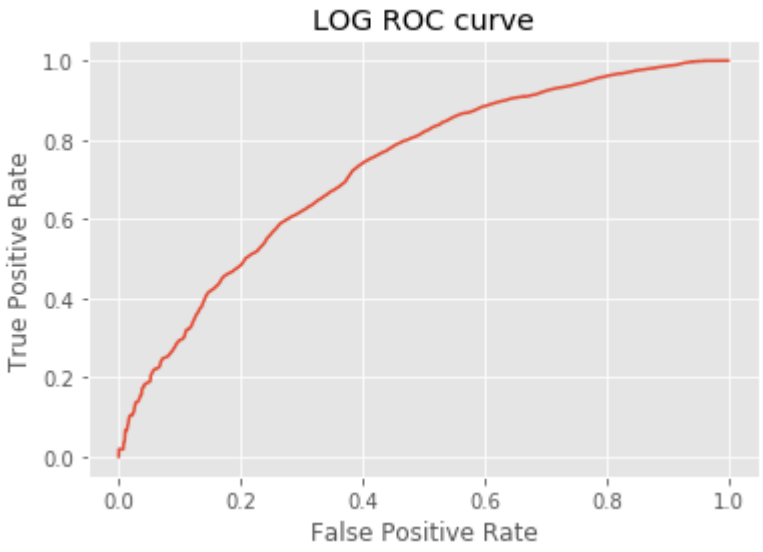
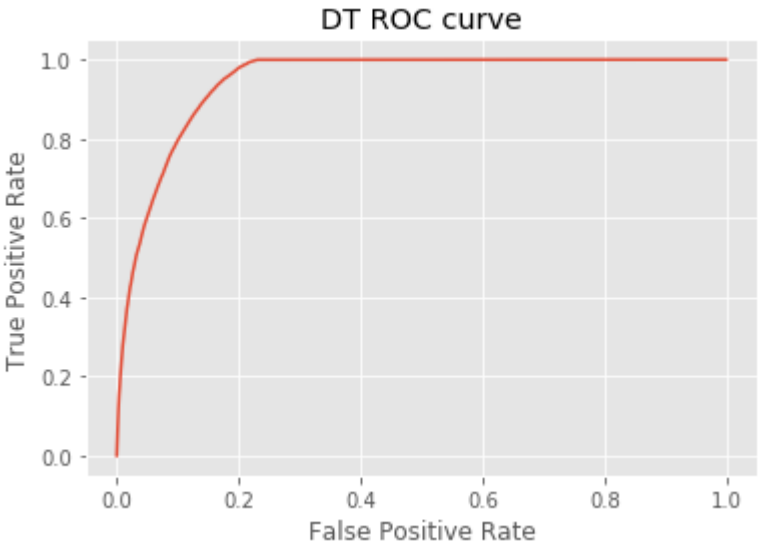
```
Python 3.6.7 |Anaconda, Inc.| (default, Oct 28 2018, 19:44:12) [MSC v.1915
64 bit (AMD64)]
Type "copyright", "credits" or "license" for more information.

IPython 6.5.0 -- An enhanced Interactive Python.

In [1]: runfile('E:/University_of_Florida/11th term/Statistical Machine
Learning/Project/purchase-prediction/purchase-class-models-gridCV.py',
wdir='E:/University_of_Florida/11th term/Statistical Machine
Learning/Project/purchase-prediction')
Importing libraries.......
Exporting data.......
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 550068 entries, 0 to 550067
Data columns (total 12 columns):
User_ID                     550068 non-null int64
Product_ID                  550068 non-null object
Gender                      550068 non-null object
Age                         550068 non-null object
Occupation                  550068 non-null int64
City_Category               550068 non-null object
Stay_In_Current_City_Years  550068 non-null object
Marital_Status              550068 non-null int64
Product_Category_1          550068 non-null int64
Product_Category_2          376430 non-null float64
Product_Category_3          166821 non-null float64
Purchase                    550068 non-null int64
dtypes: float64(2), int64(5), object(5)
memory usage: 50.4+ MB
Data Preprocessing.......
Classification Model Training.......
Logistic Regression
Fitting 3 folds for each of 3 candidates, totalling 9 fits
[CV] max_iter=50 .....................................................
[Parallel(n_jobs=1)]: Using backend SequentialBackend with 1 concurrent
workers.
[CV] ..................................... max_iter=50, total=  10.5s
[CV] max_iter=50 .....................................................
[Parallel(n_jobs=1)]: Done   1 out of   1 | elapsed:   10.6s remaining:
0.0s
[CV] ..................................... max_iter=50, total=  10.0s
[CV] max_iter=50 .....................................................
[CV] ..................................... max_iter=50, total=  10.1s
[CV] max_iter=100 ....................................................
[CV] .................................... max_iter=100, total=  19.2s
[CV] max_iter=100 ....................................................
[CV] .................................... max_iter=100, total=  19.3s
[CV] max_iter=100 ....................................................
[CV] .................................... max_iter=100, total=  18.5s
[CV] max_iter=200 ....................................................
[CV] .................................... max_iter=200, total=  36.3s
[CV] max_iter=200 ....................................................
[CV] .................................... max_iter=200, total=  40.2s
[CV] max_iter=200 ....................................................
```

