

Ranqueamento de Clientes com Classificação

Agenda:

- Contexto
- Desafio
- Desenvolvimento da Solução
- Conclusão & Demonstração
- Próximos Passos

1.0 Contexto

1.0 Contexto

- A SafeLife é uma seguradora de saúde e quer potencializar seus lucros e então adotou a estratégia de cross-sell, visando identificar clientes inclinados a contratar também um seguro de veículo.
- Foi feita uma pesquisa com seus clientes sobre o interesse nesse novo seguro.

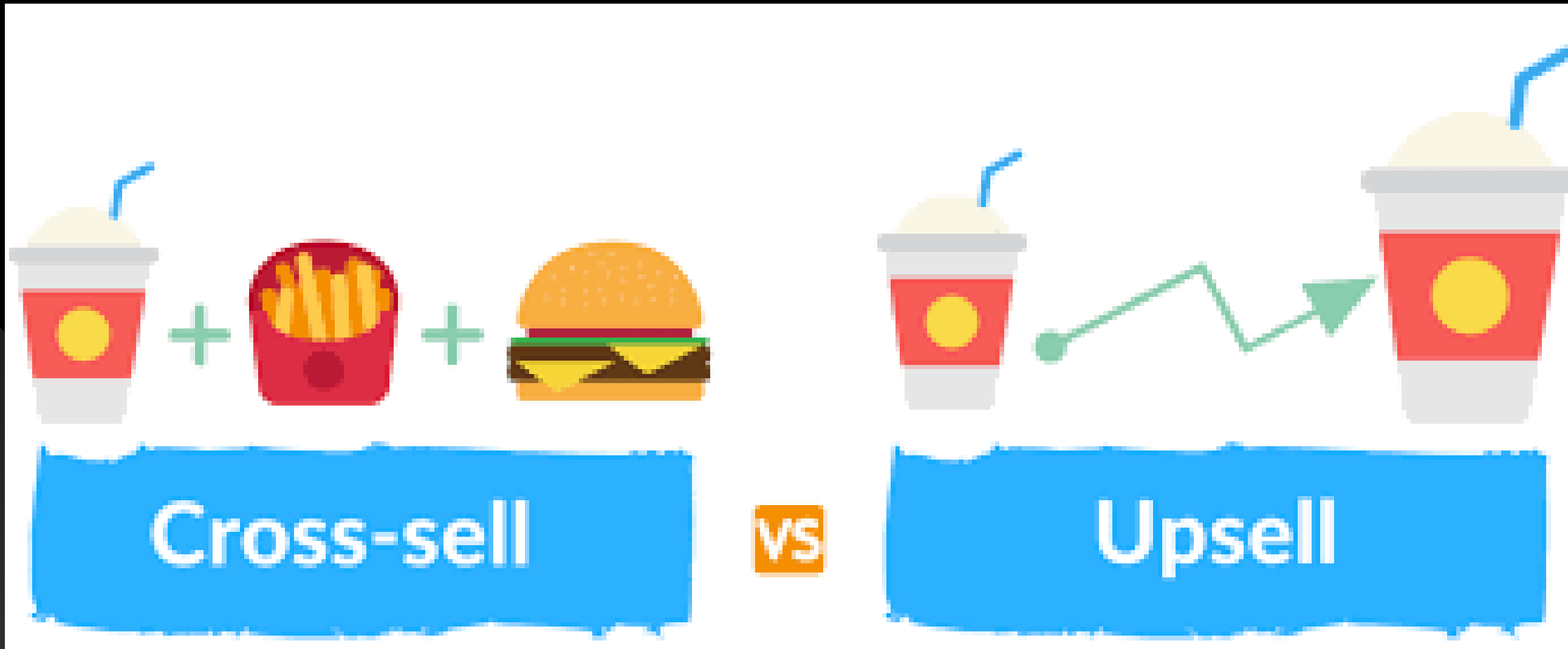
Seguro

- Acordo pelo qual uma empresa se compromete a fornecer uma garantia de compensação por perdas, danos, doenças ou morte específicos em troca do pagamento de um prêmio específico.
- Probabilidade.
- Todos compartilham o risco de todos os outros.

Seguro de Veículo

- Em caso de acidente infeliz com o veículo, a seguradora fornece uma indenização chamada de 'soma assegurada' para o consumidor.
- Fatores que influenciam no valor do prêmio:
 1. Idade
 2. Garagem
 3. Rastreador veicular
 4. Tipo do carro
 5. Local
 6. Uso do veículo

Cross-Sell



2. Desafio

2.0 Desafio

Problema:

- Definir os clientes que o time de marketing deveria entrar em contato.

2.0 Desafio

Problema:

- Definir os clientes que o time de marketing deveria entrar em contato.

Causas:

- O grande número de clientes torna desafiador o processo de seleção, dificultando a identificação dos clientes ideais para o time de marketing contatar.

2.0 Desafio

Solução:

- Usar Machine Learning para fazer o ranqueamento dos clientes.

2.0 Desafio

Solução:

- Usar Machine Learning para fazer o ranqueamento dos clientes.

Produto Final:

- Uma planilha no Google Sheets com um botão vinculado ao modelo na nuvem, permitindo realizar previsões de forma eficiente

Machine Learning



3.0 Desenvolvimento da Solução

3.0 Desenvolvimento da Solução

Descrição dos Dados:

- Number of Rows: 304887
- Number of Cols: 12

3.0 Desenvolvimento da Solução

Descrição dos Dados:

Campo	Descrição
ID	Identificador único do cliente.
Gênero	Gênero do cliente.
Idade	Idade do cliente.
Carteira de Motorista	0: O cliente não possui CNH, 1: O cliente já possui CNH.
Código de Região	Código único para a região do cliente.
Já Segurado Antes	1: O cliente já possui Seguro de Veículo, 0: O cliente não possui Seguro de Veículo.
Idade do Veículo	Idade do veículo.
Danos no Veículo	1: O cliente teve seu veículo danificado no passado, 0: O cliente não teve seu veículo danificado no passado.
Prêmio Anual	O valor que o cliente precisa pagar como prêmio durante o ano.
Canal de Vendas da Apólice	Código anonimizado para o canal de contato com o cliente, ou seja, diferentes agentes, por correio, por telefone, pessoalmente, etc.
Tempo de Associação	Número de dias que o cliente está associado à empresa.
Resposta	1: O cliente está interessado, 0: O cliente não está interessado.

