

Perfil da Inadimplência no Brasil: Análise Pré, Durante e Pós-Pandemia Covid-19

Autores: Ana Maria Alves da Silva¹; Fabrizzio Alphonsus Alves de Melo Nunes Soares^{2*}

¹Faculdade de Enfermagem. Universidade Federal de Goiás, Goiânia, GO, Brasil.

² Instituto de Informática. Universidade Federal de Goiás, Goiânia, GO, Brasil.

*Autor correspondente:

Fabrizzio Alphonsus Alves de Melo Nunes Soares, Universidade Federal de Goiás Caixa Postal 131 - CEP 74001-970 - Goiânia - GO, fabrizzio@inf.ufg.br

Resumo

Este trabalho analisa o perfil da inadimplência de pessoas físicas no Brasil em três marcos temporais: pré-pandemia (2019), durante a pandemia (2020) e pós-pandemia (2024) a partir de microdados agregados do SCR. Definiu-se inadimplência como parcela em atraso maior que 90 dias e construíram-se indicadores por modalidade, ocupação, porte/renda, UF e tipo de consolidado bancário (TCB). A estratégia empírica combina estatística descritiva (tabelas e heatmaps) com árvores de decisão (CART) para evidenciar interações entre fatores.

Palavras-chave: Inadimplência; Crédito; Pessoas Físicas; Pessoas Jurídicas; COVID-19

Introdução

O sistema financeiro exerce papel central na sociedade ao transformar poupança em investimento produtivo, administrar riscos, prover meios de pagamento e alocar crédito sob assimetria de informação. Quando cumpre bem essas funções, há ganho social via crescimento, estabilidade e inclusão; quando falha, os custos se propagam para emprego, renda e preços [5]. A pandemia de Covid-19 constituiu um choque macrofinanceiro raro, elevando incerteza, provocou impactos econômicos significativos, afetando a renda das famílias, o consumo e a capacidade de pagamento de dívidas tensionando o canal de crédito. Nesse contexto, a intermediação financeira tornou-se determinante para amortecer ou amplificar a inadimplência, por meio de condições de oferta, preços (spreads) e práticas de renegociação. À luz desse referencial, este estudo compara o perfil de inadimplência de pessoas físicas no Brasil nos períodos pré, durante e pós-pandemia, utilizando dados do SCR [1].

Na Resolução 3.721/09 do BACEM [3], em seu artigo 2º define o risco de crédito como a possibilidade de ocorrência de perdas associadas ao não cumprimento pelo tomador ou contraparte de suas respectivas obrigações financeiras nos termos pactuados, à desvalorização de contrato de crédito decorrente da deterioração na classificação de risco do tomador, à redução de ganhos ou remunerações, às vantagens concedidas na renegociação e aos custos de recuperação, isto é, o risco de crédito e a probabilidade de não cumprimento dos acordos pré estabelecidos e quando o acordo não é cumprido dizemos que o tomador está inadimplente.

Neste trabalho busca-se entender quais mudanças ocorreram na distribuição da inadimplência (90 dias) por modalidade de crédito, faixa de rendimento, unidade da federação e natureza da ocupação nos períodos pré, durante e pós-pandemia. O SCR fornece séries

mensais agregadas e detalhamentos por PF/PJ, modalidade, UF, ocupação, porte/rendimento, origem dos recursos e indexador, o que viabiliza um estudo comparativo factível e replicável com base em dados oficiais e públicos. A métrica de inadimplência do BACEN [2] considera o percentual da carteira com atraso superior a 90 dias, permitindo mensuração consistente no tempo e entre segmentos.

Método Proposto

Para a condução deste estudo, foram extraídos dados do Sistema de Informações de Créditos (SCR) e que foram processados e pós processamento foi realizado a limpeza dos dados, onde foram selecionadas as variáveis utilizadas neste estudo, essa parte é conhecida como preparo e padronização dos dados.

As variáveis selecionadas foram conduzidas por subgrupos de Pessoas Físicas (PF), comparando perfis de inadimplência por modalidade de crédito, unidade da federação (UF), porte/renda, natureza da ocupação, TCB (Tipo de Consolidado Bancário) e SR (segmento regulatório), conforme definições e agregações oficiais do SCR do Banco Central do Brasil (BCB) [1].

A população do estudo contempla operações de crédito cursadas no país e reportadas ao SCR acima do limite de identificação do Documento 3040 [2]. Foram analisados apenas registros de Pessoas Físicas (PF) e, dentro de PF, ocupações de interesse: servidor/empregado público, empregado de empresa privada, aposentado/pensionista, autônomo e MEI categorias definidas pelo SCR. O porte/renda em PF segue as faixas de salários-mínimos também padronizadas pelo SCR [2].

As variáveis usadas no estudo estão descritas na Tabela 1.

Tabela 1: Variáveis selecionadas

Carteira	Totalização de valores a vencer e vencidos (modalidades 01 a 13).
Carteira inadimplida arrastada	Somatório de operações com parcela vencida há mais de 90 dias.
Modalidade (PF)	Cartão, consignado, não consignado, habitacional, rural/agroindustrial, veículos e outros créditos, segundo agregação oficial.
UF	Total nacional e 27 UFs (PF por CEP de residência).
Ocupação (PF)	Categorias oficiais do SCR (servidor público, empregado privado, aposentado/pensionista, autônomo, MEI etc.).
Porte/Renda (PF)	Faixas de rendimentos (sem rendimento; até 1 Salário Mínimo; > 12 Salário Mínimo; ...; acima de 20 Salário Mínimo; indisponível).
TCB/SR	Classificações institucionais para tipo de consolidado bancário (bancário, não bancário, cooperativas) e segmentos S1 - S5.

Além dessas variáveis, criamos as variáveis Taxa de inadimplência que é a carteira inadimplida arrastada dividida pela carteira e Indicador binário de inadimplência que possui valores zeros ou um por célula, sendo um para quando carteira inadimplida arrastada é maior que zero e zero caso contrário.

Após o preparo e padronização dos dados realizamos um estudo descritivo dos dados extraído do Sistema de Informações de Créditos (SCR) do Banco Central do Brasil usando mapas de calor (Heatmap), onde o tipo de heatmap utilizado foi o de mapa de calor de

proporções por grupo, onde as linhas são variáveis e as colunas são os períodos observados. Esse tipo de heatmap nos permite identificar picos ou vales e suas mudanças ao longo do tempo.

Posterior ao heatmap foi realizado uma árvore de decisão (CART) para cada período afim de identificar os perfis de inadimplência para que possamos compara-los. Árvore de decisão é um modelo que aprende regras a partir dos dados para classificar ou prever valores. Ela divide o conjunto em nós usando a variável que melhor separa o alvo: em classificação, mede-se a pureza por Gini ou Entropia; em regressão, costuma-se usar MSE. O algoritmo CART escolhe o melhor ponto de corte a cada passo, criando um caminho de decisões até chegar às folhas (previsões).

A escolha em utilizar árvores de decisão é devido a sua interpretabilidade, pois se traduzem em um conjunto de regras fáceis de explicar. Além disso, árvores de decisão são não lineares, capturando naturalmente interações e efeitos de limiar porém podem sofrer risco de overfitting. O overfitting é fenômeno que ocorre quando o modelo se ajusta demais ao ruído do treino, com consequência de ajustar-se muito bem nesse conjunto, todavia, muito mal em dados novos - já que árvores muito profundas tendem a memorizar ruído controla-se isso com poda e validação cruzada, que foi realizado neste trabalho. Também são sensíveis a variações pois pequenas mudanças nos dados podem alterar a estrutura. Por fim, permitem avaliar a importância das variáveis pela redução de impureza, além de métodos pós-hoc como SHAP [6].

