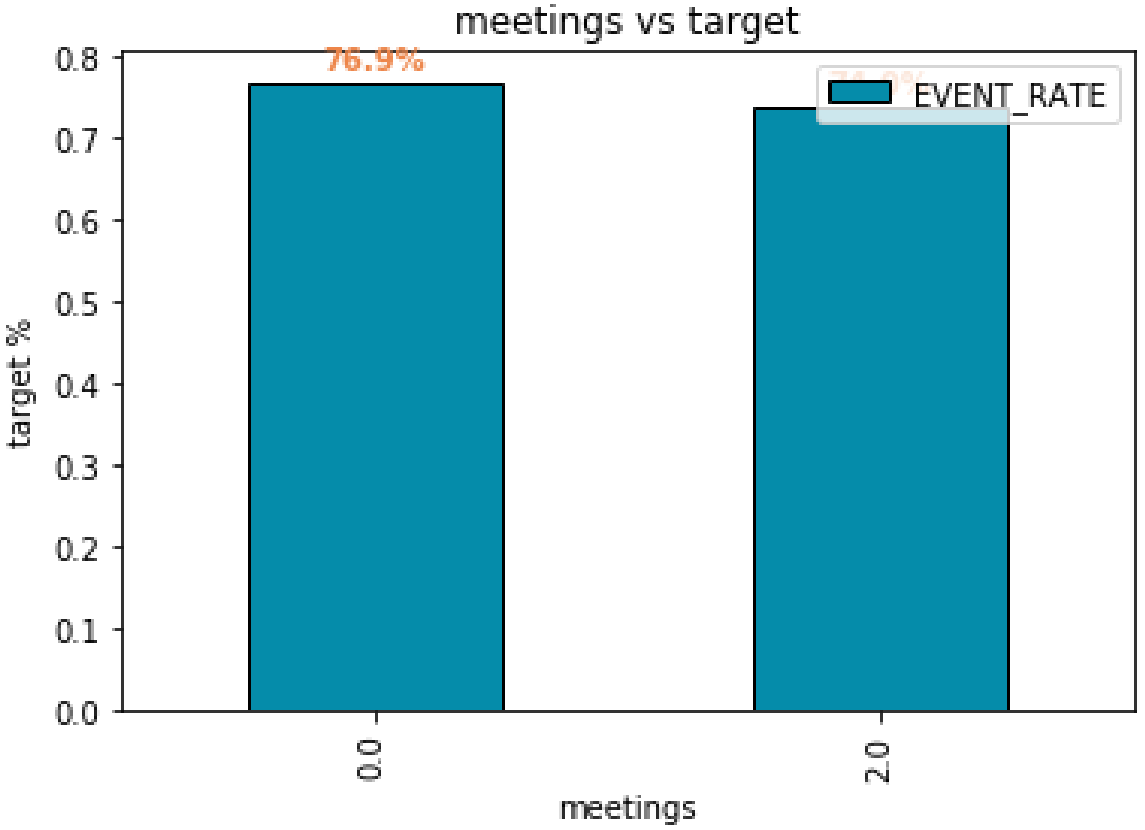
Bivariate Histograms

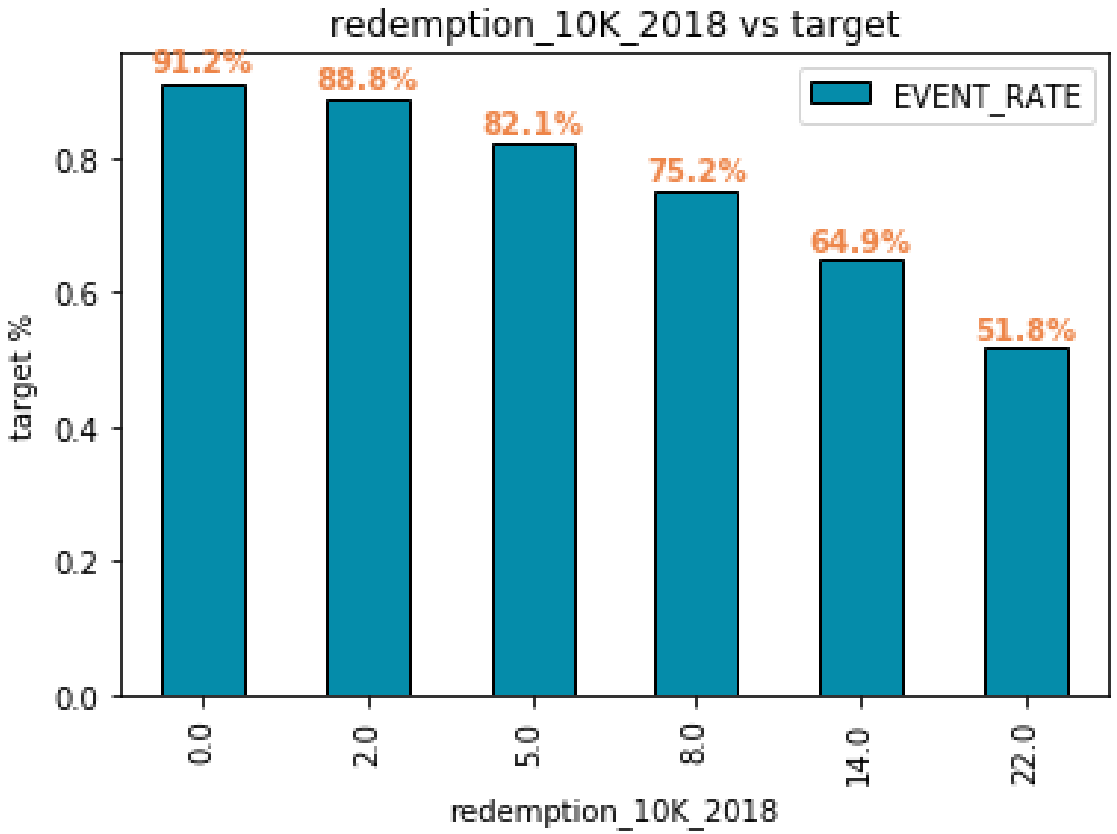




