

Spark DF, SQL, ML Exercise

```
Area under ROC = 0.773423840378484
```

```
Accuracy = 0.773423840378484
```

```
DenseMatrix([[ 5508.,  1840.],
               [  363., 2012.]])
```

```
F1 = 0.7876369354932344
```

```
Precision = 0.8365938991668678
```

```
Recall = 0.773423840378484
```

```
FPR = 0.17667422000201208
```

```
TPR = 0.773423840378484
```

Command took 25.98 seconds -- by a user at 6/15/2018, 7:00:39 AM on unknown cluster
Cmd 78



GBM - CV Model Evaluate

Cmd 79

```
1 | print_performance_metrics(gbCvPredictions)
```

```
auc = 0.8989499756467929
```

```
aupr = 0.7665167913162426
```

```
Area under PR = 0.6435054710769732
```

```
Area under ROC = 0.7508083832335329
```

```
Accuracy = 0.8484006993726216
```

```
DenseMatrix([[ 6919.,  429.],
               [ 1045., 1330.]])
```

```
F1 = 0.8401553392826959
```

```
Precision = 0.8412623026104288
```

```
Recall = 0.8484006993726216
```

```
FPR = 0.34678393290555576
```

```
TPR = 0.8484006993726216
```

Command took 21.18 seconds -- by a user at 6/15/2018, 7:01:12 AM on unknown cluster
Cmd 80



Logistic Regression - Model Explore

Cmd 81

```
1 | print('Model Intercept: ', lrCvModel.bestModel.intercept)
```

```
Model Intercept: -1.2479134417997437
```

Command took 0.01 seconds -- by a user at 6/15/2018, 7:01:36 AM on unknown cluster
Cmd 82