

Case – Analista de dados

Sophia Sales R. Gonçalves

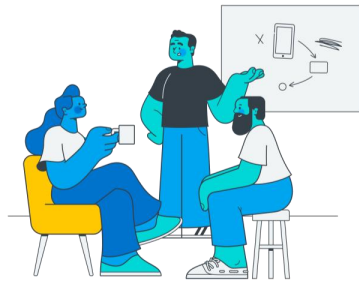
[GitHub](#)



O que você vai encontrar essa apresentação?

1

contexto e metodologia



2

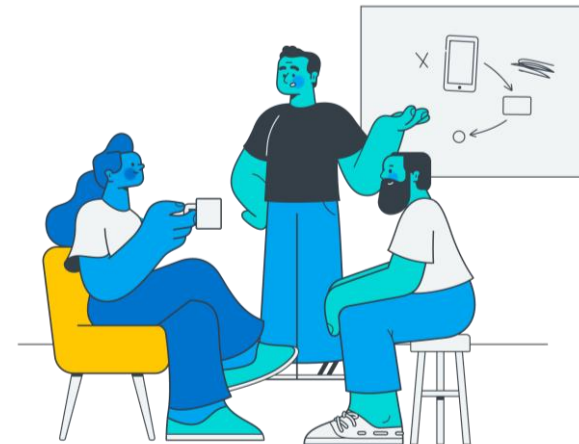
validação e respostas



3

dataviz e respostas





contexto e metodologia

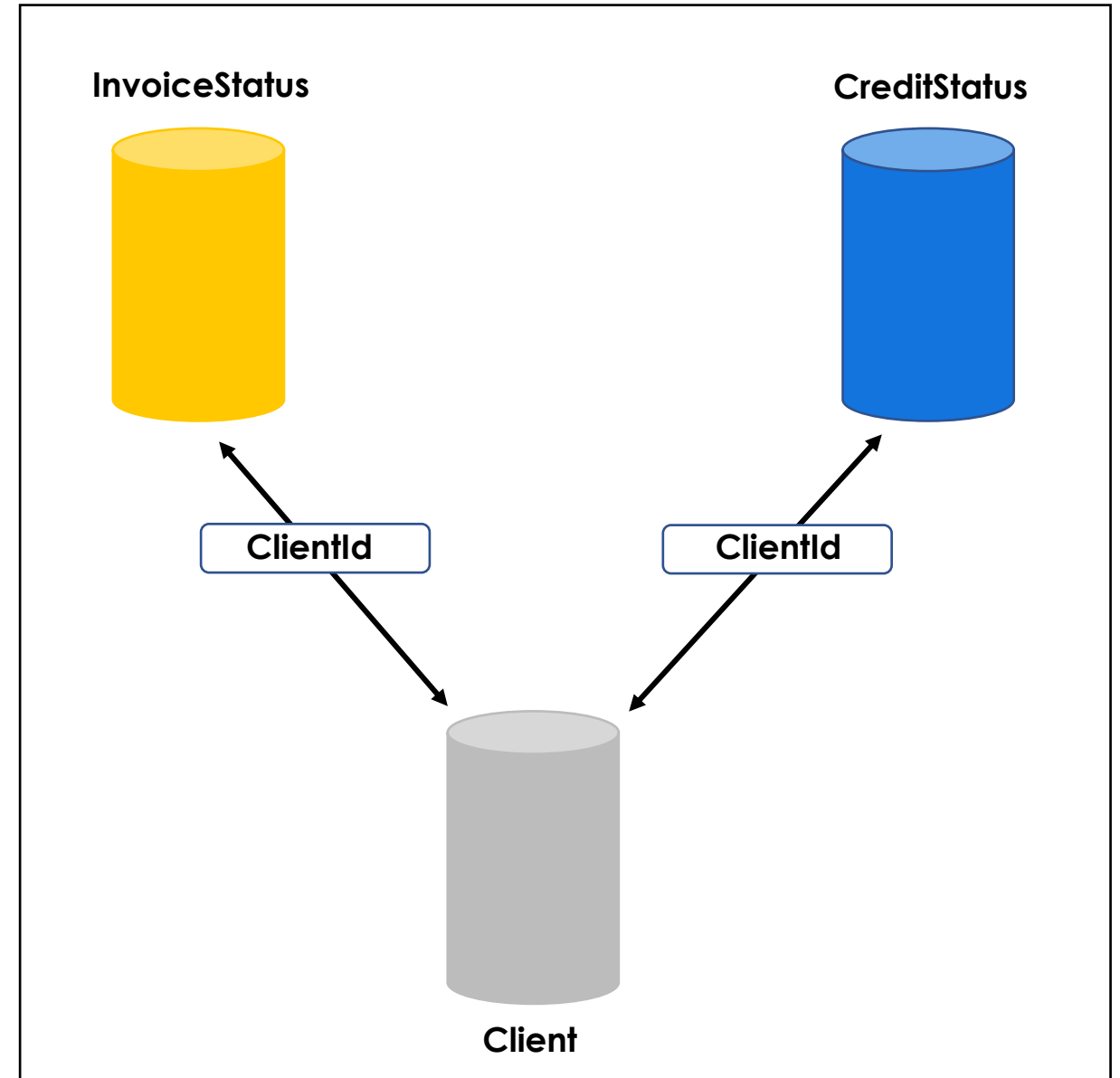
base de dados

tabelas

- Client
- CreditStatus
- InvoiceStatus

chaves

- ClientId
- Invoiceld

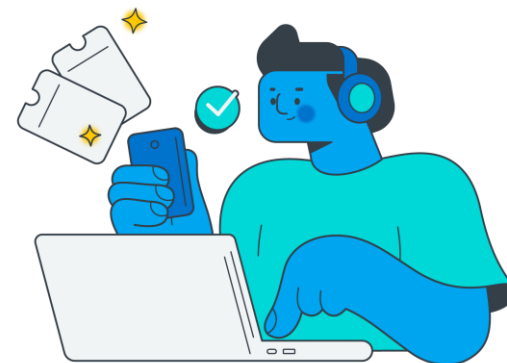


metodologia

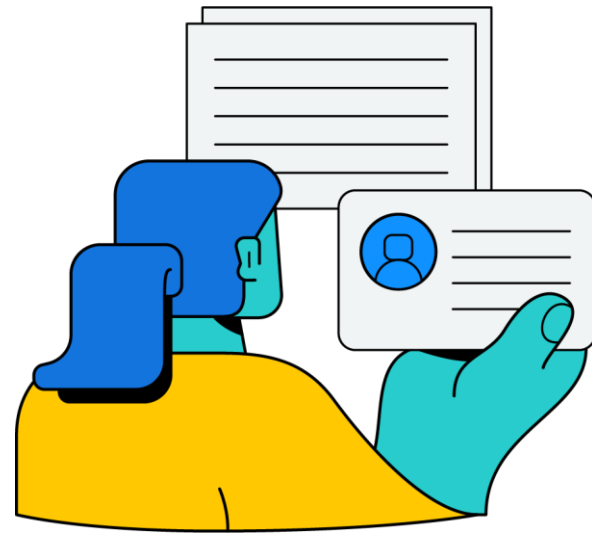
- Queries de validação
SQL
- Queries de respostas
Python e SQL
- Dataviz
Python