nuveen
A TIAA Company

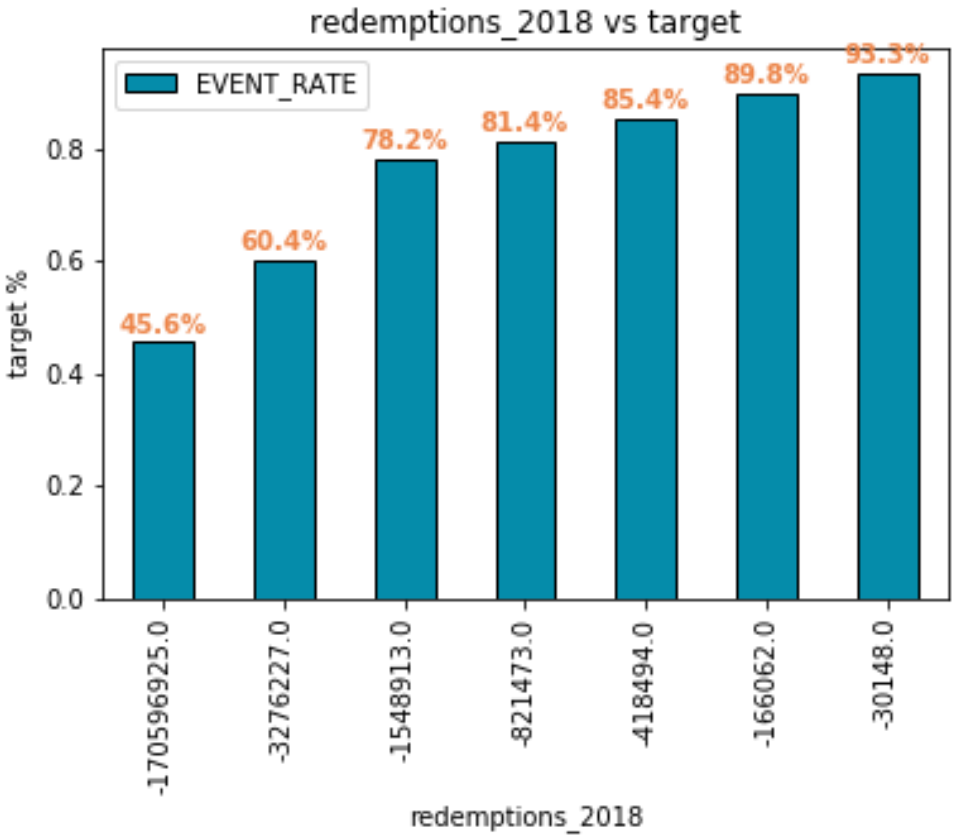