Resultados

Nesta seção, apresentamos os resultados obtidos, iniciando-se com indicadores gerais, isto é, a evolução do volume de contratos e dos indicadores de inadimplência em três pontos do tempo, pré-pandemia (2019), durante a pandemia (2020) e pós-pandemia (2024). Observa-se aumento do total de contratos de 65.957 (2019) para 89.747 (2024), equivalente a cerca de +36%. Apesar do maior volume, a proporção de contratos inadimplentes recuou de 58,6% para 54,7% (-3,9 p.p.) no mesmo período. Já a média da taxa de inadimplência (média simples entre as células analisadas) subiu em 2020 (de 12,82% para 13,85%) e caiu em 2024 (11,91%), sugerindo um pico no período crítico da pandemia, seguido de acomodação, veja Tabela 2.

Tabela 2: Indicadores gerais por ano

Ano	Total de contratos	Inadimplentes	% inadimplentes	Média da taxa de inad.
2019	65.957	38.628	58,57%	12,82%
2020	69.439	40.490	58,31%	13,85%
2024	89.747	49.066	54,67%	11,91%

A Figura 1 mostra a evolução da inadimplência por ocupação. Observa-se aumento em 2020 concentrado em categorias mais expostas ao mercado de trabalho – MEI e emprego privado (ambas em torno de 59,7%) enquanto servidores públicos e aposentados/pensionistas mantiveram níveis estáveis ou ligeiramente menores ante 2019, sinalizando maior proteção de renda no choque da Covid-19. Em 2024, há queda generalizada e níveis inferiores aos de 2019 em todas as ocupações, com reduções mais intensas em servidores públicos (5,4 p.p.) e aposentados/pensionistas (5,0 p.p.). Esses padrões são consistentes com o impacto heterogêneo da pandemia e posterior normalização, ainda que com

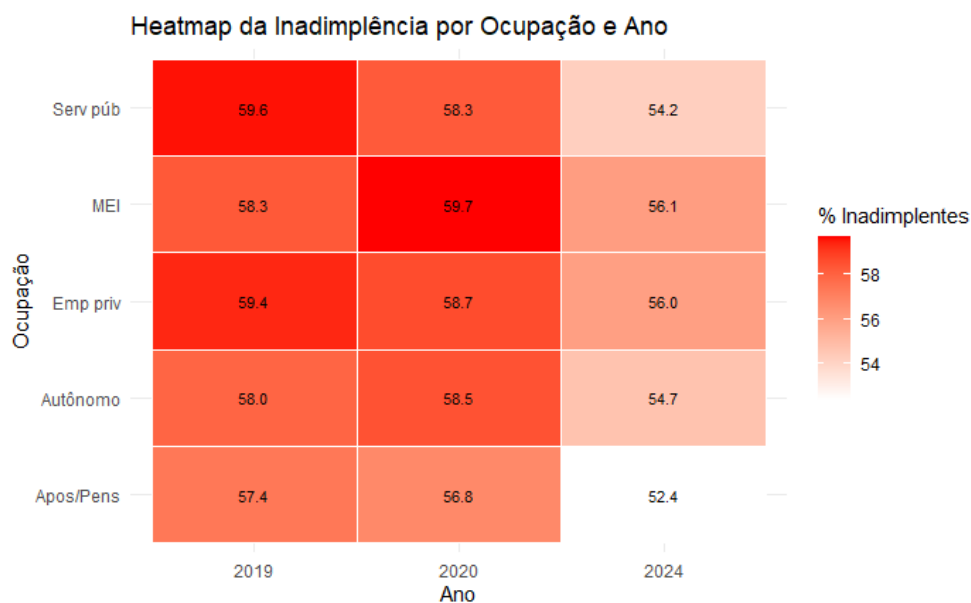


Figura 1: Heatmap da inadimplência por ocupação e ano (2019, 2020 e 2024), onde Serv púb, MEI, EMP priv, Apos/Pens significam respectivamente, Servidor Público, Micro Empreendedor Individual, Empregado em Empresa Privada, Aposentado ou Pensionista. Fonte: elaboração própria com dados do SCR.

diferenças modestas entre grupos ocupacionais.

A Figura 2 sumariza a inadimplência por modalidade. Observa-se que cartão mantém a maior proporção de inadimplentes em todos os períodos (83,3% em 2019 para 76,0% em 2024), enquanto habitação e rural apresentam os menores níveis e trajetória descendente. Entre os créditos pessoais, há redução no pós-pandemia tanto no sem consignação (7,8 p.p.) quanto no consignado (5,8 p.p.), embora permaneçam acima das modalidades com garantias. O ano de 2020 revela efeitos heterogêneos do choque: piora moderada em veículos e no consignado, leve melhora em cartão, sugerindo respostas distintas por tipo de contrato. No geral, 2024 indica normalização com melhora ampla, mas com persistência de risco relativamente maior em cartão e crédito pessoal.

A Figura 3 evidencia heterogeneidade geográfica da inadimplência. Observam-se níveis mais elevados no Norte e Nordeste, com picos em 2020 (p.ex., PA 66,0%, PE 63,9%, CE 63,7%), ao passo que Sul/Sudeste/CO apresentam patamares menores (p.ex., SC em torno de 51,53%, PR em 53,54%). Em 2024, há queda disseminada face a 2019 em vários casos entre 2 e 7 p.p. (como ES 58,7% para 51,8% e MS 58,1% para 51,4%) sinalizando normalização pós-pandemia. O choque de 2020, contudo, foi assimétrico entre estados, compatível com diferenças de composição setorial, mercado de trabalho e políticas locais.

A Figura 4 evidencia um gradiente socioeconômico consistente: faixas de menor renda apresentam maior proporção de inadimplentes em todos os períodos, ao passo que as faixas superiores mantêm níveis mais baixos (ex.: >20 SM em torno de 4,24%).

O choque de 2020 foi moderado e heterogêneo com leve aumento em 23 SM e na classe Indisponível e, em 2024, observa-se queda disseminada frente a 2019 (reduções de até 7 p.p.), ainda que o desnível por renda permaneça.

A Figura 5 compara a inadimplência por tipo de consolidado bancário (TCB). Observa-se que o segmento bancário apresenta os maiores níveis, mas exibe redução expressiva no pós-pandemia (65,2% em 2019 para 57,3% em 2024). O segmento não bancário perma-

nece relativamente estável, com leve melhora (61,2%58,7%). Já as cooperativas, embora mantenham o menor patamar entre os grupos, registram elevação ao longo do período (37,1%44,1%). Esses resultados sugerem respostas heterogêneas por TCB e possível influência de composição de carteira; por isso, no restante do estudo controlamos por modalidade e porte/renda para avaliar a significância dessas diferenças.

As árvores de decisão (6, 7 e 8) mostram, de forma consistente, uma segmentação inicial por modalidade, separando habitação/rural das demais, com menores níveis de inadimplência; ramos associados a crédito pessoal/cartão combinados a baixa renda e determinadas UFs apresentam maiores taxas. Em 2020, ramos ligados a ocupações mais expostas e faixas de renda intermediárias ganham relevância, ao passo que, em 2024, observa-se redução das taxas em diversos ramos, embora segmentos de cartão permaneçam críticos. Note que as árvores estão corroborando com os padrões dos heatmaps e evidenciando interações entre modalidade, renda, UF e TCB.

