

Super Caja- Automatizando o Processo de Análise de Risco

Objetivo:

O aumento da demanda por crédito no banco "Super Caja" está sobrecarregando a equipe de análise, que utiliza um processo manual e ineficiente. Para otimizar a análise e reduzir o risco de inadimplência, propõe-se a automação do processo com técnicas de análise de dados. O objetivo é criar um score de crédito que classifique os solicitantes por risco de inadimplência, utilizando inclusive dados de pagamentos em atraso. A automação visa aumentar a eficiência, precisão e rapidez na avaliação de crédito, além de contribuir para a solidez financeira do banco.

Análise Exploratória

Score de Crédito

Buscador Super caja

Hipóteses

Modelos de Previsão

Pontos de Atenção



Análise exploratória

Quem são os clientes analisados nesta base do Super Caja

Em números

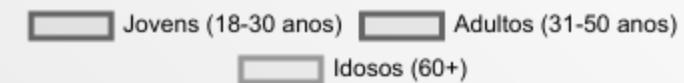
36.000

Quais informações valiosas temos no banco de dados

Quais aspectos priorizar para guiar novas soluções

taxa_de_endividamento

351,49



Atrasados a mais de 30 dias

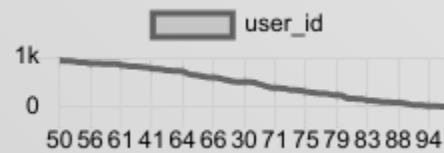
15.094

atraso_30_59_dias atraso_60_89_dias atraso_superior_90_dias

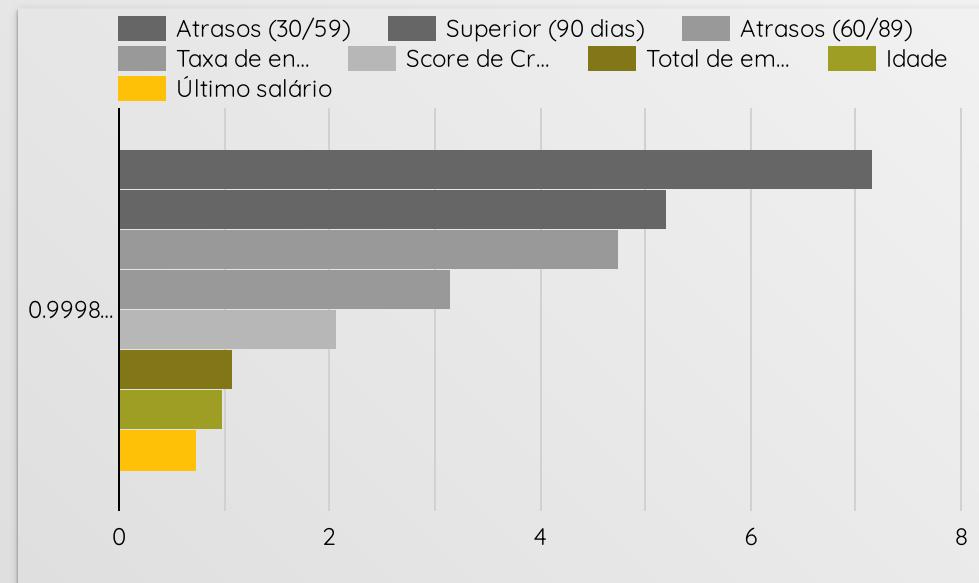
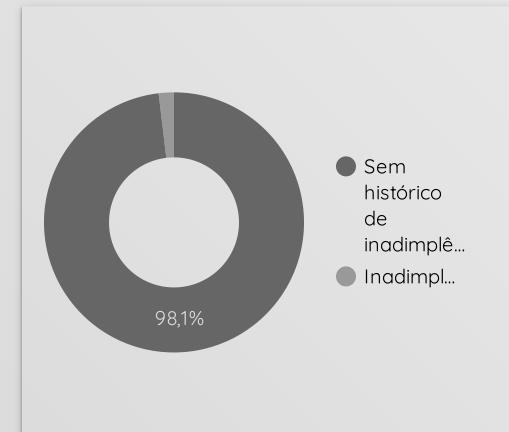
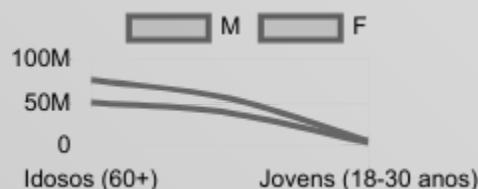
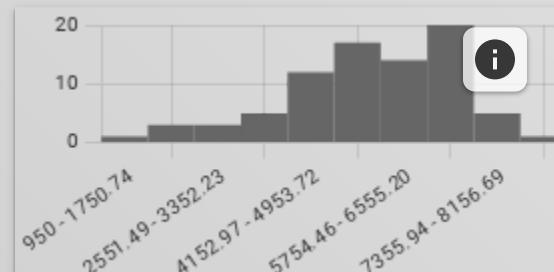
Informações faltantes encontradas:

Dependentes: 943 nulos (substituídos por 0).

Salário: 7199 nulos (mediana dos salários).



Os clientes em maior parte são adultos e idosos.

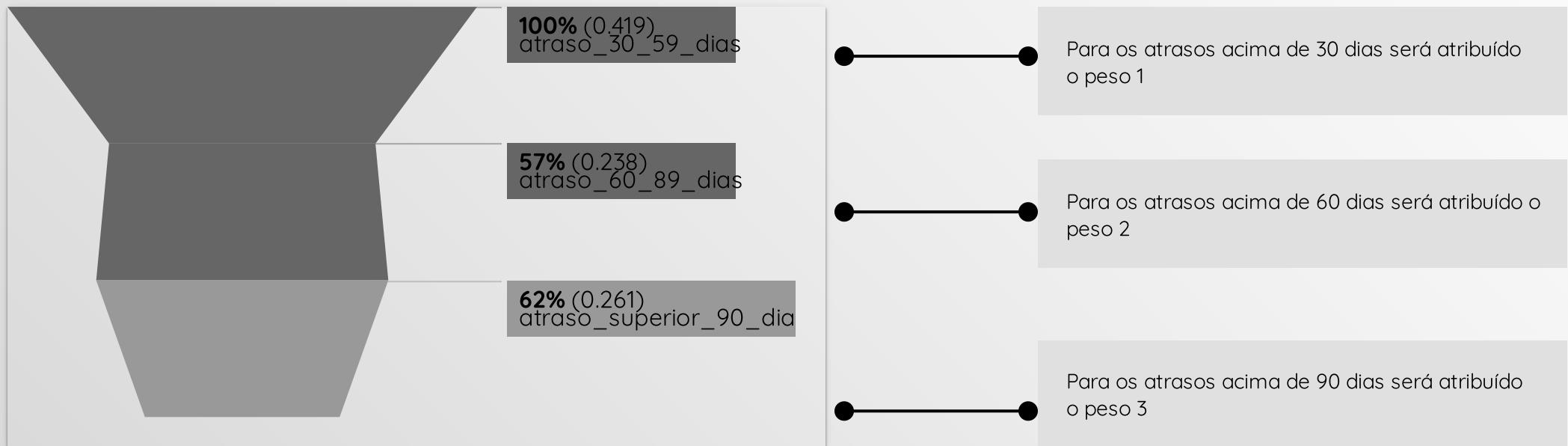


No histograma notamos a faixa salarial entre 4k e 6k. Mas, avaliando por gênero notamos que mulheres recebem menos.

