Capstone Final Presentation

Problem

How do we prevent lapses and the losses associated with them.

1. Who is lapsing and how much is it costing us?
2. Why are they lapsing?
3. How do we prevent lapses

Setting the scene

Amount $ 872 354 700
per period

Population 79 853

Total Population

Churners

Amount $ 47 978 400
per period

Population 4 998

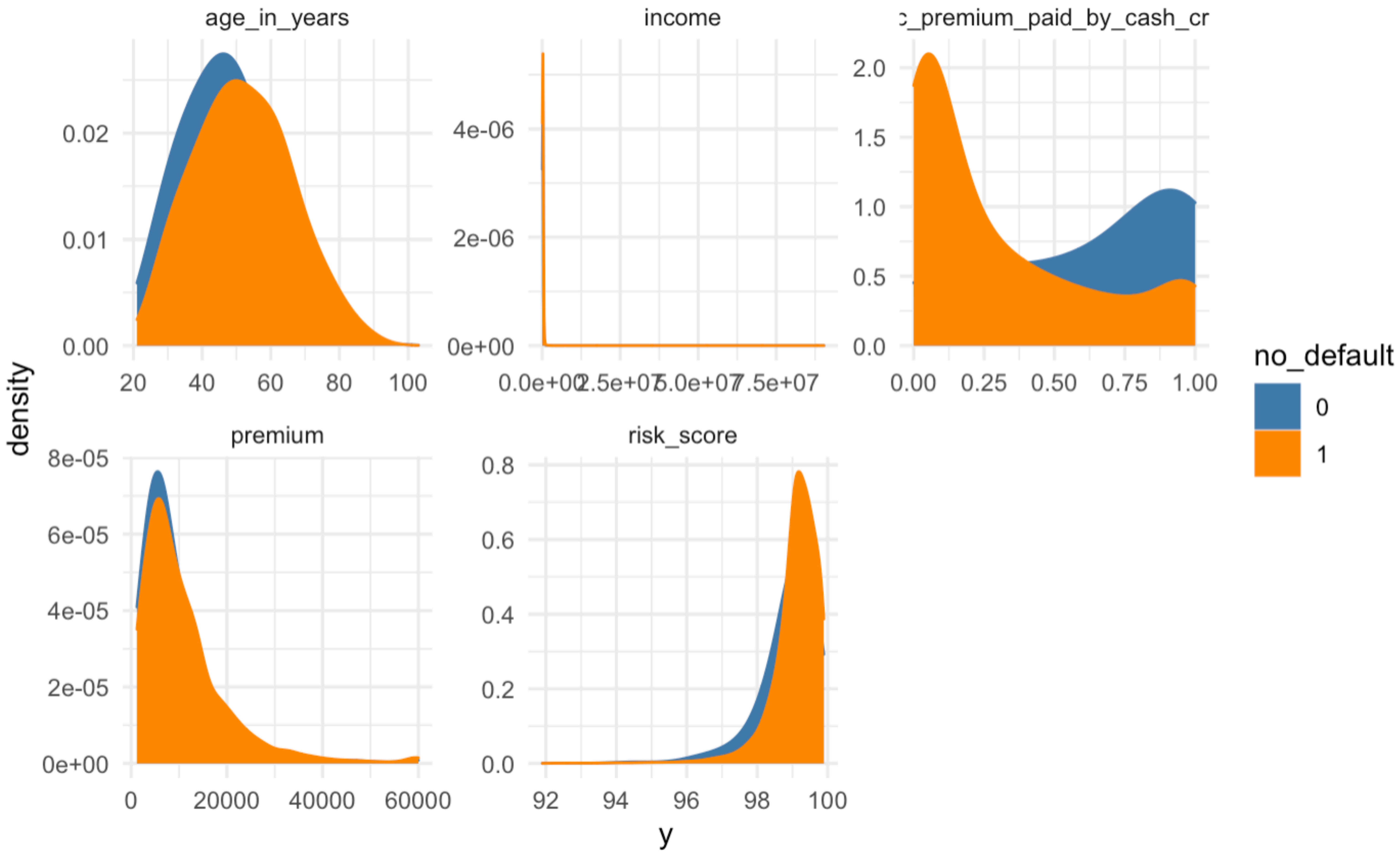
6%

Non Churners

Amount $ 824 376 300
per period

Population 74 855

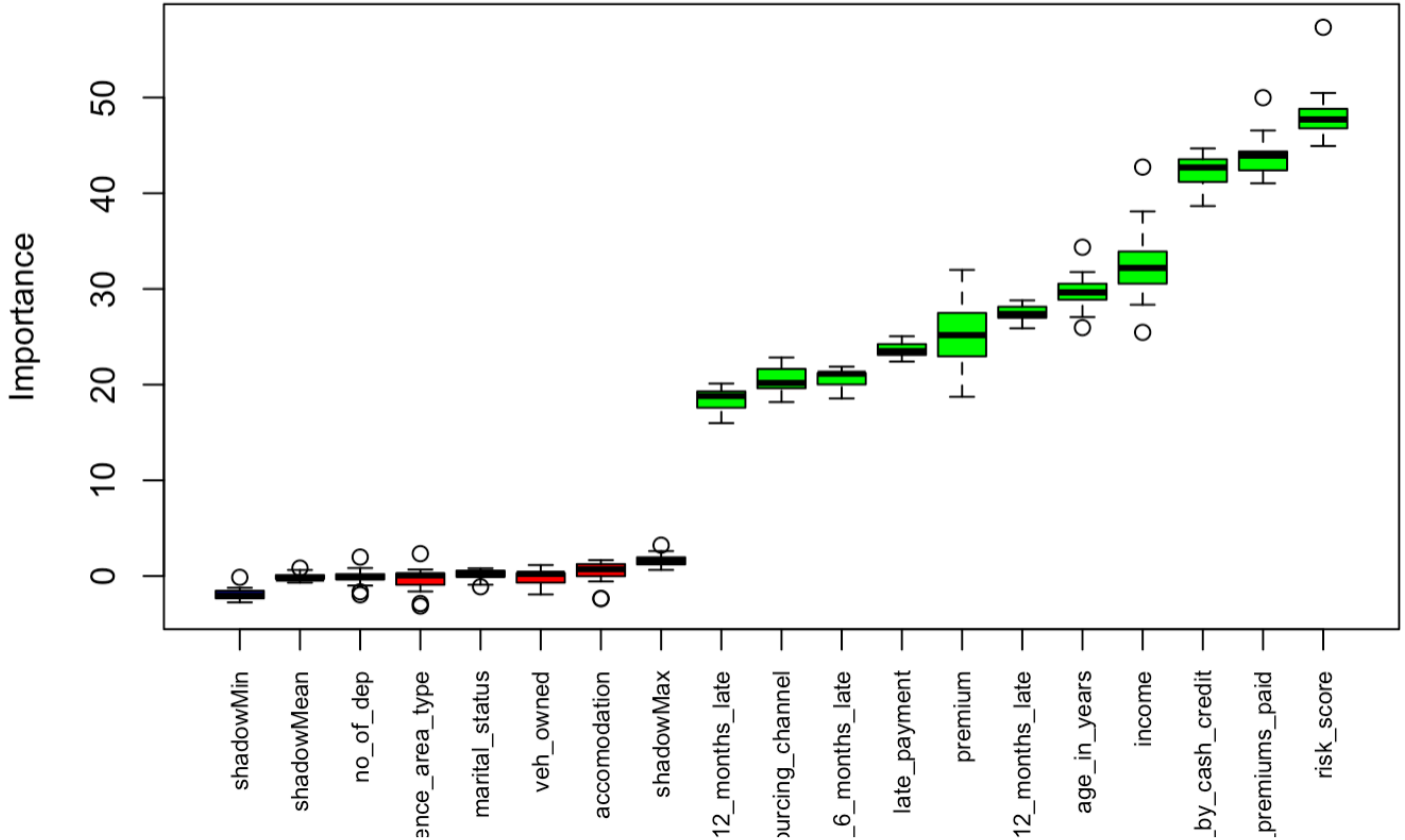
94%



Quick Insights

Individuals with a higher percentage of their premium paid by cash or credit are more likely to default

Important Variables



Quick Insights

The most important variable in predicting whether someone will leave to stay is their risk score followed by the number of premiums paid

Variables like the number of dependents, residence area type

Marital status, vehicle owned and the accommodation are not meaningful interdicting the important variables

Building the model

Picking the right model

We build four models

GLM	LOGISITC	NEURAL NETWORK	GBM
(+) Efficient to train (+) highly scalable (-) requires independent predictor features	(+) Easy to train (+) Efficient to train (-) Black box (-) Takes longer to train	(+) Robust (-) Black box (-) Takes longer to train	(+) Robust (-) Black box (-) Takes longer to train

Comparing model performance of top 3 models

Baseline models were built for the selected models and only the most accurate model was further developed
The GBM model performed the best compared to all the other models

