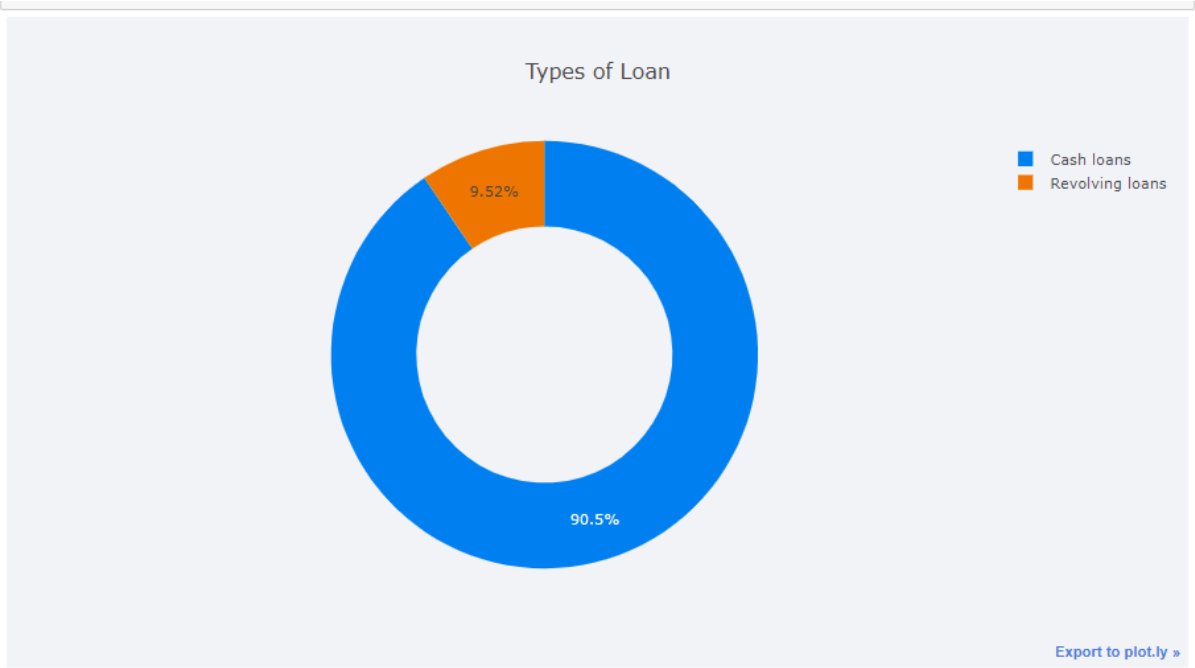
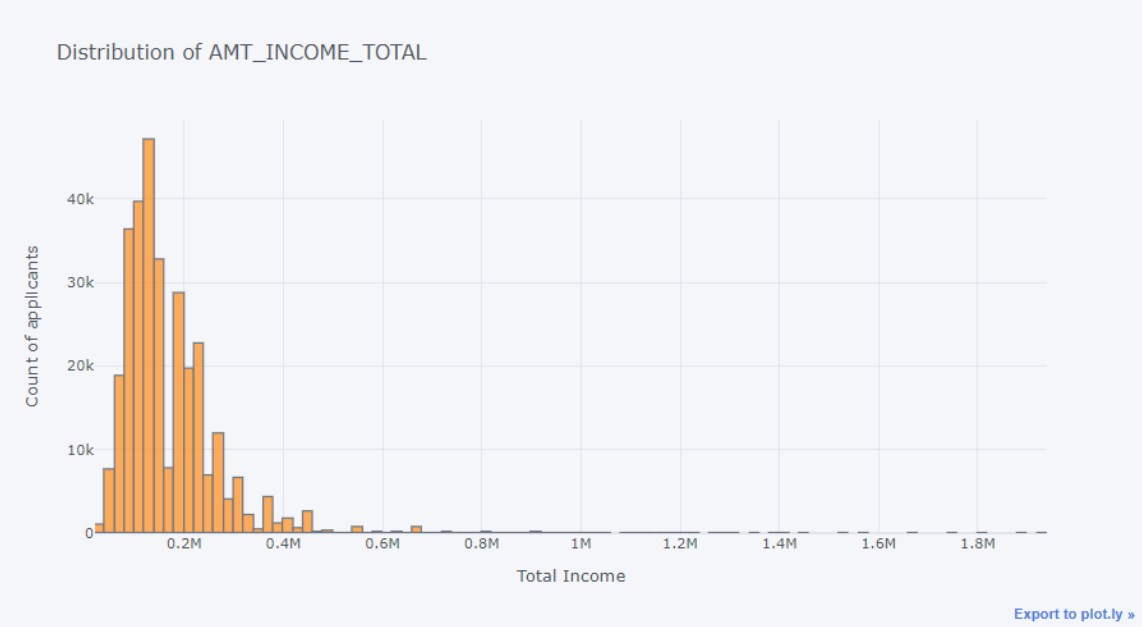
Credit Analytics Presentation

