



Modelo Preditivo de Grupos de risco de Imóveis

Produto financeiro: Consórcio



lab.data

24/10/2023

Pós-Graduação em Analytics e Inteligência Artificial – Data Science

**Nome do Aluno:**

Roberto Paulo Sianciulis Júnior

Coordenadores:

Profª Drª Alessandra de Ávila Montini

Profª Dr. Adolpho Walter Pimazoni Canton



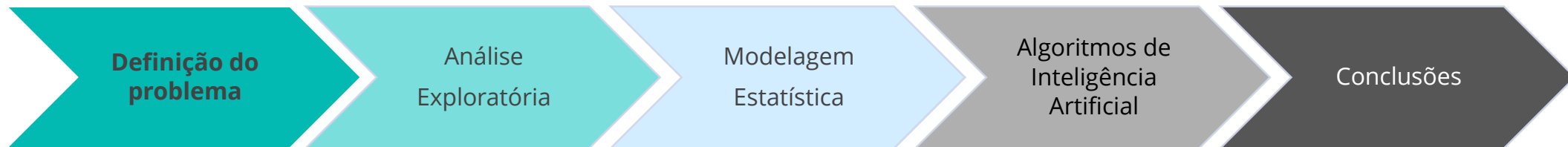
Agenda

- 1. Objetivo do Trabalho
- 2. Contextualização do Problema
- 3. Base de Dados
 - i. Bases originais
 - ii. Filtros
 - iii. Principais variáveis
 - iv. Processo de redução de variáveis
- 4. Análise Exploratória de Dados
- 5. Modelagem com Estatística Tradicional
- 6. Modelagem com Inteligência Artificial
- 7. Conclusões
- 8. Sugestão para Trabalhos Futuros



Metodologia de Análise de dados

4



Planejamento

- Objetivos
- Conceitos
- Critérios
- Histórico de dados
- Variáveis

Análise Preliminar

- Medidas resumo
- Análise bivariada
- Correlações
- Redução de variáveis

Avaliação das técnicas

- Regressão Logística
- Árvore de Decisão

Avaliação das técnicas

- Random Forest
- SVM
- Gradient Boosting Tree
- XGBoosting
- LightGBM
- CatBoost

Definição da técnica

- Validação dos resultados com negócios
- Escolha da técnica que melhor se adequa ao uso e estratégias da área de negócio



O que é **Consórcio**?

5

Juntar em grupos pessoas que desejam um mesmo bem, sendo que um sorteio define quem pode pegar o valor para comprar o bem a cada mês.

Etapas principais do **Consórcio**:



Consórcio não é como um empréstimo ou financiamento.
Nesses casos, o cliente recebe o valor do bem na hora da contratação.

Com o Consórcio, o cliente se planeja para o futuro, pagando parcelas menores e recebe o crédito após a contemplação!





Planejamento

- Objetivos
- Conceitos
- Critérios
- Histórico de dados
- Variáveis

Análise Preliminar

- Medidas resumo
- Análise bivariada
- Correlações
- Redução de variáveis

Avaliação das técnicas

- Regressão Logística
- Árvore de Decisão

Avaliação das técnicas

- Random Forest
- SVM
- Gradient Boosting Tree
- XGBoosting
- LightGBM
- CatBoost

Definição da técnica

- Validação dos resultados com negócios
- Escolha da técnica que melhor se adequa ao uso e estratégias da área de negócio



O consórcio é um **produto planejado** e uma simples alteração nos indicadores pode comprometer a perenidade de toda a operação.

As Administradoras tem verificado um **aumento expressivo em cancelamento** e em inadimplência das cotas nos últimos anos e a necessidade de **construção de um modelo preditivo** de grupos de risco de imóveis.

→ Cenário de alta competitividade no segmento de imóveis + dinâmica peculiar do produto

→ Modelo: **grande diferencial competitivo que** auxiliará a administradora a trabalhar de forma otimizada com seus grupos de imóveis.



Objetivo

Predizer grupos de risco de consórcio, especificamente da categoria de **imóveis**.

Como

Utilizando **dados históricos**, **modelos estatísticos** e **algoritmos de Inteligência Artificial**, que selecionarão as **características mais relevantes** que explicam o perfil de alto ou baixo risco de um grupo de consórcio de imóveis.

Para

Traçar **estratégias de relacionamento**, desenvolver **régua de comunicação** e **ações preventivas**,
Minimizar perdas financeiras futuras





Base Original

100.261 observações

Base de partida

Base com os registros mensais de grupos de Imóveis de Consórcio, divulgadas em jan/18 até ago/22 (56 meses).



Filtro 1 – Exclusões de observações de grupos formados antes de 2018

19.139 observações

Redução da base 1

Para criação do indicador perc_churn, foi fundamental selecionar somente grupos criados de 2018 em diante.



Filtro 2– Exclusões: erros de preenchimento

15.971 observações

Redução da base 2

Optou-se pela remoção dos registros dos grupos das administradoras da base (registros em que os valores medios de bem venda estavam zerados, ou a variação de cotas excluídas entre assembleias estava negativa ou o perc_churn estava maior que 1).



Base Final

15.971 observações – 31 Variáveis

Modelagem

Consideraremos usar 80% da base para treino (12.776 registros) e 20% para teste (3.195 registros).



Tratamentos



Detalhamento dos tratamentos



base original



base final

- 1 **Remover** espaços duplos em todos os campos
- 2 **Excluir** a última linha (se não tiver conteúdo válido)
- 3 **Mudar** o campo "#Nome_da_Administradora" para "Nome_da_Administradora"
- 4 **Criar** variável chave (nome_da_Administradora + "_" + Código_do_grupo)
- 5 F1 - **filtrar** base deixando somente grupos de administradoras que começaram em 2018 pra frente (isto é, tem Número_da_assembleia_geral_ordinária=1)
- 6 **Classificar** a base por chave (nome administradora + número do grupo) e Número_da_assembleia_geral_ordinária em ordem crescente
- 7 **Criar** o campo Variação Cotas excluídas --> fórmula: se o Número_da_assembleia_geral_ordinária for igual a 1 do campo "chave" então a Var_cotas_Excluidas é igual a 0, senão sempre fazer a linha abaixo menos a linha de cima pelo campo Quantidade_de_cotas_excluídas
- 8 **Criar** o campo parcela --> fórmula: (Valor_médio_do_bem + (Valor_médio_do_bem*Taxa_de_administração))/Prazo_do_grupo_em_meses
- 9 **Criar** o campo carteira --> fórmula: Quantidade_de_cotas_ativas_em_dia + Quantidade_de_cotas_ativas_contempladas_inadimplentes + Quantidade_de_cotas_ativas_não_contempladas_inadimplentes
- 10 **Adicionar** variável externa a base INCC_DI (utilizando o campo Data_base como chave)
- 11 **Criar** o campo percentual de Churn sobre carteira ativa --> fórmula: Var_cotas_Excluidas / Carteira
- 12 F2 - erros de preenchimento de informação: **tirar** da base grupos de administradoras com valores medios de bem venda zerados, var_cotas_excluidas negativa ou perc_churn_s_carteira >1
- 13 F3 - **criar** o campo condição (grupo "novo" criado em 2021 e 2022 e grupo "antigo" criado anterior a 2021, basta filtrar o Número_da_assembleia_geral_ordinária igual 1 e aplicar as regras segundo a data base)
- 14 **Criar** o campo Faixa_parcelas
- 15 **Criar** o campo percentual de inadimplência sobre carteira ativa --> fórmula: (Quantidade_de_cotas_ativas_contempladas_inadimplentes + Quantidade_de_cotas_ativas_não_contempladas_inadimplentes) / carteira
- 16 **Excluir** o codigo_do_segmento pois avaliaremos um segmento apenas no estudo (código segmento =1 - bens imóveis)
- 17 **Criar** o campo idade --> fórmula: se o campo Número_da_assembleia_geral_ordinária for maior do que Prazo_do_grupo_em_meses então 1, senão Número_da_assembleia_geral_ordinária / Prazo_do_grupo_em_meses
- 18 **Criar** o campo faixa_idade
- 19 **Criar** o campo faixa_prazo
- 20 **Criar** o campo Receita_potencial --> fórmula: carteira * Taxa_de_administração * Valor_médio_do_bem
- 21 **Criar** o campo Risco --> fórmula: se o Perc_churn_s_carteira <=3%, então 0 (risco baixo) senão 1 (risco alto)



Principais variáveis (31)



Variáveis cadastrais (5)

- Chave_nome_adm_e_cod_grupo
- Nome_da_Administradora
- CNPJ_da_Administradora
- Codigo_do_grupo
- Prazo_do_grupo_em_meses



Variáveis transacionais (12)

- Data_base
- Numero_da_assembleia_geral_ordinária
- Valor_medio_do_bem
- Taxa_de_administracao
- Quantidade_de_cotas_ativas_em_dia
- Quantidade_de_cotas_ativas_contempladas_inadimplentes
- Quantidade_de_cotas_ativas_nao_contempladas_inadimplentes
- Quantidade_de_cotas_ativas_contempladas_no_mes
- Quantidade_de_cotas_excluidas
- Quantidade_de_cotas_ativas_quitadas
- Quantidade_de_cotas_ativas_com_credito_pendente_de_utilizacao
- Índice de correção



Variáveis criadas (12)

- Carteira_ativa
- Carteira_inadimplente
- Parcela
- Idade
- Var_cotas_Excluidas
- Perc_churn
- Perc_inadimplencia
- Condição
- Fx_parcela
- Fx_prazo
- Fx_idade
- Receita_potencial



Variável resposta (1)

Risco:
1 = grupo de alto risco
0 = grupo de baixo risco



Variável externa (1)

- INCC_DI

Legenda:

Variáveis de identificação e controle

Variáveis Categóricas

Variáveis Numéricas





Planejamento

- Objetivos
- Conceitos
- Critérios
- Histórico de dados
- Variáveis

Análise Preliminar

- Medidas resumo
- Análise univariada e bivariada
- Correlações
- Redução de variáveis

Avaliação das técnicas

- Regressão Logística
- Árvore de Decisão

Avaliação das técnicas

- Random Forest
- SVM
- Gradient Boosting Tree
- XGBoosting
- LightGBM
- CatBoost

Definição da técnica

- Validação dos resultados com negócios
- Escolha da técnica que melhor se adequa ao uso e estratégias da área de negócio



Análise Exploratória de Dados

Medidas resumo (posição, dispersão)

Análises univariadas e bivariadas

CAT - U NUM - U

CAT - B NUM - B

Correlações

Redução de Variáveis



Referente as variáveis de identificação, destacamos na composição da base de análise:

Variáveis de identificação e controle

- Chave_nome_adm_e_cod_grupo
- Nome_da_Administradora
- CNPJ_da_Administradora
- Codigo_do_grupo
- Data_base

56 períodos mensais (**jan/18 a ago/22**), com os registros dos grupos de consórcio de imóveis.

66 Administradoras de consórcio.

755 grupos.

784 combinações distintas de nomes de Administradoras e códigos de grupos.



Raio-X da base

ANÁLISE EXPLORATÓRIA | UNIVARIADA

14

Variáveis categóricas

- Condição
- Fx_parcela
- Fx_prazo
- Fx_idade
- Índice_de_correção

Fx_parcela	Frequencia
635 - 1270	0.365663
0 - 635	0.337988
1270 - 1905	0.141444
1905 - 2540	0.100870
2540 - 3175	0.043266
3175 - 3810	0.009768
3810 - 4445	0.000626
5715 - 6350	0.000250
6350 - 6985	0.000125

Fx_idade	Frequencia
0 - 0.2	0.934632
0.2 - 0.4	0.064868
0.4 - 0.6	0.000501

Fx_prazo	Frequencia
de 15 a 20 anos	0.747480
de 10 a 15 anos	0.201991
de 5 a 10 anos	0.031557
acima de 20	0.016154
até 5 anos	0.002818

Condição	Frequencia
antigo	0.74222
novo	0.25778

Índice de correção	Frequencia
3	0.567466
99	0.358525
4	0.035001
2	0.020976
6	0.010081
1	0.007952



- O valor das parcelas está concentrado nas faixas mais baixas, entre 0 e 1.270 reais
- A maioria dos grupos ainda encontra-se no início do contrato, faixa de 0 a 20% de idade.
- A maior parte dos grupos possui prazo de duração de 15 a 20 anos de contrato
- Mais de 70% dos registros da base referem-se a grupos antigos (formados antes de 2021)
- Os grupos apresentam concentração de mais de 90% nos índices de correção 3 e 99.





Das 15.971 observações da base, 55,6% são grupos de imóveis considerados de alto risco.

Para **definição do alto ou baixo risco** foi considerado o benchmarking das administradoras que apresentaram o maior share de cotas ativas de imóveis, da base mais recente disponibilizada.

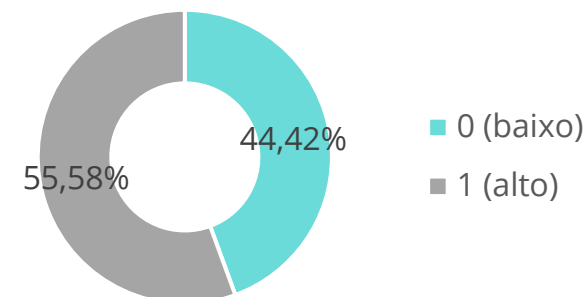
De 50 administradoras de consórcio de imóveis em Agosto/2022, **12 representam mais de 80% do share do mercado.**

Ponderando a média percentual de variação de cotas excluídas sobre a carteira ativa de cada administradora pela sua respectiva participação no mercado, foi definido que:

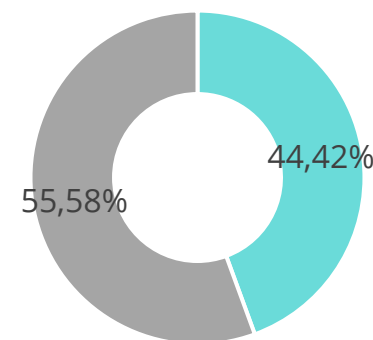
grupos de alto risco (risco =1) possuem variação de cotas excluídas sobre a carteira ativa **maior do que 3%**

grupos de baixo risco (risco=0) apresentam variação de cotas excluídas sobre a carteira ativa **menor ou igual a 3%.**

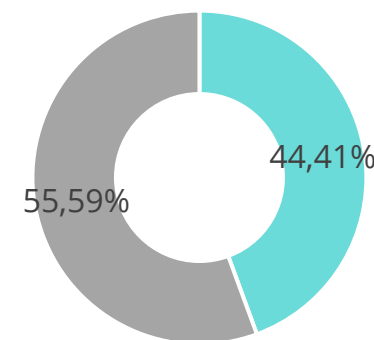
Base para Análise
15.971 observações



Treino
12.776 – 80%



Teste
3.195 – 20%



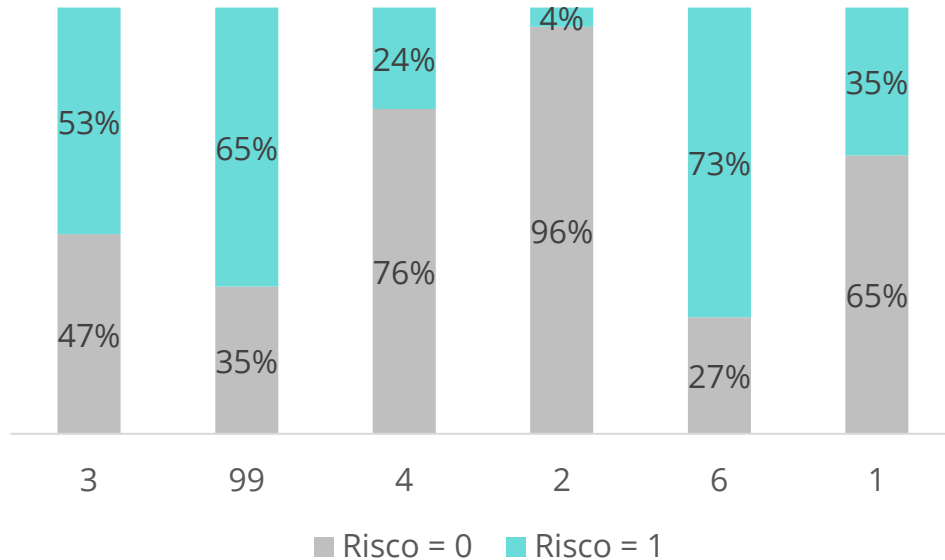
■ 0 (baixo) ■ 1 (alto)

Análise detalhada

ANÁLISE EXPLORATÓRIA | BIVARIADA

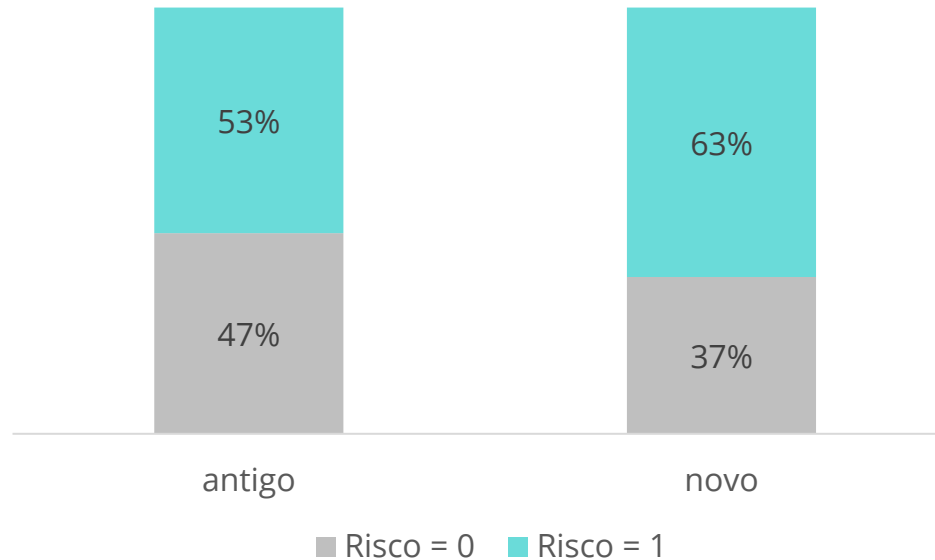
16

Índices de correção x Risco



Contratos indexados por certos índices parecem discriminar bem o alto risco, como é o caso dos índices 6 e 99, bem como o baixo risco, como é o caso dos índices 2 e 4.

Condição x Risco



A distribuição do risco para os grupos antigos é ligeiramente menor, ou seja, os grupos iniciados antes de 2021 tendem a ser um ligeiramente menos arriscados do que os grupos iniciados de 2021 em diante.

LEGENDA



Covariável parece explicar bem a resposta



Covariável parece explicar pouco a resposta



Covariável parece não explicar a resposta

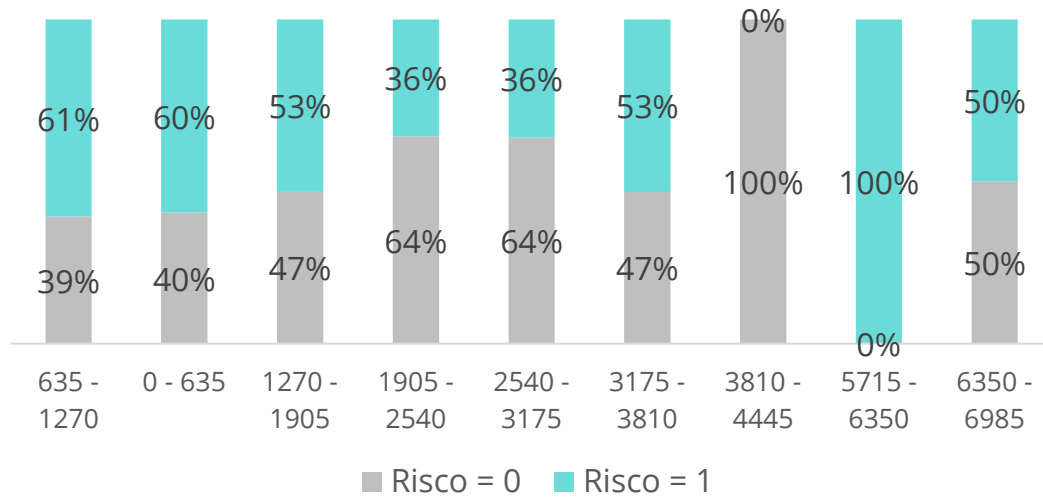


Análise detalhada

ANÁLISE EXPLORATÓRIA | BIVARIADA

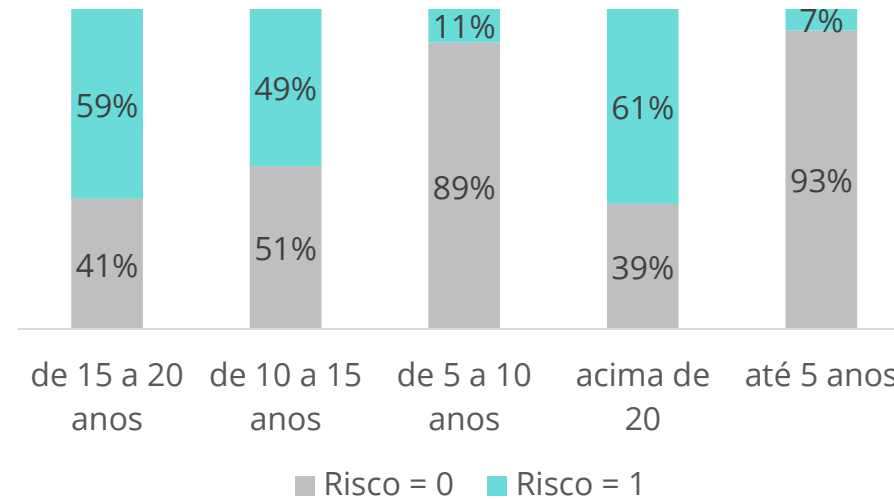
17

Faixas de Parcelas x Risco



Contratos com parcelas menores tendem a ser mais arriscados do que contratos com parcelas maiores.

Faixas de Prazos x Risco



O prazo dos contratos dos grupos parece discriminar bem o risco dos grupos. Prazos mais curtos tendem a apresentar um risco mais baixo do que grupos com contratos de longo prazo (acima de 15 anos)

LEGENDA



Covariável parece explicar bem a resposta



Covariável parece explicar pouco a resposta



Covariável parece não explicar a resposta

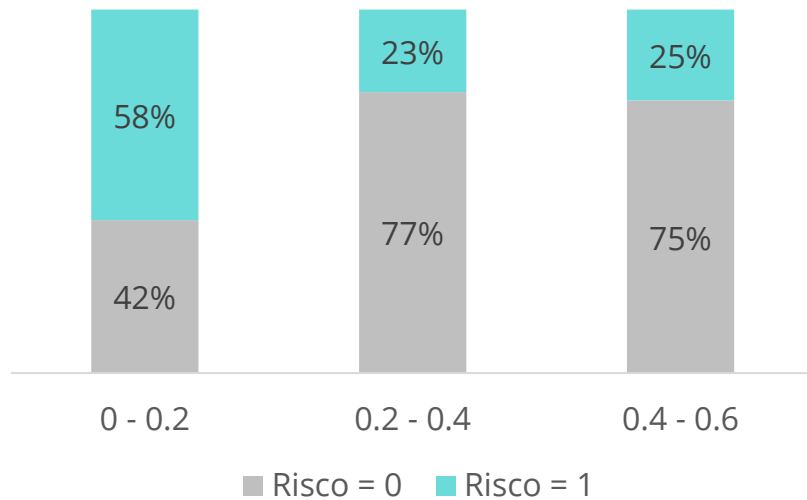


Análise detalhada




ANÁLISE EXPLORATÓRIA | BIVARIADA

18


Faixas de Idade x Risco



LEGENDA

-  Covariável parece explicar bem a resposta
-  Covariável parece explicar pouco a resposta
-  Covariável parece não explicar a resposta



 A composição da idade parece discriminar a variável Risco. Grupos com menos de 20% de cumprimento do prazo planejado, tendem a apresentar mais de 50% de chance de terem risco alto.



Referente as 20 variáveis numéricas, destacamos na composição da base de análise:



Variáveis Numéricas

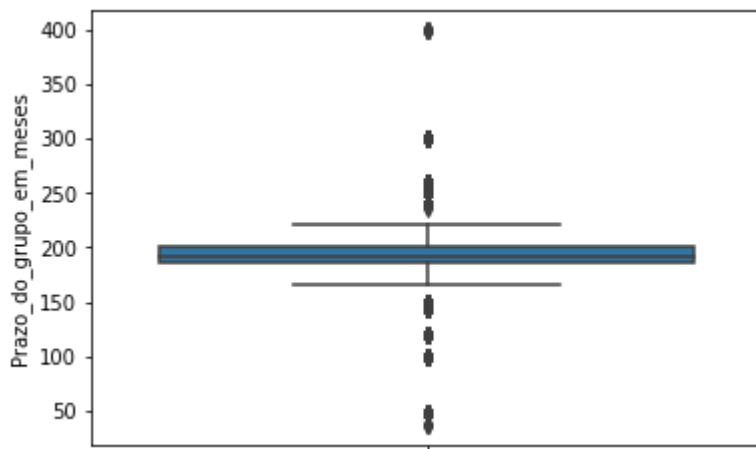
- [Prazo do grupo em meses](#)
- [INCC_DI](#)
- [Numero da assembleia geral ordinária](#)
- [Valor medio do bem](#)
- [Taxa de administracao](#)
- [Quantidade de cotas ativas em dia](#)
- [Quantidade de cotas ativas contempladas inadimplentes](#)
- [Quantidade de cotas ativas nao contempladas inadimplentes](#)
- [Quantidade de cotas ativas contempladas no mes](#)
- [Quantidade de cotas excluidas](#)
- [Quantidade de cotas ativas quitadas](#)
- [Quantidade de cotas ativas com credito pendente de utilizacao](#)
- [Carteira ativa](#)
- [Carteira inadimplente](#)
- [Parcela](#)
- [Idade](#)
- [Var cotas Excluidas](#)
- [Perc churn](#)
- [Perc inadimplencia](#)
- [Receita potencial](#)



Detalhes das análises

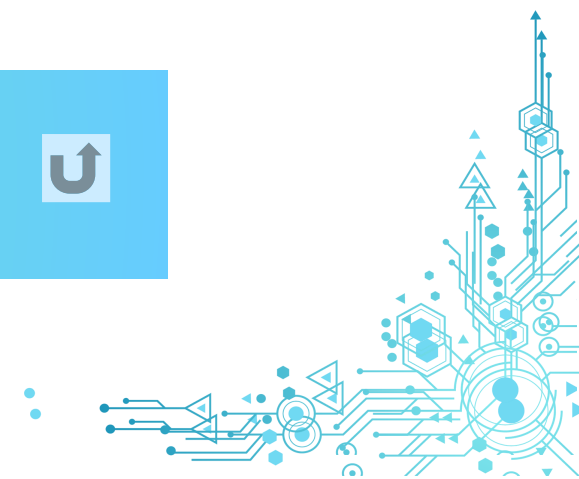
Variáveis numéricas | UNIVARIADA

20



Medida	Valor
Contagem	15971
Mínimo	36
1º Quartil	186
Mediana	192
Média	195.95
3º Quartil	200
Máximo	400
Amplitude	364
Desvio P.	26.09
Coef. Var.	13.31
Assimetria	-0.26
Curtose	7.86

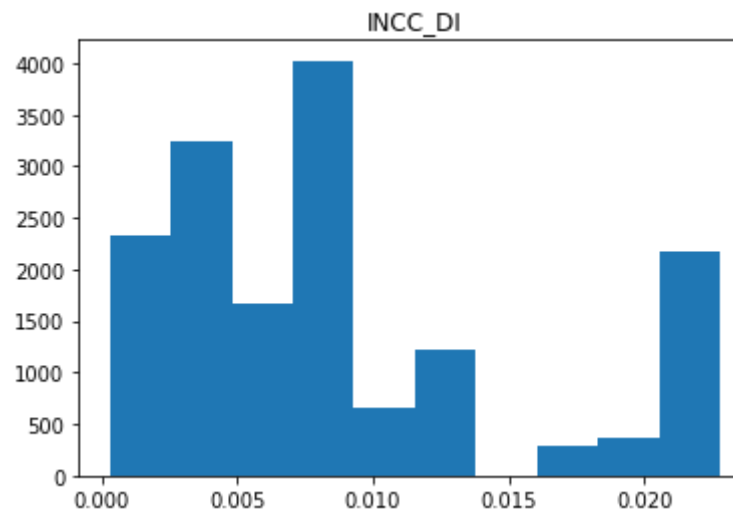
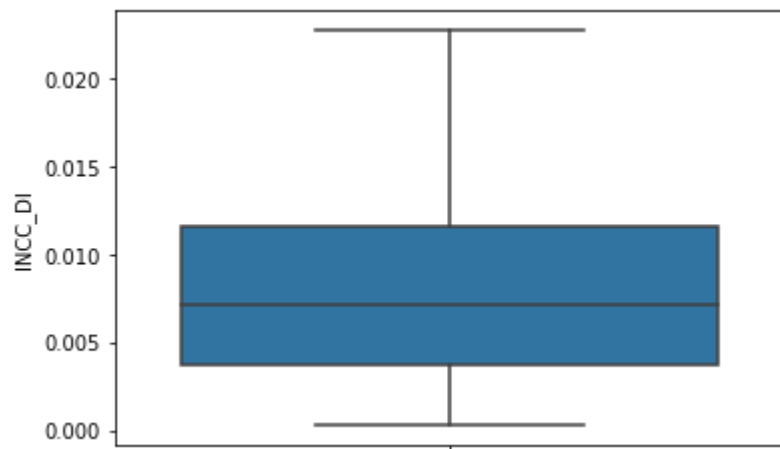
A base apresenta metade dos grupos de consórcio de imóveis com prazos até 192 meses e 75% dos grupos com prazos abaixo de 200 meses. Os grupos mais curtos possuem 36 meses de duração e os mais longos 400 meses. Verificamos uma distribuição assimétrica à direita, sem muita dispersão dos dados em relação à média, indicando poucos grupos com prazos maiores de 200 meses.



Detalhes das análises

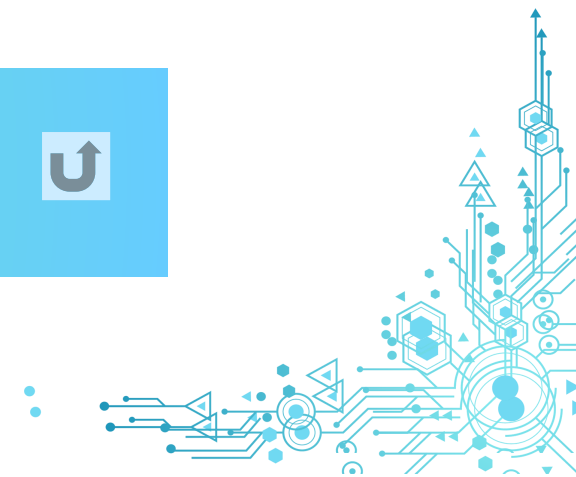
Variáveis numéricas | UNIVARIADA

21



Medida	Valor
Contagem	15971
Mínimo	0.0003
1º Quartil	0.0038
Mediana	0.0072
Média	0.0088
3º Quartil	0.0116
Máximo	0.0228
Amplitude	0.0225
Desvio P.	0.0065
Coef. Var.	74.5951
Assimetria	0.9450
Curtose	-0.1809

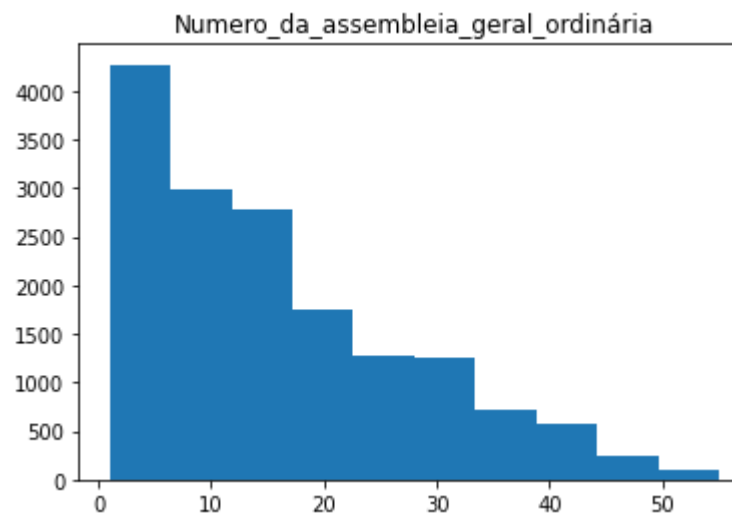
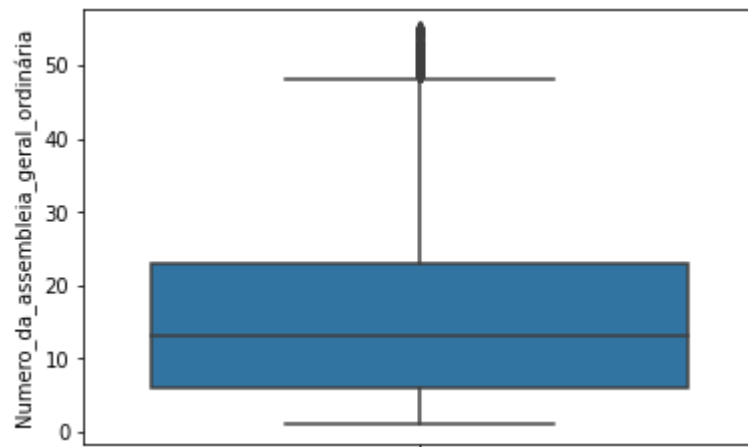
Da base analisada, 50% apresenta índice nacional de custo de construção até 0,7% e somente 25% acima de 1,16%. Essa variável apresenta uma distribuição assimétrica à direita, com uma dispersão grande dos valores em relação à média, sendo o menor valor 0,03% e o maior 2,28%



Detalhes das análises

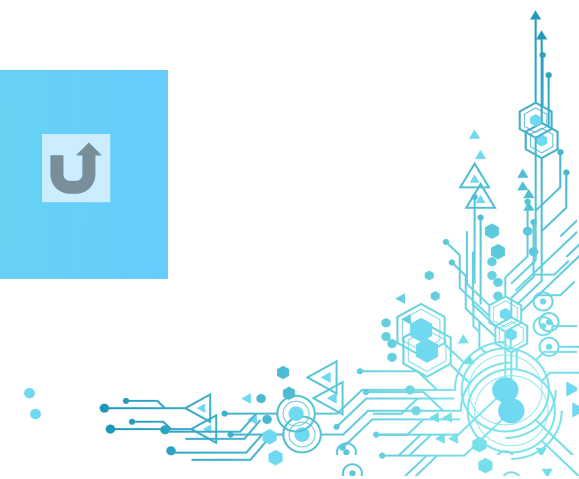
Variáveis numéricas | UNIVARIADA

22



Medida	Valor
Contagem	15971
Mínimo	1
1º Quartil	6
Mediana	13
Média	15.79
3º Quartil	23
Máximo	55
Amplitude	54
Desvio P.	11.93
Coef. Var.	75.54
Assimetria	0.85
Curtose	-0.05

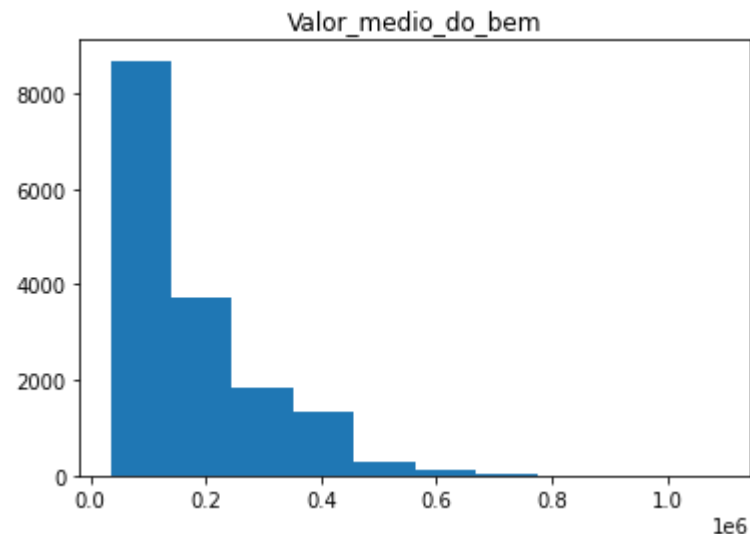
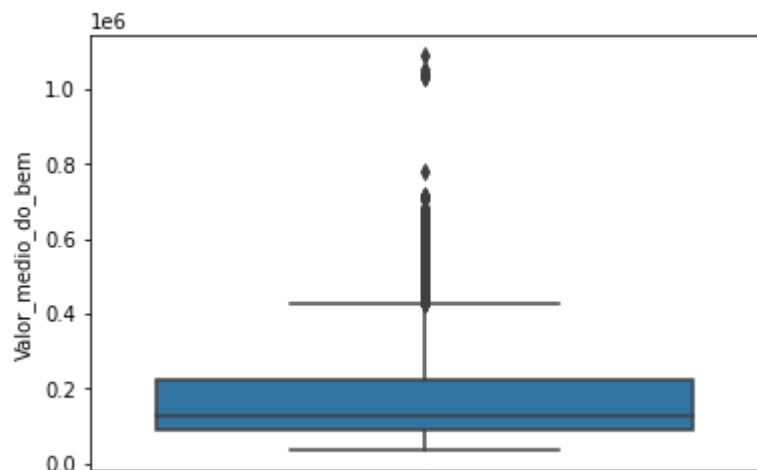
Dos grupos presentes na análise, 50% passaram pela 13ª assembleia e somente 25% já passaram da 23ª assembleia. Tem grupos que estão na 1ª assembleia e grupos mais antigos que já estão na 55ª assembleia. Essa variável apresenta uma distribuição assimétrica à direita, com uma dispersão grande dos valores em relação à média.



Detalhes das análises

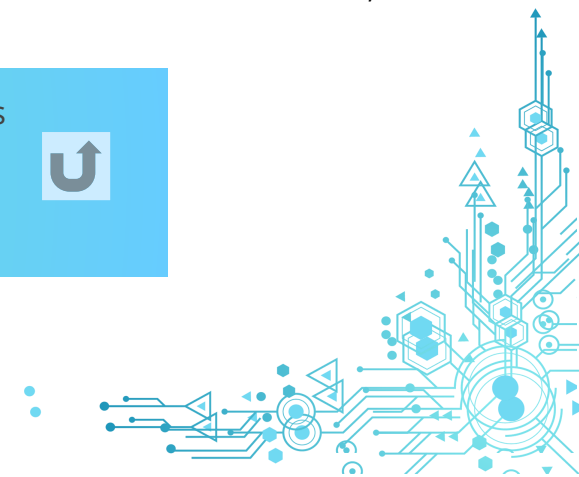
Variáveis numéricas | UNIVARIADA

23



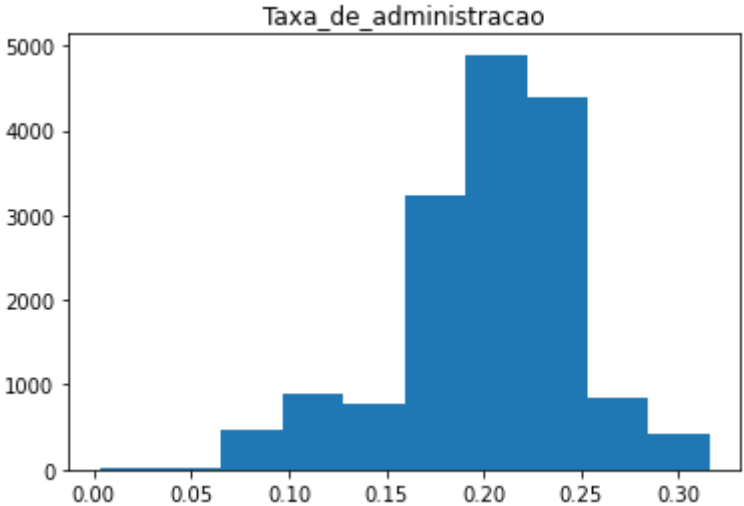
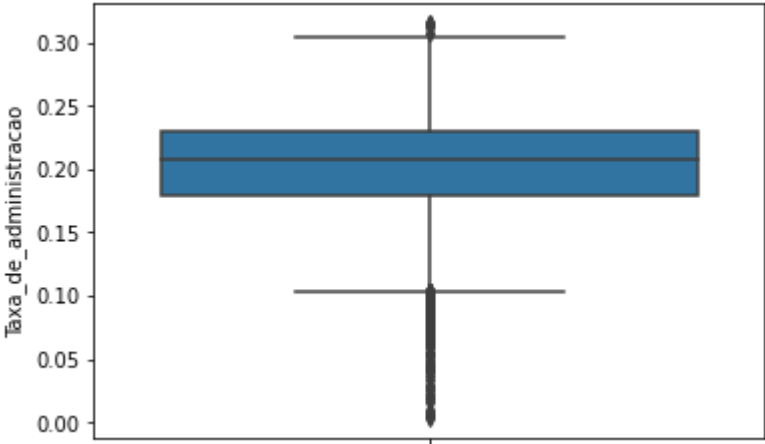
Medida	Valor
Contagem	15.971
Mínimo	34.287,44
1º Quartil	90.393,42
Mediana	126.970,44
Média	176.199,47
3º Quartil	225.792,85
Máximo	1.091.620,88
Amplitude	1.057.333,44
Desvio P.	117.332,65
Coef. Var.	66,59
Assimetria	1,49
Curtose	2,48

50% dos valores das cartas de crédito dos grupos avaliados estão concentrados em uma faixa até R\$126 mil e 25% dos maiores valores, em um faixa acima de R\$225 mil. A menor carta de crédito apresenta valor de R\$ 34 mil e a maior valor de R\$1,09 milhões. Essa variável apresenta uma distribuição assimétrica à direita, com uma dispersão grande dos valores em relação à média.



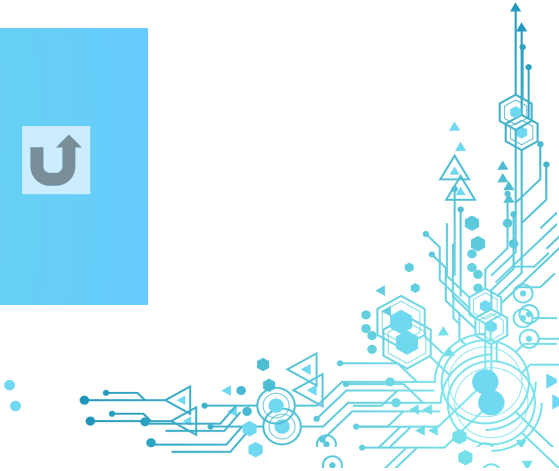
Detalhes das análises

Variáveis numéricas | UNIVARIADA



Medida	Valor
Contagem	15971
Mínimo	0.0028
1º Quartil	0.18
Mediana	0.2081
Média	0.2024
3º Quartil	0.2307
Máximo	0.3159
Amplitude	0.3131
Desvio P.	0.0461
Coef. Var.	22.7873
Assimetria	-0.7144
Curtose	0.7759

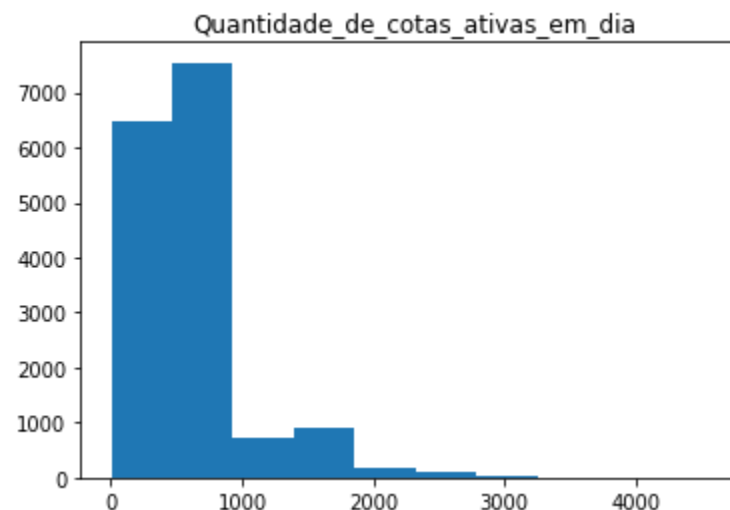
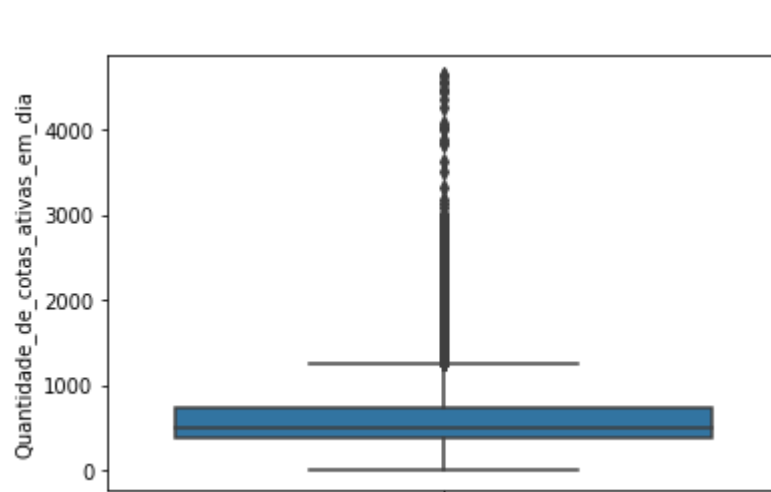
Essa variável apresenta uma distribuição assimétrica à esquerda, com a presença de mais outliers inferiores do que superiores. 50% das taxas de administração dos grupos analisados está acima de 20,81%. A menor taxa do conjunto de dados é 0,28% e a maior é 31,59%. A definição da taxa de administração é bastante estratégica para esse produto, pois representa a receita bruta que cada empresa espera obter com a venda do produto. Pode variar bastante em virtude do momento que a empresa está, de seus objetivos, de campanhas, de pressão dos concorrentes, entre outros fatores.



Detalhes das análises

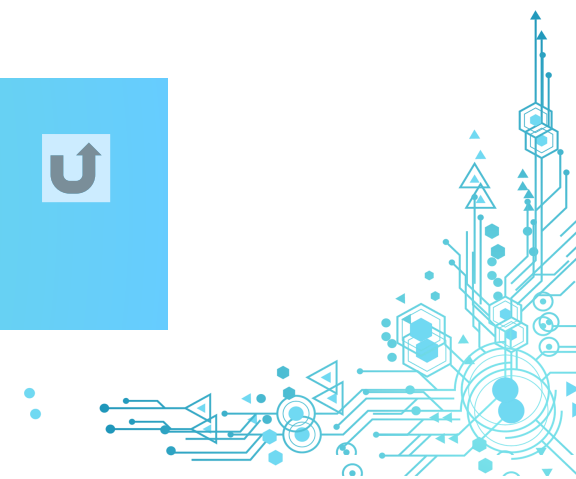
Variáveis numéricas | UNIVARIADA

25



Medida	Valor
Contagem	15971
Mínimo	1
1º Quartil	380
Mediana	506
Média	621.7876
3º Quartil	735
Máximo	4.646
Amplitude	4.645
Desvio P.	433.9867
Coef. Var.	69.7966
Assimetria	2.493
Curtose	9.7315

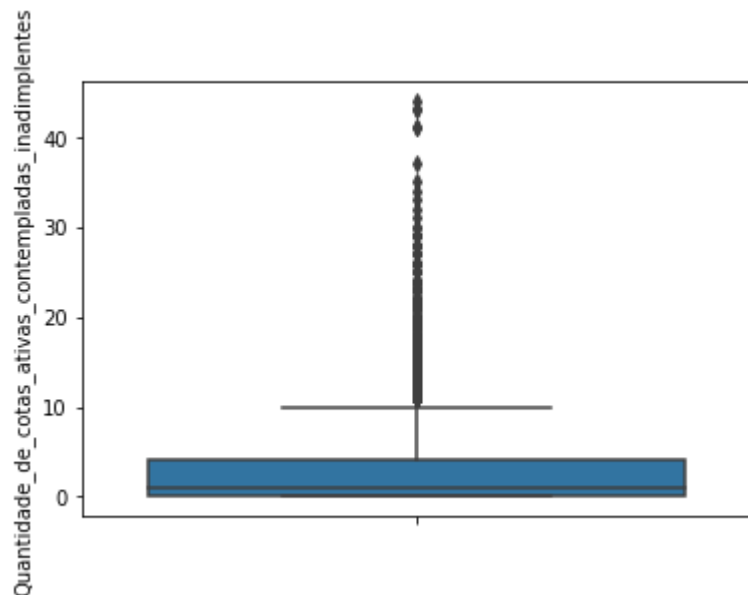
A quantidade de cotas ativas em dia é muito importante do ponto de vista de sustentabilidade do negócio. Para os dados avaliados, 50% dos grupos apresentam até 506 cotas ativas em dia e somente 25% tem mais de 735 cotas ativas em dia. Tem grupos apenas com 1 cota ativa em dia (são grupos que acabaram de iniciar as assembléias ordinárias) e grupos com mais de 1,3 mil cotas ativas em dia, isto é, de empresas com maior market share e não necessariamente de grupos no início do contrato, por exemplo, Banco do Brasil e Caixa. Essa variável apresenta uma distribuição assimétrica à direita, com uma dispersão grande dos valores em relação à média e bastante outliers superiores.



Detalhes das análises

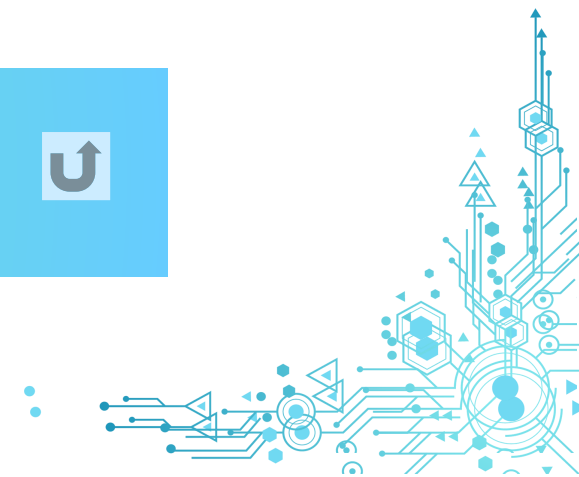
Variáveis numéricas | UNIVARIADA

26



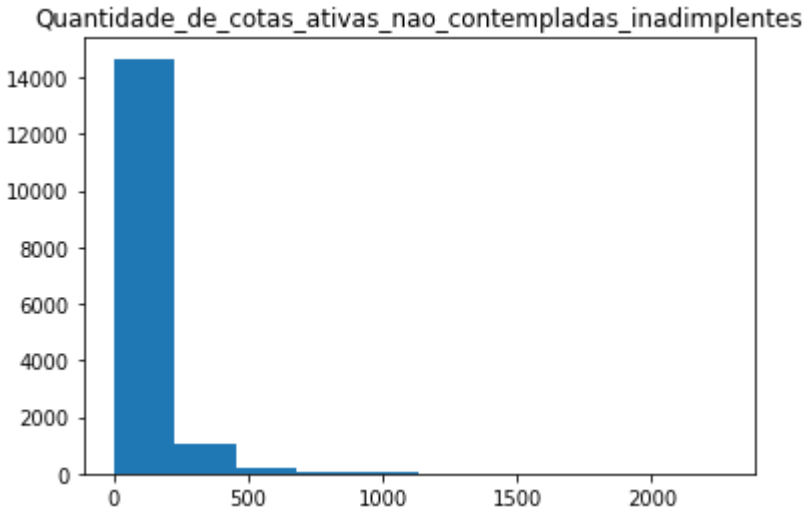
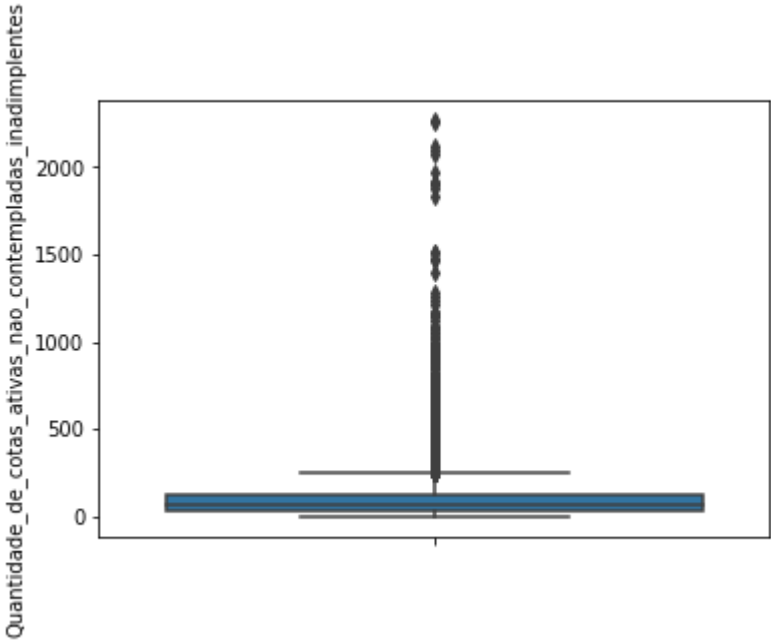
Medida	Valor
Contagem	15971
Mínimo	0
1º Quartil	0
Mediana	1
Média	2.7026
3º Quartil	4
Máximo	44
Amplitude	44
Desvio P.	3.9192
Coef. Var.	145.015
Assimetria	2.7254
Curtose	11.1773

A quantidade de cotas ativas contempladas inadimplentes é baixa, sendo de até 1 cota para 50% dos dados analisados. Os grupos com mais cotas contempladas inadimplentes possuem 44 cotas. Essa variável apresenta uma distribuição assimétrica à direita, com uma dispersão muito grande dos valores em relação à média, com muitos outliers superiores.



Detalhes das análises

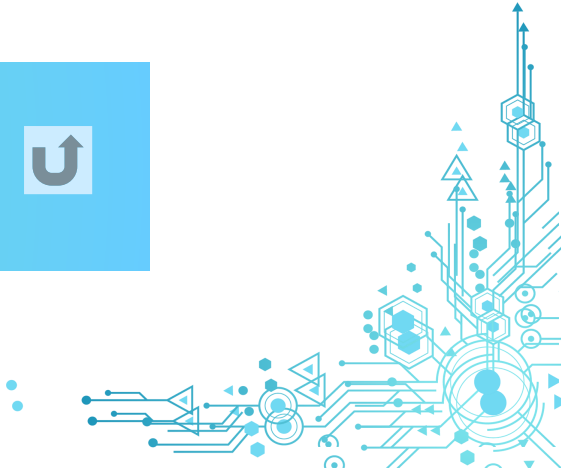
Variáveis numéricas | UNIVARIADA



Medida	Valor
Contagem	15971
Mínimo	0
1º Quartil	31
Mediana	62
Média	95.4355
3º Quartil	118
Máximo	2.267
Amplitude	2.267
Desvio P.	122.2741
Coef. Var.	128.1222
Assimetria	5.5934
Curtose	59.804

Essa variável apresenta uma distribuição assimétrica à direita, com uma dispersão grande dos valores em relação à média e muitos outliers superiores.

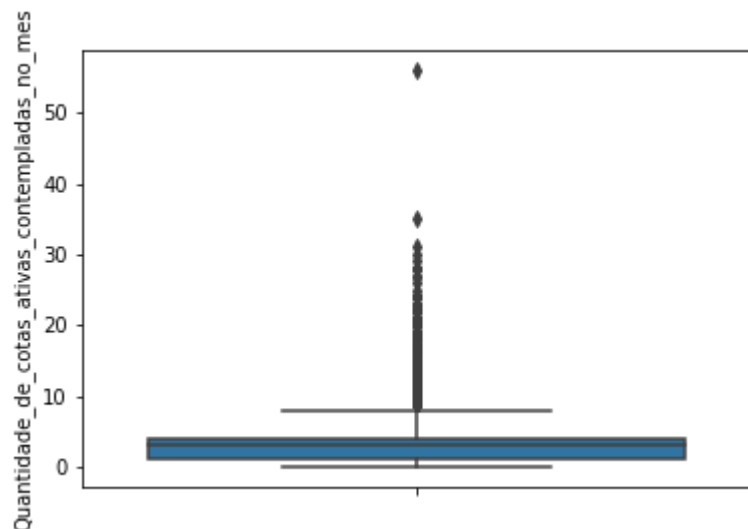
Diferente das cotas ativas contempladas inadimplentes, essa variável apresenta 50% da base com até 62 cotas não contempladas inadimplentes e 25% dos grupos com mais de 118 cotas não contempladas inadimplentes. Observamos maior quantidade de cotas inadimplentes em grupos não contemplados.



Detalhes das análises

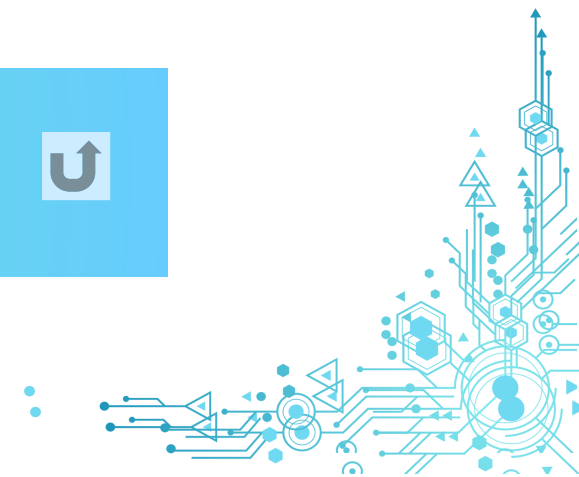
Variáveis numéricas | UNIVARIADA

28



Medida	Valor
Contagem	15971
Mínimo	0
1º Quartil	1
Mediana	3
Média	3.3556
3º Quartil	4
Máximo	56
Amplitude	56
Desvio P.	2.7569
Coef. Var.	82.1585
Assimetria	2.7338
Curtose	18.8212

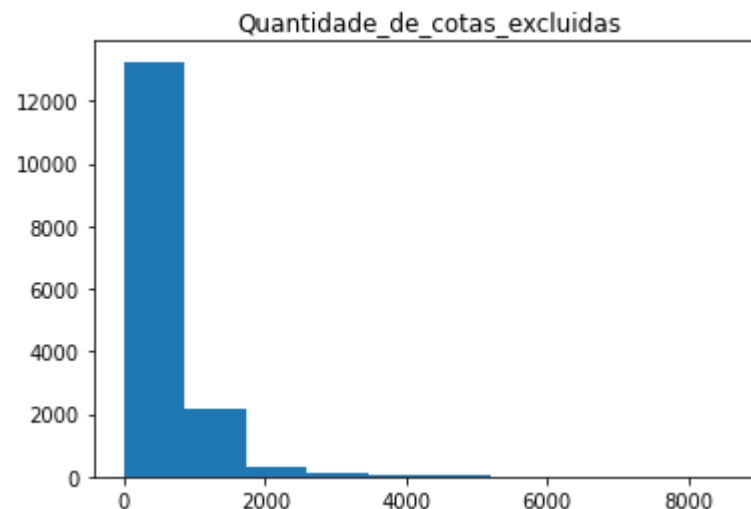
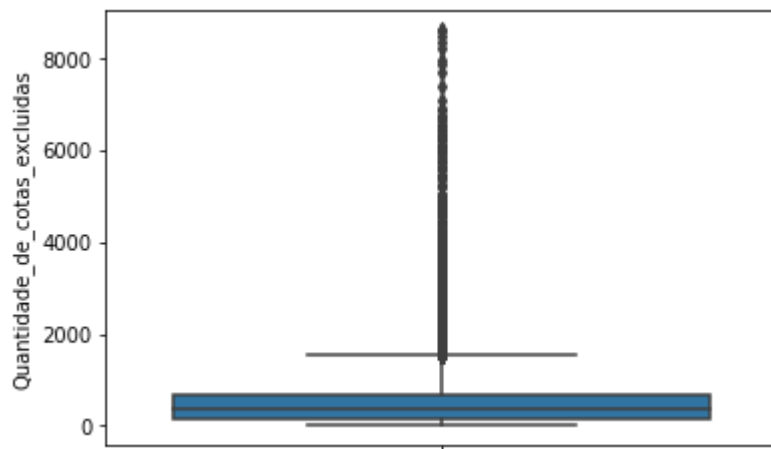
Essa variável apresenta uma distribuição assimétrica à direita, com uma dispersão grande dos valores em relação à média e muitos outliers superiores, sendo o maior de 56 cotas ativas contempladas em um mês, muito provavelmente por conta de alguma campanha em um grupo com prazos menores e valores mais baixos de imóveis.



Detalhes das análises

Variáveis numéricas | UNIVARIADA

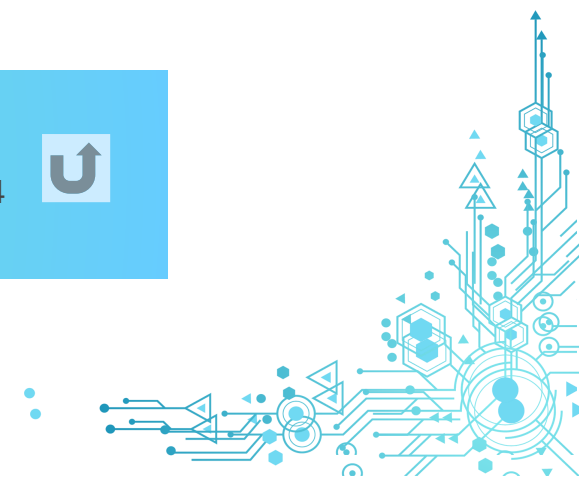
29



Medida	Valor
Contagem	15971
Mínimo	0
1º Quartil	137
Mediana	366
Média	517.4595
3º Quartil	694
Máximo	8640
Amplitude	8640
Desvio P.	626.5488
Coef. Var.	121.0817
Assimetria	4.2661
Curtose	31.8933

Essa variável apresenta uma distribuição assimétrica à direita, com uma dispersão grande dos valores em relação à média e muitos outliers superiores.

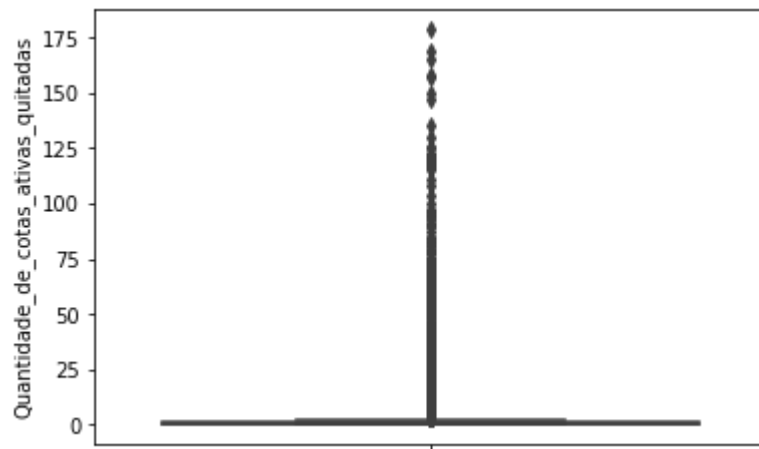
50% dos dados de cotas excluídas estão abaixo de 366 cotas, sendo que em 25% dos dados, os valores são maiores do que 694 cotas, chegando em alguns casos em mais de 8.500 cotas excluídas.



Detalhes das análises

Variáveis numéricas | UNIVARIADA

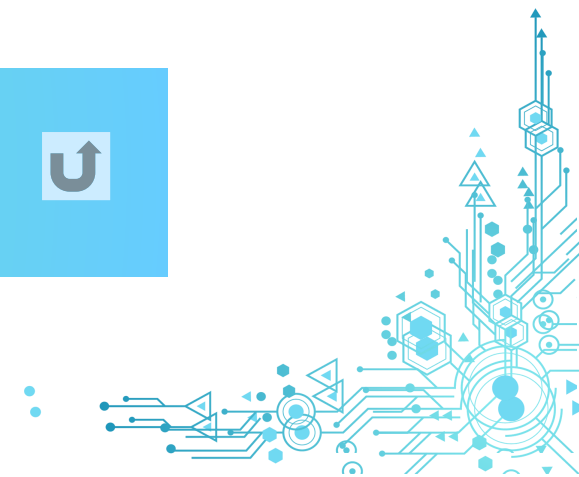
30



Medida	Valor
Contagem	15971
Mínimo	0
1º Quartil	0
Mediana	0
Média	2.5002
3º Quartil	1
Máximo	179
Amplitude	179
Desvio P.	9.3461
Coef. Var.	373.8211
Assimetria	8.7634
Curtose	99.8638

Essa variável apresenta uma distribuição assimétrica à direita, com uma dispersão enorme dos valores em relação à média, muito provavelmente por conta dos outliers superiores.

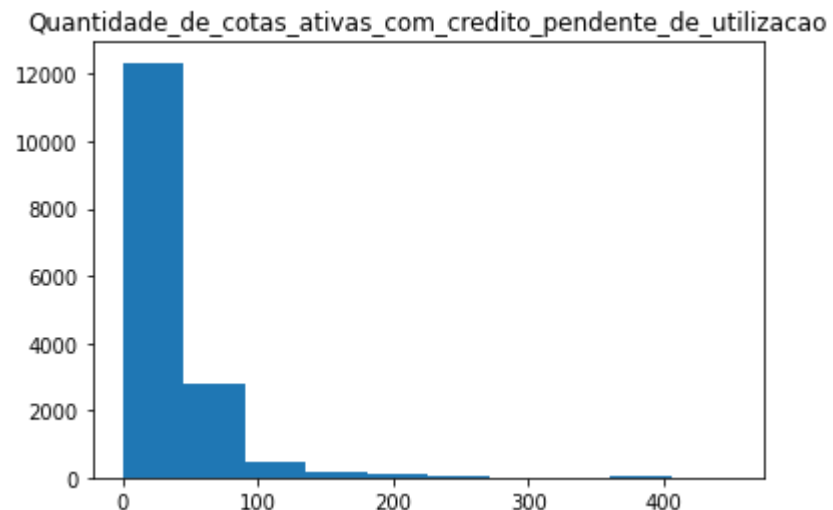
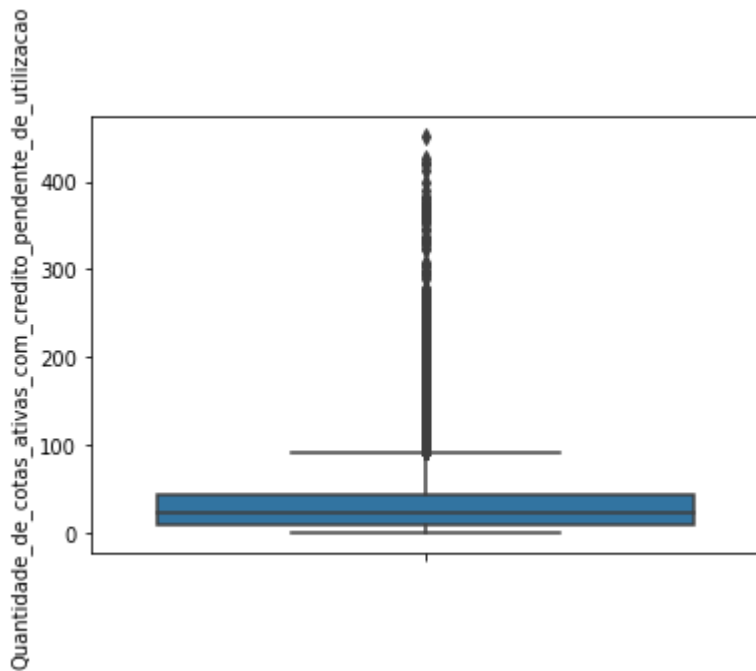
50% dos dados apresentam nenhuma cota ativa quitada, o que pode ser explicado pela idade dos grupos presentes na base (iniciados em 2018) e pelos prazos contratados de imóveis que normalmente são longos.



Detalhes das análises

Variáveis numéricas | UNIVARIADA

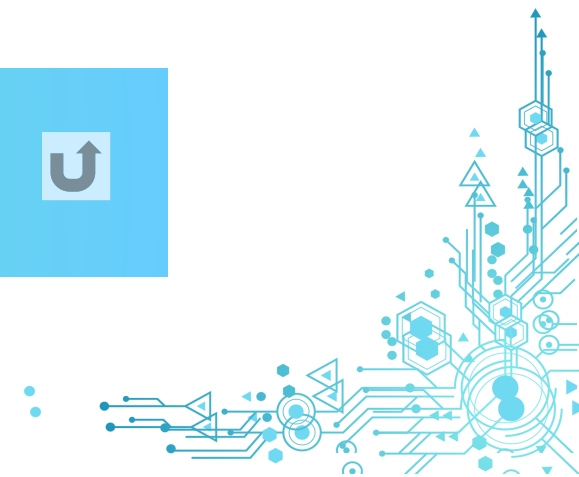
31



Medida	Valor
Contagem	15971
Mínimo	0
1º Quartil	10
Mediana	23
Média	33.0073
3º Quartil	43
Máximo	451
Amplitude	451
Desvio P.	37.7951
Coef. Var.	114.5054
Assimetria	3.796
Curtose	23.1051

Essa variável apresenta uma distribuição assimétrica à direita, com uma dispersão enorme dos valores em relação à média, muito provavelmente por conta dos outliers superiores.

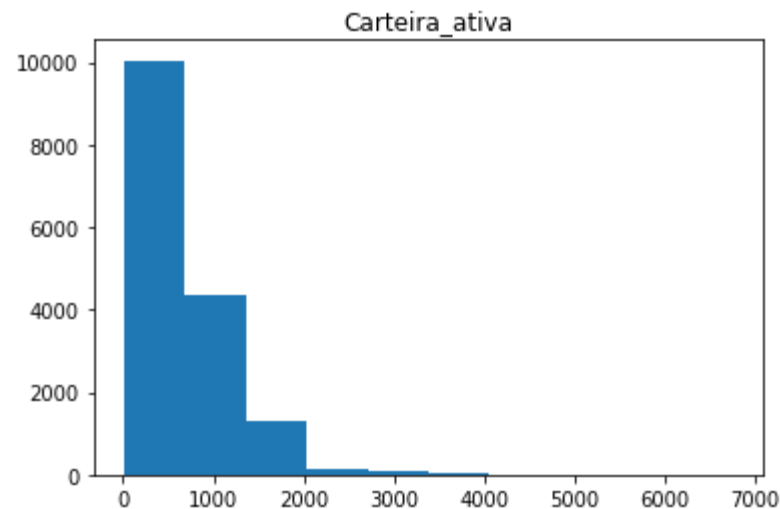
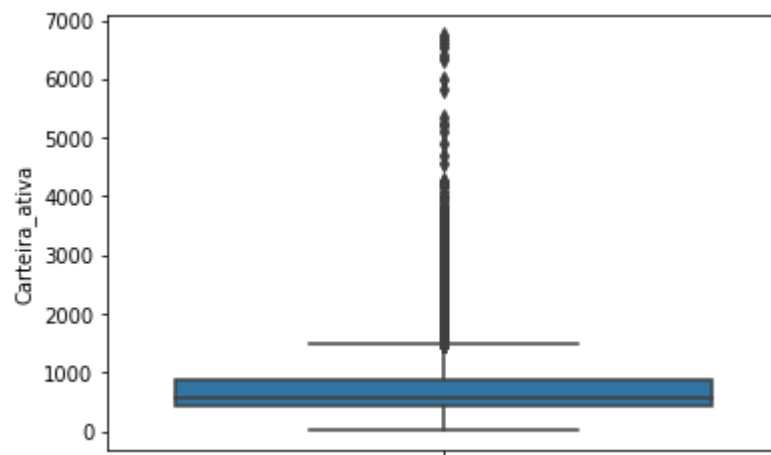
75% da base avaliada apresenta até 43 cotas ativas com crédito disponível para utilização de seus clientes e somente 25% apresentam uma quantidade superior a 43 cotas, chegando a grupos com 451 cotas disponíveis.



Detalhes das análises

Variáveis numéricas | UNIVARIADA

32

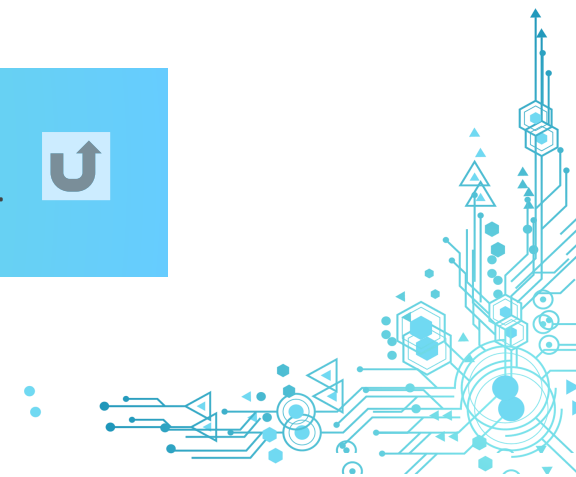


Medida	Valor
Contagem	15971
Mínimo	7
1º Quartil	447
Mediana	584
Média	719.9257
3º Quartil	866
Máximo	6755
Amplitude	6748
Desvio P.	511.3824
Coef. Var.	71.0327
Assimetria	2.9651
Curtose	15.9921

Essa variável apresenta uma distribuição assimétrica à direita, com uma dispersão grande dos valores em relação à média e a presença de muitos outliers superiores.

75% da carteira ativa dos grupos analisados possuem até 866 cotas, sendo que existem grupos com mais de 6.500 cotas ativas.

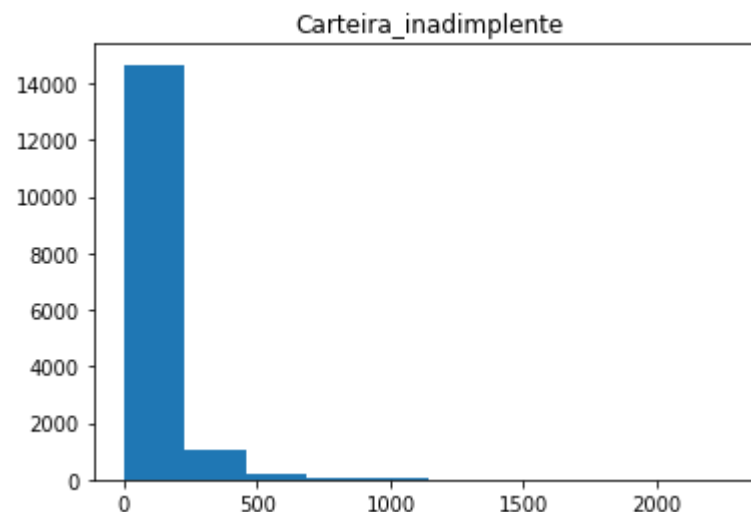
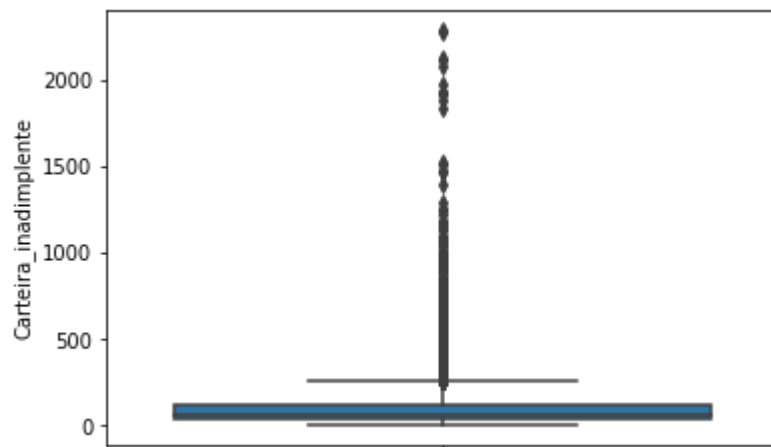
Essa variável parece indicar que existem grupos de empresas com grande participação de mercado nesse segmento.



Detalhes das análises

Variáveis numéricas | UNIVARIADA

33

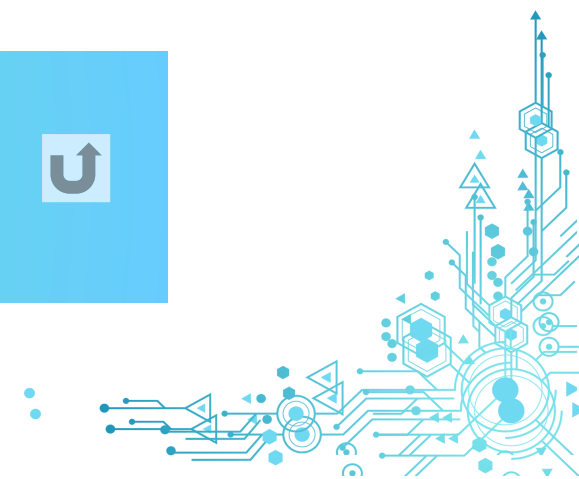


Medida	Valor
Contagem	15971
Mínimo	0
1º Quartil	34
Mediana	65
Média	98.1382
3º Quartil	122
Máximo	2288
Amplitude	2288
Desvio P.	122.8028
Coef. Var.	125.1326
Assimetria	5.5673
Curtose	59.4579

Essa variável apresenta uma distribuição assimétrica à direita, com uma dispersão grande dos valores em relação à média e a presença de muitos outliers superiores.

A média da carteira inadimplente da base analisada é de 98 cotas.

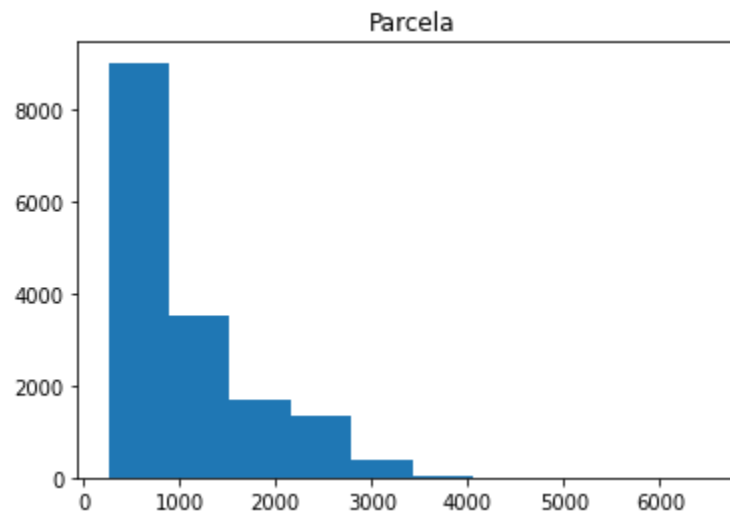
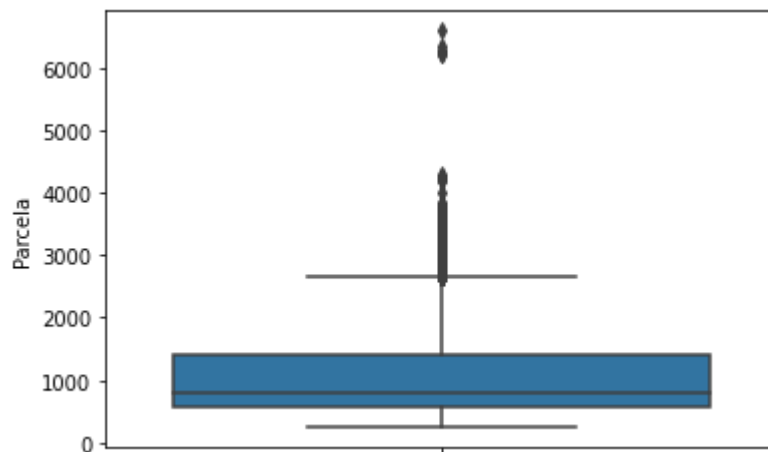
75% da carteira inadimplente dos grupos analisados possuem até 122 cotas, sendo que existem grupos com mais de 2.200 cotas inadimplentes.



Detalhes das análises

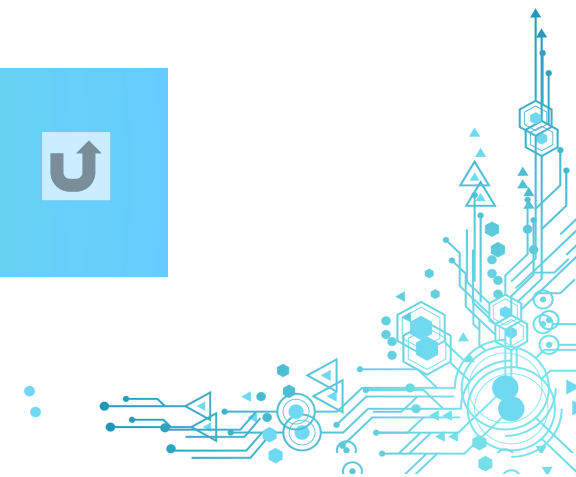
Variáveis numéricas | UNIVARIADA

34



Medida	Valor
Contagem	15971
Mínimo	256.6609
1º Quartil	569.67
Mediana	787.3954
Média	1083.2268
3º Quartil	1403.678
Máximo	6594.5846
Amplitude	6337.9237
Desvio P.	704.0149
Coef. Var.	64.9924
Assimetria	1.4068
Curtose	1.9969

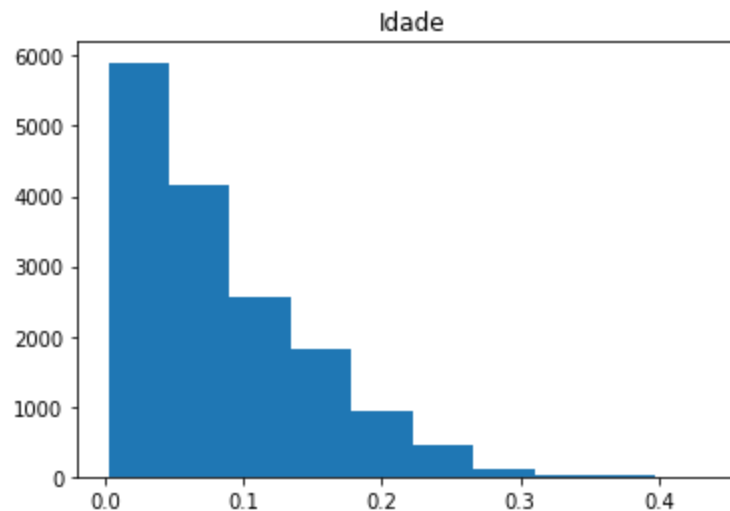
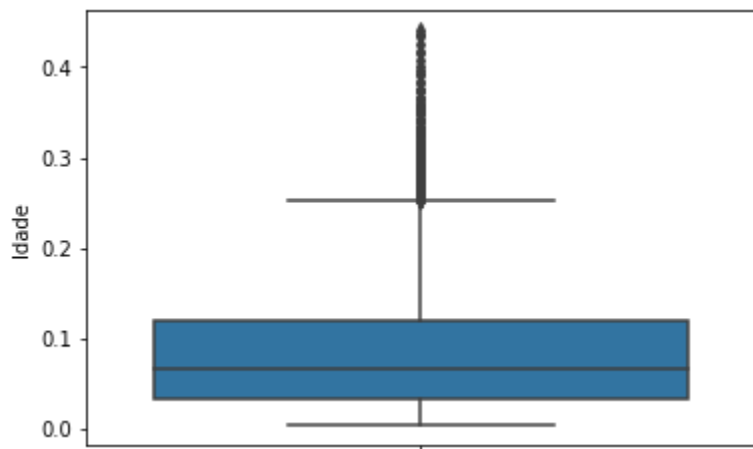
50% dos valores das parcelas para adquirir imóveis encontra-se até R\$ 787, tendo o valor mínimo de R\$ 256. Somente 25% dos valores são superiores a R\$ 1.400, podendo alcançar até R\$ 6.594, dependendo do imóvel desejado. Essa variável apresenta uma distribuição assimétrica à direita, com uma dispersão grande dos valores em relação à média.



Detalhes das análises

Variáveis numéricas | UNIVARIADA

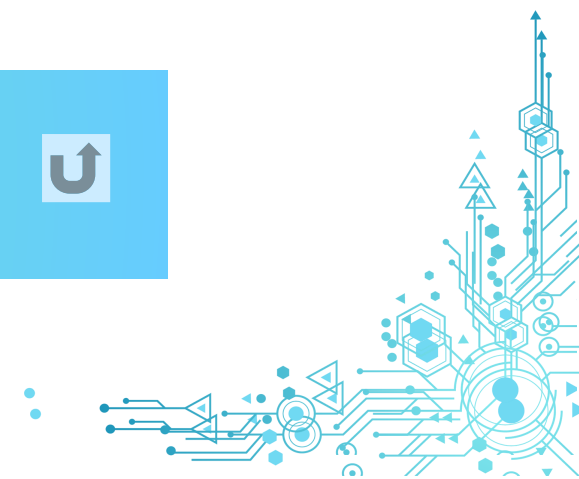
35



Medida	Valor
Contagem	15971
Mínimo	0.0025
1º Quartil	0.0312
Mediana	0.0667
Média	0.0828
3º Quartil	0.12
Máximo	0.4417
Amplitude	0.4392
Desvio P.	0.0652
Coef. Var.	78.6528
Assimetria	1.0636
Curtose	0.9609

50% dos grupos avaliados já cumpriram até 6,67% da idade planejada.

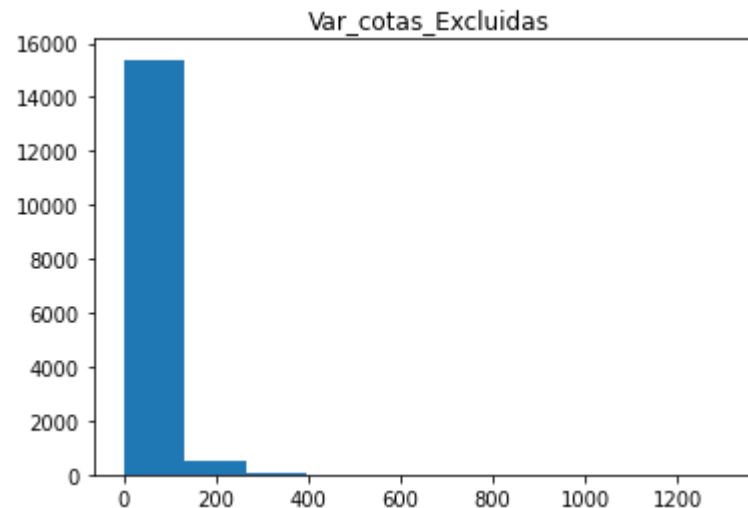
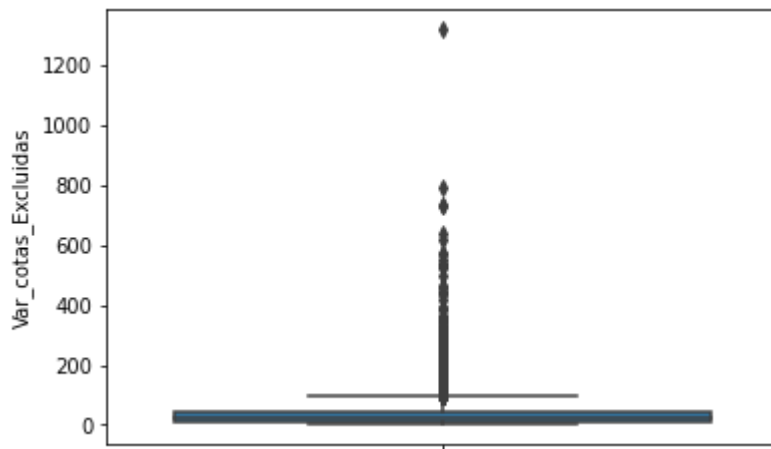
O grupo mais jovem da base apresenta 0,25% de cumprimento do plano estipulado para o grupo e o mais velho apresenta 44,17%. Essa variável apresenta uma distribuição assimétrica à direita, com uma dispersão grande dos valores em relação à média.



Detalhes das análises

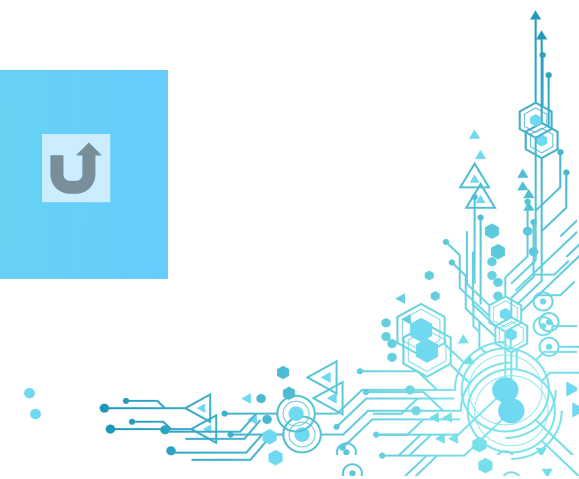
Variáveis numéricas | UNIVARIADA

36



Medida	Valor
Contagem	15971
Mínimo	0
1º Quartil	9
Mediana	20
Média	33.7518
3º Quartil	43
Máximo	1322
Amplitude	1322
Desvio P.	45.8521
Coef. Var.	135.8509
Assimetria	5.3771
Curtose	70.1697

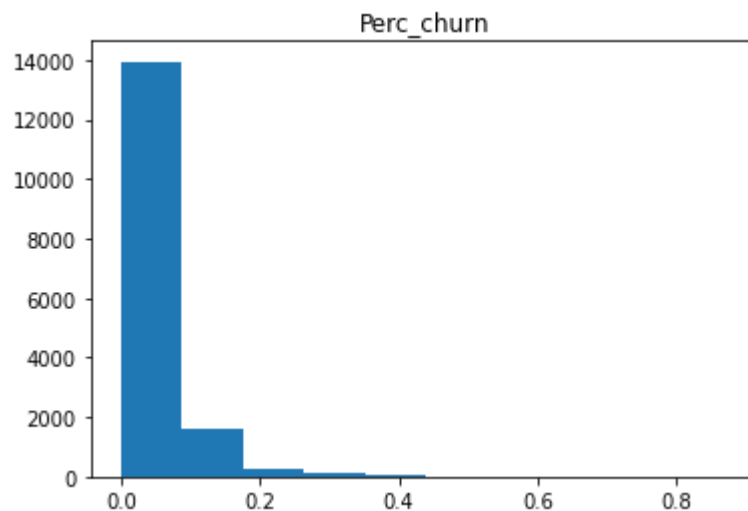
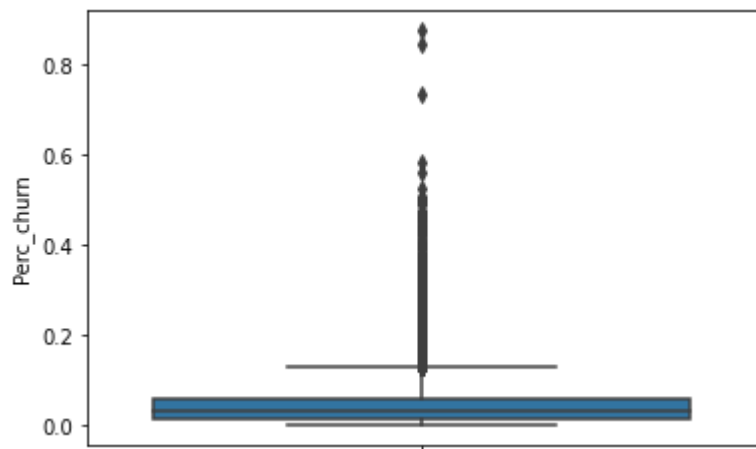
Essa variável apresenta uma distribuição assimétrica à direita, com uma dispersão grande dos valores em relação à média e com muitos outliers superiores. Enquanto alguns grupos apresentaram em alguns períodos variação nula de cotas excluídas, outros chegaram a patamares de mais de 1300 cotas. A média da variação de cotas excluídas da base é de 33 cotas!



Detalhes das análises

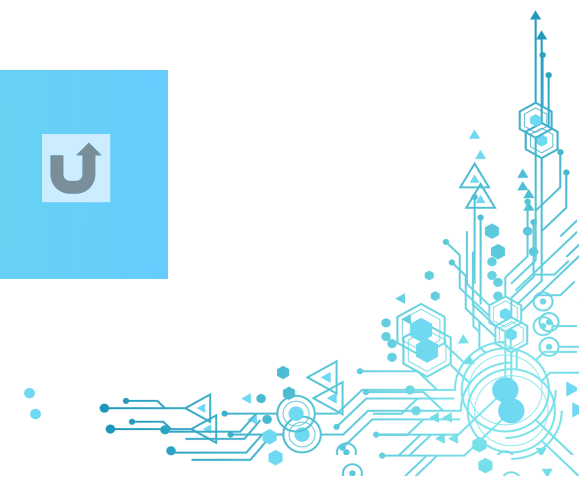
Variáveis numéricas | UNIVARIADA

37



Medida	Valor
Contagem	15971
Mínimo	0
1º Quartil	0.0158
Mediana	0.0343
Média	0.0471
3º Quartil	0.061
Máximo	0.8778
Amplitude	0.8778
Desvio P.	0.0515
Coef. Var.	109.4352
Assimetria	3.5614
Curtose	24.2632

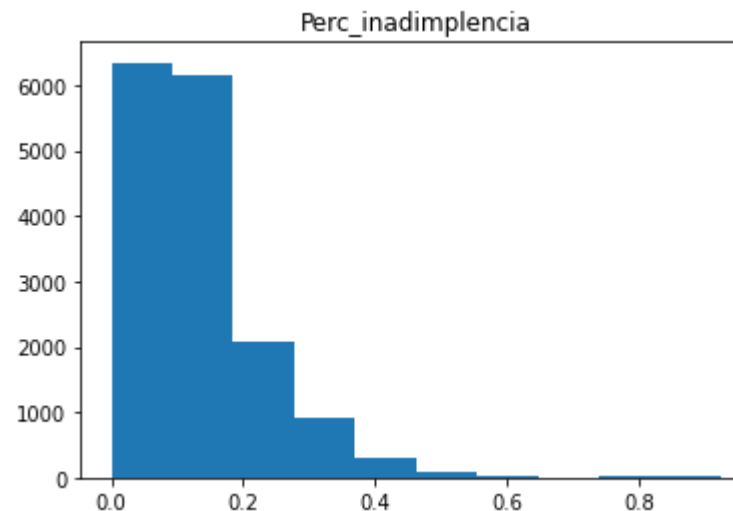
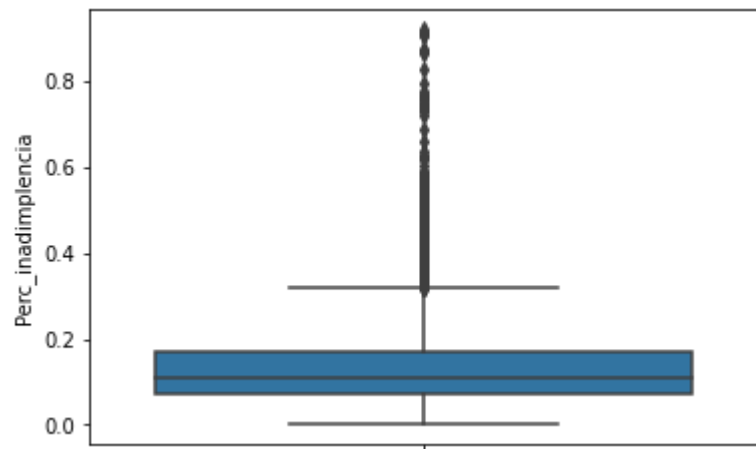
Essa variável apresenta uma distribuição assimétrica à direita, com uma dispersão grande dos valores em relação à média. 50% dos percentuais de churn estão até 3,43%. Somente 25% dos valores estão acima de 6,1%, sendo o valor máximo de 87,78%



Detalhes das análises

Variáveis numéricas | UNIVARIADA

38

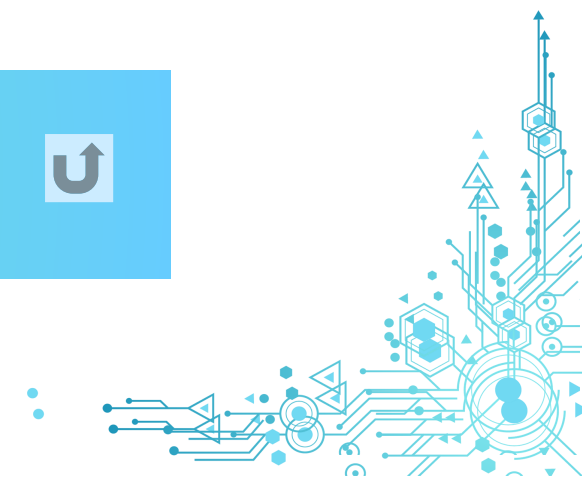


Medida	Valor
Contagem	15971
Mínimo	0
1º Quartil	0.0711
Mediana	0.1081
Média	0.132
3º Quartil	0.1707
Máximo	0.9231
Amplitude	0.9231
Desvio P.	0.0981
Coef. Var.	74.2823
Assimetria	1.7273
Curtose	5.5455

50% da base apresenta uma carteira inadimplente de até 10,81%.

25% apresenta uma carteira inadimplente superior a 17%, sendo o valor máximo 92,31%.

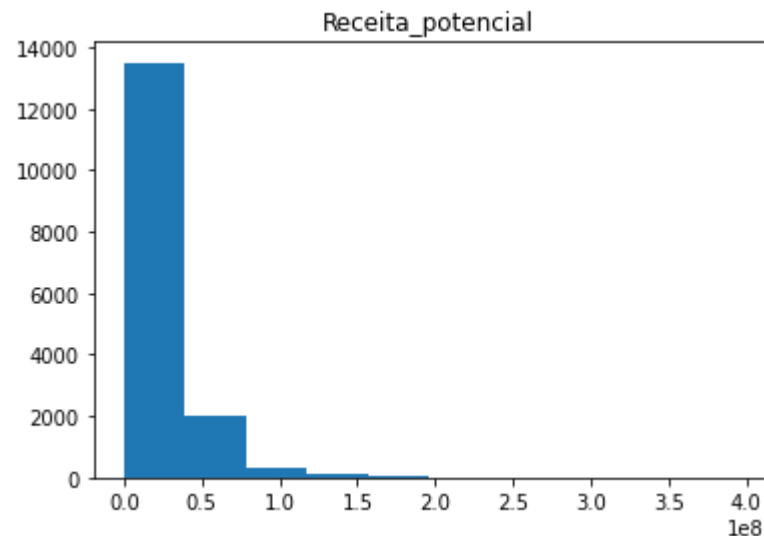
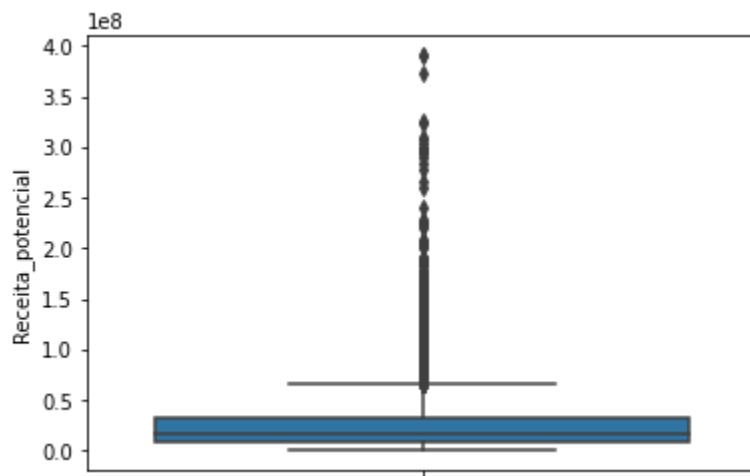
Essa variável apresenta uma distribuição assimétrica à direita, com uma dispersão grande dos valores em relação à média.



Detalhes das análises

Variáveis numéricas | UNIVARIADA

39

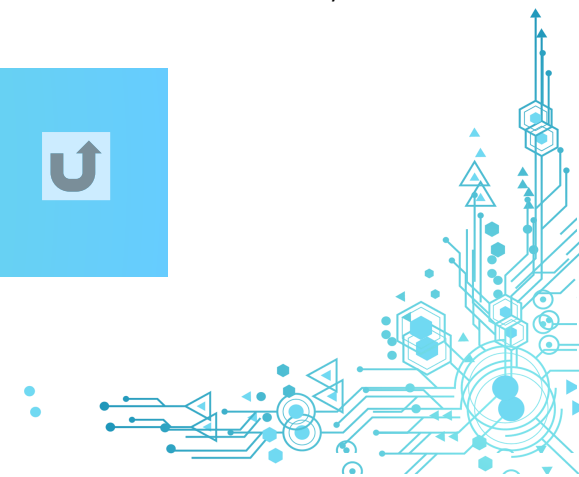


Medida	Valor
Contagem	15971
Mínimo	96.482,5786
1º Quartil	9.393.350,34
Mediana	16.695.438,91
Média	24.066.870,93
3º Quartil	32.324.424,79
Máximo	391.475.726,7
Amplitude	391.379.244,2
Desvio P.	23.801.844,61
Coef. Var.	98,8988
Assimetria	4,0885
Curtose	33,3299

50% dos valores de receita potencial estão até R\$ 16,69 milhões.

Somente 25% ds valores estão acima de R\$ 32,3 milhões, sendo o maior valor de R\$ 391,4 milhões.

Essa variável apresenta uma distribuição assimétrica à direita, com uma dispersão grande dos valores em relação à média.



Referente as 20 variáveis numéricas, destacamos na composição da base de análise:

Variáveis Numéricas




- Prazo_do_grupo_em_meses
- INCC_DI
- Numero_da_assembleia_geral_ordinária
- Valor_medio_do_bem
- Taxa_de_administracao
- Quantidade_de_cotas_ativas_em_dia
- Quantidade_de_cotas_ativas_contempladas_inadimplentes
- Quantidade_de_cotas_ativas_nao_contempladas_inadimplentes
- Quantidade_de_cotas_ativas_contempladas_no_mes
- Quantidade_de_cotas_excluidas
- Quantidade_de_cotas_ativas_quitadas
- Quantidade_de_cotas_ativas_com_credito_pendente_de_utilizacao
- Carteira_ativa
- Carteira_inadimplente
- Parcela
- Idade
- Var_cotas_Excluidas
- Perc_churn
- Perc_inadimplencia
- Receita_potencial



Análise detalhada

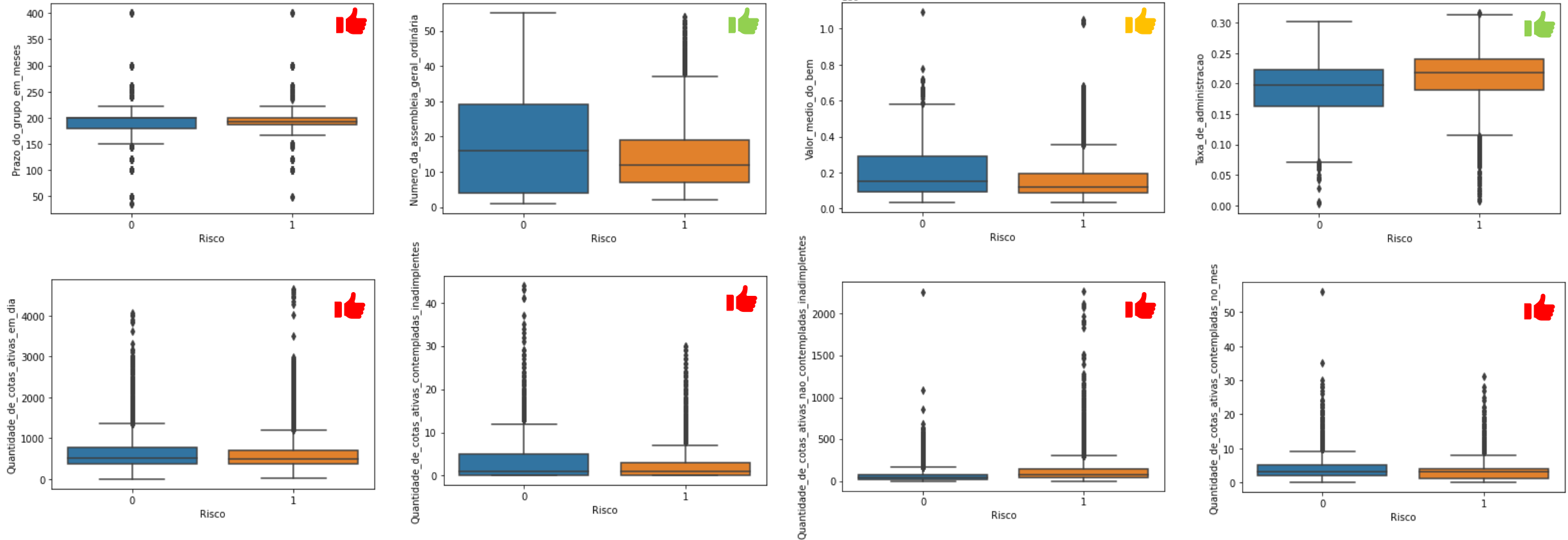
ANÁLISE EXPLORATÓRIA | BIVARIADA

LEGENDA

-  Covariável parece explicar bem a resposta
-  Covariável parece explicar pouco a resposta
-  Covariável parece não explicar a resposta



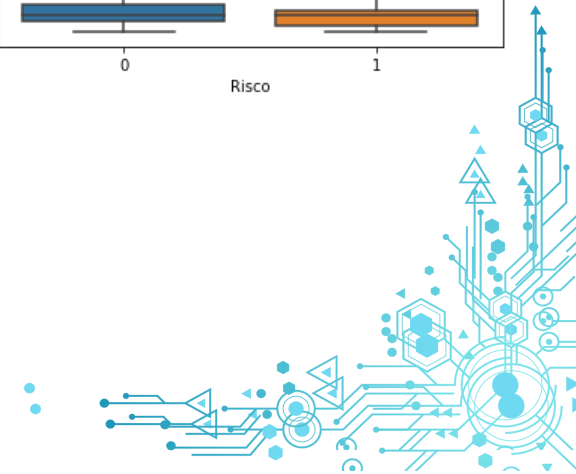
41



Grupos com taxas maiores de administração tendem a serem mais arriscados.
Mas grupos com mais número de assembleias não necessariamente tendem a ter risco maior.






Grupos com valores maiores do bem não necessariamente tendem a ter um risco maior.



Análise detalhada

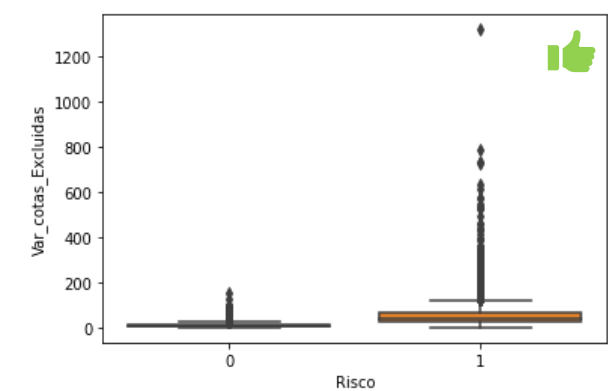
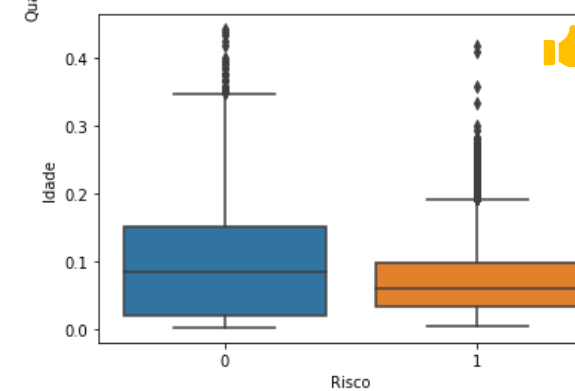
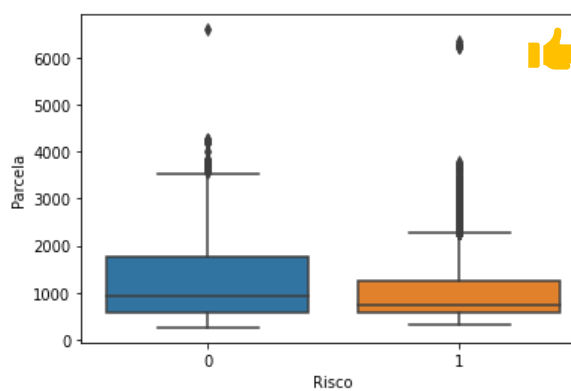
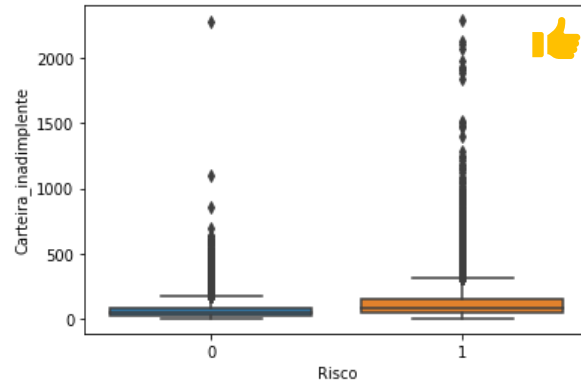
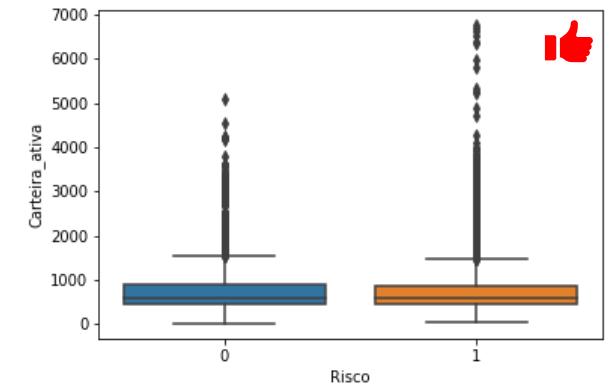
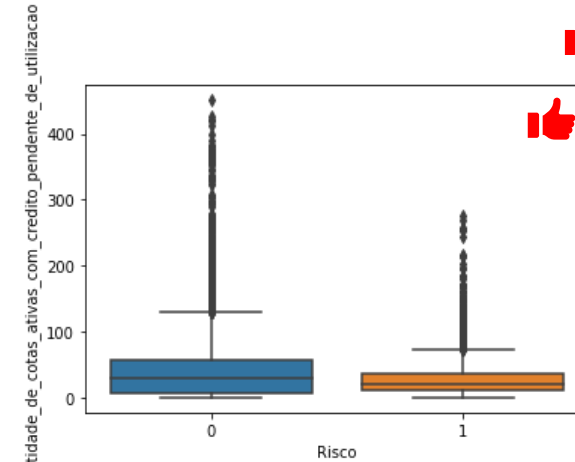
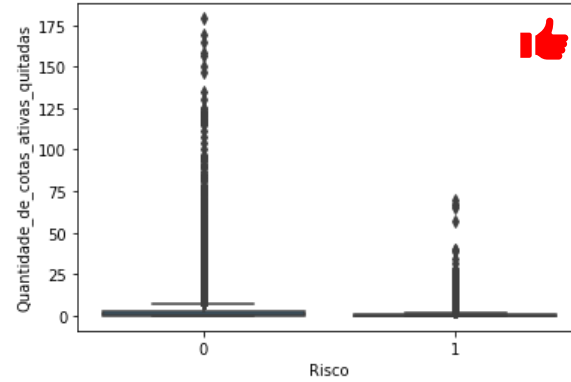
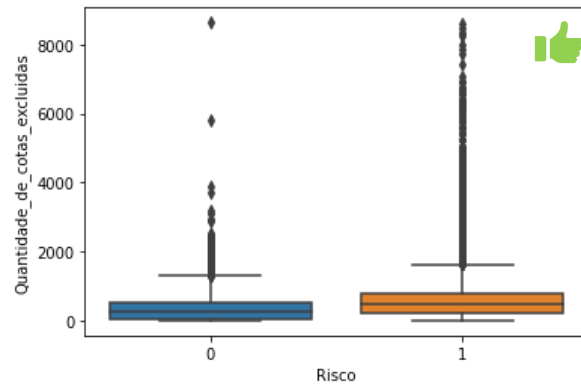
ANÁLISE EXPLORATÓRIA | BIVARIADA



LEGENDA

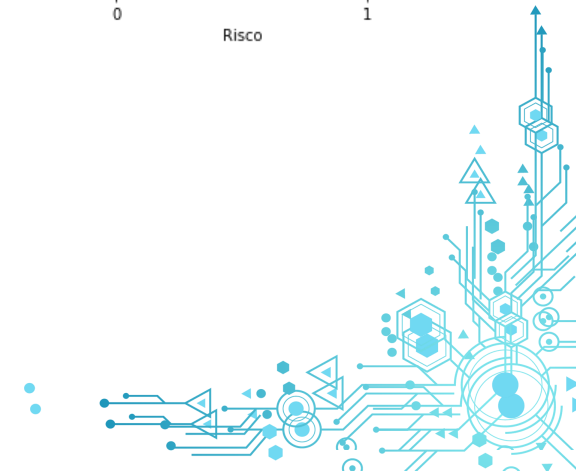
-  Covariável parece explicar bem a resposta
-  Covariável parece explicar pouco a resposta
-  Covariável parece não explicar a resposta



42






-  Grupos com mais quantidade de cotas excluídas e maiores variações de cotas excluídas tendem a serem mais arriscados.
-  Grupos com mais tempo decorrido (idade) e valores de parcelas maiores não necessariamente tendem a apresentar risco alto. Ao passo que grupos com carteira inadimplente maior tendem a apresentar risco maior.



Análise detalhada

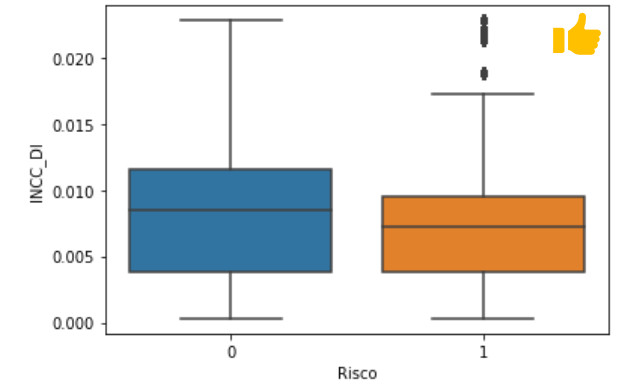
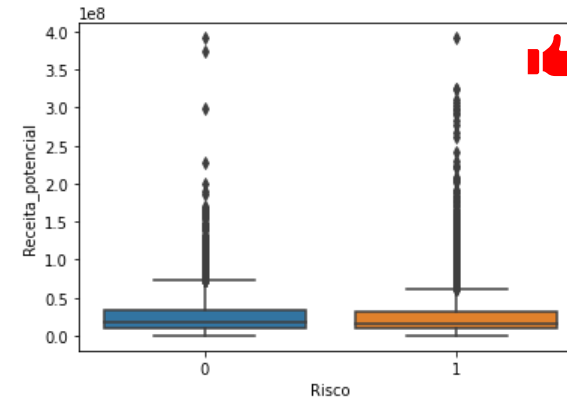
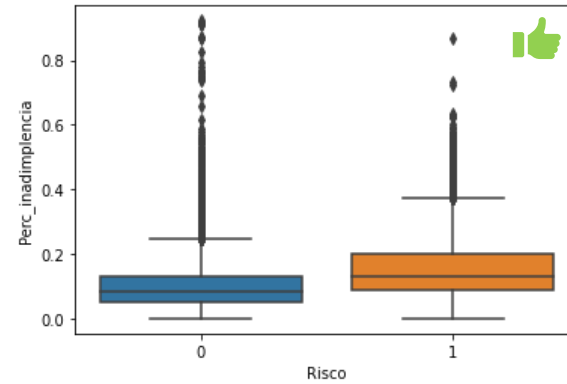
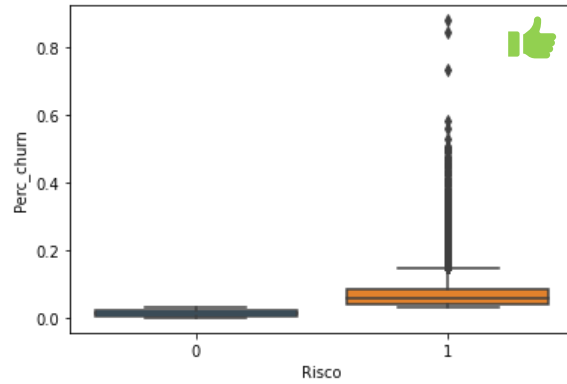
ANÁLISE EXPLORATÓRIA | BIVARIADA


LEGENDA

-  Covariável parece explicar bem a resposta
-  Covariável parece explicar pouco a resposta
-  Covariável parece não explicar a resposta



43



 Grupos com percentuais de churn e de inadimplência mais elevados tendem a ter risco alto.

 Grupos com maiores índices nacionais de custo de construção não necessariamente tendem a ser mais arriscados.



Análise das covariáveis x Risco

ANÁLISE EXPLORATÓRIA | BIDIMENSIONAL

44



Principais variáveis que parecem explicar o evento de Risco Alto:

Variável	Intepretação em relação ao Risco
Índice de correção	6, 99, 3
Faixas de Parcelas	Valores mais baixos
Faixas de prazos	Prazos mais longos
Faixas de Idade	Grupos mais novos
Quantidade de cotas excluídas	Maior quantidade de cancelamentos
Taxa de Administração	Taxas maiores
Número de assembleia geral ordinária	Menor número de assembléias
Var_cotas_excluidas	Variações grandes de cotas excluídas
Perc_Inadimplencia	Percentuais maiores de inadimplência
Perc_Churn	Percentuais maiores de churn

Detalhes das análises



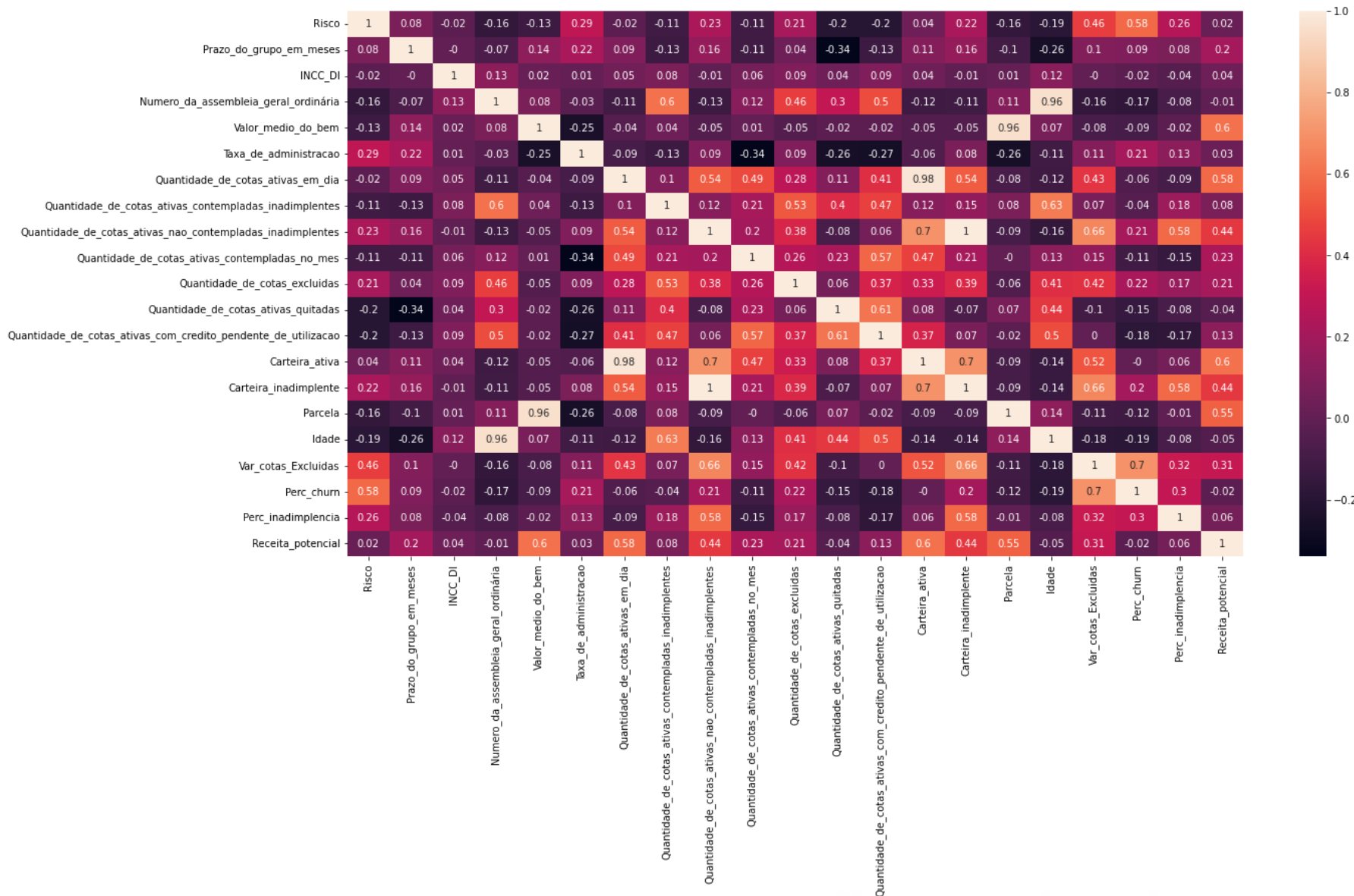
Todas as informações foram validadas pela área de negócio e fazem sentido para o comportamento de risco alto.



Detalhes das análises

Correlação variáveis numéricas | variável resposta

45



As variáveis **Var_cotas_excluidas** e **perc_churn** apresentaram correlação alta com a variável resposta, indicando vazamento de informação, muito provavelmente por conta que apoiaram a construção da variável resposta.

Apesar de saber que variáveis com alta correlação podem impactar os modelos, inicialmente não retiraremos nenhuma variável, para obter como cada modelo se comporta com as variáveis da base, e posteriormente, vamos retirar uma a uma, para observar os impactos



Metodologia de Análise de dados



Desempenho dos modelos

MODELAGEM COM ESTATÍSTICA TRADICIONAL | COMPARAÇÃO ENTRE TÉCNICAS

Definição do problema

Análise Exploratória

Modelagem Estatística

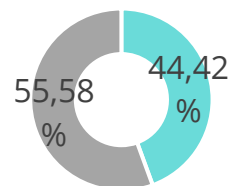
Algoritmos de Inteligência Artificial

Conclusões

47

Tratamento das bases de dados para modelagem

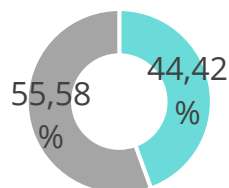
1. Balanceamento da resposta: amostra aleatória
2. 80% aleatório para treino e 20% para teste



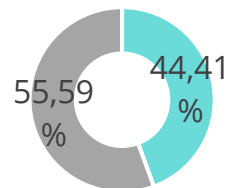
■ 0 (baixo)
■ 1 (alto)

Base para Análise
15.971 observações

Treino
12.776 – 80%



Teste
3.195 – 20%



3. Regressão Logística

4. Árvore de decisão

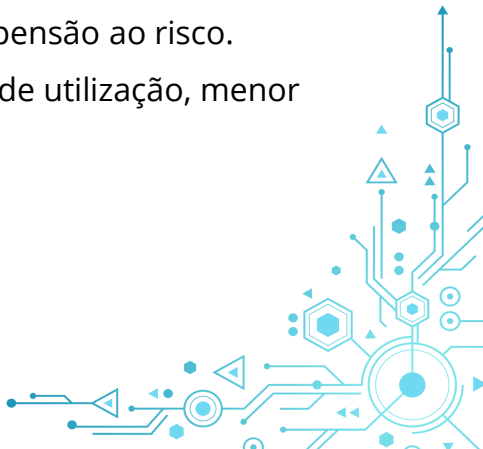
- Ambas as técnicas apresentaram **bom acerto** preditivo, com acerto geral de 68%.
- A **Regressão logística** apresentou uma quantidade maior de acertos de todos os casos que de fato eram alto risco (71%), frente a 66% da **Árvore de Decisão**.
- A **Regressão logística** também apresentou **mais estabilidade** quando comparamos os desempenhos das bases de treino e teste.
- A AUC da Regressão logística de 0.77 significa que a classe positiva tem uma chance de 0.77 de receber um score maior do que da classe negativa



O modelo seleciona as variáveis mais relevantes e estima um peso para cada uma de suas categorias, atribuindo para cada grupo de imóveis de consórcio a probabilidade de ser de alto risco.

$$p = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4}}$$

Variável	Coeficiente (β)	Interpretação em relação ao Risco
Intercepto	-0.61	Quanto maior o intercepto, menor a propensão ao risco
Prazo do grupo em meses	-1.87	Quanto maior o prazo, menor a propensão ao risco.
Valor médio do bem	-2.19	Quanto maior o prazo, menor a propensão ao risco.
Taxa de Administração	2.94	Quanto maior a taxa, maior o risco.
Quantidade de cotas ativas em dia	-4.91	Quanto maior a quantidade de cotas ativas em dia, menor a propensão ao risco.
Quantidade de cotas ativas contempladas inadimplentes	-0.95	Quanto maior a quantidade de cotas contempladas inadimplentes, menor a propensão ao risco.
Quantidade de cotas ativas não contempladas inadimplentes	17.23	Quanto maior a quantidade de cotas não contempladas inadimplentes, maior a propensão ao risco.
Quantidade de cotas ativas contempladas no mês	2.16	Quanto maior a quantidade de cotas contempladas, maior a propensão ao risco.
Quantidade de cotas ativas quitadas	-18.48	Quanto maior a quantidade de cotas quitadas, menor a propensão ao risco.
Quantidade de cotas ativas com crédito pendente de utilização	-2.16	Quanto maior a quantidade de cotas com crédito pendente de utilização, menor a propensão ao risco.
Percentual de inadimplência	1.13	Quanto maior a inadimplência, maior o risco.
Receita potencial	2.82	Quanto maior a receita, maior o risco.

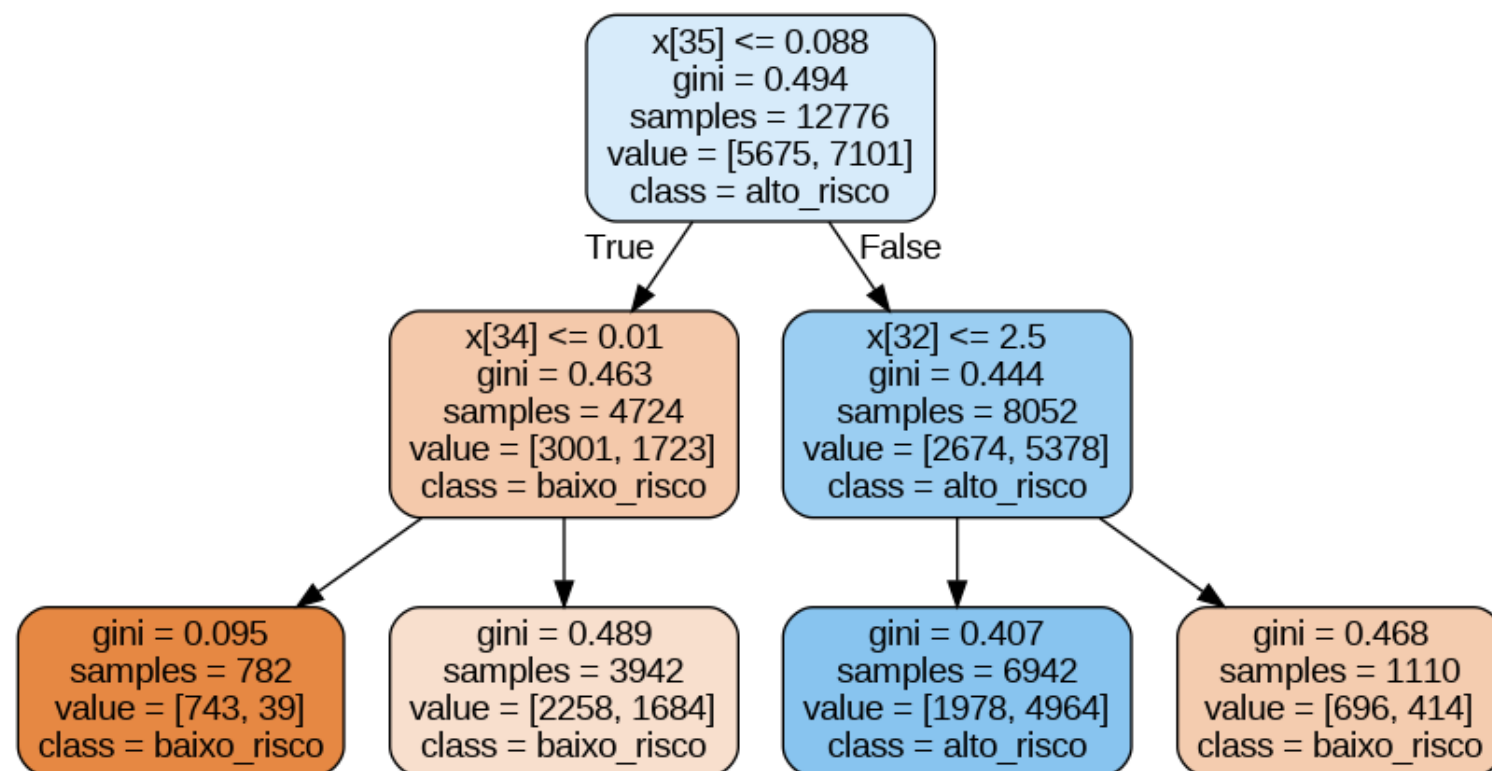


Árvore de decisão

MODELAGEM COM ESTATÍSTICA TRADICIONAL | INTERPRETAÇÃO DAS VARIÁVEIS

49

Classifica as observações pela combinação de características, por meio de uma árvore de classificação, que explique o evento de Risco alto.



- **Intepretação:**

a variável mais importante é a variável idade, seguida pela quantidade de cotas ativas com crédito pendente de utilização e pela quantidade de cotas ativas contempladas no mês.

- **Grupo de menor risco:**

Com idade menor ou igual a 8,8% do prazo planejado do grupo e quantidade de cotas ativas com crédito pendente de utilização menor ou igual a 0.01.

- **Grupo de maior risco:**

Com idade menor ou igual a 8,8% do prazo planejado do grupo e quantidade de cotas ativas contempladas no mês menor ou igual do que 2.5.

Regressão Logística

O desempenho do modelo, tanto nas bases de treino como de teste, apresentou ótimo acerto preditivo, próximo a 72% no percentual geral de classificação correta.

O percentual de acerto do evento de alto risco ficou em torno de 72% e para o evento de baixo risco 71,5%.

As bases de treino e validação apresentaram resultados similares, com uma pequena variação no acerto na base de teste, dentro do esperado para este tipo de análise.

Árvore de decisão

O desempenho do modelo, tanto nas bases de treino como de teste, apresentou ótimo acerto preditivo, próximo a 67% no percentual geral de classificação correta.

O percentual de acerto do evento de risco alto ficou em torno de 71% e 63% para o evento de risco baixo.

Nota-se que há uma **queda muito expressiva nos indicadores da base de teste**, que deve ser investigada no sentido de tentar utilizar outras técnicas que contornem um possível problema de overfitting no treino.

Regressão Logística		Indicador	Árvore de decisão	
Base de Treino	Base de Teste		Base de Treino	Base de Teste
0.7193	0.7142	Percentual de classificação Correta	0.6779	0.6613
0.8090	0.8001	Sensibilidade	0.6991	0.6881
0.7204	0.7180	Especificidade	0.7151	0.6983
0.7837	0.7681	AUC Roc	0.6961	0.6781



Metodologia de Análise de dados



Planejamento

- Objetivos
- Conceitos
- Critérios
- Histórico de dados
- Variáveis

Análise Preliminar

- Medidas resumo
- Análise univariada e bivariada
- Correlações
- Redução de variáveis

Avaliação das técnicas

- Regressão Logística
- Árvore de Decisão

Avaliação das técnicas

- Random Forest
- SVM
- Gradient Boosting Tree
- XGBoosting
- LightGBM
- CatBoost

Definição da técnica

- Validação dos resultados com negócios
- Escolha da técnica que melhor se adequa ao uso e estratégias da área de negócio

Catboost é uma biblioteca desenvolvida em 2017 pela Yandex, que fornece uma outra implementação do algoritmo *Gradient Boosting*, um método *Ensemble* que combina vários preditores fracos em um forte (preditor fraco significa que sua classificação é apenas um pouco melhor do que adivinhações aleatórias).

A ideia geral é treinar sequencialmente os preditores, cada um tentando corrigir seu antecessor.

Esse algoritmo possui dois grandes objetivos: evitar o *overfitting* e de forma padrão oferecer bons hiperparâmetros.

Modelos	Acuracidade	Precisão	Recall	f1	Roc_auc
CatBoost	0.8553	0.8624	0.8803	0.8712	0.9362
LightGBM	0.8545	0.8633	0.8773	0.8701	0.9360
XGBoosting	0.8510	0.8601	0.8741	0.8670	0.9334
Random Forest	0.8502	0.8589	0.8742	0.8664	0.9297
Gradient Boosting	0.8461	0.8503	0.8778	0.8638	0.9276
SVM	0.8142	0.8307	0.8364	0.8334	0.8970
Regressão Logística	0.7674	0.7722	0.8248	0.7976	0.8458
Árvore de Decisão	0.7981	0.8217	0.8134	0.8175	0.7962

O melhor modelo foi o **Catboost**, com o **maior Roc_auc** e **melhor acurácia**;

Os modelos tradicionais (Regressão Logística e Árvore de decisão) tiveram uma melhora no desempenho frente ao observado anteriormente (0,77 e 0,68), principalmente pela inclusão da técnica de **StratifiedKFold** no novo processamento.

Todos os modelos apresentaram um **Roc_auc** bom, pois um classificador com precisão de apenas 0.5 não é melhor do que adivinhar aleatoriamente.



Metodologia de Análise de dados



Planejamento

- Objetivos
- Conceitos
- Critérios
- Histórico de dados
- Variáveis

Análise Preliminar

- Medidas resumo
- Análise univariada e bivariada
- Correlações
- Redução de variáveis

Avaliação das técnicas

- Regressão Logística
- Árvore de Decisão

Avaliação das técnicas

- Randon Forest
- SVM
- Gradient Boosting Tree
- XGBoosting
- LightGBM
- CatBoost

Definição da técnica

- Validação dos resultados com negócios
- Escolha da técnica que melhor se adeque ao uso e estratégias da área de negócio

- Acurácia: **Catboost (85,5%)** apresentou melhor desempenho em comparação com o **LightGBM (85,4%)**

Shap Values: revelou as features mais importantes do Catboost.

- **Quantidade de cotas excluídas**
- **Numero_da_assembleia_geral_ordinária.**



- Ambos os modelos podem servir como base para predizer grupos de alto risco de consórcio de imóveis.

Exemplo: alimentação de um **trigger para geração de alertas as áreas de negócios.**

Squad de vendas: buscar novos consorciados para reposição em grupos específicos com alta taxa de cancelamento;

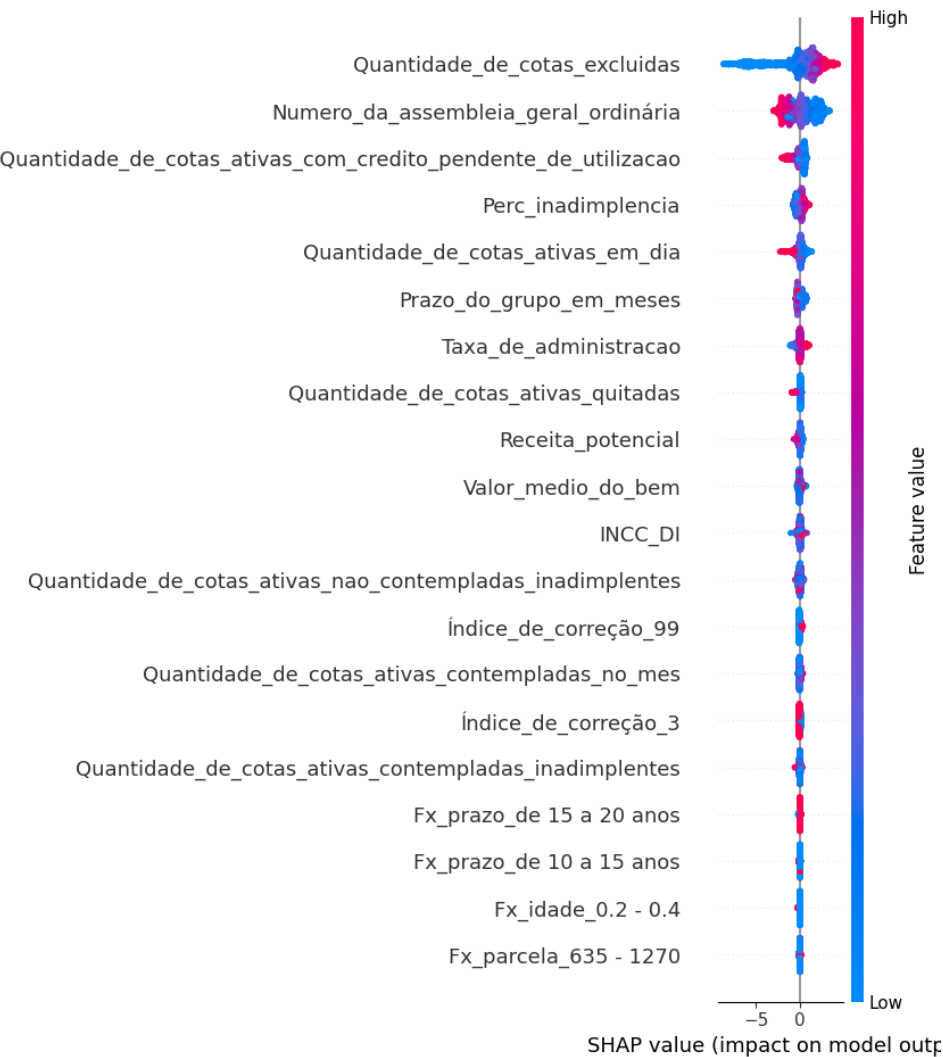
Squad de Cancelamentos: fazer ações de retenção com os clientes dos grupos de maior risco:

- propor alternativas de redução da carta de crédito,
- renegociação da dívida pelo prazo restante,
- ou até mesmo a recompra da cota pela Administradora.

Com isso buscar a melhoria contínua e garantir a sustentabilidade/perenidade do negócio.

Shap Values - Feature importance

Modelagem com inteligência artificial



Variáveis + importantes obtidas pelo modelo Catboost

X

Variáveis que pareciam explicar o risco na AED

Variável	Intepretação em relação ao Risco
Índice de correção	6, 99, 3
Faixas de Parcelas	Valores mais baixos
Faixas de prazos	Prazos mais longos
Faixas de Idade	Grupos mais novos
1 Quantidade de cotas excluídas	Maior quantidade de cancelamentos
Taxa de Administração	Taxas maiores
2 Número da assembleia geral ordinária	Menor número de assembléias
Var_cotas_excluidas	Variações grandes de cotas excluídas
4 Perc_Inadimplencia	Percentuais maiores de inadimplência
Perc_Churn	Percentuais maiores de churn

Sugestão para Trabalhos Futuros

Realizar a **experimentação de outros modelos**;

Explorar **técnicas de otimização dos hiperparâmetros** dos modelos;

Buscar uma **quantidade maior de dados retroativos** e também adicionar **novas variáveis**;



Descrição das variáveis – Variáveis Cadastrais

57

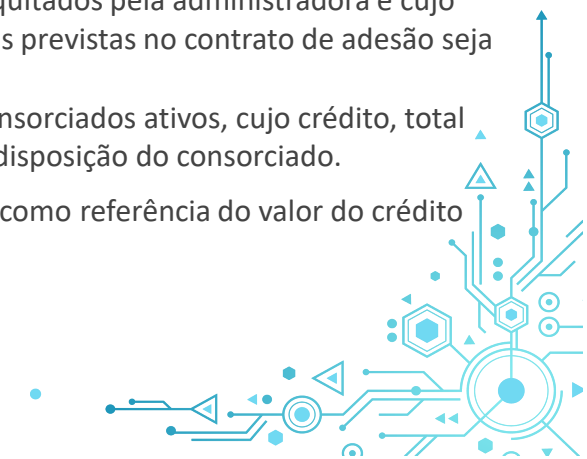
Variável	Descrição
Chave_nome_adm_e_cod_grupo	União dos campos "Nome_da_Administradora" e "Codigo_do_grupo" para facilitar análises
Nome_da_Administradora	Nome reduzido da instituição no Unacad
CNPJ_da_Administradora	Raiz do Cadastro Nacional da Pessoa Jurídica - CNPJ com 8 dígitos da Administradora de Consórcio
Codigo_do_grupo	Código do grupo. Identifica o grupo no sistema da administradora.
Prazo_do_grupo_em_meses	Prazo do grupo em meses. É prazo de duração do grupo, em meses.



Descrição das variáveis – Variáveis Transacionais

58

Variável	Descrição
Data_base	Data-base do documento, com formato AAAA-MM, onde AAAA = ano e MM = mês (01 a 12)
Numero_da_assembleia_geral_ordinária	Número de assembléias gerais ordinárias realizadas até a data-base.
Valor_medio_do_bem	Média aritmética simples dos preços dos bens adotados como referência do valor do crédito no contrato de participação dos consorciados ativos do grupo, na data da assembleia geral ordinária imediatamente anterior à data-base
Taxa_de_administracao	Percentual correspondente à média aritmética simples das taxas de administração contratuais (incluídas as antecipações) das cotas de consorciados ativos na data-base
Quantidade_de_cotas_ativas_em_dia	Total de cotas, contempladas ou não, de consorciados ativos cujo percentual em atraso é inferior ao percentual de amortização mensal, na data-base.
Quantidade_de_cotas_ativas_contempladas_inadimplentes	Total de cotas contempladas de consorciados ativos cujo percentual em atraso é igual ou superior ao percentual de amortização mensal, na data-base.
Quantidade_de_cotas_ativas_nao_contempladas_inadimplentes	Total de cotas não contempladas de consorciados ativos cujo percentual em atraso é igual ou superior ao percentual de amortização mensal, na data-base.
Quantidade_de_cotas_ativas_contempladas_no_mes	Total de cotas contempladas (por sorteio ou por lance) no mês da data-base, de consorciados ativos
Quantidade_de_cotas_excluidas	Total, na data-base, de cotas de consorciados excluídos (por desistência declarada ou por deixar de cumprir as obrigações financeiras contratuais)
Quantidade_de_cotas_ativas_quitadas	Total de cotas, contempladas ou não, de consorciados ativos considerados quitados pela administradora e cujo saldo remanescente para quitação integral das responsabilidades financeiras previstas no contrato de adesão seja inferior ao valor equivalente a uma prestação.
Quantidade_de_cotas_ativas_com_credito_pendente_de_utilizacao	Total, na data-base, de cotas contempladas (por sorteio ou por lance) de consorciados ativos, cujo crédito, total ou parcial, se encontra depositado em conta vinculada a contemplações, à disposição do consorciado.
Índice_de_correção	Índice contratado como critério para atualização do preço do bem adotado como referência do valor do crédito



Descrição das variáveis – Variáveis Criadas

59

Variável	Descrição
Carteira_ativa	Total de cotas ativas do grupo no período
Carteira_inadimplente	Total de cotas ativas inadimplentes do grupo no período
Parcela	Valor estimado da parcela do grupo
Idade	Faixa de idade em que o grupo se encontra (calculado com base no numero de assembleias ordinárias decorridas e prazo de duração do grupo)
Var_cotas_Excluidas	Variação da quantidade de cotas excluídas mês a mês de cada cada grupo, de cada administradora
Perc_churn	Faixa de churn em que o grupo se encontra (baseado na quantidade de cotas excluídas e cotas totais do grupo)
Perc_inadimplencia	Faixa de inadimplencia em que o grupo se encontra (baseado na quantidade de cotas inadimplentes e cotas ativas do grupo)
Condição	Grupos novos (iniciados em 2021 e 2022) e grupos antigos (antes de 2021)
Fx_parcela	Faixas de parcelas que os grupos apresentam
Fx_prazo	Faixas de prazos de duração que os grupos apresentam
Fx_idade	Faixas de idade do grupo baseada no número da assembleia geral ordinária pelo prazo de cada grupo
Receita_potencial	Receita potencial do grupo (carteira ativa * taxa de administração * valor medio do bem)



Descrição das variáveis – Variável Externa

60

Variável	Descrição
INCC_DI	Índice Nacional de Custo de Construção — Disponibilidade Interna. Essa modalidade tem relação com a alteração dos custos na construção civil do primeiro ao último dia do mês de referência. Referência: http://indiceeconomicos.secovi.com.br/indicadormensal.php?idindicador=59



Descrição das variáveis – Variável Resposta

61

Variável	Descrição
Risco	0 – grupo de imóveis de consórcio é de baixo risco 1 – grupo de imóveis de consórcio é de alto risco

Para **definição do alto ou baixo risco** foi considerado o benchmarking das administradoras que apresentaram o maior share de cotas ativas de imóveis, da base mais recente disponibilizada (202208).

De 50 administradoras de consórcio de imóveis em Agosto/2022, **12 representam mais de 80% do share do mercado.**

Ponderando a média percentual de variação de cotas excluídas sobre a carteira ativa de cada administradora pela sua respectiva participação no mercado, foi definido que:

grupos de alto risco (risco =1) possuem variação de cotas excluídas sobre a carteira ativa **maior do que 3%**

grupos de baixo risco (risco=0) apresentam variação de cotas excluídas sobre a carteira ativa **menor ou igual a 3%.**