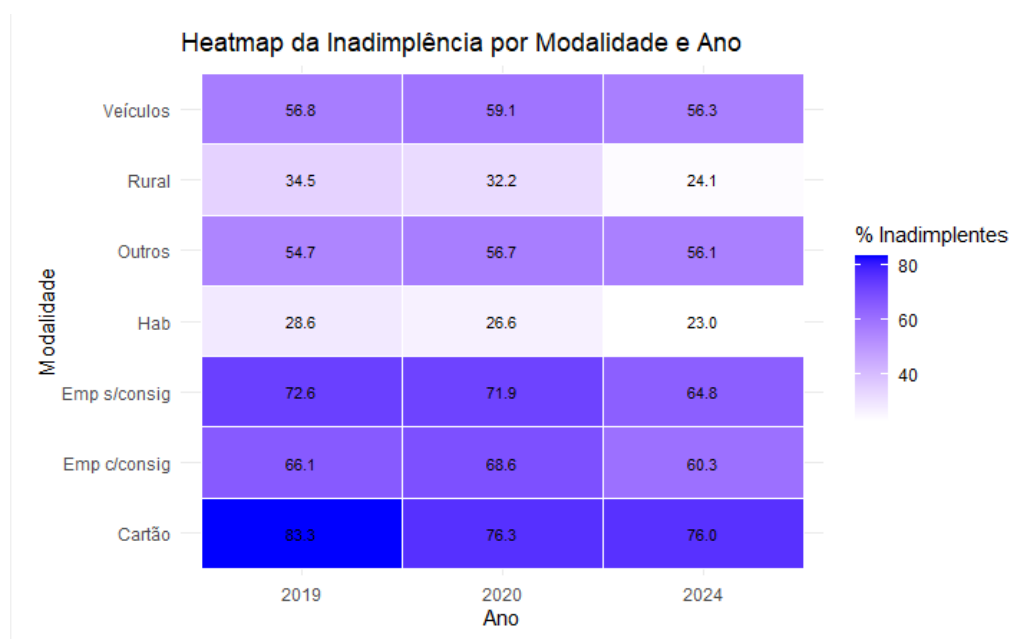


Figura 2: Heatmap da inadimplência por modalidade e ano (2019, 2020 e 2024), onde Emp s/consig e Emp c/consig significam Empréstimo Sem Consignado e Empréstimo Consignado respectivamente.

Fonte: elaboração própria com dados do SCR.

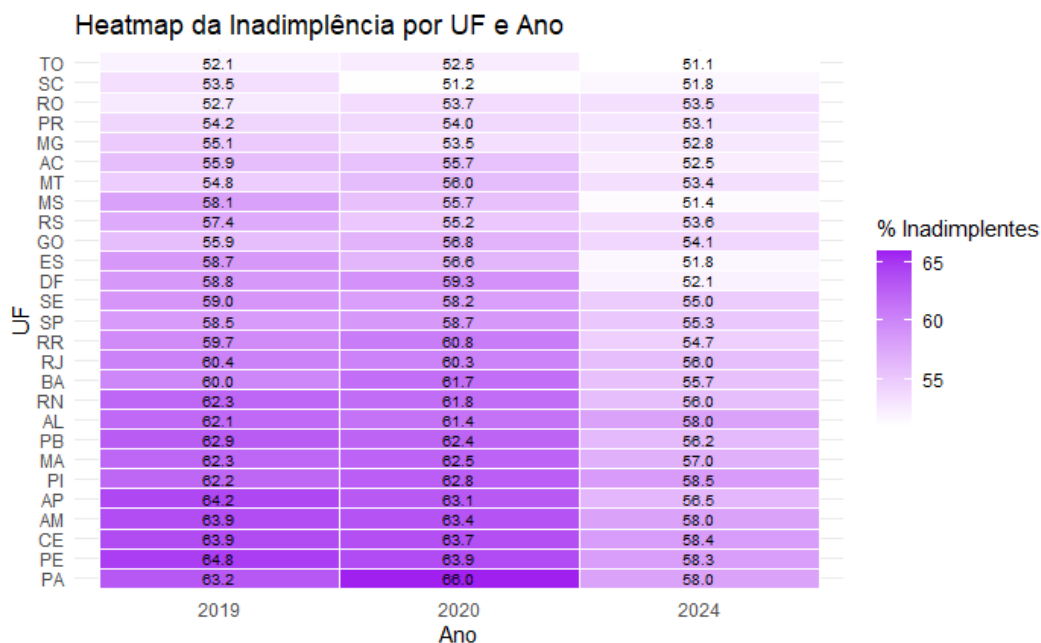


Figura 3: Heatmap da inadimplência por UF e ano (2019, 2020 e 2024).
Fonte: elaboração própria com dados do SCR.

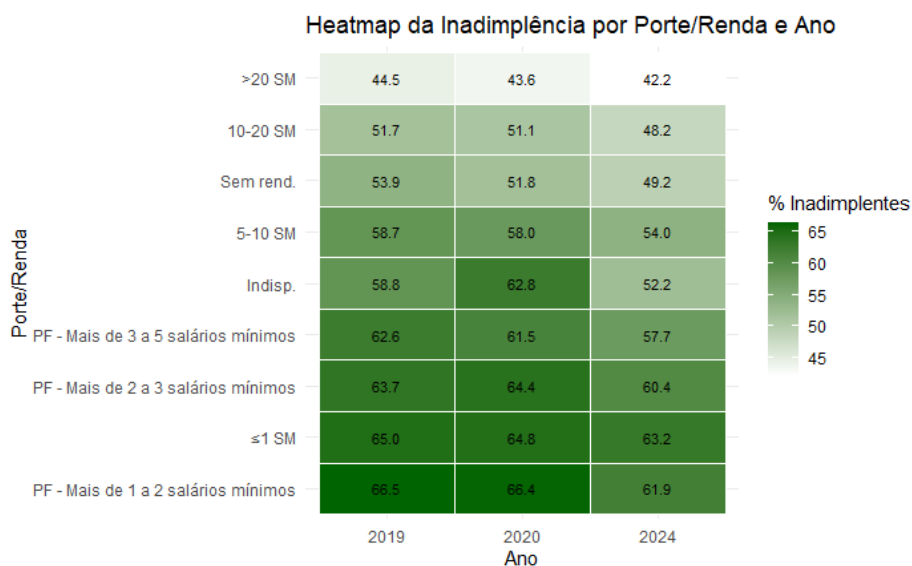


Figura 4: Heatmap da inadimplência por porte/renda e ano (2019, 2020 e 2024), onde SM significa Salário Mínimo, Sem rend. significa sem renda.
Fonte: elaboração própria com dados do SCR.