3.0 Desenvolvimento da Solução

Estatística Descritiva:

Feature	Unique	Non-Null Count	Dtype
id	304887	304887 non-null	int64
gender	2	304887 non-null	object
age	66	304887 non-null	int64
driving_license	2	304887 non-null	int64
region_code	53	304887 non-null	float64
previously_insured	2	304887 non-null	int64
vehicle_age	3	304887 non-null	object
vehicle_damage	2	304887 non-null	object
annual_premium	46479	304887 non-null	float64
policy_sales_channel	154	304887 non-null	float64
vintage	290	304887 non-null	int64
response	2	304887 non-null	int64

3.0 Desenvolvimento da Solução

Estatística Descritiva:

features	min	max	range	mean	median	std	skew	kurtosis
age	20.0	85.0	65.0	38.826897	36.0	15.515274	0.671710	-0.568526
annual_premium	2630.0	540165.0	537535.0	30591.308311	31697.0	17239.257418	1.721299	32.191697
vintage	10.0	299.0	289.0	154.392214	154.0	83.670175	0.002134	-1.200642

Outliers

policybazaar.com
HAR FAMILY HOGI INSURED

Term Insurance Investment Plans Health Insurance Motor Insurance

Above-mentioned details are for a 7-year-old Maruti WAGON R AVANCE LXI (998 CC) registered in Gurgaon and 15 days before expiry of previous policy

IDV Calculator

LAMBORGHINI Urus Petrol Twin Turbo V8 (3996 cc) 2023 DL1

Just one more step

Did you file any claims in the previous policy year?


☐ Yes ☒ No

IDV	Premium	
₹ 2,55,00,000	₹ 3,07,020	<div>↻ Calculate again</div> <div>View Plan</div>

This premium has been calculated based on the standard available discount.


Lamborghini Urus 4.0 V8 LP 650-4 2023

Lamborghini Urus 2023 | Lamborghini Urus a venda | Simulador de financiamento | Vender este carro



icarros
Preço 0Km
N/D
Preço sugerido pela montadora.


fipe
Preço FIPE
R\$ 3.751.458
Preço que representa a média de veículos no mercado nacional.



Simulador de Financiamento

Faça uma simulação e veja o valor do seu próximo carro, online e sem compromisso.

Simular Financiamento



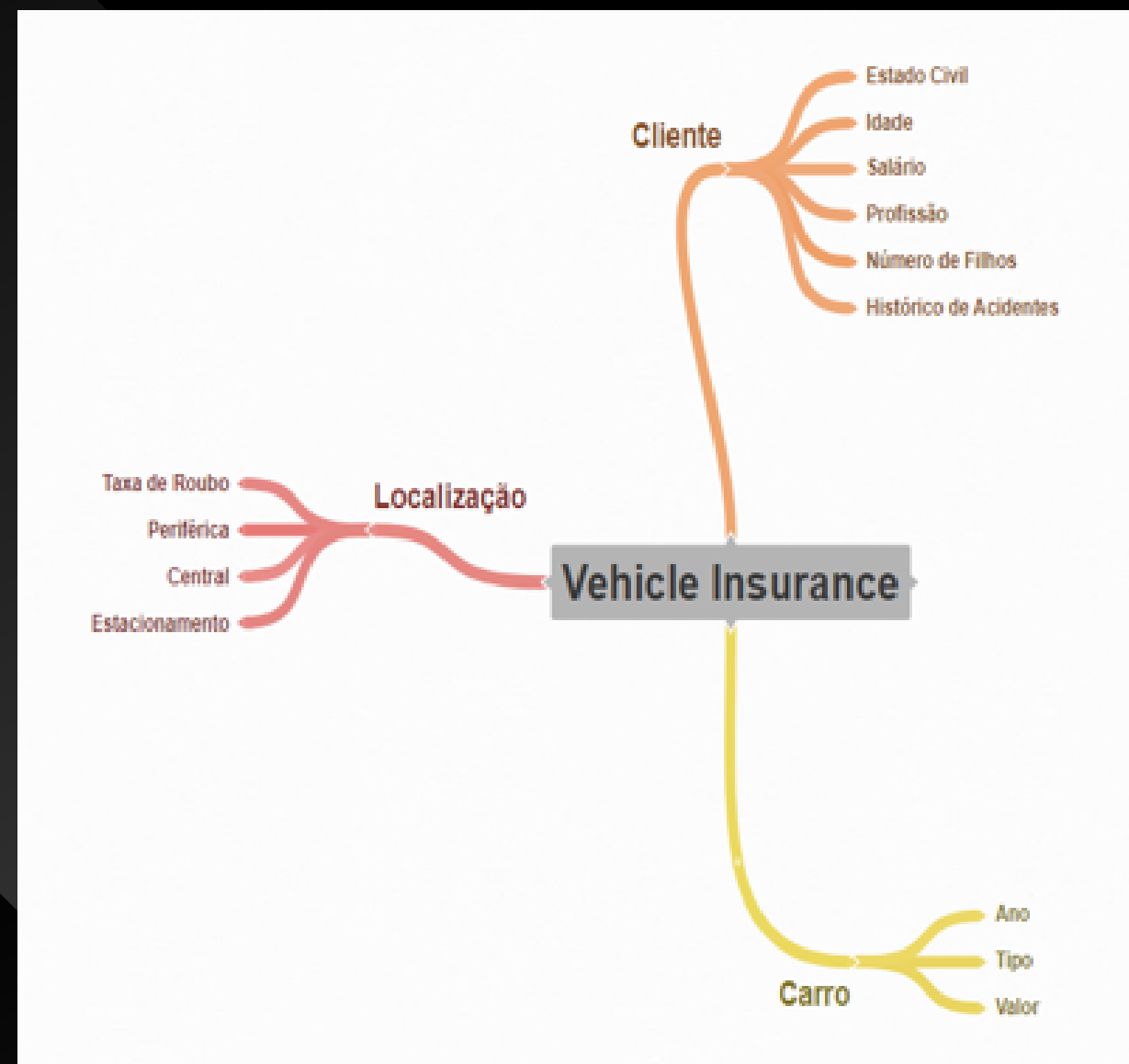
Simule o consórcio Klubi, indicado pelo icarros

Realize seu sonho com facilidade, sem juros e sem entrada.

Simular consórcio

3.0 Desenvolvimento da Solução

Mapa Mental de Hipóteses:



3.0 Desenvolvimento da Solução

Hipóteses da Análise Exploratória:

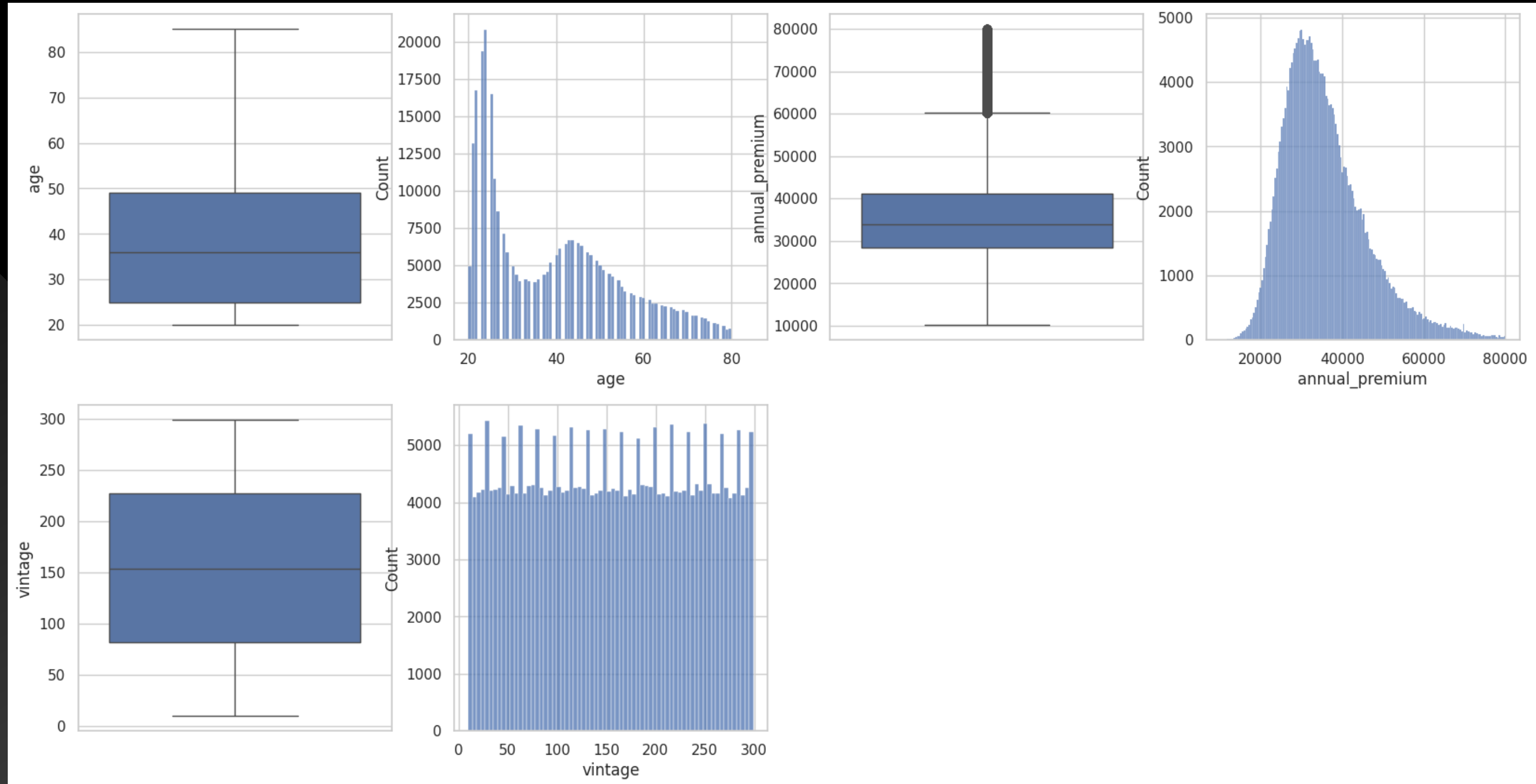
1. Clientes que já tiveram seu veículo danificado podem ter maior interesse em adquirir o seguro de veículo.
2. Clientes que não tem seguro de carro podem estar mais interessados em um seguro de veículo.
3. Clientes há mais tempo são mais prováveis de adquirirem o seguro.
4. A idade pode ser um fator diferencial, com clientes mais jovens tendo menos interesse.
5. Clientes com carros novos podem ter maior interesse em aderir ao seguro de carro.
6. O gênero do cliente pode ser um fator importante na aquisição do seguro.
7. Clientes que moram em regiões específicas podem ter maior interesse no seguro.
8. Diferentes métodos de contato podem impactar o interesse em seguro de veículos.
9. O valor cobrado pode impactar no interesse do seguro de veículo.
10. Clientes que tem CNH são mais prováveis de adquirir o seguro.

3.0 Desenvolvimento da Solução

Análise Exploratória de Dados:

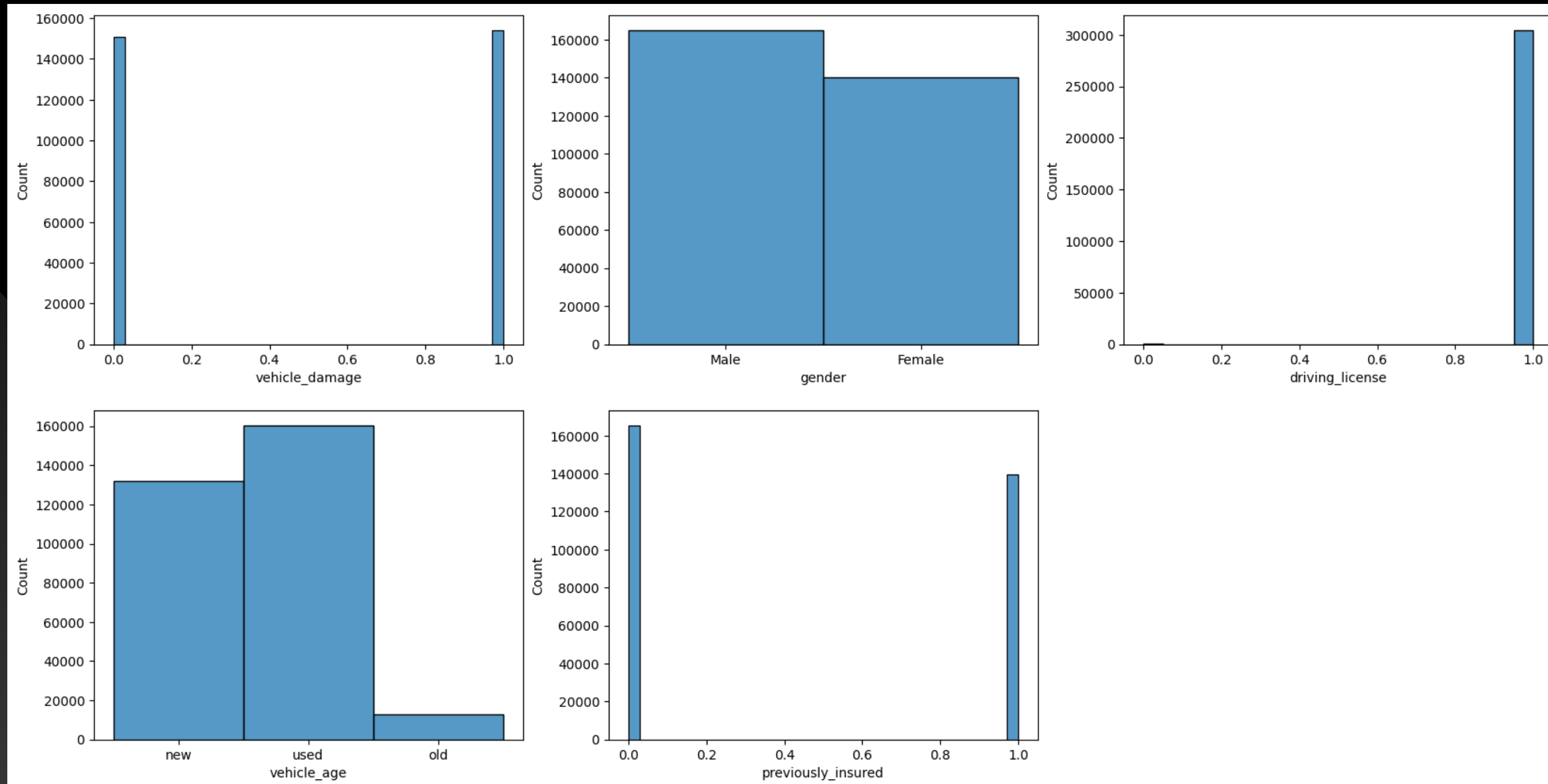
Análise Exploratória de Dados:

Variável Numérica:



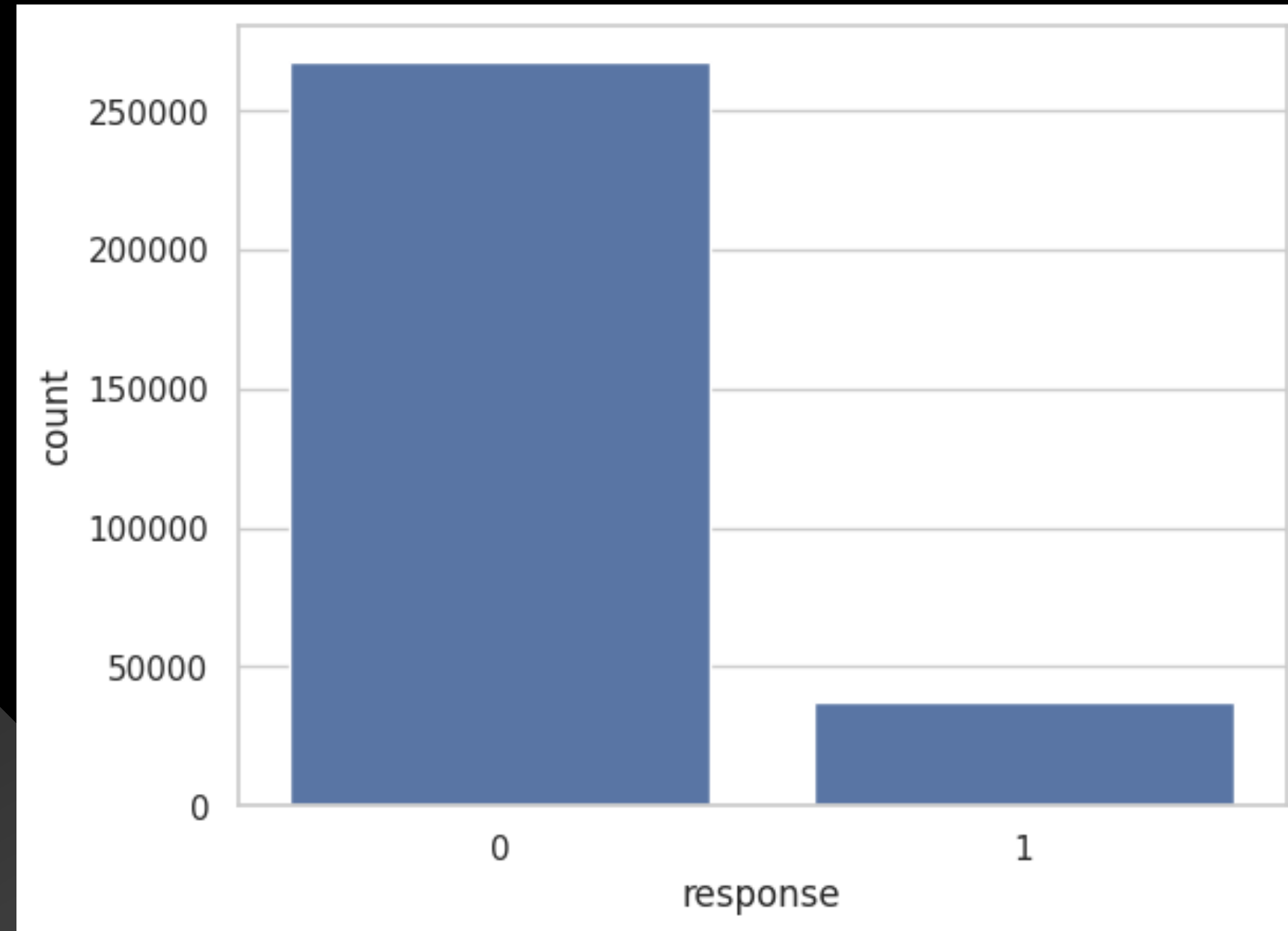
Análise Exploratória de Dados:

Variável Categórica:



Análise Exploratória de Dados:

Variável Resposta:

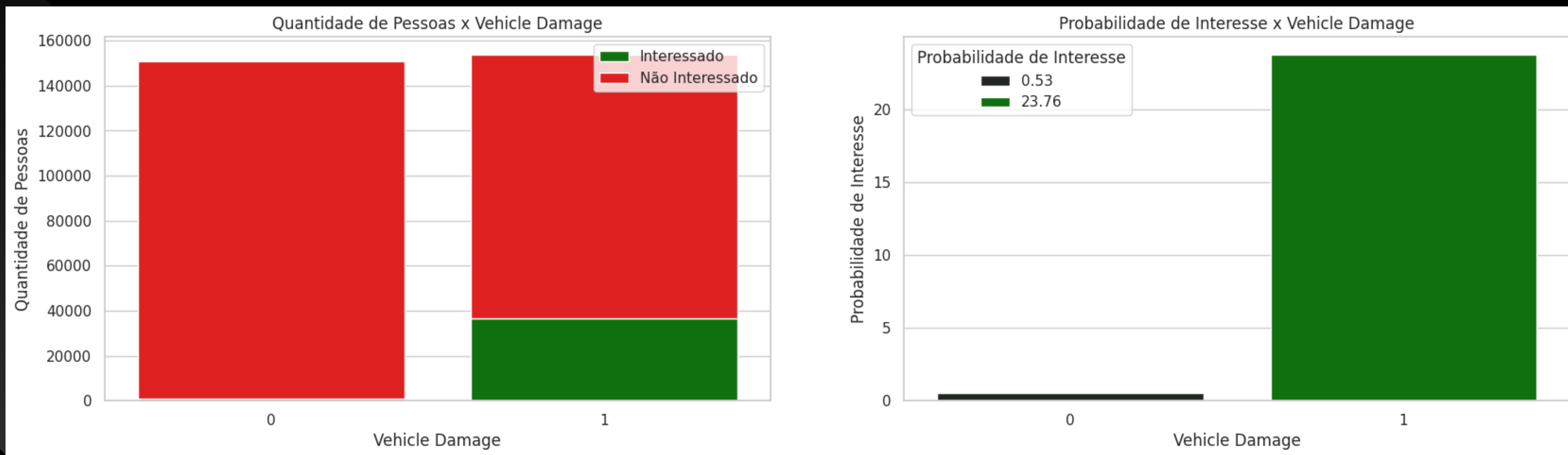


3.0 Desenvolvimento da Solução

Validação de Hipóteses:

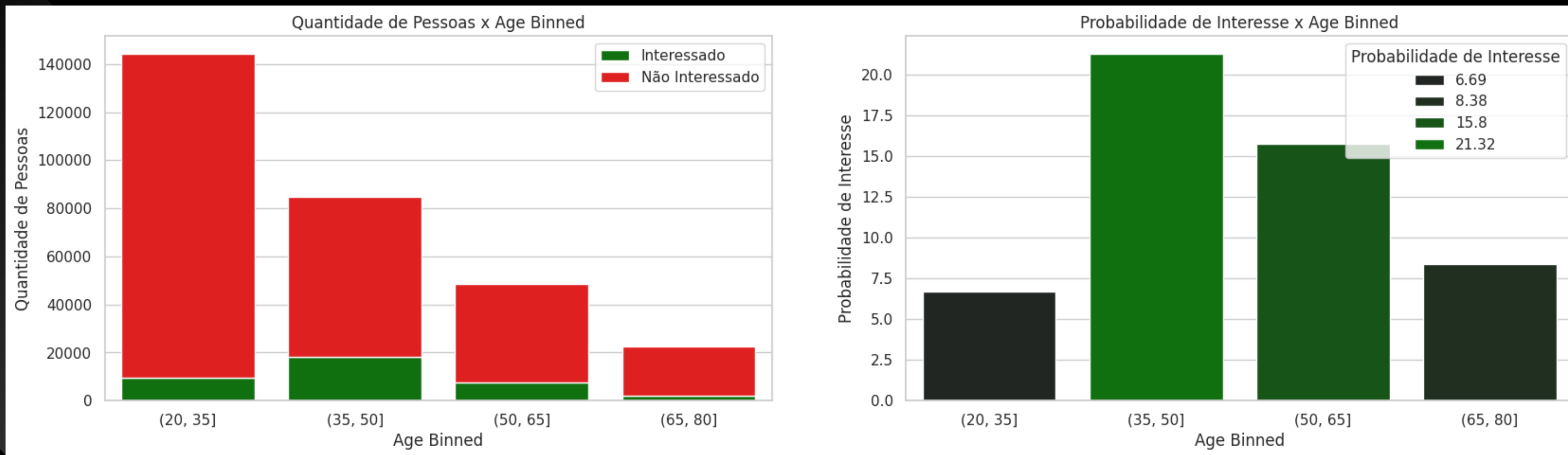
H1. Clientes que já tiveram seu veículo danificado podem ter maior interesse em adquirir o seguro de veículo.

VERDADEIRO – A ocorrência de incidentes anteriores exerce uma influência significativa no aumento do interesse pelo seguro automotivo.

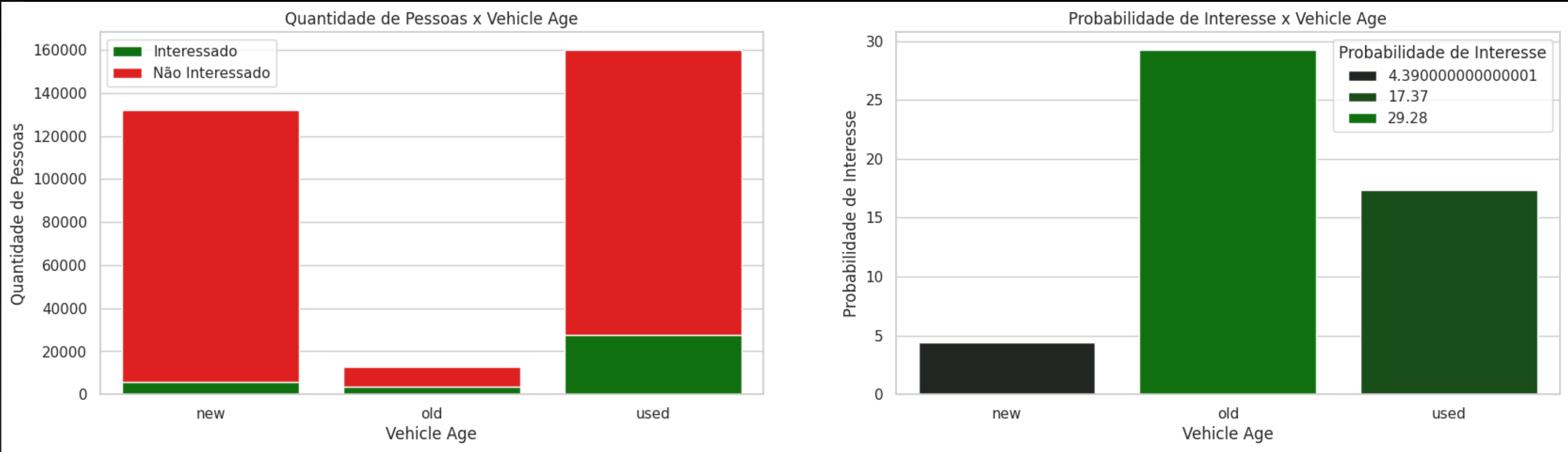


H2. A idade pode ser um fator diferencial, com clientes mais jovens tendo menos interesse.

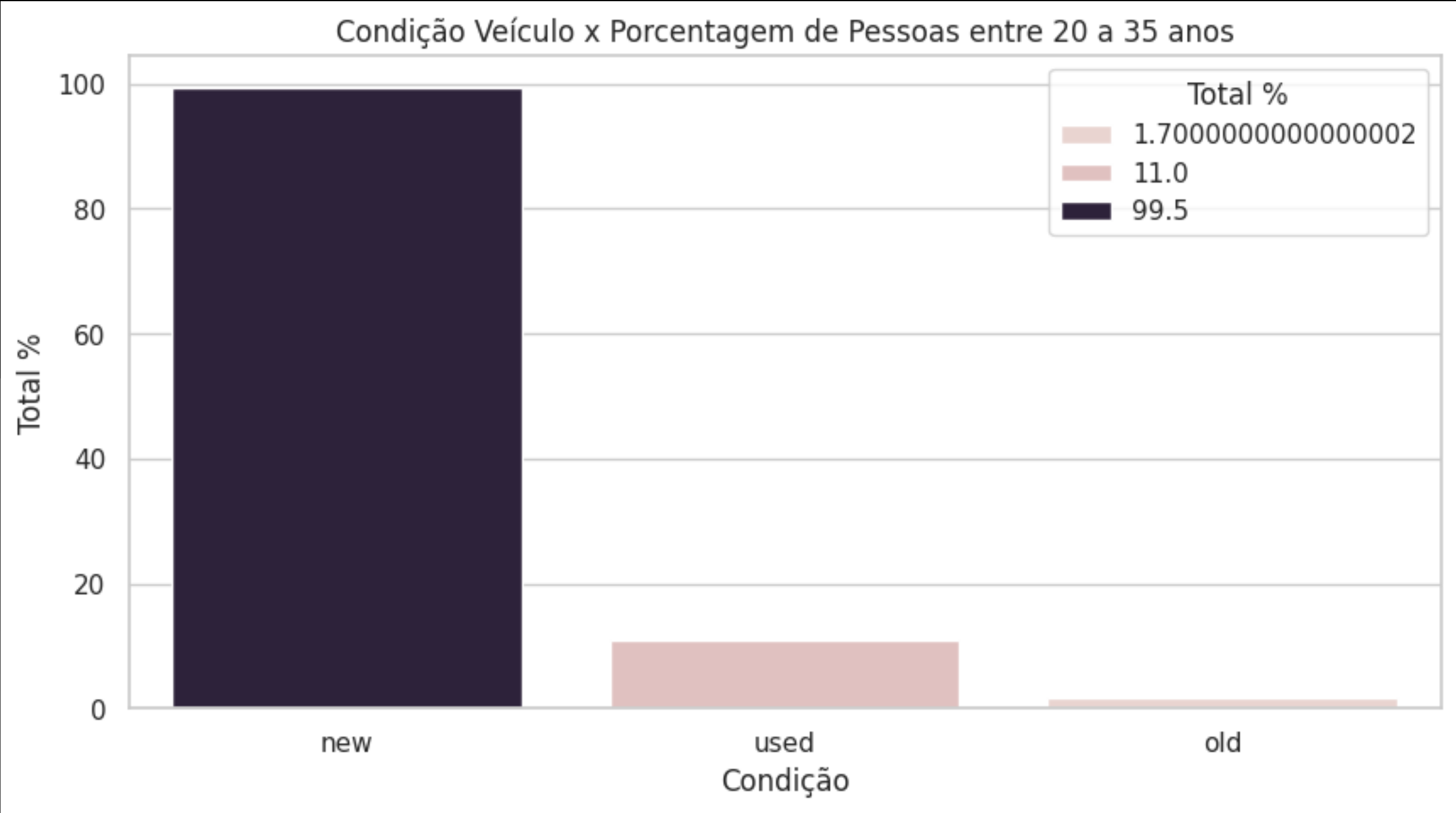
VERDADEIRO – Clientes com idade entre 35 e 50 são os mais interessados.



H3. Clientes com carros novos tem maior interesse em aderir ao seguro de carro.
FALSO– Clientes com carros usados e velhos tem maior interesse.

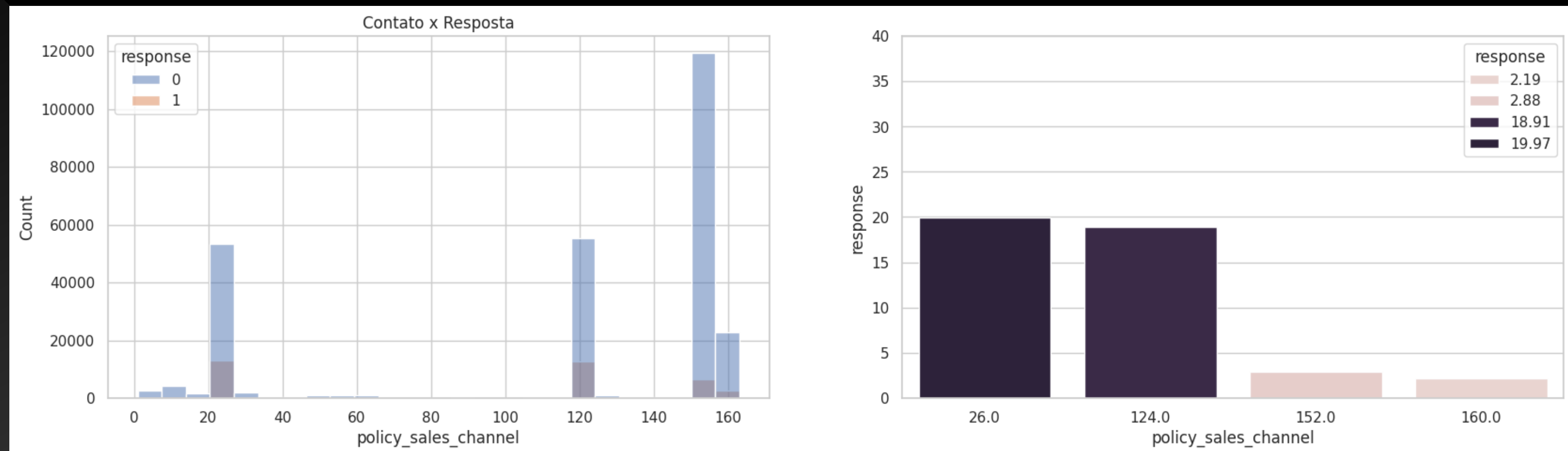


H3. Explicação



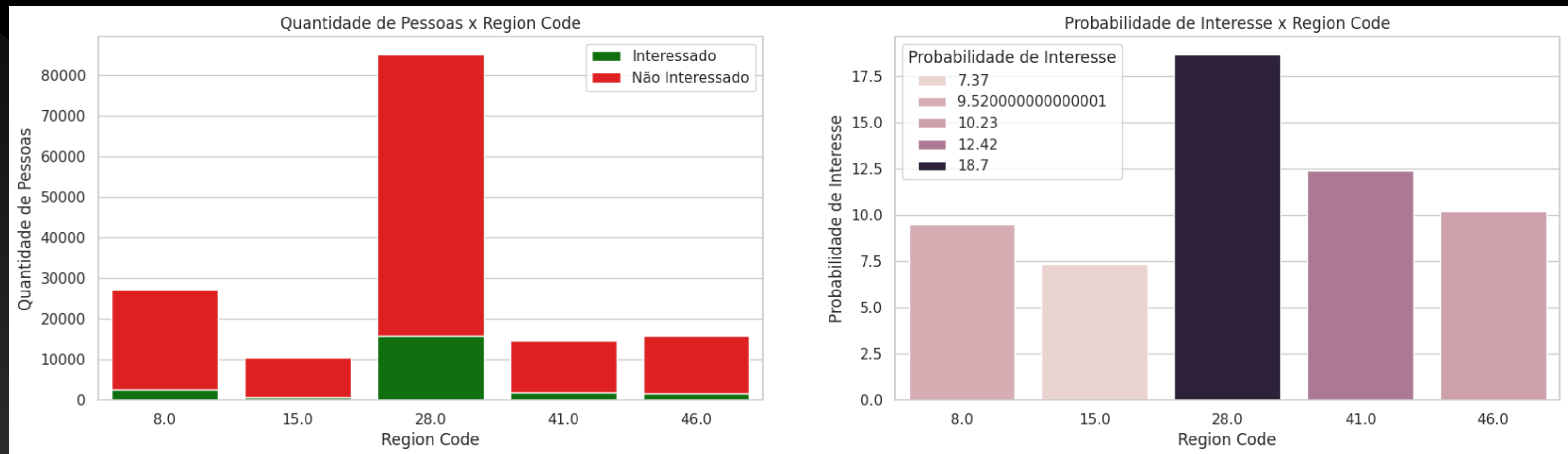
H4. Diferentes métodos de contato podem impactar o interesse em seguro de veículos.

