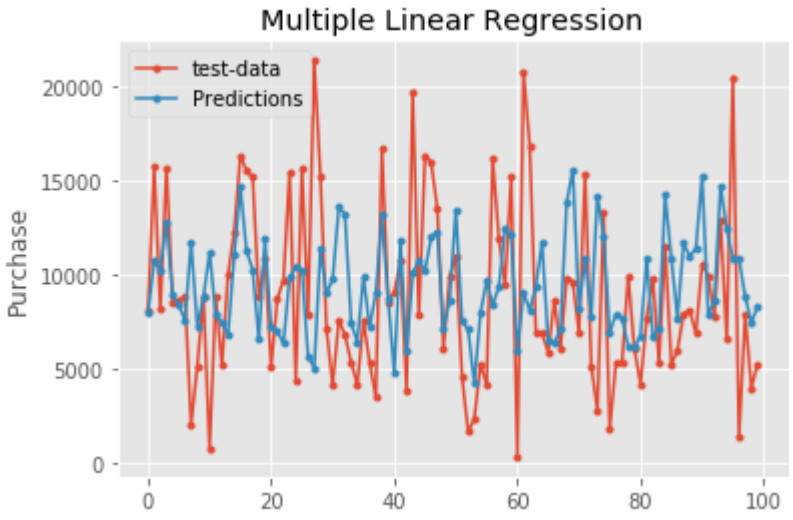
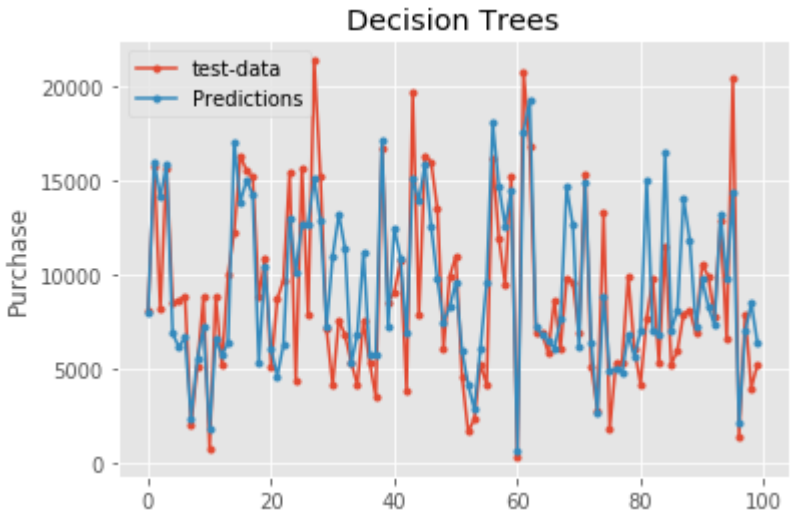
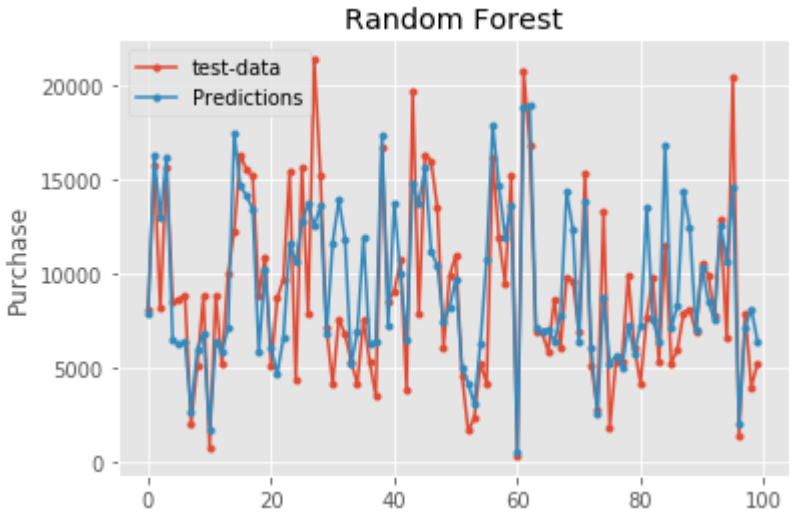
```
Python 3.6.7 |Anaconda, Inc.| (default, Oct 28 2018, 19:44:12) [MSC v.1915
64 bit (AMD64)]
Type "copyright", "credits" or "license" for more information.

IPython 6.5.0 -- An enhanced Interactive Python.

Restarting kernel...
```

