Part II – Classification Problem

Let's jump into some business.

Dream House Finance Business Needs

- **Pain Point:** Loan Applications need to be manually reviewed based on information provided by the customer. This ties up company resources…

- **Solution:** Company wants to implement a machine learning solution to automate the loan eligibility process based on customer detail provided while filling the online application form and give instant feedback.

- Key details are Gender, Marital Status, Education, Number of Dependents, Income, Loan Amount, Credit History, and others.

Outcome

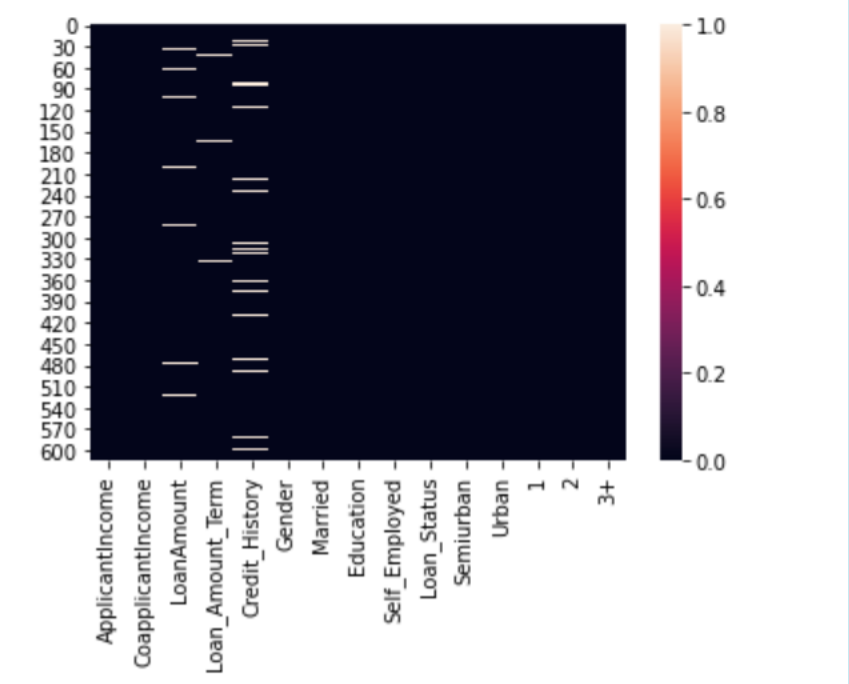
- **Random Forest With Grid Search** was the best solution due to its increased accuracy rate our business needs.

	precision	recall	f1-score	support
0	0.79	0.31	0.45	35
1	0.80	0.97	0.88	99
accuracy			0.80	134
macro avg	0.79	0.64	0.66	134
weighted avg	0.80	0.80	0.76	134

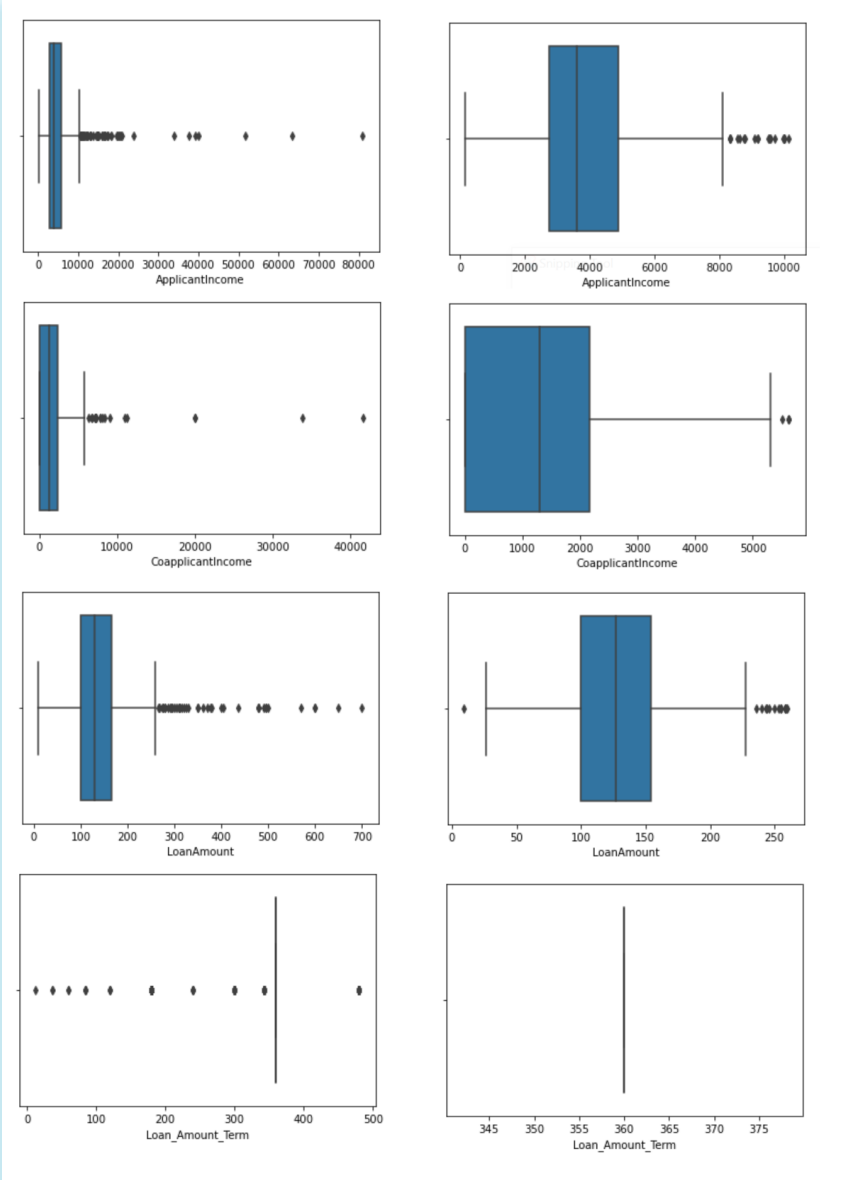
	Predicted False	Predicted True
Actual False	11	24
Actual True	3	96

