

Universidade Federal do Paraná
Setor de Ciências Exatas
Departamento de Estatística
Programa de Especialização em *Data Science* e *Big Data*

Maria Eduarda Kumlehn

**Análise de probabilidade de inadimplência em
uma empresa de energia**

**Curitiba
2022**

Maria Eduarda Kumlehn

Análise de probabilidade de inadimplência em uma empresa de energia

Monografia apresentada ao Programa de Especialização em Data Science e Big Data da Universidade Federal do Paraná como requisito parcial para a obtenção do grau de especialista.

Orientador: Walmes Marques Zeviani

Curitiba
2022

Análise de probabilidade de inadimplência em uma empresa de energia

Maria Eduarda Kumlehn¹
Walmes Marques Zeviani²

Resumo

O setor de energia no Brasil é marcado por sua alta taxa de inadimplência e seu baixo custo de multa e juros, fazendo com que a conta de luz seja a primeira a não ser paga caso o pequeno negócio precise equilibrar o fluxo de caixa. O objetivo deste estudo foi entender se existem características que possuem uma maior correlação com as taxas de inadimplência dentro de uma empresa de energia que atua em mais de um estado, para compreender características que hoje são subjetivas para o negócio. Dentro do estudo, serão identificadas 8 variáveis que influenciam de maneira mais clara a variável resposta, de todas as 29 potenciais iniciais.

Palavras-chave: Inadimplência, energia, correlação

Abstract

The energy sector in Brazil is marked by its high rate of default and its low cost of fines and interest, making the electricity bill the first not to be paid if the small business needs to balance the cash flow. The objective of this study was to understand if there are characteristics that have a greater correlation with the delinquency rates within an energy company that operates in more than one state, to understand characteristics that are currently subjective for the business. Within the study, 8 variables will be identified that most clearly influence the response variable, out of all 29 initial potentials.

Keywords: Non-payment, energy, correlation *palavras-chave.*

1 Introdução

1.1 Contexto

O Brasil hoje possui um dos maiores índices de famílias endividadas e inadimplentes do mundo e este cenário não é muito diferente quando olhamos para o mercado de micro e pequenas empresas. Segundo no Serasa S/A [1] a inadimplência atingiu 5,44 milhões de micro e pequenos negócios em fevereiro deste ano, um aumento de 0,5% em comparação ao mesmo período de 2021.

Após o cenário pandêmico derivado do Covid-19, as micro e pequenas empresas ainda sofrem para colocar suas contas em dia e recuperar os danos causados pela ausência de consumo ou prestação de serviço. Além disso, as altas taxas de juros e o aumento da inflação dificultam para que o fluxo de caixa das empresas termine de maneira positiva.

Um dos setores mais impactados pela alta da inflação foi o de energia. Segundo a Associação Brasileira de Defesa do Consumidor (Proteste) [2], mais de 40% dos entrevistados durante uma entrevista tem dificuldade de pagar suas contas de luz e optam por desligar aparelhos para proteger suas finanças.

Mesmo com a alta dos preços e custos relacionados ao setor de energia, o mercado não para de crescer ano após ano. Conhecemos comumente o setor a partir das distribuidoras de energia, porém um novo mercado promissor desperta o interesse de empresas, o mercado da geração distribuída.

O modelo de geração, transmissão e distribuição de energia elétrica que conhecemos já não atende mais às necessidades da população e muito menos o bolso das pequenas empresas. Segundo Ricardo Sauaia [3], o setor elétrico encontra-se em meio a um importante processo de transformação, similar ao vivenciado pelo setor de telecomunicações há 30 anos.

A geração distribuída vem para reduzir o impacto ambiental causado pelas grandes usinas no Brasil e dar a oportunidade para as empresas consumidores de fontes mais sustentáveis e descentralizadas. Porém o principal valor a ser avaliado considerando o momento atual do mercado, é a economia gerada para os pequenos negócios, tornando a conta de luz uma dor menor para o orçamento destes.

1.2 A empresa

A Lemon Energia é um marketplace de energia que funciona dentro do mercado de geração distribuída. A empresa conecta geradores de energia sustentável com clientes finais, oferecendo uma nova forma de consumir energia limpa e mais barata, sem taxas, obras ou instalações.

Fundada em 2019, a empresa atende cerca de 3 mil pequenos negócios e traz uma economia que pode chegar a 20% na conta de luz, segundo os sócios [4]. Desde a fundação, a empresa faz uma estimativa de R\$ 4,3 mi-

¹Aluno do programa de Especialização em Data Science & Big Data, mekumlehn@gmail.com.

²Professor do Departamento de Estatística - DEST/UFPR.

lhões em custos evitados para os consumidores, assim como 3 mil toneladas de CO2 a menos na atmosfera.

1.3 O problema

Apesar do mercado promissor, a mudança tecnológica do setor, o cenário do não pagamento de contas do Brasil e do serviço não passa sem impactar a empresa. A inadimplência passou a ser um risco para o negócio, financeiro e também em termos de experiência para o consumidor.

Para a Lemon, a inadimplência se torna uma das métricas mais críticas para a organização e que vários times podem apoiar e ajudar. Além disso, tem consumido um valor considerável da margem bruta do negócio com um custo que não é suportável para o crescimento.

Apesar disso, existe pouca clareza sobre a métrica, além do momento atual do mercado que pode contribuir para uma tomada de decisão por parte das pessoas internamente para que o valor seja melhor administrado.

A *inadimplência* é calculada considerando o valor em aberto das contas já vencidas dividido pelo valor total emitido das contas que já passaram a data de vencimento:

$$\text{Inadimplência: } I = \frac{\text{Valoremaberto}}{\text{Valortotalemitido}}$$

Para a Lemon a inadimplência atual cresce a cada mês e se torna um risco grande para o fluxo de caixa, a relação com o parceiro e o impacto na receita. Durante os últimos 12 meses (mar/21 a fev/22) a Lemon deixou de receber R\$ 1.05 M por conta dos clientes inadimplentes, do total de R\$ 17.3 M emitidos, representando uma inadimplência de 6%, conforme a Figura 1.

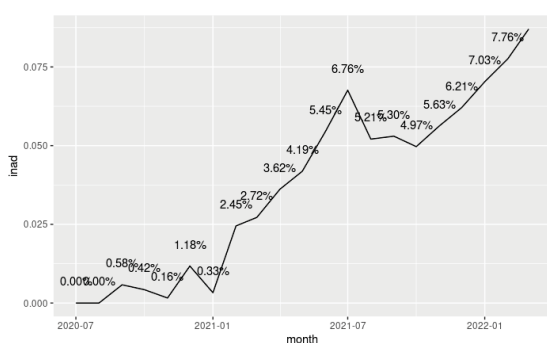


Figura 1: Percentual de inadimplência da empresa no últimos 12 meses.

Observando os valores acumulados de resultado, conforme Figura 2, percebe-se na comparação entre o GMV (receita) e o valor de inadimplência em reais ano após ano (março vs março), que o crescimento foi muito mais acentuado para a métrica do problema, do que para a receita. O crescimento do valor percentual de inadimplência, cresceu 22,2x comparando março de 2021 a março de 2022, enquanto a receita apenas 7,26x.

month	gmv	inad	inad_reais
2022-03-01	\$3.051,235	8,70%	\$263,351
2022-02-01	\$2.619,110	7,76%	\$197,112
2022-01-01	\$2.553,841	7,03%	\$176,679
2021-12-01	\$2.299,711	6,21%	\$141,335
2021-11-01	\$2.094,369	5,63%	\$117,226
2021-10-01	\$2.013,018	4,97%	\$99,879
2021-09-01	\$1.483,896	5,30%	\$78,867
2021-08-01	\$991,806	5,21%	\$51,660
2021-07-01	\$878,987	6,76%	\$59,100
2021-06-01	\$791,181	5,45%	\$43,066
2021-05-01	\$607,347	4,19%	\$25,367
2021-04-01	\$573,050	3,62%	\$20,717
2021-03-01	\$419,793	2,72%	\$11,421

Figura 2: Tabela com valores de receita emitidos (GMV) versus valor de inadimplência dos últimos 12 meses

Considerando os números apresentados, a importância do entendimento da inadimplência é crucial para o momento atual e vital para a sobrevivência no longo prazo da empresa. Portanto, a partir de informações do cliente, sua jornada com a Lemon e do mercado, o estudo se propõe a descobrir as características que estão melhor correlacionadas a inadimplência dentro da empresa.

2 Materiais e Métodos

Acredita-se que existem diversos fatores para o crescimento da inadimplência e o principal foco da análise é entender se existem fatores que levam um cliente a ser mais inadimplente ou não dentro da empresa e assim tomar ações para reduzir o impacto para a saúde financeira da empresa e experiência dos clientes.

2.1 O conjunto de dados

A base única de dados do estudo foi baseada em três datasets que a empresa possui e normalizada para uma única base com 2351 clientes, seus dados de inadimplência consolidados e 30 potenciais variáveis que podem ou não ter impacto na inadimplência da empresa.

- Fonte de variáveis resposta: Base financeira da empresa, com todos os boletos emitidos por cliente e suas informações de vencimento, referência e pagamento.
- Fonte de perfil de cliente: Informações relacionadas ao ciclo de vida do cliente durante a sua jornada de compra do serviço.
- Fonte de jornada: Indicadores relacionados ao perfil de contrato e energia que o cliente recebe em seu estabelecimento.

A primeira fonte (A) utilizada foi a base financeira da empresa, com todos os boletos emitidos por cliente e suas informações de vencimento, referência e pagamento. A partir dessas informações foi criada a referência por cód. dos clientes, que será utilizado apenas para união das bases e outros 9 indicadores que poderão ser utilizados como preditores para as análises de inadimplência. São eles:

Tabela 1: Dicionário do conjunto de dados fonte A

Variável	Descrição
status cliente	informação se o cliente é adimplente em todas as suas contas, inadimplente em todas as suas contas ou inadimplente em apenas alguma de suas contas.
boletos adimplentes	número de boletos adimplentes do cliente.
boletos inadimplentes	número de boletos inadimplentes do cliente.
boletos totais	total de boletos emitidos para o cliente.
percentual inadimplente boletos	valor percentual de boletos inadimplentes, considerando o total de boletos emitidos.
receita adimplente	soma do valor adimplente de todos os boletos pagos pelo cliente.
receita inadimplente	soma do valor inadimplente de todos os boletos não pagos pelo cliente.
receita total	soma do valor de todos os boletos emitidos para o cliente.
percentual inadimplente reais	valor percentual inadimplente em reais, comparado ao total emitido para o cliente.

A segunda fonte (B) que foi compilada com a base de inicial foi de dados de perfil e informações relacionadas ao ciclo de vida do cliente durante a sua jornada de compra do serviço. A principal hipótese a ser validada, é o entendimento se existem características de perfil ou jornada que impactam para a inadimplência ou não. Outras 23 variáveis foram adicionadas ao modelo e serão validadas se possuem ou não correlação com o problema. São elas:

Tabela 2: Dicionário do conjunto de dados fonte B

Variável	Descrição
estado	estado de origem da primeira conta de luz importada no momento do cadastro na plataforma.
disco	distribuidora de origem da primeira conta de luz importada no momento do cadastro na plataforma.
cidade	cidade de origem da primeira conta de luz importada no momento do cadastro na plataforma.
email type	provedor de email utilizado pelo cliente.

Tabela 3: Dicionário do conjunto de dados fonte B (continuação)

Variável	Descrição
email grouptype	agrupamento do provedor de email considerando maturidade tecnológica em: genérico, business e lemon.
cpf cnpj	tipo de documento cadastrado na distribuidora de energia como titular da conta de luz.
media historica kwh	média histórica em kwh de consumo de energia dos últimos 12 meses no momento do cadastro da conta na empresa.
group of media	agrupamento de média histórica de consumo de energia do momento de cadastro.
cdd campaign _a mbev	informação de origem da campanha feita pela Ambev para indicar clientes.
cdd cidade	informação da cidade de origem da empresa que indicou o cliente para a Lemon.
canal lemon	canal de origem por onde o cliente conheceu a Lemon.
source group	grupo de source de origem por onde o cliente conheceu a Lemon.
source	source de origem por onde o cliente conheceu a Lemon.
medium	medium de origem por onde o cliente conheceu a Lemon.
vendedor	vendedor que fez a venda durante o processo de cadastramento da conta na Lemon.
economia proposta	valor ofertado de economia anual para o cliente no momento da venda.
data lead	data de entrada do cadastro do cliente na Lemon.
data assinatura	data de assinatura do contrato do cliente na Lemon.
data onboarding	data de finalização do onboarding do cliente após a assinatura do contrato.
data usina	data de inclusão do cliente na usina de energia para recebimento da primeira conta Lemon.
ciclo venda	tempo em dias entre o cadastro e a assinatura de contrato.
ciclo onboarding	tempo em dias entre a venda e a finalização do onboarding.
ciclo usina	tempo em dias entre a finalização do onboarding e a adição do cliente na usina de energia.

Por fim, foi adicionada uma terceira base (C) com outros 6 indicadores relacionados ao perfil de contrato e energia que o cliente recebe em seu estabelecimento. Abrindo uma hipótese mais ampla sobre o que ofertamos de economia para o cliente na conta e o que o modo de consumo de energia pode nos trazer de impacto para a inadimplência. Os dados adicionados foram:

Variável	Descrição
classe	classe de energia utilizada pelo cliente padrão da distribuidora.
tipo conexao	tipo de conexão utilizado pelo cliente conforme seu volume de consumo.
modalidade tarifaria	tipo de modalidade tarifária utilizada pelo cliente.
prazo fidelidade	modelo de fidelidade acordado com cliente no momento da assinatura do contrato com a Lemon.
protecao bandeira	benefício acordado com o cliente no momento da assinatura do contrato.
desconto cliente	percentual de desconto negociado com o cliente no momento da assinatura do contrato.

A partir das três bases de dados, a base de análise conta com uma base de clientes de julho de 2020 a março de 2021, com 2.351 clientes com 1 a N boletos emitidos e que servirá para entender através de suas características as informações que possuem maior correlação ou impacto na inadimplência.

2.2 A escolha da variável resposta

A base de dados inicial (A) possuía algumas variáveis que poderiam se tornar respostas, portanto foi avaliado a melhor escolha para o momento, não excluindo a possibilidade de fazer no futuro mais análises com as outras alternativas.

A primeira exclusão se deu, considerando a regra de negócio que era entender melhor o perfil de inadimplentes independente do valor da inadimplência, para que assim um cliente que possuísse apenas um boleto em atraso, porém com um alto valor não impactasse a análise e trouxesse conclusões que não ajudariam.

Foi feita a análise univariada de três variáveis possíveis para entender o que encaixava melhor com a premissa do estudo e que poderia ser melhor aplicada na análise de correlação com as outras variáveis.

A primeira variável foi a *status cliente*, que tem como definição o agrupamento de clientes considerando sua inadimplência em três classificações: inadimplente, alguma inadimplência, adimplente. Conforme a Figura

3, é possível perceber uma alta concentração de cliente dentro do status adimplente e cerca de 13% nos grupos restantes, sendo aproximadamente 10% em alguma inadimplência e 3.7% em inadimplência.

No	Variable	Stats / Values	Freqs (% of Valid)	Graph	Valid	Missing
1	status_cliente [factor]	1. adimplente 2. alguma inadimplência 3. inadimplente	2028 (86.3%) 236 (10.0%) 87 (3.7%)		2351 (100.0%)	0 (0.0%)

Figura 3: Valores, frequência e distribuição da variável status do cliente.

A segunda análise feita foi para a variável *número de boletos inadimplentes*. Na Figura 4 pode se observar que existem 15 possíveis boletos e uma concentração mais agrupada de clientes em nenhum boleto inadimplente.

No	Variable	Stats / Values	Freqs (% of Valid)	Graph	Valid	Missing
1	boletos_inadimplentes [numeric]	Mean (sd) : 0.5 (1.5) min ≤ med ≤ max: 0 ≤ 0 ≤ 22 IQR (CV) : 0 (3.3)	15 distinct values		2351 (100.0%)	0 (0.0%)

Figura 4: Valores, frequência e distribuição da variável número de boletos inadimplentes.