Bivariate Histograms

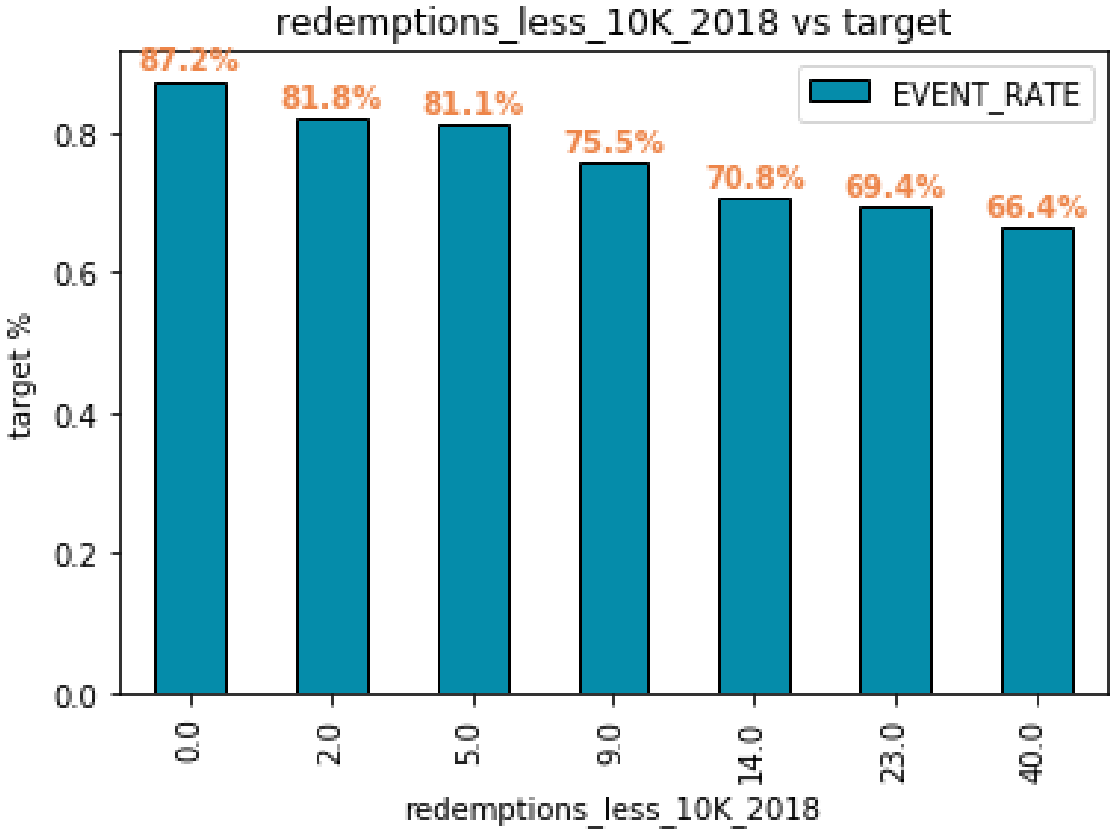




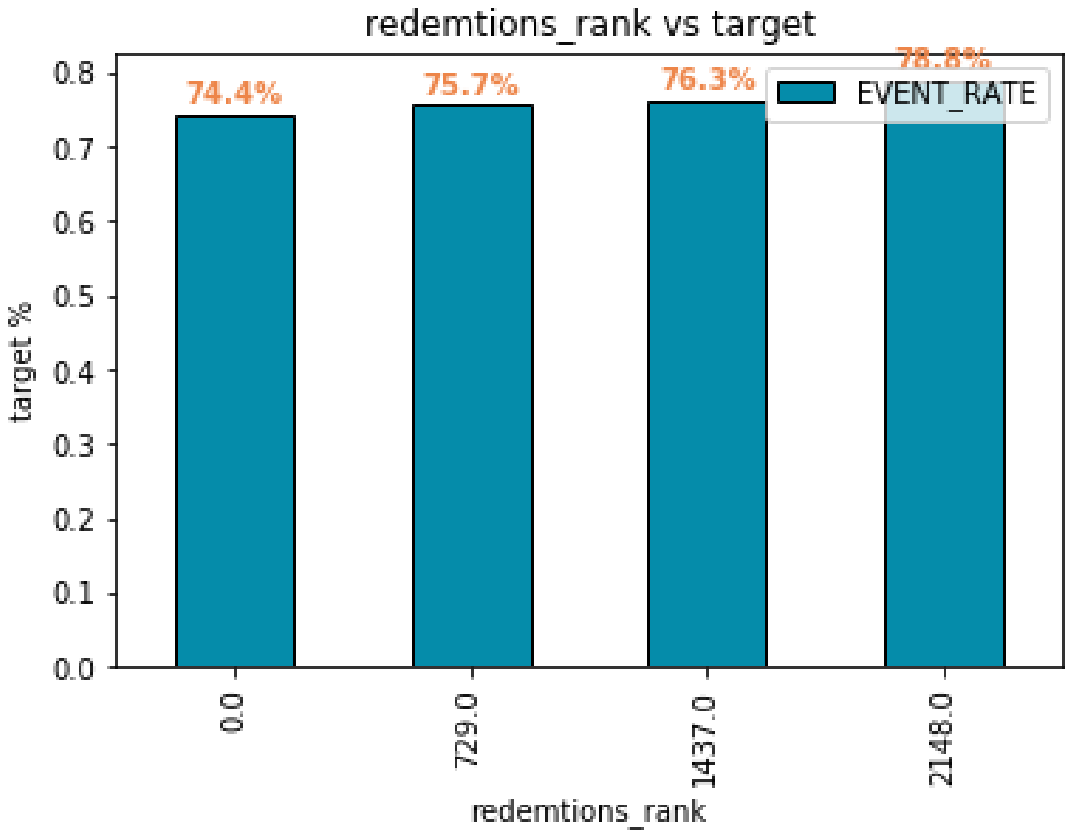
