Crescent Bank

Shiqi Wang

Hao Sui

Hassan Aziz

Manjun Li

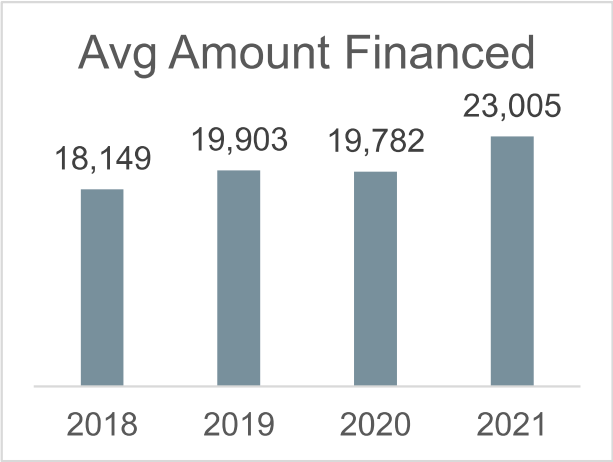
A.B. Freeman School of Business

Tulane University

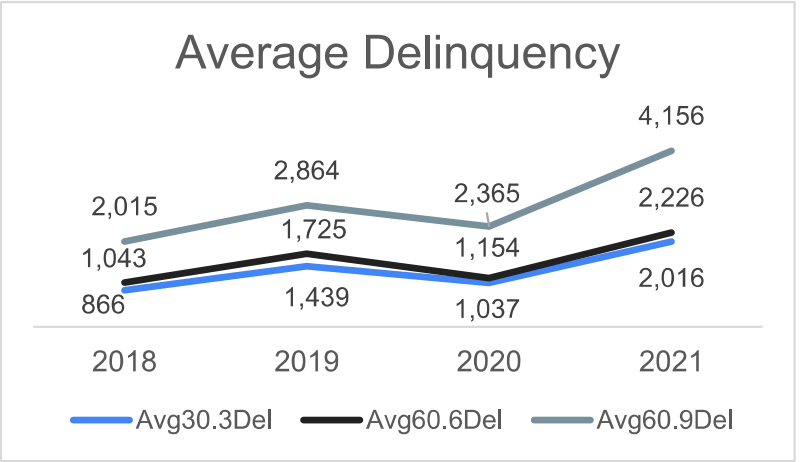
Introduction to Dataset

- The auto loan industry was affected heavily by Covid – Less Applicants in 2020
- Finance Amount increases linearly with inflation in the US
- Average Delinquency is linked with number of applicants – Less Applicants -> More scrutiny by bank
- *Takeaway -> Greater scrutiny by banks into applicants would reduce average count of delinquency*

Number of Applicants drop in 2020 due to Covid



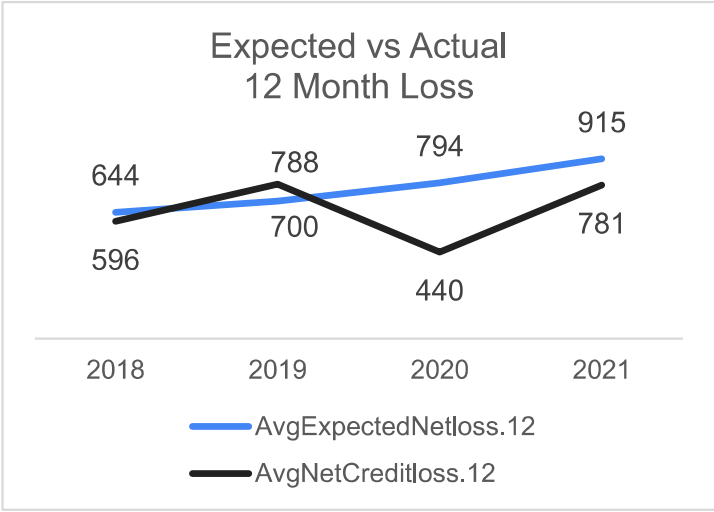
Avg Amount Financed increases each year with inflation (No effect of Covid) - *New Car inflation jumped to 5% from 2020 to 2021*



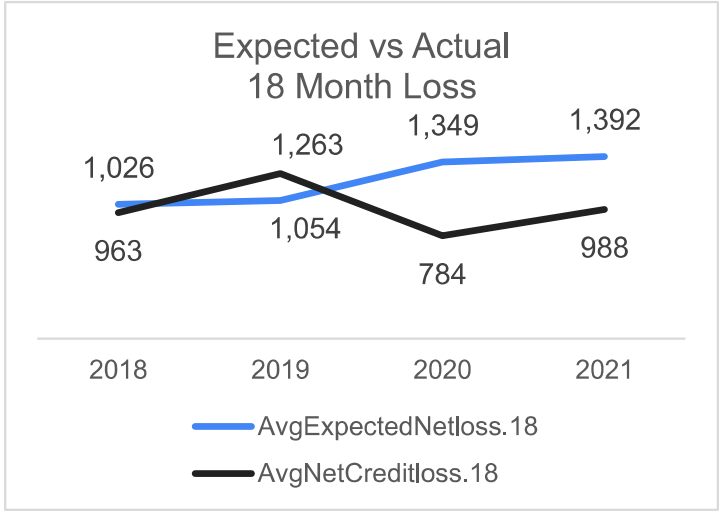
2020 has the lowest count of average delinquencies – 2021 has the highest count of average delinquencies

Introduction to Dataset

- Crescent overestimates expected Loss for all years except 2019 - Tendency to under forecast to show good performance
- Lowest Actual Losses in 2020 reinforce the idea that scrutinizing customers leads to lower losses
- If a customer is delinquent for 60 days in 6 Months, then likely to be delinquent for 60 days in 9 months too
- *Takeaway -> If a customer is delinquent for 60 days in 6 months, Crescent should start calling the customer to ensure payment (One of the problem statements expressed by bank in site visit)*



12 Month actual Loss is Lowest in Covid (2020)



18 Month actual Loss is Lowest in Covid (2020)

Variable 1	Variable 2	Correlation
Ever30.3	Ever60.6	Medium
Ever30.3	Ever60.9	Medium
Ever60.6	Ever60.9	High
Delinquency	Amount Financed	Low
Delinquency	Month	Low
Delinquency	Year	Low
Actual Loss	Amount Financed	Low
Delinquency	Actual Loss	Medium

Delinquency is correlated with predicting Actual Loss (for both 12 & 18 Months)

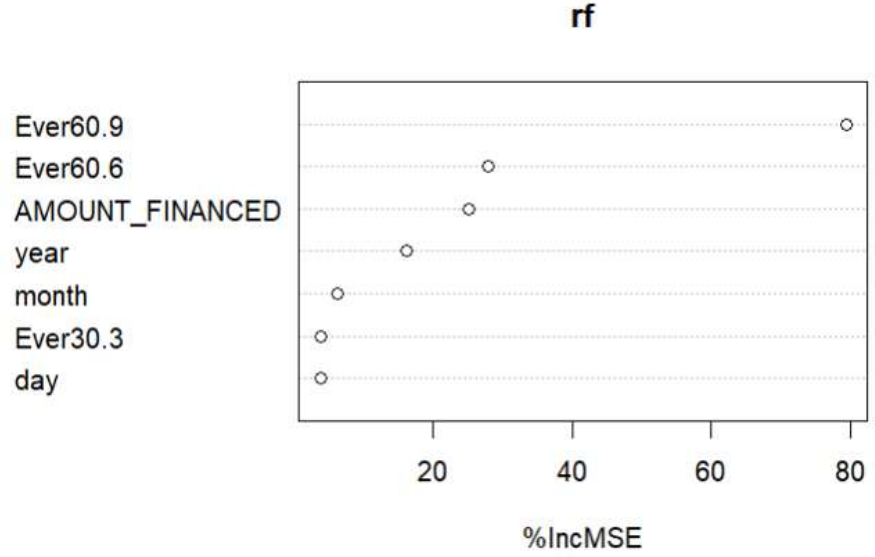
Random Forest

Random Forest Variables

Predictor variables chosen for the Random Forest were :

- Ever30.3
- Ever60.6
- Ever60.9
- Amount Financed
- Day
- Month
- Year
- Actual Loss 12 Months (for the 18 Month Model only)

Important Variable Plot for modelling 12 Month Loss



Important variables for predicting both Actual Loss 12 Months and Actual Loss 18 Months were :

1- Ever 60.9

2- Ever 60.6

3- Amount Financed

4- Actual 12 Month Loss (for the 18 Month Model Only)

Modelling for Expected Loss

To take account of Covid data, which is an anomaly, and most recent data - 3 models were compared and the one with the highest decrease in RMSE was chosen:

- Model 1 – All Dataset
- Model 2 – Excluding 2020 (Covid)
- Model 3 – Only 2021 data (Most Recent)

Actual Loss 12 Months Results

	Model 1	Model 2	Model 3
Bank RMSE	3198	3319	3684
Model RMSE	2730	2798	3291
RMSE Decrease	468	521	393

Actual Loss 12 Months – 2021 Only

	Model 1	Model 2	Model 3
Bank RMSE	3809	3652	3684
Model RMSE	3355	3230	3291
RMSE Decrease	454	422	393

Actual Loss 18 Months Results*

	Model 1	Model 2	Model 3
Bank RMSE	3633	3860	4274
Model RMSE	2826	2992	3681
RMSE Decrease	807	868	593

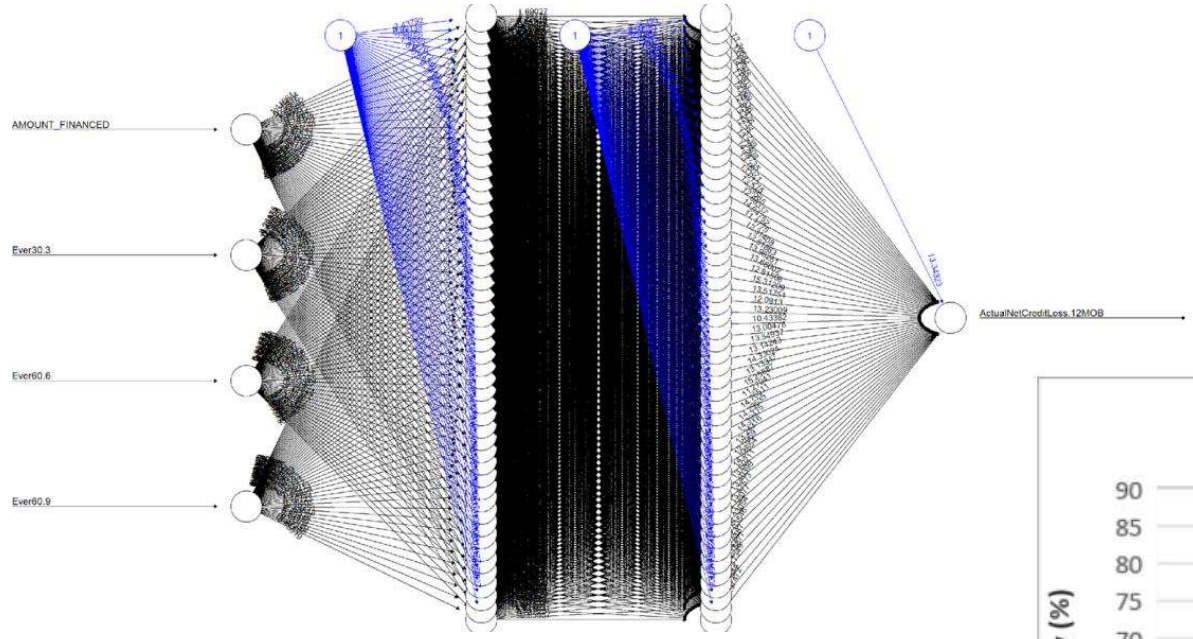
*2021 Results not compared due to lack of significant datapoints (3 Months data only)

Conclusion:

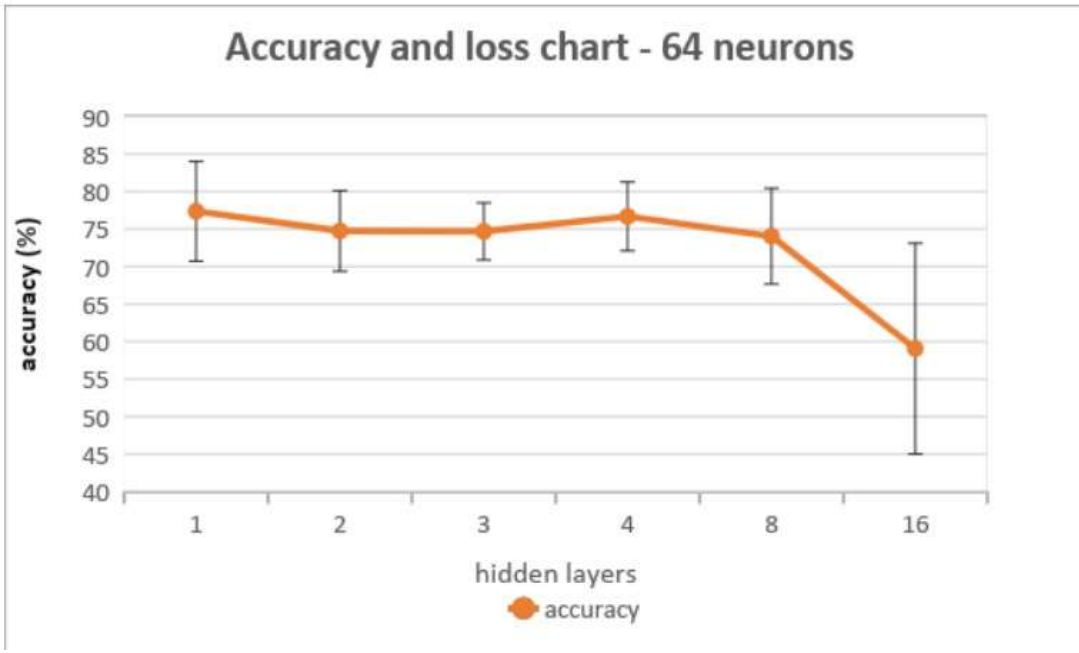
- To model losses for 12 Months – Although Model 2 predicts the dataset the best, Model 1 predicts the most recent data (2021 applicants) better -> Use Model 1
- To model losses for 18 Months -> Use Model 2 as it has the highest decrease in RMSE

Neural Network

Overfitting and Huge O(N)

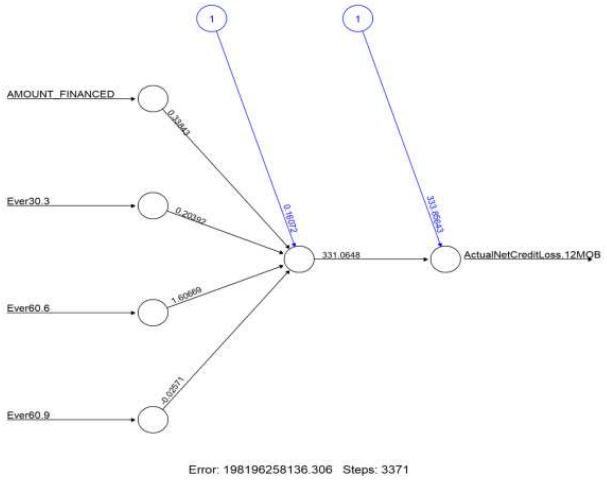


Accuracy score:
0.000000001952477



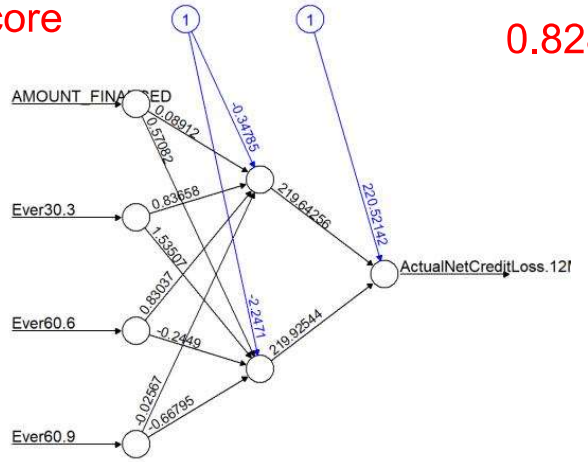
Rule of thumb:
Hidden layers < 3

One hidden layers vs. Two hidden layers



0.911 Accuracy score

```
whole.prediction_1 ExpectedNetLoss.12MOB ActualNetCreditLoss.12MOB
          37264181                42739592                 37965453
```

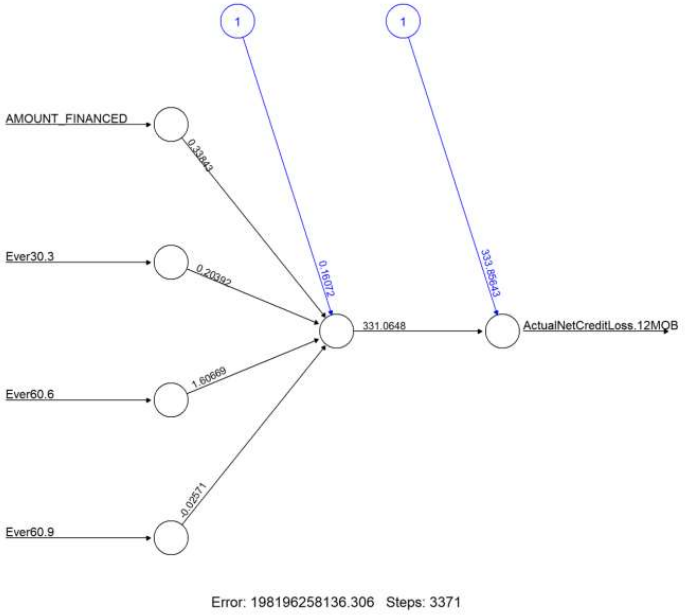


0.828 Accuracy score

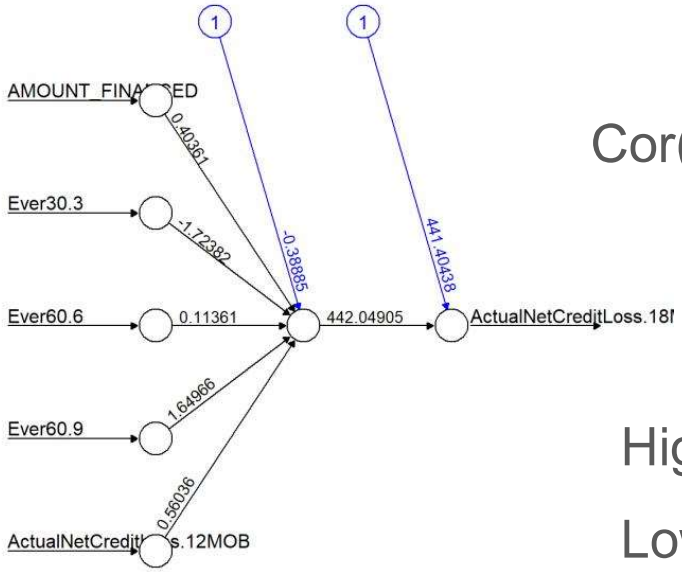
```
whole.prediction_1 ExpectedNetLoss.12MOB ActualNetCreditLoss.12MOB
          38309978                42739592                 37965453
```

