EMERITUS

nuveen
A TIAA Company

Nuveen Sales Data Analysis Project

PGDMLAI CAPSTONE

JULIAN RENZ

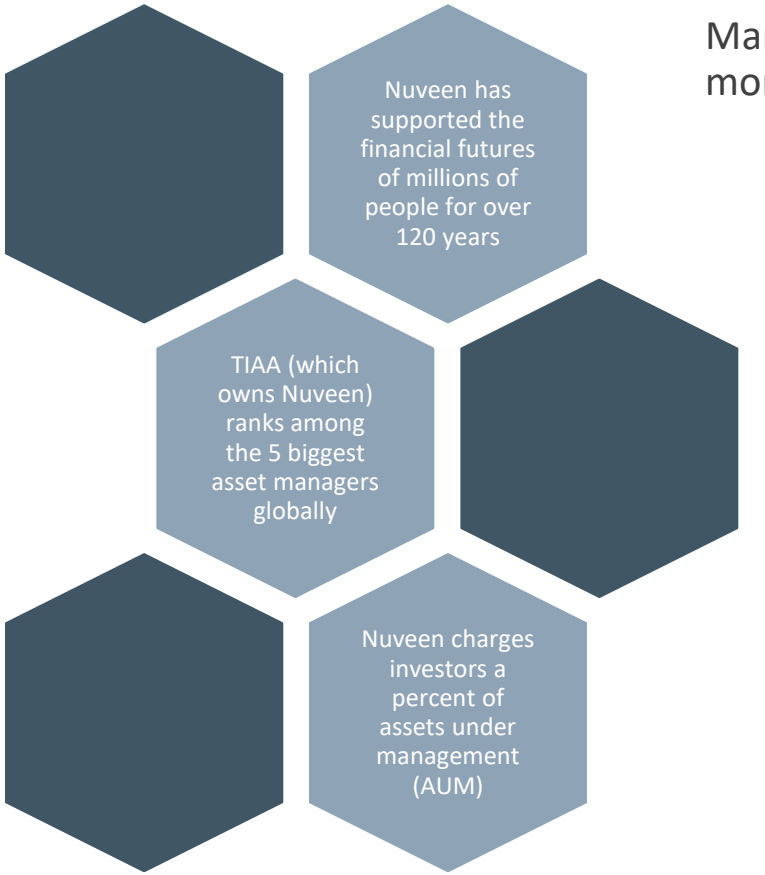
nuveen
A TIAA Company

Tables of Contents

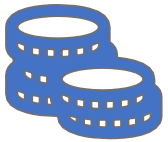
EMERITUS

nuveen
A TIAA Company

1. Background



Management, sales and redemption fees increase with more asset amounts, more funds added and longer investment duration



Decreasing Cost of Acquisition

Sales channels:
- In-person salesforce (wholesalers)
- Meetings and events
- Telephone salesforce (internal wholesalers)
- Email and direct mail marketing
- Social media

Cost of acquisition vs fees earned

Resource Allocation

EMERITUS

nuveen
A TIAA Company

2. Objectives – Predict & Recommend on ADR

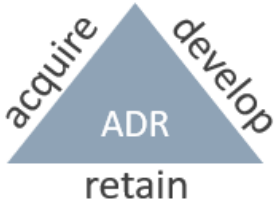
Predict next year's sales per advisor using previous year's data

- Break the sales down into segments
- Compare to actual sales and feature correlation

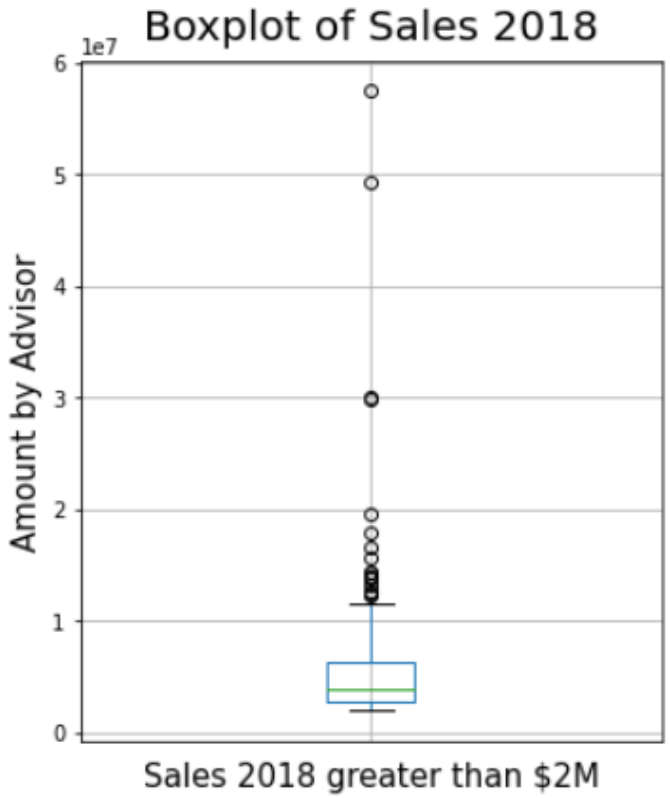
Predict if advisors will add funds in the next year

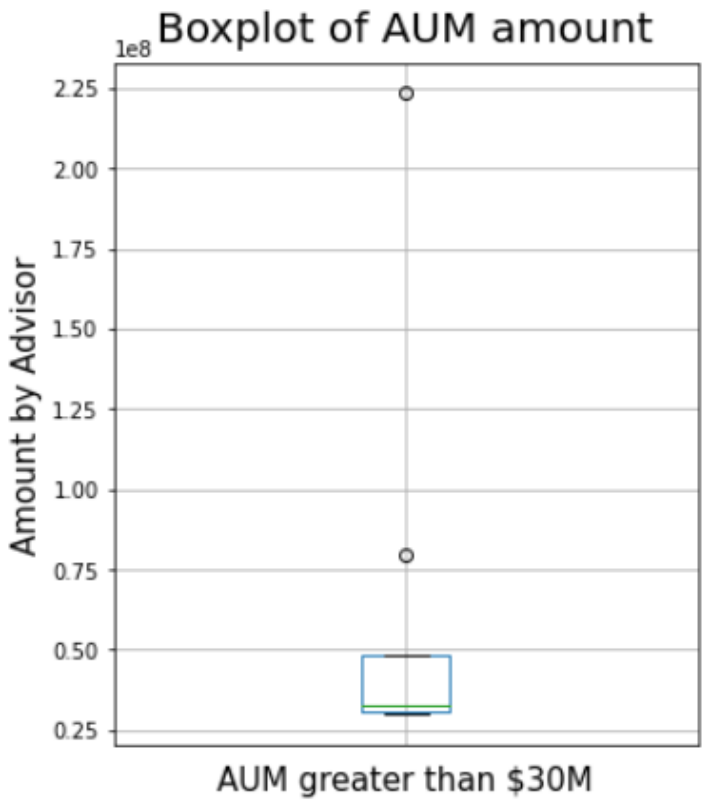
- Find feature correlations to deciles
- Examine which segments make up these deciles

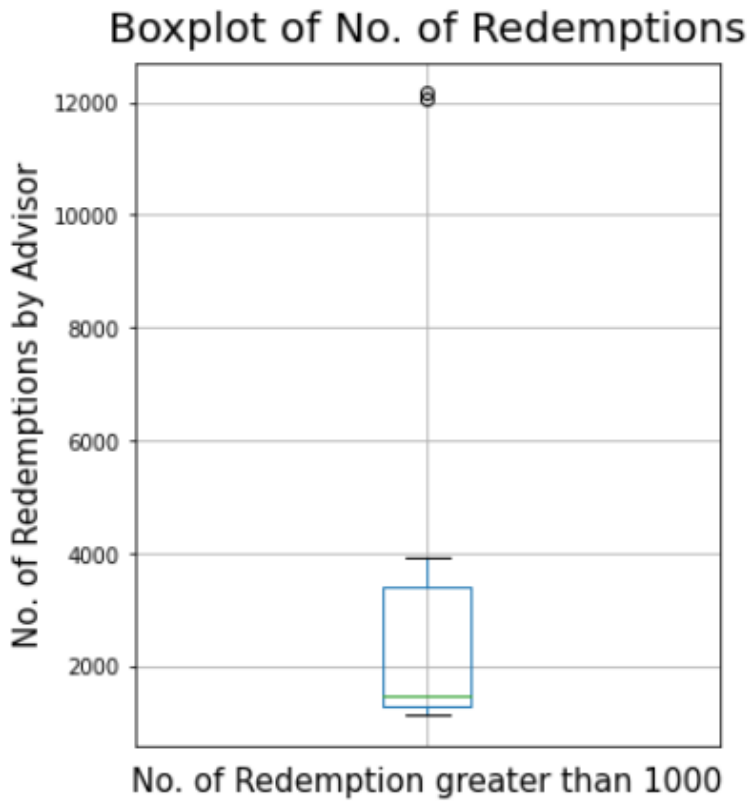
A good model will point out key advisor groups and channels to focus sales efforts on, and thus help saving costs