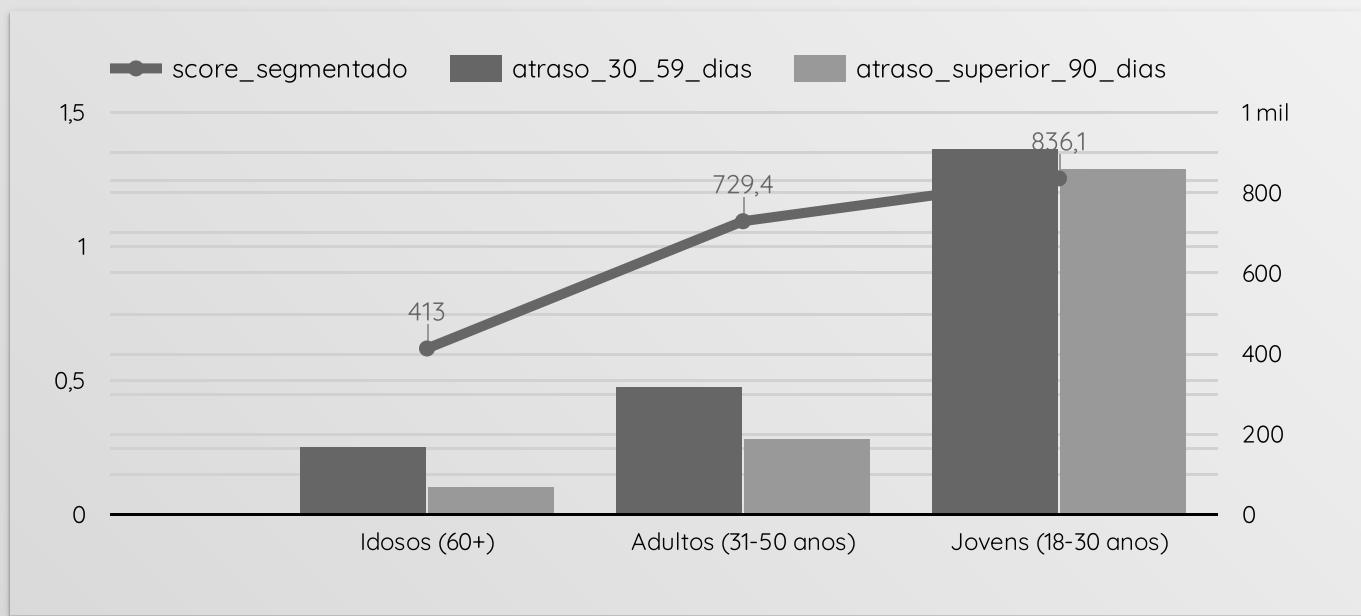
A consequência da diferença salarial ressoa na taxa de endividamento maior para mulheres.

Alguns aspectos são fundamentais para dar sequência na criação de um score adequado, os dias em atraso e a taxa de endividamento são variáveis fundamentais.

Definindo o Score de Crédito



O resultado dessas multiplicações é então somado a taxa de endividamento e o **Score de Crédito** é criado



Através desse score de crédito definimos o perfil de pagador por quartil para segmentar os perfis de alto ou baixo risco

genero

faixa_etaria

user_id

Insira um valor

Buscador Super-Caja

	user_id	faixa_etaria	classificacao_de_inadimplencia	perfil_pagador	Score de crédito
1.	36000	Jovens (18-30 anos)	Sem histórico de inadimplência	mau_pagador	3
2.	35999	Idosos (60+)	Sem histórico de inadimplência	mau_pagador	3
3.	35998	Idosos (60+)	Sem histórico de inadimplência	mau_pagador	1
4.	35997	Idosos (60+)	Sem histórico de inadimplência	bom_pagador	0
5.	35996	Idosos (60+)	Sem histórico de inadimplência	mau_pagador	4