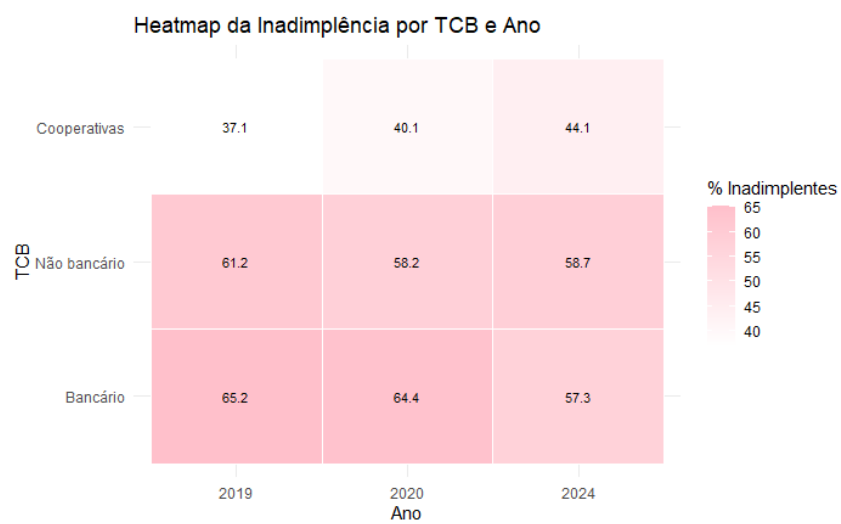
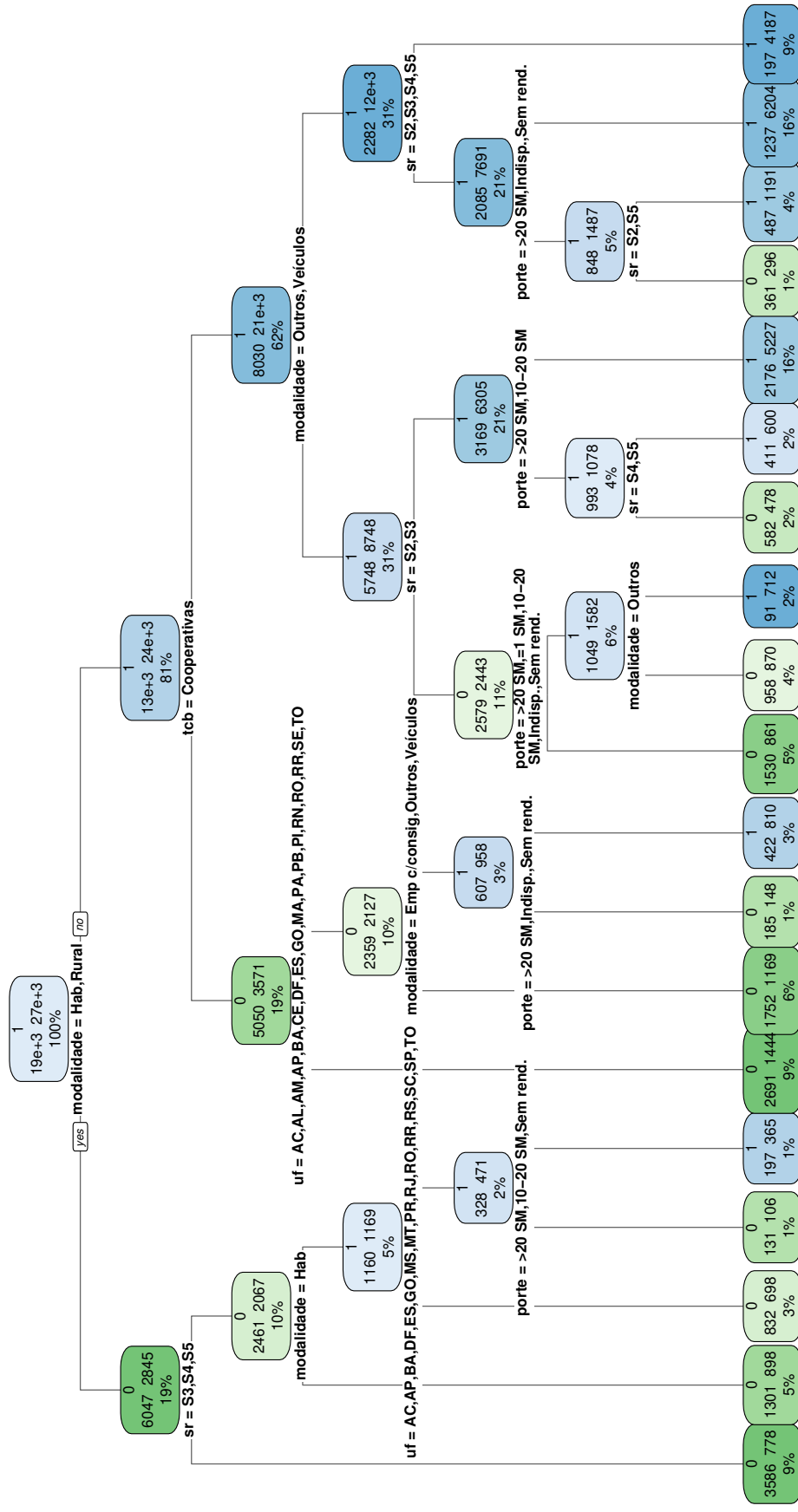


Figura 5: Heatmap da inadimplência por TCB e ano (2019, 2020 e 2024).
Fonte: elaboração própria com dados do SCR.



