



## 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

Condição	Frequencia
antigo	0.74222
novo	0.25778

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

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.

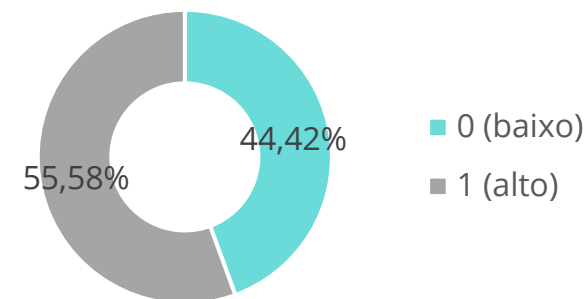
De 50 administradoras de consórcio de imóveis em Agosto/2022, **12 representam mais de 80% do share do mercado.**

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:

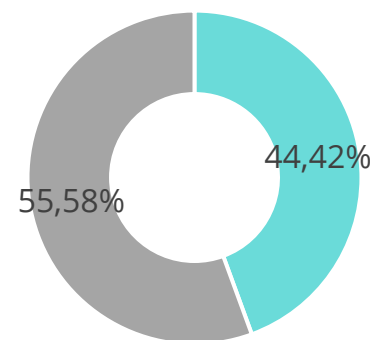
**grupos de alto risco (risco =1)** possuem variação de cotas excluídas sobre a carteira ativa **maior do que 3%**

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

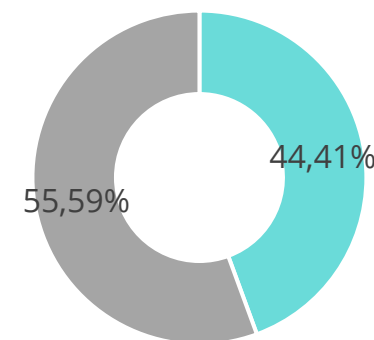
**Base para Análise**  
15.971 observações



**Treino**  
12.776 – 80%



**Teste**  
3.195 – 20%



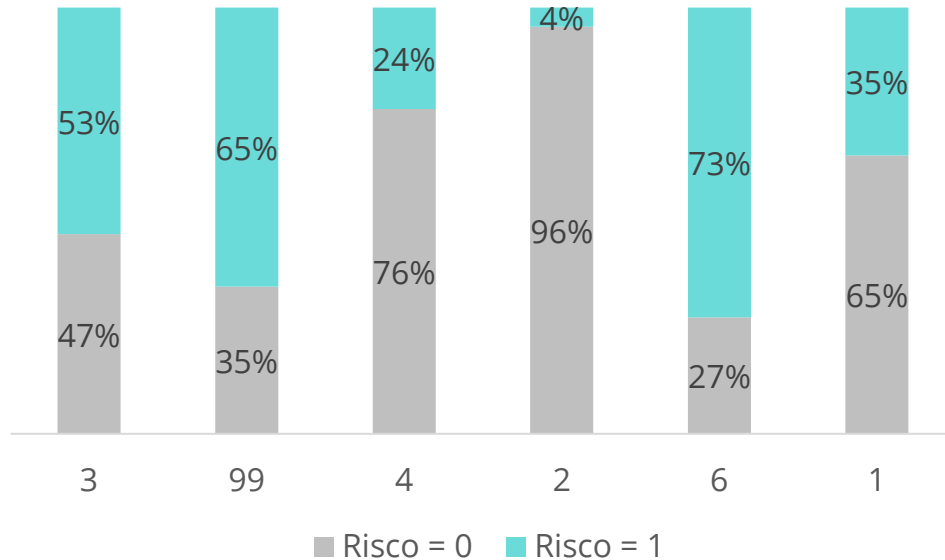
■ 0 (baixo) ■ 1 (alto)