EMERITUS

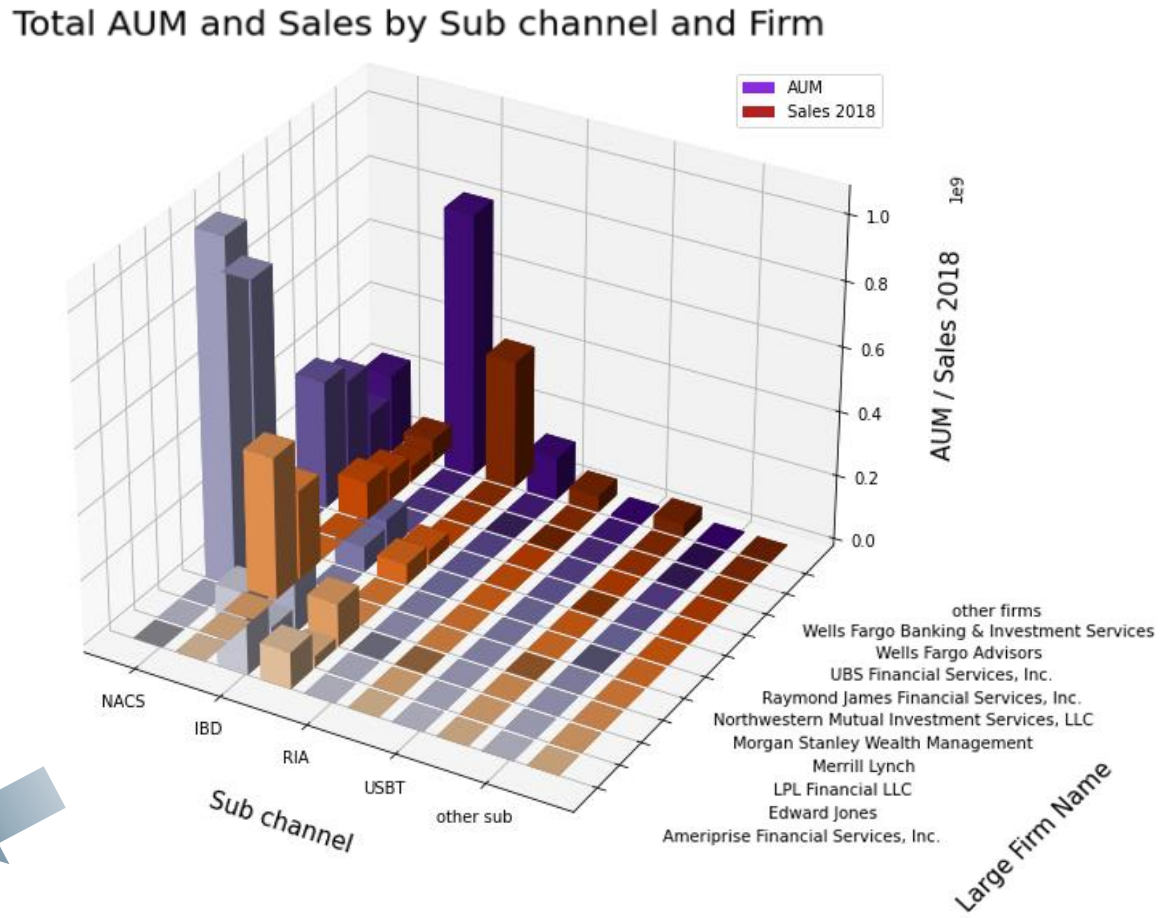






3. Many zero values & outliers may present challenges for predictions (outliers may contain valuable insights)

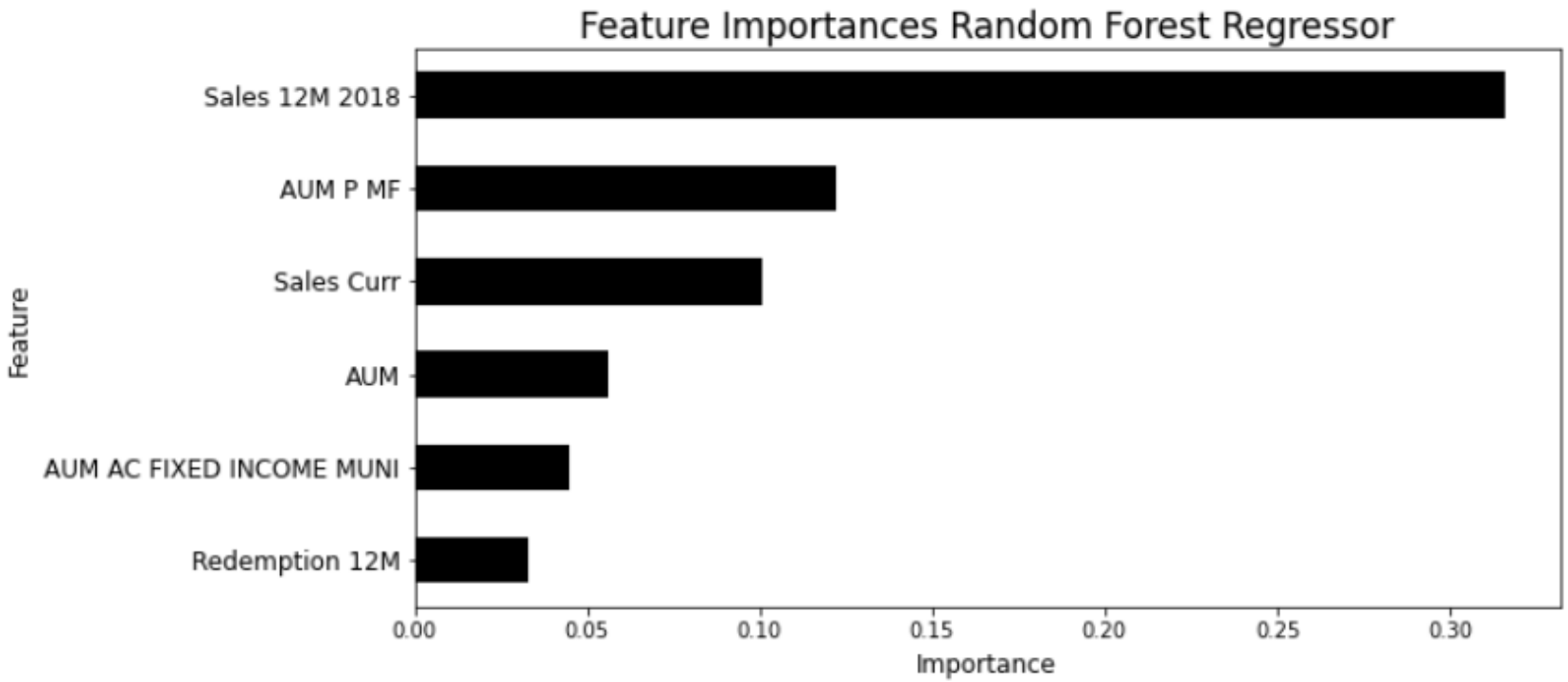
3. Massively segmented data set - Expectation to gain valuable insights by dividing data into Sub channels and largest advisor firms



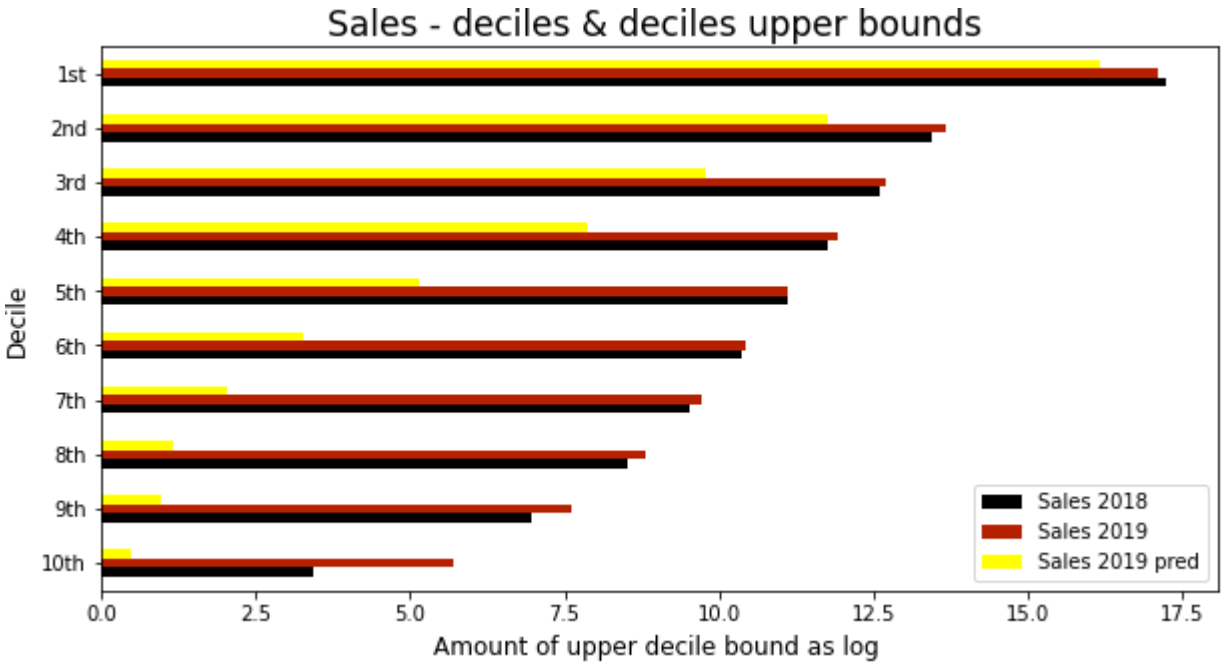
e.g. Morgan Stanley and Merrill Lynch make up the greatest part of AUM/Sales within NACS