ferramentas

- DB Browser
SQLite3
- Python
Jupyter Notebook
- Editor
VS Code



validação e respostas



validação

- tamanho das bases
- dados nulos e duplicados
- intervalo de tempo

```
----- QUERIES de validação dos dados -----
----- conhecendo a base de dados CreditStatus
SELECT *
FROM CreditStatus;

--- tamanho da base de dados CreditStatus
SELECT COUNT(1)
FROM CreditStatus;
-- Resposta: 34.763

--- clientes nulos
SELECT COUNT(ClientId)
FROM CreditStatus
WHERE ClientId IS NULL;
-- Resposta: 0

--- clientes únicos
SELECT COUNT (DISTINCT ClientId)
FROM CreditStatus;
-- Resposta: 34.763, tal como o tamanho da base

----- conhecendo a base de dados FirstInvoiceDate
SELECT *
FROM FirstInvoiceDate;

--- tamanho da base de dados FirstInvoiceDate
SELECT COUNT(*)
FROM FirstInvoiceDate;
-- Resposta: 114.559

--- faturas nulas
SELECT COUNT(InvoiceId)
FROM FirstInvoiceDate
WHERE InvoiceId IS NULL;
-- Resposta: 0

--- quantidade de clientes únicos na base FirstInvoiceDate
SELECT COUNT(DISTINCT ClientId)
FROM FirstInvoiceDate;
-- Resposta: 22.116. Ou seja, os clientes se repetem 92.443.

--- período analisado considerando as safras
SELECT MIN(FirstInvoiceDate), MAX(FirstInvoiceDate)
FROM FirstInvoiceDate;
-- Resposta: a safra inicia em 01/01/2019 e finaliza em 25/12/2019

.
----- conhecendo a base de dados Client
SELECT *
FROM Client;

--- tamanho da base de dados FirstInvoiceDate
SELECT COUNT()
FROM Client;
-- Resposta: 58005

--- clientes únicos e não nulos
SELECT sum(CASE WHEN ClientId IS NULL THEN 1 ELSE 0 END) AS Nulos, COUNT(ClientId) AS NaoNulos
FROM Client;
-- Resposta: Nulos: 18.089 e Não Nulos: 39.916

--- clientes únicos
SELECT COUNT(DISTINCT ClientId)
FROM Client;
-- Resposta: 39.916. Igual a quantidade de valores não nulos

--- quantidade de clientes que já tiveram seu registro aprovado
SELECT MIN(RegisterApprovalDate), MAX(RegisterApprovalDate)
FROM Client;
-- Resposta: os registros vão de 01/01/2017 a 31/12/2019
```

CreditStatus

- 34.763 observações
- sem dados ausentes ou duplicados

InvoiceStatus

- 114.559 observações
- 22.116 clientes únicos
- safra - período 01/01/2019 a 25/12/19
- relacionamento com CreditStatus gera inconsistência nas informações relativas a CreditStatus

Client

- 58.005 observações
- 18.089 clientes identificados e 39.916 nulos
- registro - período 01/01/2017 a 31/12/2019





dataviz e respostas

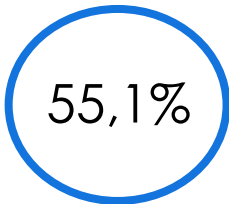
etapa 1

perguntas

- Qual a taxa de aprovação dos clientes?
- Quais características dos clientes apresentam maior correlação com a resposta de aprovação do motor de crédito?

respostas: conhecendo os aprovados

taxa de aprovação de crédito

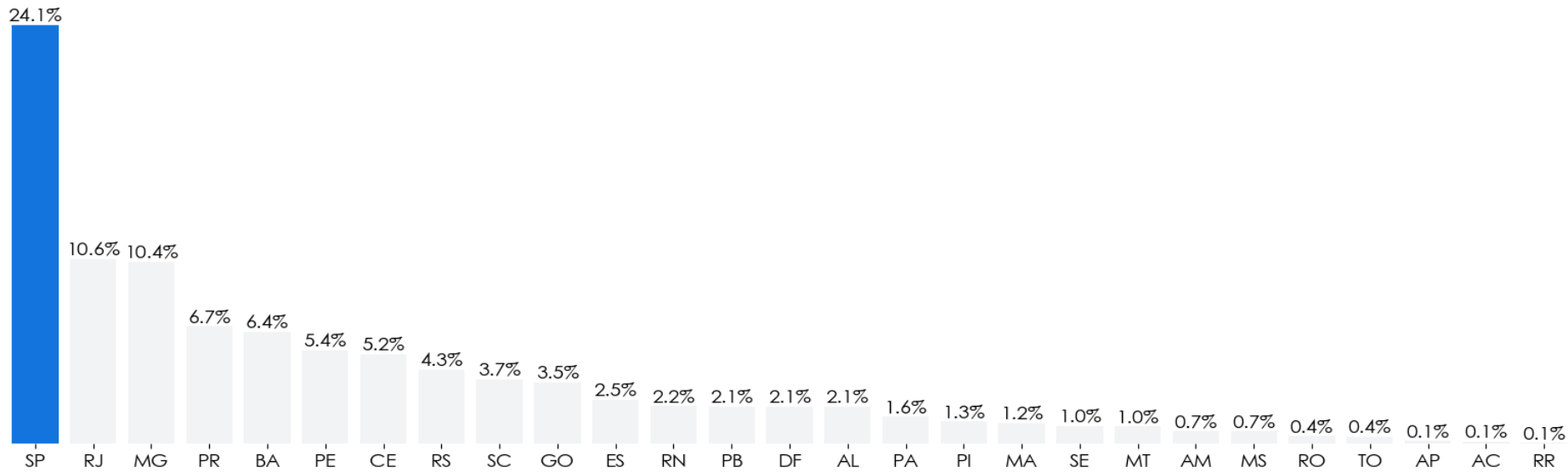


taxa de aprovação por gênero

masculino	<u>51,3%</u>
feminino	45,6%
desconhecido	3,1%

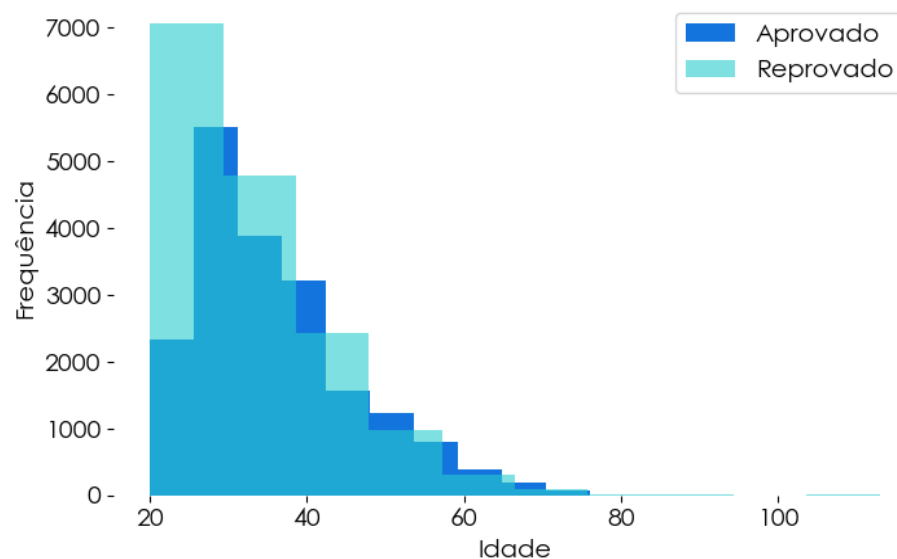


taxa de aprovação por estados

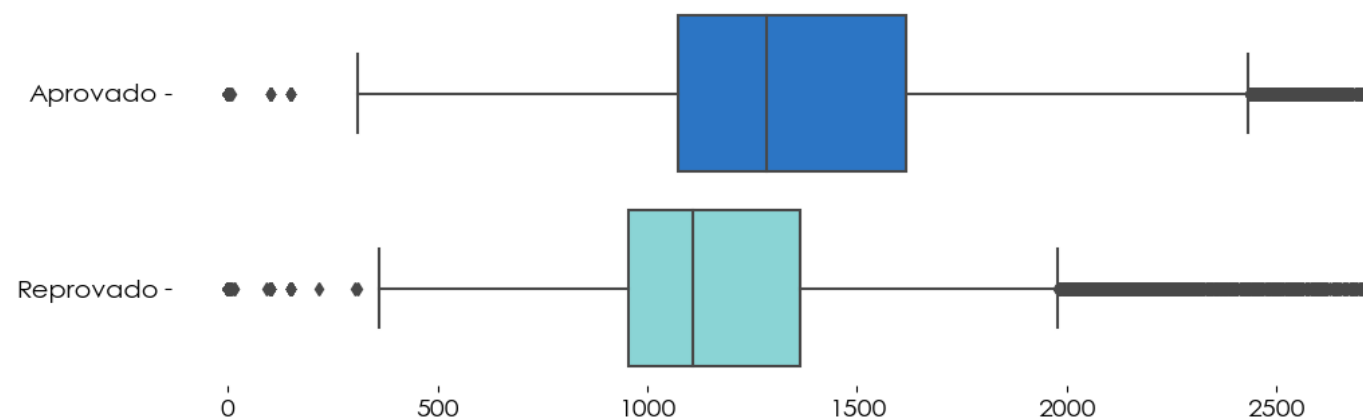


respostas: conhecendo os aprovados

diferença de idade entre aprovado e reprovado



diferença de renda entre aprovado e reprovado



etapa 2

perguntas

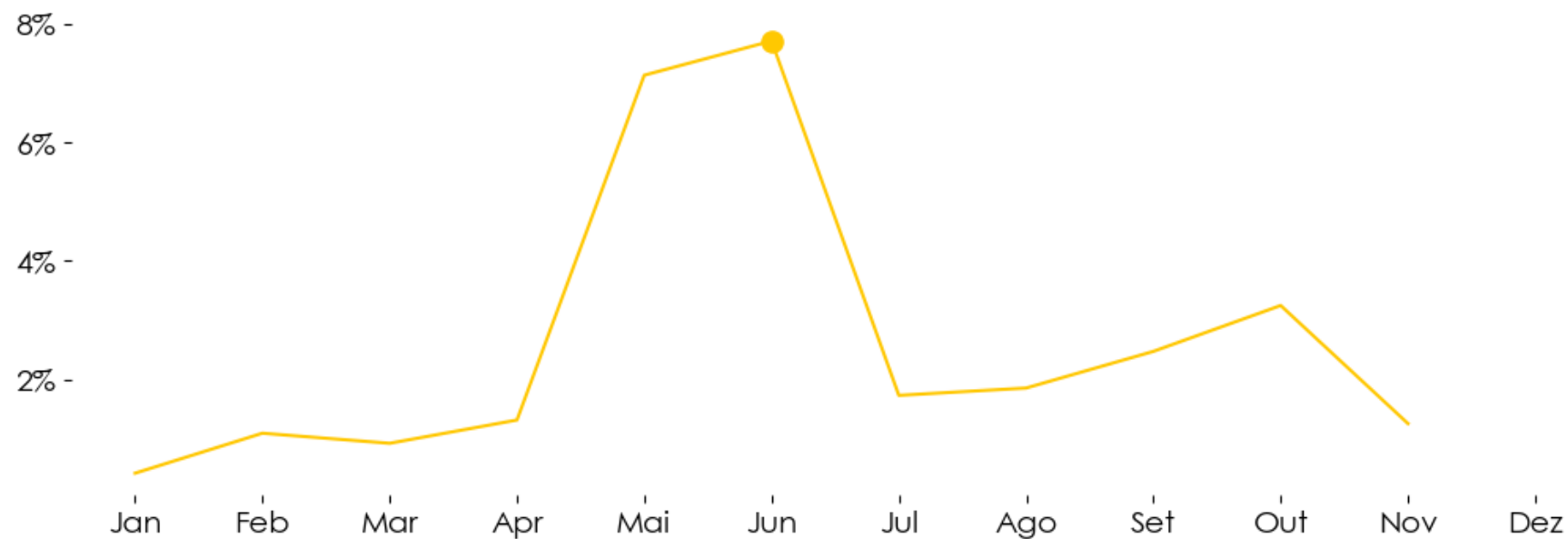
- Qual o Over 30 mob3 para a Neon em uma visão geral?
- Qual o Over 30 mob3 para cada uma das safras (mensais)?
- Utilizando as informações fornecidas neste documento, avalie quais são as variáveis mais relevantes para a taxa de inadimplência do cartão de crédito Neon.

respostas: taxa de over 30 mob3

taxa de over 30 mob3 geral

3,19%

taxa de over 30 mob3 por safras

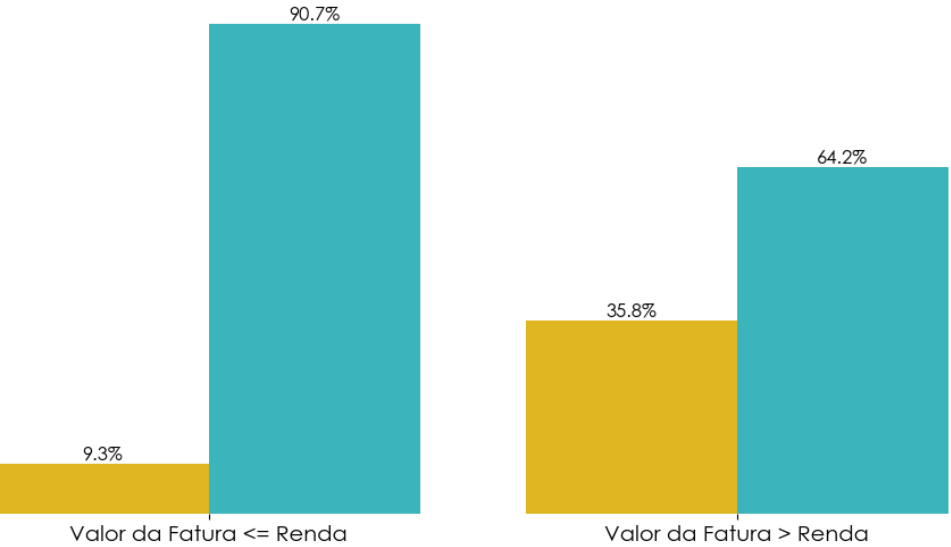


respostas: inadimplência e características associadas

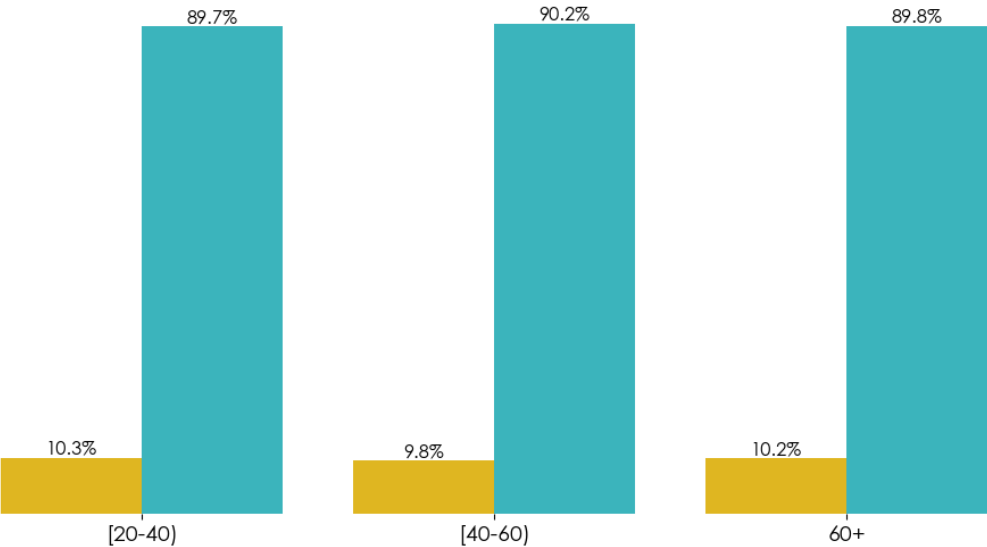
taxa de inadimplência geral 10,1%



valor da fatura x renda presumida

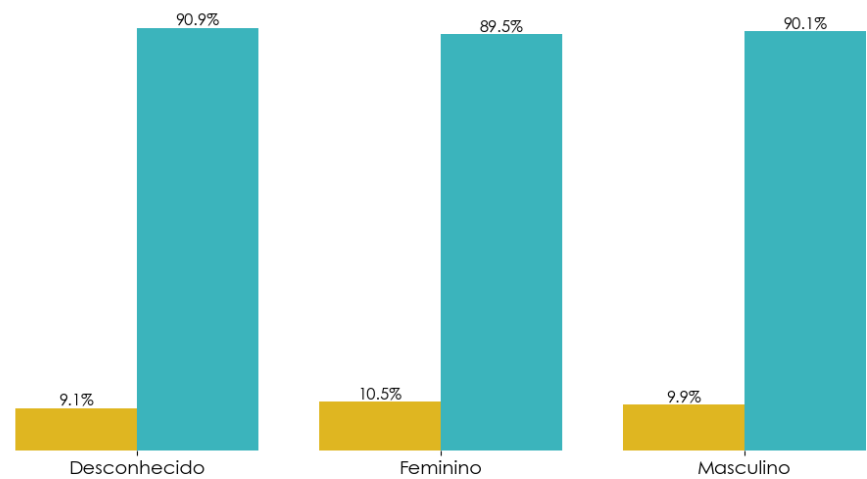


faixa de idade



respostas: inadimplência e características associadas

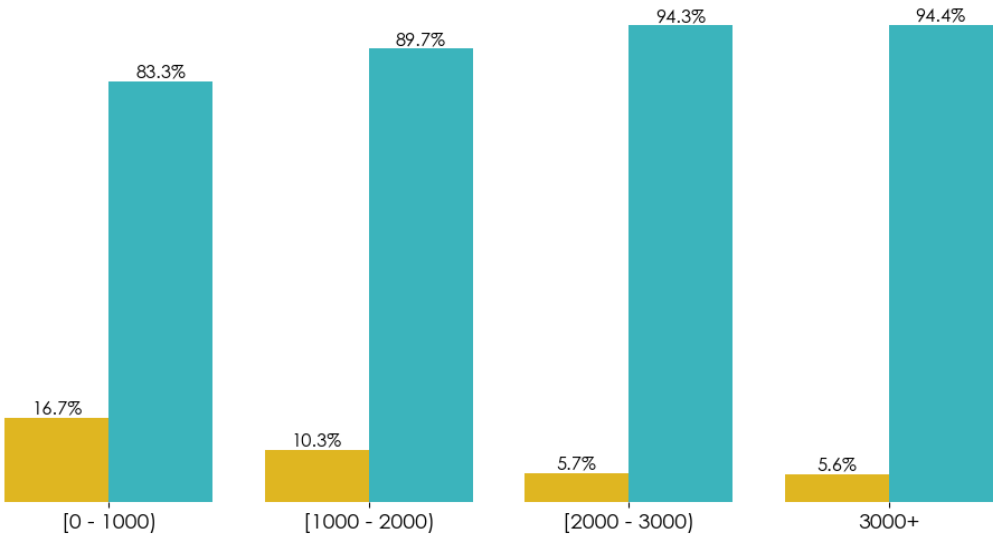
gênero



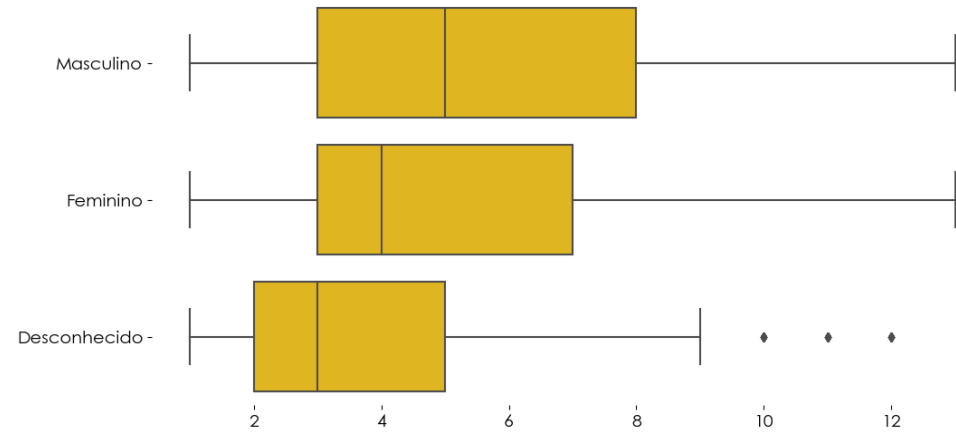
legenda



faixa de renda



recorrência de faturas inadimplentes



etapa 3

perguntas

“CLIENTES DAS SAFRAS A PARTIR DE JULHO/2019 POSSUEM UMA TENDÊNCIA MAIOR À INADIMPLÊNCIA”

- A percepção do time de Cartão de Crédito está correta?
