

# Results from Original File

## Logistic Regression with SMOTE

logistic regression model - SMOTE RFM

training set

	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
weighted avg	0.85	0.61	0.68	1500

logistic regression model - SMOTE CLV

training set

	precision	recall	f1-score	support
0	0.59	0.62	0.61	3073
1	0.60	0.58	0.59	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	0.61	0.74	1330
1	0.17	0.62	0.27	170
accuracy			0.61	1500
macro avg	0.55	0.62	0.50	1500
weighted avg	0.84	0.61	0.69	1500

## XGBoost with SMOTE

XGBoost model - SMOTE RFM

[0] validation\_0-auc:0.577145  
Will train until validation\_0-auc hasn't improved in 5 rounds.

[1] validation\_0-auc:0.666252  
[2] validation\_0-auc:0.64722  
[3] validation\_0-auc:0.643182  
[4] validation\_0-auc:0.651329  
[5] validation\_0-auc:0.666252  
[6] validation\_0-auc:0.664854  
Stopping. Best iteration:  
[1] validation\_0-auc:0.666252

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

**Tuning  
CLV**

Best AUC Score: 0.671839531581116  
Accuracy: 0.48733333333333334  
[[594 736]  
[ 33 137]]

	precision	recall	f1-score	support
0	0.95	0.45	0.61	1330
1	0.16	0.81	0.26	170
accuracy			0.49	1500
macro avg	0.55	0.63	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.

[1] validation\_0-auc:0.670705  
[2] validation\_0-auc:0.68444  
[3] validation\_0-auc:0.684042  
[4] validation\_0-auc:0.687351  
[5] validation\_0-auc:0.690741  
[6] validation\_0-auc:0.692315  
[7] validation\_0-auc:0.69236  
[8] validation\_0-auc:0.686809  
[9] validation\_0-auc:0.68937  
[10] validation\_0-auc:0.690741  
[11] validation\_0-auc:0.69335  
[12] validation\_0-auc:0.693412  
[13] validation\_0-auc:0.693538  
[14] validation\_0-auc:0.694135  
[15] validation\_0-auc:0.694847  
[16] validation\_0-auc:0.69479  
[17] validation\_0-auc:0.694034  
[18] validation\_0-auc:0.693919  
[19] validation\_0-auc:0.694715  
[20] validation\_0-auc:0.694622  
Stopping. Best iteration:  
[15] validation\_0-auc:0.694847

training set

	precision	recall	f1-score	support
0	0.68	0.56	0.62	3073
1	0.63	0.74	0.68	3073
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
0	0.94	0.57	0.71	1330
1	0.18	0.72	0.28	170
accuracy			0.58	1500
macro avg	0.56	0.64	0.40	1500
weighted avg	0.85	0.58	0.66	1500

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

**Let's start...**

# Results from My File

## Logistic Regression SMOTE - RFM

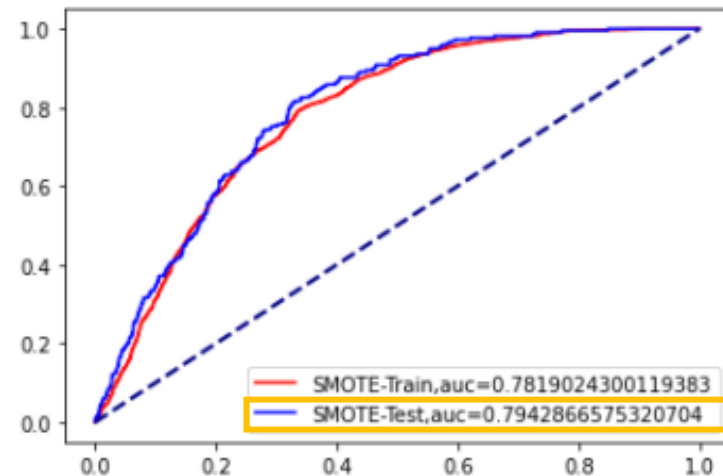
logistic regression model - SMOTE RFM  
training set

	precision	recall	f1-score	support
0	0.72	0.69	0.71	4389
1	0.70	0.72	0.71	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	0.77	0.35	218
accuracy			0.69	2066
macro avg	0.59	0.73	0.57	2066
weighted avg	0.88	0.69	0.75	2066