GLM

Reference

Prediction

	Error%
Default	99%
No_Default	0...%
Totals	6%

AUC: 0.83
AUCPR: 0.98

Neural Network

Reference

Prediction

	Error%
Default	36%
No_Default	13%
Totals	25%

AUC: 0.83
AUCPR: 0.98

GBM Model

Reference

Prediction

	Error%
Default	29%
No_Default	16%
Totals	22%

AUC: 0.84
AUCPR: 0.98

Quick Insights

- Further enhancements to the modelling process include removing outliers and SMOTE variable oversampling
- The GBM model performs the best compared to all the other models and has a lower likelihood of misclassifying a person as a defaulter
- The AUC and AUCPR are quite high due to the class imbalance and the best model is judged on how well it does when it comes to predicting the clients who default (that’s what we are mostly concerned with)

Comparing the models

Analysing the results

Confusion Matrix

	Actual = Yes	Actual = No
Predicted = Yes	TP	FP
Predicted = No	FN	TN

Metric Considerations

- The point of the model is to make the best prediction of the individuals who default. This means that we want to minimise the number of people we wrongly predict to default(false positives) and reduce the false negatives as well (i.e. people we predict to default)
- Because of the highly imbalance nature of the data and the fact that there are more people who do not default, we need to choose the model that has the lowest false positive rate.

Top 3 performing models

Comparing results

GBM Model

Prediction

	Default	No Default	Error	Error%
Default	707	297	297/1004	29%
No Default	2715	12251	2715/14966	18%
Total	11	116	3012/15970	19%