A última variável resposta possível analisada foi o *percentual de inadimplência por boletos emitidos*. Conforme Figura 5 a distribuição ainda é concentrada no valor zero, porém abre-se a oportunidade de comparação com 63 outros valores para análises mais robustas.

No	Variable	Stats / Values	Freqs (% of Valid)	Graph	Valid	Missing
1	percentual_inadimplente_boletos [formattable, numeric]	Mean (sd) : 10.00% (0.2) min ≤ med ≤ max: 0.00% ≤ 0.00% ≤ 100.00% IQR (CV) : 0 (3)	64 distinct values		2351 (100.0%)	0 (0.0%)

Figura 5: Valores, frequência e distribuição da variável percentual de inadimplência em boletos.

Ao fazer a análise univariada das possíveis variáveis resposta, entendendo frequência, setores, moda para a variável qualitativa e frequência, histograma, moda, mediana e métricas de dispersão para as variáveis quantitativas, a melhor escolha é a análise considerando o percentual de inadimplência dos boletos.

A variável trará comparações e agrupamentos mais robustos e assim, tornar mais profundo o plano de ação que a empresa espera para a indicador que é tão crítico.

2.3 Limpeza e preparo dos dados

Após a análise descritiva das variáveis respostas foi necessário realizar a análise descritiva de todas as variáveis preditoras, afim de entender se alguns fatores não se sobrepunham ou se o nível de distribuição trazia informações que não agregavam para as correlações que seriam feitas nas análises bivariadas.

As variáveis preditoras estado e disco possuem a mesma distribuição, portanto apenas a variável disco será utilizada na comparação com a inadimplência.

As variáveis cidade, email type, media historica, cdd campaign ambev, cdd cidade, source, medium e vendedor foram retirados da análise bivariada, pelo nível de distribuição dos dados ou comparação com dados já existentes em outros grupo.

É possível ver a análise de todas as variáveis de maneira detalhada nas Figuras 6, 7, 8 e 9. Variáveis com uma distribuição muito ampla, com valores únicos acima de dezenas, valores como 'outros' dentro de algumas das opções com maior distribuição ou com distribuições muito próximas de outras variáveis não foram consideradas, pois não possuem conclusões melhores do que as outras escolhidas.

Após a análise descritiva com as variáveis preditoras e ajustes para base de descrição bivariada reduziu-se o número de variáveis para 16.

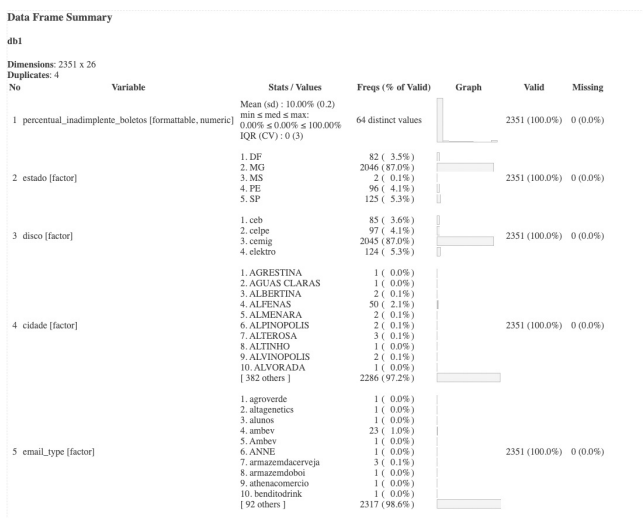


Figura 6: Valores, frequência e distribuição das variáveis preditoras (1)

