Results from Original File

Logistic Regression with SMOTE

logistic regr training set	ession model	- SMOTE	RFM	
	precision	recall	f1-score	support
0	0.60	0.62	0.61	3073
1	0.60	0.59	0.60	3073
accuracy			0.60	6146
macro avg	0.60	0.60	0.60	6146
weighted avg	0.60	0.60	0.60	6146
test set				
	precision	recall	f1-score	support
0	0.93	0.60	0.73	1330
1	0.17	0.65	0.27	170
accuracy			0.61	1500
macro avg	0.55	0.63	0.50	1500
veighted avg	0.85	0.61	0.68	1500

logistic training	_	ession model	L - SMOTE	CLV	
		precision	recall	f1-score	support
	0	0.59	0.62	0.61	3073
	1	0.60	0.58	0.59	3073
accur	acv			0.60	6146
macro		0.60	0.60		
weighted	_	0.60	0.60	0.60	6146
test set					
		precision	recall	f1-score	support
	0	0.93	0.61	0.74	1330
	1	0.17	0.62	0.27	170
				0.61	1500
accur					1500
macro		0.55	0.62		1500
weighted	avg	0.84	0.61	0.69	1500

XGBoost with SMOTE

Will train un [1] valid [2] valid [3] valid [4] valid [5] valid [6] valid Stopping. Bes	ation_0-auc:0 til validatio ation_0-auc:0 ation_0-auc:0 ation_0-auc:0 ation_0-auc:0 ation_0-auc:0	on_0-auc l 0.666252 0.64722 0.643182 0.651329 0.666252 0.664854	hasn't imp	roved in 9	5 rounds
+					
training set	precision	recall	f1-score	support	
0	0.64	0.57	0.61	3073	
1	0.61	0.68	0.65	3073	
accuracy			0.63	6146	
macro avg	0.63	0.63	0.63	6146	
weighted avg	0.63	0.63	0.63	6146	
test set					
	precision	recall	f1-score	support	
0	0.93	0.57	0.71	1330	
1	0.17	0.66	0.27	170	
accuracy			0.58	1500	
macro avg	0.55	0.62	0.49	1500	
weighted avg	0.84	0.58	0.66	1500	
	Best AUC Scor Accuracy: 0.4 [[594 736] [33 137]] test set	8733333333	333334		
Tuning		precision	recall	f1-score	support
CLV	0	0.95	0.45	0.61	1330
CLV	1	0.16		0.26	170
	accuracy			0.49	1500
	macro avg	a 55		0.43	1500
	weighted avg	0.86	0.49	0.57	1500

```
XGBoost model - SMOTE CLV
[0] validation 0-auc:0.638452
Will train until validation 0-auc hasn't improved in 5 rounds.
      validation 0-auc:0.670705
      validation_0-auc:0.68444
      validation_0-auc:0.684042
      validation 0-auc:0.687351
      validation_0-auc:0.690741
      validation 0-auc:0.692315
      validation 0-auc:0.69236
      validation 0-auc:0.686809
       validation_0-auc:0.68937
      validation 0-auc:0.690741
      validation_0-auc:0.69335
       validation 0-auc:0.693412
      validation 0-auc:0.693538
      validation 0-auc:0.694135
      validation_0-auc:0.694847
      validation 0-auc:0.69479
      validation 0-auc:0.694034
      validation 0-auc:0.693919
      validation 0-auc:0.694715
      validation 0-auc:0.694622
Stonning Rest iteration:
      validation 0-auc:0.694847
```

Let's start...

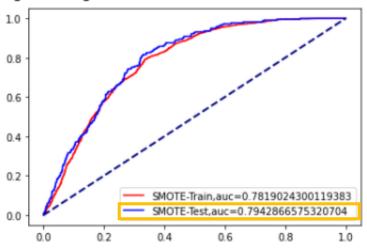
- What to do 1. Drop March 2013 - March 2015
- 2. Fix imbalanced data with undersampling and oversampling
- 3. Calculate AUC
- 4. Compare!

