LOGISTIC REGRESSION MODEL - SMOTE RFM

2011 - 2015

logistic regression model - SMOTE RFM
training set

	precision	recall	f1-score	support
0	0.68	0.63	0.65	1883
1	0.65	0.70	0.67	1883
accuracy			0.66	3766
macro avg	0.66	0.66	0.66	3766
weighted avg	0.66	0.66	0.66	3766

test set

	precision	recall	f1-score	support
0	0.95	0.63	0.75	4354
1	0.16	0.68	0.26	465
accuracy			0.63	4819
macro avg	0.55	0.65	0.51	4819
weighted avg	0.87	0.63	0.71	4819

2013 - 2015

logistic regression model - SMOTE RFM
training set

	precision	recall	f1-score	support
0	0.72	0.69	0.70	1883
1	0.70	0.73	0.71	1883
accuracy			0.71	3766
macro avg	0.71	0.71	0.71	3766
weighted avg	0.71	0.71	0.71	3766

test set

	precision	recall	f1-score	support
0	0.96	0.68	0.79	4354
1	0.19	0.73	0.31	465
accuracy			0.68	4819
macro avg	0.58	0.70	0.55	4819
weighted avg	0.89	0.68	0.75	4819

2014 - 2015

logistic regression model - SMOTE RFM
training set

	precision	recall	f1-score	support
0	0.70	0.67	0.68	1871
1	0.68	0.72	0.70	1871
accuracy			0.69	3742
macro avg	0.69	0.69	0.69	3742
weighted avg	0.69	0.69	0.69	3742

test set

	precision	recall	f1-score	support
0	0.95	0.67	0.78	4328
1	0.18	0.67	0.28	465
accuracy			0.67	4793
macro avg	0.56	0.67	0.53	4793
weighted avg	0.87	0.67	0.74	4793

LOGISTIC REGRESSION MODEL - SMOTE CLV

2011 - 2015

logistic regression model - SMOTE CLV
training set

	precision	recall	f1-score	support
0	0.69	0.63	0.66	1883
1	0.66	0.72	0.69	1883
accuracy			0.67	3766
macro avg	0.67	0.67	0.67	3766
weighted avg	0.67	0.67	0.67	3766

test set

	precision	recall	f1-score	support
0	0.95	0.63	0.76	4354
1	0.16	0.66	0.26	465
accuracy			0.63	4819
macro avg	0.55	0.65	0.51	4819
weighted avg	0.87	0.63	0.71	4819

2013 - 2015

logistic regression model - SMOTE CLV
training set

	precision	recall	f1-score	support
0	0.72	0.70	0.71	1883
1	0.71	0.72	0.71	1883
accuracy			0.71	3766
macro avg	0.71	0.71	0.71	3766
weighted avg	0.71	0.71	0.71	3766

test set

	precision	recall	f1-score	support
0	0.96	0.69	0.80	4354
1	0.20	0.73	0.31	465
accuracy			0.69	4819
macro avg	0.58	0.71	0.56	4819
weighted avg	0.89	0.69	0.76	4819

2014 - 2015

logistic regression model - SMOTE CLV
training set

	precision	recall	f1-score	support
0	0.72	0.67	0.69	1871
1	0.69	0.74	0.71	1871
accuracy			0.70	3742
macro avg	0.70	0.70	0.70	3742
weighted avg	0.70	0.70	0.70	3742

test set

	precision	recall	f1-score	support
0	0.95	0.66	0.78	4328
1	0.18	0.68	0.28	465
accuracy			0.66	4793
macro avg	0.56	0.67	0.53	4793
weighted avg	0.88	0.66	0.73	4793

XGBOOST MODEL - SMOTE RFM

2011 - 2015

```
XGBoost model - SMOTE RFM
[0]     validation_0-auc:0.582917
Will train until validation_0-auc hasn't improved in 5 rounds.
[1]     validation_0-auc:0.678763
[2]     validation_0-auc:0.659691
[3]     validation_0-auc:0.648997
[4]     validation_0-auc:0.669309
[5]     validation_0-auc:0.677161
[6]     validation_0-auc:0.67469
Stopping. Best iteration:
[1]     validation_0-auc:0.678763
```

training set

	precision	recall	f1-score	support
0	0.83	0.76	0.80	1883
1	0.78	0.85	0.81	1883
accuracy			0.80	3766
macro avg	0.81	0.80	0.80	3766
weighted avg	0.81	0.80	0.80	3766

test set

	precision	recall	f1-score	support
0	0.94	0.71	0.81	4354
1	0.17	0.56	0.26	465
accuracy			0.69	4819
macro avg	0.55	0.63	0.53	4819
weighted avg	0.86	0.69	0.75	4819

2013 - 2015

```
[8]     validation_0-auc:0.710467
[9]     validation_0-auc:0.716909
[10]    validation_0-auc:0.718758
[11]    validation_0-auc:0.720424
[12]    validation_0-auc:0.715269
[13]    validation_0-auc:0.71471
[14]    validation_0-auc:0.710705
[15]    validation_0-auc:0.715432
[16]    validation_0-auc:0.717446
Stopping. Best iteration:
[11]    validation_0-auc:0.720424
```

training set

	precision	recall	f1-score	support
0	0.94	0.93	0.93	1883
1	0.93	0.94	0.93	1883
accuracy			0.93	3766
macro avg	0.93	0.93	0.93	3766
weighted avg	0.93	0.93	0.93	3766

test set

	precision	recall	f1-score	support
0	0.93	0.85	0.89	4354
1	0.20	0.35	0.26	465
accuracy			0.80	4819
macro avg	0.57	0.60	0.57	4819
weighted avg	0.86	0.80	0.83	4819

2014 - 2015

```
[14]    validation_0-auc:0.701971
[15]    validation_0-auc:0.703378
[16]    validation_0-auc:0.704289
[17]    validation_0-auc:0.702937
Stopping. Best iteration:
[12]    validation_0-auc:0.705624
```

training set

	precision	recall	f1-score	support
0	0.94	0.94	0.94	1871
1	0.94	0.94	0.94	1871
accuracy			0.94	3742
macro avg	0.94	0.94	0.94	3742
weighted avg	0.94	0.94	0.94	3742

test set

	precision	recall	f1-score	support
0	0.92	0.86	0.89	4328
1	0.20	0.32	0.25	465
accuracy			0.81	4793
macro avg	0.56	0.59	0.57	4793
weighted avg	0.85	0.81	0.83	4793

XGBOOST MODEL - SMOTE CLV'

2011 - 2015

```
XGBoost model - SMOTE CLV
[0]     validation_0-auc:0.641503
Will train until validation_0-auc hasn't improved in 5 rounds.
[1]     validation_0-auc:0.650185
[2]     validation_0-auc:0.659211
[3]     validation_0-auc:0.662832
[4]     validation_0-auc:0.657371
[5]     validation_0-auc:0.661992
[6]     validation_0-auc:0.653851
[7]     validation_0-auc:0.652762
[8]     validation_0-auc:0.649685
Stopping. Best iteration:
[3]     validation_0-auc:0.662832
```

training set

	precision	recall	f1-score	support
0	0.97	0.92	0.94	1883
1	0.92	0.97	0.95	1883
accuracy			0.95	3766
macro avg	0.95	0.95	0.95	3766
weighted avg	0.95	0.95	0.95	3766

test set

	precision	recall	f1-score	support
0	0.92	0.84	0.88	4354
1	0.16	0.29	0.21	465
accuracy			0.79	4819
macro avg	0.54	0.56	0.54	4819
weighted avg	0.84	0.79	0.81	4819

2013 - 2015

```
[3]     validation_0-auc:0.718795
[4]     validation_0-auc:0.721801
[5]     validation_0-auc:0.721163
[6]     validation_0-auc:0.71842
[7]     validation_0-auc:0.715367
[8]     validation_0-auc:0.709786
[9]     validation_0-auc:0.715573
Stopping. Best iteration:
[4]     validation_0-auc:0.721801
```

training set

	precision	recall	f1-score	support
0	0.97	0.96	0.96	1883
1	0.96	0.97	0.96	1883
accuracy			0.96	3766
macro avg	0.96	0.96	0.96	3766
weighted avg	0.96	0.96	0.96	3766

test set

	precision	recall	f1-score	support
0	0.92	0.90	0.91	4354
1	0.22	0.27	0.24	465
accuracy			0.84	4819
macro avg	0.57	0.58	0.58	4819
weighted avg	0.85	0.84	0.85	4819

2014 - 2015

```
[9]     validation_0-auc:0.690573
Stopping. Best iteration:
[4]     validation_0-auc:0.690764
```

training set

	precision	recall	f1-score	support
0	0.96	0.94	0.95	1871
1	0.94	0.96	0.95	1871
accuracy			0.95	3742
macro avg	0.95	0.95	0.95	3742
weighted avg	0.95	0.95	0.95	3742

test set

	precision	recall	f1-score	support
0	0.92	0.87	0.89	4328
1	0.18	0.26	0.21	465
accuracy			0.81	4793
macro avg	0.55	0.57	0.55	4793
weighted avg	0.84	0.81	0.83	4793

AFTER PIPELINE & HYPERPARAMETER TUNING - CLV

2011 - 2015

```
Best AUC Score: 0.7328937313118873
Accuracy: 0.6526250259389915
[[2853 1501]
 [ 173  292]]
test set
              precision    recall  f1-score   support

           0       0.94      0.66      0.77      4354
           1       0.16      0.63      0.26       465

    accuracy                           0.65      4819
   macro avg       0.55      0.64      0.52      4819
weighted avg       0.87      0.65      0.72      4819
```

2013 - 2015

```
Best AUC Score: 0.7722449050026589
Accuracy: 0.6955799958497614
[[3027 1327]
 [ 140  325]]
test set
              precision    recall  f1-score   support

           0       0.96      0.70      0.80      4354
           1       0.20      0.70      0.31       465

    accuracy                           0.70      4819
   macro avg       0.58      0.70      0.56      4819
weighted avg       0.88      0.70      0.76      4819
```

2014 - 2015

```
Best AUC Score: 0.7720679007189519
Accuracy: 0.6951804715209681
[[3033 1295]
 [ 166  299]]
test set
              precision    recall  f1-score   support

           0       0.95      0.70      0.81      4328
           1       0.19      0.64      0.29       465

    accuracy                           0.70      4793
   macro avg       0.57      0.67      0.55      4793
weighted avg       0.87      0.70      0.76      4793
```

Summary:

- Logistic regression model accuracy is affected by different time duration. Therefore, this creates difficulties in finding exactly appropriated duration in which the model will work best. To conclude, this is not the most appropriate model to be used in the case where time duration effects model effectiveness in its nature.

- Timeframe should be relevant. Shouldn't be too further away or aged

- When the appropriate timeframe is selected, you will see that no matter the duration is – result should be at the same effectiveness level (Recommend Pipeline parameter tuning or XGBoost SMOTE)

- However, in case of selecting XGBoost SMOTE – we need to also consider on features used. The selected feature should not be affected by duration and should have enough dimensions which brings in stability and accuracy in itself. (Recommend RFM rather than CLV)