- Por qual motivo o time de Cartão de Crédito pode ter chego a esta percepção?
- Quais são possíveis causas do aumento da inadimplência observada em meses específicos?

respostas: inadimplência e os diferentes cálculos

por safra

Ano Safra	Mês Safra	% Inadimplência
2019	Jan	6.72
2019	Feb	7.79
2019	Mar	7.47
2019	Apr	7.65
2019	Mai	18.59
2019	Jun	19.18
2019	Jul	6.90
2019	Ago	6.16
2019	Set	6.26
2019	Out	6.03
2019	Nov	4.29
2019	Dec	0.95

10,6%

9,48%

por data da fatura

Ano Fatura	Mês Fatura	% Inadimplência
2019	Feb	2.29
2019	Mar	5.93
2019	Apr	5.48
2019	Mai	5.01
2019	Jun	10.94
2019	Jul	20.83
2019	Ago	17.84
2019	Set	10.12
2019	Out	8.24
2019	Nov	7.87
2019	Dec	7.22
2020	Jan	6.60

6,97%

11,23%

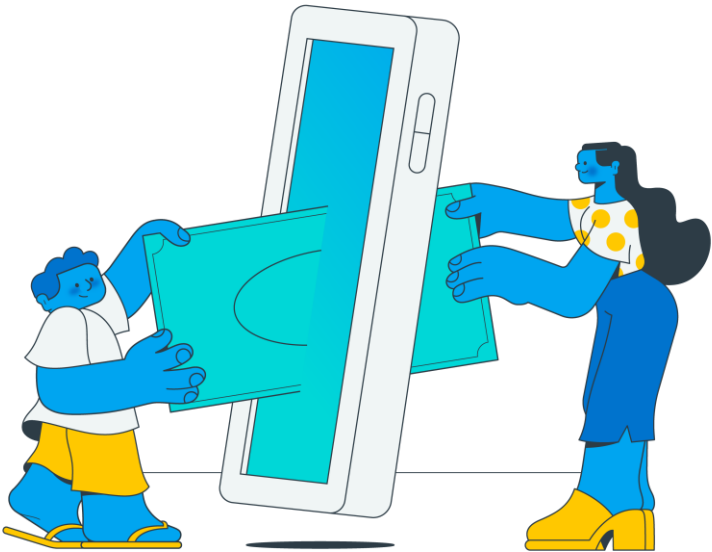
respostas: hipóteses para explicar a inadimplência

H1 – houve um volume maior de clientes com registro aprovado nos meses com maior número de inadimplência?

Mês Registro	Clientes
Ago	1985
Apr	1952
Set	1889
Mai	1800
Mar	1732
Jul	1585
Jan	1467
Jun	1467
Out	1388
Nov	1094
Feb	1046
Dec	65

H2 – clientes que se registraram em 2019 e que tiveram data de aprovação de registro até maio tiveram taxas maiores de inadimplência?

Mês Safra	Jan	Feb	Mar	Apr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dec
Registro												
Registro em 2019 até maio	8.71	9.81	8.67	8.37	19.82	19.26	5.76	4.02	4.65	4.36	4.99	1.10
Registro em outro período	6.65	6.41	5.43	5.48	15.71	18.94	7.80	6.62	6.49	6.24	4.23	0.94



Obrigada!