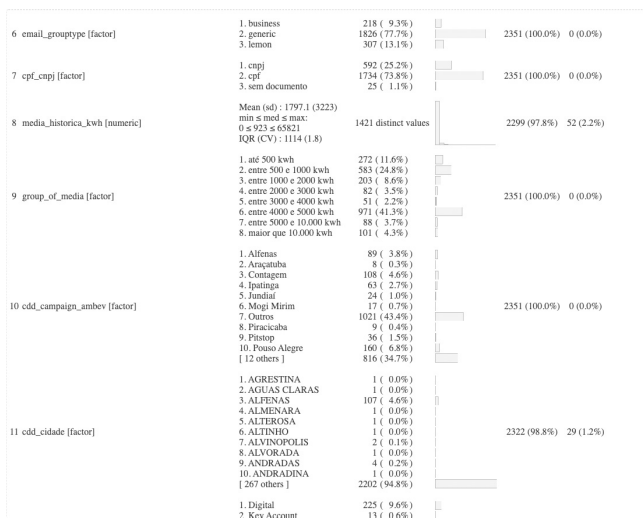


Figura 7: Valores, frequência e distribuição das variáveis preditoras (2)

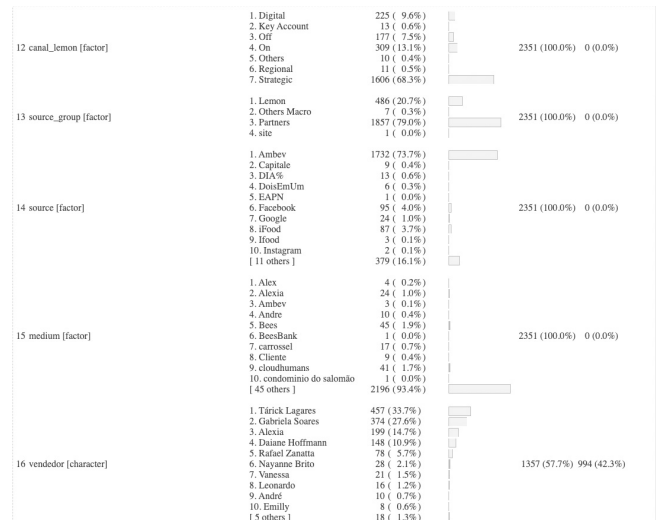


Figura 8: Valores, frequência e distribuição das variáveis preditoras (3)

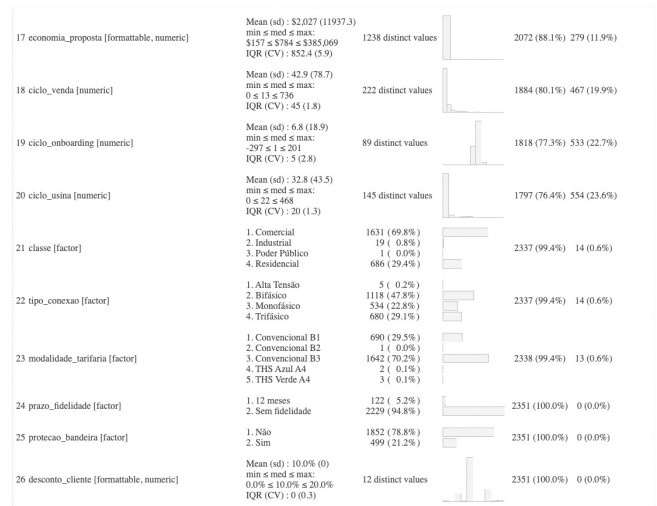


Figura 9: Valores, frequência e distribuição das variáveis preditoras (4)

2.4 Metodologia de análise

A partir da nova base de dados e as variáveis escolhidas, deu-se início a análise descritiva bivariada dentro do software R [5] utilizando diversos pacotes e modelos de visualização para facilitar a apresentação e compartilhando dos resultados com a empresa do estudo.

3 Resultados e Análises

Conforme Cunha (1968, p. 3) [6],

[...] método experimental é a observação do fato e o controle das causas que possam influir no mesmo, uma por uma, até se identificar a causa principal ou se poder avaliar a influência que

cada uma delas possa exercer sobre o comportamento do fenômeno, chegando-se a um conhecimento que, se for verdadeiro, poderá ser comprovado por novas experiências.

Apesar da análise não ter permitido novos experimentos, as conclusões aqui apresentadas serão levadas em consideração para serem testadas e apresentadas como hipóteses para estudo e testes dentro da empresa.