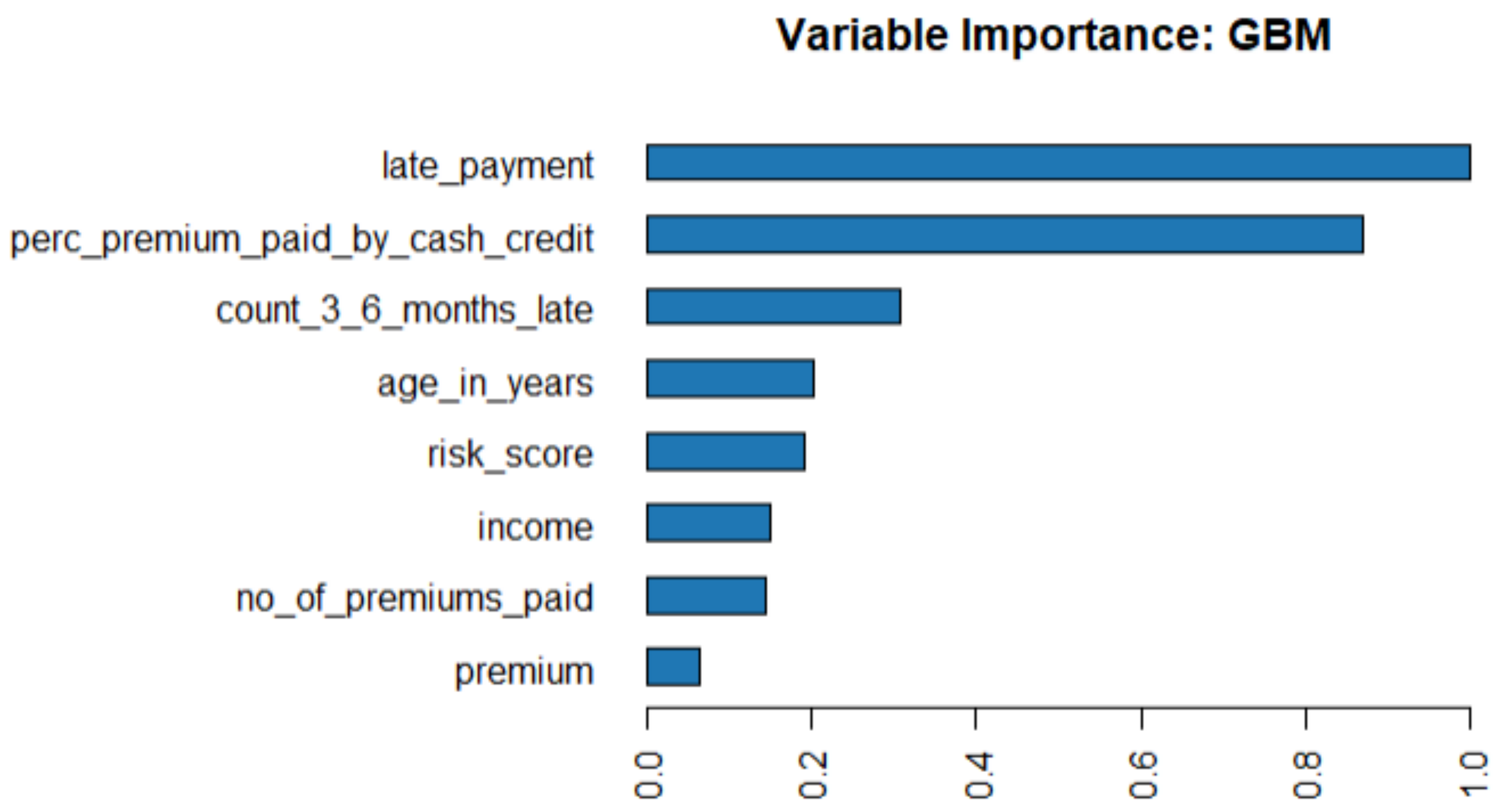
Quick Insights

- The best model misclassifies 29% of people who default as if they do not.
- A risk rating based on the model output has been produced and as such there will be 3 tiers of clients (High risk, Medium Risk and Low risk)
- The clients with high risk are those most likely to default.
- Action should be taken once specific conditions are met

Conclusion

Conclusion

- The most important variable in determining default is whether a person missed a premium or not
- Clients have been tiered into three groups based on their probability of default
- The risk bands are high, medium and low
- If an individual has missed a payment and their risk level is high or medium the business should intervene
- If the person has a higher percentage of their premium paid by credit or cash (higher than 50%) then the business should keep an eye on the clients

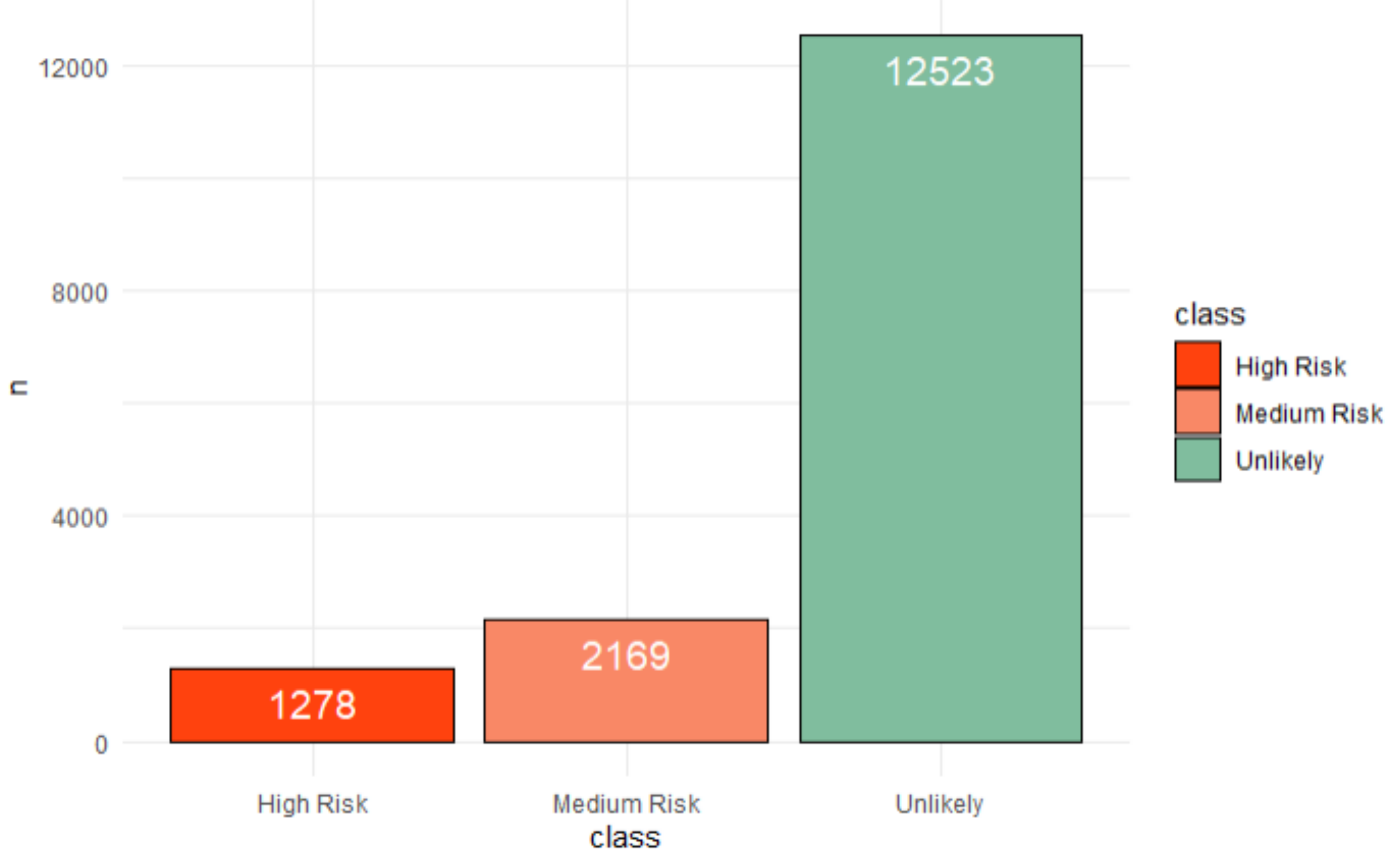


Important Variables

Quick Insights

To lower losses due to defaulting, the business should:

- Avoid individuals who pay most of their premiums by cash credit
- Further clustering of the clients according to their demographics will be pursued



Numbers are based on the test set

How this plays out in reality

1

Preventative

New clients are filtered according to business rules in order to get good clients with a lower likelihood of defaulting

Importance of preventing:-

Avoid unnecessary costs associated with setting up the insurance contract for individuals who are going to default

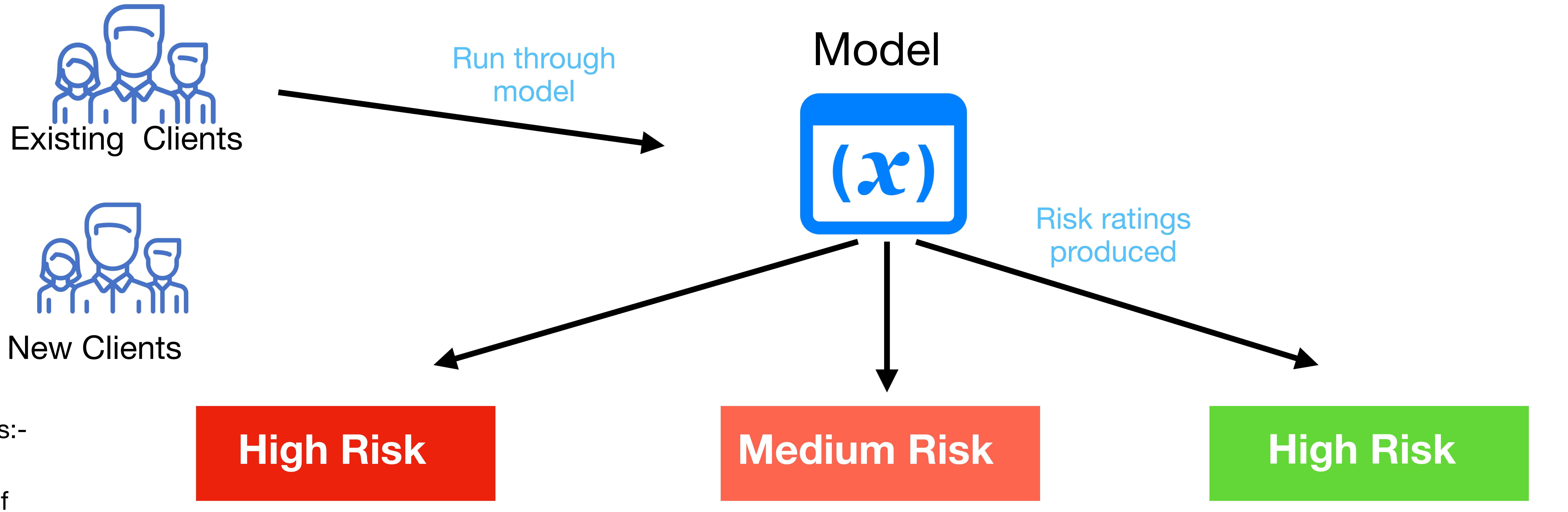
Avoid major changes to bottomline income due to unexpected lapses

2

Maintenance

Existing clients are passed through the model and are given a probability to lapse and grouped into High, Medium and Low.

These clients are then treated according to their risk to lapse profile



Apply business rules:-

- the percentage of premium paid by cash or credit should not be greater than 50%

-

Customers to be watched and possibly premiums lowered.

Get brokers or agents to call and assess health of account and consider adjusting contract or premiums

Clients to be passively monitored

Clients we can up sell to as they are least likely to leave at this point and there is less risk of default by up selling

Future optimisations:

- cluster the clients in the different groups and see the profile of clients who default and those who won't and also come up wit a marketing strategy to prevent lapse based on the client profile

- Predict the expected tenure and thus customer lifetime value of clients using a survival model so we can tier the client interventions and get the customer lifetime value of our clients

- Produce clearer customer segments to help with updating the preventative business rules