Logistic regression model - SMOTE RFM - AUC



## Logistic Regression Undersampling - RFM

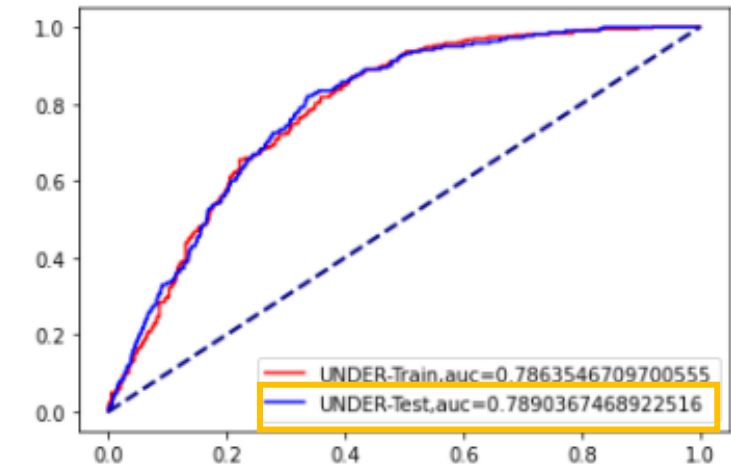
logistic regression model - Undersampling RFM  
training set

	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

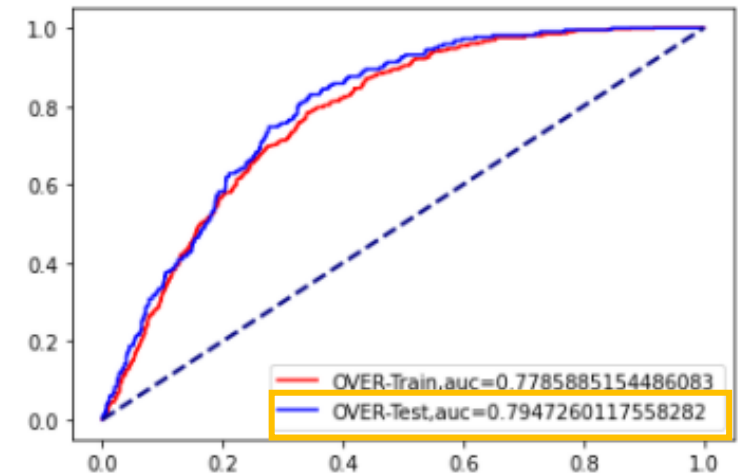
logistic regression model - Oversampling RFM  
training set

	precision	recall	f1-score	support
0	0.71	0.69	0.70	4389
1	0.70	0.72	0.71	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 avg	0.59	0.73	0.57	2066
weighted avg	0.88	0.69	0.75	2066

Logistic regression model - Oversampling RFM - AUC



# Results from My File

## Logistic Regression SMOTE - CLV

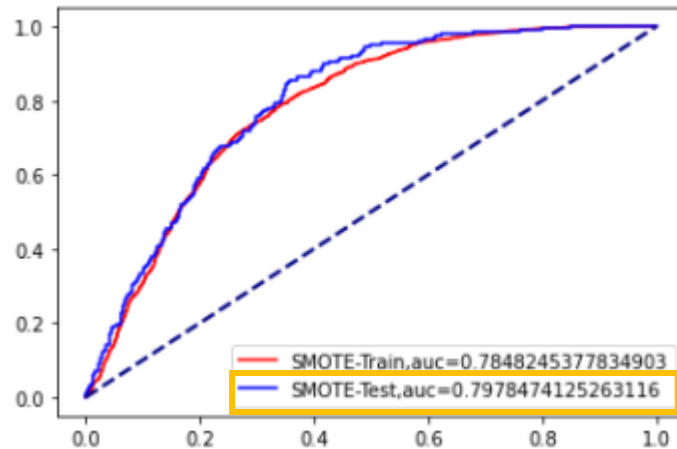
logistic regression model - SMOTE CLV  
training set

	precision	recall	f1-score	support
0	0.73	0.69	0.71	4389
1	0.71	0.75	0.73	4389
accuracy			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
accuracy			0.69	2066
macro avg	0.59	0.73	0.57	2066
weighted avg	0.88	0.69	0.75	2066

Logistic regression model - SMOTE clv - AUC



## Logistic Regression Undersampling - CLV

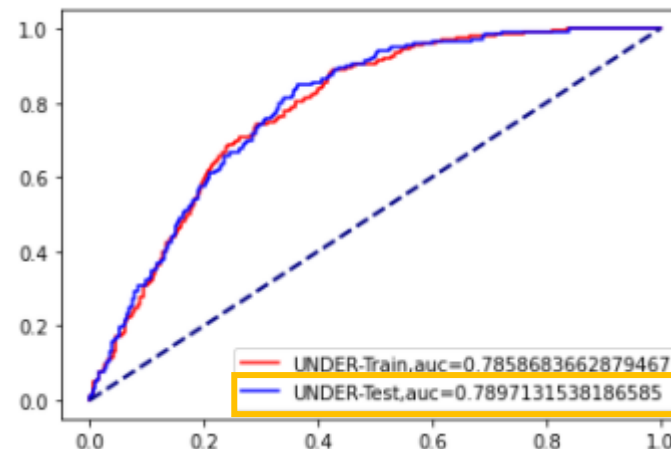
logistic regression model - Undersampling CLV  
training set

	precision	recall	f1-score	support
0	0.73	0.70	0.71	429
1	0.71	0.74	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	0.69	0.80	1848
1	0.22	0.76	0.34	218
accuracy			0.70	2066
macro avg	0.59	0.72	0.57	2066
weighted avg	0.88	0.70	0.75	2066

Logistic regression model - Undersampling CLV - AUC



## Logistic Regression Oversampling - CLV

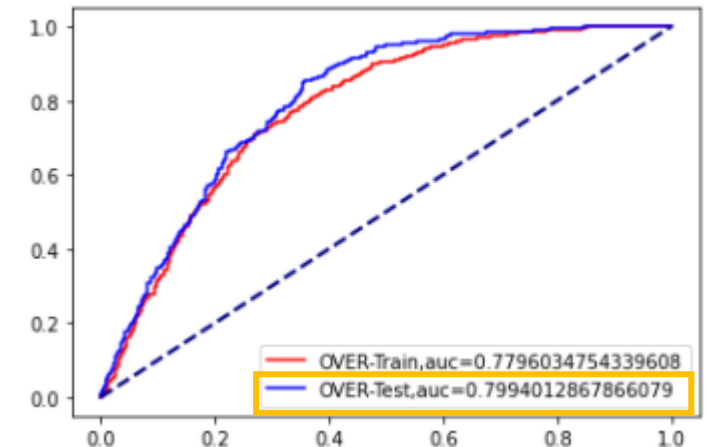
logistic regression model - Oversampling CLV  
training set

	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 avg	0.59	0.73	0.57	2066
weighted avg	0.88	0.69	0.75	2066

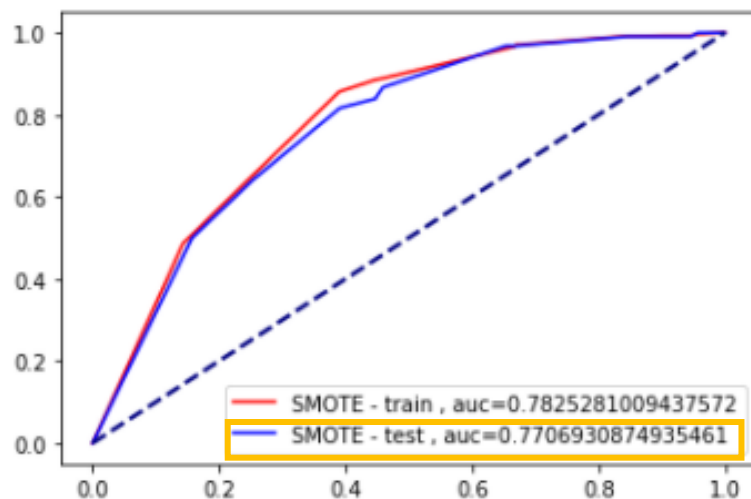
Logistic regression model - Oversampling CLV - AUC



# Results from My File

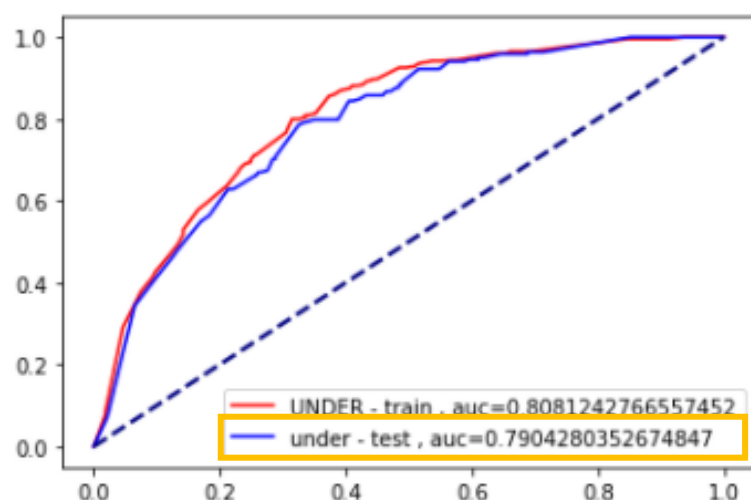
## XGBoost SMOTE - RFM

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



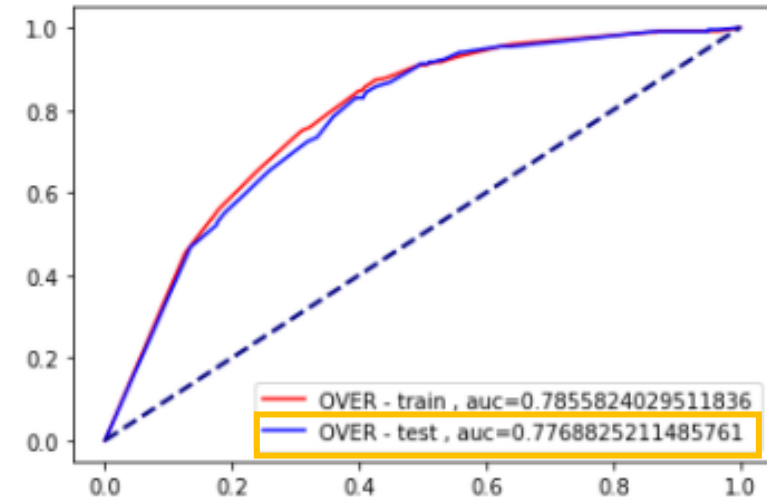
## XGBoost Undersampling - RFM

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

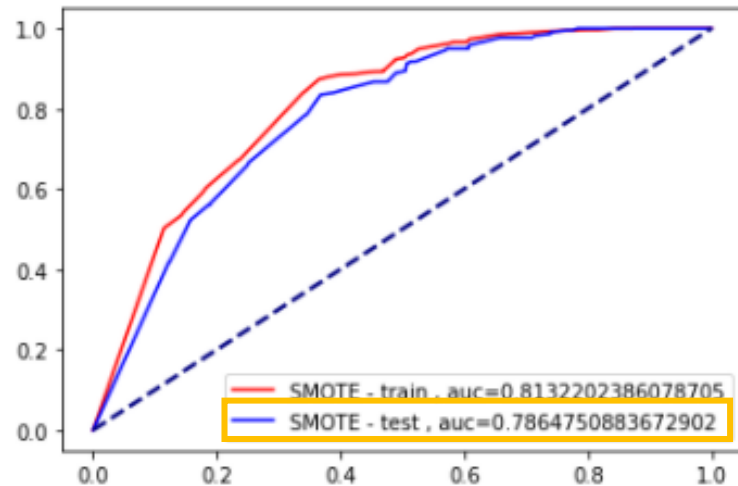
training set				
	precision	recall	f1-score	support
0	0.79	0.60	0.69	4389
1	0.68	0.84	0.75	4389
accuracy			0.72	8778
macro avg	0.74	0.72	0.72	8778
weighted avg	0.74	0.72	0.72	8778
test set				
	precision	recall	f1-score	support
0	0.97	0.60	0.74	1848
1	0.20	0.83	0.32	218
accuracy			0.63	2066
macro avg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