The primary objective of this study is to assist Home Credit in deciding which loan applications should be disbursed, and which should be rejected, based on the applicant's past behaviour and application information.

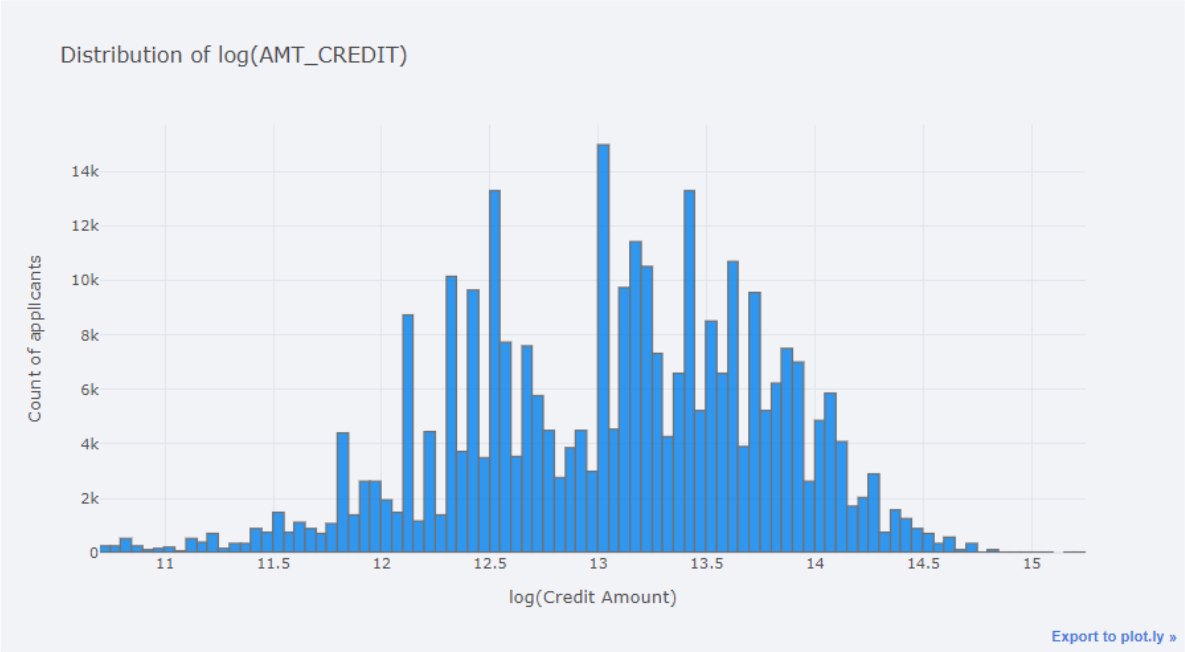
EDA





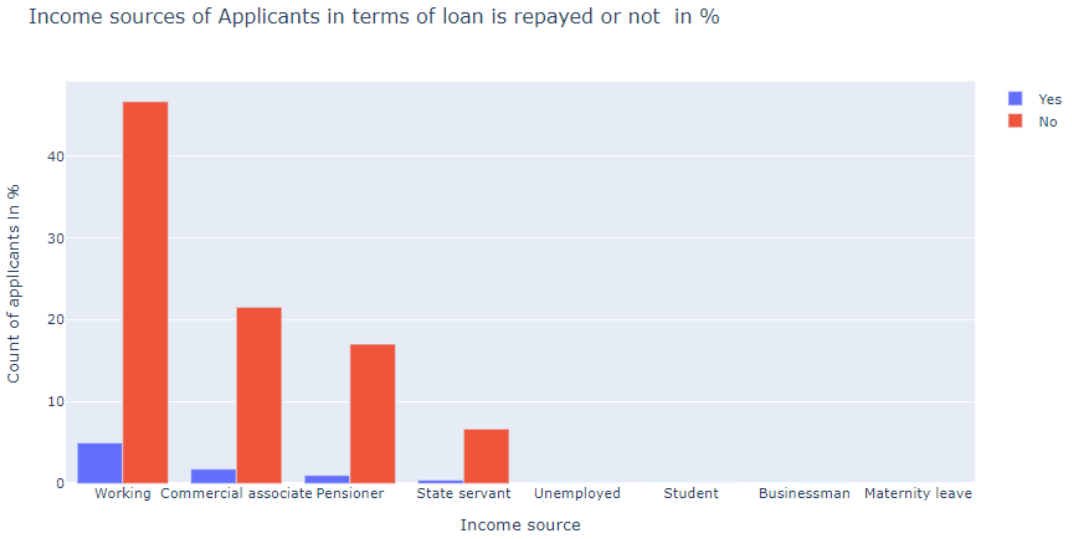
Observations:

1. The distribution is right skewed and there are extreme values, we can apply log distribution.
2. People with high income(>1000000) are likely to repay the loan.

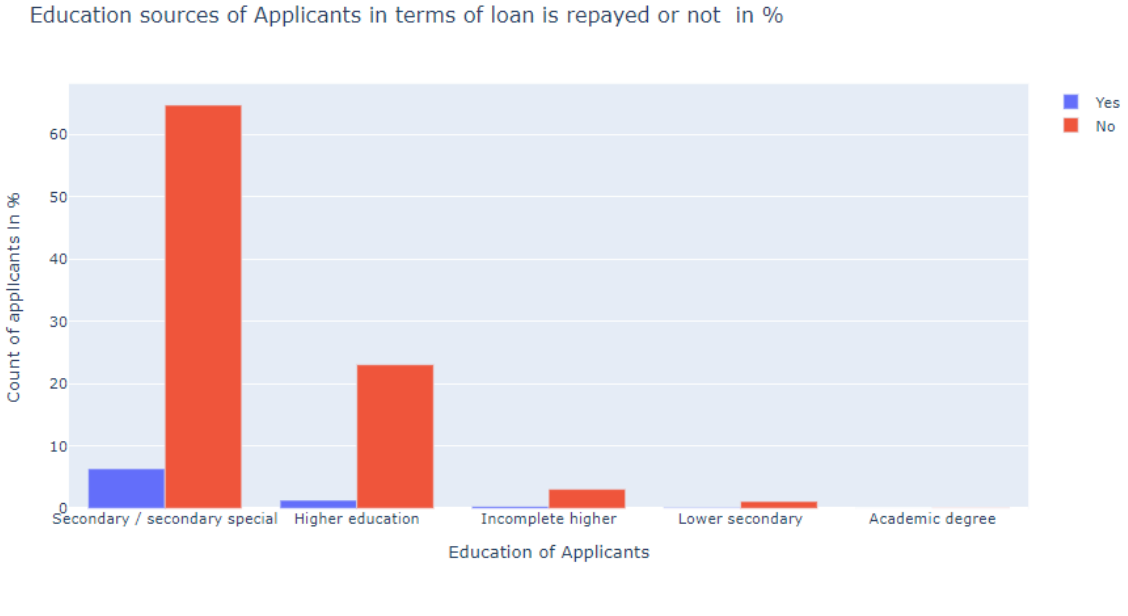


Observations:

1. People who are taking credit for large amount are very likely to repay the loan.
2. Originally the distribution is right skewed, we used log transformation to make it normal distributed.