	1 hidden layer	2 hidden layers
Accuracy	91.80%	82.80%

Variables for 12 Month vs. 18 Month



Error: 198196258136.306 Steps: 3371

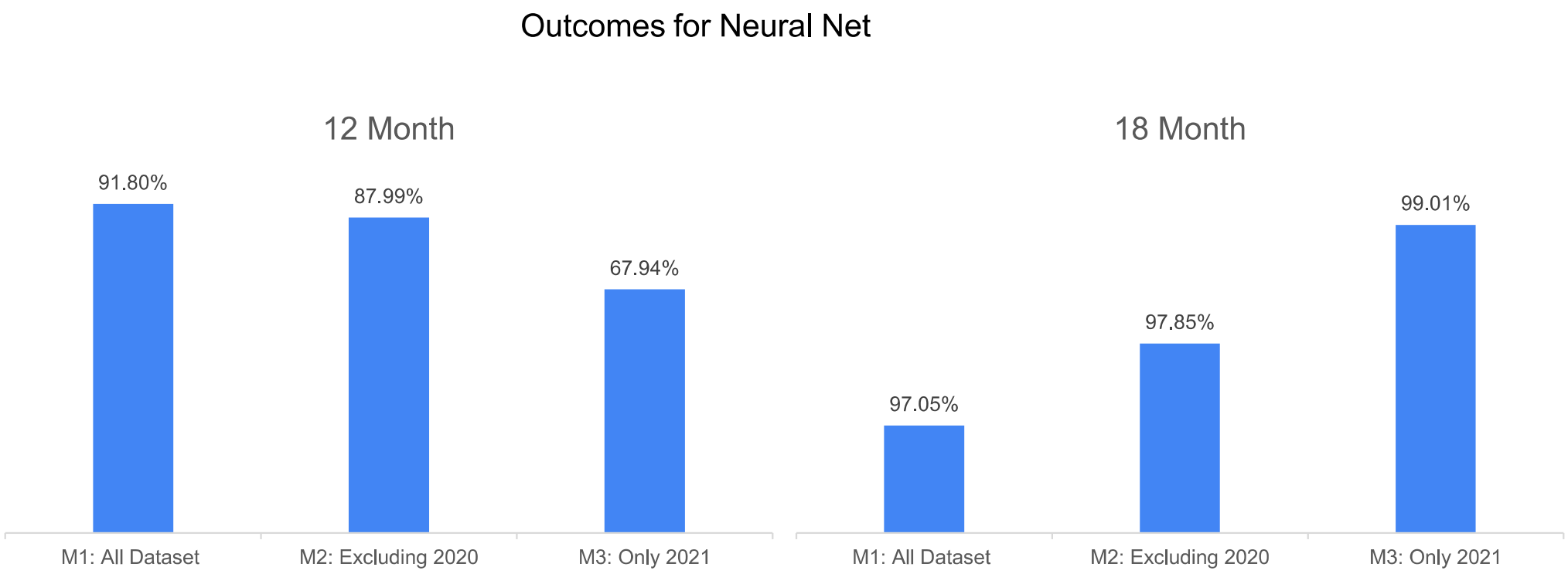


Error: 233008651594.207 Steps: 4486

Cor(12,18) = 0.536

Higher Accuracy

Lower RMSE



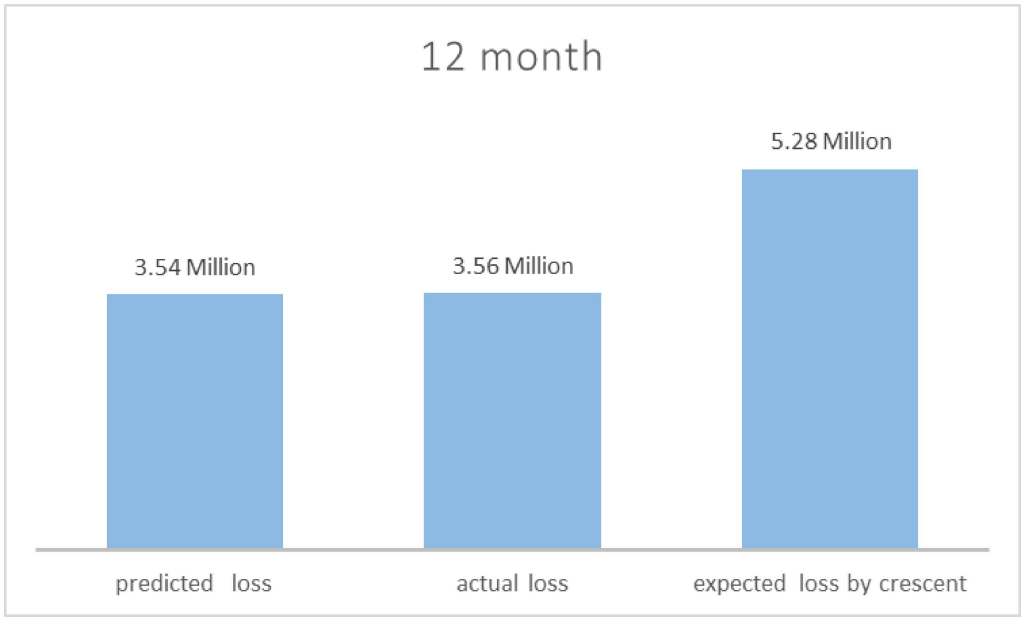
Compared to the bank model

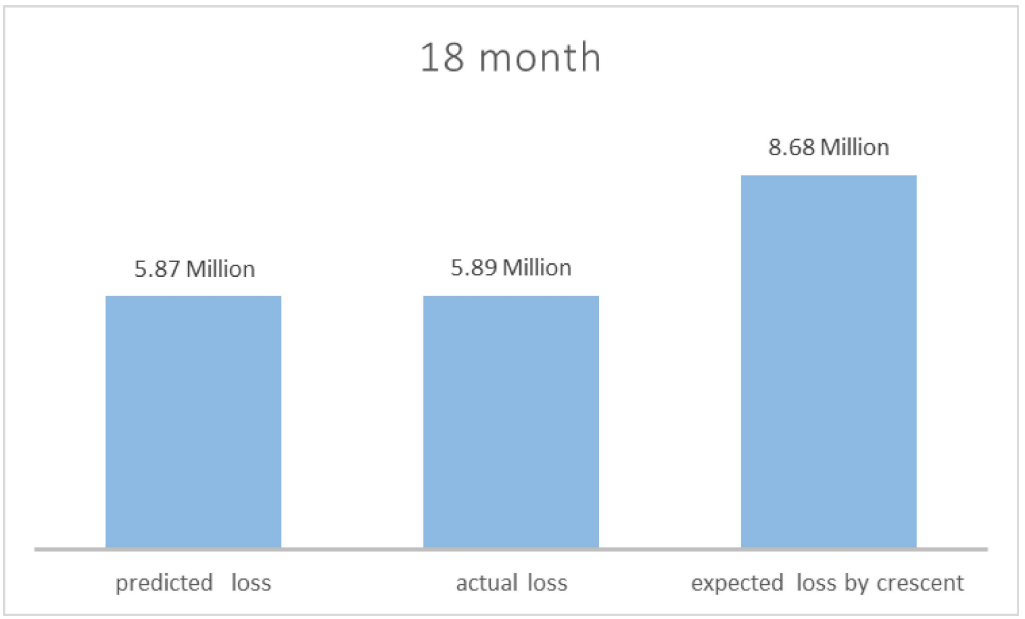
Model 2 has the best model accuracy for predicting 12 month loss, which improved 87.99% of loss
Model 3 has the best model accuracy for predicting 18 month loss

Logistic Regression &
Multiple Linear Regression

Logistic & Multiple Linear Regression

Linear Regression Outcome

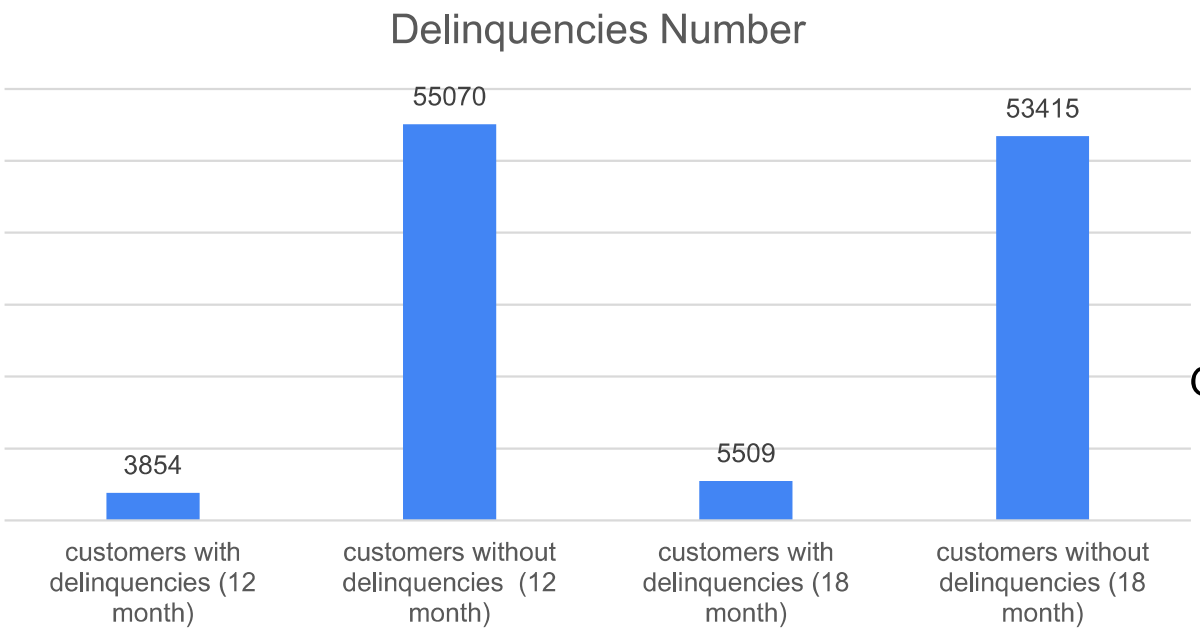




RMSE decreased 300 for 12 month and 100 for 18 month.

Logistic & Multiple Linear Regression