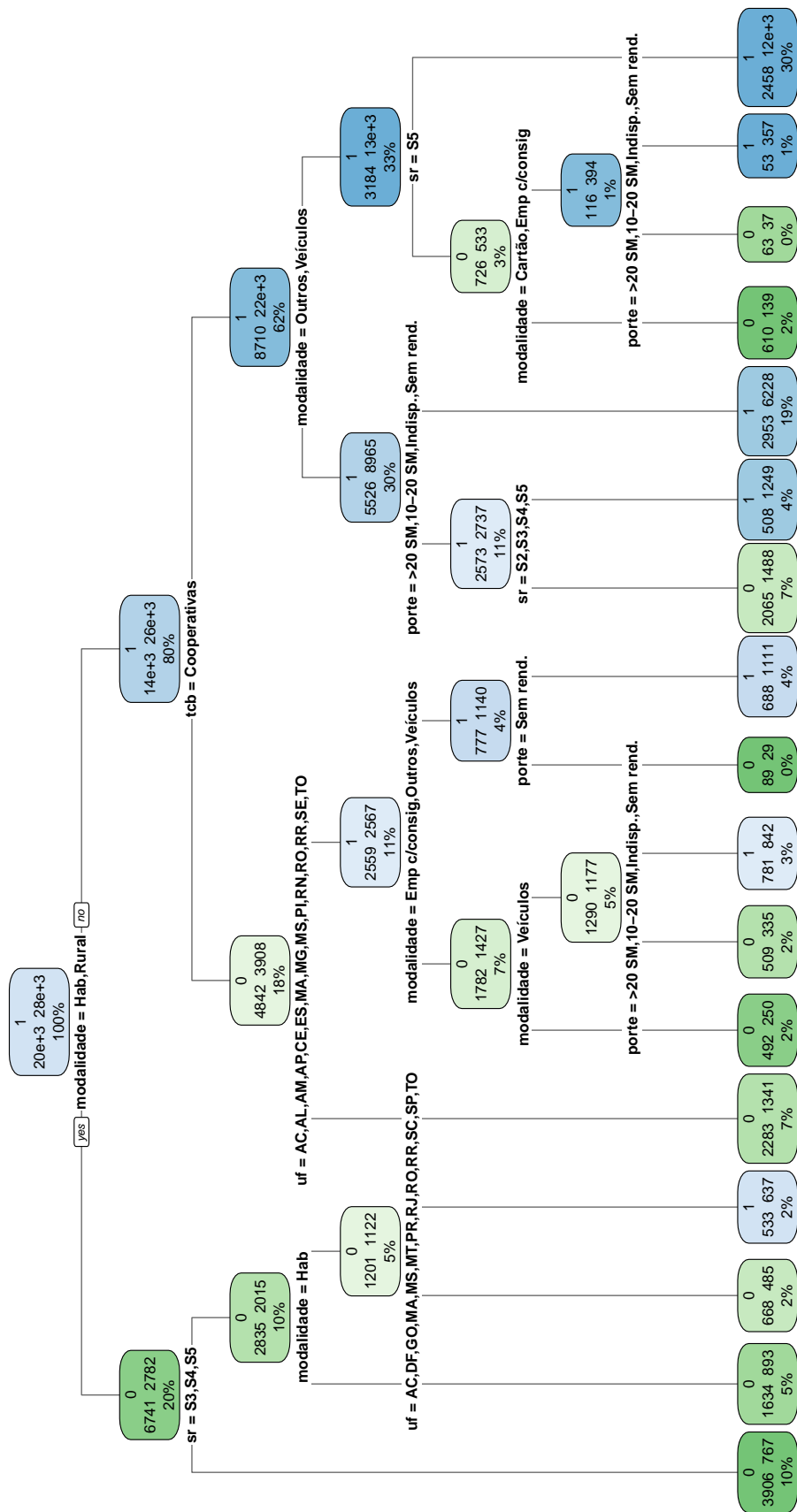
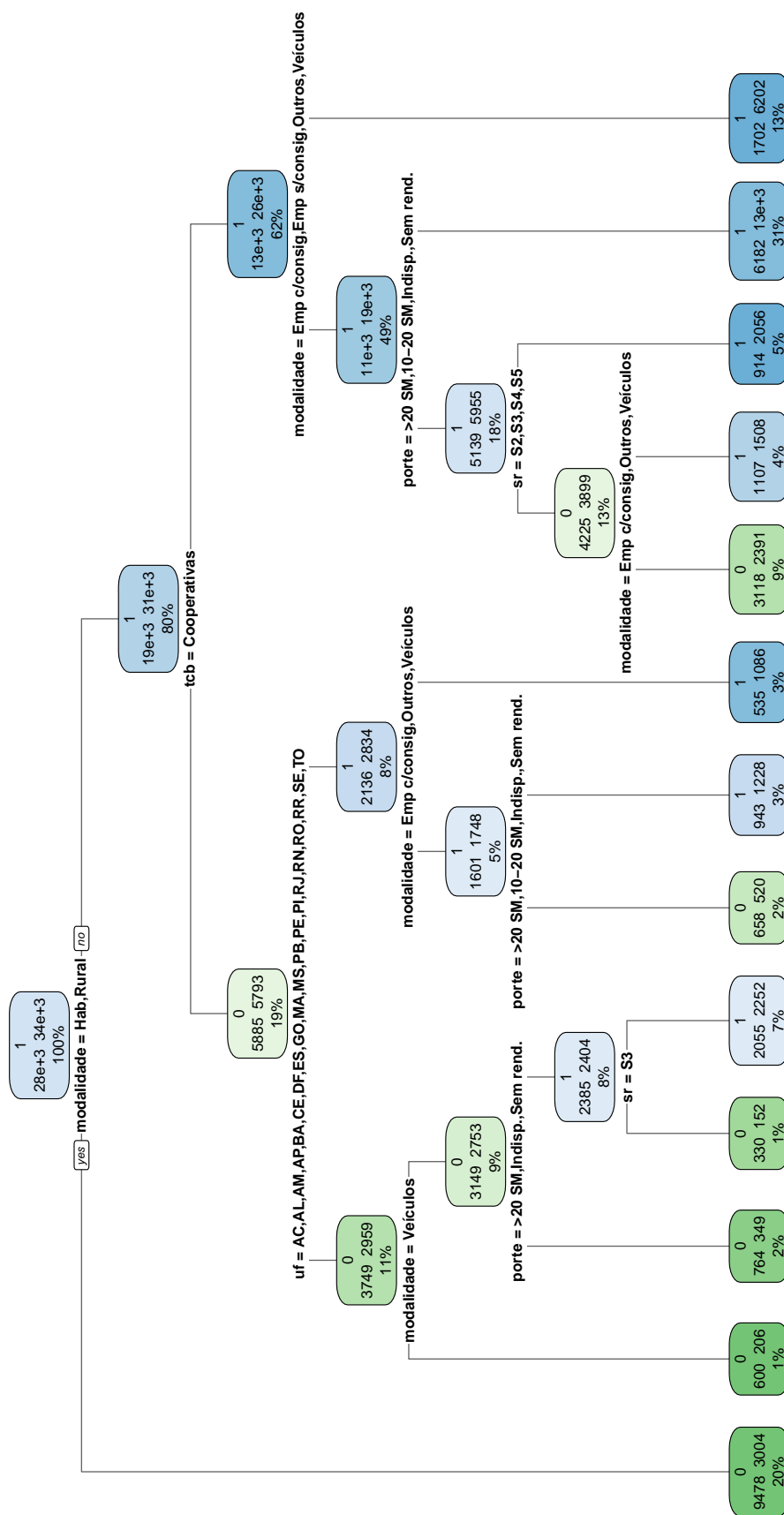


Figura 7: Árvore de decisão (CART), 2020.
Fonte: elaboração própria com dados do SCR.



Discussão

Os resultados apontam três regularidades, o pico de inadimplência em 2020 e normalização em 2024; uma forte heterogeneidade por modalidade, perfil socioeconômico (renda/ocupação) e UF; e diferenças entre tipos de consolidado bancário (TCB). À luz do arcabouço de intermediação financeira, sob assimetria de informação, desenho contratual e colateral condicionam risco e preço do crédito, de modo que choques agregados (como a Covid-19) operam por canais de juros, crédito, balanços e expectativas, afetando a inadimplência (5). As definições operacionais (carteira, atraso >90 dias, agregações por modalidade, ocupação e UF) seguem o SCR e sua metodologia oficial (1, 2); o pano de fundo regulatório e prudencial é dado pela Resolução nº 3.721/2009 e pela Resolução CMN nº 4.908/2021 a qual autoriza a prorrogação de parcelas dos financiamentos realizados com recursos do Fundo Constitucional de Financiamento do Norte (FNO), do Fundo Constitucional de Financiamento do Nordeste (FNE) e do Fundo Constitucional de Financiamento do Centro-Oeste (FCO) para atender os setores e atividades mais afetados pela pandemia da Covid-19. (3, 4), tornando mais difícil identificar contratos inadimplentes devido a sua prorrogação.

Nos heatmaps (Figuras 1, 2, 3, 4, 5) e, sobretudo, nas árvores de decisão de 2019, 2020 e 2024 (6, 7, 8), emerge uma segmentação consistente: habitação e rural compõem os ramos de menor inadimplência; ramos com cartão e crédito pessoal (sem consignação) combinados a baixa renda e a determinadas UFs concentram maiores taxas; em 2020, ramos ligados a ocupações mais expostas (p. ex., MEI e empregados do setor privado) ganham relevância; em 2024, há redução generalizada das taxas em diversos ramos, embora cartão permaneça crítico. As diferenças por TCB mostram patamares mais elevados no segmento bancário (com queda em 2024), estabilidade no não bancário e níveis historicamente menores em cooperativas, ainda que com aumento recente sinais compatíveis com composição de portfólio e práticas institucionais (4). Esse conjunto de evidências dialoga com a literatura de Mishkin (5), segundo a qual instituições e contratos modulam a transmissão de choques para a economia real (renda e emprego), refletindo-se na dinâmica de inadimplência.

Os padrões observados são coerentes com quatro mecanismos interligados: primeiro, a composição de portfólio, na qual maior peso relativo de cartão e crédito pessoal sem consignação produtos de menor colateral e mais sensíveis à renda eleva a inadimplência média dos grupos correspondentes (5); segundo, colateral e desenho contratual, pois habitação e rural tendem a apresentar menor inadimplência, em linha com garantias e subsídios/arranjos de pagamento mais estáveis (2, 5); terceiro, mercado de trabalho e renda, já que o pico de 2020 recaiu sobretudo sobre ocupações mais expostas ao ciclo (MEI e empregados privados), com preservação relativa de servidores e aposentados (5); e, por fim, o ambiente prudencial e as práticas de renegociação, em que o enquadramento normativo e a supervisão (Res. 3.721/2009; Res. CMN 4.908/2021) condicionam buffers de capital, monitoramento e incentivos de renegociação, contribuindo para a normalização observada em 2024 (3, 4).