	training set				
		precision	recall	f1-score	support
	0	0.68	0.56	0.62	3073
	•	0.00	0.50	0.02	3073
	1	0.63	0.74	0.68	3073
				0.65	C146
	accuracy			0.65	6146
	macro avg	0.65	0.65	0.65	6146
	weighted avg	0.65	0.65	0.65	6146
	test set				
		precision	recall	f1-score	support
					4224
	0	0.94	0.57	0.71	1330
	1	0.18	0.72	0.28	170
	accuracy			0.58	1500
ı	#3500 3V5	A E6	0.64	0.40	1500
	weighted avg	0.85	0.58	0.66	1500

Logistic Regression SMOTE - RFM

logistic re training se	egression mod et	el - SMOTE	RFM	
	precision	recall	f1-score	support
	0 0.72	0.69	0.71	4389
	1 0.70	0.72	0.71	4389
accurac	су		0.71	8778
macro av	/g 0.71	0.71	0.71	8778
weighted av	/g 0.71	0.71	0.71	8778
test set				
	precision	recall	f1-score	support
	0 0.96	0.68	0.80	1848
	1 0.22	0.77	0.35	218
accurac	су		0.69	2066
macro av	/g 0.59	0.73	0.57	2066
weighted av	/g 0.88	0.69	0.75	2066

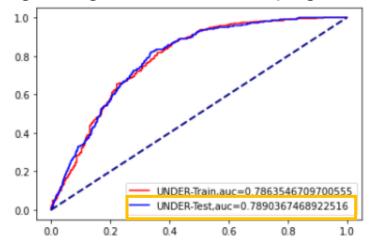
Logistic regression model - SMOTE RFM - AUC



Logistic Regression Undersampling - RFM

logistic regr training set	ession model	- Unders	ampling RFM	1
_	precision	recall	f1-score	support
0	0.74	0.68	0.71	429
1	0.70	0.76	0.73	429
accuracy			0.72	858
macro avg	0.72	0.72	0.72	858
weighted avg	0.72	0.72	0.72	858
test set				
	precision	recall	f1-score	support
0	0.96	0.67	0.79	1848
1	0.22	0.79	0.34	218
accuracy			0.68	2066
macro avg	0.59	0.73	0.57	2066
weighted avg	0.89	0.68	0.74	2066

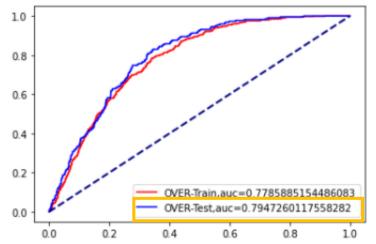
Logistic regression model - Undersampling RFM - AUC



Logistic Regression Oversampling - RFM

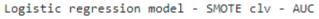
logistic training	_	ession model	- Oversa	mpling RFM	
		precision	recall	f1-score	support
	0	0.71	0.69	0.70	4389
	1	0.70	0.72	0.71	4389
accur	23.61/			0.71	8778
	-				
macro	avg	0.71	0.71	0.71	8778
weighted	avg	0.71	0.71	0.71	8778
test set					
		precision	recall	f1-score	support
	0	0.96	0.68	0.80	1848
	_				
	1	0.22	0.78	0.35	218
accur	acv			0.69	2066
		0.50	0.72		
macro		0.59	0.73	0.57	2066
weighted	avg	0.88	0.69	0.75	2066

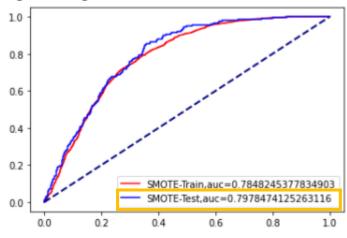
Logistic regression model - Oversampling RFM - AUC



Logistic Regression SMOTE - CLV

logistic training	_	ession model	- SMOTE	CLV	
		precision	recall	f1-score	support
	0	0.73	0.69	0.71	4389
	1	0.71	0.75	0.73	4389
accur	acy			0.72	8778
macro	avg	0.72	0.72	0.72	8778
weighted	avg	0.72	0.72	0.72	8778
test set					
		precision	recall	f1-score	support
	0	0.96	0.68	0.80	1848
	1	0.22	0.78	0.35	218
accur	acy			0.69	2066
macro	avg	0.59	0.73	0.57	2066
weighted	avg	0.88	0.69	0.75	2066