Training Data Overview

- Categorical data that needed to be converted.
- Missing values in the data set.
- Dropped Loan ID.
- Extreme outliers were found.

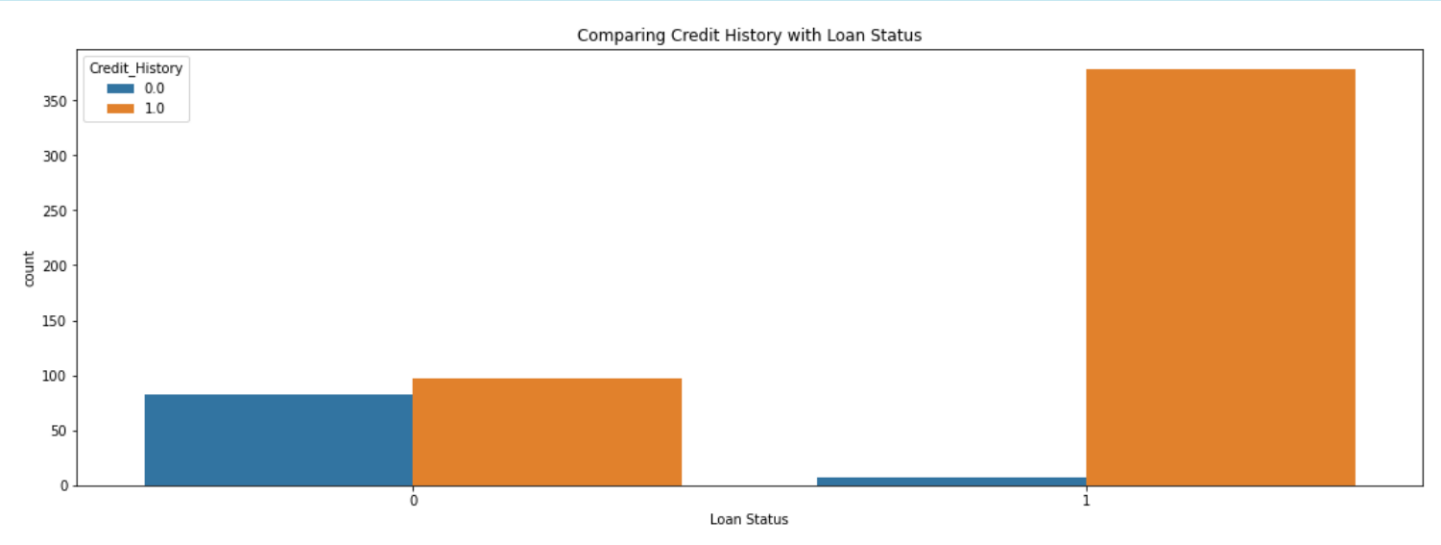


Outliers (Before and After)

