Credit Risk Modelling

Data and Techniques Used in the UK Banking Industry

Fabrizio Russo

24/10/2019

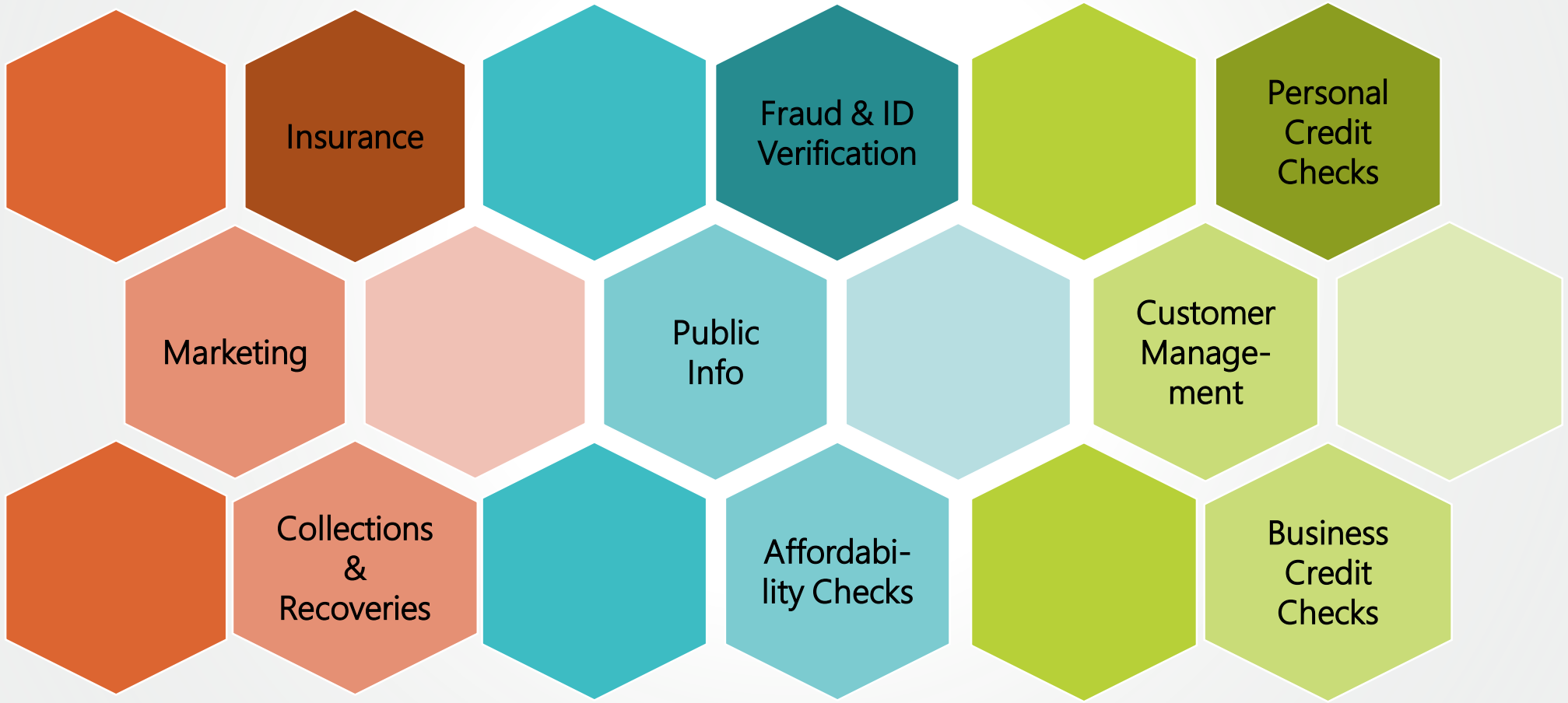
Executive Summary

- Credit Reference Agency (CRA) data significantly increases Credit Risk models' performance
- Machine Learning uses data more effectively and outperforms traditional Scorecards
- Advancements in ML transparency has removed the perceived barriers to the adoption of more advanced techniques

Contents

What Services Do CRAs Offer?

Because of the wealth of information held by the CRAs, their data can be utilised in many ways

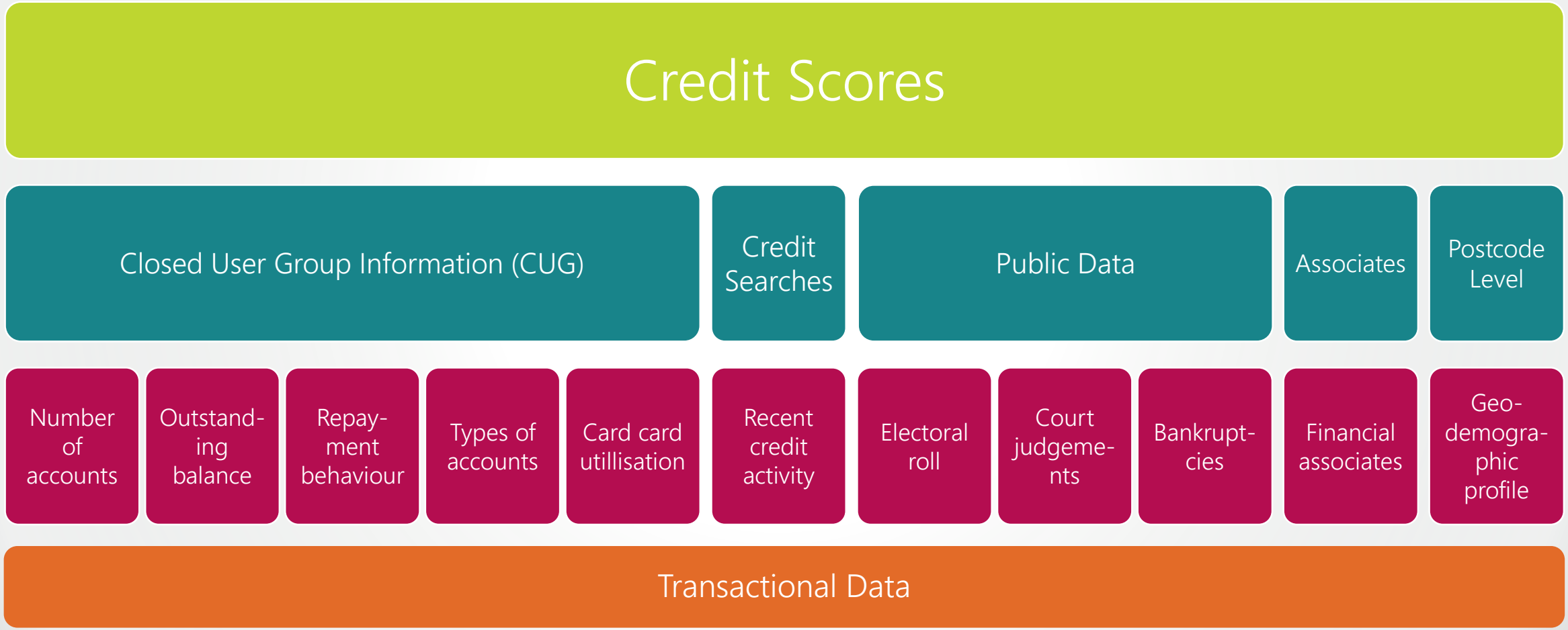


Credit Risk Projects with Extensive Use of CRA Data

- Origination Strategy
- Reject Inference
- Customer Management Strategy
- Regulatory Impacts
- New Products

Application Credit Checks Data

There are many different types of data that are available as part of a credit check:



Benefits of Using CRA Data in Credit Risk Modelling

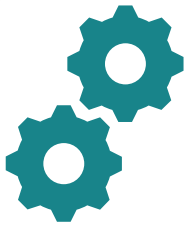
Credit Assessment at Point of Application

Objective

Simplification of risk decisions

Increased automation

Consistent decisions



Target Variable

Customer Risk

Assessment Criteria

Model performance

Auditability and transparency

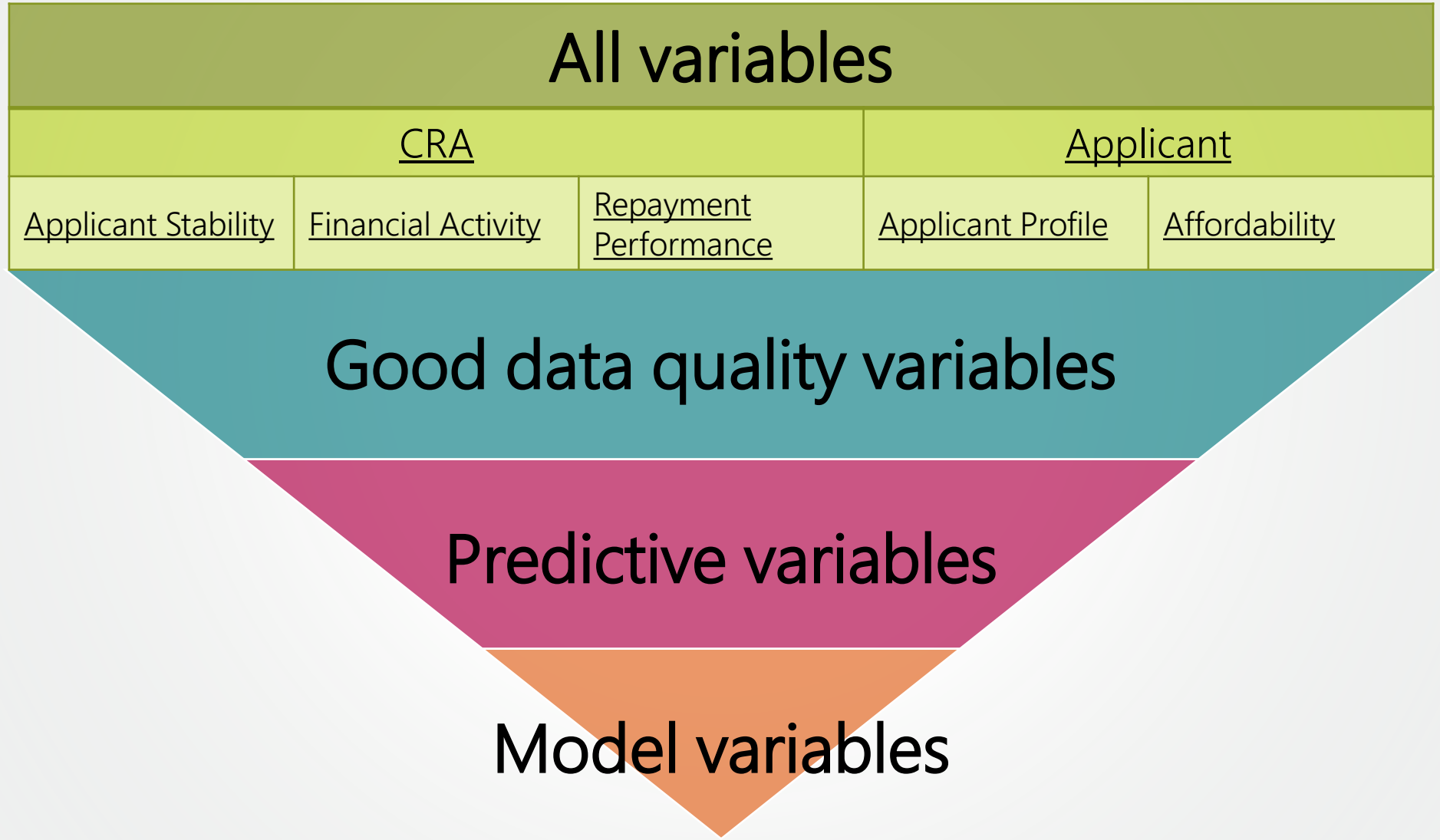
Approach

Data analysis

Build

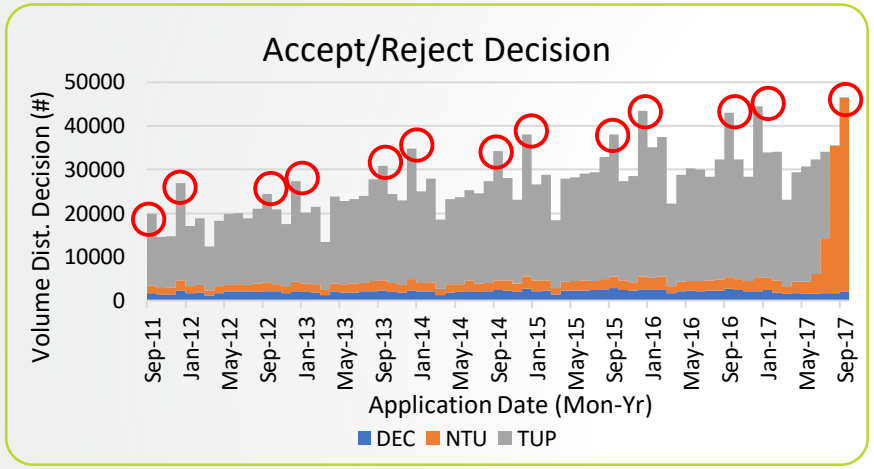
Validation

Data Analysis – Quantity & Types



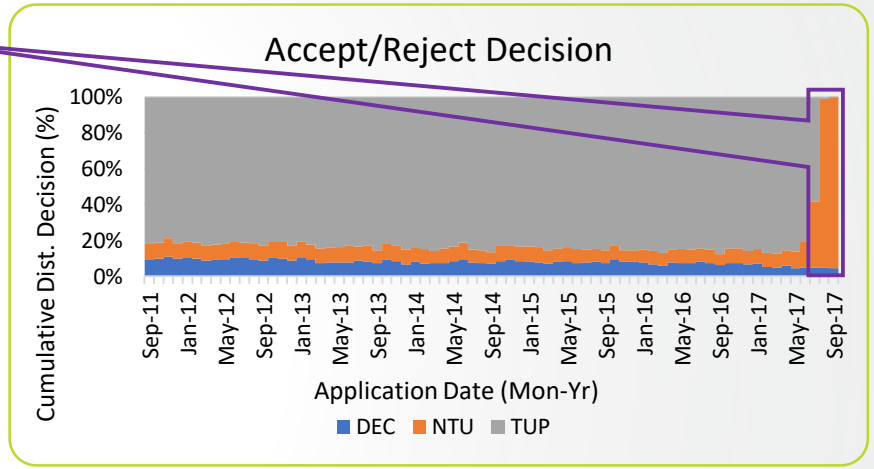
Data Analysis – Quality

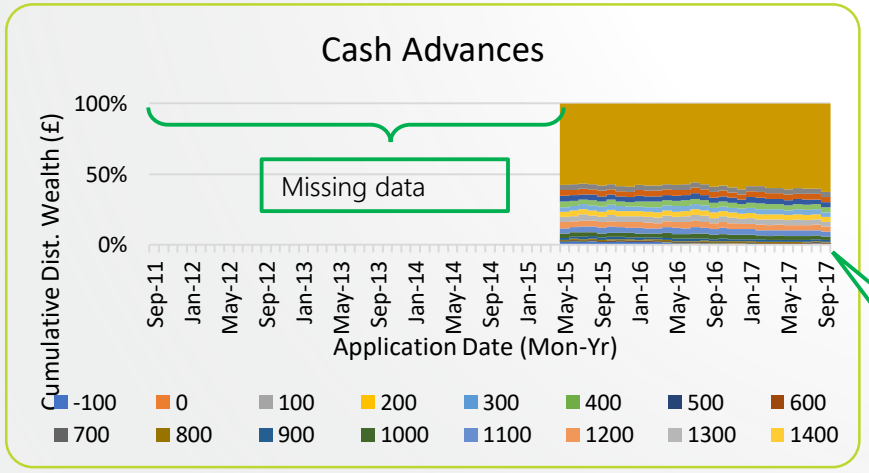
Review data to identify trends or items for exclusion



Take-up can take up to 3 months

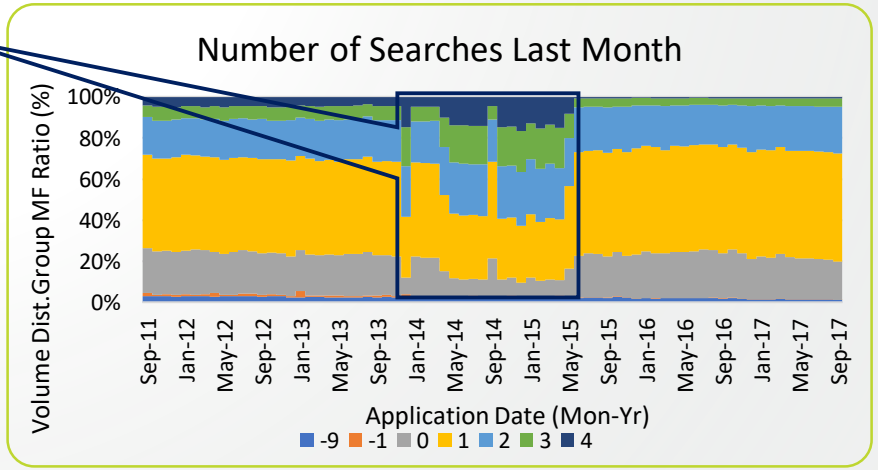
Application spikes in September and December





Unusual behaviour from December '13 to May '15

~200 cases per month with default value of -100



Data Analysis – Predictiveness & Suitability

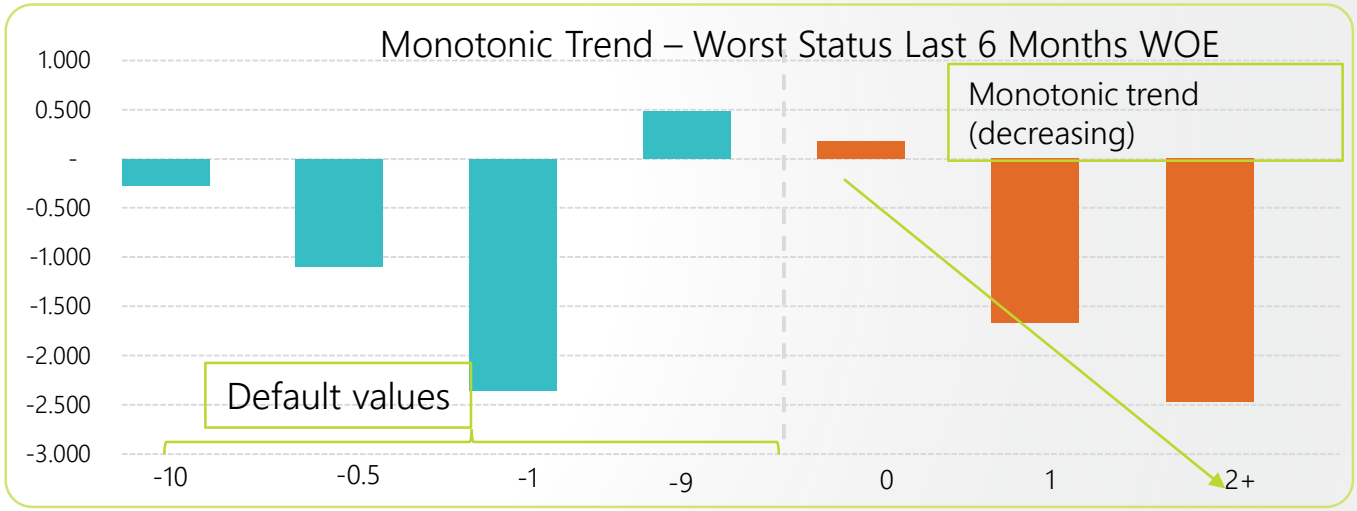
Variable	Overall IV	Data Group	Potential for Modelling
Worst Status Last 6 Months	1.22	CUG	y
Number of Delinquent Accounts	1.22	CUG	y
Value of Delinquent Accounts	1.22	CUG	maybe
Months Since Delinquency	1.19	CUG	y
Value of Unsecured Delinquent Debt	1.18	CUG	no
Number of Unsecured Delinquencies	1.18	CUG	Y
Time Since Most Recent Default	1.05	CUG	Y
Value of Defaults	1.03	CUG	no
Number of Defaults	1.03	CUG	Y
Months Since Mortgage Default	1.00	CUG	y
Value of Mortgage Default	0.99	CUG	maybe
Number of Mortgage Defaults	0.99	CUG	y
Confirmed at Address	0.31	ER	y
Number of Judgements	0.28	Public	y
Tine Since Judgement	0.28	Public	y
Time on ER at Current Address	0.27	ER	y
Number of All Public Judgement Records	0.26	Public	y
Time Since Bankruptcy	0.26	Public	y
Value of Bankruptcy	0.26	Public	y
Applicant Age	0.25	Internal	y
Confirmed at Current Address	0.18	ER	y
Worst Status of Active Accounts Last 12 Months	0.92	CUG	y
Credit Limit Utilisation	0.92	CUG	y
Worst Current Status	0.89	CUG	y
Worst Status Last 3 Motnhs	0.83	CUG	y
Months Since Most Recent Delinquency	0.78	CUG	y
Age of Oldest Active Account	0.51	CUG	y
Age of Youngest Active Account	0.38	CUG	y
Exisiting Customer Worst Status	0.36	Internal	y
Number of Mortgage Accounts	0.35	CUG	y
Number of Settled Accounts	0.12	CUG	maybe
Number of Active Accounts	0.10	CUG	y

$$WoE = LnOdds(attribute) - LnOdds(population)$$

$$IV = Avg_{Good}(WoE) - Avg_{Bad}(WoE)$$

Data Analysis – Predictiveness & Suitability

Variable	Overall IV	Data Group	Potential for Modelling
Worst Status Last 6 Months	1.22	CUG	y
Number of Delinquent Accounts	1.22	CUG	y
Value of Delinquent Accounts	1.22	CUG	maybe
Months Since Delinquency	1.19	CUG	y
Value of Unsecured Delinquent Debt	1.18	CUG	no
Number of Unsecured Delinquencies	1.18	CUG	Y
Time Since Most Recent Default	1.05	CUG	Y
Value of Defaults	1.03	CUG	no
Number of Defaults	1.03	CUG	Y
Months Since Mortgage Default	1.00	CUG	y
Value of Mortgage Default	0.99	CUG	maybe
Number of Mortgage Defaults	0.99	CUG	y
Confirmed at Address	0.31	ER	y
Number of Judgements	0.28	Public	y
Tine Since Judgement	0.28	Public	y
Time on ER at Current Address	0.27	ER	y
Number of All Public Judgement Records	0.26	Public	y
Time Since Bankruptcy	0.26	Public	y
Value of Bankruptcy	0.26	Public	y
Applicant Age	0.25	Internal	y
Confirmed at Current Address	0.18	ER	y
Worst Status of Active Accounts Last 12 Months	0.92	CUG	y
Credit Limit Utilisation	0.92	CUG	y
Worst Current Status	0.89	CUG	y
Worst Status Last 3 Motnhs	0.83	CUG	y
Months Since Most Recent Delinquency	0.78	CUG	y
Age of Oldest Active Account	0.51	CUG	y
Age of Youngest Active Account	0.38	CUG	y
Exisiting Customer Worst Status	0.36	Internal	y
Number of Mortgage Accounts	0.35	CUG	y
Number of Settled Accounts	0.12	CUG	maybe
Number of Active Accounts	0.10	CUG	y



Data Analysis – Predictiveness & Suitability

Variable	Overall IV	Data Group	Potential for Modelling	Marginal IV with Worst_status_L6M
Credit Limit Utilisation	0.92	CUG	y	0.91
Time Since Most Recent Default	1.05	CUG	Y	0.89
Value of Delinquent Accounts	1.22	CUG	maybe	0.86
Number of Mortgage Defaults	0.99	CUG	y	0.85
Months Since Delinquency	1.19	CUG	y	0.82
Number of Delinquent Accounts	1.22	CUG	y	0.77
Value of Mortgage Default	0.99	CUG	maybe	0.76
Number of Defaults	1.03	CUG	Y	0.71
Value of Defaults	1.03	CUG	no	0.68
Worst Status of Active Accounts Last 12 Months	0.92	CUG	y	0.52
Age of Oldest Active Account	0.51	CUG	y	0.46
Value of Unsecured Delinquent Debt	1.18	CUG	no	0.37
Months Since Mortgage Default	1	CUG	y	0.28
Time Since Bankruptcy	0.26	Public	y	0.26
Number of Judgements	0.28	Public	y	0.20
Number of Mortgage Accounts	0.35	CUG	y	0.19
Exisiting Customer Worst Status	0.36	Internal	y	0.15
Time on ER at Current Address	0.27	ER	y	0.14
Value of Bankruptcy	0.26	Public	y	0.13
Months Since Most Recent Delinquency	0.78	CUG	y	0.13
Number of All Public Judgement Records	0.26	Public	y	0.13
Confirmed at Current Address	0.18	ER	y	0.11
Tine Since Judgement	0.28	Public	y	0.09
Applicant Age	0.25	Internal	y	0.09
Confirmed at Address	0.31	ER	y	0.08
Worst Current Status	0.89	CUG	y	0.08
Number of Unsecured Delinquencies	1.18	CUG	Y	0.08
Age of Youngest Active Account	0.38	CUG	y	0.04
Number of Settled Accounts	0.12	CUG	maybe	0.02
Number of Active Accounts	0.1	CUG	y	0.02
Worst Status Last 3 Months	0.83	CUG	y	0.01
Worst Status Last 6 Months	1.22	CUG	y	0

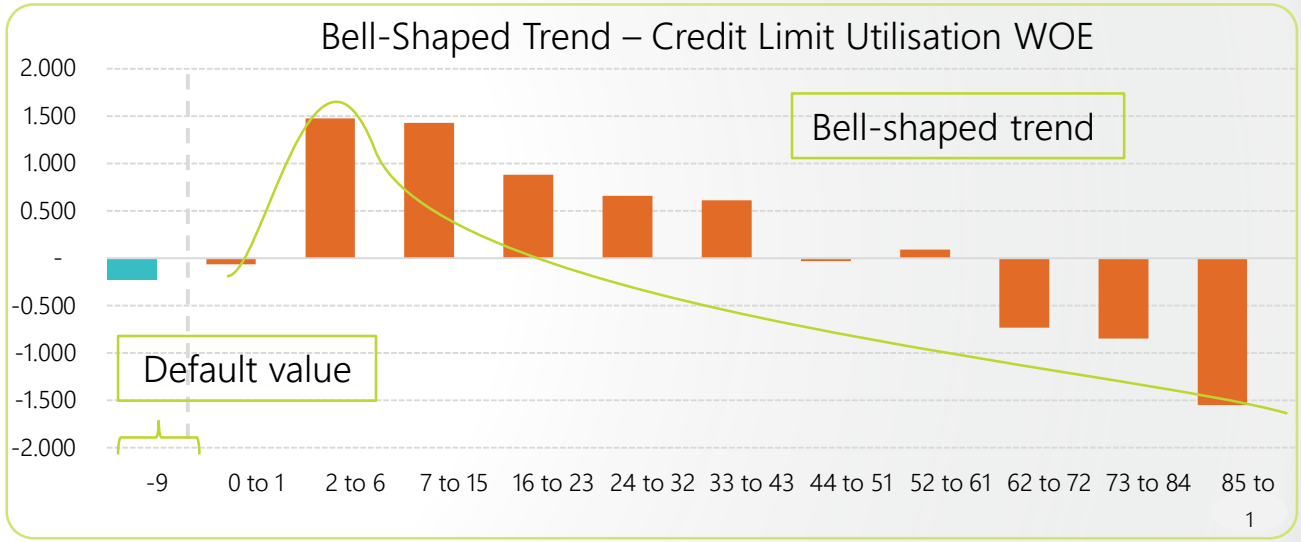
$$\text{WoE} = \text{LnOdds(attribute)} - \text{LnOdds(population)}$$

$$\text{Delta Score} = \text{Observed WoE} - \text{Expected WoE}$$

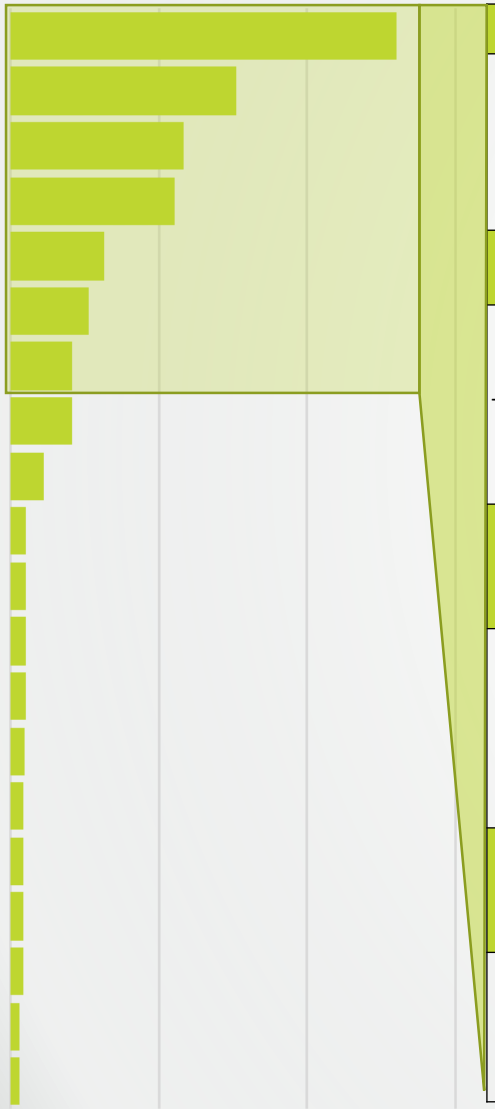
$$\text{MIV} = \text{Avg}_{\text{Good}}(\text{Delta Score}) - \text{Avg}_{\text{Bad}}(\text{Delta Score})$$

Data Analysis – Predictiveness & Suitability

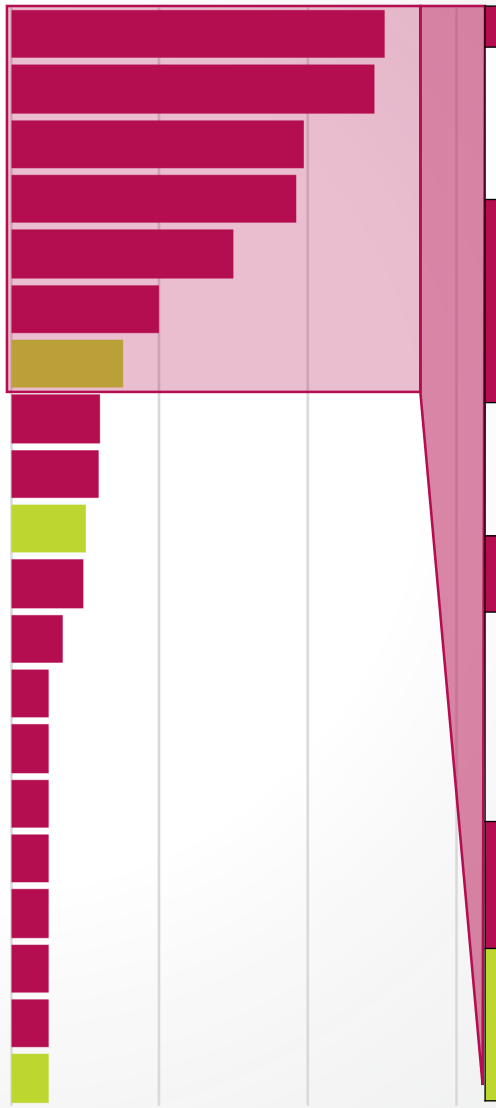
Variable	Overall IV	Data Group	Potential for Modelling	Marginal IV with Worst_status_L6M
Credit Limit Utilisation	**0.92**	**CUG**	**y**	**0.91**
Time Since Most Recent Default	1.05	CUG	Y	0.89
Value of Delinquent Accounts	1.22	CUG	maybe	0.86
Number of Mortgage Defaults	0.99	CUG	y	0.85
Months Since Delinquency	1.19	CUG	y	0.82
Number of Delinquent Accounts	1.22	CUG	y	0.77
Value of Mortgage Default	0.99	CUG	maybe	0.76
Number of Defaults	1.03	CUG	Y	0.71
Value of Defaults	1.03	CUG	no	0.68
Worst Status of Active Accounts Last 12 Months	0.92	CUG	y	0.52
Age of Oldest Active Account	0.51	CUG	y	0.46
Value of Unsecured Delinquent Debt	1.18	CUG	no	0.37
Months Since Mortgage Default	1	CUG	y	0.28
Time Since Bankruptcy	0.26	Public	y	0.26
Number of Judgements	0.28	Public	y	0.20
Number of Mortgage Accounts	0.35	CUG	y	0.19
Exisiting Customer Worst Status	0.36	Internal	y	0.15
Time on ER at Current Address	0.27	ER	y	0.14
Value of Bankruptcy	0.26	Public	y	0.13
Months Since Most Recent Delinquency	0.78	CUG	y	0.13
Number of All Public Judgement Records	0.26	Public	y	0.13
Confirmed at Current Address	0.18	ER	y	0.11
Tine Since Judgement	0.28	Public	y	0.09
Applicant Age	0.25	Internal	y	0.09
Confirmed at Address	0.31	ER	y	0.08
Worst Current Status	0.89	CUG	y	0.08
Number of Unsecured Delinquencies	1.18	CUG	Y	0.08
Age of Youngest Active Account	0.38	CUG	y	0.04
Number of Settled Accounts	0.12	CUG	maybe	0.02
Number of Active Accounts	0.1	CUG	y	0.02
Worst Status Last 3 Months	0.83	CUG	y	0.01
Worst Status Last 6 Months	1.22	CUG	y	0



Traditional Scorecard - Internal & CRA Data



Variable	Values	Score
Loan to Value	Low to 25	24
	26 - 40	10
	41 - 50	5
	51 - 60	2
	61 - 69	-3
	70 - 79	-7
	80 to high	-11
Good Existing Customer	New Customer	0
	Yes	10
	No	-15
Time in Employment (MM)	No Info	-10
	Low to 36	-10
	37 - 66	-7
	67 - 91	-3
	92 - 120	-2
	121 - 143	2
	144 - 184	5
	185 to high	13
Residential Status	No Info	-1
	Public Tenant	-5
	Living with Parents	5
	Private Tenant	10
	Owner	15
Applicant Age (YY)	Low to 22	-6
	23 - 25	-5
	26 - 29	-3
	30 - 33	-2
	34 - 37	0
	38 - 42	2
	43 - 48	5
	49 - high	9
Declared Unsecured Debt to Income	No Info	-2
	Low to 15	4
	16 - 29	-2
	30 - 49	-6
	50 to high	-10
Employment Status	No info	2
	Unemployed	-10
	House Person	-3
	Contractor	6
	Part Time	8
	Full Time	12

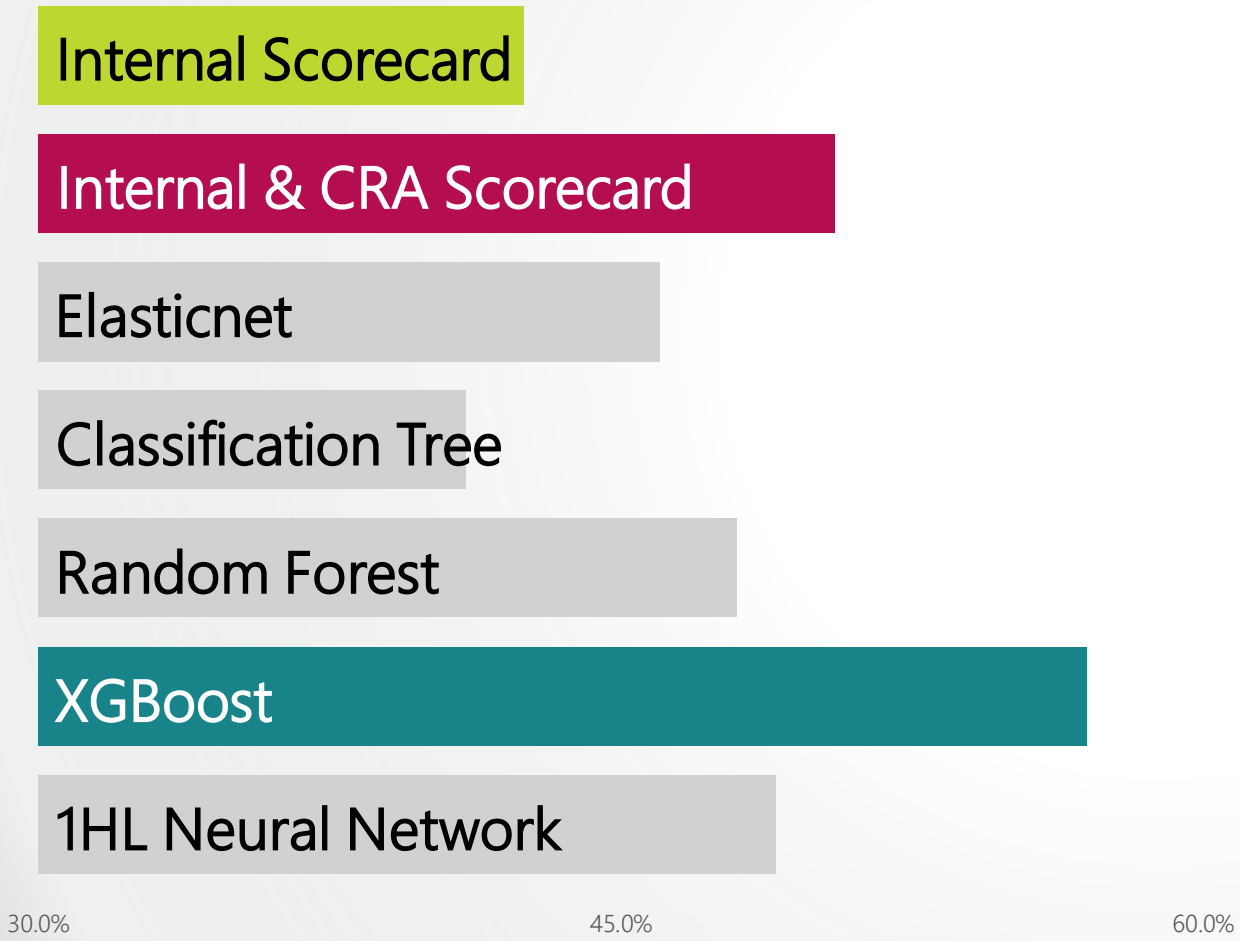


Variable	Values	Score
Worst Payment Status Last 6 Months	No Accounts	-11
	0	6
	1	-11
	2	-21
	3	-28
	4 to high	-30
Credit Limit Utilisation (%)	No Credit Card	-5
	0	-1
	Jan-16	28
	17 - 33	15
	34 - 48	6
	49 - 63	-2
	64 - 81	-13
	82 to high	-24
Time Since Last Delinquency (MM)	No Delinq Accounts	5
	Low to 12	-21
	13 - 36	-17
	37 - 50	-13
	51 to high	-9
Number of Defaulted Accounts	No Default	4
	01-Feb	-13
	3 to high	-16
Age of Oldest Active Account	No Active Account	-11
	Low to 52	-12
	53 - 90	-8
	91 - 113	-5
	114 - 132	-1
	133 - 152	2
	153 - 179	4
	180 to high	5
Time Since Opening Mortgage Account	No Mortgage	-8
	low to 23	19
	24 - 54	14
	55 - 91	10
	92 to high	7
Time Since Missed Payment (Existing Customers)	New Customer	0
	None	15
	Low to 6	-10
	7 – 12	-5
	13 – 36	-1
	37 to high	0

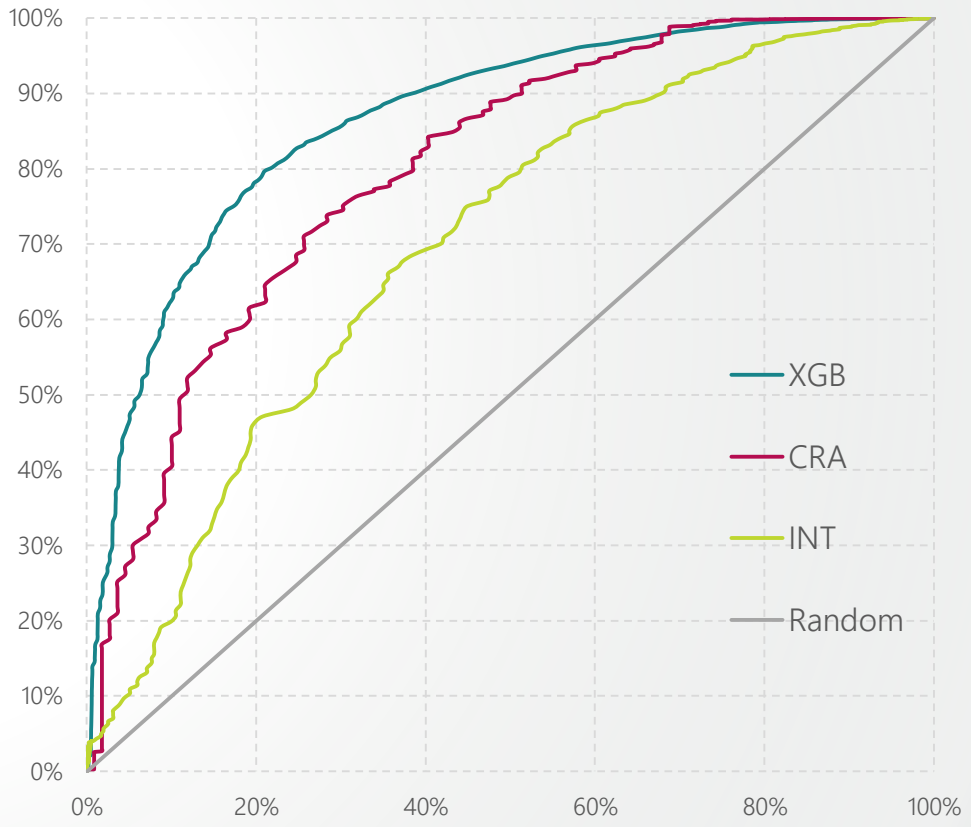
From Regression to Machine Learning Models

Algorithm Comparison

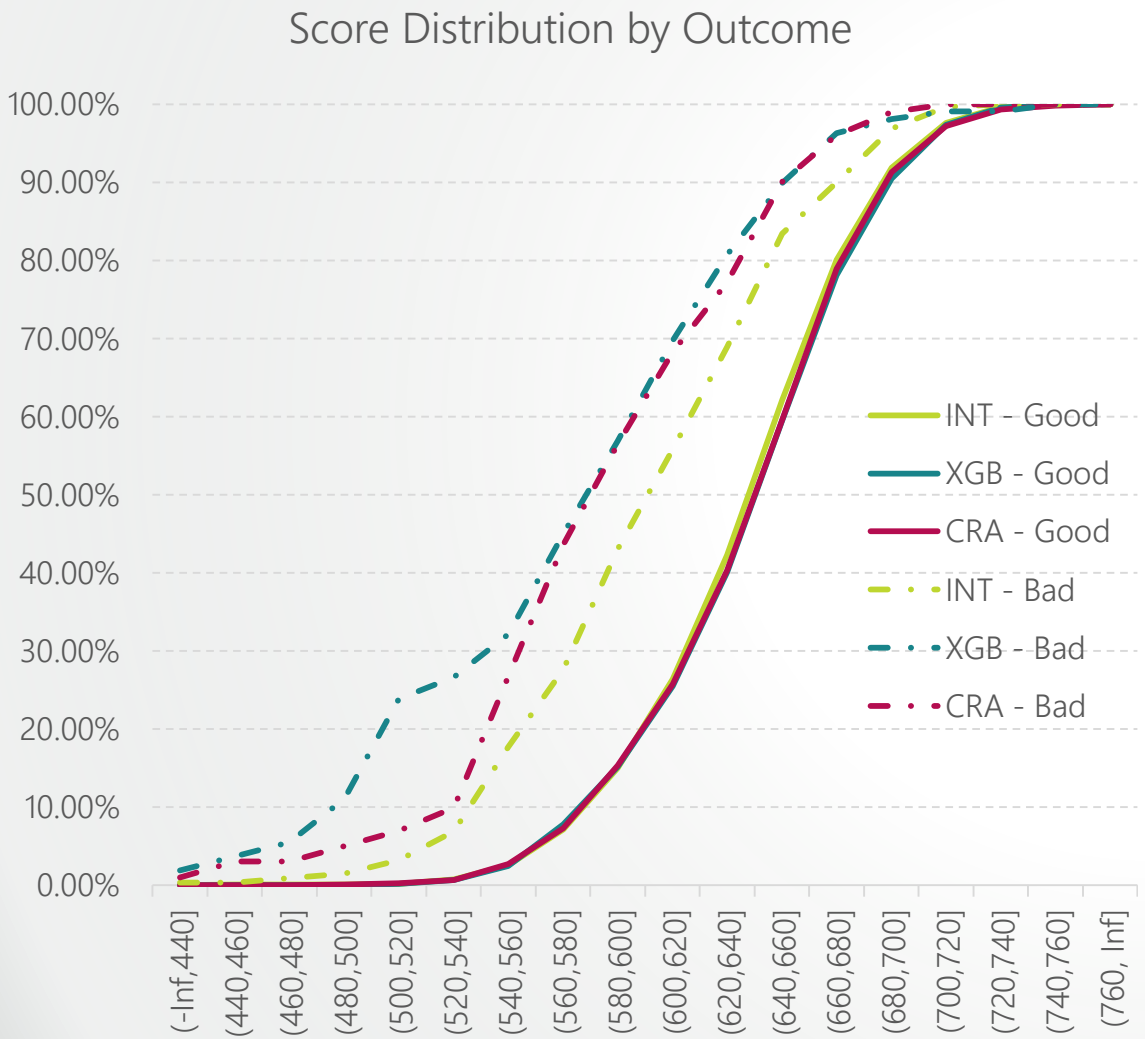
GINI Comparison on Test Sample

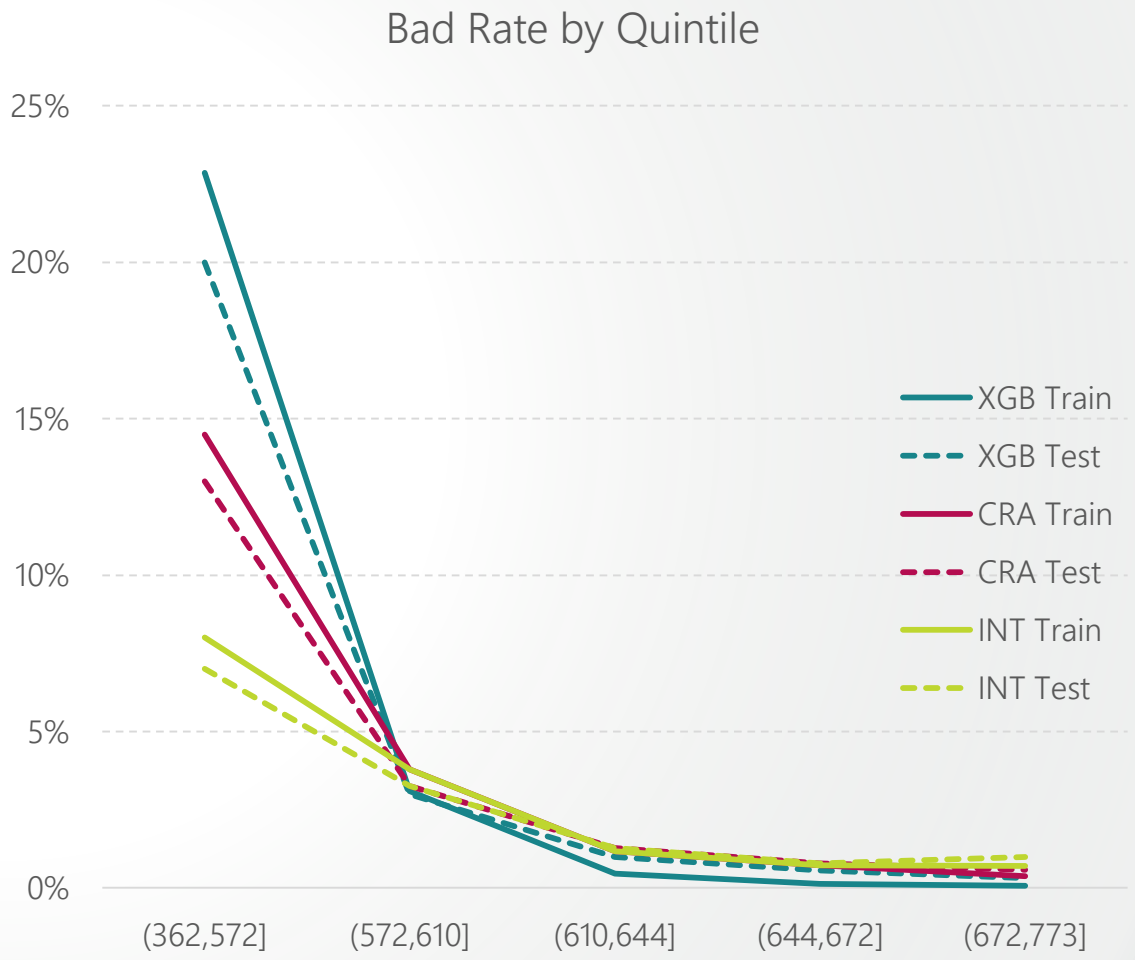


ROC Curves

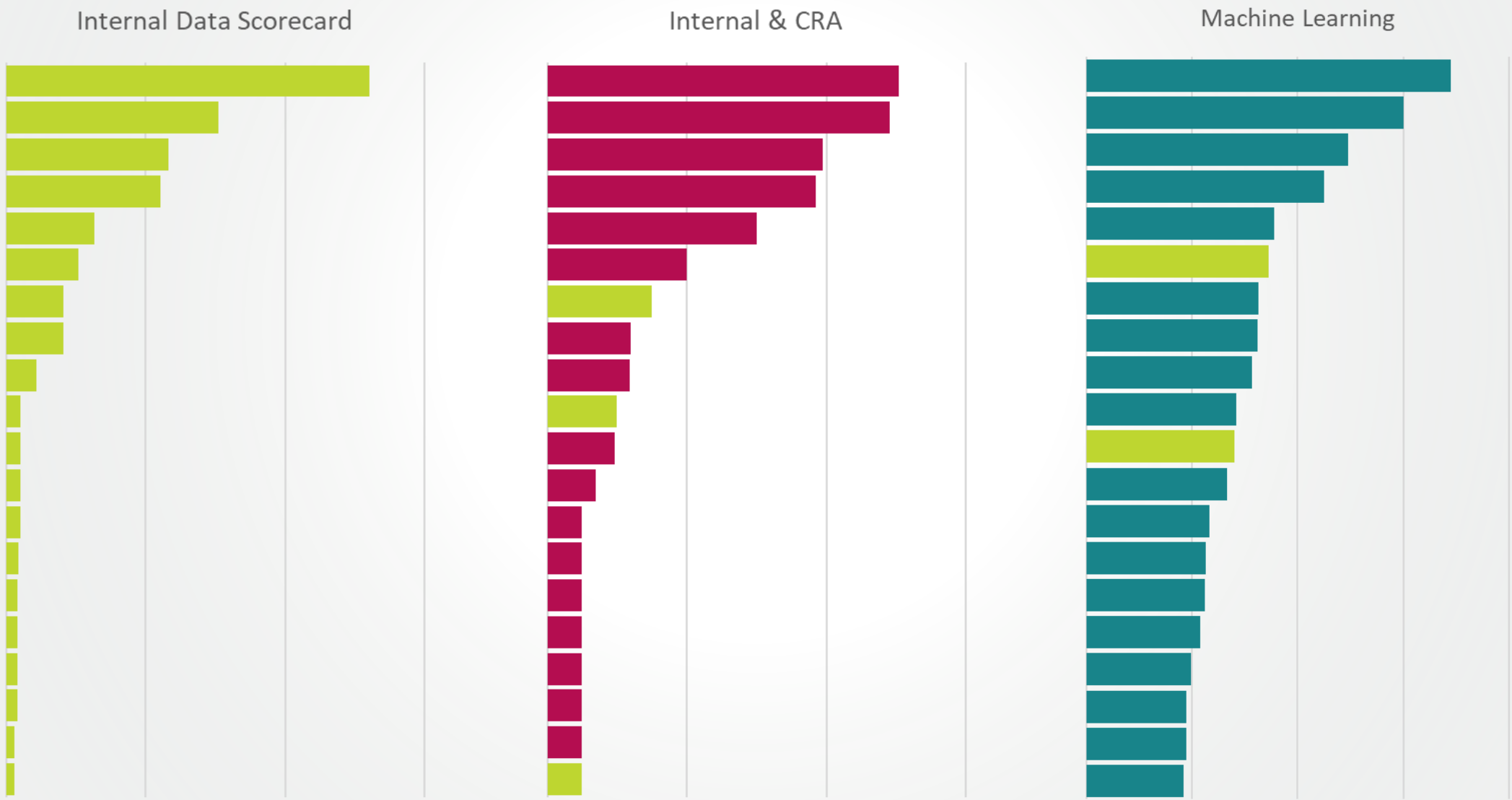


Performance Comparison





Variable Importance Comparison

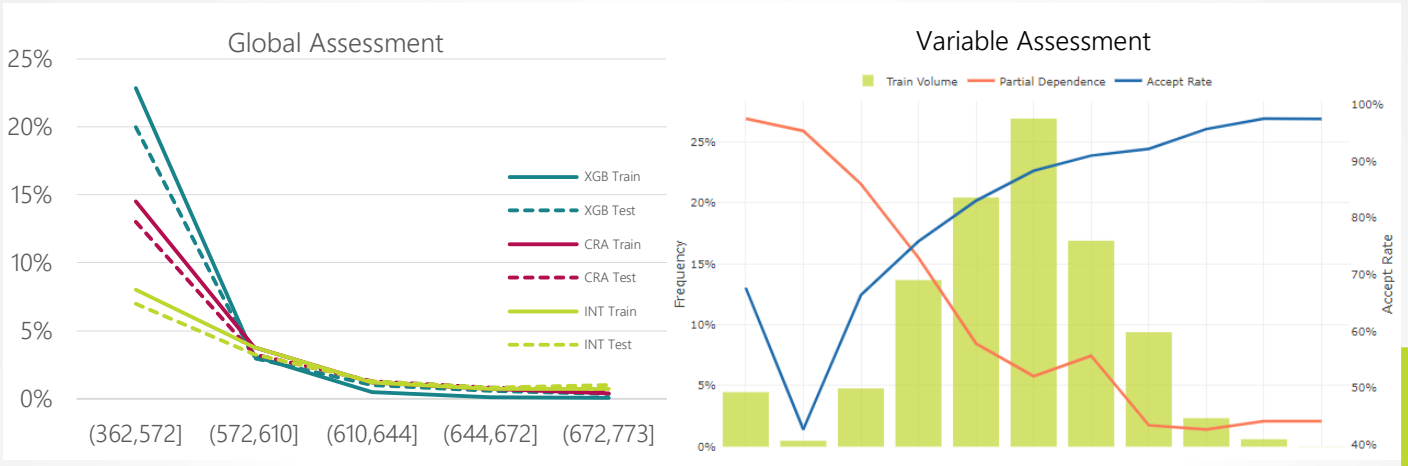


Considerations and Summary

Regulatory Considerations

Governance

Transparency



Consistent Decisions & Treat Customers Fairly

Questions?