1 - 100 / 36000 < >

taxa_de_endividamento

351,49

atraso_30_59_dias

15.094

total_de_emprestimo_por_id

8,48

atraso_60_89_dias

8.563

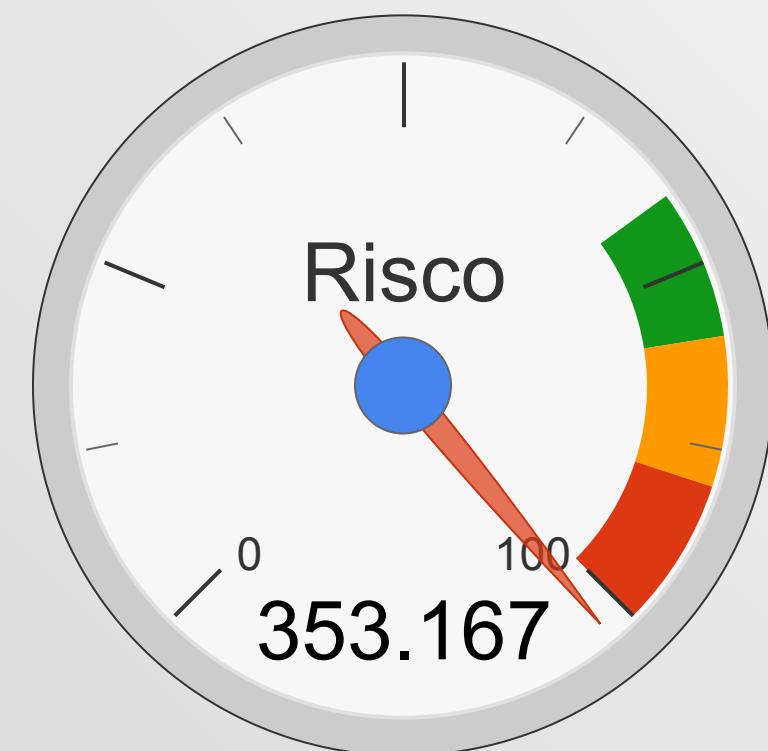
ultimo_salario_informado

231.122.774

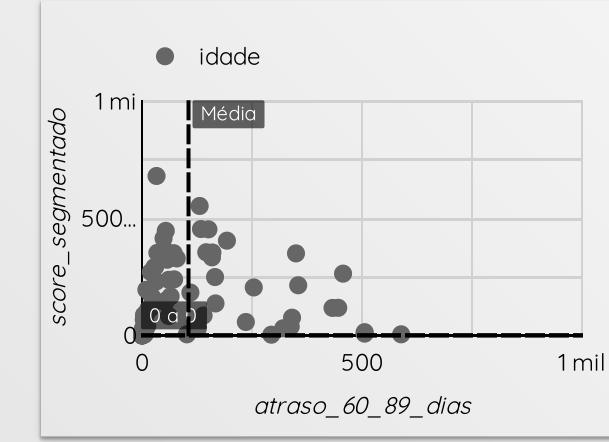