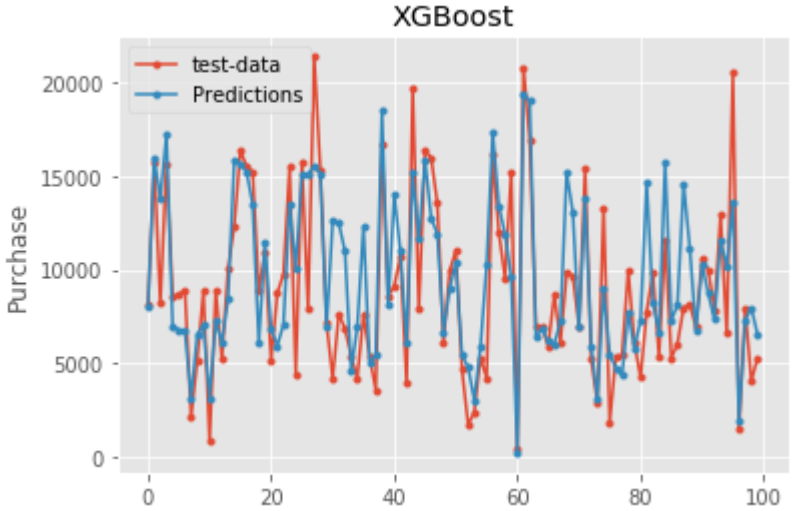
```
In [1]: runfile('E:/University_of_Florida/11th term/Statistical Machine
Learning/Project/purchase-prediction/purchase-reg-models-gridCV.py',
wdir='E:/University_of_Florida/11th term/Statistical Machine
Learning/Project/purchase-prediction')
Importing libraries.......
Exporting data.......
<class '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.......
Regression Model Training.......
Multiple Linear Regression

 Training and Testing Performance Report
Train RMSE : 4381
Test RMSE : 4365
CV Score : Mean - 4381 | Std - 13.42
Decision Tree Regression
Fitting 3 folds for each of 8 candidates, totalling 24 fits
[Parallel(n_jobs=1)]: Using backend SequentialBackend with 1 concurrent
workers.
[Parallel(n_jobs=1)]: Done  24 out of  24 | elapsed:   24.8s finished
0.6993336829333154
DT Train RMSE : 2698
DT Test RMSE : 2746
Random Forest Regression
Fitting 3 folds for each of 18 candidates, totalling 54 fits
[Parallel(n_jobs=1)]: Using backend SequentialBackend with 1 concurrent
workers.
```

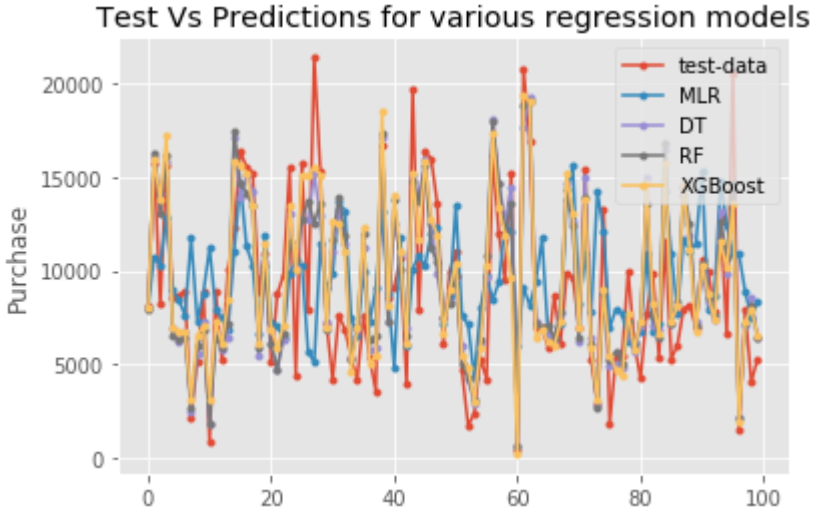
```
[Parallel(n_jobs=1)]: Done  54 out of  54 | elapsed:  2.8min finished
0.7075660672647061
RF Train RMSE : 2616
RF Test RMSE : 2708
Extreme Gradient Boosting Decision Tree Regression
Fitting 3 folds for each of 60 candidates, totalling 180 fits
[Parallel(n_jobs=1)]: Using backend SequentialBackend with 1 concurrent
workers.
[Parallel(n_jobs=1)]: Done 180 out of 180 | elapsed: 563.0min finished
0.7472936977981175
XGB Train RMSE : 2009
XGB Test RMSE : 2518
Model prediction graphs.......
```

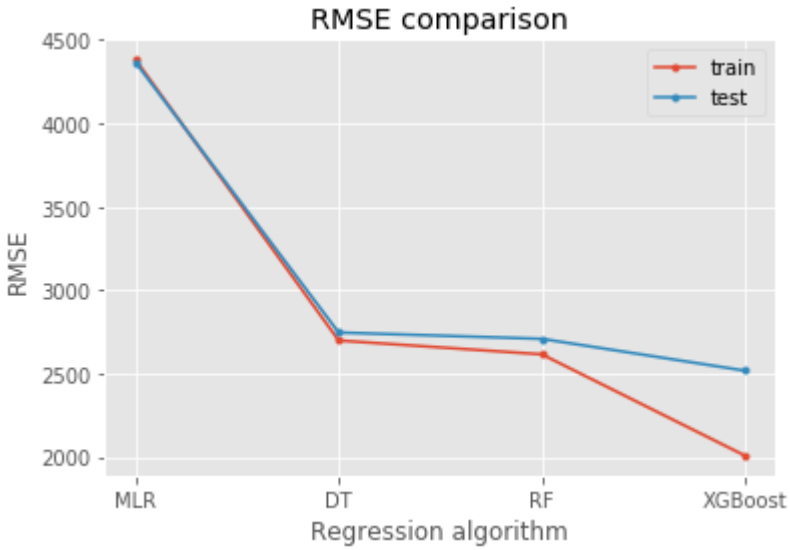





Random Forest
test-data
Predictions
Purchase
0
5000
10000
15000
20000
0
20
40
60
80
100


XGBoost
test-data
Predictions
Purchase
0
5000
10000
15000
20000
0
20
40
60
80
100


Test Vs Predictions for various regression models
test-data
MLR
DT
RF
XGBoost
Purchase
0
5000
10000
15000
20000
0
20
40
60
80
100


RMSE comparison
4500
4000
3500
3000
2500
2000
RMSE
train
test
MLR
DT
RF
XGBoost
Regression algorithm

In [2]: