

Olá, Prezados.

Venho por meio deste relatório trazer os principais pontos encontrados no estudo solicitado pelo setor de crédito. Aqui irei relatar a metodologia utilizada, principais insights, e recomendações a partir das conclusões.

METODOLOGIA

Para este projeto utilizei a metodologia CRISPDM, onde o problema foi quebrado em algumas etapas, e o início da próxima etapa está ligado algum tipo de conclusão da etapa anterior. Segui as seguintes etapas:

- 1 - Entendimento de Negócio - Definição dos objetivos chaves do estudo para satisfazer o problema de negócio.
- 2 - Entendimento dos Dados - Uma primeira compreensão dos dados, quais tratamentos necessários, e quais utilizar ou não.
- 3 - Tratamento dos Dados - Preparação da base para o estudo.
- 4 - Desenvolvimento - Desenvolvimento do estudo e extração de conclusões.

ENTENDIMENTO DE NEGÓCIO

Aqui tracei alguns objetivos e hipóteses para serem respondidas, conforme os possíveis problemas levantados pelo time de crédito.

OBJETIVOS

- 1) Avaliar a relação entre a política de crédito atual e o comportamento de inadimplência dos clientes, verificando se ela impede que clientes com alta probabilidade de inadimplência recebam crédito.
- 2) Identificar as principais variáveis que impactam o índice de inadimplência dos clientes e analisar como essas variáveis se relacionam com o modelo de score atual, visando propor melhorias na política de crédito.
- 3) Com base nas análises anteriores, avaliar a possibilidade de oferecer suporte aos clientes que tiveram crédito negado, visando reduzir a inadimplência.

ENTENDIMENTO DOS DADOS E TRATAMENTO

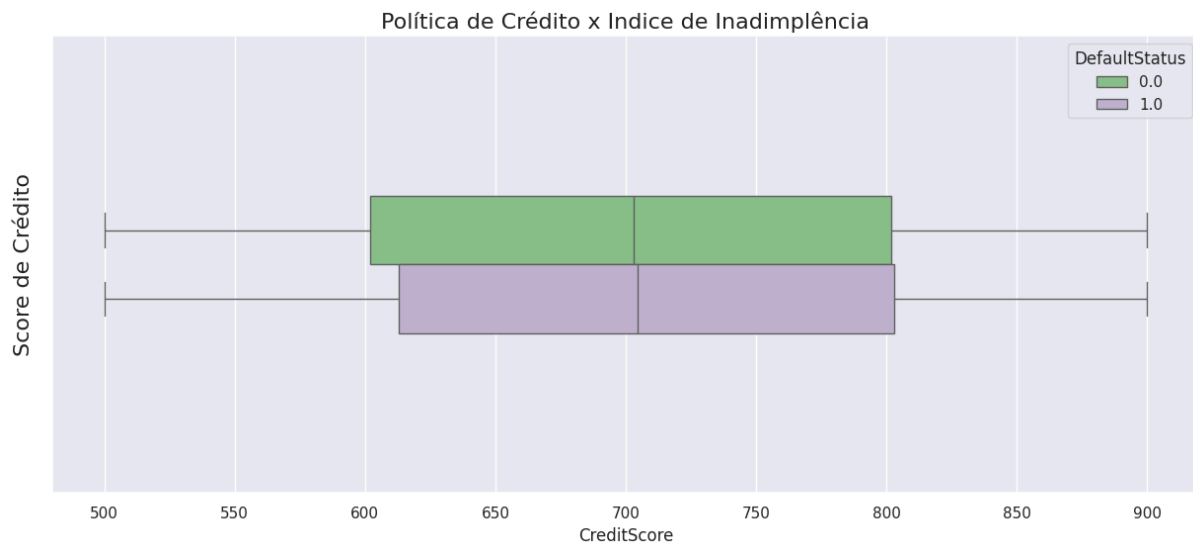
Antes de seguir com análise identifiquei alguns problemas nos dados que precisavam ser corrigidos. O primeiro foi o formato do dataset que estava em wide format e passei para long format, unindo a variável renda e receita, já que existe uma coluna para esse tipo de distinção de pessoa física e jurídica. Após esse ajuste fiz a verificação e tratamento de nulos, duplicados e zeros.