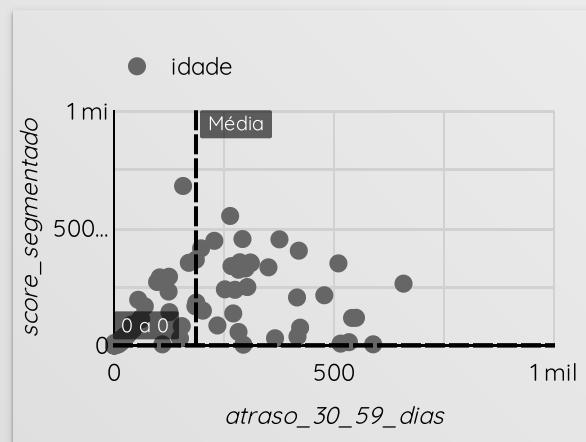
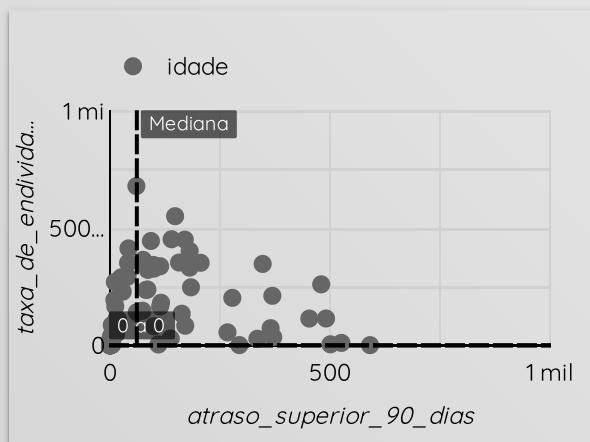
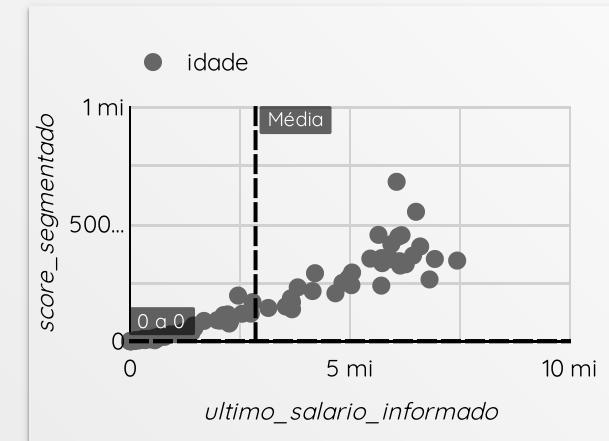
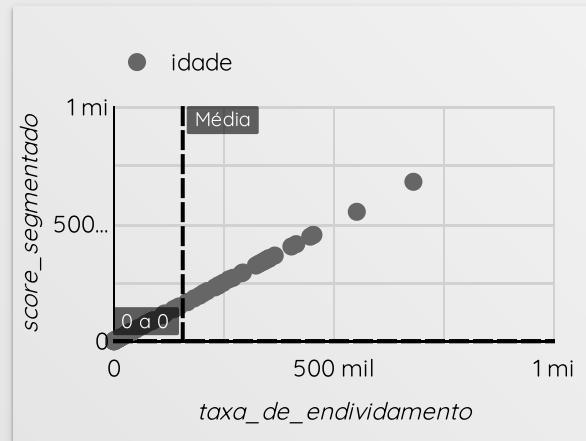
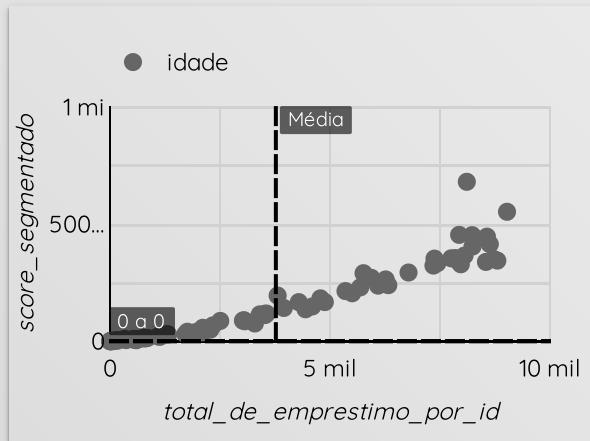
atraso_superior_90_dias

