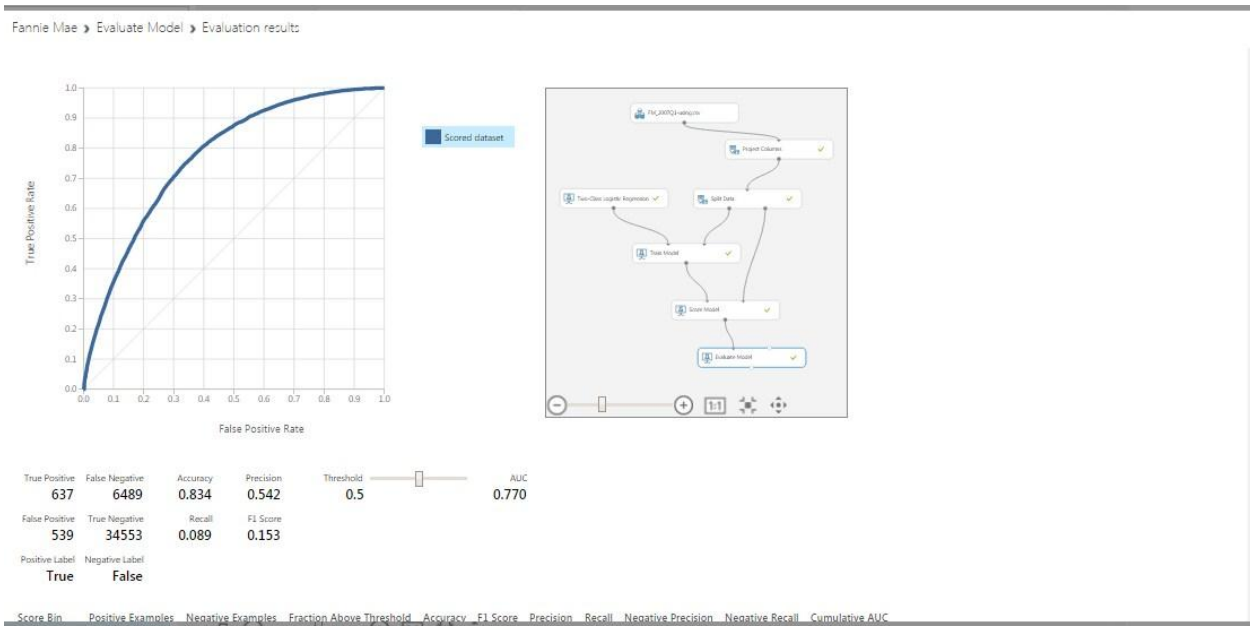
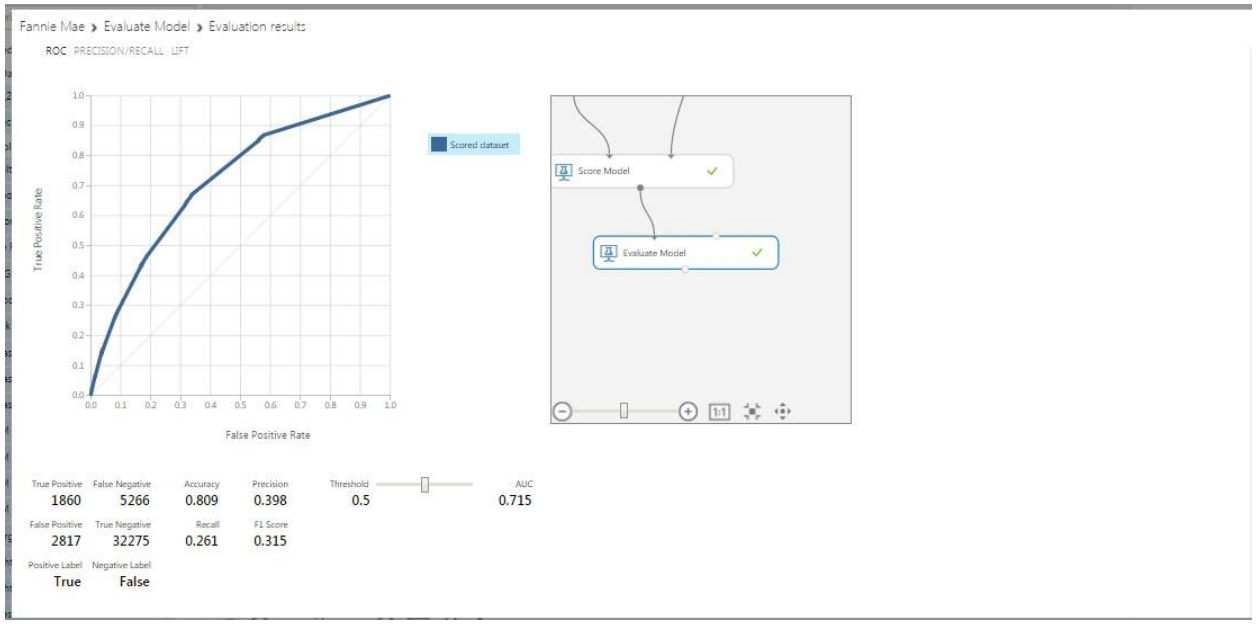
Fannie Mae Report – MS Azure Modelling