DESENVOLVIMENTO

Para essa etapa transformei os objetivos em hipóteses.

1) A POLÍTICA DE CRÉDITO ATUAL ESTÁ CORRELACIONADA COM O COMPORTAMENTO DE INADIMPLÊNCIA DOS CLIENTES E IMPEDE QUE CLIENTES COM ALTA PROBABILIDADE DE INADIMPLÊNCIA RECEBAM CRÉDITO?

Para responder essa problemática primeiramente busquei entender a proporção de clientes adimplentes (0 > 81.42%) e inadimplentes (1 > 18.58 %). Em seguida fiz um gráfico para entender se visualmente era possível ver uma diferença de score entre bons e maus pagadores.



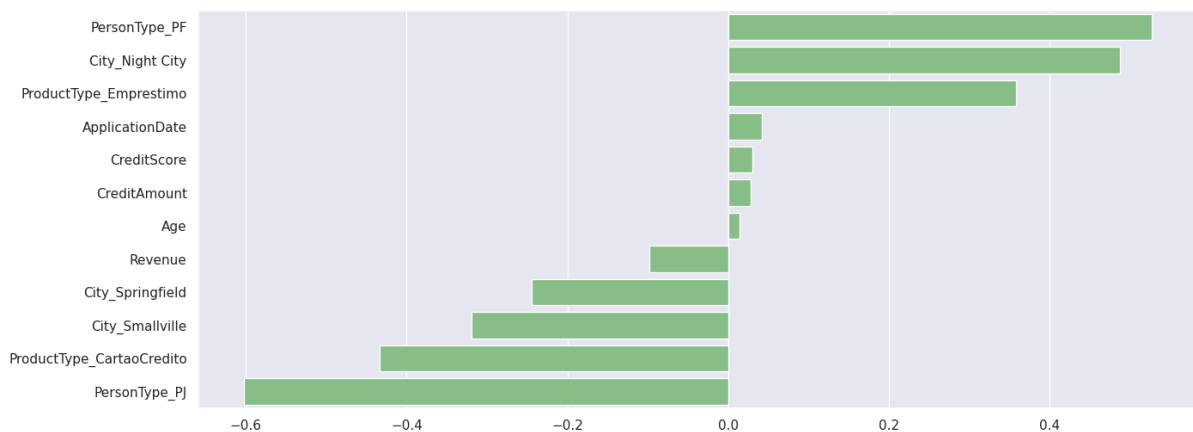
Como o comportamento ficou bem próximo nos dois tipos do cliente com a variável tendo uma dispersão bem parecida (tamanho das caixas) e a mediana (traço no meio da caixa) se encontra quase no mesmo ponto, ficou entendido que a política de crédito não tem correlação com comportamento de inadimplência, e não está impedindo que maus pagadores recebam crédito, assim como não está ajudando bons pagadores a receber crédito. Para confirmar essa conclusão, realizei um teste de hipóteses bicaudal com objetivo de afirmar estatisticamente se existe ou diferença entre nível de score de clientes adimplentes ou não.

Nesse teste usei um nível de confiança de 99%, e como resposta tive a confirmação de que não existe evidência estatística suficiente para afirmar que há uma diferença (e impacto) da política de crédito entre clientes adimplentes e inadimplentes.

2) QUAIS AS PRINCIPAIS VARIÁVEIS AJUDAM A IDENTIFICAR CLIENTES COM TENDÊNCIA À INADIMPLÊNCIA INADIMPLENTES? EXISTE ALGUMA RELAÇÃO DESSAS VARIÁVEIS COM O SCORE DE CRÉDITO?