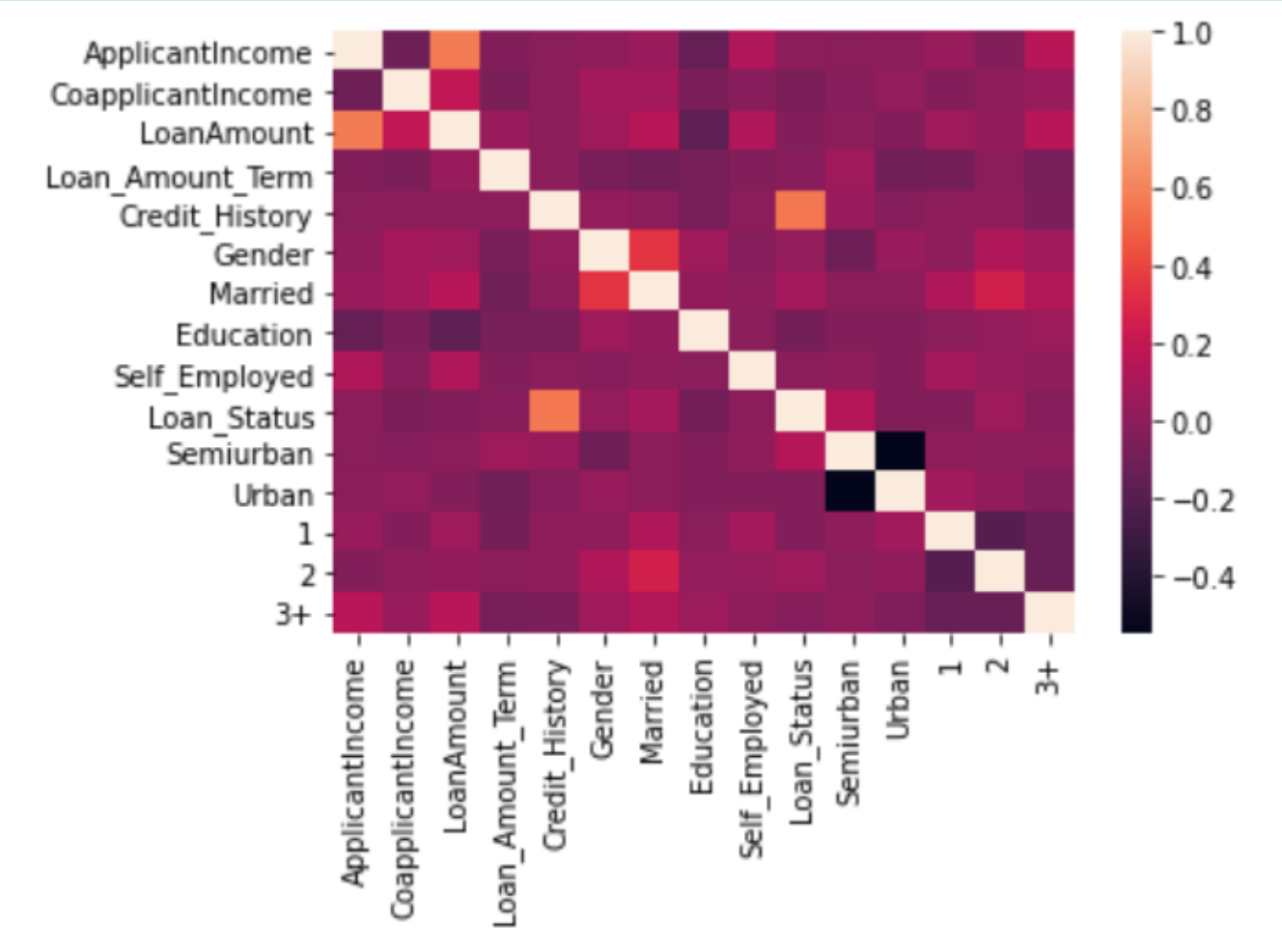


Some of the deciding factors

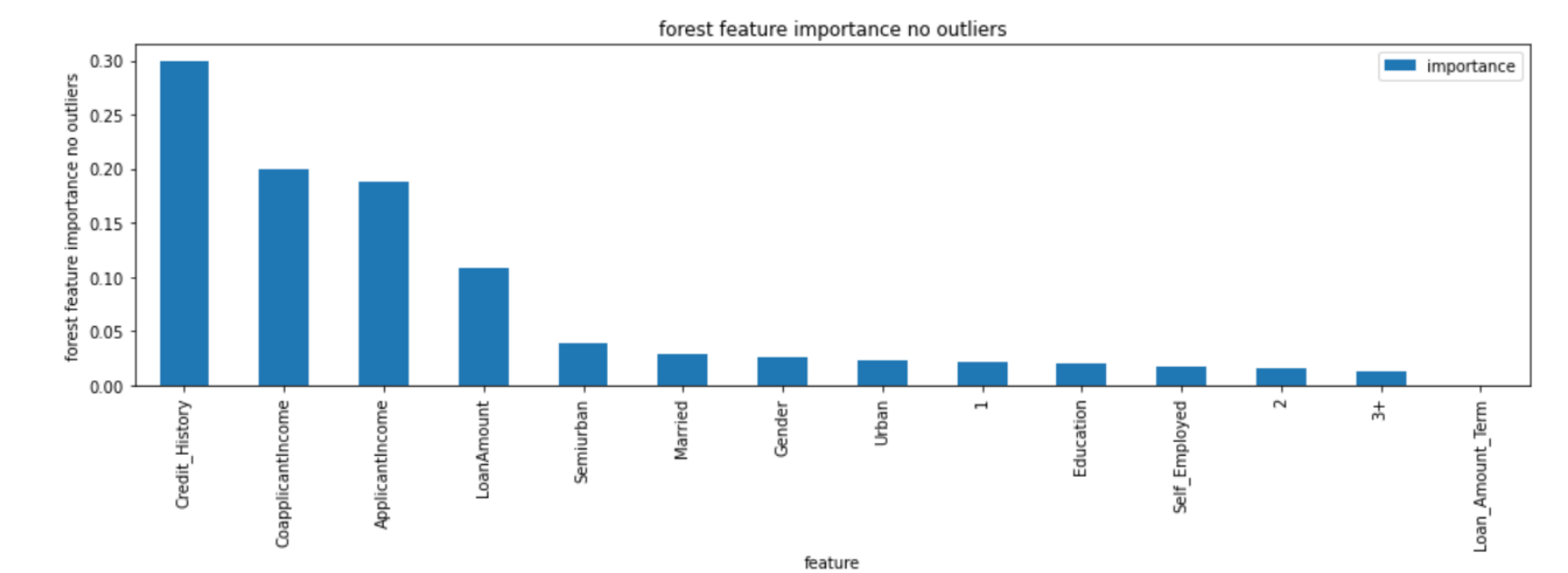
- Credit History played a significant role in determining loan application approval.