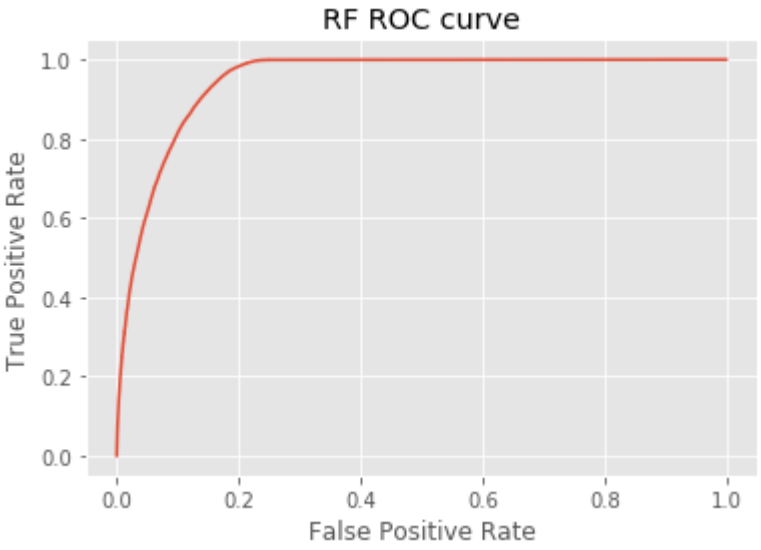
```
[CV] ..................................... max_iter=200, total=  37.4s
[Parallel(n_jobs=1)]: Done   9 out of   9 | elapsed:  3.4min finished
F1-score:  0.4918597639762254
Decision Tree Regression
Fitting 3 folds for each of 6 candidates, totalling 18 fits
[CV] max_depth=5, min_samples_leaf=50 ................................
[Parallel(n_jobs=1)]: Using backend SequentialBackend with 1 concurrent
workers.
[CV] .................. max_depth=5, min_samples_leaf=50, total=   0.4s
[CV] max_depth=5, min_samples_leaf=50 ................................
[Parallel(n_jobs=1)]: Done   1 out of   1 | elapsed:    0.5s remaining:
0.0s
[CV] .................. max_depth=5, min_samples_leaf=50, total=   0.4s
[CV] max_depth=5, min_samples_leaf=50 ................................
[CV] .................. max_depth=5, min_samples_leaf=50, total=   0.4s
[CV] max_depth=5, min_samples_leaf=100 ...............................
[CV] .................. max_depth=5, min_samples_leaf=100, total=   0.4s
[CV] max_depth=5, min_samples_leaf=100 ...............................
[CV] .................. max_depth=5, min_samples_leaf=100, total=   0.4s
[CV] max_depth=5, min_samples_leaf=100 ...............................
[CV] .................. max_depth=5, min_samples_leaf=100, total=   0.4s
[CV] max_depth=10, min_samples_leaf=50 ...............................
[CV] .................. max_depth=10, min_samples_leaf=50, total=   0.5s
[CV] max_depth=10, min_samples_leaf=50 ...............................
[CV] .................. max_depth=10, min_samples_leaf=50, total=   0.5s
[CV] max_depth=10, min_samples_leaf=50 ...............................
[CV] .................. max_depth=10, min_samples_leaf=50, total=   0.5s
[CV] max_depth=10, min_samples_leaf=100 ..............................
[CV] ................. max_depth=10, min_samples_leaf=100, total=   0.5s
[CV] max_depth=10, min_samples_leaf=100 ..............................
[CV] ................. max_depth=10, min_samples_leaf=100, total=   0.5s
[CV] max_depth=10, min_samples_leaf=100 ..............................
[CV] ................. max_depth=10, min_samples_leaf=100, total=   0.6s
[CV] max_depth=20, min_samples_leaf=50 ...............................
[CV] .................. max_depth=20, min_samples_leaf=50, total=   0.7s
[CV] max_depth=20, min_samples_leaf=50 ...............................
[CV] .................. max_depth=20, min_samples_leaf=50, total=   0.7s
[CV] max_depth=20, min_samples_leaf=50 ...............................
[CV] .................. max_depth=20, min_samples_leaf=50, total=   0.7s
[CV] max_depth=20, min_samples_leaf=100 ..............................
[CV] ................. max_depth=20, min_samples_leaf=100, total=   0.6s
[CV] max_depth=20, min_samples_leaf=100 ..............................
[CV] ................. max_depth=20, min_samples_leaf=100, total=   0.6s
[CV] max_depth=20, min_samples_leaf=100 ..............................
[CV] ................. max_depth=20, min_samples_leaf=100, total=   0.6s
[Parallel(n_jobs=1)]: Done  18 out of  18 | elapsed:   12.6s finished
F1-score:  0.7556259347233218
Random Forest Regression
Fitting 3 folds for each of 6 candidates, totalling 18 fits
[CV] max_depth=5, n_estimators=5 .....................................
[Parallel(n_jobs=1)]: Using backend SequentialBackend with 1 concurrent
workers.
[CV] ...................... max_depth=5, n_estimators=5, total=   1.0s
[CV] max_depth=5, n_estimators=5 .....................................
```

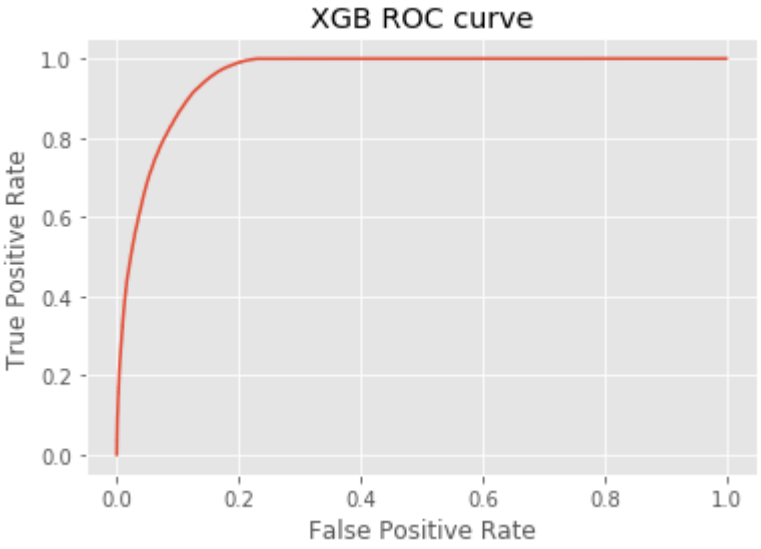
```
[Parallel(n_jobs=1)]: Done   1 out of   1 | elapsed:    1.2s remaining:
0.0s
[CV] ..................... max_depth=5, n_estimators=5, total=   1.0s
[CV] max_depth=5, n_estimators=5 .....................................
[CV] ..................... max_depth=5, n_estimators=5, total=   0.9s
[CV] max_depth=5, n_estimators=10 ....................................
[CV] ..................... max_depth=5, n_estimators=10, total=   1.5s
[CV] max_depth=5, n_estimators=10 ....................................
[CV] ..................... max_depth=5, n_estimators=10, total=   1.4s
[CV] max_depth=5, n_estimators=10 ....................................
[CV] ..................... max_depth=5, n_estimators=10, total=   1.6s
[CV] max_depth=10, n_estimators=5 ....................................
[CV] ..................... max_depth=10, n_estimators=5, total=   1.2s
[CV] max_depth=10, n_estimators=5 ....................................
[CV] ..................... max_depth=10, n_estimators=5, total=   1.2s
[CV] max_depth=10, n_estimators=5 ....................................
[CV] ..................... max_depth=10, n_estimators=5, total=   1.1s
[CV] max_depth=10, n_estimators=10 ...................................
[CV] ..................... max_depth=10, n_estimators=10, total=   2.0s
[CV] max_depth=10, n_estimators=10 ...................................
[CV] ..................... max_depth=10, n_estimators=10, total=   2.0s
[CV] max_depth=10, n_estimators=10 ...................................
[CV] ..................... max_depth=10, n_estimators=10, total=   1.9s
[CV] max_depth=20, n_estimators=5 ....................................
[CV] ..................... max_depth=20, n_estimators=5, total=   1.1s
[CV] max_depth=20, n_estimators=5 ....................................
[CV] ..................... max_depth=20, n_estimators=5, total=   1.2s
[CV] max_depth=20, n_estimators=5 ....................................
[CV] ..................... max_depth=20, n_estimators=5, total=   1.1s
[CV] max_depth=20, n_estimators=10 ...................................
[CV] ..................... max_depth=20, n_estimators=10, total=   2.1s
[CV] max_depth=20, n_estimators=10 ...................................
[CV] ..................... max_depth=20, n_estimators=10, total=   2.1s
[CV] max_depth=20, n_estimators=10 ...................................
[CV] ..................... max_depth=20, n_estimators=10, total=   2.2s
[Parallel(n_jobs=1)]: Done  18 out of  18 | elapsed:   32.8s finished
F1-score:  0.7674964772193519
Extreme Gradient Boosting Decision Tree Regression
Fitting 3 folds for each of 6 candidates, totalling 18 fits
[CV] max_depth=5, n_estimators=500 ...................................
[Parallel(n_jobs=1)]: Using backend SequentialBackend with 1 concurrent
workers.
[CV] ..................... max_depth=5, n_estimators=500, total=  47.6s
[CV] max_depth=5, n_estimators=500 ...................................
[Parallel(n_jobs=1)]: Done   1 out of   1 | elapsed:   48.9s remaining:
0.0s
[CV] ..................... max_depth=5, n_estimators=500, total=  52.4s
[CV] max_depth=5, n_estimators=500 ...................................
[CV] ..................... max_depth=5, n_estimators=500, total=  51.6s
[CV] max_depth=5, n_estimators=1000 ..................................
[CV] ..................... max_depth=5, n_estimators=1000, total= 1.7min
[CV] max_depth=5, n_estimators=1000 ..................................
[CV] ..................... max_depth=5, n_estimators=1000, total= 1.7min
[CV] max_depth=5, n_estimators=1000 ..................................
```

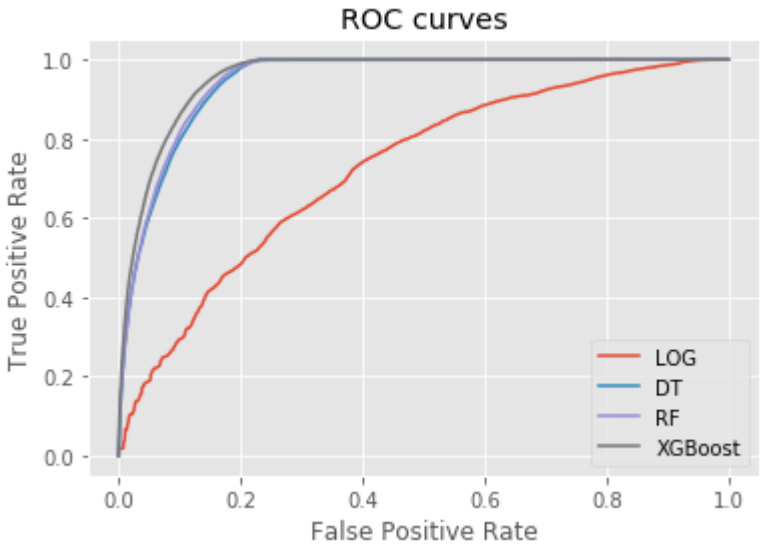
```
[CV] ................. max_depth=5, n_estimators=1000, total= 1.5min
[CV] max_depth=10, n_estimators=500 ..................................
[CV] ................. max_depth=10, n_estimators=500, total= 1.5min
[CV] max_depth=10, n_estimators=500 ..................................
[CV] ................. max_depth=10, n_estimators=500, total= 1.5min
[CV] max_depth=10, n_estimators=500 ..................................
[CV] ................. max_depth=10, n_estimators=500, total= 1.5min
[CV] max_depth=10, n_estimators=1000 .................................
[CV] ................. max_depth=10, n_estimators=1000, total= 3.0min
[CV] max_depth=10, n_estimators=1000 .................................
[CV] ................. max_depth=10, n_estimators=1000, total= 2.9min
[CV] max_depth=10, n_estimators=1000 .................................
[CV] ................. max_depth=10, n_estimators=1000, total= 3.1min
[CV] max_depth=20, n_estimators=500 ..................................
[CV] ................. max_depth=20, n_estimators=500, total= 3.2min
[CV] max_depth=20, n_estimators=500 ..................................
[CV] ................. max_depth=20, n_estimators=500, total= 3.1min
[CV] max_depth=20, n_estimators=500 ..................................
[CV] ................. max_depth=20, n_estimators=500, total= 3.1min
[CV] max_depth=20, n_estimators=1000 .................................
[CV] ................. max_depth=20, n_estimators=1000, total= 6.3min
[CV] max_depth=20, n_estimators=1000 .................................
[CV] ................. max_depth=20, n_estimators=1000, total= 6.4min
[CV] max_depth=20, n_estimators=1000 .................................
[CV] ................. max_depth=20, n_estimators=1000, total= 6.3min
[Parallel(n_jobs=1)]: Done  18 out of  18 | elapsed: 52.6min finished
F1-score:  0.7897432299705346
Model prediction graphs.......
No handles with labels found to put in legend.
```

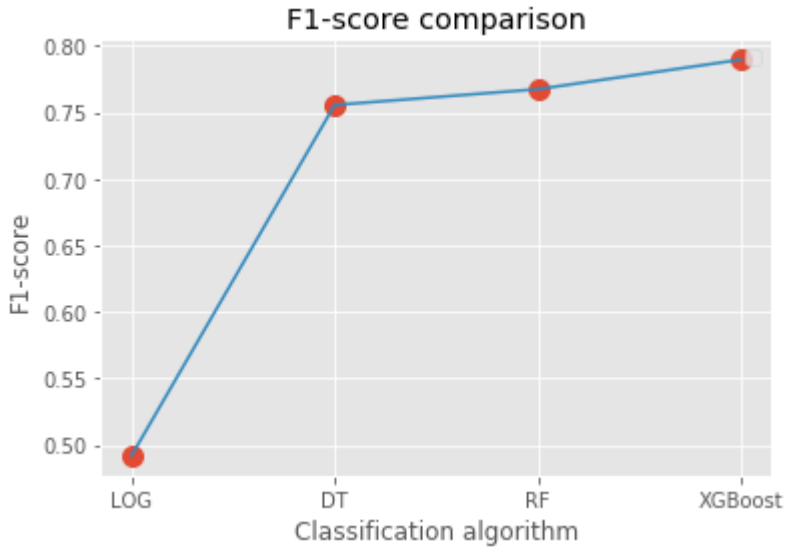



DT ROC curve
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0


RF ROC curve
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0


XGB ROC curve
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0


ROC curves
True Positive Rate
False Positive Rate
LOG
DT
RF
XGBoost
0.0
0.2
0.4
0.6
0.8
1.0


F1-score comparison
F1-score
Classification algorithm
LOG
DT
RF
XGBoost
0.50
0.55
0.60
0.65
0.70
0.75
0.80

In [2]: