Sacombank

CASA Cross-sell Framework

September 2022

Visa Confidential

Sacombank

Technical Document for Cross-sell Risk Scorecard

Version 2.0 – September 2022

Document Name	Sacombank CASA Cross-Sell Risk Scoring Framework
Version	2.0
Version Date	11/10/2022
Status	□ Working Draft √ Draft for Review
Author	Visa Consulting & Analytics

Version control

VERSION	DATE	CREATOR	COMMENTS
1.0	30-09-2022	Visa Consulting & Analytics	Initial draft for review by Sacombank
2.0	11-10-2022	Visa Consulting & Analytics	Initial draft for review by Sacombank

Document Distribution

NAME	TITLE	ROLE
		Sign-off authority
		Sign-off authority
		Sign-off authority

Approval

REVIEWER NAME	DATE OF REVIEW	STATUS	SIGNATURE
		Accepted / Comments	

CONTENTS

1. EXECUTIVE SUMMARY

The objective of this project is to develop "Customer-centric Cross-sell risk framework" for New to Card (NTC) customer segment of Sacombank. Sacombank will leverage the scorecards to prioritise targeting of customers for cross-selling of credit cards to existing CASA customers. Sacombank will use the scorecards in addition to internal policies and parameters for the acquisition of customers.

The new cross-sell risk strategy aims to benefit Sacombank by -

- Optimising customer targeting by using customer-level cross-sell risk scores for credit card issuance
- Increase credit card acquisitions in the market

1.1 RECOMMENDATIONS

The cross-sell risk scorecard will be added in the existing underwriting process and Sacombank will need to amend the policy documents.

The following should be considered or investigated by the Sacombank

- Assess policy rules - The existing policy rules should be assessed to determine if there are any significant changes due to the introduction of the new score.
- Limit assignment – Sacombank needs to further analyse the initial line assignment to the customers.
- Monitoring - Close monitoring is recommended of the scorecard as Sacombank is actively conducting cross-sell and the incoming CASA segment profile may have changed. Monitoring reports should be tracked from the first month of implementation.
- Monitoring - Should the scorecards experience shift upon implementations in the short term it is recommended that the cut-off be reviewed and adjusted to match desired acceptance rates and follow delinquency rates closely. In the long term, a validation of the scorecard should be considered once the scorecards have been implemented for at least 12 months.

1.2 BACKGROUND AND SCOPE

Sacombank wants to transform its customer acquisition process and focus on New-to-Card (NTC) customers. The primary objective of this engagement with Visa Consulting and Analytics is to

1) Increase credit expansion by targeting NTC customers

2) Lower risk exposures through effective targeting of the customers

This project aims to develop Cross-sell Risk scorecards for the NTC segment by using Sacombank's internally available data for customers. The scorecards will be based on 360^o view of customer's relationship with Sacombank which include data on liability accounts, credit cards and loans, casa account transactions, debit card transactions and demographics. These scores will be used to

1) Predict customer's risk of default post taking up the Credit Card

This document focuses on the technical documentation of model design for the Cross-sell Risk scorecard.

1.3 FINAL SCORECARD

1.3.1 OVERVIEW

Cross-sell Risk scorecards were developed for New-to-Card segment and customers who had at least three months of CASA relationship at the time of card opening month were included. Model development cohorts were created based on sampling from 12 different snapshots of time and 2 other cohorts were selected for out-of-time validation. The final model of choice is lasso logistic regression available in Python.

Table 1 details the final scorecard for new-to-card customers and the beta coefficient for each of the predictors is provided in Table 2.

Table 1: Cross-sell Risk Scorecards for NTC customers

Customer Segment	**NTC with at least 3 months of CASA on book**
Average Default Rate	1.1%
Lift in top 10%	3.19
Discriminatory power	KS = 0.37 on test set, 0.36 on validation set
Direction	Higher score = Higher probability of default in the next 12 months following card opening date

Table 2. CASA segment scorecard characteristics

No.	Predictor	Description	Beta Coefficient
1	num_of_productholdings_d	Number of Product Holdings as of cohort month	-0.6142
2	mon_average_balance_qtr0	Monthly Average Balance in cohort month	-0.624554
3	mon_average_balance_qtr2	Monthly Average Balance in L2Q	-0.310559
4	mon_average_balance_qtr3	Monthly Average Balance in L3Q	-0.286934
5	earliest_casa_vintage_to_cohort_d	Earliest CASA account vintage to cohort month (in years)	-0.336656
6	earliest_casa_before_active_inyears_d	Earliest CASA account activation date from open date (in years)	-0.506613
7	length_of_employment_lessthan5yrs	Length of employment less than 5 years	-0.751471
8	avg_remaining_tenure_d	Average remaining tenure of loans	-0.717041
9	avg_interest_rate_d	Average interest rate of loans	-0.895748
10	mon_avg_bal_decrease_count_d_qtr1	Number of times monthly average balance decrease in L1Q	0.308668
11	tot_debit_count_ratio_l3m	Ratio of total debit count to total credit and debit count in L3M from cohort month in CASA account	-0.618776
12	tot_debit_amt_ratio_l3m	Ratio of total debit amount to total credit and debit amount in L3M from cohort month in CASA account	-0.424831
13	num_of_mths_debit_count_l3m_qtr0	Debit transaction flag in cohort month	0.95086
14	num_of_mths_credit_count_l3m_qtr0	Credit transaction flag in cohort month	-0.609936
15	num_of_mths_credit_gt_debit_amt_l3m_qtr1	Number of months with credit greater than debit amount in L1Q	-0.549035
16	tot_credit_count_qtr1	Total credit count in L1Q	1.721773
17	qtr_end_bal_ratio_l3mp3m_d	Ratio of Month end balances in L3M to Month end balance in P3M	-0.536562
18	qtr_end_bal_ratio_l6mp6m_d	Ratio of Month end balances in L6M to Month end balance in P6M	-0.433537
19	qtr0_mon_avg_over_end_bal	Ratio of Monthly average balance to month end balance in cohort month	-1.004293
20	(Intercept)	-	-2.21742676

*LxQ: last x quarter(s). i.e. L1Q means last 1 quarter, given an observation date of Jul 2022, L1Q is April to June 2022

*LxM: last x month(s). i.e. L3M means last 3 months, given an observation date of Jul 2022, L3M is April to June 2022

*PxM: past x month(s). i.e. P3M is the previous 3 months before L3M, given an observation date of July 2022, P3M is Jan to Mar 2022

1.3.2 PERFORMANCE & VALIDATION

Table 3 below shows the various performance metrics of the scorecard. The validation result conducted on out-of-time test set indicates relative stability of the model. It is recommended that Sacombank conduct independent out-of-time validation of the scorecard after implementation as well.

Table 3. Performance & validation of the scorecard

Performance metric	Train Set	Validation	Test Set
KS	0.36	0.36	0.37
ROC	0.75	0.74	0.74
PSI	-	0.008	0.005

1.4 LIMITATIONS OF MODEL AND DATA

Sacombank should consider the following scorecard limitations –

1. Risk scorecards have been developed using sampled customer data, which was representative of the total population for the duration of Feb 2019 to March 2022.
2. The model relies on Sacombank extracted tables and data fields for scoring and assumes the availability of data format and extraction logic in sustainable form.
3. Data availability also limits the features that can be used as inputs for the model. For example, missing rate of income from the customer demographics table was at ~50%, and channel type of CASA transactions (branch, ATM, online, mobile banking etc) were not available for modelling.
4. The scorecards have been developed on historical information of customers and are vulnerable to changing customer behaviour with time and any shift in portfolio trends.
5. The models are developed in context of current business and operating model of Sacombank and should be monitored with changes in operating scenario.
6. Regulatory changes and/or economic shifts might impact the model performance and should be closely monitored.
7. As the scorecards use machine learning techniques, they may lack the ability to generalize conditions and direct causality effects.

1.4.1 MODEL PERFORMANCE MONITORING

Visa recommends close monitoring of the scorecards to be conducted by Sacombank for any shift in performance.

Sacombank should be evaluating the models on below criteria –

1. Discriminatory power – KS statistic
2. Accuracy of prediction – Predicted vs Actual bads
3. Stability of the model – Population stability index and validation on latest time period

Sacombank may use the following guidelines for monitoring and evaluation of models –

Kolmogorov – Smirnov (KS) Statistic	
If change in KS statistic is +/- 5%	The model is performing as expected
If change in KS statistic is +/- 5-15%	Sacombank should investigate the cause and take action accordingly
If change in KS statistic is >15%	Sacombank should consider recalibrating/redeveloping the models

Population stability index	
If PSI < 0.1	No change, the model is performing as expected
If PSI >= 0.1 and < 0.2	Sacombank should investigate the cause and take action accordingly
If PSI >= 0.2	Sacombank should consider recalibrating/redeveloping the models

2. MODEL APPROACH

This section provides an overview of model design and framework which discusses on data sources, data quality checks, model segmentation and sampling design.

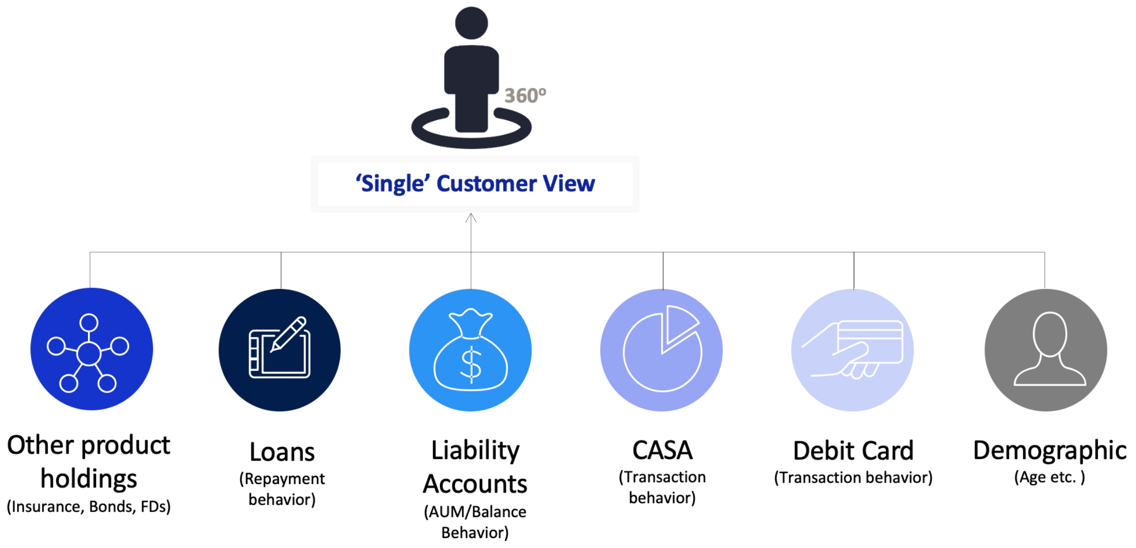
2.1 DATA SOURCES

The model leverage on Sacombank data only, collected from multiple sources as listed below.

1. CASA Account: Historical records from Feb 2019 to March 2021
2. Customer demographics: Historical records from Feb 2019 to March 2021
3. Product Holdings: Historical records from Feb 2019 to March 2021
4. Loan Asset: Historical records from Feb 2019 to March 2022
5. Card characteristics: Historical records from Feb 2019 to March 2022
6. Debit transaction: Historical records from Feb 2019 to March 2021
7. Credit transaction: Historical records from Feb 2019 to March 2021
8. Billing data: Historical records from Feb 2019 to March 2021

Note that datasets on debit transaction are not used in the final model although it was used during the modelling process and iterations. Also, credit transactions and billing data are excluded since the model objectives of targeting non-carded customers would result in information leakage.

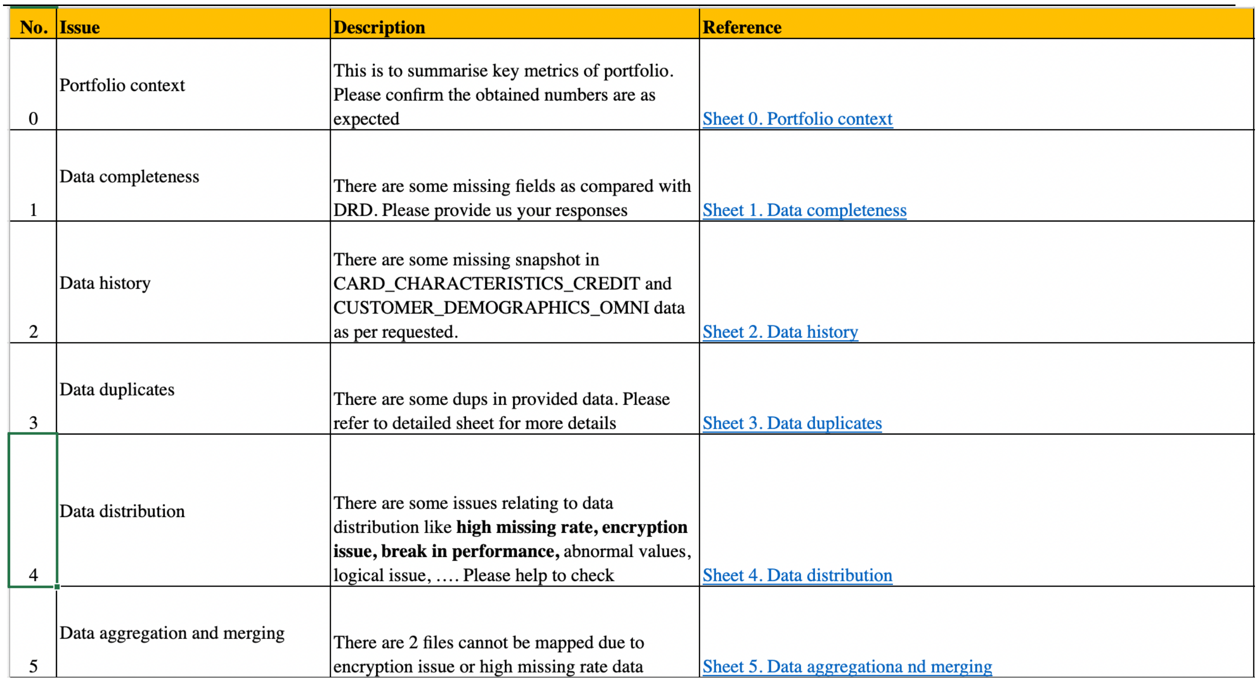
Based on the provided datasets, various information can be derived to understand customer behaviour and banking relationship.



2.2 DATA QUALITY & INTEGRITY

While working through all the components of the model design; the data used (and manipulated) needs to be tested to ensure the integrity of the information used to develop the scorecards. A detailed level of quality of data was conducted. As a process, all the data validation results were shared with Sacombank and confirmation on data was shown in July 2022. (Appendix A)

Figure 2.2.1: DIDQ Report sample output for data sources



No.	Issue	Description	Reference
0	Portfolio context	This is to summarise key metrics of portfolio. Please confirm the obtained numbers are as expected	Sheet 0. Portfolio context
1	Data completeness	There are some missing fields as compared with DRD. Please provide us your responses	Sheet 1. Data completeness
2	Data history	There are some missing snapshot in CARD_CHARACTERISTICS_CREDIT and CUSTOMER_DEMOGRAPHICS_OMNI data as per requested.	Sheet 2. Data history
3	Data duplicates	There are some dups in provided data. Please refer to detailed sheet for more details	Sheet 3. Data duplicates
4	Data distribution	There are some issues relating to data distribution like **high missing rate, encryption issue, break in performance,** abnormal values, logical issue, …. Please help to check	Sheet 4. Data distribution
5	Data aggregation and merging	There are 2 files cannot be mapped due to encryption issue or high missing rate data	Sheet 5. Data aggregationa nd merging

In conclusion, overall good data quality is seen at a customer level. There are also some data limitations and future improvement areas for consideration:

1. Limited breadth of information on CASA accounts
    a. Data provided were aggregated at a customer-account level, thus there may be loss of information from channel type (Online, Mobile, ATM, Branch etc.) which are typically stronger predictors for default labelling.
2. Low coverage for debit transaction, product holdings and loans table
    a. For new-to-card customers, only 32% of them have debit transactions, 36% with loans and 16% with other product holdings.
3. Small sample size and low bad rate
    a. After defining new-to-card customers, the available customer base for model development was less than 50k customers and the bad rate was at 1.1%.

2.2 DEFINING NTC CUSTOMER BASE

Defining the New-to-Card segment of the customer base is necessary to ensure that the customer profile used for modelling is representative of potential group of new customers Sacombank would be issuing cards to. This also minimises overfitting of the model if existing-to-card customers were to be included in the scorecard development.

Model development population – NTC Customers

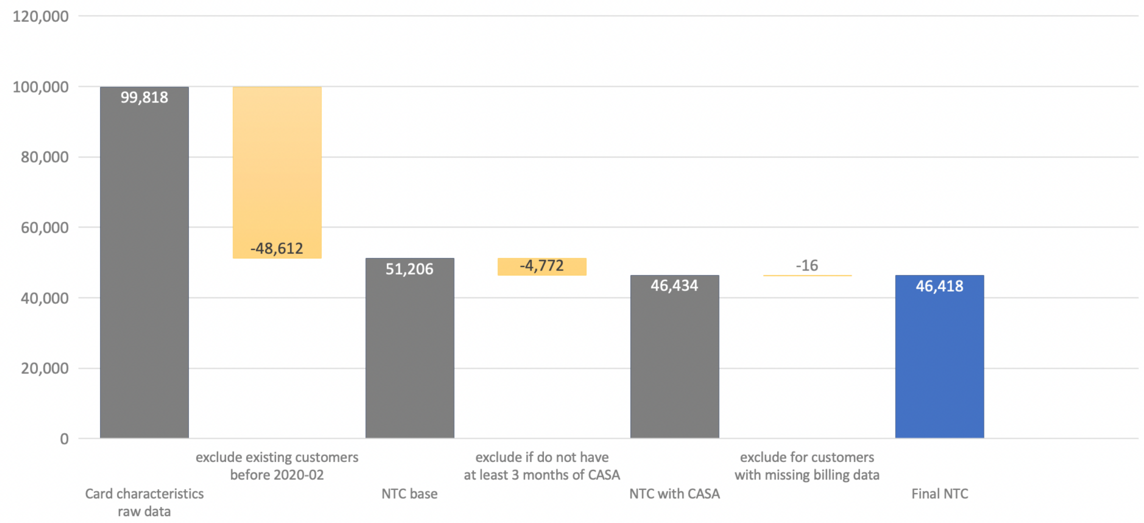
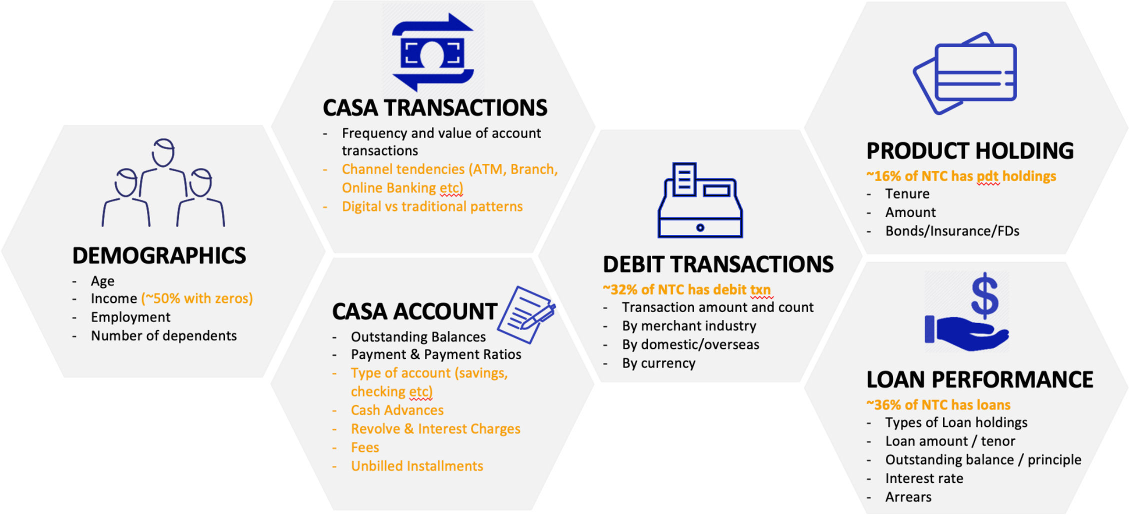


Figure 2.2.1. Waterfall chart describing the breakdown of new-to-card customers derived from the card characteristics table provided by Sacombank from the period of February 2019 to March 2022

2.3 DATA PREPARATION

2.3.1 FEATURES FROM DATA SOURCES

Customer demographics, CASA balances and transaction, product holdings, debit card transaction and loans data were extracted by Sacombank and shared. Listed below are the data streams used for model development and sample features associated with the datasets.



