Credit Score

- Results -

Data Science Credit

Score | Summary

Objective

Development of customized credit risk scoring models for the customer journey.

relationship time (months)

0   1   3   12

---

25 features

+8.9 p.p of performance*
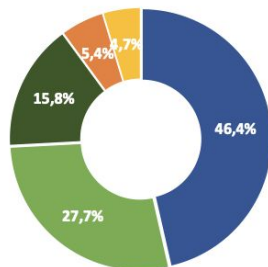
- SERASA
- CADASTRO
- TRANSAÇÃO
- CADASTRO CARTAO
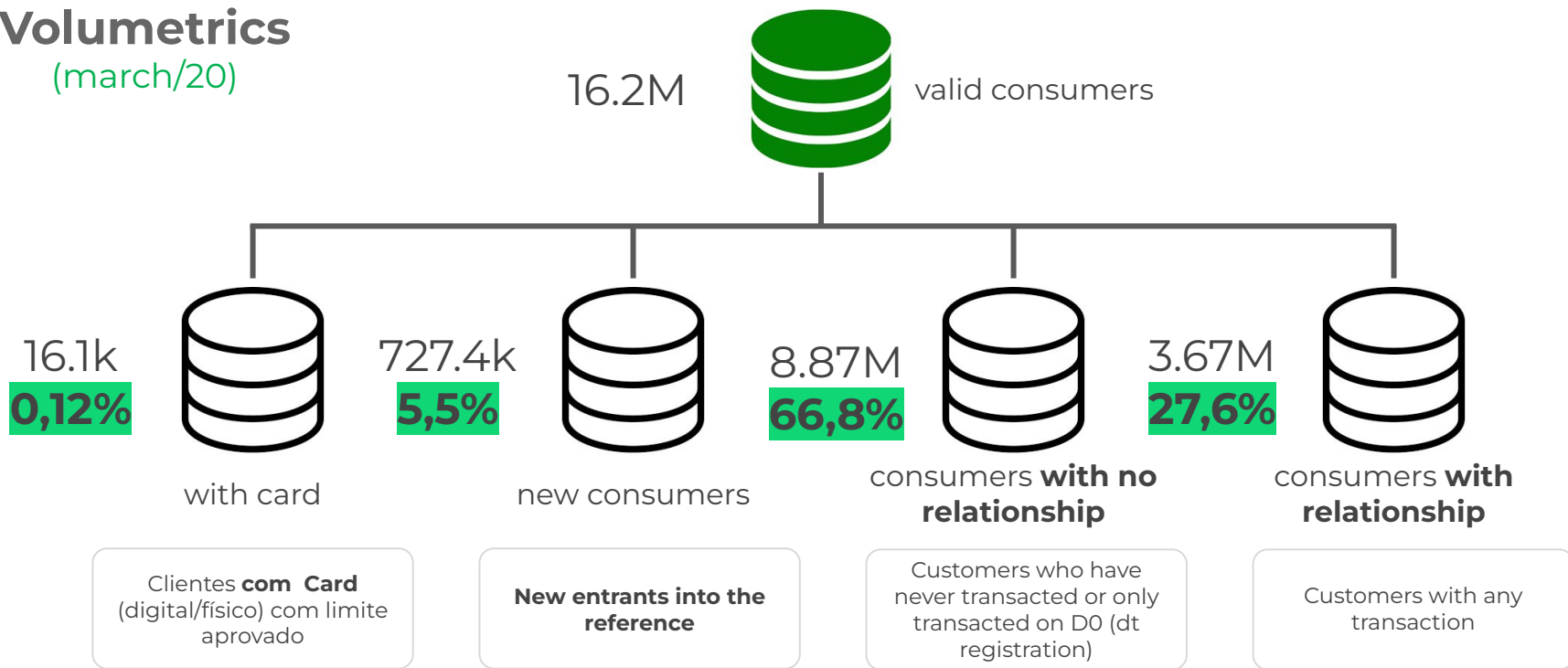- BEHAVIOR SEGMENTATION

46,4%
27,7%
15,8%
5,4%
4,7%

Information sources

*baseline comparison

Score | Public Sample

Consumers were categorized into 4 main groups related to the degree of relationship they have with the company

Volumetrics
(march/20)

16.2M **valid consumers**

16.1k
0,12%
with card

Clientes **com Card** (digital/físico) com limite aprovado

727.4k
5,5%
new consumers

New entrants into the reference

8.87M
66,8%
consumers **with no relationship**

Customers who have never transacted or only transacted on D0 (dt registration)

3.67M
27,6%
consumers **with relationship**

Customers with any transaction

Score | Timeline

- Rationale used to create public harvests, performance and information lags

Public definitions
Bad at the origin
Transaction book
Registration information, etc.

Information
SERASA

M-2

M-1

201911

M0

Vision: if the customer has owed at least 100 reais for 60 days or more within 6 months on the market

M+1

M+6

20191001

Picture of the wallet

20191201

External performance
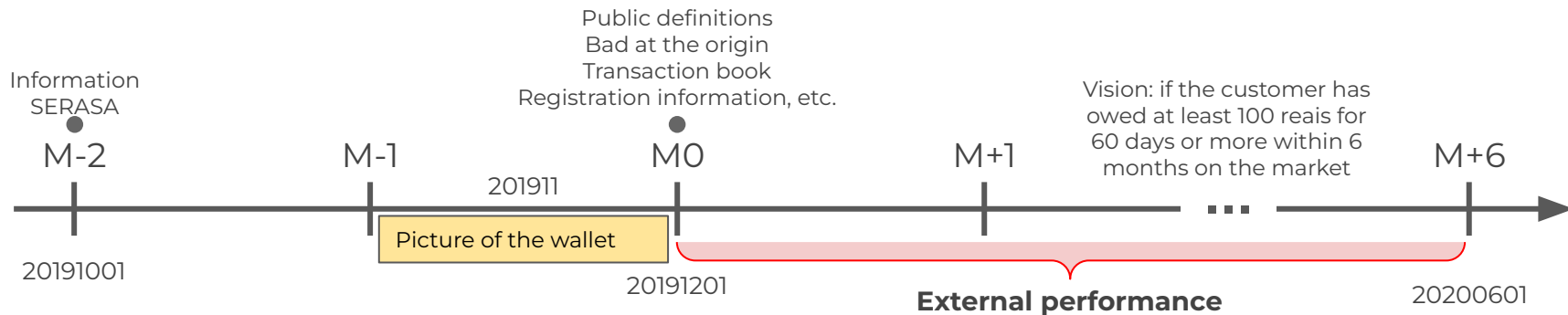
20200601

SAFRAS	201910	201911	201912	202001	202002	202003	202004	202005	202006	202007	202008	202009	202010	202011
		Picture of the wallet	P1	P2	P3	P4	P5	P6						
DESENV				Picture of the wallet	P1	P2	P3	P4	P5	P6				
						Picture of the wallet	P1	P2	P3	P4	P5	P6		
VAL. OOT							Picture of the wallet	P1	P2	P3	P4	P5	P6	

Score | Information sources

- Both Serasa market information and internal variables were tested

external

n. features

Score Negativo
Model C2BA**

Score Positivo + Negativo
Model HSPA

2

Registration
Features related to the consumer's registration, device and credit card

+100

internal information

Payments and Receiving
Transaction variable book

+700

Cognitive Biases
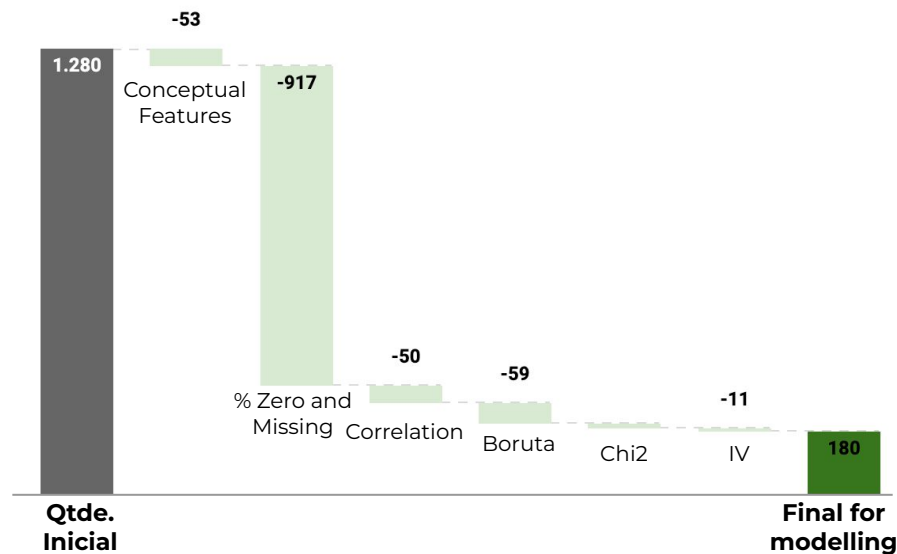Consumers' latent bias scores
(Behavioral Science)

+400

**score baseline

Score | Modeling Filters

		Bads on M0	Scr = 0	Scr = 1	No performance	Inactive clients	Age < 18	Ok Public
#	1.527.592	573.641	50.600	12.665	205	716	3.394	**886.371**
%	100%	37,6%	3,3%	0,8%	0,0%	0,1%	0,2%	**58,3%**
BR %	47,5%	96,0%	0,4%	NA	NA	16,5%	8,4%	**22,4%**

Score | Feature selection

PS: A clean up was made after the modeling part because of the feature importances so we could get to a minor number of features

Critérios

- Removal of conceptual variables (timestamp, without a clear or anti-business description)
- Proportion of missing and zeros > 95%
- Boruta (random forest with randomization of features) and binary correlation (5% with the target)
- Information value (IV), WoE and Chi2

Score | Modelagem

Treino dos modelos

5 Técnicas
Logistic regression
Random Forest
GBoosting
LightGBM
XGBoost

3 Combinações
Internal Features
C2BA + Internal Features
HSPA + Internal Features

50 trials
Optuna
(Bayesian optimization
for grid search)

Total

+ 750 models tested

Performance of the models - KS%

Legend: C2BA, HSPA, Internas, C2BA + Int., HSPA + Int.

DEV:
- C2BA: 36,7%
- HSPA: 42,5%
- Internas: 16,8%
- C2BA + Int.: 43,7%
- HSPA + Int.: 45,6%

VAL OOS:
- C2BA: 36,3%
- HSPA: 42,5%
- Internas: 16,4%
- C2BA + Int.: 43,5%
- HSPA + Int.: 45,7%
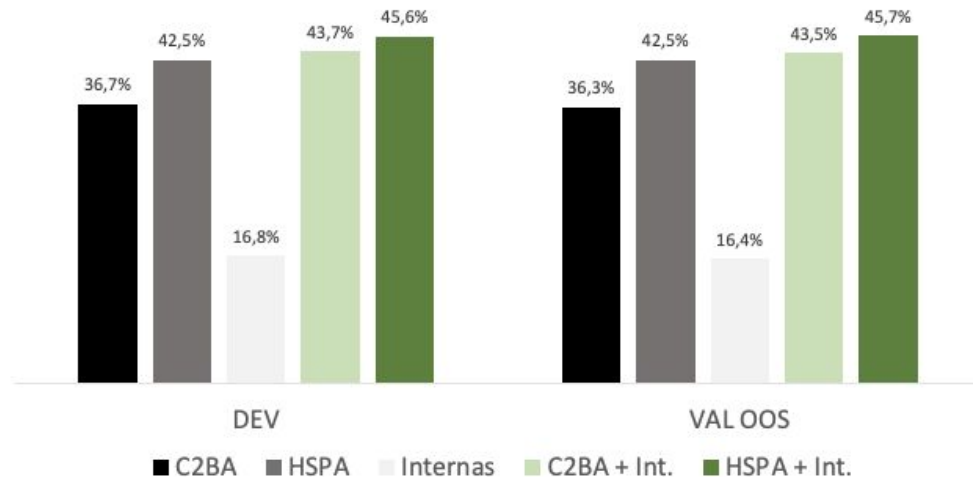
Score

Performance Δ p.p / Δ %

Metric/Model	Desenv. (80%)		Val. OOS (20%)		Val. OOT (-)	
C2BA	36,7%		36,3%		37,6%	
KS HSPA	42,5%	+3,1 / +8,9	42,5%	+3,2 / +9,4	43,4%	+2,8 / +8,9
Score	45,6%	+7,3 / +24,3	45,7%	+7,5 / +25,9	46,2%	+6,5 / +22,9

Feature Importance (Relative)

25 Features

46,4%
27,7%
15,8%
5,4%
4,7%

■ SERASA ■ CADASTRO ■ TRANSAÇÃO ■ CADASTRO CARTAO ■ BEHAVIOR SEGMENTATION

Feature Importance (Relative)

0,00% 10,00% 20,00% 30,00% 40,00% 50,00%

- HSPA
- idade_em_anos
- ddd_state
- tempo_registro_em_meses
- days_since_last_transaction
- agrupamento_instituicao_cartao_ref
- am_max_pay_12m_trx
- modelo_celular_ref
- mental_accounting_score
- peer_effect_score
- email_ajuste
- am_max_recp2p_12m_trx
- qtde_emissores_distintos
- digital_account_status
- am_max_payp2pbal_12m_trx
- am_min_paypavcred_06m_trx
- am_tot_paypavbal_06m_trx
- am_max_paypavgood_12m_trx
- ct_tot_payp2pcred_12m_trx
- am_tot_payp2pbal_06m_trx
- bandeira_cartao_ref
- am_tot_paypavmix_12m_trx
- ct_min_paypav_12m_trx
- am_tot_paypavcred_12m_trx
- qtde_cartoes_ativos_cadastrados
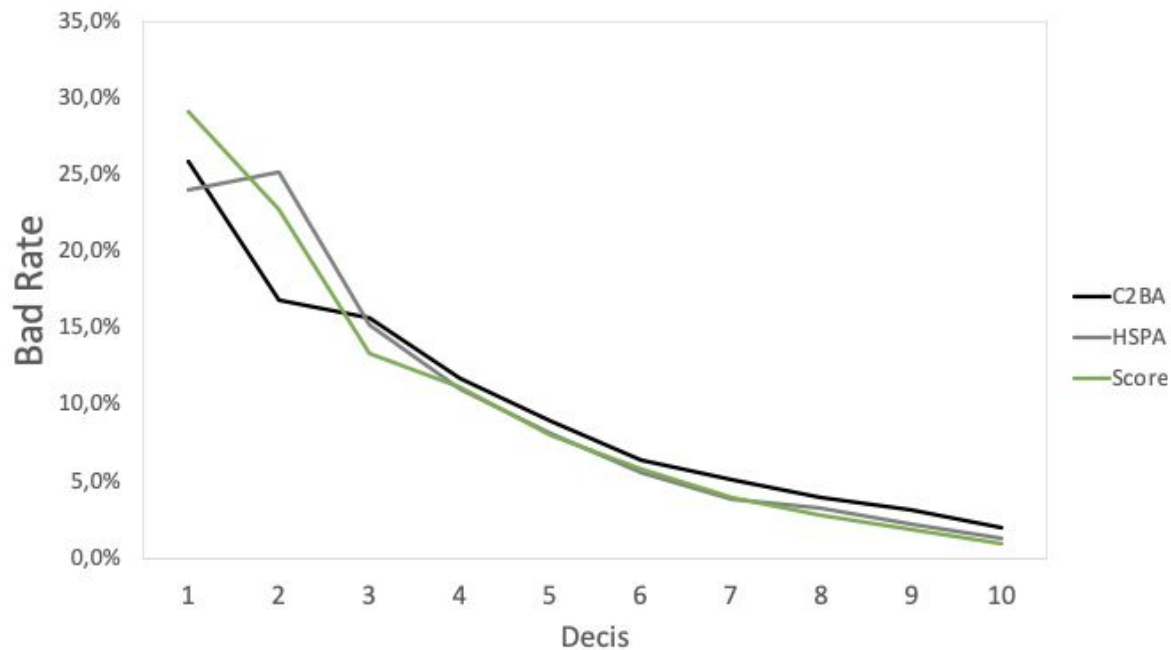
Score

📊 Ordenação

Total (100%)			
Decis	**C2BA**	**HSPA**	**Score**
1	24,7%	22,7%	28,7%
2	17,1%	24,0%	21,5%
3	15,7%	15,5%	14,9%
4	11,8%	11,9%	11,2%
5	9,2%	8,6%	8,3%
6	6,9%	6,2%	5,8%
7	5,4%	4,1%	4,2%
8	4,0%	3,5%	2,8%
9	3,2%	2,3%	1,8%
10	2,0%	1,2%	0,9%
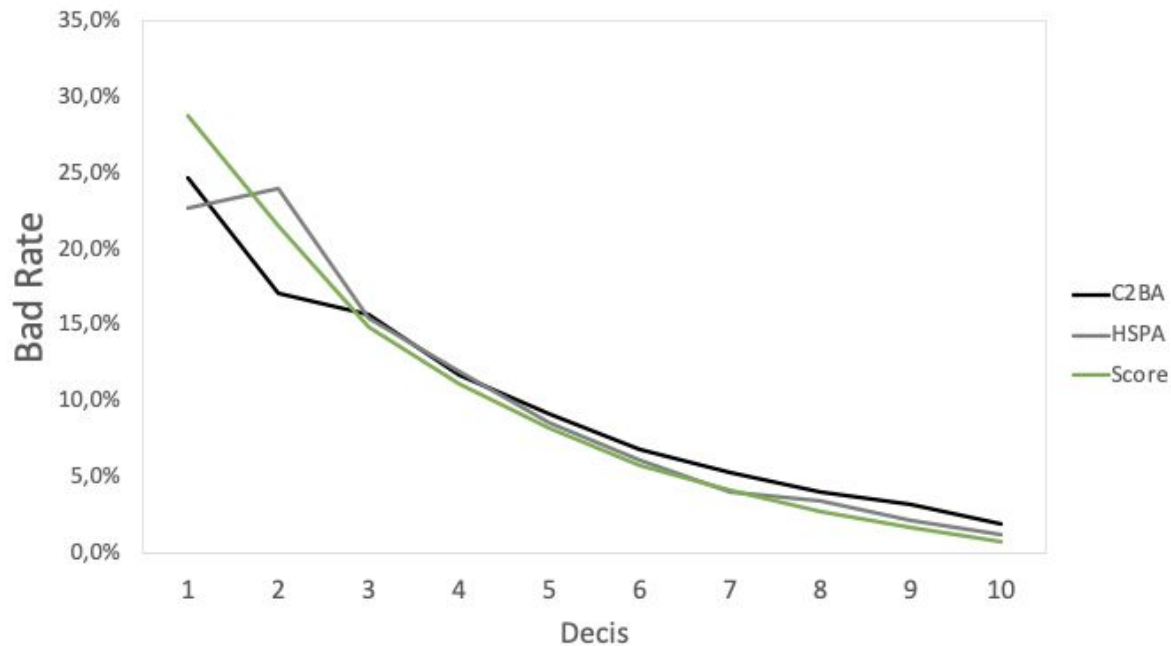
Score

📊 Ordenação

Out-of-Time			
Decis	**C2BA**	**HSPA**	**Score**
1	25,9%	24,1%	29,1%
2	16,9%	25,2%	22,8%
3	15,7%	15,3%	13,4%
4	11,8%	11,1%	11,2%
5	9,0%	8,1%	8,0%
6	6,5%	5,6%	5,8%
7	5,2%	3,9%	4,0%
8	4,0%	3,3%	2,8%
9	3,1%	2,2%	1,9%
10	2,0%	1,3%	0,9%

Score

Public cuts - Analisys (KS% - val. oot)

C2BA 38%   HSPA 41%   Score 43%

Público	Segmentation			Performance
Score C2BA	> 360	V: 86%	B: 15%	C2BA 31% — HSPA 39% — Score 41%
	<= 360	V: 14%	B: 40%	C2BA 17% — HSPA 24% — Score 38%
Card Registration	with card	V: 18%	B: 14%	C2BA 42% — HSPA 48% — Score 50%
	no card	V: 82%	B: 20%	C2BA 35% — HSPA 41% — Score 45%
Idade	> 25 yo	V: 80%	B: 19%	C2BA 40% — HSPA 44% — Score 44%
	<= 25 yo	V: 20%	B: 15%	C2BA 35% — HSPA 43% — Score 49%
Transactions in history	with transactions	V: 80%	B: 19%	C2BA 36% — HSPA 41% — Score 44%
	no transactions	V: 20%	B: 15%	C2BA 41% — HSPA 47% — Score 49%
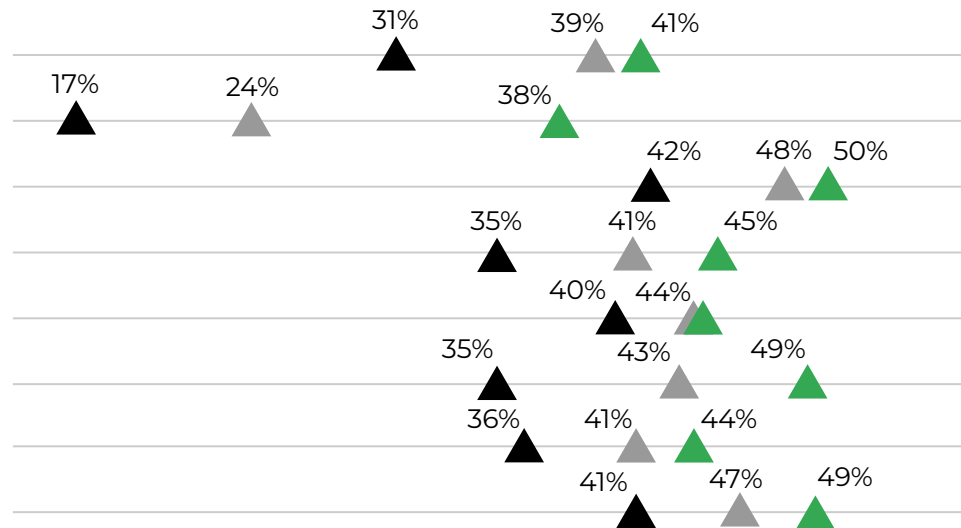
Consumer's journey

Model Journey | Monthly Model

- How the core model connects with customer relationship length

consumers

vision M0
monthly batch

scored consumers → decision

geração da carteira de clientes válidos

scoring customers using the batch information available on the date

with scoring, inputs are generated to generate pre-approved products

Model Journey | Consumer's moment in life

- How the core model connects with customer relationship length

external information (Serasa)

external information + internal information

internal information with relationship

0 3 12 ••• relationship time in months

Model application