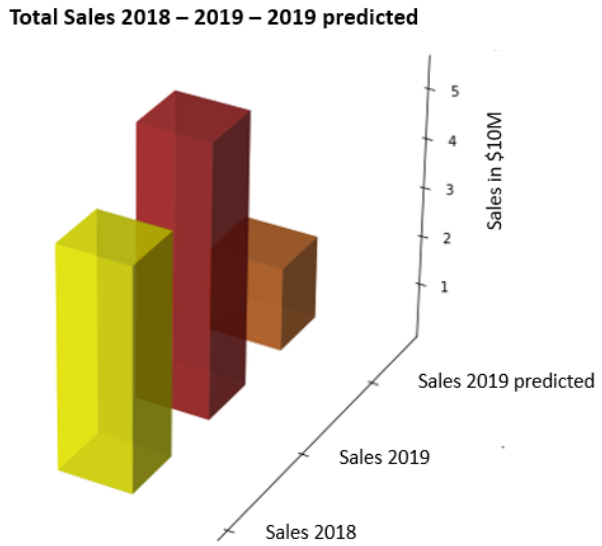
EMERITUS

4. Regression Results - Most Important Variables condense into Sales, AUM-related and Redemption "amount-features"



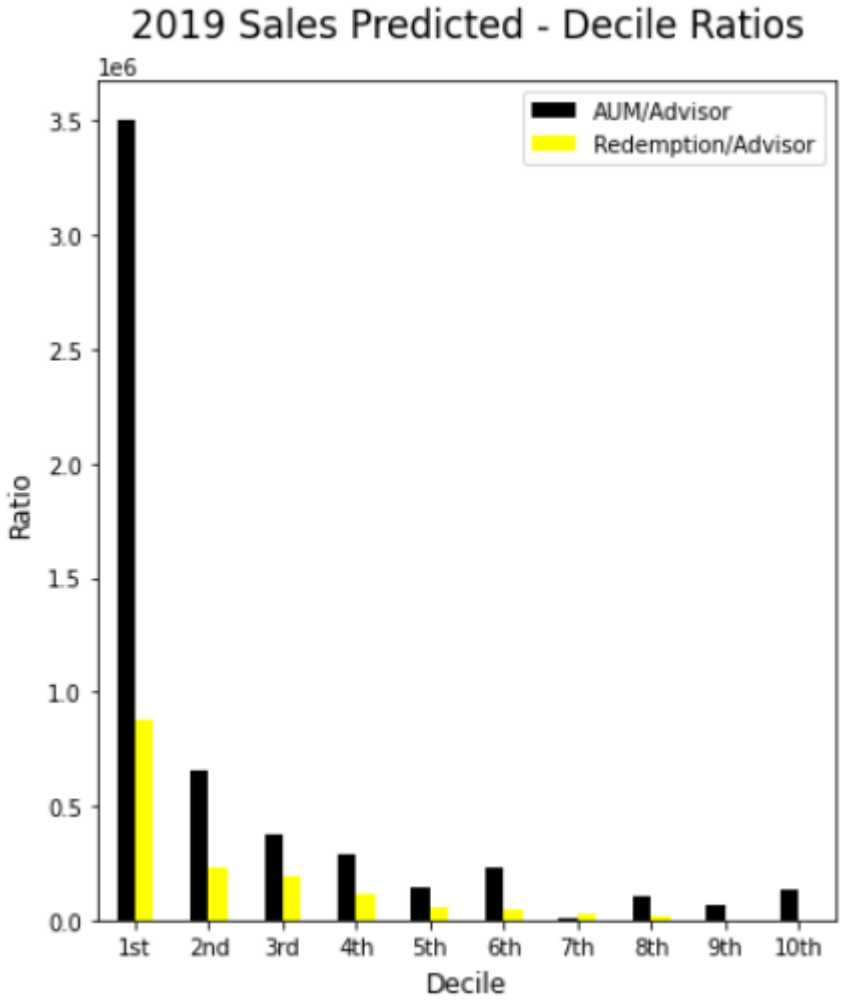
4. High-sales-deciles predicted well despite total differences in model





- Model total difference actually provides valuable insights laid out in the following

- The model does better in predicting the higher sales advisors
- As evident in EDA, the many zero values present a challenge for prediction (underprediction)
- Actual sales 2018 & 2019 have much higher counts in the lowest deciles (many zero values)



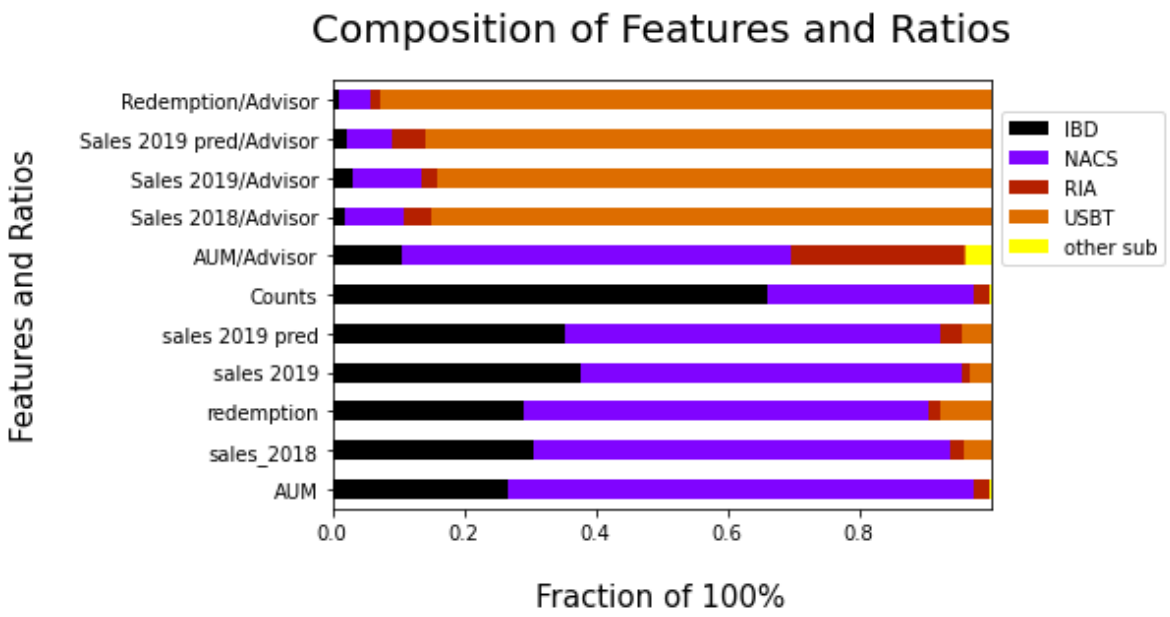
4. More active advisors also generate higher sales

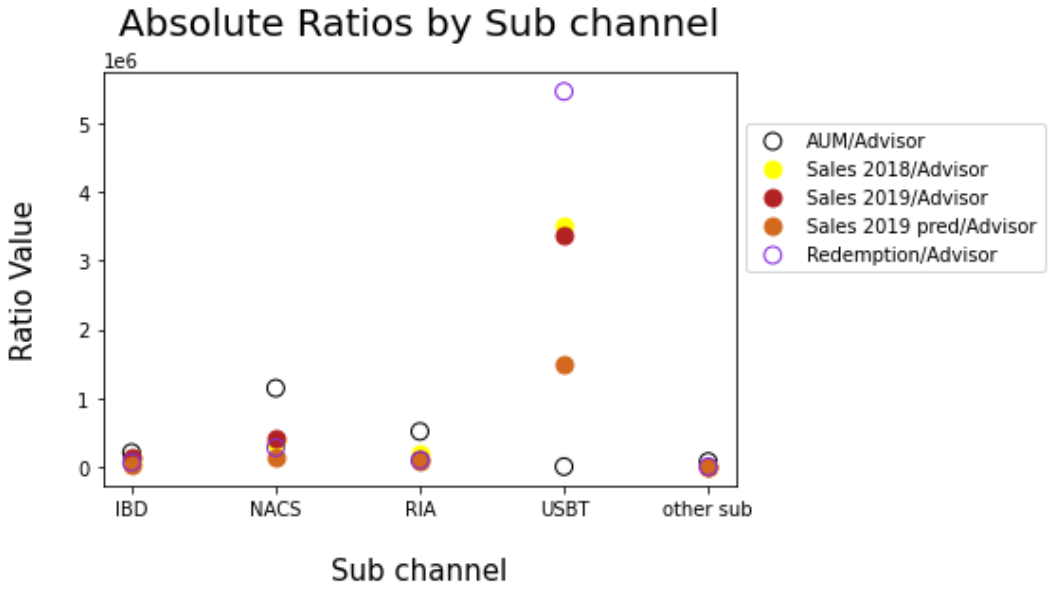
- Advisors in the high sales' deciles also manage a higher AUM and have *higher* redemption amounts
- Noteworthy: redemption is *positively correlated* to amount of sales

- The top decile makes up about 91% of total sales predicted and its composition of sub channels and firms reflects overall segmentation of sales, too.