CASA TRANSACTIONS
- Frequency and value of account transactions
- Channel tendencies (ATM, Branch, Online Banking etc)
- Digital vs traditional patterns

PRODUCT HOLDING
~16% of NTC has pdt holdings
- Tenure
- Amount
- Bonds/Insurance/FDs

DEMOGRAPHICS
- Age
- Income (~50% with zeros)
- Employment
- Number of dependents

DEBIT TRANSACTIONS
~32% of NTC has debit txn
- Transaction amount and count
- By merchant industry
- By domestic/overseas
- By currency

CASA ACCOUNT
- Outstanding Balances
- Payment & Payment Ratios
- Type of account (savings, checking etc)
- Cash Advances
- Revolve & Interest Charges
- Fees
- Unbilled Installments

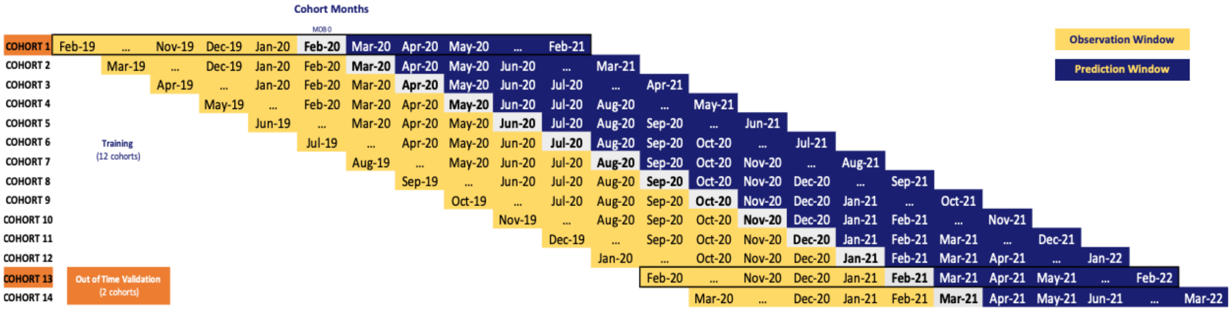
LOAN PERFORMANCE
~36% of NTC has loans
- Types of Loan holdings
- Loan amount / tenor
- Outstanding balance / principle
- Interest rate
- Arrears

Figure 2.3.1. Data sources used for creating a 'Single Customer View' and limitations (highlighted in orange)

The details of final features used in the model and corresponding derivation logic from the data sources shown above can be found in the excel sheet (sacombank_feature_documentation_202209.xlsx) accompanying this document.

2.3.2 TIMEFRAMES

Observation and prediction windows of 12 months were used. Training and validation data contained observations on a sliding window basis from different time snapshots. Two out-of-time cohorts (Feb 2020 and Feb 2021) were used for test purposes to capture the pre- and post-covid behavioural changes that customers may display.



Cohort Months

						MOB 0																		
COHORT 1	Feb-19	...	Nov-19	Dec-19	Jan-20	Feb-20	Mar-20	Apr-20	May-20	...	Feb-21													
COHORT 2		Mar-19	...	Dec-19	Jan-20	Feb-20	Mar-20	Apr-20	May-20	Jun-20	...	Mar-21												
COHORT 3			Apr-19	...	Jan-20	Feb-20	Mar-20	Apr-20	May-20	Jun-20	Jul-20	...	Apr-21											
COHORT 4				May-19	...	Feb-20	Mar-20	Apr-20	May-20	Jun-20	Jul-20	Aug-20	...	May-21										
COHORT 5					Jun-19	...	Mar-20	Apr-20	May-20	Jun-20	Jul-20	Aug-20	Sep-20	...	Jun-21									
COHORT 6						Jul-19	...	Apr-20	May-20	Jun-20	Jul-20	Aug-20	Sep-20	Oct-20	...	Jul-21								
COHORT 7							Aug-19	...	May-20	Jun-20	Jul-20	Aug-20	Sep-20	Oct-20	Nov-20	...	Aug-21							
COHORT 8								Sep-19	...	Jun-20	Jul-20	Aug-20	Sep-20	Oct-20	Nov-20	Dec-20	...	Sep-21						
COHORT 9									Oct-19	...	Jul-20	Aug-20	Sep-20	Oct-20	Nov-20	Dec-20	Jan-21	...	Oct-21					
COHORT 10										Nov-19	...	Aug-20	Sep-20	Oct-20	Nov-20	Dec-20	Jan-21	Feb-21	...	Nov-21				
COHORT 11											Dec-19	...	Sep-20	Oct-20	Nov-20	Dec-20	Jan-21	Feb-21	Mar-21	...	Dec-21			
COHORT 12												Jan-20	...	Oct-20	Nov-20	Dec-20	Jan-21	Feb-21	Mar-21	Apr-21	...	Jan-22		
COHORT 13													Feb-20	...	Nov-20	Dec-20	Jan-21	Feb-21	Mar-21	Apr-21	May-21	...	Feb-22	
COHORT 14														Mar-20	...	Dec-20	Jan-21	Feb-21	Mar-21	Apr-21	May-21	Jun-21	...	Mar-22

Figure 2.3.2 Modelling window design

Distribution of good and bad labels by train, validation and test sets are as follows:

Assigned Set | Target Label	Bad	Good
Train	316	28725
Validation	152	12295
Test	47	4883

2.3.3 BAD DEFINITION

Vintage Analysis was performed to understand the maturity of credit card portfolio with age and establish the prediction window. Roll Rates of various bad definitions were also investigated to identify the proportion of accounts that eventually are written off in 12 months. Also, the proportion of eventual write-offs that would be captured by the bad definition was calculated. Vintage and roll rate analysis output are shown in Appendix B.

The ideal bad definition is one that captures a high percentage of eventual write offs without misclassifying good accounts as bad.

Bad definition was agreed with Sacombank.

The target definition chosen for bad:
Customer is classified as 'bad' if ever 30+ DPD on new credit cards in the next 12 months from card opening date.

There are no in-determinates in this definition. All accounts are classified as 'bad' or 'good'. The table below provides volume of cases identified as 'Good', 'Bad'.

Target label	**Good (label: 0)**	**Bad (label: 1)**	**Default rate (%)**
Cohort month			
2020-02	3455	31	0.89
2020-03	3610	47	1.29
2020-04	1108	10	0.89
2020-05	2337	29	1.23
2020-06	2822	26	0.91
2020-07	4943	46	0.92
2020-08	5001	58	1.15
2020-09	3815	66	1.70
2020-10	3795	48	1.25
2020-11	3731	37	0.98
2020-12	3137	34	1.07
2021-01	2740	26	0.94
2021-02	1428	16	1.11
2021-03	3981	41	1.02
Total	**45903**	**515**	
Grand Total	**46418**		**Average 1.10%**

Table 2.3.3.1 Distribution of "good" vs "bad" labels for NTC customers by cohort month

2.3.4 EXCLUSIONS

Some groups of accounts should be excluded from scorecard development in order to avoid potential biases in the final scorecard. Accounts which are excluded from the development are classified as such:

Observation Exclusions - Accounts that will not be evaluated using the scorecard or for which accurate historical data is not available. The main observation exclusions are customer who do not have CASA vintage of minimum 3 months on book.

2.4 TARGET OVERSAMPLING

In order to bias the classification of rare event, which is "bad" customers in this model, we over-sample the rare event. In the training sample, we put a higher proportion of rare-event observations than the proportion that exists in the actual population.

Below are the target rates used for training the models –

Segment	**Training data – Target classification rate**
NTC Customers	10%

3. MODEL DEVELOPMENT AND RESULTS

3.1 FINE CLASSING

Fine Classing involves analysing the data at a granular level. It is the first step in the two-step process of classing. Fine classing assists in determining how each characteristic is represented in the scorecard.

For continuous variables, a maximum of 20 equally sized classes was created; for discrete variables one class per value was created (unless there were many possible values – in this case grouping values based on good / bad odds was done). Filtering criteria was done using information value (IV) where the formula is as such. Fine IV threshold was set at 0.01 thus features with IV less than this value is removed as they are very weak predictors.

$$Information\,Value\;(IV) = \;\ln\left(\frac{\%good}{\%bad}\right) - (\%good - \%bad)$$

3.2 COARSE CLASSING

Coarse classing is the aggregation of the data into stable and predictive groups. The predictive strength of the characteristic is measured using the information value (IV), both pre and post coarse classing. This forms a monotonic bad rate trend in relation to the target variable.

Table below shows the Fine and Coarse IV by segment for all characteristics included in the scorecard.

Table 3.2.1 Fine and Coarse IV

No.	Predictor	Coarse IV	Fine IV
1	length_of_employment_lessthan5yrs	0.092572	0.092572
2	avg_remaining_tenure_d	0.029611	0.055899
3	avg_interest_rate_d	0.065859	0.102998
4	num_of_productholdings_d	0.209423	0.164772
5	earliest_casa_vintage_to_cohort_d	0.109082	0.108255
6	earliest_casa_before_active_inyears_d	0.07355	0.083018
7	mon_average_balance_qtr0	0.519631	0.597256
8	mon_average_balance_qtr2	0.362446	0.408225

9	mon_average_balance_qtr3	0.329444	0.335944
10	mon_avg_bal_decrease_count_d_qtr1	0.065805	0.067702
11	qtr_end_bal_ratio_l3mp3m_d	0.04532	0.104464
12	qtr_end_bal_ratio_l6mp6m_d	0.041141	0.102698
13	num_of_mths_credit_count_l3m_qtr0	0.061764	0.061764
14	num_of_mths_debit_count_l3m_qtr0	0.065986	0.065986
15	num_of_mths_credit_gt_debit_amt_l3m_qtr1	0.045144	0.052942
16	tot_credit_count_qtr1	0.049132	0.059252
17	tot_debit_count_ratio_l3m	0.115223	0.189179
18	tot_debit_amt_ratio_l3m	0.148972	0.080539
19	qtr0_mon_avg_over_end_bal	0.080044	0.224585

IV drops after coarse classing, as fine class characteristics have less irregular pattern in bad rate.

3.3 VARIABLE CLUSTERING

Variable clustering is performed as another feature reduction technique. Cluster analysis groups features that have high explanatory power within the cluster and low outside the cluster. Features that have high R-square within its cluster and high IV values are selected. Additional expert judgement is also part of the model development process to include features that are believed to have good variance and explainability.

3.4 TUNING PARAMETERS IN LOGISTIC REGRESSION MODEL

This section specifies the input parameters to the logistic regression model, which were used in the final models.

Parameter	Set value
Penalty	L1
Solver	Lib-linear
C	1
Random state	0

After fitting the logistic regression model to the selected features, an additional p-value check was done to ensure that the features are statistically significant (p-value < 0.05).

3.5 BINARY CLASSIFICATION

Logistic regression model was used to build the final scorecard with weights of evidence (WOE) as the independent variables. The scorecard uses the weight of evidence (WOE) of the coarse classed characteristics to predict the target label.