nuveen
A TIAA Company

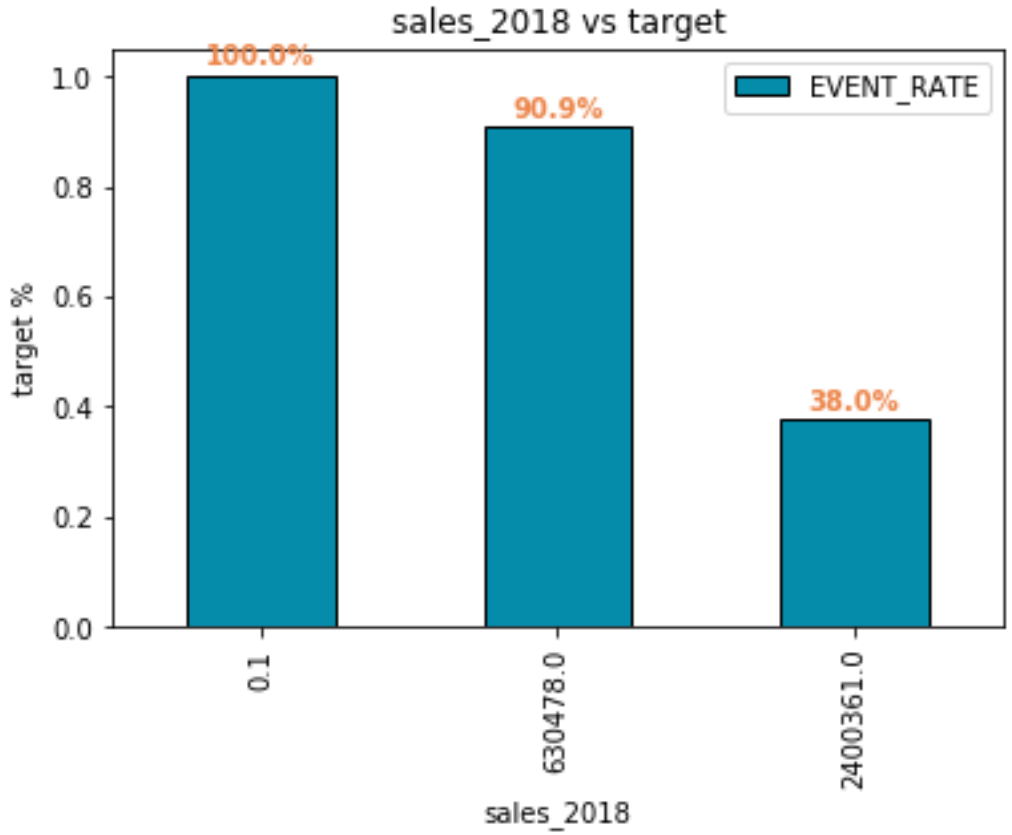
Bivariate Histograms