# Análise detalhada

ANÁLISE EXPLORATÓRIA | BIVARIADA

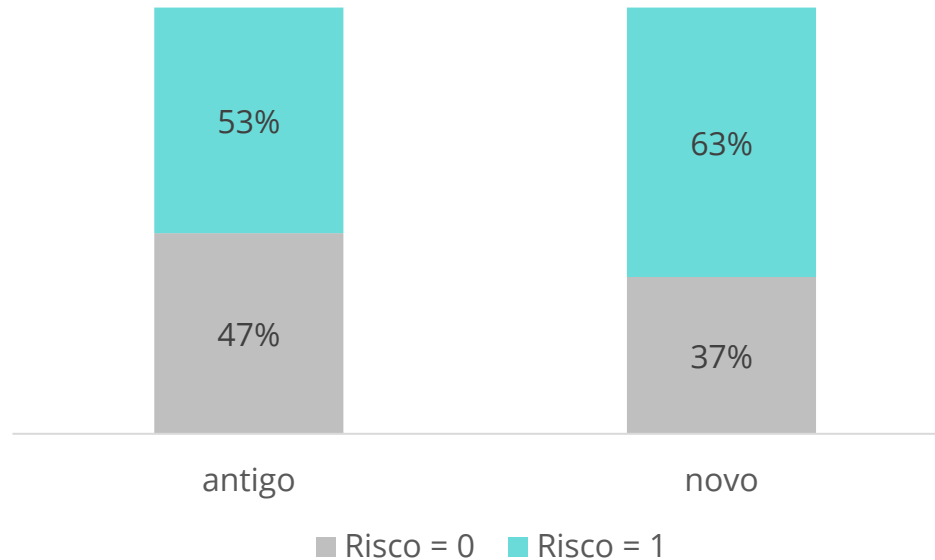
16

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

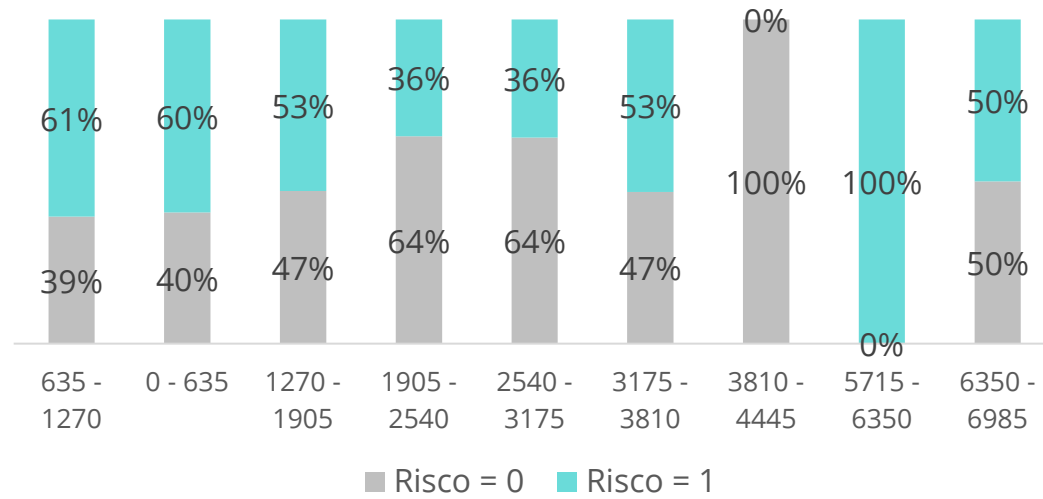


# Análise detalhada

ANÁLISE EXPLORATÓRIA | BIVARIADA

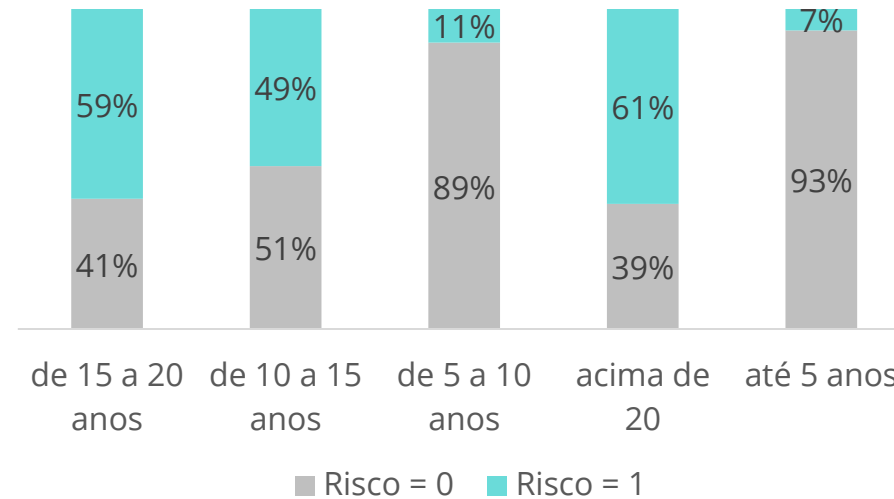
17

## 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

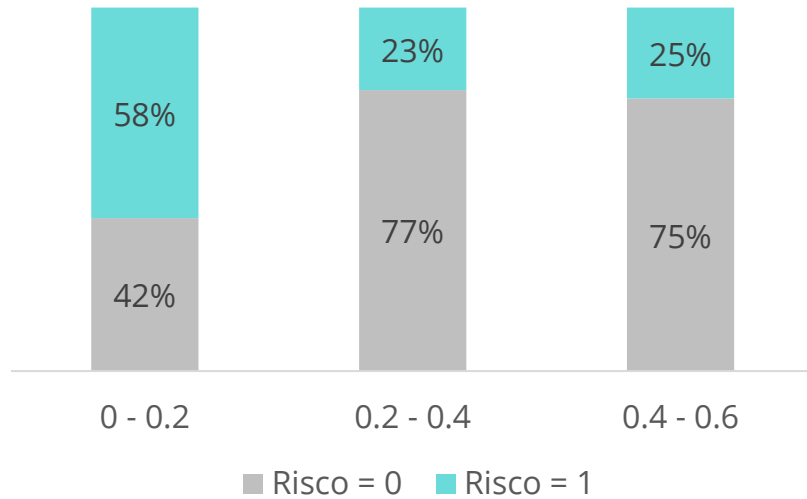


# Análise detalhada




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

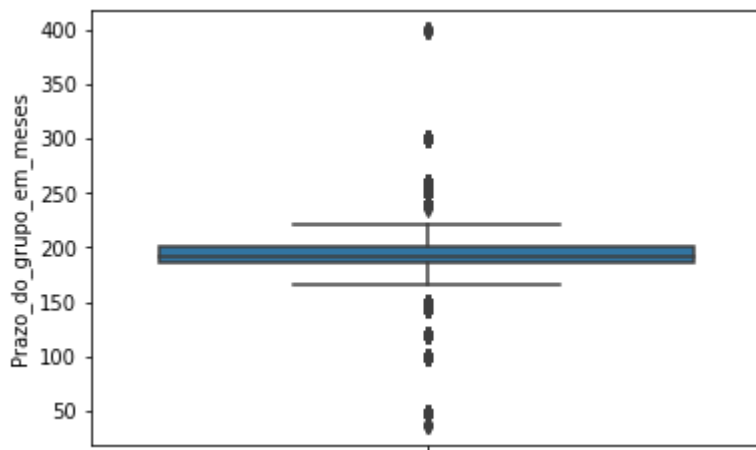
- [Prazo do grupo em meses](#)
- [INCC\\_DI](#)
- [Numero da assembleia geral ordinária](#)
- [Valor medio do bem](#)
- [Taxa de administracao](#)
- [Quantidade de cotas ativas em dia](#)