9.389

Risco



Hipóteses



A diferença nas médias é extrema, sugerindo que a taxa de endividamento é um forte indicador de risco. Atraso Superior a 90 Dias, Atraso de 30-59 Dias, Atraso de 60-89 Dias:

Esses atrasos são significativamente maiores em "maus pagadores", destacando a relevância de histórico de pagamentos. Score Segmentado: A enorme diferença na média indica que o score segmentado é um fator crucial para determinar o risco de inadimplência. Total de Empréstimos por ID: A diferença é menor, indicando que essa variável tem um impacto menos dramático no risco relativo comparado às outras.

Conclusões Variáveis com Alto Risco Relativo: Taxa de Endividamento, Atrasos e Score Segmentado são críticos para prever o risco de inadimplência. Indicadores Importantes: Variáveis relacionadas ao histórico de crédito e comportamento de pagamento são fortes indicadores.

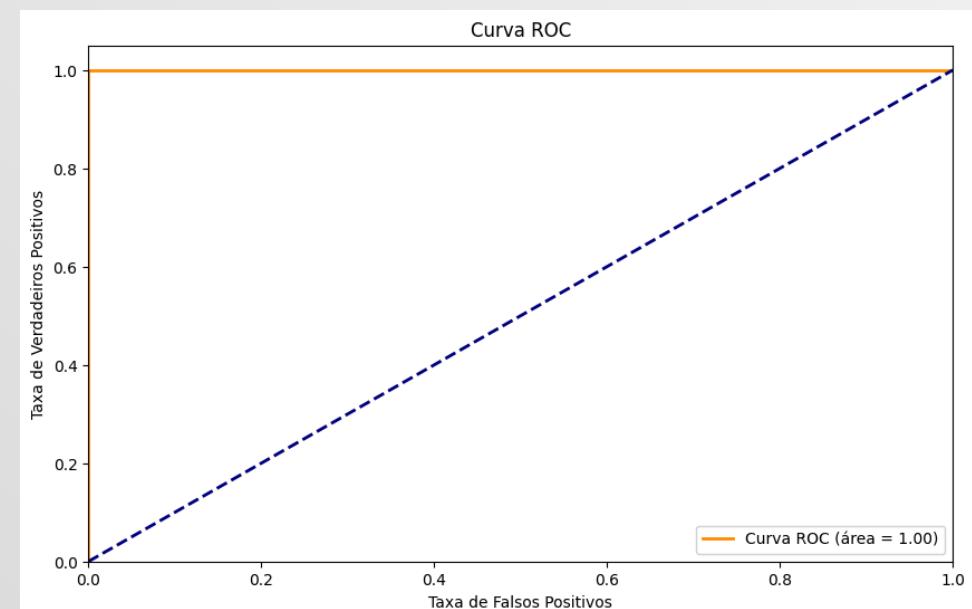
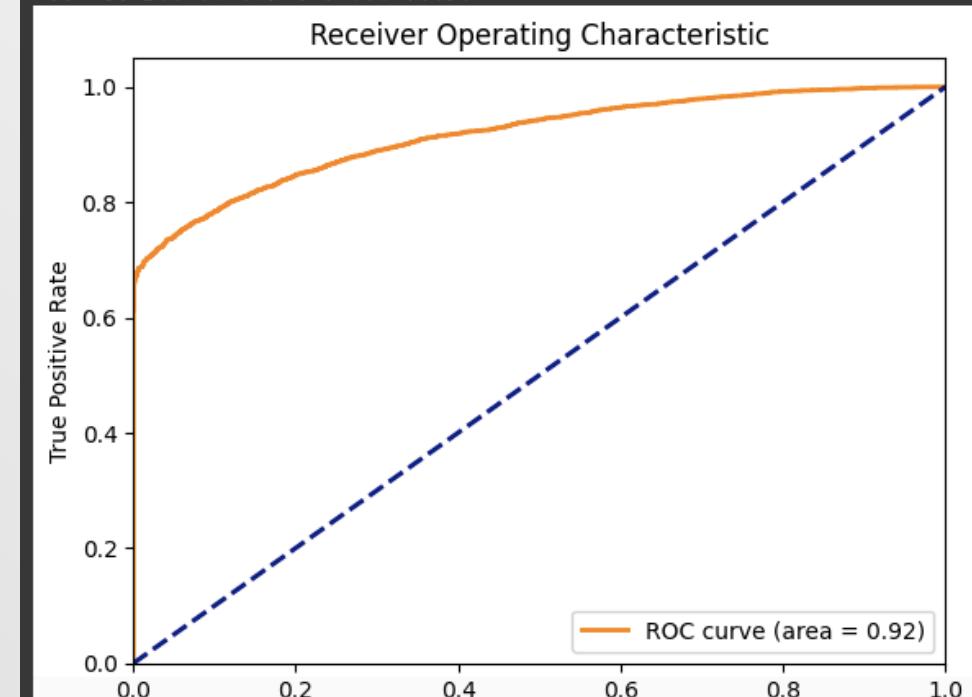
A idade, o número de dependentes e o último salário informado são hipóteses refutadas, não tem impacto sobre o nível de inadimplência.

Para testar o modelo diversas técnicas foram utilizadas, o modelo ajustado performa com uma acurácia de 82% e prevê corretamente 85% dos clientes com risco de inadimplência, enquanto os clientes sem risco de inadimplência são previstos corretamente 79% das vezes.

O modelo foi ajustado novamente e no re-ajuste seu desempenho geral alcança 97% de acurácia para identificar o perfil de pagamento dos clientes do Super Caja.

Relatório de Classificação Ajustado:				
	precision	recall	f1-score	support
bom_pagador	0.84	0.79	0.82	3612
mau_pagador	0.80	0.85	0.83	3588
accuracy			0.82	7200
macro avg	0.82	0.82	0.82	7200
weighted avg	0.82	0.82	0.82	7200

ROC-AUC Score: 0.9162317081300904



Pontos de Atenção

Na análise exploratória descobrimos que mulheres recebem menos e esse pode ser um fator para a maior taxa de inadimplência, dito isto, podemos pensar linhas de crédito para mulheres e programas que auxiliem a sair da inadimplência, com menor taxa de juros, facilitação no pagamento e menores parcelas.

Para que o desempenho do modelo preditivo melhore, podemos utilizar outras informações que não estão no banco de dado, como a data de efetivação do empréstimo, o valor de cada empréstimo, assim formamos um Score de Crédito ainda mais eficiente para classificar os clientes, criar diferentes tipos de score que não segmente apenas bons e maus pagadores, pode melhorar a classificação.