A análise bivariada será a grande entrega de valor para a empresa que optou por não querer um modelo e sim conclusões que apoiassem de maneira fácil a tomada de decisão.

Para a análise foram considerados dois modelos:

- ▶ Análise descritiva bivariada - qualitativa x quantitativa: todas as variáveis foram comparadas com a variável resposta percentual de inadimplência por número de boletos: canal lemon, classe, cpf cnpj, disco, email grouptype, modalidade tarifaria, prazo fidelidade, protecao bandeira, source group, tipo conexao, group of media. As análises feitas foram frequência (boxplot/bars), resumo médio, mediana desvio padrão por categoria.

A partir dos histogramas, médias e medianas das variáveis para o primeiro modelo, percebeu-se que algumas variáveis não apresentavam nenhuma variação para as suas distribuições considerando a variável resposta escolhida. Dentro elas, a variável tipo de conexão, por exemplo apresentou valores zerados para todos os quadrantes, assim como também percentuais de distribuição iguais para todas as possíveis opções preditoras versus a variável resposta, conforme Figura 10.

```

tapply(db1$percentual_inadimplente_boletos, db1$tipo_conexao, mean)

## Alta Tensão      Bifásico      Monofásico      Trifásico
## 0.00000000  0.07988264  0.07262057  0.07793303

tapply(db1$percentual_inadimplente_boletos, db1$tipo_conexao, sd)

## Alta Tensão      Bifásico      Monofásico      Trifásico
## 0.00000000  0.2367342  0.2216793  0.2302379

tapply(db1$percentual_inadimplente_boletos, db1$tipo_conexao, quantile)

## $`Alta Tensão`
##  0%  25%  50%  75% 100%
## 0.00% 0.00% 0.00% 0.00% 0.00%
##
## $Bifásico
##  0%  25%  50%  75% 100%
## 0.00% 0.00% 0.00% 0.00% 100.00%
##
## $Monofásico
##  0%  25%  50%  75% 100%
## 0.00% 0.00% 0.00% 0.00% 100.00%
##
## $Trifásico
##  0%  25%  50%  75% 100%
## 0.00% 0.00% 0.00% 0.00% 100.00%

```

Figura 10: Modelo de análise bivariada com variável tipo de conexão

Considerando essa mesma análise, outras duas variáveis se comportaram da mesma maneira e foram consideradas como irrelevantes para a variável resposta que

foram: prazo de validade e modalidade tarifaria. As outras variáveis do estudo, considerando o modelo de análise qualitativa x quantitativa, possuem alguma variável dentro das possíveis preditoras vs a variável resposta, o que leva a empresa a querer testar oportunidades de hipóteses e ações para melhor compreensão do cliente.

Análise descritiva bivariada - quantitativa x quantitativa: todas as variáveis foram comparadas com a variável resposta percentual de inadimplência por número de boletos: desconto cliente, economia proposta, percentual inadimplente boletos, ciclo onboarding, ciclo usina, ciclo venda. Para as descrições quantitativas, frequência e diagramas de dispersão foram as formas mais fáceis de compreender as correlações. Além das análises numéricas de coeficiente de correlação linear de Pearson.

Para a análise descritiva bivariada no modelo quantitativa vs quantitativa, as variáveis analisadas passaram por análises quantitativas, de frequência, além da correlação linear de Pearson.

Todas as variáveis preditoras quantitativas não tiveram bons resultados na análise de correlação, com valores baixos e com P-value maiores que 0,1 e pouca ou nenhuma evidência contra a variável resposta, conforme Figuras 11, 12, 13, 14 e 15.

```

Pearson's product-moment correlation

data: db1$percentual_inadimplente_boletos and db1$ciclo_onboarding
t = -0.37842, df = 1816, p-value = 0.7052
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.05483030  0.03710861
sample estimates:
 cor
-0.008879609

```

Figura 11: Análise de Pearson para variável ciclo de onboarding versus variável resposta percentual de inadimplência

```

Pearson's product-moment correlation

data: db1$percentual_inadimplente_boletos and db1$ciclo_usina
t = 3.674, df = 1795, p-value = 0.0002458
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.04031301 0.13210618
sample estimates:
 cor
0.08639295