- [Quantidade de cotas ativas contempladas inadimplentes](#)
- [Quantidade de cotas ativas nao contempladas inadimplentes](#)
- [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

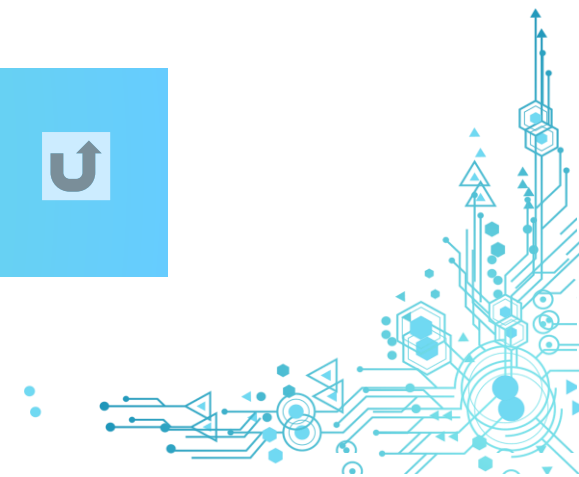
Variáveis numéricas | UNIVARIADA

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

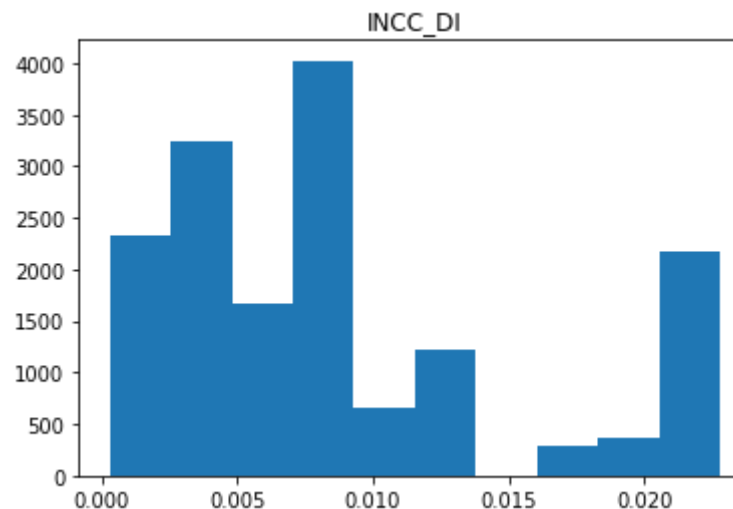
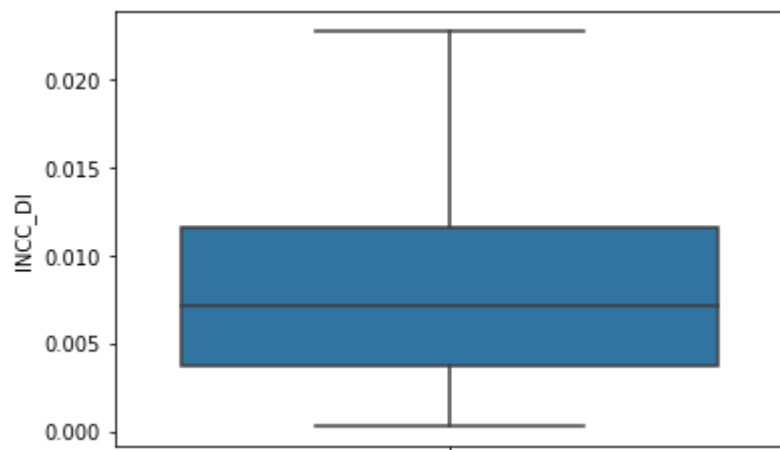
A base apresenta metade dos grupos de consórcio de imóveis com prazos até 192 meses e 75% dos grupos com prazos abaixo de 200 meses. Os grupos mais curtos possuem 36 meses de duração e os mais longos 400 meses. Verificamos uma distribuição assimétrica à direita, sem muita dispersão dos dados em relação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



# Detalhes das análises

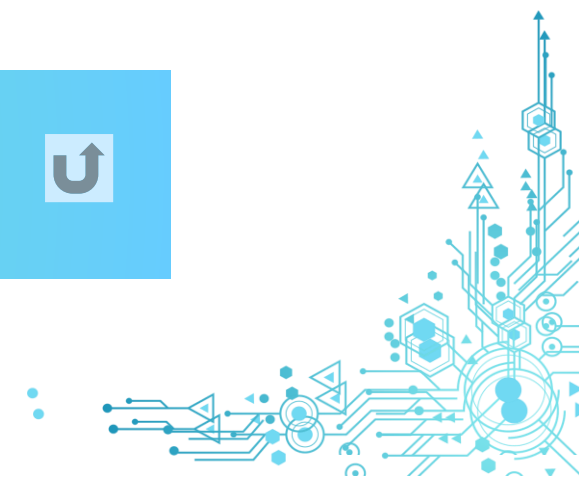
Variáveis numéricas | UNIVARIADA

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

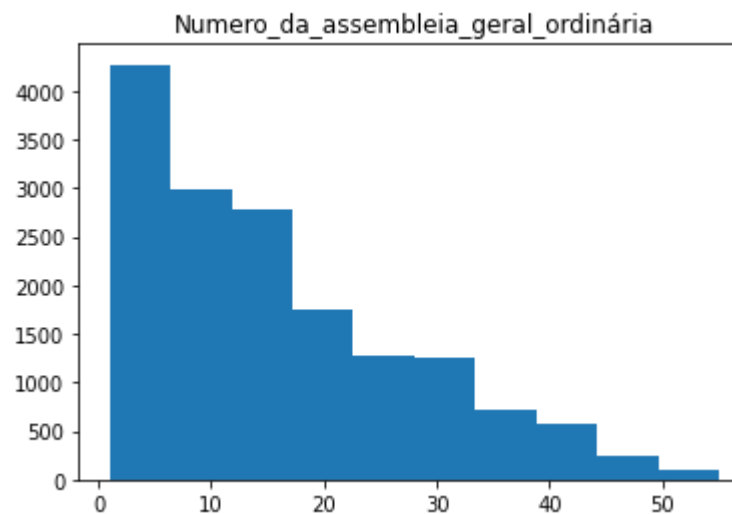
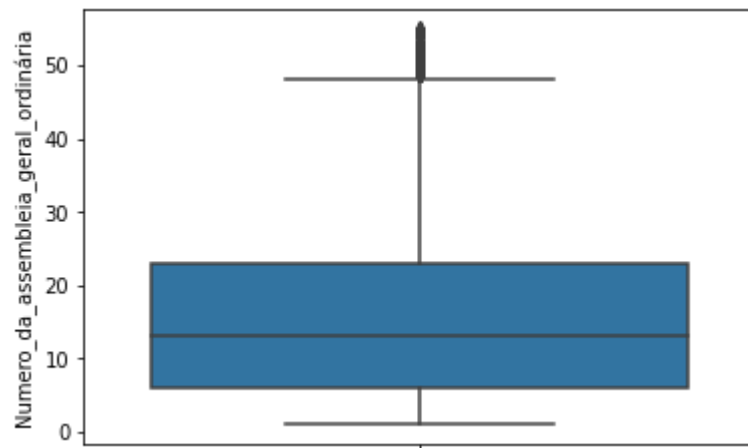
Da base analisada, 50% apresenta índice nacional de custo de construção até 0,7% e somente 25% acima de 1,16%. Essa variável apresenta uma distribuição assimétrica à direita, com uma dispersão grande dos valores em relação à média, sendo o menor valor 0,03% e o maior 2,28%



# Detalhes das análises

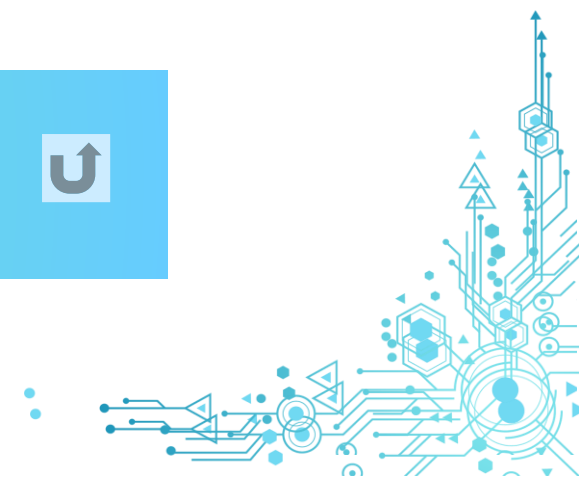
Variáveis numéricas | UNIVARIADA

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

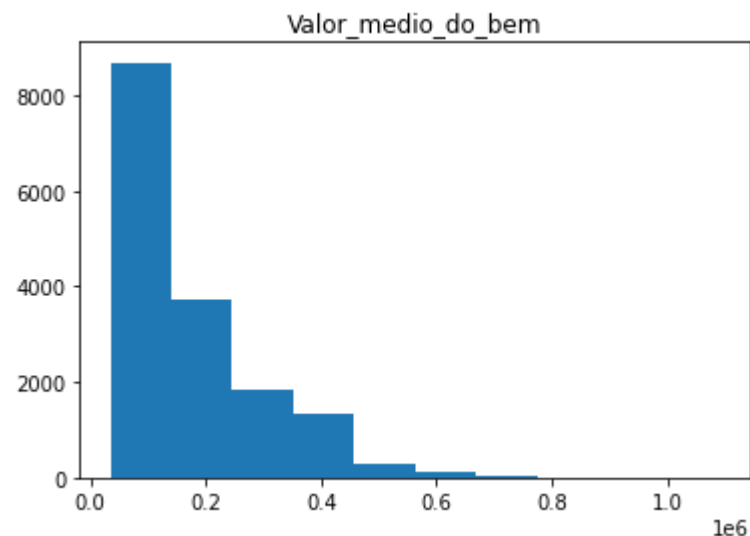
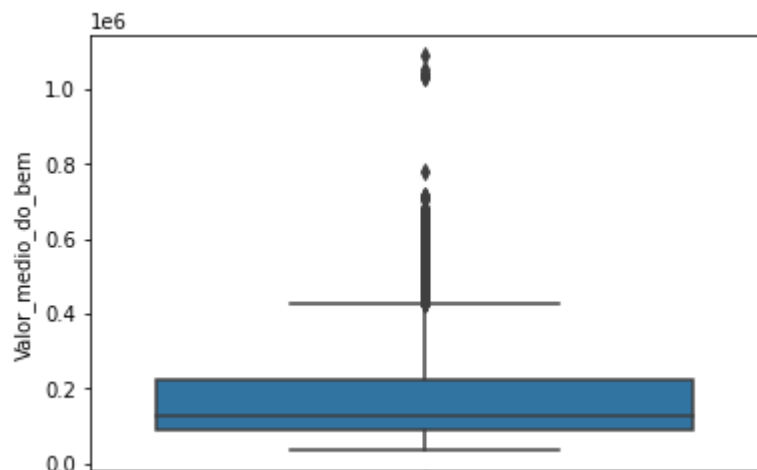
Dos grupos presentes na análise, 50% passaram pela 13ª assembleia e somente 25% já passaram da 23ª assembleia. Tem grupos que estão na 1ª assembleia e grupos mais antigos que já estão na 55ª assembleia. Essa variável apresenta uma distribuição assimétrica à direita, com uma dispersão grande dos valores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



# Detalhes das análises

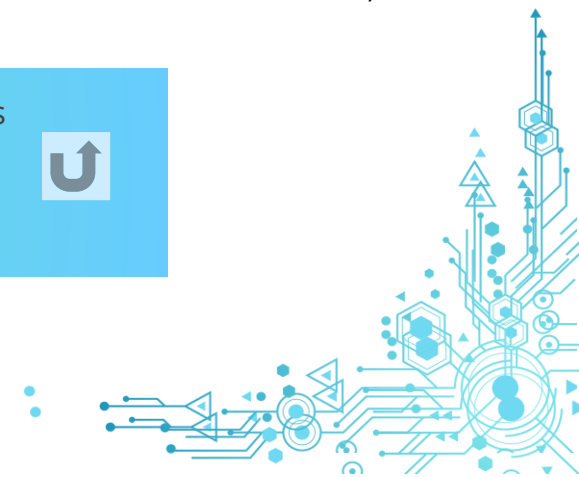
Variáveis numéricas | UNIVARIADA

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

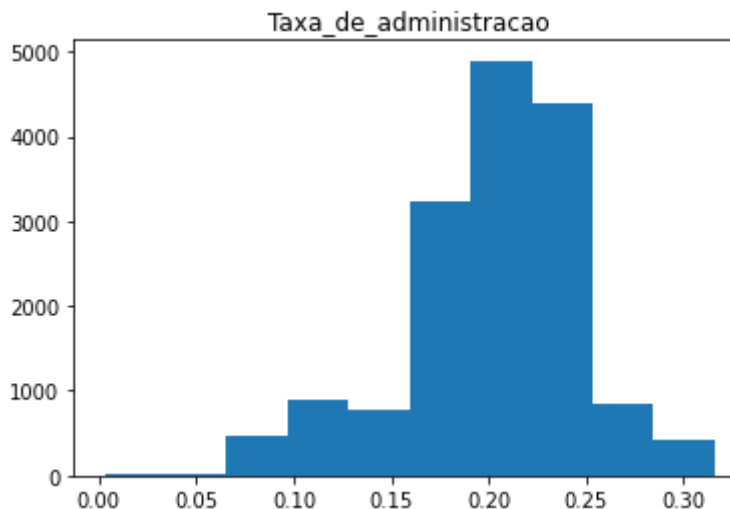
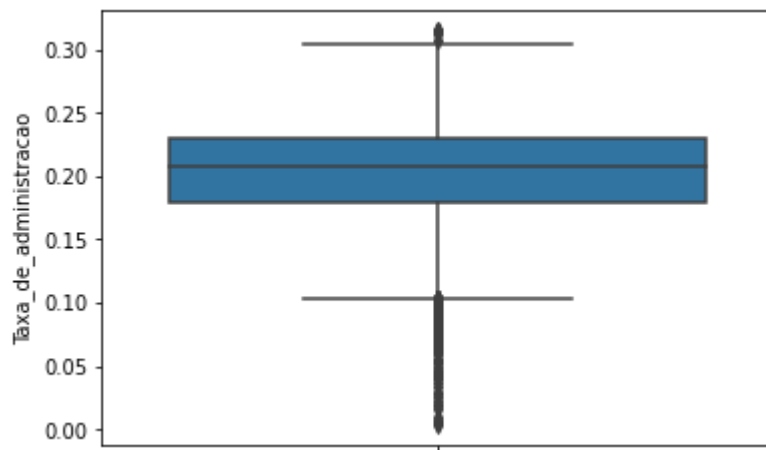
50% dos valores das cartas de crédito dos grupos avaliados estão concentrados em uma faixa até R\$126 mil e 25% dos maiores valores, em um faixa acima de R\$225 mil. A menor carta de crédito apresenta valor de R\$ 34 mil e a maior valor de R\$1,09 milhões. Essa variável apresenta uma distribuição assimétrica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



# Detalhes das análises

Variáveis numéricas | UNIVARIADA

24

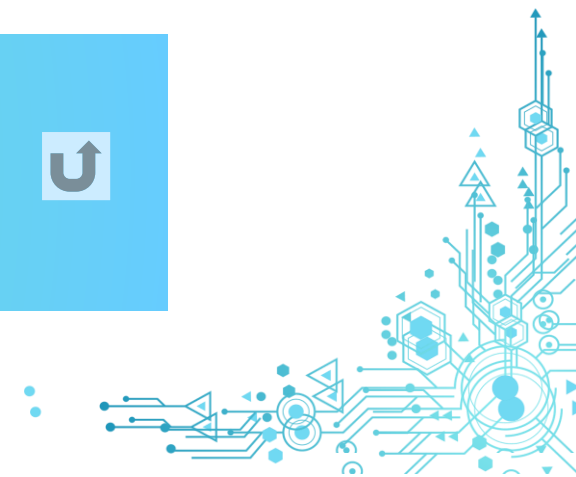


Medida	Valor
Contagem	15971
Mínimo	0.0028
1º Quartil	0.18
Mediana	0.2081
Média	0.2024
3º Quartil	0.2307
Máximo	0.3159
Amplitude	0.3131
Desvio P.	0.0461
Coef. Var.	22.7873
Assimetria	-0.7144
Curtose	0.7759

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%.

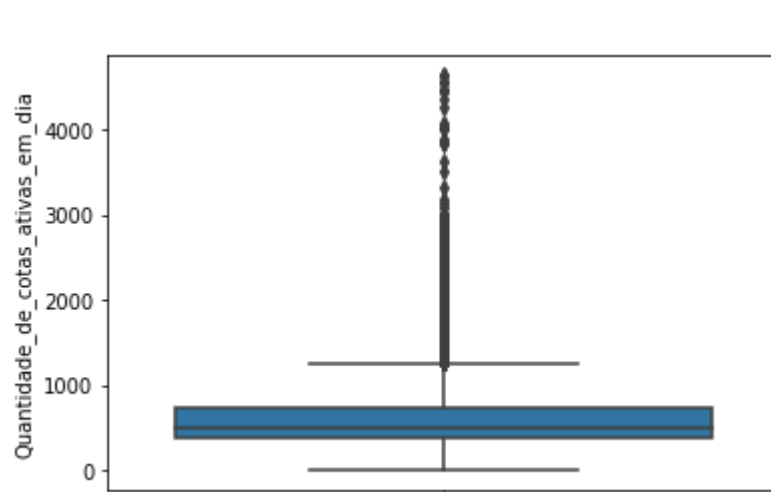
A definição da taxa de administração é bastante estratégica para esse produto, pois representa a receita bruta que cada empresa espera obter com a venda do produto. Pode variar bastante em virtude do momento que a empresa está, de seus objetivos, de campanhas, de pressão dos concorrentes, entre outros fatores.