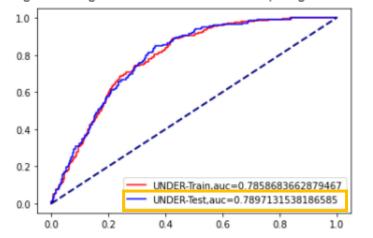




Logistic Regression Undersampling - CLV

regr set	ession model	- Unders	ampling CLV	
	precision	recall	f1-score	support
0	0.73	0.70	0.71	429
1	0.71	0.74	0.73	429
racy			0.72	858
avg	0.72	0.72	0.72	858
avg	0.72	0.72	0.72	858
	precision	recall	f1-score	support
0	0.96	0.69	0.80	1848
1	0.22	0.76	0.34	218
acy			0.70	2066
avg	0.59	0.72	0.57	2066
avg	0.88	0.70	0.75	2066
	o 1 acy avg avg 1 acy avg	precision 0 0.73 1 0.71 Pacy avg 0.72 avg 0.72 precision 0 0.96 1 0.22 Pacy avg 0.59	precision recall 0 0.73 0.70 1 0.71 0.74 Pacy avg 0.72 0.72 avg 0.72 0.72 precision recall 0 0.96 0.69 1 0.22 0.76 Pacy avg 0.59 0.72	precision recall f1-score 0 0.73 0.70 0.71 1 0.71 0.74 0.73 Pacy 0.72 0.72 avg 0.72 0.72 0.72 precision recall f1-score 0 0.96 0.69 0.80 1 0.22 0.76 0.34 Pacy avg 0.59 0.72 0.72

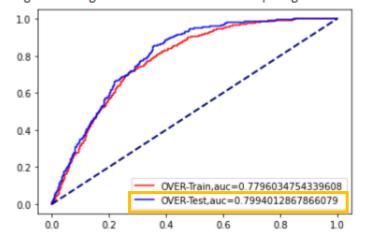
Logistic regression model - Undersampling CLV - AUC



Logistic Regression Oversampling - CLV

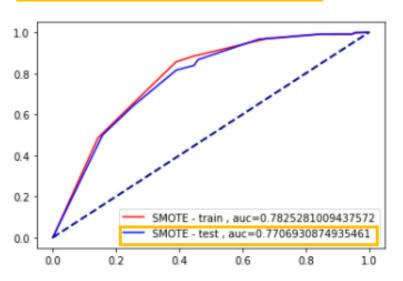
logistic regr	ession model	- Oversa	mpling CLV	,
er danang bee	precision	recall	f1-score	support
0	0.73	0.69	0.71	4389
1	0.70	0.74	0.72	4389
accuracy			0.71	8778
macro avg	0.71	0.71	0.71	8778
weighted avg	0.71	0.71	0.71	8778
test set				
	precision	recall	f1-score	support
0	0.96	0.68	0.80	1848
1	0.22	0.78	0.35	218
accuracy			0.69	2066
macro ave	0.59	0.73	0.57	2066
weighted avg	0.88	0.69	0.75	2066

Logistic regression model - Oversampling CLV - AUC



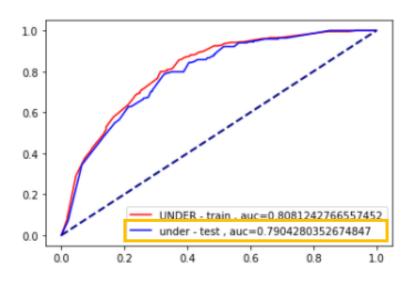
XGBoost SMOTE - RFM

training set				
	precision	recall	f1-score	support
0	0.81	0.61	0.70	4389
1	0.69	0.86	0.76	4389
accuracy			0.73	8778
macro avg	0.75	0.73	0.73	8778
weighted avg	0.75	0.73	0.73	8778
test set				
	precision	recall	f1-score	support
0	0.97	0.61	0.75	1848
1	0.20	0.82	0.32	218
accuracy			0.63	2066
macro avg	0.58	0.71	0.53	2066
weighted avg	0.88	0.63	0.70	2066