```

Figura 12: Análise de Pearson para variável ciclo de usina versus variável resposta percentual de inadimplência

```

Pearson's product-moment correlation

data: db1$percentual_inadimplente_boletos and db1$ciclo_venda
t = -1.1458, df = 1882, p-value = 0.252
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.07147694  0.01878112
sample estimates:
      cor
-0.02640172

```

Figura 13: Análise de Pearson para variável ciclo de venda versus variável resposta percentual de inadimplência

```

Pearson's product-moment correlation

data: db1$percentual_inadimplente_boletos and db1$desconto_cliente
t = -4.4444, df = 2349, p-value = 9.225e-06
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.13125801 -0.05107884
sample estimates:
      cor
-0.09131642

```

Figura 14: Análise de Pearson para variável desconto cliente versus variável resposta percentual de inadimplência

```

Pearson's product-moment correlation

data: db1$percentual_inadimplente_boletos and db1$economia_proposta
t = 1.9544, df = 2070, p-value = 0.05079
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.0001466924  0.0858199656
sample estimates:
      cor
0.04291607

```

Figura 15: Análise de Pearson para variável economia proposta versus variável resposta percentual de inadimplência

4 Conclusões

A partir do estudo obteve-se 8 variáveis que influenciam a variável resposta de maneira mais clara e que a empresa atuará de maneira clara para tentar impactar e projetar melhor as métricas para o futuro.

Algumas das variáveis possuem plano de ação claro interno dentro da empresa e outras servirão como apoio para a alocação de um crédito mais sustentável e menos arriscado para a empresa. Como o caso CNPJ x CPF, onde cliente CPF possuem maior inadimplência e recebem menor energia Lemon de entrada, assim reduzindo o risco de inadimplência em termos de valores monetários para a empresa.

Entende-se que o estudo comporte variações e outras análises complementares que podem levar a conclusões diferentes das encontradas. Dentro dos estudos relevantes podem estar:

- Análise descritivas e modelos com base em outras variáveis resposta, como o agrupamento do clientes por status ou o valor de inadimplência financeira.

- Para as variáveis preditoras com conhecimento no momento de entrada dos leads, ajustes começaram a ser feitos no crédito disponibilizado para as empresas.
- Novos estudos e modelos focados em entender o processo de pagamento e informações históricas de pagamento do cliente, para score ao longo da vida do cliente.
- Entendimento de outras variáveis e modelos para produção de um modelo que possa ser utilizado por outros times.
- Separação e identificação considerando apenas a base de inadimplência, para redução de ruídos.

Agradecimentos

Gostaria de agradecer a empresa, por ter fornecido uma base de dados detalhada e completa, com várias informações que não apenas podem ser utilizadas para o estudo em questão, mas também para novas conclusões, análises e modelos futuros.

Além disso, gostaria de agradecer a atenção e orientação do meu orientador durante o estudo, que se fez disponível para tirar minhas dúvidas e apoiar quando necessário.

Por fim, gostaria de agradecer meu noivo e família por me apoiarem incondicionalmente a cada entrega, etapa e novo aprendizado, não apenas durante esse projeto, mas principalmente, durante toda a minha vida acadêmica.

Referências

- [1] Serasa Experian. *Inadimplência das micro e pequenas empresas cresceu 0,5% em fevereiro, revela Serasa Experian*. Serasa Experian, 2022.
- [2] Filipe Puente, Beatriz Puente e Brasilda. *Inflação: Mais de 90% dos brasileiros mudaram hábitos de consumo, aponta estudo*. CNN Brasil, 2022.
- [3] Ricardo Sauaia. *Geração distribuída solar fotovoltaica: o novo sempre vem*. Absolar, 2019.
- [4] Rafael Vignoli. *Lemon Energia atrai Kaszek e primeiro cheque da Lowercarbon no Brasil*. Pipeline Valor, 2022.
- [5] Tom Fawcett and Foster Provost. *R: A language and environment for statistical computing*, 2010.
- [6] S. E. CUNHA. *Estatística descritiva: na psicologia e na educação*. Rio de Janeiro: Forense, 1968.