a.) Logistic Regression Model



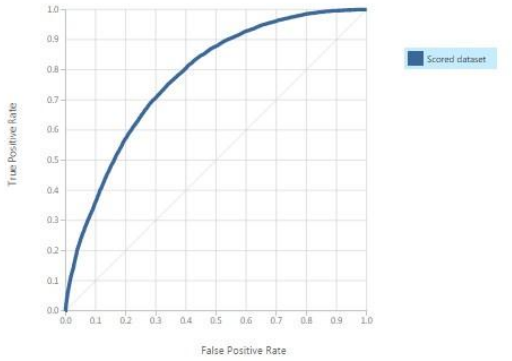
b.) Random Forests

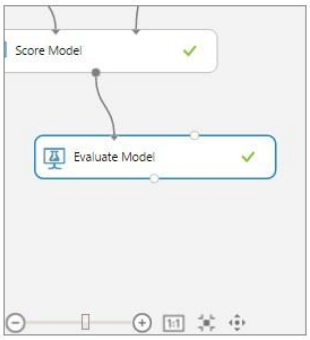


c.) Random Jungle

Fannie Mae › Evaluate Model › Evaluation results

ROC PRECISION/RECALL LIFT





True Positive	False Negative	Accuracy	Precision	Threshold	AUC
464	6662	0.835	0.605	0.5	0.775

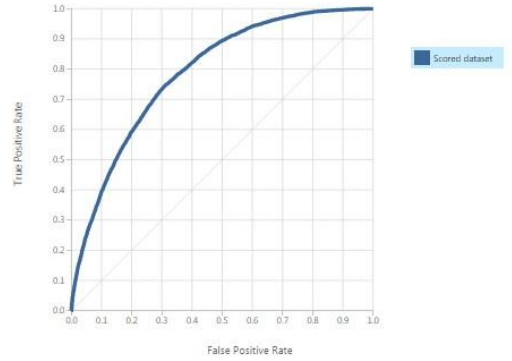
False Positive	True Negative	Recall	F1 Score
303	34789	0.065	0.118

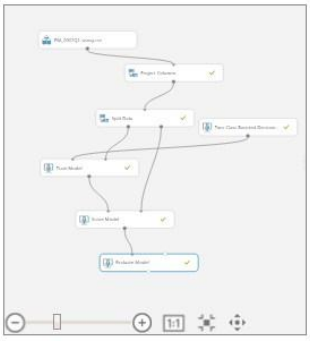
Positive Label	Negative Label
True	False

d.) Boosted Trees

Fannie Mae › Evaluate Model › Evaluation results

ROC PRECISION/RECALL LIFT





True Positive	False Negative	Accuracy	Precision	Threshold	AUC
1236	5890	0.836	0.542	0.5	0.788

False Positive	True Negative	Recall	F1 Score
1043	34049	0.173	0.263

Positive Label	Negative Label
True	False

e.) Neural Networks

1.)

	Precision	Recall(Sensitivity)	Specificity	F1	AUC	Accuracy
Logistic Regression	0.542	0.089	0.98	0.153	0.77	0.834
Random Forest	0.398	0.261	0.91	0.315	0.715	0.809
Random Jungle	0.605	0.065	0.99	0.118	0.775	0.835
Boosted Trees	0.542	0.173	0.97	0.263	0.788	0.836
Neural Networks	0.573	0.056	0.99	0.102	0.779	0.834

From doing some background research, we could find that:

Average loan amount per borrower in first quarter = $202,297 (average amount from the data)
Average interest rate per loan = 6.25 APR (average interest from the given data)
= 6.25/12= 0.5% per month

Average borrowing period = 360 months (provided in the data)

Thus, over the period of 360 months we realize that the every **FP** which is **compounded interest on principal amount** costs around 500% of the current principal amount. Hence, every false negative would cost us approximately $1M in opportunity cost over the period of 30 years. Using, current inflation

adjusted cost at 5% rate of interest would be $230,000. On the other hand, every FN, which is unexpected default or delinquency, would cost me $202,297(Principal amount default).
From the above calculation, we realize that every FP and FN approximately proves equally costly for Fannie Mae and one should focus on model that helps reduce these values. A model with F1 value closer to 1 would successfully help reduce such losses and increase the scope of earning profits through opportunity cost.

Thus, out of the given models, **Boosted trees** model has the **highest F1 value** compared to other models, and is a better model.

2.) With FN = 5890, FP = 1043, TP = 1236 and TN = 34049

Therefore,

a.) Unexpected defaulting of borrower (loss due to FN) = $1.19 Billion
b.) Opportunity loss(loss due to FP) = $2.4 Billion
c.) Amount Profited due to proper lending to the right customer (Profit due to TN) = $7.8 Billion d.) TP neither results in profits or loss.

Hence, total profit is 7.8 – (2.4 + 1.19) = $4.21 Billion

The given model predicts a profitable scenario where in Fannie Mae earns a substantial profit. For insolvency the losses should be more than the profits earned, which does not happen in this case. Therefore, the model could have successfully prevented Fannie Mae insolvency.

Conclusion: This model successfully provides a profitable prediction and it could have prevented Fannie Mae insolvency.