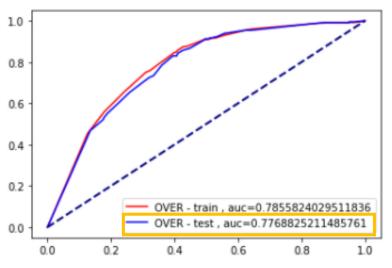
XGBoost Undersampling - RFM

training set				
	precision	recall	f1-score	support
0	0.78	0.65	0.71	429
1	0.70	0.81	0.75	429
accuracy			0.73	858
macro avg	0.74	0.73	0.73	858
weighted avg	0.74	0.73	0.73	858
test set				
	precision	recall	f1-score	support
0	0.96	0.61	0.75	1848
1	0.20	0.80	0.31	218
accuracy			0.63	2066
macro avg	0.58	0.71	0.53	2066
weighted avg	0.88	0.63	0.70	2066



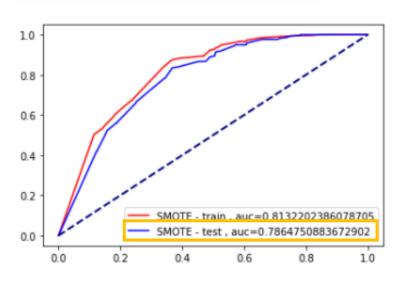
XGBoost Oversampling - RFM

				training set
support	f1-score	recall	precision	
4389	0.69	0.60	0.79	0
4389	0.75	0.84	0.68	1
8778	0.72			accuracy
8778	0.72	0.72	0.74	macro avg
8778	0.72	0.72	0.74	weighted avg
				test set
support	f1-score	recall	precision	
1848	0.74	0.60	0.97	0
218	0.32	0.83	0.20	1
2066	0.63			accuracy
2066	0.53	0.72	0.58	macro avg
2066	0.70	0.63	0.89	weighted avg



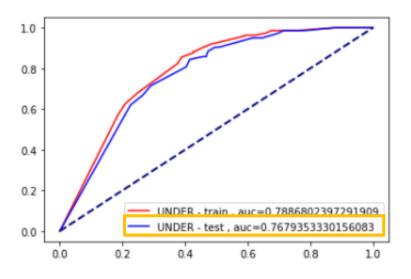
Logistic Regression SMOTE - CLV

training set				
	precision	recall	f1-score	support
0	0.84	0.60	0.70	4389
1	0.69	0.89	0.77	4389
accuracy			0.74	8778
macro avg	0.76	0.74	0.74	8778
weighted avg	0.76	0.74	0.74	8778
test set				
	precision	recall	f1-score	support
0	0.97	0.59	0.73	1848
1	0.20	0.85	0.32	218
accuracy			0.62	2066
macro avg	0.58	0.72	0.53	2066
weighted avg	0.89	0.62	0.69	2066



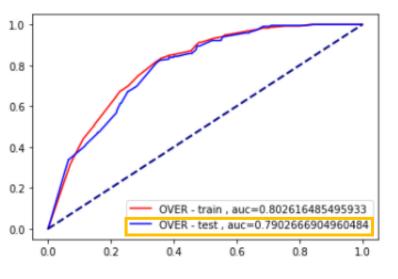
Logistic Regression Undersampling - CLV

0.81		f1-score	support
	0.61		
0.60		0.70	429
0.69	0.86	0.76	429
		0.73	858
0.75	0.73	0.73	858
0.75	0.73	0.73	858
recision	recall	f1-score	support
0.97	0.58	0.73	1848
0.19	0.84	0.31	218
		0.61	2066
0.58	0.71	0.52	2066
0.89	0.61	0.69	2066
	0.69 0.75 0.75 precision 0.97 0.19	0.69 0.86 0.75 0.73 0.75 0.73 0recision recall 0.97 0.58 0.19 0.84	0.69 0.86 0.76 0.73 0.75 0.73 0.73 0.75 0.73 0.73 0recision recall f1-score 0.97 0.58 0.73 0.19 0.84 0.31 0.61 0.58 0.71 0.52