EMERITUS

4. Great consistency of predicted/actual features and ratios with one exception

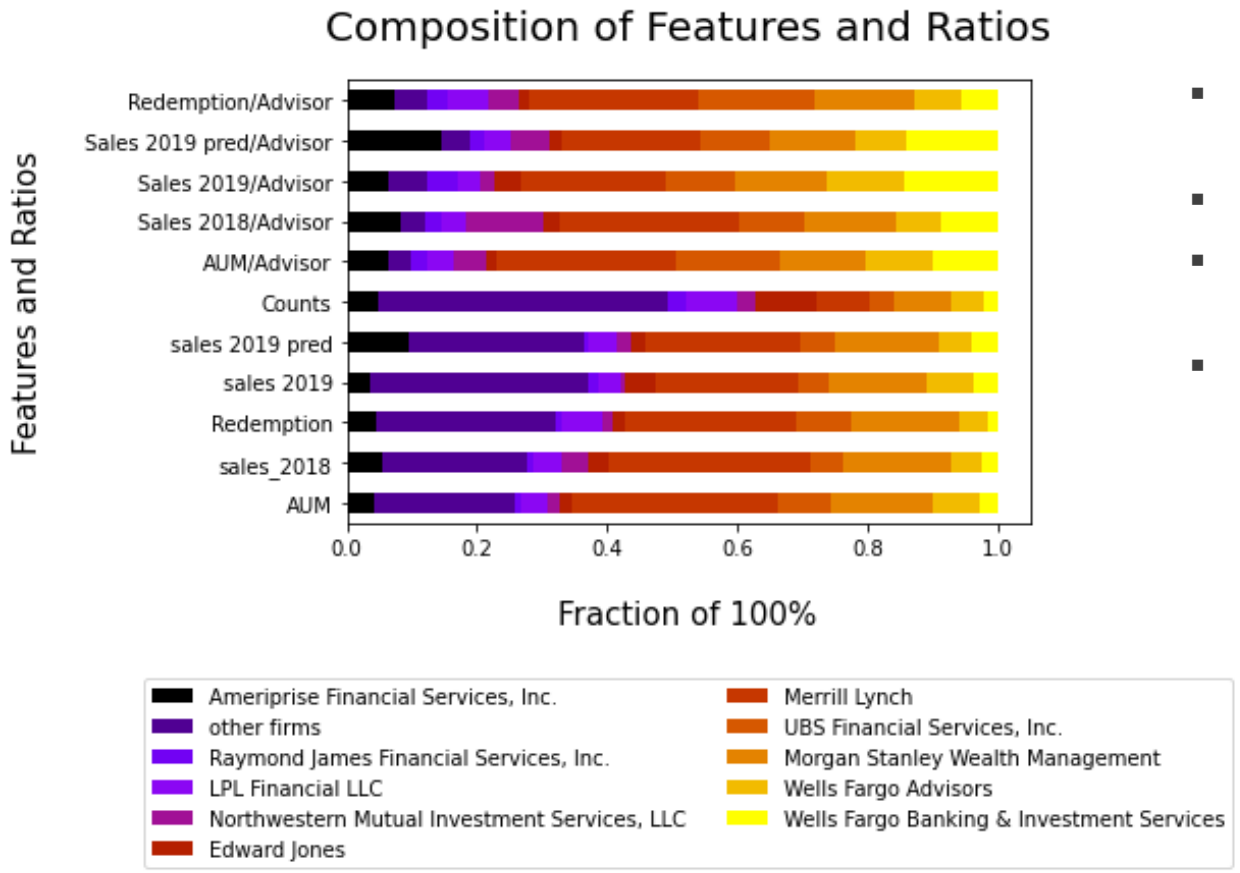




- Advisors in the smaller sub channels have a higher fraction of overall AUM (RIA), Sales & Redemption (USBT) per advisor

- The model underpredicts high performing advisors in USBT (US Bank Private Wealth Management)
- A reason for total model deviation lies here

4. Relative ratios also accurate for firm segmentation

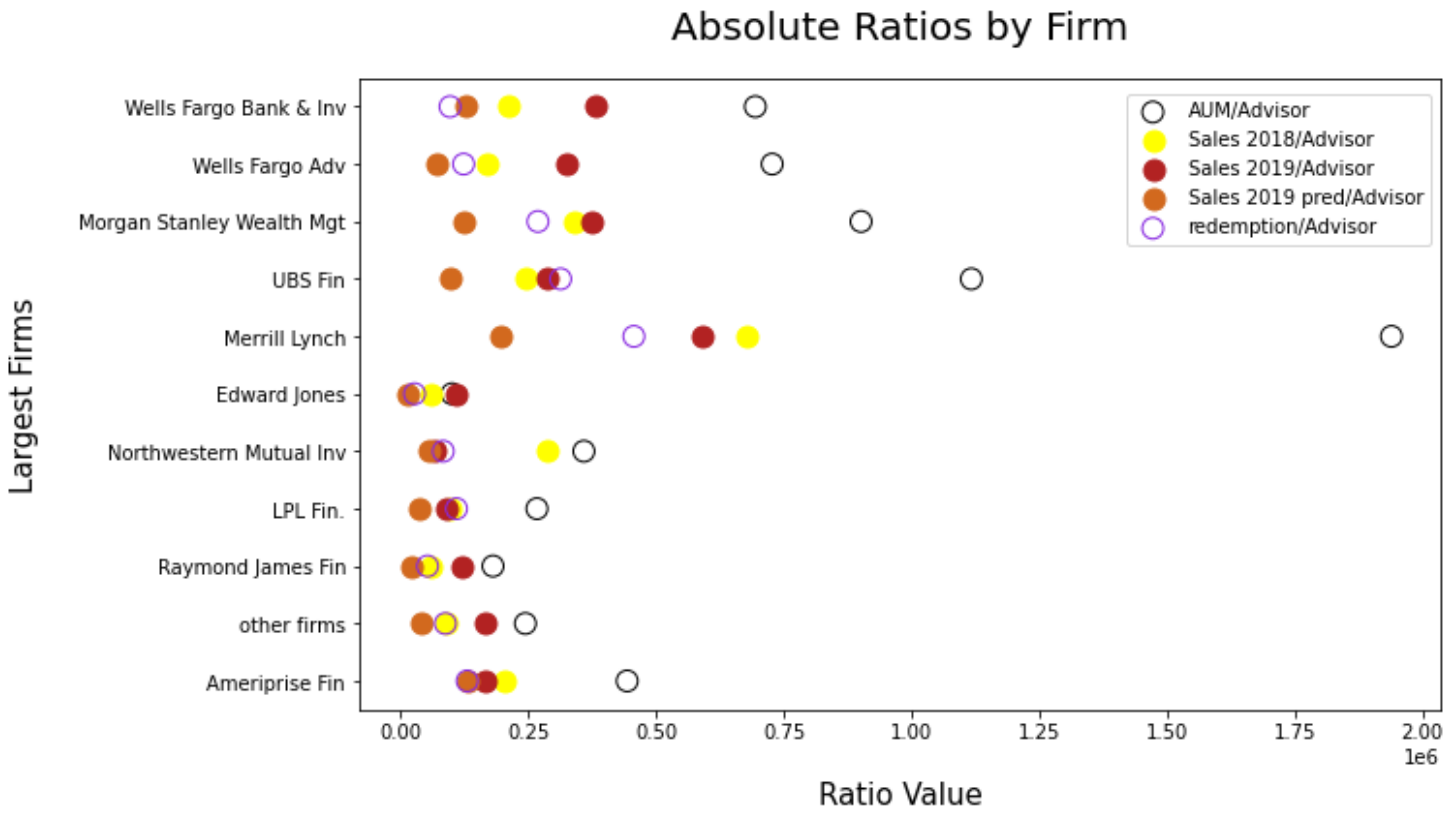


- E.g. decrease in Northwestern Mutual for 2019 sales / advisor predicted correctly by model
- Merrill Lynch, Ameriprise overpredicted
- Wells Fargo Bank. & Fin. And Morgan Stanley perform better per advisor
- Other firms have high proportion of absolute numbers, but small per advisor

EMERITUS

nuveen
A TIAA Company

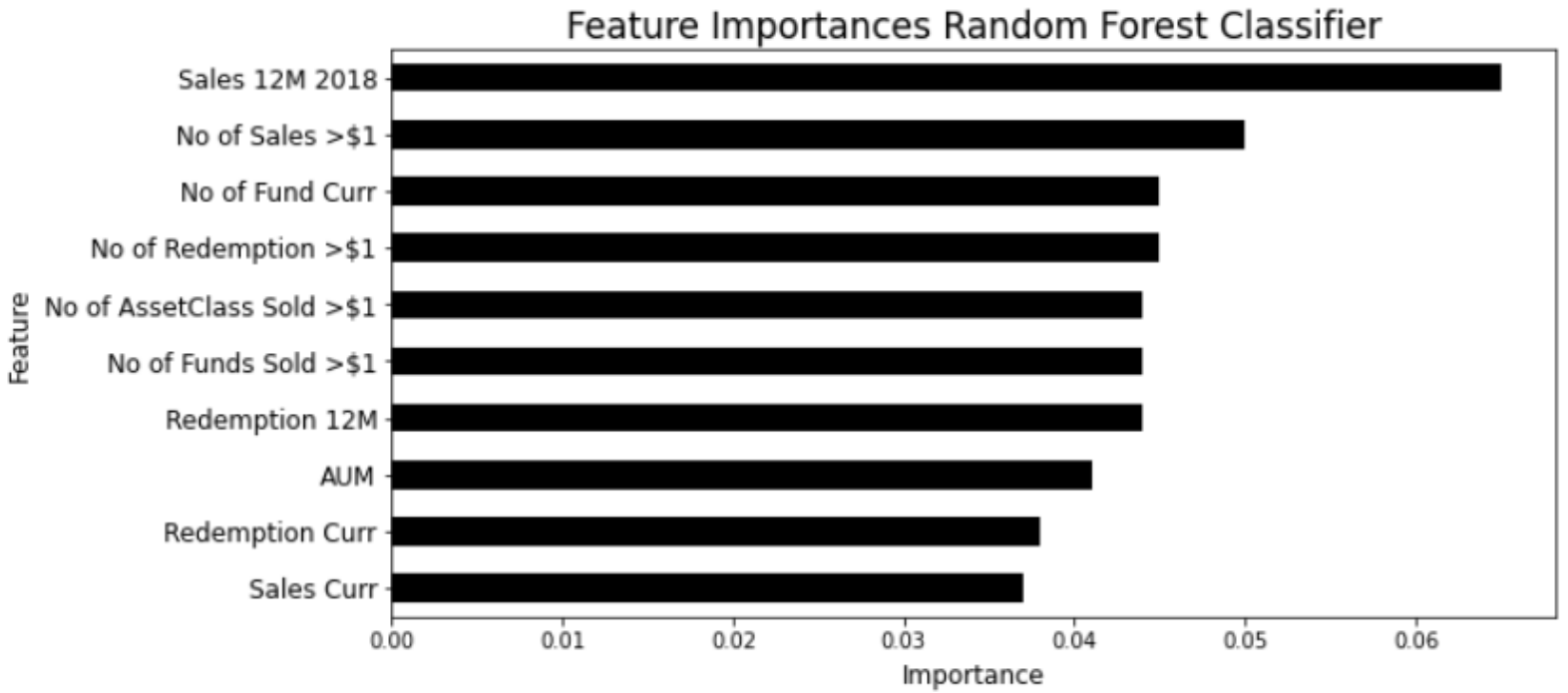
4. Absolute predicted ratios show the right trend