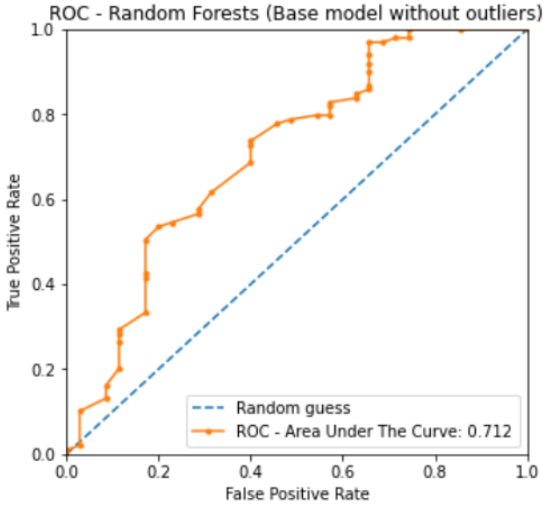
Correlation Heatmap

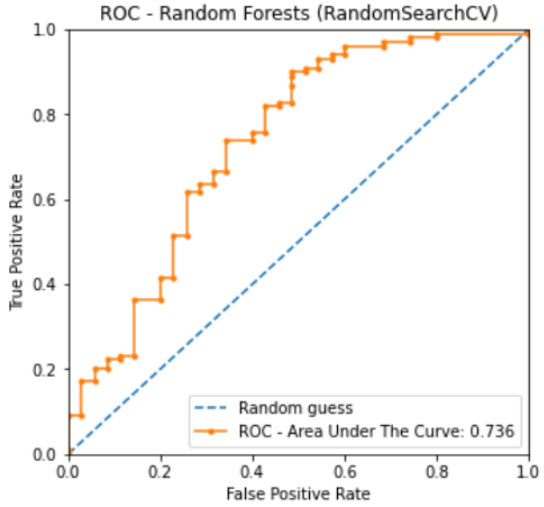


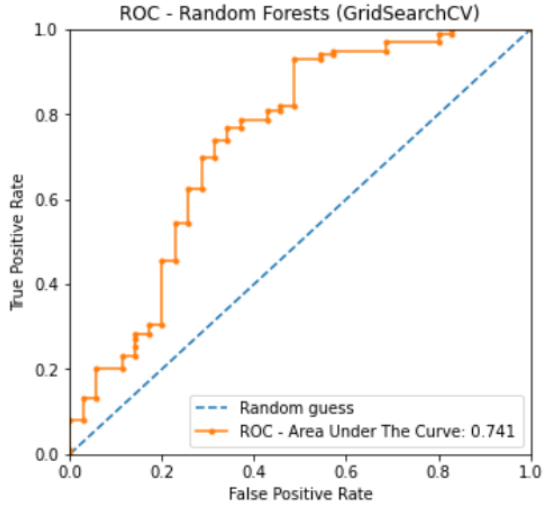
Feature Importance

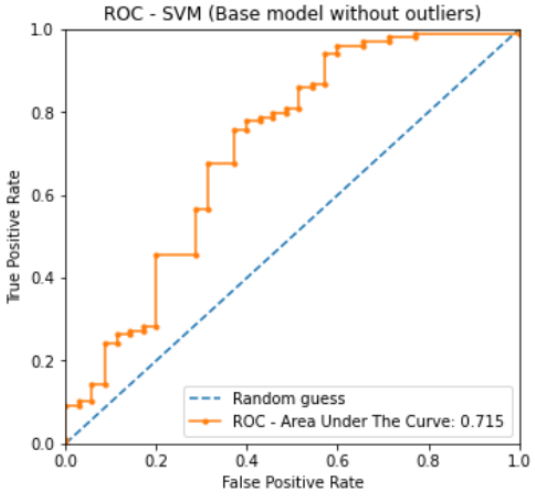


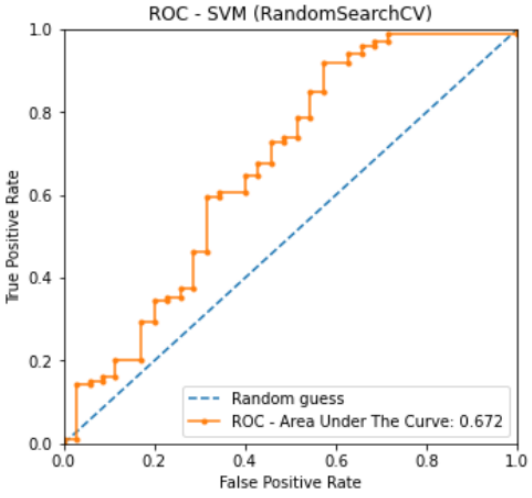
Training Models

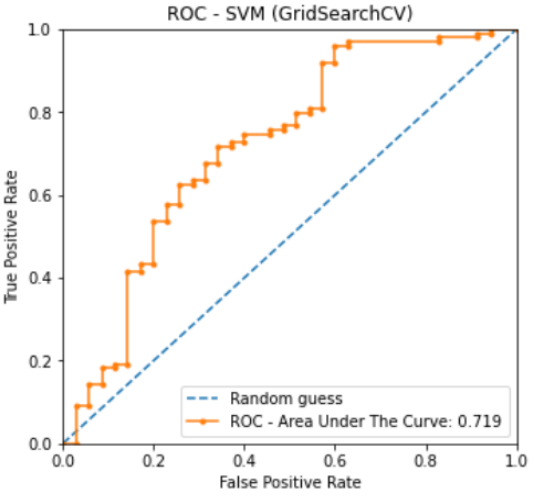












Candidate Model

	precision	recall	f1-score	support
0	0.79	0.31	0.45	35
1	0.80	0.97	0.88	99
accuracy			0.80	134
macro avg	0.79	0.64	0.66	134
weighted avg	0.80	0.80	0.76	134

	Predicted False	Predicted True
Actual False	11	24
Actual True	3	96

- The Random Forest with Grid Search offers the best ROC area curve and with a training time approx. 2.5 seconds. Compared to SVM at 0.02 seconds.
- Accuracy was the deciding factor for our business needs.

Outcome on Provided Test Data

- The candidate model has determined from the test data set that 84% of the applications are flagged to be approved while 16% of the test data was flagged to be rejected.
- This machine learning solution will resolve the business's pain point on determining loan eligibility.

Thank you for your time

Questions? Feedback?