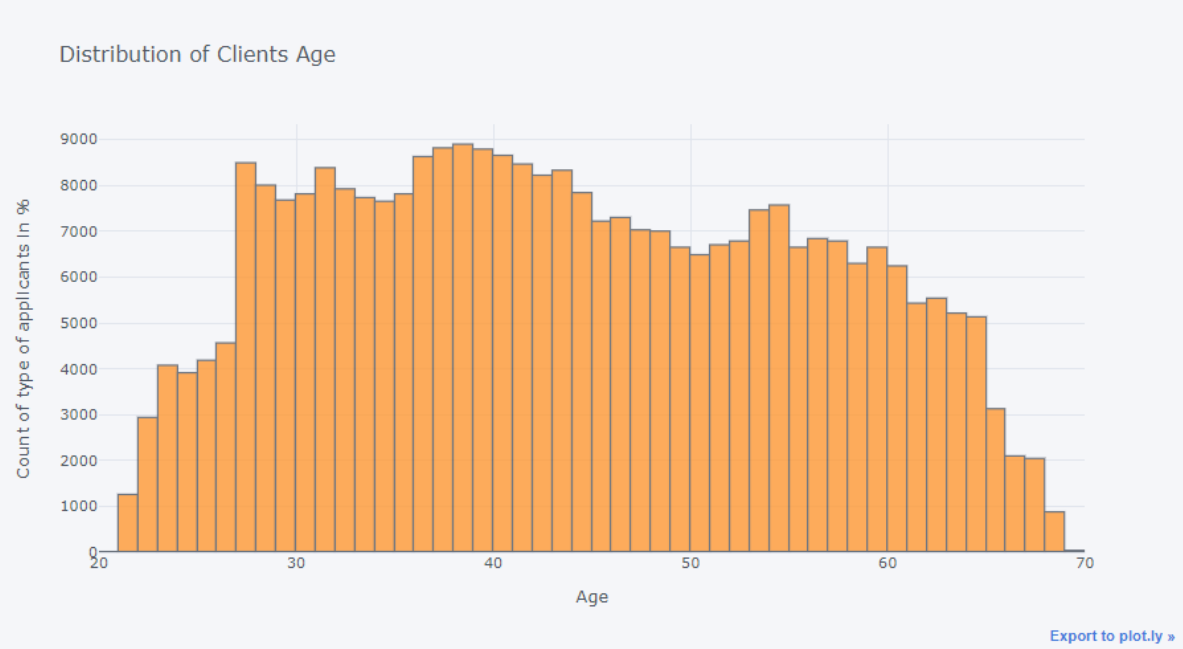


Observations: All Students and Businessman are repaying the loan.



Observations:

1. People with Academic Degree are more likely to repay the loan(Out of 164, only 3 applicants are not able to repay)

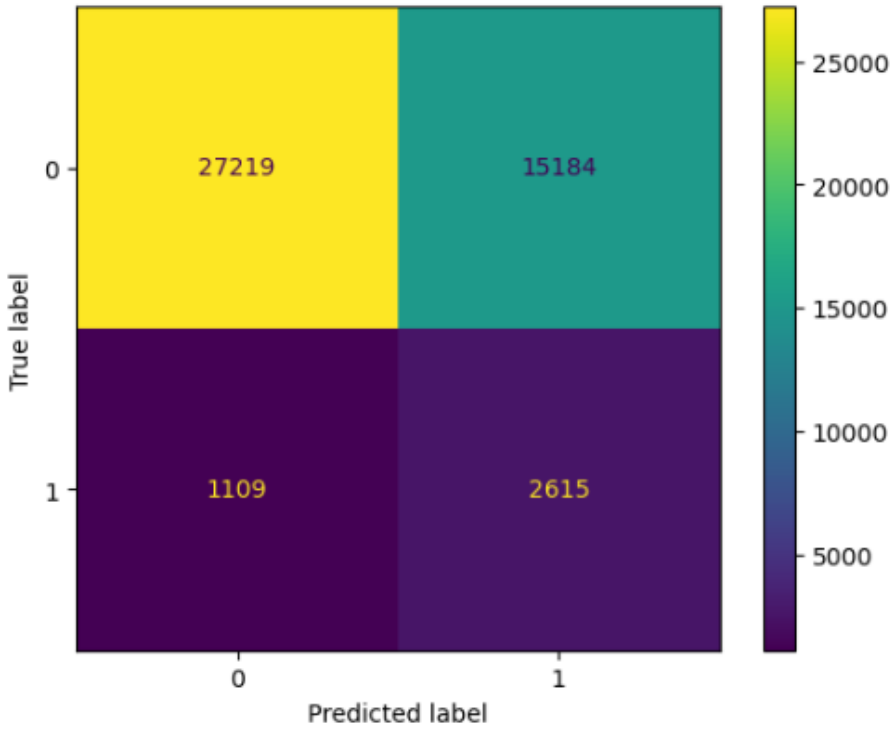


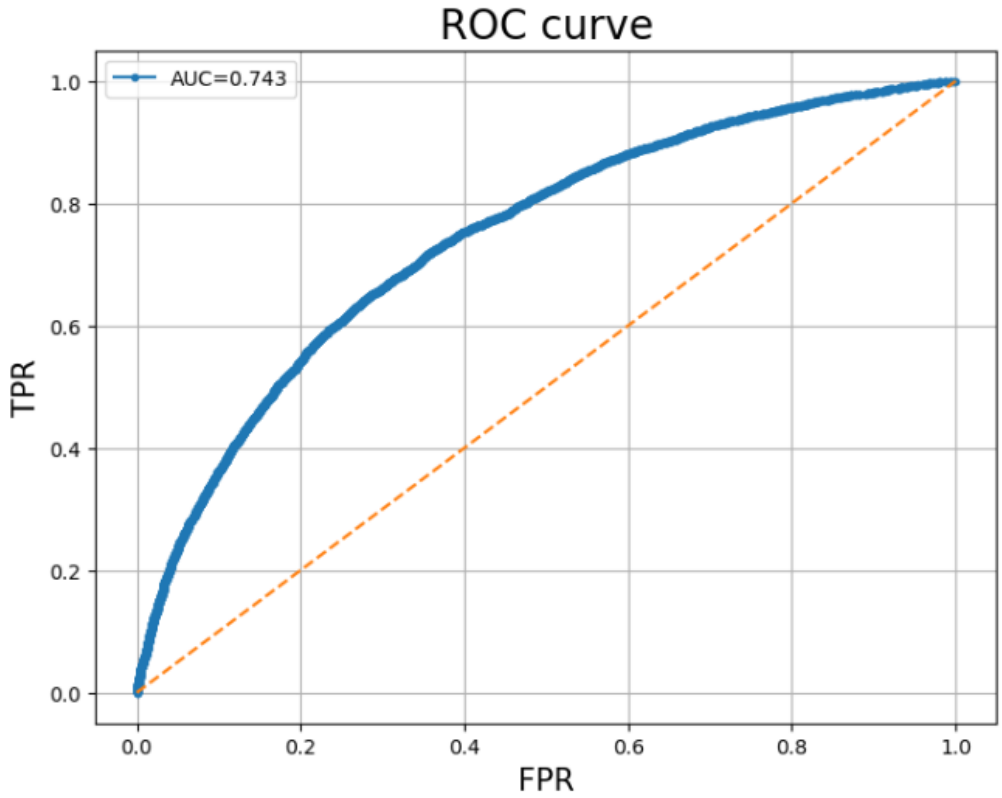
Machine Learning Models:

Cross validation results and plot for Logistic Regression model.

```
For best alpha 0.0001, The Train AUC score is 0.7448744907447847
For best alpha 0.0001, The Cross validated AUC score is 0.7439458144858475
For best alpha 0.0001, The Test AUC score is 0.7433035005807024
The test AUC score is : 0.7433035005807024
The percentage of misclassified points 35.32% :
```