# Detalhes das análises

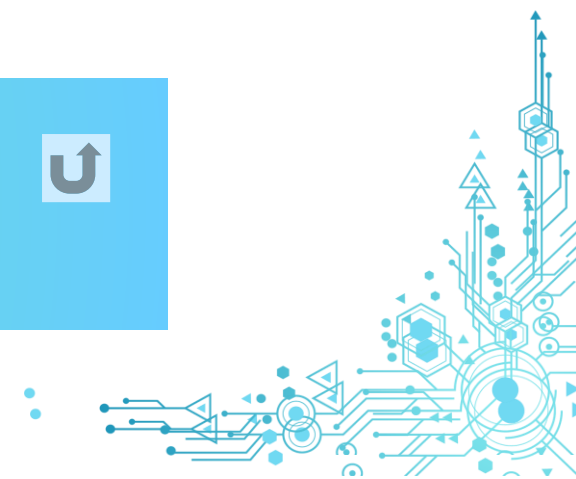
Variáveis numéricas | UNIVARIADA

25



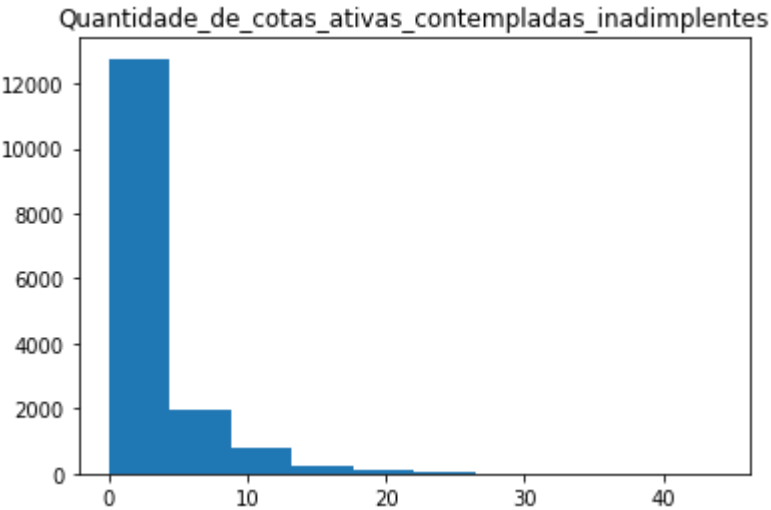
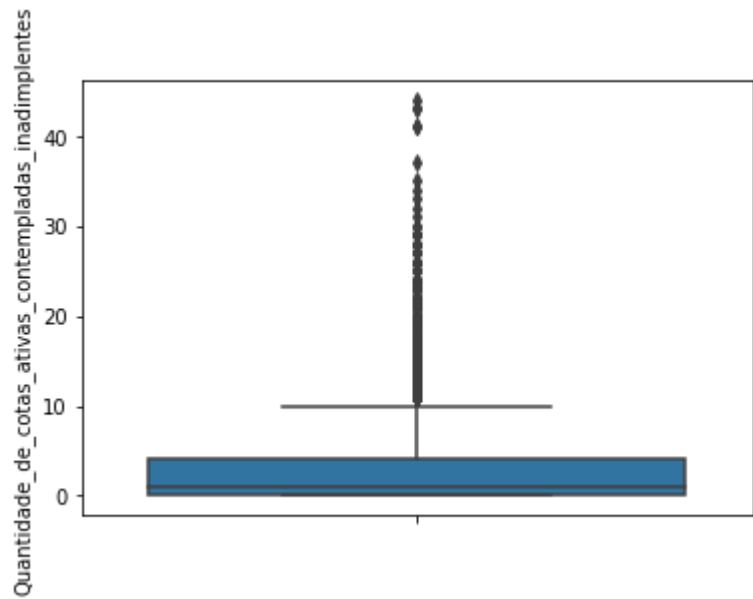
Medida	Valor
Contagem	15971
Mínimo	1
1º Quartil	380
Mediana	506
Média	621.7876
3º Quartil	735
Máximo	4.646
Amplitude	4.645
Desvio P.	433.9867
Coef. Var.	69.7966
Assimetria	2.493
Curtose	9.7315

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.



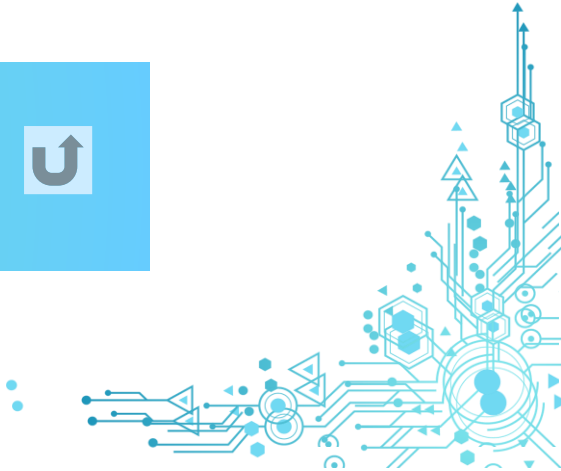
# Detalhes das análises

Variáveis numéricas | UNIVARIADA



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

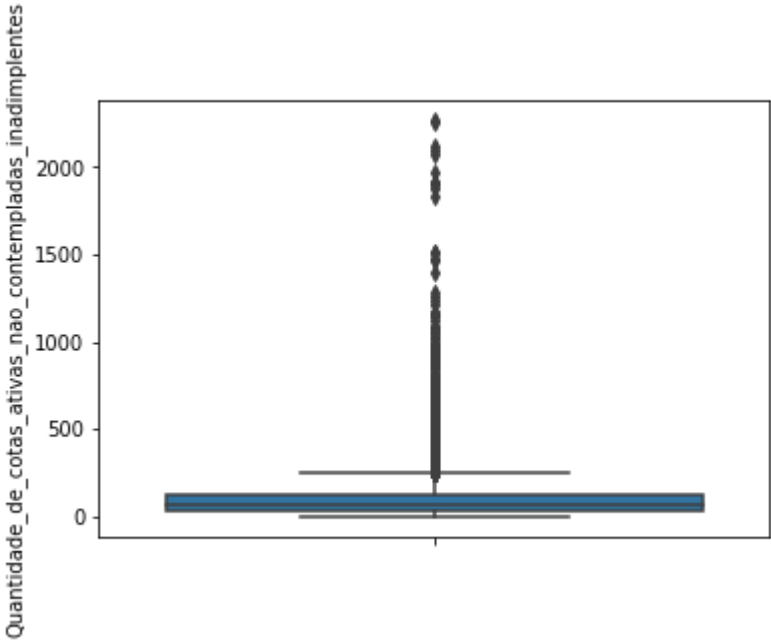
A quantidade de cotas ativas contempladas inadimplentes é baixa, sendo de até 1 cota para 50% do dados analisados. Os grupos com mais cotas contempladas inadimplentes possuem 44 cotas. Essa variável apresenta uma distribuição assimétrica à direita, com uma dispersão muito grande dos valores em relação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





# Detalhes das análises

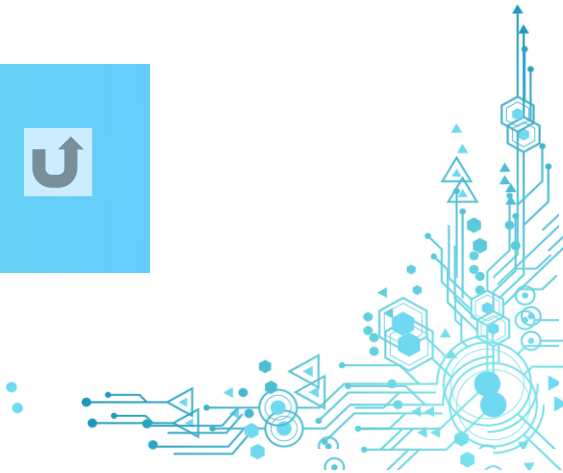
Variáveis numéricas | UNIVARIADA



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.

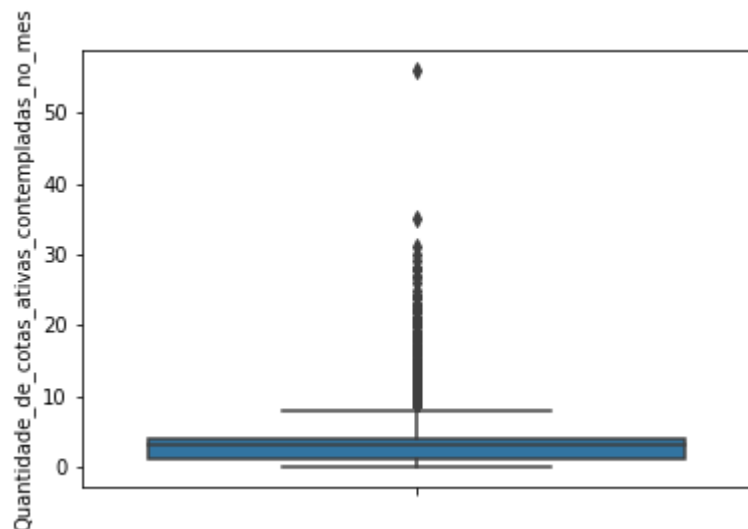
Diferente das cotas ativas contempladas inadimplentes, essa variável apresenta 50% da base com até 62 cotas não contempladas inadimplentes e 25% dos grupos com mais de 118 cotas não contempladas inadimplentes. Observamos maior quantidade de cotas inadimplentes em grupos não contemplados.



# 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.