# Results from My File

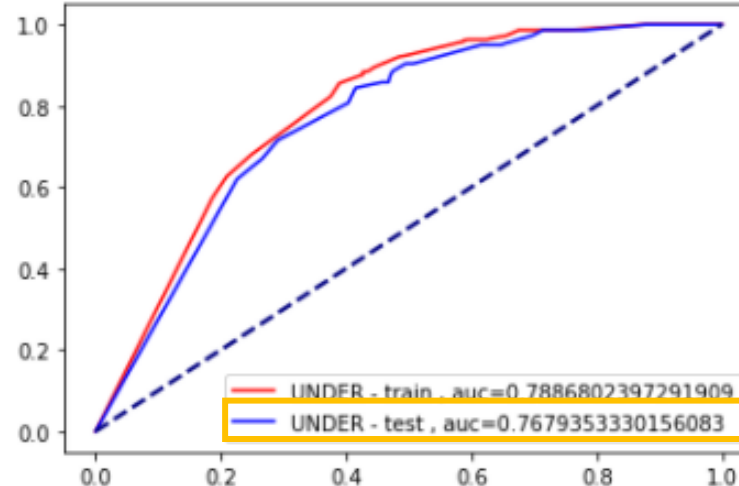
## Logistic Regression SMOTE - CLV

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



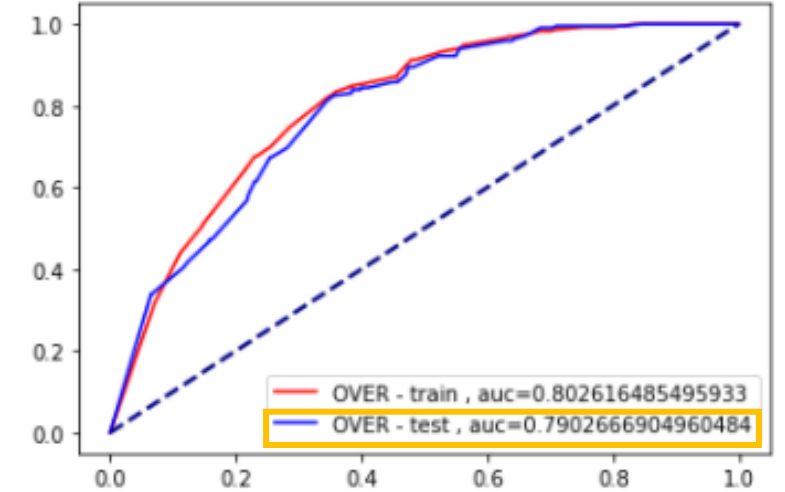
## Logistic Regression Undersampling - CLV

training set				
	precision	recall	f1-score	support
0	0.81	0.61	0.70	429
1	0.69	0.86	0.76	429
accuracy			0.73	858
macro avg	0.75	0.73	0.73	858
weighted avg	0.75	0.73	0.73	858
test set				
	precision	recall	f1-score	support
0	0.97	0.58	0.73	1848
1	0.19	0.84	0.31	218
accuracy			0.61	2066
macro avg	0.58	0.71	0.52	2066
weighted avg	0.89	0.61	0.69	2066



## 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

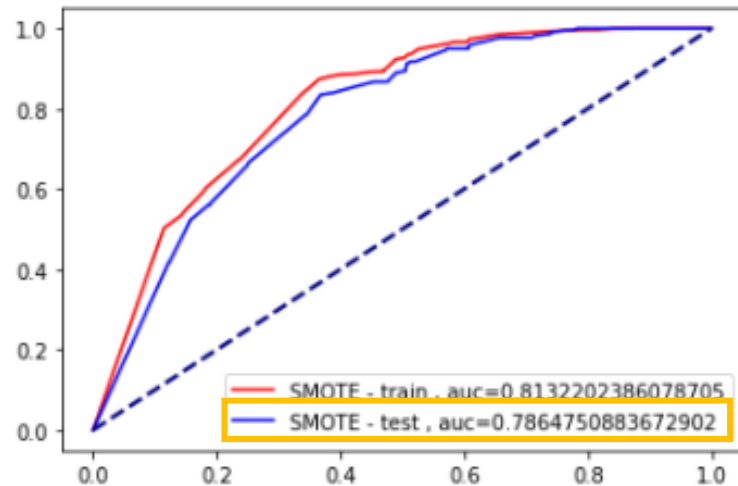




# Results from My File

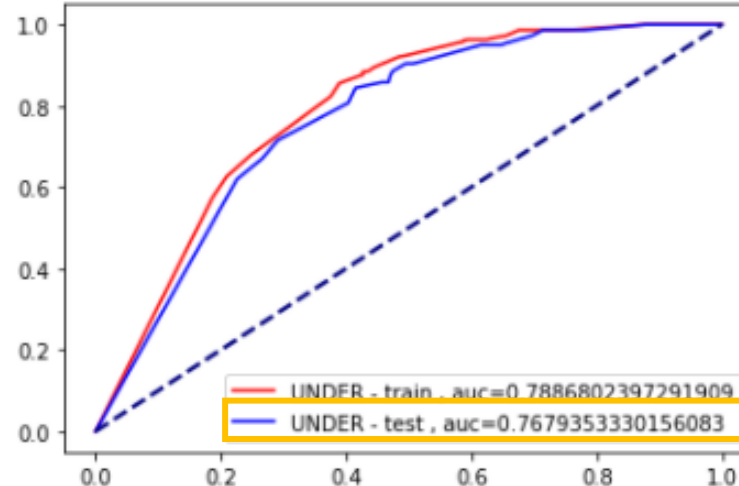
## Logistic Regression SMOTE - CLV

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



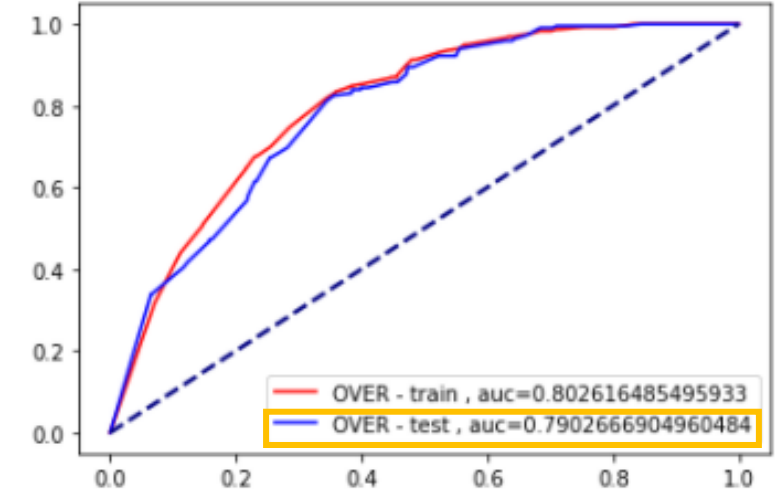
## Logistic Regression Undersampling - CLV

training set				
	precision	recall	f1-score	support
0	0.81	0.61	0.70	429
1	0.69	0.86	0.76	429
accuracy			0.73	858
macro avg	0.75	0.73	0.73	858
weighted avg	0.75	0.73	0.73	858
test set				
	precision	recall	f1-score	support
0	0.97	0.58	0.73	1848
1	0.19	0.84	0.31	218
accuracy			0.61	2066
macro avg	0.58	0.71	0.52	2066
weighted avg	0.89	0.61	0.69	2066



## 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



# Results from My File

## Tuning - RFM

Best AUC Score: 0.7626080744558595  
Accuracy: 0.6892545982575025

[[1265 583]  
[ 59 159]]

test set

	precision	recall	f1-score	support
0	0.96	0.68	0.80	1848
1	0.21	0.73	0.33	218
accuracy			0.69	2066
macro avg	0.58	0.71	0.56	2066
weighted avg	0.88	0.69	0.75	2066

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

## 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)