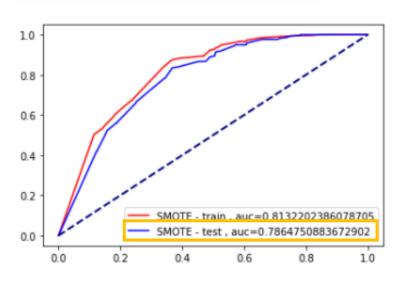
Logistic Regression Oversampling - CLV

training set				
	precision	recall	f1-score	support
0	0.81	0.56	0.66	4389
1	0.66	0.87	0.75	4389
accuracy			0.71	8778
macro avg	0.73	0.71	0.70	8778
weighted avg	0.73	0.71	0.70	8778
test set				
	precision	recall	f1-score	support
0	0.97	0.55	0.70	1848
1	0.18	0.86	0.30	218
accuracy			0.58	2066
macro avg	0.58	0.70	0.50	2066
weighted avg	0.89	0.58	0.66	2066



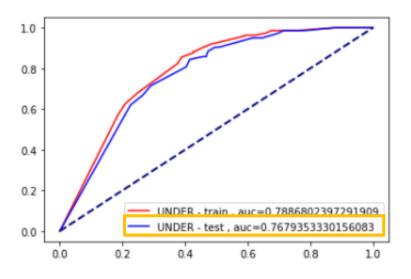
Logistic Regression SMOTE - CLV

training set				
	precision	recall	f1-score	support
0	0.84	0.60	0.70	4389
1	0.69	0.89	0.77	4389
accuracy			0.74	8778
macro avg	0.76	0.74	0.74	8778
weighted avg	0.76	0.74	0.74	8778
test set				
	precision	recall	f1-score	support
0	0.97	0.59	0.73	1848
1	0.20	0.85	0.32	218
accuracy			0.62	2066
macro avg	0.58	0.72	0.53	2066
weighted avg	0.89	0.62	0.69	2066



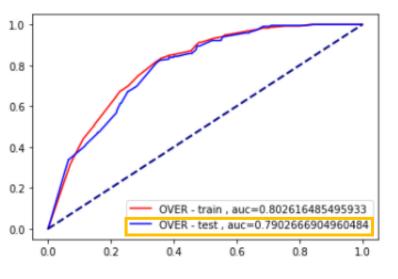
Logistic Regression Undersampling - CLV

0.81		f1-score	support
	0.61		
0.60		0.70	429
0.69	0.86	0.76	429
		0.73	858
0.75	0.73	0.73	858
0.75	0.73	0.73	858
recision	recall	f1-score	support
0.97	0.58	0.73	1848
0.19	0.84	0.31	218
		0.61	2066
0.58	0.71	0.52	2066
0.89	0.61	0.69	2066
	0.69 0.75 0.75 precision 0.97 0.19	0.69 0.86 0.75 0.73 0.75 0.73 0recision recall 0.97 0.58 0.19 0.84	0.69 0.86 0.76 0.73 0.75 0.73 0.73 0.75 0.73 0.73 0recision recall f1-score 0.97 0.58 0.73 0.19 0.84 0.31 0.61 0.58 0.71 0.52



Logistic Regression Oversampling - CLV

training set				
	precision	recall	f1-score	support
0	0.81	0.56	0.66	4389
1	0.66	0.87	0.75	4389
accuracy			0.71	8778
macro avg	0.73	0.71	0.70	8778
weighted avg	0.73	0.71	0.70	8778
test set				
	precision	recall	f1-score	support
0	0.97	0.55	0.70	1848
1	0.18	0.86	0.30	218
accuracy			0.58	2066
macro avg	0.58	0.70	0.50	2066
weighted avg	0.89	0.58	0.66	2066



Tuning - RFM

```
Best AUC Score: 0.7626080744558595
Accuracy: 0.6892545982575025
[[1265 583]
 [ 59 159]]
test set
                         recall f1-score support
             precision
                  0.96
                            0.68
                                     0.80
                                               1848
                  0.21
                           0.73
                                     0.33
                                                218
                                     0.69
                                               2066
    accuracy
   macro avg
                  0.58
                            0.71
                                     0.56
                                               2066
```

The result wasn't better than other models So, I decided not to do other steps.

0.69

0.75

2066

0.88

weighted avg

Summary

- Dropped the timeframe before 2013-03-01 provided the better results so, we could imply that the effectiveness of models depended on appropriate timeframe.
- Evaluated with by 'accuracy' was not enough, calculated ROC curve would be a good choice to figure out the exact best result.
- Based on The accuracy and AUC Score, the best model was presented by Logistic Regression Oversampling – CLV at Accuracy = 0.69 (better than original model at 0.61) at AUC = 0.799 (0.666)