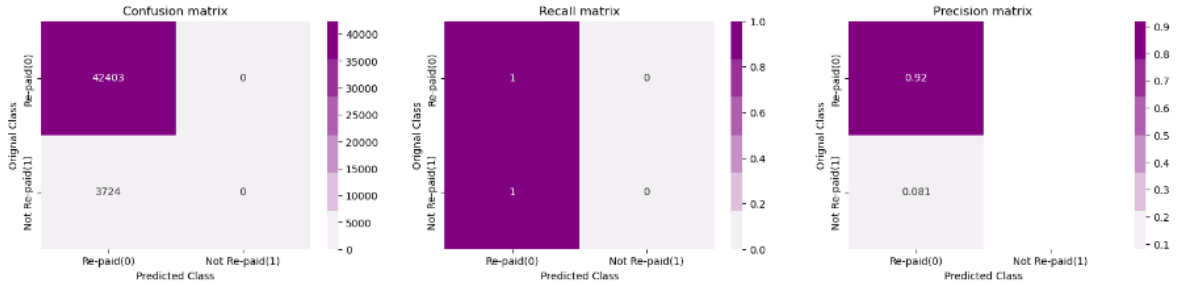
]: `<sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x249017cf490>`

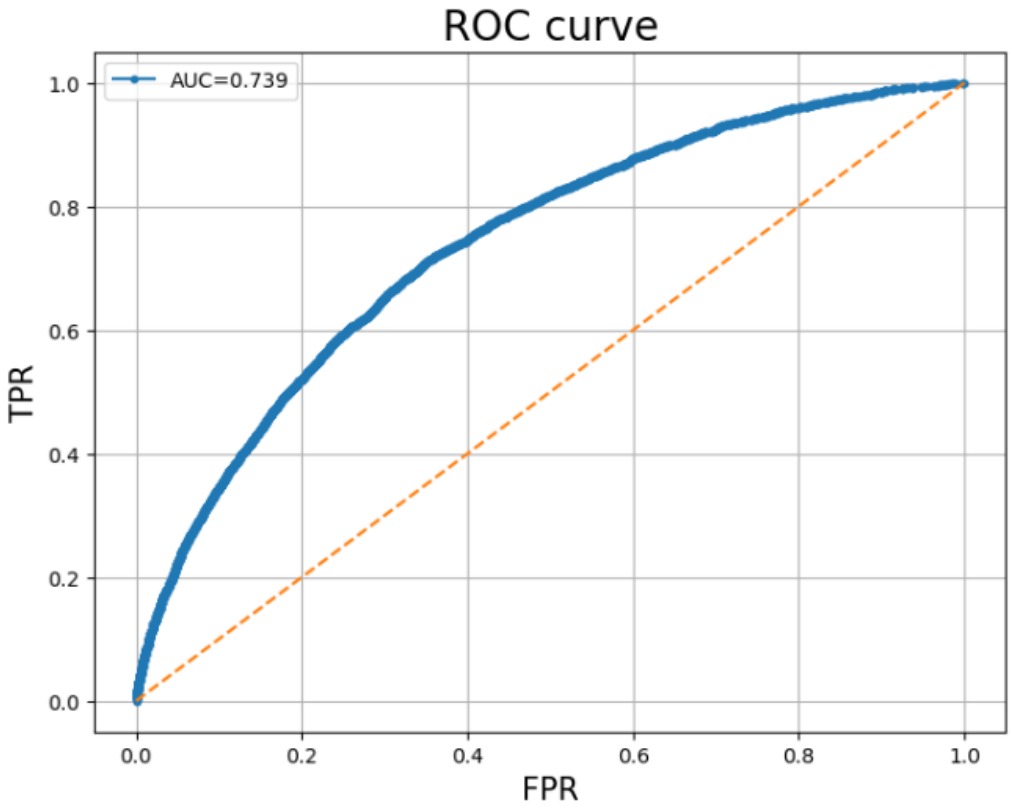




Cross validation results for Random Forest model.

```
For best n_estimators 500 best max_depth 10, The Train AUC score is 0.8182380712097698
For best n_estimators 500 best max_depth 10, The Validation AUC score is 0.7400160897964554
For best n_estimators 500 best max_depth 10, The Test AUC score is 0.7385312767804945
The test AUC score is : 0.7385312767804945
The percentage of misclassified points 08.07% :
```





Overview of Results

Logistic regression with Selected Features:

```
Train AUC:0.744
Valication AUC:0.743
Test AUC:0.743
```

Random Forest with Selected Features:

```
Train AUC:0.818
Valication AUC:0.740
Test AUC:0.738
```

Logistic Regression gives the best performance and it is also faster to train when compared to Random Forest.