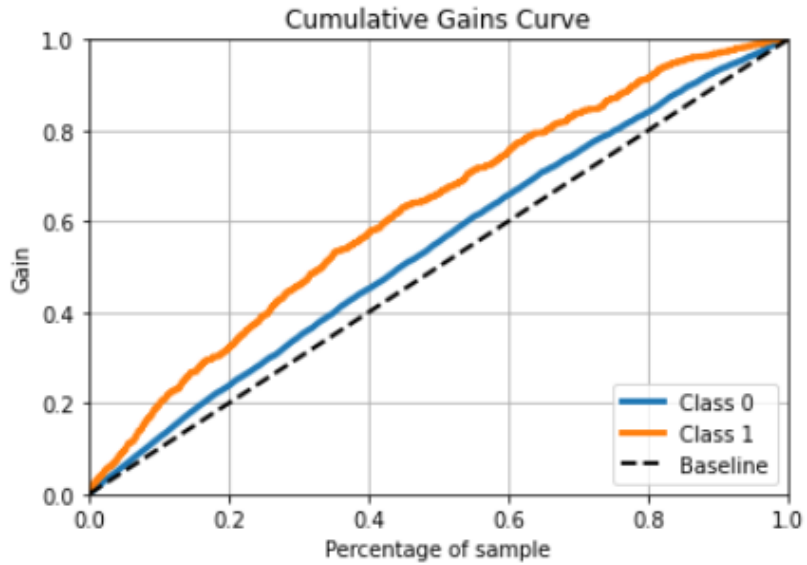
- Overall good predictions
- Higher performing advisors are underpredicted in model
- Among the 10 most frequent client companies, the ones with higher sales also have higher AUM

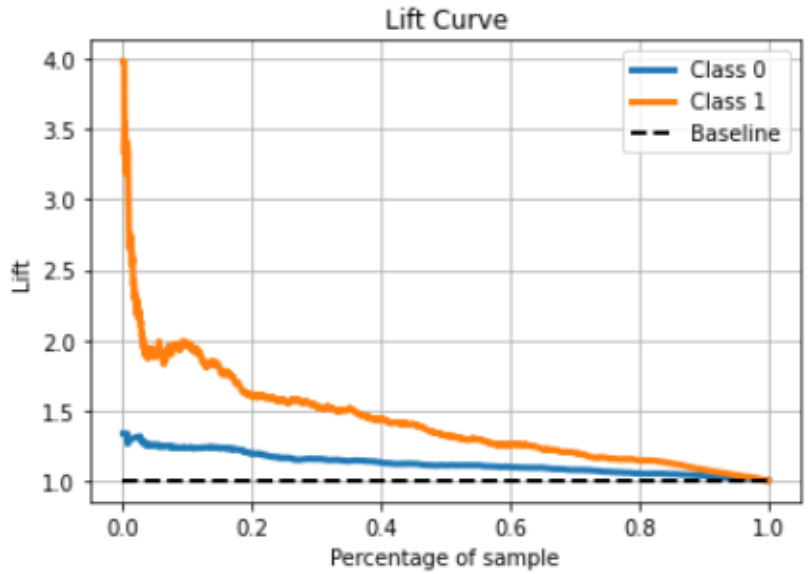
EMERITUS

4. Classification Results - Most Important Variables are more evenly distributed and contain "No. of .." features



4. About 40% of advisors have a lift of 50% or higher





- The model is especially helpful for Class 1 predictions with a max gain of about 20%
- About 40% of the data has a lift of 50% or higher over the baseline for Class 1

EMERITUS

nuveen
A TIAA Company

4. Table confirms Lift plot insight with exactly 38% of advisors at 49% cum. lift

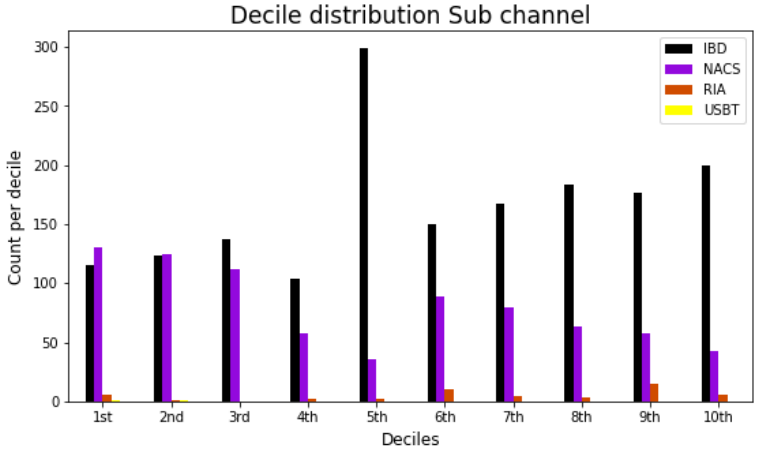
Decile	Number	Lift_positive	Lift_positive Index	Num. Cum.	Cum. Lift positive	Cum. Lift
1st	251	0.49	91%	251	0.49	91%
2nd	250	0.4	56%	501	0.45	74%
3rd	250	0.38	48%	751	0.42	65%
4th	250	0.26	2%	1,001	0.38	49%
5th	250	0.19	-26%	1,251	0.34	34%
6th	337	0.25	-2%	1,588	0.32	27%
7th	163	0.22	-14%	1,751	0.31	23%
8th	250	0.2	-22%	2,001	0.30	17%
9th	250	0.11	-57%	2,251	0.28	9%
10th	251	0.06	-77%	2,502	0.26	0%
Total	2,502	0.26				

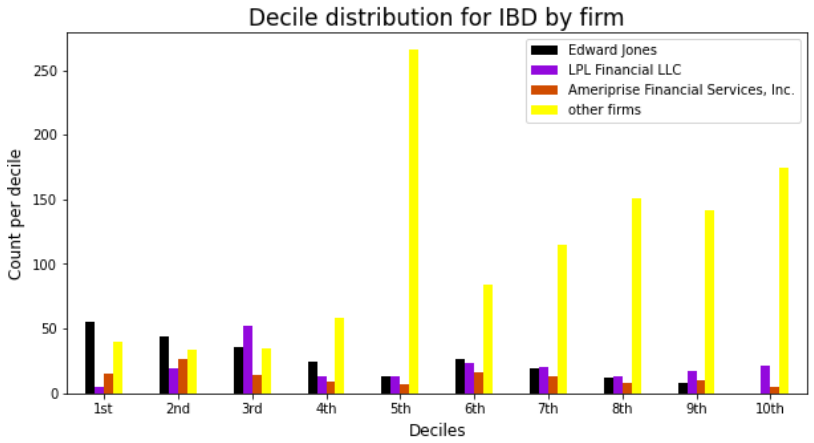
- Average probability of adding a fund is 26%
- The top 3 deciles carry the whole population (above average lift index)
  - This means that for the top 751 advisors, 488 (65%) more correct predictions on funds added can be made as opposed to randomly contacting these advisors
- Except for the lowest 2 deciles it is worth pursuing the probability indication with the associated advisors (i.e. apply sales measures beyond mass email blasts)

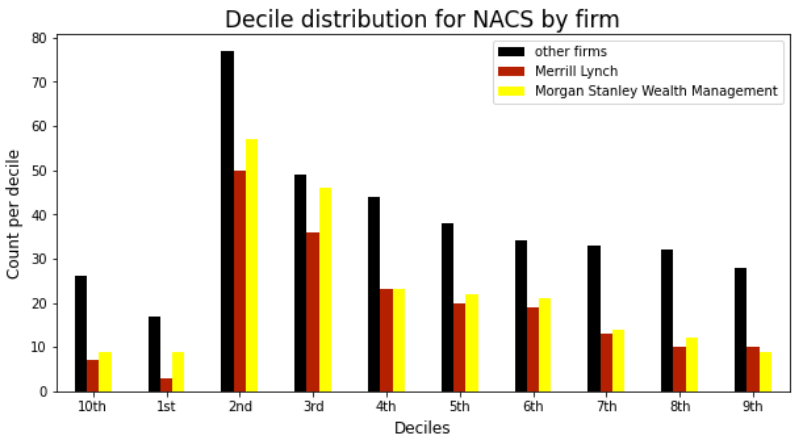
EMERITUS

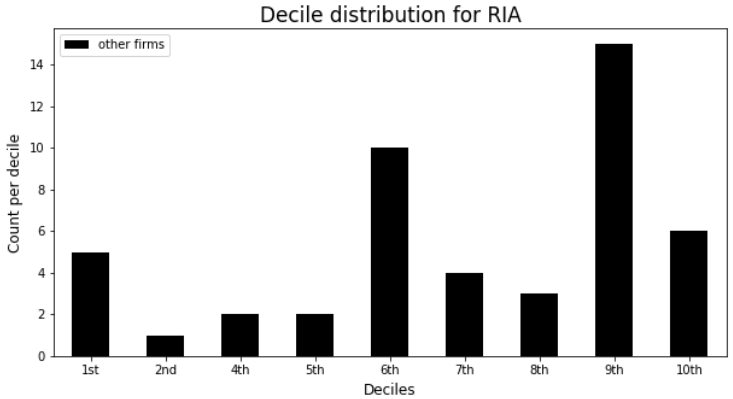
nuveen
A TIAA Company

4. Smaller firms gained through IBD are of the lower probability type, compared to NACS





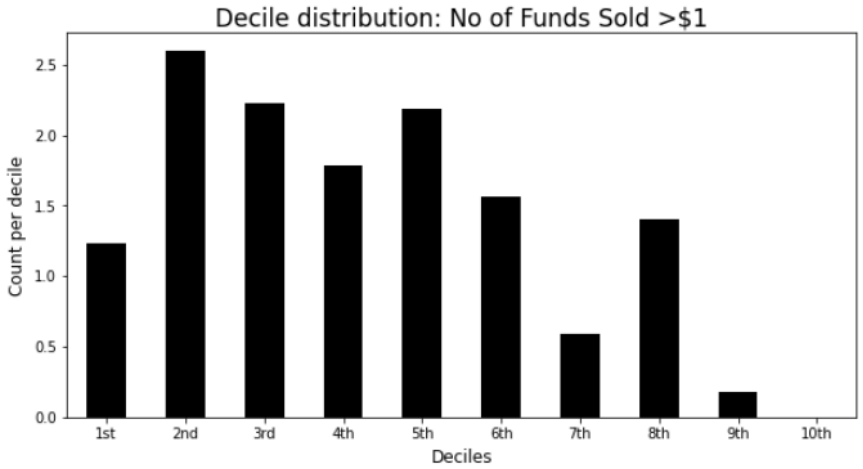


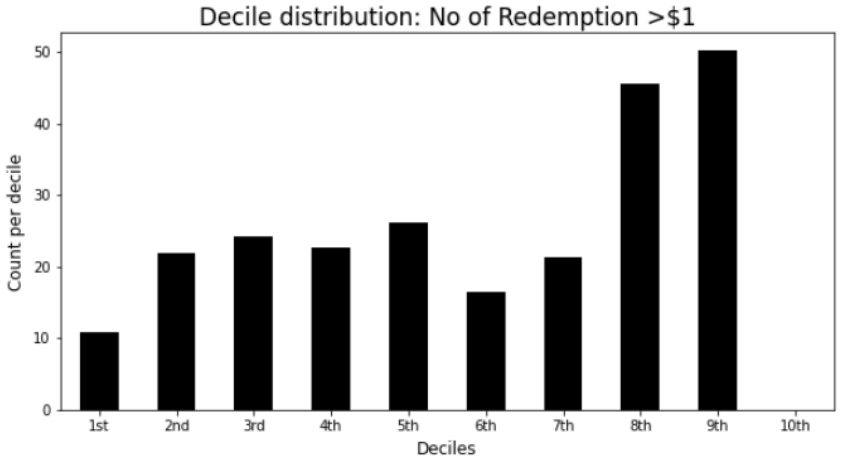


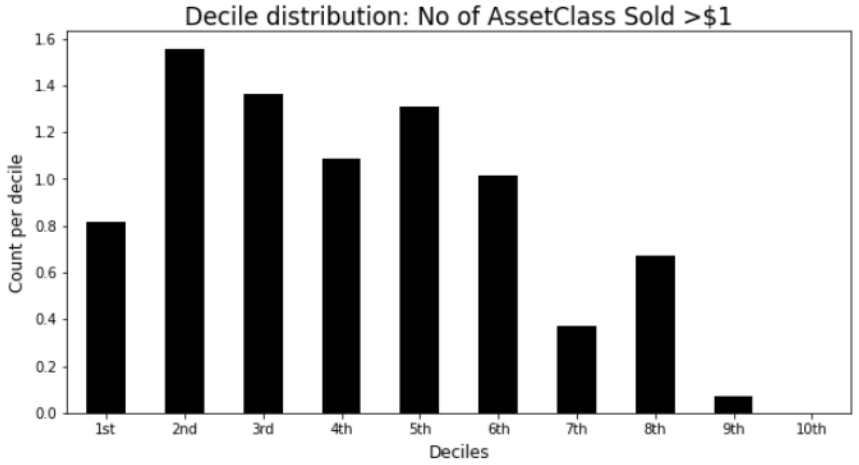
- IBD has more advisors in the lower deciles
- Clients of the largest firms are more likely to add a fund within IBD
- NACS has higher quality “other firms” (allocate more resources)
- RIA tends to have more advisors that don’t add sales

EMERITUS

4. “Number of..” features classify funds-added better







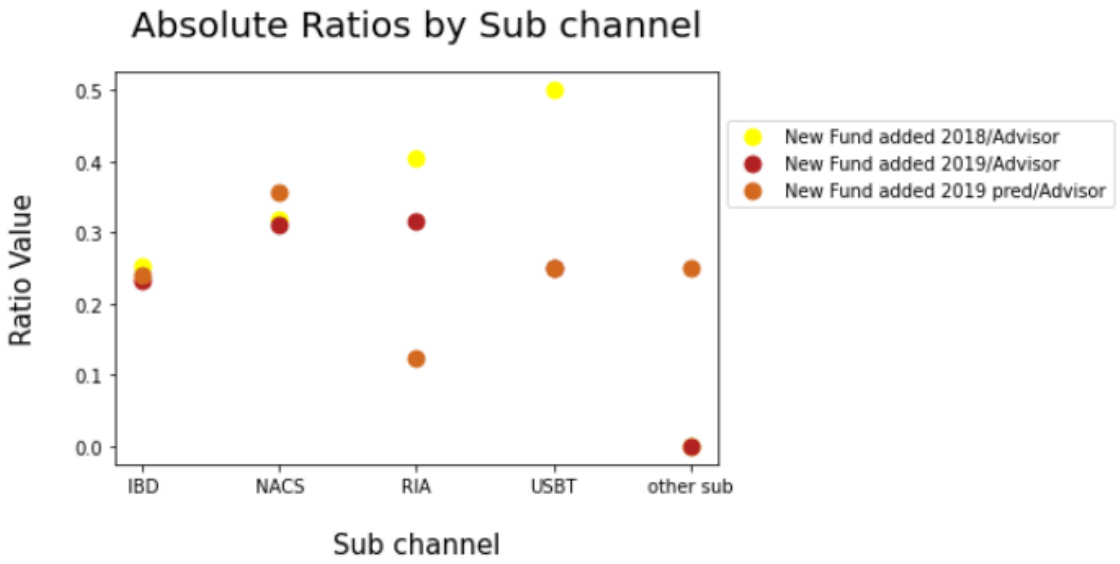
- Only 3 variables show a clear trend
- If advisors sold more funds / asset classes or have less *no. of* redemptions in 2018, they are more likely to add a new fund in 2019

No. of redemption is negatively correlated

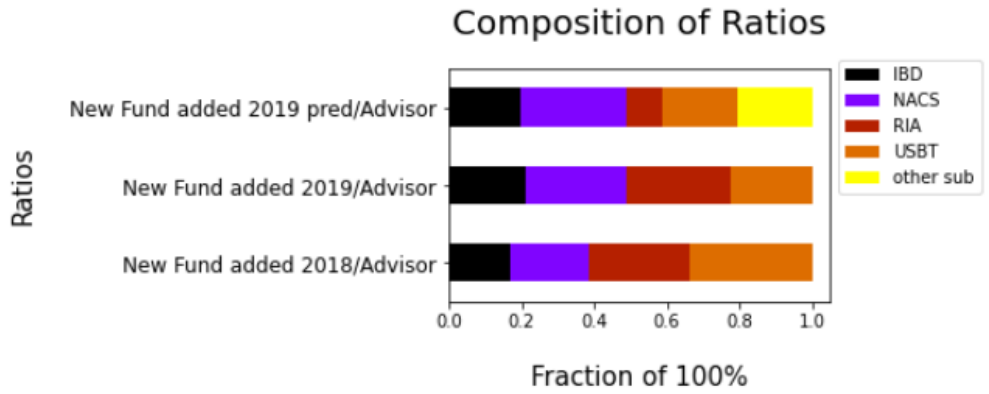
EMERITUS

nuveen
A TIAA Company

4. Average ratios don't show much differences, but NACS tends do better than IBD here ,too



- RIA is underpredicted, but the direction is predicted correctly (i.e. less funds added per advisor in the following year)

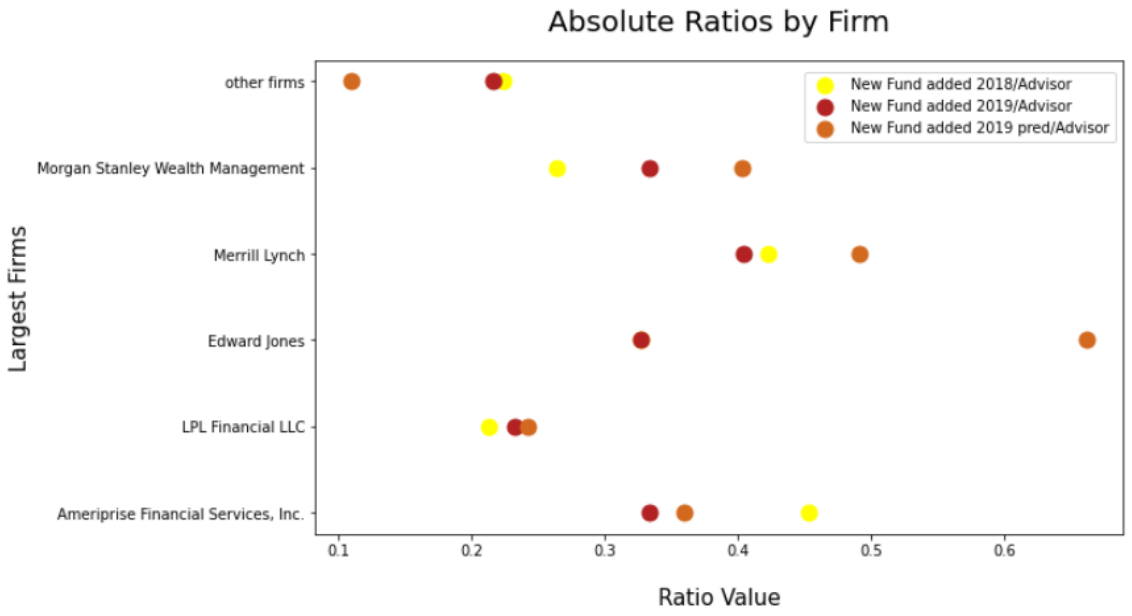


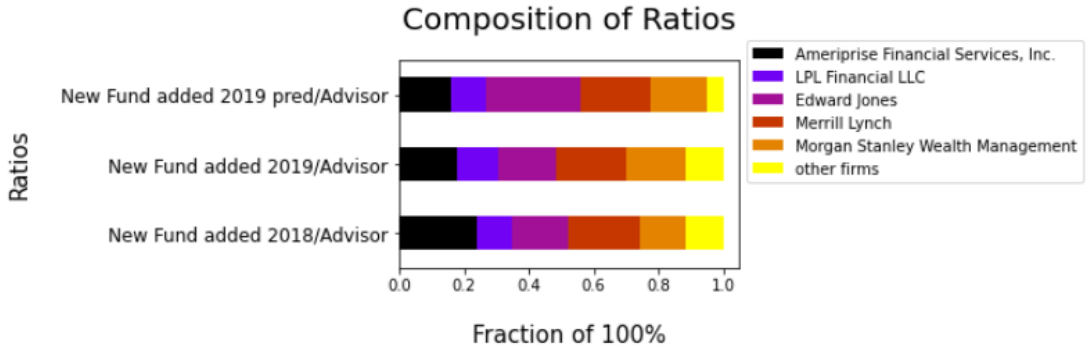
- Higher sensitivity for ratios with low counts (USBT, other sub)

EMERITUS

nuveen
A TIAA Company

4. About half of the big firms are more likely to add a fund, but higher sensitivity of results





- Morgan Stanley, Merrill Lynch, Ameriprise and Edward Jones advisors are more likely to add a fund
- Edward Jones is overpredicted

- Relative ratios are well predicted
- Other firms less likely to add a fund on average

EMERITUS

5. Recommendations I-III

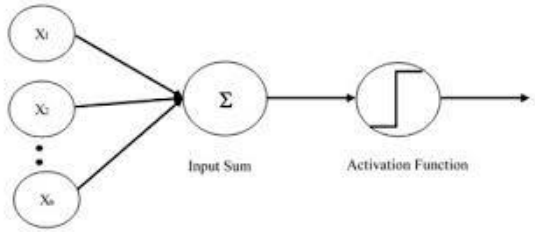
- Pay attention to previous year's sales, redemption and AUM for sales prediction.
  - Relative segment ratios and overall direction are very accurate predictors, more than the exact absolute amount because of the many zero values.

- In-person or phone marketing for high sales predictions, for high probability deciles, for the RIA channel (high AUM) and definitely the USBT outliers!

- If advisors made high sales or had high AUM the previous year, don't put resources into preventing redemptions.
  - Redemptions are positively correlated to generating sales and bring in fees, too.

5. Recommendations IV-VII

- Number of redemptions is negatively correlated to funds added.
  - 2 more variables to monitor during the year for predicting funds added: Number of funds and asset classes sold (positively correlated).
- Amount features more relevant to sales, “Number of features” more relevant for funds-added predictions.
  - This is logical, marketing has to keep this in mind when contacting advisors.
- Create virtual training sessions for low advisor ratios (smaller) firms.
- Nourish the relationship with big name firms, and nurture smaller firms within NACS as these have more potential.



Thank you !

nuveen
A TIAA Company

Appendix - Data Processing

Data: 10005 advisors	
	Sample Features
Transaction Data 2018	no. & amt of funds & asset classes sold & redeemed, AUM of different products
Transaction Data 2019	12M sales & no. funds added
Firm Information	Firm names, Sales (Sub-) Channels
Sub channels	IBD 6679, NACS 3075, RIA 219, USBT 16, other Sub channels 16

One-Hot Encoding			
Criteria for firms: Largest by number of advisors (most data points)			
	# Sub channels	# Channels	# Firms
Regression	4	--	10
Classification	4	4	5

Preliminary Data Operations
Merging of DataFrames
Convert redemption_12M and redemption_curr to positive values
Setting all remaining negative values in DataFrame to zero
Sales_curr not added to sales_12M as assumed last month relevant

EMERITUS

nuveen
A TIAA Company

Appendix - Feature & Model Selection

Feature Selection	Features	Heatmap (I) (corr>0.59)	Recursive Feature Elimination (RFE) (II)	RF Feature Importance (III) <0.01	Non-categorical features removed (count: I+II+III>1)
Regression	52	16	26 (Linear Regression)	7+ all categorial	10
Classification	57	16	26 (RF)	15 + 12 categorial	9

Model Selection (Regression)

Method	Base Model	RF	RF alt.
Data	SS, only 'Sales 2018'	PT, 42 features	PT, 6 features (feat. Imp.>0.03), (III)
Model	LR	Max_depth=9	Max_depth =9
R2	0.33	0.52	0.53
RMSE	0.66	0.47	0.48
Avg CV=5	0.36	0.51	0.49
Other well performing regressors modeled: AdaBoost, XGBoost, NN Other effective Data Preparation methods: PCA			

Model Selection (Classification)

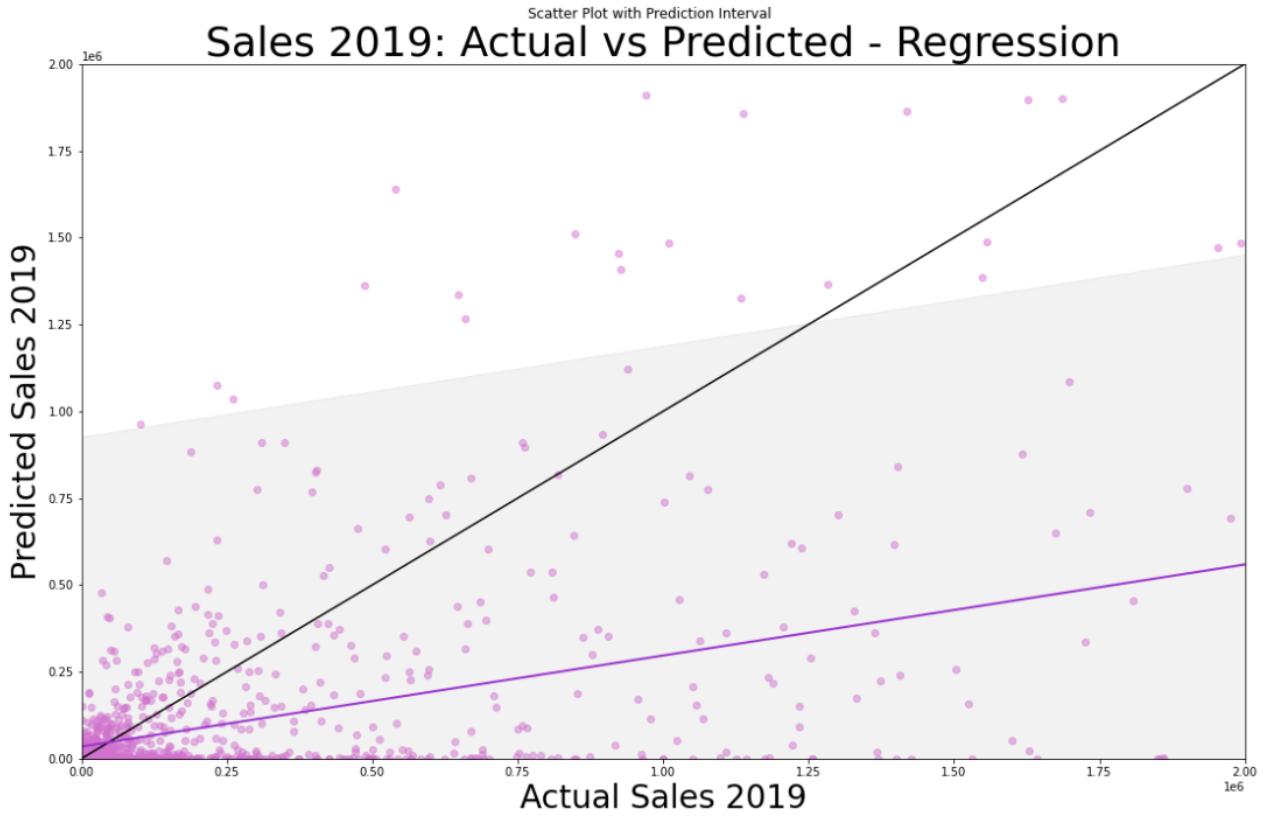
Method	RF	RF alt.
Data	SMOTE (ratio=0.35), PT, 48 features	SMOTE()PT, 10 features (feat. Imp.>0.03),(III)
Param.	Max_depth=13	Max_depth =13
Validation Score (Test Data)	0.76 over 0.66 (Class 0 ratio)	0.74 over 0.64 (Class 0 ratio)
Avg CV=5	0.74	0.74
Other well performing classifiers modeled: Boosting Classifiers, NN Other effective Data Preparation methods: RandomOverSampler		

Random Forest RF, Linear Regression LR, StandardScaler SS, PowerTransformer PT
NN Neural Network

EMERITUS

nuveen
A TIAA Company

Appendix - Actual vs predicted sales show some values are underpredicted but most values lie in prediction interval

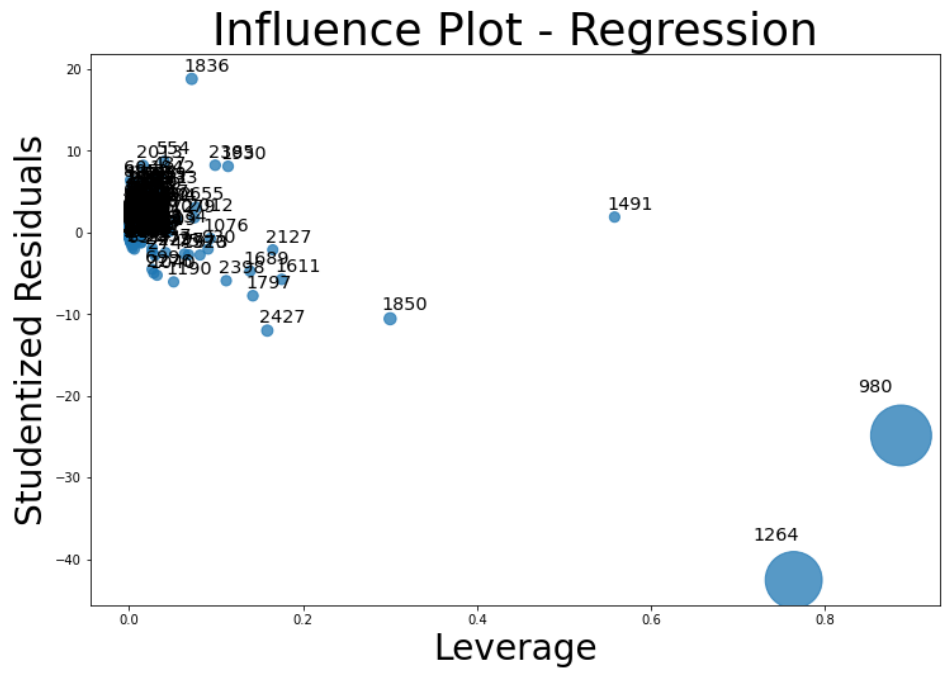


- Graph shows actual vs predicted values for sales 2019
- Black line shows ideal trend
- Orchard line shows predicted trend
- Shaded area shows predicted interval
  - A single new advisor predicted is expected to lie in this interval 95 out of 100 times
  - For the data given, this means that overpredicted values are scarce
- Some points are underpredicted, these are often advisors with zero 2018 sales

EMERITUS

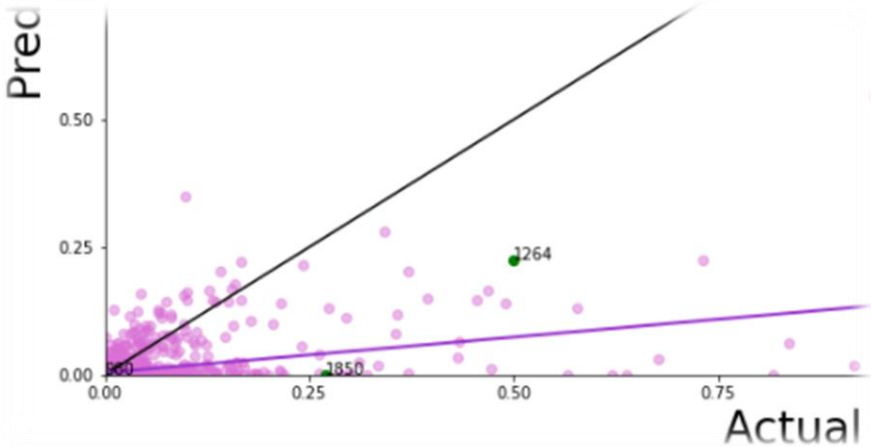
nuveen
A TIAA Company

Appendix – Examining the removal of regression leverage points to improve model metrics



- Leverage points take extreme values in the independent variables / distort regression parameters
- Two iterations of removing leverage points don't improve the metrics, therefore these data points were kept.

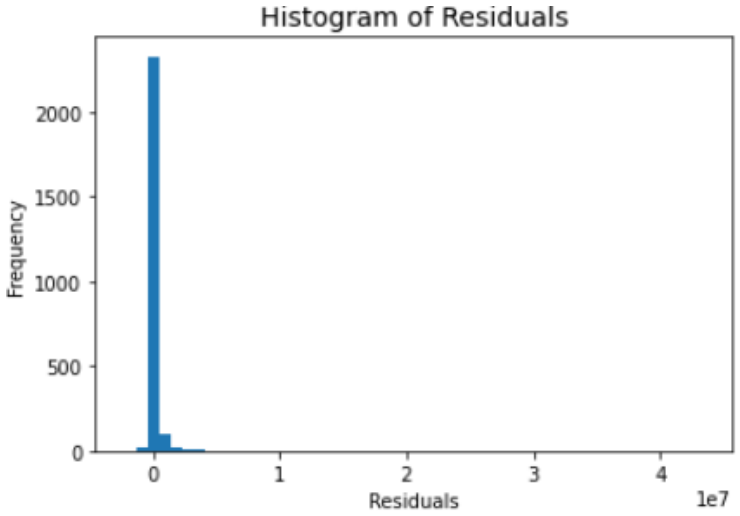
Scatter plot (Actual-Predicted sales) snippet: Tracking Leverage points



	RF w/o leverage points
R2	0.46
RMSE	0.54
Avg CV=5	0.42

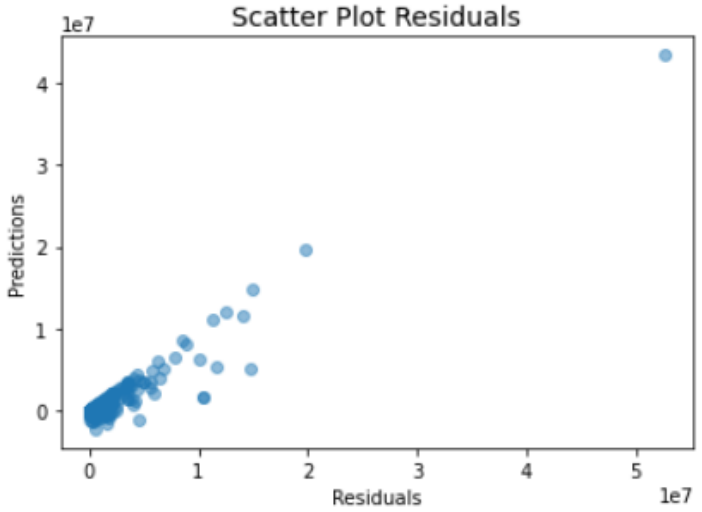
- Unlike outliers, leverage points may follow the regression trend

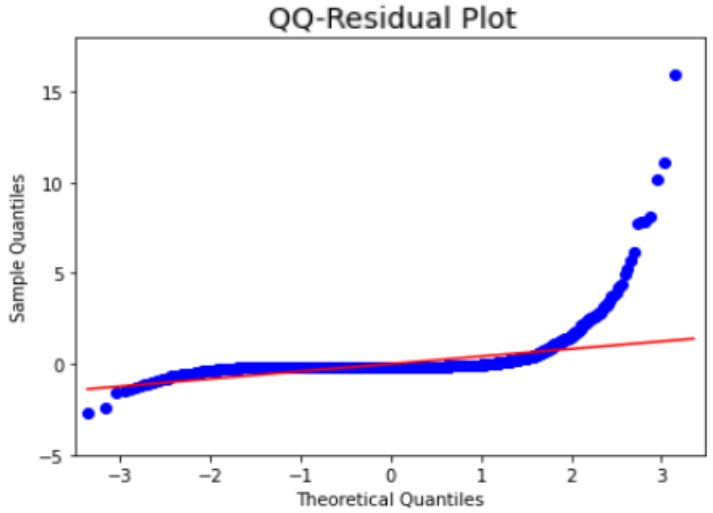
EMERITUS



Appendix - Regression Residuals

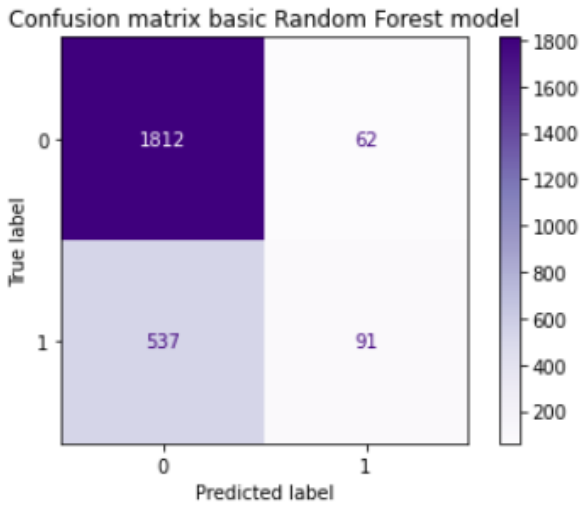
- Residuals are overwhelmingly normal by visual inspection
- This adds to the validity of the model

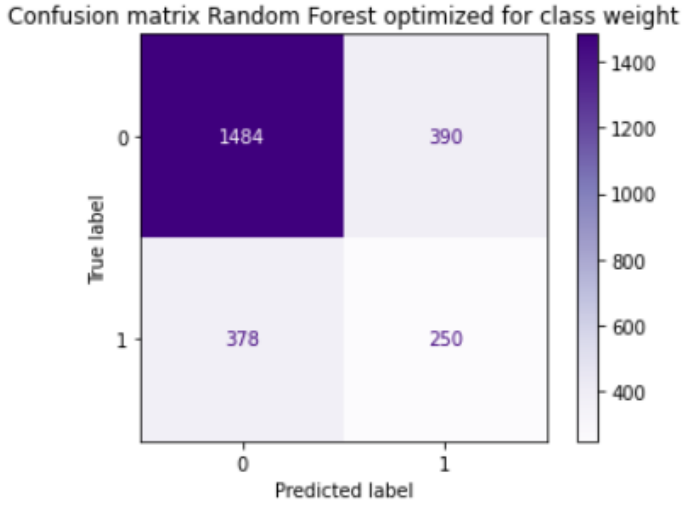




nuveen
A TIAA Company

Appendix - Confusion Matrices





	RF base
Validation Score (Test Data)	0.76 over 0.66
Avg CV=5	0.74

Updated Result	RF opt
Validation Score (Test Data)	0.69 over 0.66
Avg CV=5	0.70

Credit: https://www.kaggle.com/eikedehling/exploring-class-imbalance-resampling-and-weights

```
weights=np.linspace(0.01, 0.99, 20)

gsc = GridSearchCV(
    estimator=RandomForestClassifier(n_estimators=100, max_depth=13),
    param_grid={
        'class_weight':
        [{0: x, 1: 1.0-x} for x in weights]
    },
    scoring='f1_macro'

)
grid_result = gsc.fit(Xtr_sp,ytr_s.ravel())

print("Best parameters : %s" % grid_result.best_params_)
```

```
Best parameters : {'class_weight': {0: 0.2163157894736842, 1: 0.7836842105263158}}
```

- The optimized model was used for results because of the better ability to predict 1s correctly

EMERITUS