Compare Customers with and without Delinquencies



Delinquencies less than 10%

Goal

Predict whether the customer will have delinquencies at 12 month and 18 month

Logistic & Multiple Linear Regression

Cutoff = 0.15 for 12 month

```
Confusion Matrix and Statistics

          Reference
Prediction     0     1
         0 13445    94
         1  1323   769

               Accuracy : 0.9093
```

Cutoff = 0.16 for 18 month

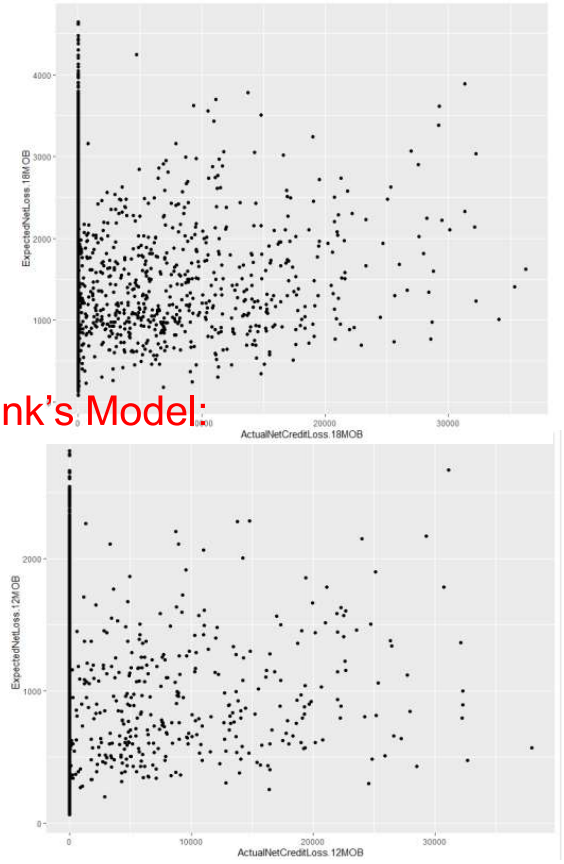
```
Confusion Matrix and Statistics

          Reference
Prediction     0     1
         0 13087   534
         1   987  1023

               Accuracy : 0.9027
```

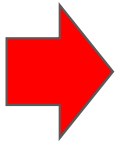
Improve the accuracy and minimize type I error at the same time.
Banks would be more likely to assume customers will have delinquencies.

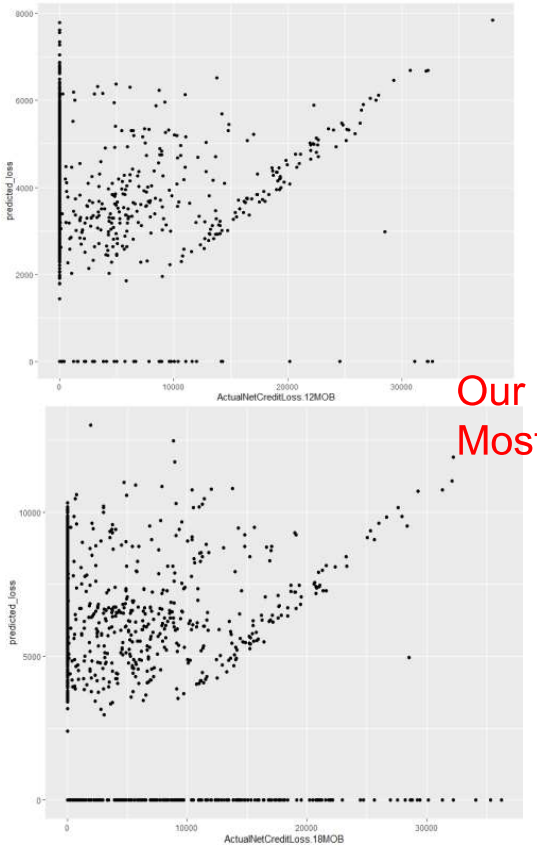
Logistic & Multiple Linear Regression

Change of 12 and 18 month

Crescent Bank's Model:
No zero.





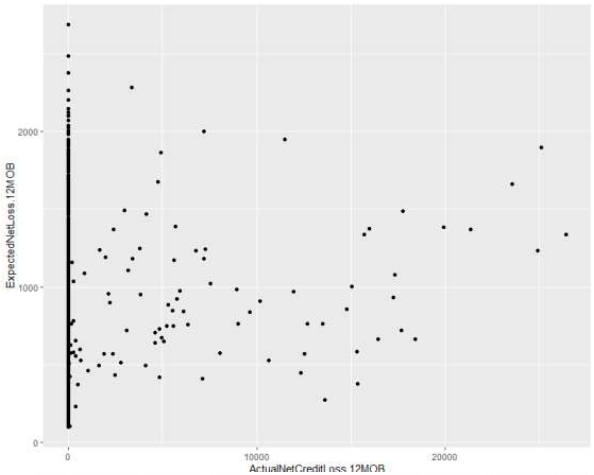


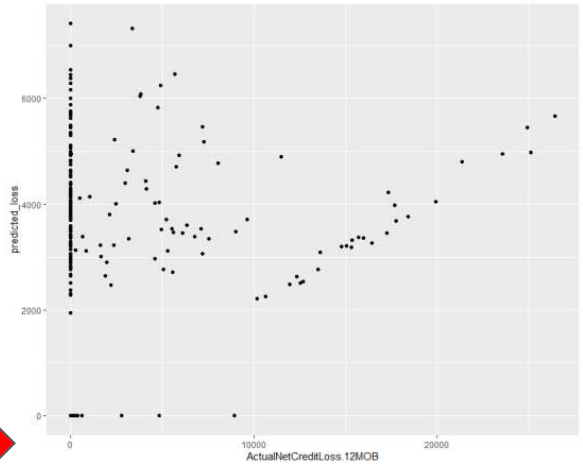
Our Model:
Most numbers are zero.

Logistic & Multiple Linear Regression

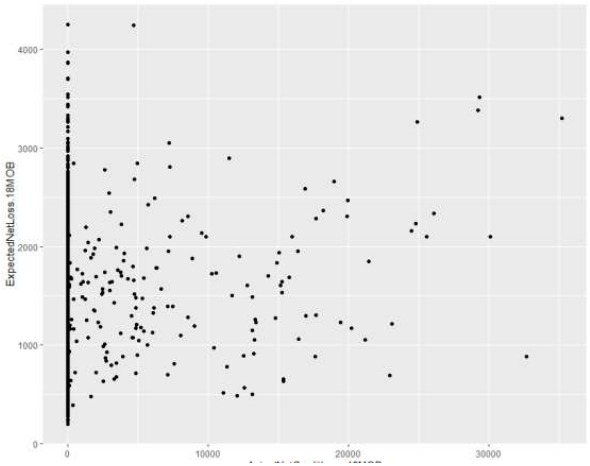
Year 2020—COVID

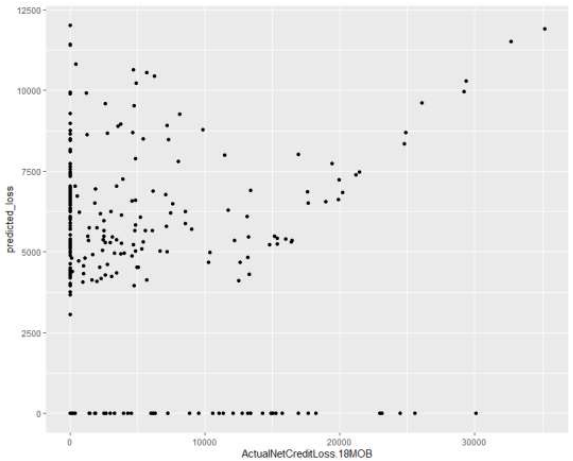
12M





18M



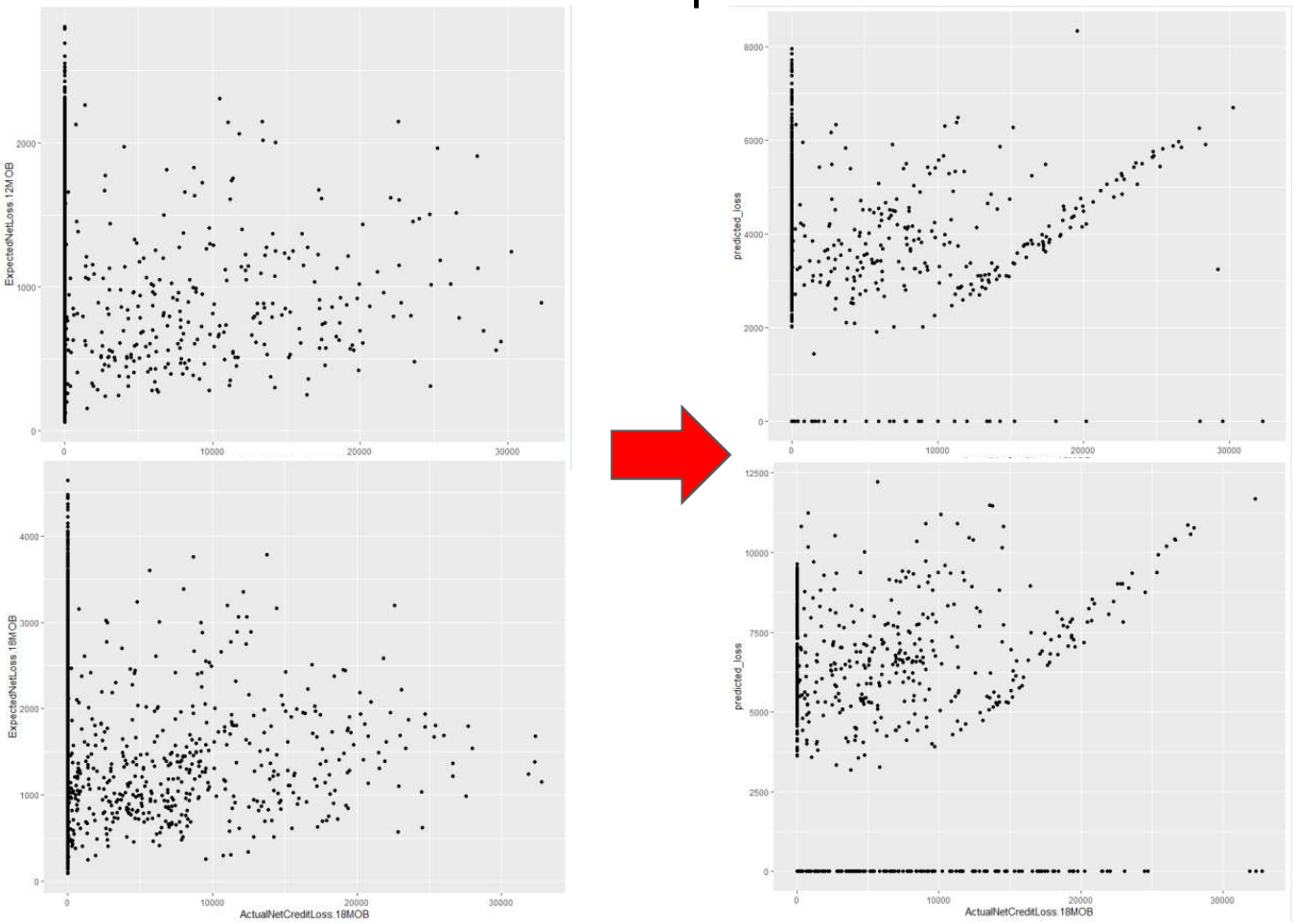


Logistic & Multiple Linear Regression

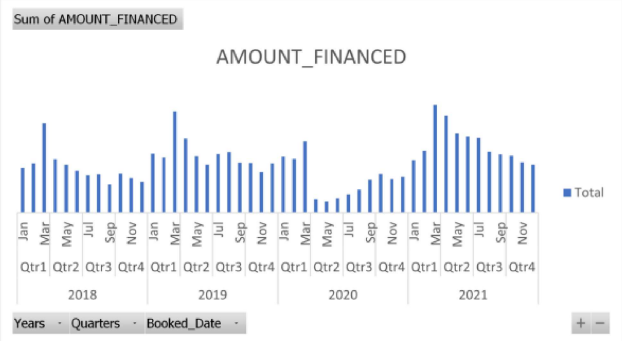
Years except 2020

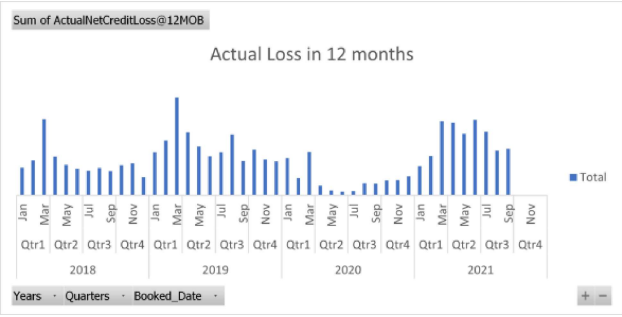
12M

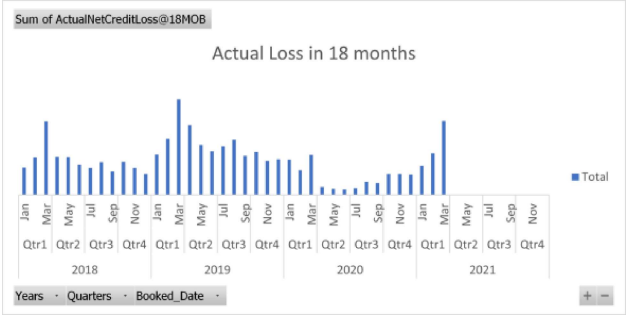
18M



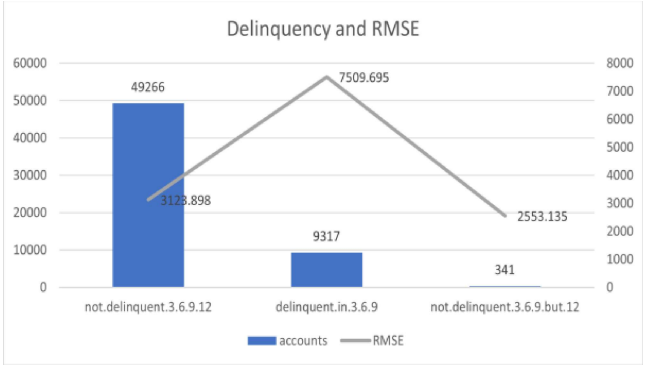
Quarter-Based Prediction



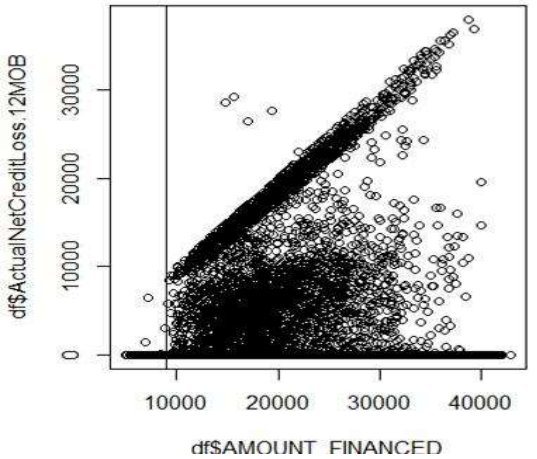


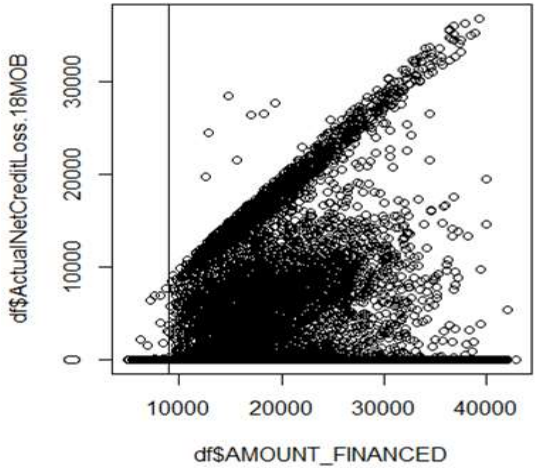


Delinquency-Based Prediction



Amount-Based Prediction





AMOUNT_FINANCED=9000

Variable Selection

Variables Selection (12 months)

	Predicted Loss	Actual Loss	Predicted Loss-Actual Loss
(Overall)			
Ever30.3	38,208,108		761,500
Ever30.3+Ever60.6	36,735,575	37,446,608	711,033
Ever30.3+Ever60.6+Ever60.9	41,414,704		3,968,096
(2021)			
Ever30.3	10,738,302		633,076
Ever30.3+Ever60.6	11,088,695	11,371,378	282,683
Ever30.3+Ever60.6+Ever60.9	14,426,533		3,055,155

Variables in 12 months:

Booked_Date+AMOUNT_FINANCED+Ever30.3+Ever60.6

Variables Selection (18 months)

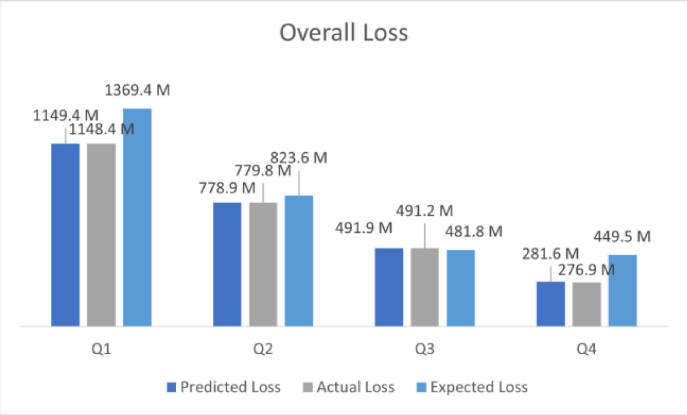
	Predicted Loss	Actual Loss	Predicted Loss -Actual Loss
(Overall)			
Ever30.3	50,099,787		3,240,493
Ever30.3+Ever60.6	49,944,233	48,892,903	3,263,093
Ever30.3+Ever60.6+Ever60.9	51,820,864		2,653,837
Ever30.3+Ever60.6+Ever60.9+12 month	49,742,252		2,576,301
(2021)			
Ever30.3	1,943,147		1,206,884
Ever30.3+Ever60.6	1,920,547	5,183,640	1,051,330
Ever30.3+Ever60.6+Ever60.9	2,529,803		2,927,961
Ever30.3+Ever60.6+Ever60.9+12 month	2,607,339		849,349

Variables in 18 months:

Booked_Date+AMOUNT_FINANCED+Ever30.3+Ever60.6
+Ever60.9+ActualNetCreditLoss.12MOB

Quarter-Based Prediction

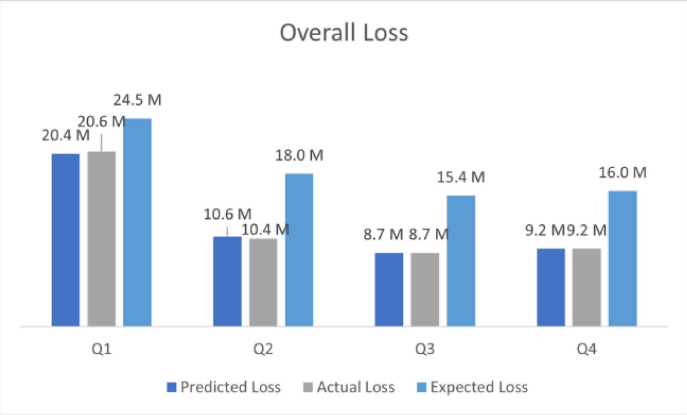
12 months



Overall Average Improved Loss: 96.86%

2021 Average Improved Loss: 92.51%

18 months



Overall Average Improved Loss: 97.57%

2021 Average Improved Loss: 92.61%

Delinquency-Based Prediction

Overall Average Improved Loss: 97.38%

2021 Average Improved Loss: 57.56%

Amount-Based Prediction

Overall Average Improved Loss: 97.21%

2021 Average Improved Loss: 74.34%

Logistic & Multiple Linear Regression

Overall Average Improved Loss: 92.81%

2021 Average Improved Loss: 65.96%

Conclusion

Compared to Crescent Bank model, all our models can significantly improve prediction accuracy.

Therefore, by computing the average of all the techniques, the prediction will be more accurate and stable.