As limitações do estudo envolvem cinco aspectos principais: primeiro, o desenho descritivo, pois embora as árvores (CART) revelem interações, elas não estabelecem causalidade e os resultados podem refletir composição e variáveis não observadas (5); segundo, agregação e ponderação, uma vez que os indicadores por células não ponderam a exposição (saldo/EAD), de modo que os achados podem se alterar com ponderação ou estratificação por valor (2); terceiro, a janela temporal, dado que três cortes (2019, 2020 e 2024)

não capturam dinâmicas intra-anuais nem o intervalo de 2021-2023; quarto, a definição de inadimplência, em que o corte >90 dias segue o SCR, sendo desejável conduzir análises de robustez com 30/60/120 dias, que podem modificar as comparações (2); e por fim, a limitação de informações socioeconômicas dos dados, isso é, informações sobre profissão, escolaridade, idade, estado civil e outras que pudessem nos dar um perfil mais completo sobre a inadimplência no país.

Pesquisas futuras

Como agenda de aprofundamento, propomos: modelagem inferencial, estimando regressões com interações ano \times modalidade/renda/UF/TCB, modelos hierárquicos com efeitos por UF/instituição e controles macro (desemprego, Selic, inflação) (5, 2); análise da dinâmica de transição, via modelos de sobrevivência e matrizes de migração (15→30→60→90 dias), testes de robustez e mensuração, ponderando por saldo/EAD, avaliando reestruturações e estratificando por garantias/prazos (2); e, por fim, aprendizado de máquina explicável, comparando CART a modelos preditivos (logística, Random Forest, GBM) e interpretando-os por importância de variáveis/SHAP nos períodos pré/-durante/pós (5).

Conclusão

Este estudo analisou o perfil da inadimplência de pessoas físicas no Brasil em três marcos temporais: pré-pandemia (2019), durante a pandemia (2020) e pós-pandemia (2024) a partir de microdados agregados do SCR (1, 2) e do enquadramento regulatório pertinente, interpretando os achados à luz do arcabouço de intermediação financeira de Mishkin (5). Observou-se um ciclo com pico em 2020 e normalização em 2024, coerente com a transmissão de choques por juros, crédito, balanços e expectativas.

A heterogeneidade por modalidade é marcada pelo cartão e crédito pessoal sem consignação onde se concentram maiores proporções de inadimplentes em todos os períodos, enquanto habitação e rural exibem níveis persistentemente menores.

Verificou-se um gradiente socioeconômico consistente pois faixas de menor renda e ocupações mais expostas ao ciclo (MEI e empregado privado) apresentam patamares superiores e maior reação no pico; servidores públicos e aposentados/pensionistas permanecem relativamente mais protegidos.

No plano territorial, estados do Norte e Nordeste tendem a níveis mais altos, com queda disseminada em 2024. No recorte institucional (TCB), o segmento bancário parte de níveis mais elevados, mas recua de forma expressiva em 2024; cooperativas mantêm os menores patamares, embora com alta recente, sugerindo efeitos de composição de carteira e práticas institucionais.

As árvores de decisão reforçam esses padrões, evidenciando interações entre produto, renda, UF e TCB e mostrando que condições contratuais (colateral, prazos) e características socioeconômicas são determinantes conjuntos do risco, em linha com a teoria e com a metodologia oficial do SCR.

Os resultados sinalizam priorizar gestão de risco por produto (cartão e pessoal sem consignação), segmentação socioeconômica nas políticas de concessão/cobrança e ajustes territoriais por UF. Do ponto de vista prudencial, reforçam a importância do uso de modelos de machine learning para reduzir a inadimplência antes (Credit Score) e durante

(Behavior Score) a concessão do crédito, bem como modelos para a recuperação de crédito (Collection Score), bem como o uso de testes de estresse e boas práticas de renegociação, coerentes com o arcabouço vigente.

Em conjunto, os achados reforçam que instituições e contratos importam para a propagação de choques e para a estabilidade do crédito, e que a combinação de evidências descritivas (heatmaps) e segmentação multivariada (árvores de decisão) oferecem um quadro operativo para monitorar risco e orientar intervenções, em consonância com a teoria e com o marco regulatório.

Referências

- [1] BANCO CENTRAL DO BRASIL. Sistema de Informações de Créditos SCR. Disponível em: https://dadosabertos.bcb.gov.br/dataset/scr_data. Acesso em: 4 ago. 2025.
- [2] BANCO CENTRAL DO BRASIL. Metodologia. Disponível em: https://www.bcb.gov.br/content/estabilidadefinanceira/scr/SCR/scr_data_metodologia.pdf. Acesso em: 4 ago. 2025.
- [3] Banco Central do Brasil (2009, 30 abril). Resolução N^o 3721. Disponível em: https://www.bcb.gov.br/pre/normativos/res/2009/pdf/res_3721_v1_0.pdf. Acesso em: 12 set. 2025.
- [4] Banco Central do Brasil Resolução CMN N^o 4.908 de 29/4/2021. Disponível em: <https://www.bcb.gov.br/estabilidadefinanceira/exibenormativo?tipo=Resolu%C3%A7%C3%A3o%20CMN&numero=4908>. Acesso em: 12 set. 2025.
- [5] MISHKIN, Frederic S. The Economics of Money, Banking and Financial Markets. 4th Canadian ed.
- [6] Breiman, L.; Friedman, J.; Olshen, R.; Stone, C. (1984). Classification and Regression Trees. Chapman Hall/CRC.

Apêndice

Nesta seção trazemos os códigos utilizados em linguagem de programação R durante o desenvolvimento desta pesquisa.

A.1 Preparo e padronização dos dados

Listing 1: Leitura, filtros e criação de indicadores

```
1 library(readr)
2 library(dplyr)
3
4 # Caminho CORRETO (com S em Bases_Dados)
5 caminho <- ""
6
7 # (opcional) conferir se a pasta existe e listar CSVs
8 stopifnot(dir.exists(caminho))
9 print(list.files(caminho, pattern = "\\\\.csv$", ignore.case =
    TRUE))
10
11 # Ler os arquivos (CSV padrão BR: separador ; e decimal ,)
12 df_2019 <- read_csv2(file.path(caminho, "planilha_201904.csv"),
13                       locale = locale(decimal_mark = ",",
14                                         grouping_mark = "."))
15 df_2020 <- read_csv2(file.path(caminho, "planilha_202004.csv"),
16                       locale = locale(decimal_mark = ",",
17                                         grouping_mark = "."))
18 df_2024 <- read_csv2(file.path(caminho, "planilha_202404.csv"),
19                       locale = locale(decimal_mark = ",",
20                                         grouping_mark = "."))
21
22 ocupacoes_validas <- c(
23   "PF - Servidor ou empregado público",
24   "PF - Empregado de empresa privada",
25   "PF - Aposentado/pensionista",
26   "PF - Autônomo",
27   "PF - MEI"
28 )
29
30 # Filtrar cada dataframe
31 df_2019_filtrado <- df_2019 %>%
32   filter(cliente == "PF", ocupacao %in% ocupacoes_validas)
33
34 df_2020_filtrado <- df_2020 %>%
35   filter(cliente == "PF", ocupacao %in% ocupacoes_validas)
36
37 df_2024_filtrado <- df_2024 %>%
38   filter(cliente == "PF", ocupacao %in% ocupacoes_validas)
```

```

38     mutate(taxa_inadimplencia_contrato =
39           carteira_inadimplida_arrastada / carteira_ativa)
40 df_2020_filtrado <- df_2020_filtrado %>%
41   mutate(taxa_inadimplencia_contrato =
42         carteira_inadimplida_arrastada / carteira_ativa)
43 df_2024_filtrado <- df_2024_filtrado %>%
44   mutate(taxa_inadimplencia_contrato =
45         carteira_inadimplida_arrastada / carteira_ativa)
46 df_2019_filtrado <- df_2019_filtrado %>%
47   mutate(
48     inadimplencia = carteira_inadimplida_arrastada /
49       carteira_ativa,
50     inadimplente = if_else(carteira_inadimplida_arrastada > 0,
51       1, 0)
52   )
53 df_2020_filtrado <- df_2020_filtrado %>%
54   mutate(
55     inadimplencia = carteira_inadimplida_arrastada /
56       carteira_ativa,
57     inadimplente = if_else(carteira_inadimplida_arrastada > 0,
58       1, 0)
59   )
60 df_2024_filtrado <- df_2024_filtrado %>%
61   mutate(
62     inadimplencia = carteira_inadimplida_arrastada /
63       carteira_ativa,
64     inadimplente = if_else(carteira_inadimplida_arrastada > 0,
65       1, 0)
66   )
67 df_2019_filtrado <- df_2019_filtrado %>%
68   mutate(
69     modalidade = recode(modalidade,
70       "PF - Cartão de crédito" = "Cartão",
71       "PF - Empréstimo sem consignação em folha" = "Emp
72         s/consig",
73       "PF - Empréstimo com consignação em folha" = "Emp
74         c/consig",
75       "PF - Outros créditos" = "Outros",
76       "PF - Habitacional" = "Hab",
77       "PF - Veículos" = "Veículos",
78       "PF - Rural e agroindustrial" = "Rural"
79     ),
80     ocupacao = recode(ocupacao,
81       "PF - Empregado de empresa privada" = "Emp priv",
82       "PF - Servidor ou empregado público" = "Serv púb",