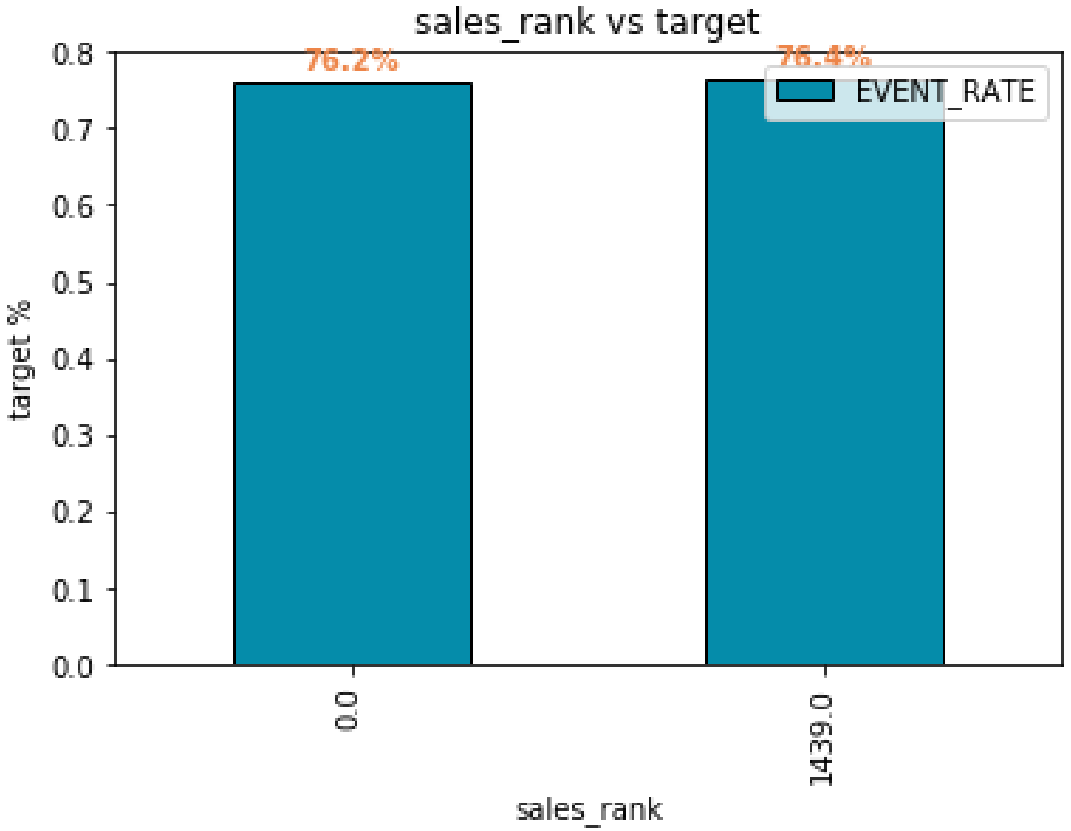
Bivariate Histograms

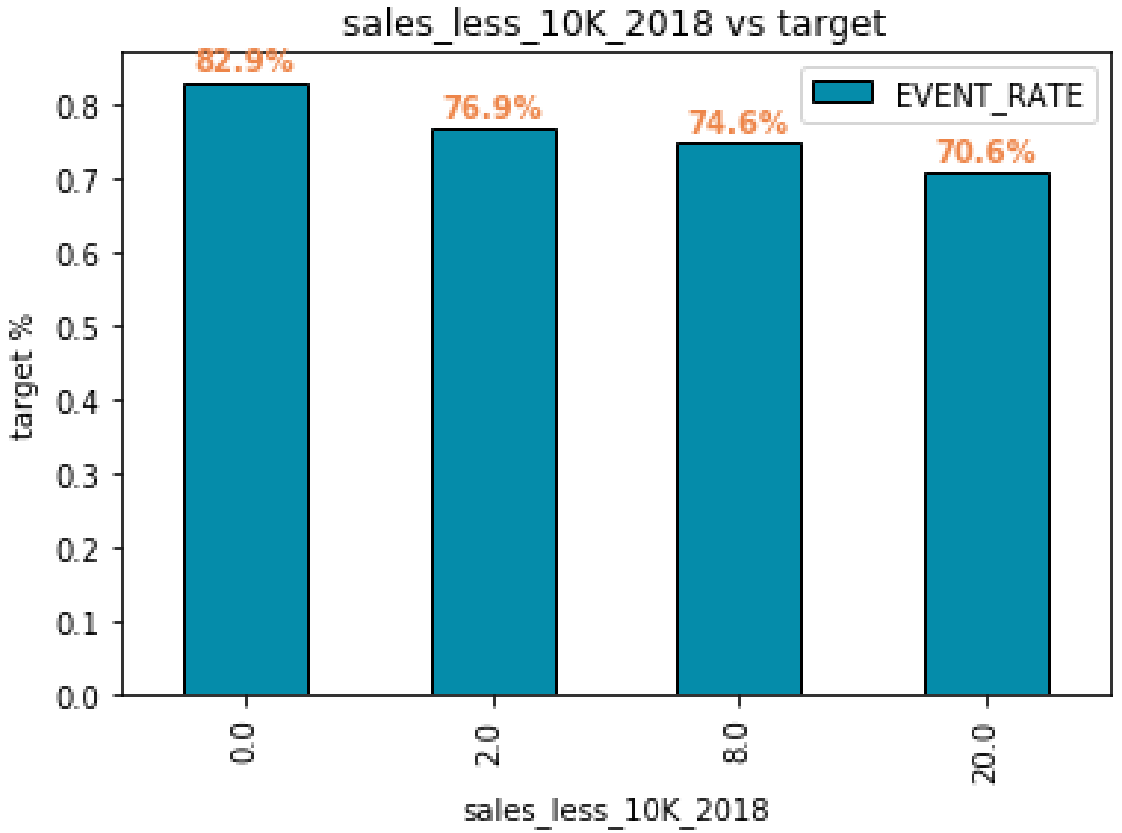




nuveen
A TIAA Company

Bivariate Histograms





Bivariate Histograms