A ideia aqui foi entender quais das variáveis explicativas disponibilizadas impactam na probabilidade do cliente ser inadimplente ou não, e ver se essas variáveis eram também as que mais impactam no score. para isso construí um modelo de regressão logística para entender a relação da inadimplência, e uma regressão linear para entender a relação do score.

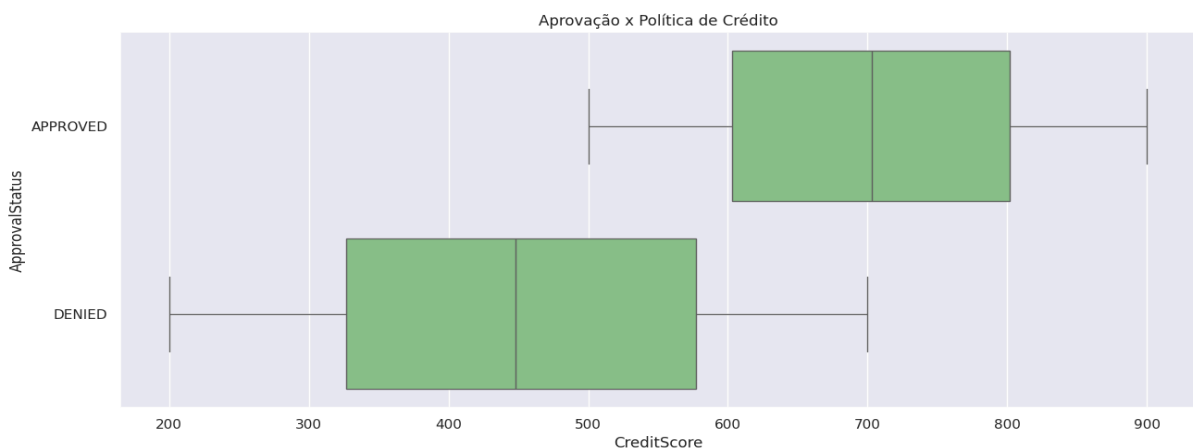
Após análise foi possível notar que essas eram as principais variáveis relacionadas com a tendência a inadimplência, sendo as positivas que mais impactam na tendência do cliente ser inadimplente, e as negativas as que mais impactam em ele ser adimplente.



Um ponto aqui interessante aqui é ver que o Score é quase nulo na tentativa de explicar a tendência a inadimplência, o que confirma a análise anterior. Ao fazer um modelo de regressão para entender se essas mesmas variáveis explicam o score, porém ao fazer a regressão com as mesmas variáveis, apenas tornando o score na variável dependente, vi que os valores eram capaz de explicar apenas 0.08% da variável dependente, ou seja, essas variáveis têm pouco impacto na política de crédito atual. Recomendo o time de crédito realinhar a política levando em consideração a força das variáveis explicativas de tendência de inadimplência.

3) É POSSÍVEL AUXILIAR CLIENTES QUE TIVERAM CRÉDITO NEGADO, COM BASE NAS DESCOBERTAS SOBRE A POLÍTICA DE CRÉDITO E OS INDICADORES DE INADIMPLÊNCIA?

Com as duas questões respondidas busquei entender se os clientes que tiveram crédito negado no último ano foram impactados pela política de crédito, para isso fiz um novo gráfico parecido com o da primeira análise.



Visualmente foi possível notar que clientes que tiveram crédito negado tem um score mais baixo, busquei confirmar fazendo um novo teste de hipóteses com confiança em 99%, e o efeito da política de crédito no aceite da solicitação foi comprovado estatisticamente. A partir daí decidi criar um modelo de classificação mais robusto, com objetivo de prever quais desses clientes poderiam passar por uma reanálise, ajudando o time de crédito aumentar a concessão, e também na satisfação dos clientes, desenvolvi o modelo e estou deixando em anexo essa base.