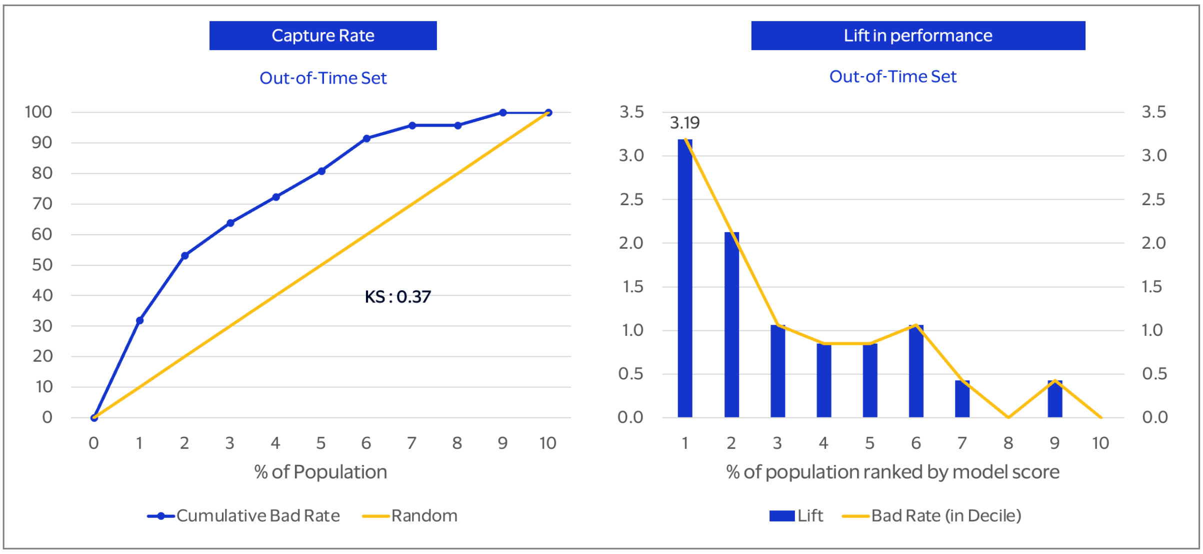
$$WOE = ln\left(\frac{\%good}{\%bad}\right)$$

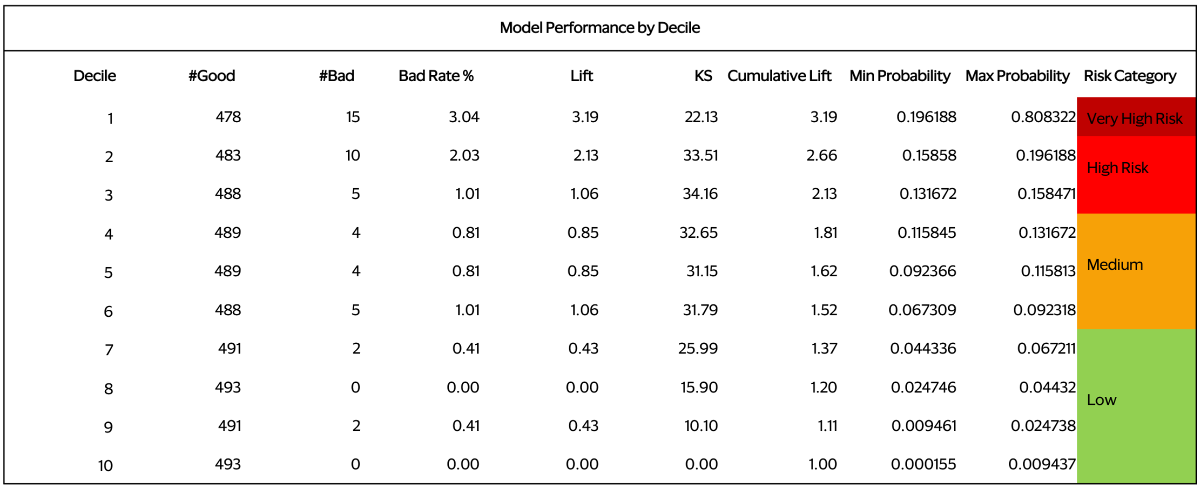
Breakdown of the selected features and their corresponding WOE transformation values are in a separate excel document (coarse_bins_selected_features.csv) as shared with Sacombank.

3.6 MODEL VALIDATION

MODEL VALIDATION AT VISA

The model was validated on out-of-time sample of NTC accounts for the cohort months of Feb 2020 and Feb 2021. Overall the model shows good separation strength from the good and bad labels.





Model Performance by Decile

Decile	#Good	#Bad	Bad Rate %	Lift	KS	Cumulative Lift	Min Probability	Max Probability	Risk Category
1	478	15	3.04	3.19	22.13	3.19	0.196188	0.808322	Very High Risk
2	483	10	2.03	2.13	33.51	2.66	0.15858	0.196188	High Risk
3	488	5	1.01	1.06	34.16	2.13	0.131672	0.158471	
4	489	4	0.81	0.85	32.65	1.81	0.115845	0.131672	Medium
5	489	4	0.81	0.85	31.15	1.62	0.092366	0.115813	
6	488	5	1.01	1.06	31.79	1.52	0.067309	0.092318	
7	491	2	0.41	0.43	25.99	1.37	0.044336	0.067211	Low
8	493	0	0.00	0.00	15.90	1.20	0.024746	0.04432	
9	491	2	0.41	0.43	10.10	1.11	0.009461	0.024738	
10	493	0	0.00	0.00	0.00	1.00	0.000155	0.009437	

Table 3.4.1 KS Statistics on out-of-time test set

As shown in the KS table above, the lift at the top decile after ranking customers by their predicted probability is at 3.19. This means that the bad rate at the top decile is 3.19 times higher than the average bad rate across all deciles. Correspondingly, we can divide the deciles into 4 different risk categories, and target the customers in the low risk category for cross selling of new cards.

MODEL BACKTESTING BY SACOMBANK TEAM

Additionally, Sacombank team also conducted independent model backtesting.

---------------------- Results to be added by Sacombank team -----------------------------

4. APPENDIX

4.1 APPENDIX A: DATA QUALITY & INTEGRITY REPORT

Detailed report shared with Sacombank

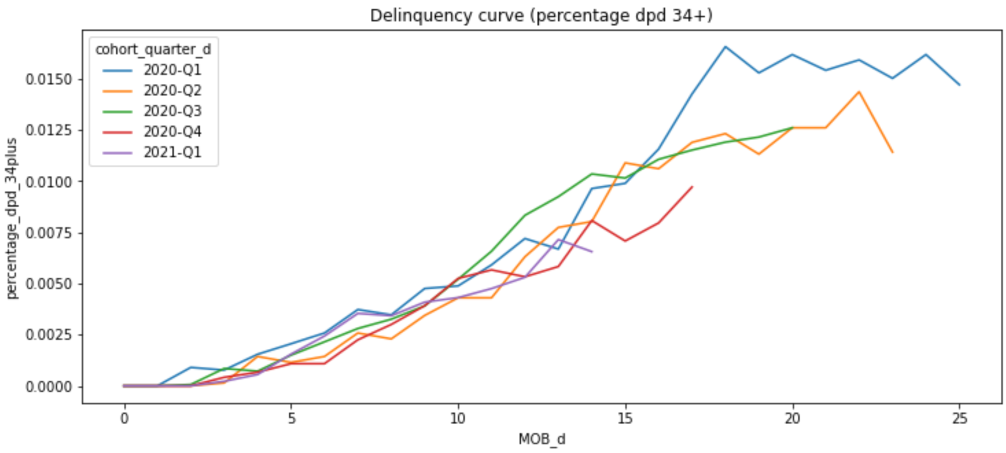
- STB_Data Validation Result_Full_Data_v1 (version 1)_from STB 31.07.2022.xlsx

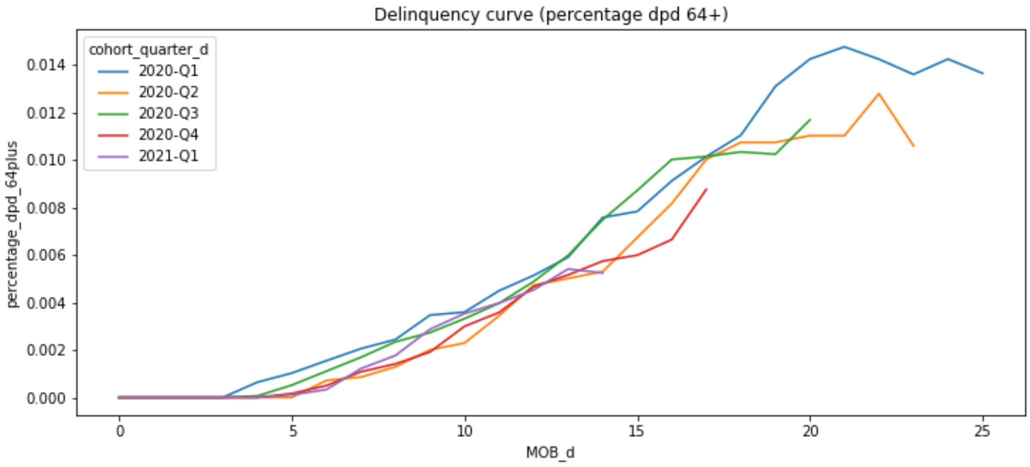
4.2 APPENDIX B: VINTAGE & ROLL RATE ANALYSIS

4.2.1 Roll Rate Analysis for NTC Customers

Portfolio	% Population	Charge Off Rate	Charge Off Capture	Mean Value
30+ in 6 months	0.32%	76.51%	23.87%	36.39%
30+ in 9 months	0.68%	67.34%	44.17%	53.35%
30+ in 12 months	1.10%	63.70%	67.29%	65.45%
60+ in 6 months	0.08%	100.00%	8.08%	14.96%
60+ in 9 months	0.26%	100.00%	24.81%	39.76%
60+ in 12 months	0.48%	100.00%	46.05%	63.06%
Num. of times 30+ in 12 months = [1,2,3]	0.75%	46.86%	33.65%	39.17%
Num. of times 30+ in 12 months = [4,5,6]	0.18%	98.90%	16.92%	28.89%
Num. of times 30+ in 12 months = [7,8,9]	0.13%	100.00%	12.59%	22.37%
Num. of times 30+ in 12 months = [10,11,12]	0.03%	100.00%	3.20%	6.19%
Num. of times 60+ in 12 months = [1,2,3]	0.21%	100.00%	20.49%	34.01%
Num. of times 60+ in 12 months = [4,5,6]	0.15%	100.00%	14.85%	25.86%
Num. of times 60+ in 12 months = [7,8,9]	0.09%	100.00%	9.02%	16.55%
Num. of times 60+ in 12 months = [10,11,12]	0.01%	100.00%	1.13%	2.23%

4.2.2 Vintage Analysis for NTC Customers





4.3 APPENDIX C: PYTHON CODES & DOCUMENTATION

1. Documentation for feature list and data processing
   - Sacombank_feature_documentation_202209.xlsx

2. Python Codes
   i) Jupyter Notebook
      - Sacombank_xsell_risk_model_202209.ipynb

   ii) Loan mapping lookup file
      - Loan_type_mapping_20220817_fromSacombank.csv

   iii) Model pickle file
      - Lasso_modelv2_trained.pkl

   iv) WOE Transformation lookup file
      - coarse_bins_selected_features_modelv2.csv
3. Areas of Jupyter Notebook that requires inputs from Sacombank
   a. Section 1: Data Inputs, change to relevant filepaths
   b. Section 2a: Defining New-to-Card customers
   c. Section 4: Modeling, selecting time periods for scoring customers