```

```

78     "PF - Aposentado/pensionista" = "Apos/Pens",
79     "PF - Autônomo" = "Autônomo",
80     "PF - MEI" = "MEI"
81   ),
82   porte = recode(porte,
83     "PF - Até 1 salário mínimo" = "1 SM",
84     "PF - Mais de 1 a 5 salários mínimos" = "1-5 SM",
85     "PF - Mais de 5 a 10 salários mínimos" = "5-10 SM",
86     "PF - Mais de 10 a 20 salários mínimos" = "10-20 SM",
87     "PF - Acima de 20 salários mínimos" = ">20 SM",
88     "PF - Indisponível" = "Indisp.",
89     "PF - Sem rendimento" = "Sem rend."
90   )
91 )
92
93 df_2020_filtrado <- df_2020_filtrado %>%
94   mutate(
95     modalidade = recode(modalidade,
96       "PF - Cartão de crédito" = "Cartão",
97       "PF - Empréstimo sem consignação em folha" = "Emp
98         s/consig",
99       "PF - Empréstimo com consignação em folha" = "Emp
100         c/consig",
101       "PF - Outros créditos" = "Outros",
102       "PF - Habitacional" = "Hab",
103       "PF - Veículos" = "Veículos",
104       "PF - Rural e agroindustrial" = "Rural"
105     ),
106     ocupacao = recode(ocupacao,
107       "PF - Empregado de empresa privada" = "Emp priv",
108       "PF - Servidor ou empregado público" = "Serv púb",
109       "PF - Aposentado/pensionista" = "Apos/Pens",
110       "PF - Autônomo" = "Autônomo",
111       "PF - MEI" = "MEI"
112     ),
113     porte = recode(porte,
114       "PF - Até 1 salário mínimo" = "1 SM",
115       "PF - Mais de 1 a 5 salários mínimos" = "1-5 SM",
116       "PF - Mais de 5 a 10 salários mínimos" = "5-10 SM",
117       "PF - Mais de 10 a 20 salários mínimos" = "10-20 SM",
118       "PF - Acima de 20 salários mínimos" = ">20 SM",
119       "PF - Indisponível" = "Indisp.",
120       "PF - Sem rendimento" = "Sem rend."
121     )
122   )
123
124 df_2024_filtrado <- df_2024_filtrado %>%
125   mutate(
126     modalidade = recode(modalidade,
127       "PF - Cartão de crédito" = "Cartão",

```



```

126     "PF - Empréstimo sem consignação em folha" = "Emp
127         s/consig",
128     "PF - Empréstimo com consignação em folha" = "Emp
129         c/consig",
130     "PF - Outros créditos" = "Outros",
131     "PF - Habitacional" = "Hab",
132     "PF - Veículos" = "Veículos",
133     "PF - Rural e agroindustrial" = "Rural"
134 ),
135 ocupacao = recode(ocupacao,
136     "PF - Empregado de empresa privada" = "Emp priv",
137     "PF - Servidor ou empregado público" = "Serv púb",
138     "PF - Aposentado/pensionista" = "Apos/Pens",
139     "PF - Autônomo" = "Autônomo",
140     "PF - MEI" = "MEI"
141 ),
142 porte = recode(porte,
143     "PF - Até 1 salário mínimo" = "1 SM",
144     "PF - Mais de 1 a 5 salários mínimos" = "1-5 SM",
145     "PF - Mais de 5 a 10 salários mínimos" = "5-10 SM",
146     "PF - Mais de 10 a 20 salários mínimos" = "10-20 SM",
147     "PF - Acima de 20 salários mínimos" = ">20 SM",
148     "PF - Indisponível" = "Indisp.",
149     "PF - Sem rendimento" = "Sem rend."
150 )
151 )
152 df_todos <- bind_rows(
153     df_2019_filtrado %>% mutate(ano = 2019),
154     df_2020_filtrado %>% mutate(ano = 2020),
155     df_2024_filtrado %>% mutate(ano = 2024)
156 )
157 df_todos %>%
158     group_by(ano) %>%
159     summarise(
160         media_inad = mean(taxa_inadimplencia_contrato, na.rm = TRUE),
161         mediana_inad = median(taxa_inadimplencia_contrato, na.rm =
162             TRUE),
163         max_inad = max(taxa_inadimplencia_contrato, na.rm = TRUE),
164         .groups = "drop"
165     )

```

A.2 Visualizações (linhas e heatmaps)

Listing 2: Gráficos por modalidade, origem, ocupação, UF, porte/renda e TCB

```

1 library(dplyr)
2 library(ggplot2)
3 # opcional: para reordenar e lidar com labels
4 library(forcats)

```

```

5
6 # =====
7 # Função auxiliar: sumariza % inadimplentes por ano + dimensão
8 # =====
9 sumariza_perc <- function(.data, var){
10   .data %>%
11     group_by(ano, {{ var }}) %>%
12     summarise(perc_inadimplentes =
13       mean(inadimplente, na.rm = TRUE) * 100,
14       .groups = "drop")
15   }
16
17 # 1) Linha: evolução por modalidade
18 evol_modalidade <- sumariza_perc(df_todos, modalidade)
19 ggplot(evol_modalidade, aes(x = ano, y = perc_inadimplentes,
20   color = modalidade, group = modalidade)) +
21   geom_line(size = 1.2) +
22   geom_point(size = 2) +
23   labs(title = "Evolução da Inadimplência por Modalidade",
24    x = "Ano", y = "% Inadimplentes") +
25   theme_minimal() +
26   theme(legend.position = "bottom")
27
28 # 1b) Linha: evolução por origem
29 evol_origem <- sumariza_perc(df_todos, origem)
30 ggplot(evol_origem, aes(x = ano, y = perc_inadimplentes,
31   color = origem, group = origem)) +
32   geom_line(size = 1.2) +
33   geom_point(size = 2) +
34   labs(title = "Evolução da Inadimplência por Origem",
35    x = "Ano", y = "% Inadimplentes") +
36   theme_minimal() +
37   theme(legend.position = "bottom")
38
39 # 2) Heatmap: Ocupação E Ano
40 heat_ocup <- sumariza_perc(df_todos, ocupacao)
41 ggplot(heat_ocup, aes(x = factor(ano), y = ocupacao, fill =
42   perc_inadimplentes)) +
43   geom_tile(color = "white") +
44   geom_text(aes(label = sprintf("%.1f", perc_inadimplentes)),
45    color = "black", size = 3) +
46   scale_fill_gradient(low = "white", high = "red") +
47   labs(title = "Heatmap da Inadimplência por Ocupação e Ano",
48    x = "Ano", y = "Ocupação", fill = "% Inadimplentes") +
49   theme_minimal()
50
51 # 3) Heatmap: Modalidade E Ano
52 heat_modalidade <- evol_modalidade
53 ggplot(heat_modalidade, aes(x = factor(ano), y = modalidade,
54   fill = perc_inadimplentes)) +
55   geom_tile(color = "white") +

```

```

54 geom_text(aes(label = sprintf("%.1f", perc_inadimplentes)),
55           color = "black", size = 3) +
56 scale_fill_gradient(low = "white", high = "blue") +
57 labs(title = "Heatmap da Inadimplência por Modalidade e Ano",
58       x = "Ano", y = "Modalidade", fill = "% Inadimplentes") +
59 theme_minimal()
60
61 # 4) Heatmap: UF & Ano
62 heat_uf <- sumariza_perc(df_todos, uf)
63 ordem_uf <- heat_uf %>% group_by(uf) %>% summarise(m =
        mean(perc_inadimplentes, na.rm = TRUE),
        .groups = "drop") %>%
64   arrange(desc(m)) %>% pull(uf)
65 heat_uf <- heat_uf %>% mutate(uf = factor(uf, levels =
        ordem_uf))
66 ggplot(heat_uf, aes(x = factor(ano), y = uf, fill =
        perc_inadimplentes)) +
67   geom_tile(color = "white") +
68   geom_text(aes(label = sprintf("%.1f", perc_inadimplentes)),
        color = "black", size = 2.7) +
69   scale_fill_gradient(low = "white", high = "purple") +
70   labs(title = "Heatmap da Inadimplência por UF e Ano",
71        x = "Ano", y = "UF", fill = "% Inadimplentes") +
72   theme_minimal()
73
74 # 5) Heatmap: Porte/Renda & Ano
75 heat_renda <- sumariza_perc(df_todos, porte)
76 ordem_renda <- heat_renda %>% group_by(porte) %>% summarise(m =
        mean(perc_inadimplentes, na.rm = TRUE), .groups =
        "drop") %>%
77   arrange(desc(m)) %>% pull(porte)
78 heat_renda <- heat_renda %>% mutate(porte = factor(porte,
        levels = ordem_renda))
79 ggplot(heat_renda, aes(x = factor(ano), y = porte, fill =
        perc_inadimplentes)) +
80   geom_tile(color = "white") +
81   geom_text(aes(label = sprintf("%.1f", perc_inadimplentes)),
        color = "black", size = 3) +
82   scale_fill_gradient(low = "white", high = "darkgreen") +
83   labs(title = "Heatmap da Inadimplência por Porte/Renda e
        Ano",
84        x = "Ano", y = "Porte/Renda", fill = "% Inadimplentes") +
85   theme_minimal()
86
87 # 6) Heatmap: TCB & Ano
88 # Se a coluna vier como "TCB", use: df_todos <- df_todos %>%
        rename(tcb = TCB)
89 heat_tcb <- sumariza_perc(df_todos, tcb)
90 ordem_tcb <- heat_tcb %>% group_by(tcb) %>% summarise(m =
        mean(perc_inadimplentes, na.rm = TRUE), .groups =
        "drop") %>%

```

```

91     arrange(desc(m)) %>% pull(tcb)
92     heat_tcb <- heat_tcb %>% mutate(tcb = factor(tcb, levels =
      ordem_tcb))
93     ggplot(heat_tcb, aes(x = factor(ano), y = tcb, fill =
      perc_inadimplentes)) +
94     geom_tile(color = "white") +
95     geom_text(aes(label = sprintf("%.1f", perc_inadimplentes)),
      color = "black", size = 3) +
96     scale_fill_gradient(low = "white", high = "pink") +
97     labs(title = "Heatmap da Inadimplência por TCB e Ano",
98           x = "Ano", y = "TCB", fill = "% Inadimplentes") +
99     theme_minimal()

```

A.3 Árvores de decisão (CART)

Listing 3: Modelagem CART e podas por ano

```

1  library(rpart)
2  library(rpart.plot)
3
4  # --- 2019
5  set.seed(123)
6  idx <- sample(seq_len(nrow(df_2019_filtrado)), size =
      0.7*nrow(df_2019_filtrado))
7  treino_2019 <- df_2019_filtrado[idx, ]
8  teste_2019  <- df_2019_filtrado[-idx, ]
9
10 arvore_2019 <- rpart(inadimplente ~ ocupacao + modalidade + uf +
      porte + tcb + sr,
11     data = treino_2019,
12     method = "class",
13     control = rpart.control(cp = 0.001, minsplit = 30, maxdepth
      = 6))
14 printcp(arvore_2019)
15 best_cp <-
      arvore_2019$cptable[which.min(arvore_2019$cptable[, "xerror"]),
      "CP"]
16 arvore_podada_2019 <- prune(arvore_2019, cp = best_cp)
17 rpart.plot(arvore_podada_2019, main = "Árvore Podada - 2019",
      type = 3, extra = 104, under = TRUE)
18 pred_2019 <- predict(arvore_podada_2019, newdata = teste_2019,
      type = "class")
19 arvore_2019$variable.importance
20
21 # --- 2020
22 set.seed(123)
23 idx <- sample(seq_len(nrow(df_2020_filtrado)), size =
      0.7*nrow(df_2020_filtrado))
24 treino_2020 <- df_2020_filtrado[idx, ]
25 teste_2020  <- df_2020_filtrado[-idx, ]
26

```

```

27 arvore_2020 <- rpart(inadimplente ~ ocupacao + modalidade + uf +
    porte + tcb + sr,
28     data = treino_2020,
29     method = "class",
30     control = rpart.control(cp = 0.001, minsplit = 30, maxdepth
        = 6))
31 printcp(arvore_2020)
32 best_cp <-
    arvore_2020$cptable[which.min(arvore_2020$cptable[, "xerror"]),
        "CP"]
33 arvore_podada_2020 <- prune(arvore_2020, cp = best_cp)
34 rpart.plot(arvore_podada_2020, main = "Árvore Podada - 2020",
    type = 3, extra = 104, under = TRUE)
35 pred_2020 <- predict(arvore_podada_2020, newdata = teste_2020,
    type = "class")
36 arvore_2020$variable.importance
37
38 # --- 2024
39 set.seed(123)
40 idx <- sample(seq_len(nrow(df_2024_filtrado)), size =
    0.7*nrow(df_2024_filtrado))
41 treino_2024 <- df_2024_filtrado[idx, ]
42 teste_2024 <- df_2024_filtrado[-idx, ]
43
44 arvore_2024 <- rpart(inadimplente ~ ocupacao + modalidade + uf +
    porte + tcb + sr,
45     data = treino_2024,
46     method = "class",
47     control = rpart.control(cp = 0.001, minsplit = 30, maxdepth
        = 6))
48 printcp(arvore_2024)
49 best_cp <-
    arvore_2024$cptable[which.min(arvore_2024$cptable[, "xerror"]),
        "CP"]
50 arvore_podada_2024 <- prune(arvore_2024, cp = best_cp)
51 rpart.plot(arvore_podada_2024, main = "Árvore Podada - 2024",
    type = 3, extra = 104, under = TRUE)
52 pred_2024 <- predict(arvore_podada_2024, newdata = teste_2024,
    type = "class")
53 arvore_2024$variable.importance

```