VERDADEIRO – Os métodos de contato 26.0 e 124.0 apresentam a maior taxa de conversão de clientes interessados em seguro de veículos.



H5. Clientes que moram em regiões específicas podem ter maior interesse no seguro.

VERDADEIRO– Pessoas que moram na região 28 tem maior interesse no seguro.



Insights Acionáveis:

Se o projeto chegasse a esse ponto e não envolvesse a parte de Machine Learning, seria uma recomendação válida entrar em contato com os seguintes grupos de clientes:

1. Clientes que já tiveram dano em seus veículos.
2. Clientes que não tem seguro de carro.
3. Aqueles com faixa etária entre 35 e 50 anos.
4. Clientes residentes da região 28.

E se atentar aos seguintes pontos:

- Clientes com carros novos e que são jovens tem pouquíssimo interesse no seguro.
- Priorizar os métodos de contato 26 e 124.
- A longevidade do cliente na empresa não é um fator importante.

Essas recomendações vão direcionar as estratégias de contato e vendas para maximizar as chances de sucesso.

3.0 Desenvolvimento da Solução

Machine Learning:

Comparação da Performance em 15% dos Dados de Treino usando Cross-Validation (5 kfold):

Models CV	Precision at k	Recall at k	F1-Score at k
XGBoost	0.370	0.453	0.407
LogisticRegression	0.343	0.420	0.378
Random Forest	0.340	0.416	0.374
KNN	0.317	0.389	0.349

3.0 Desenvolvimento da Solução

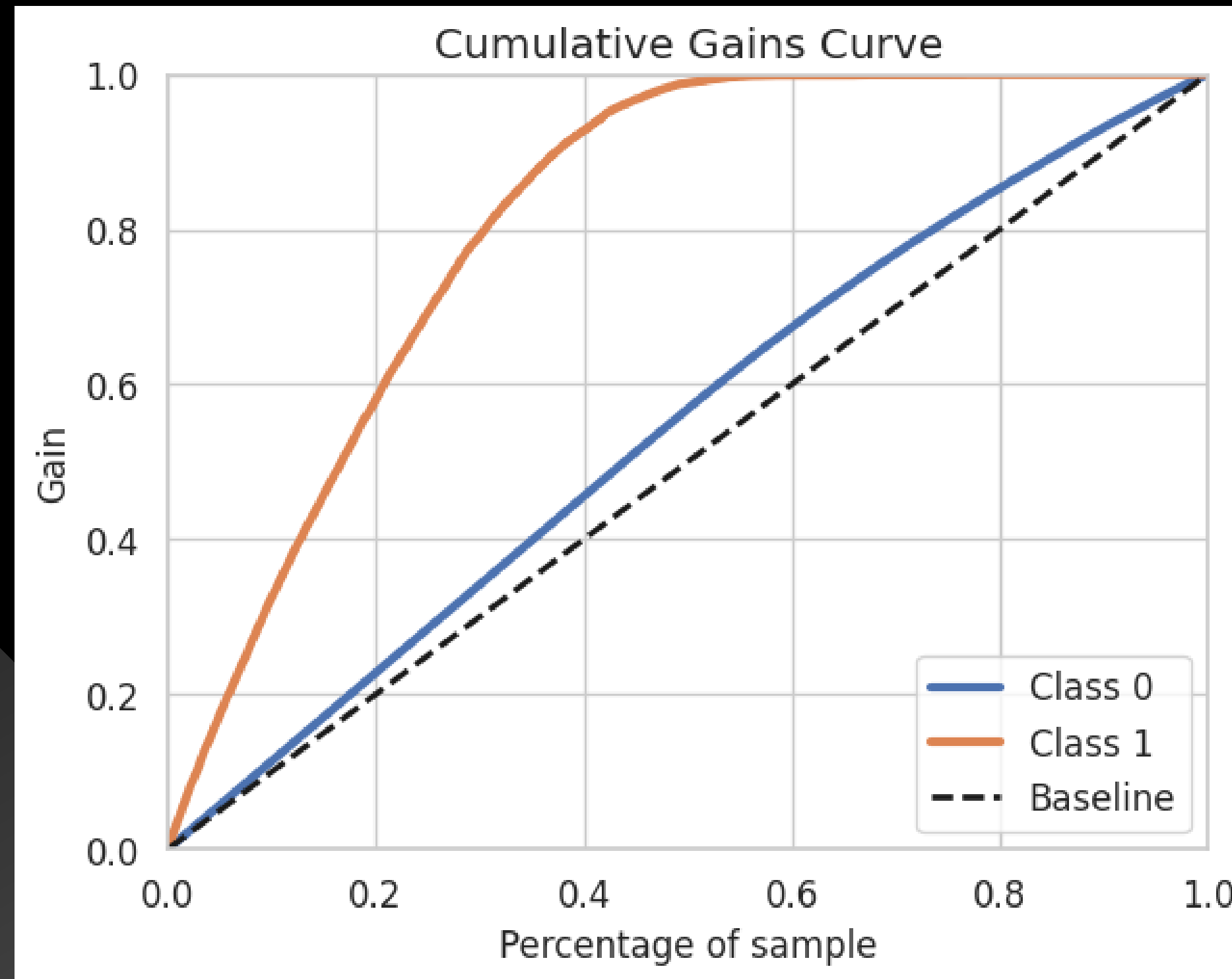
Machine Learning:

Comparação da Performance em 15% dos Dados de Validação:

Model	Precision at k	Recall at k	F1-Score at k
XGBoost	0.374	0.458	0.412

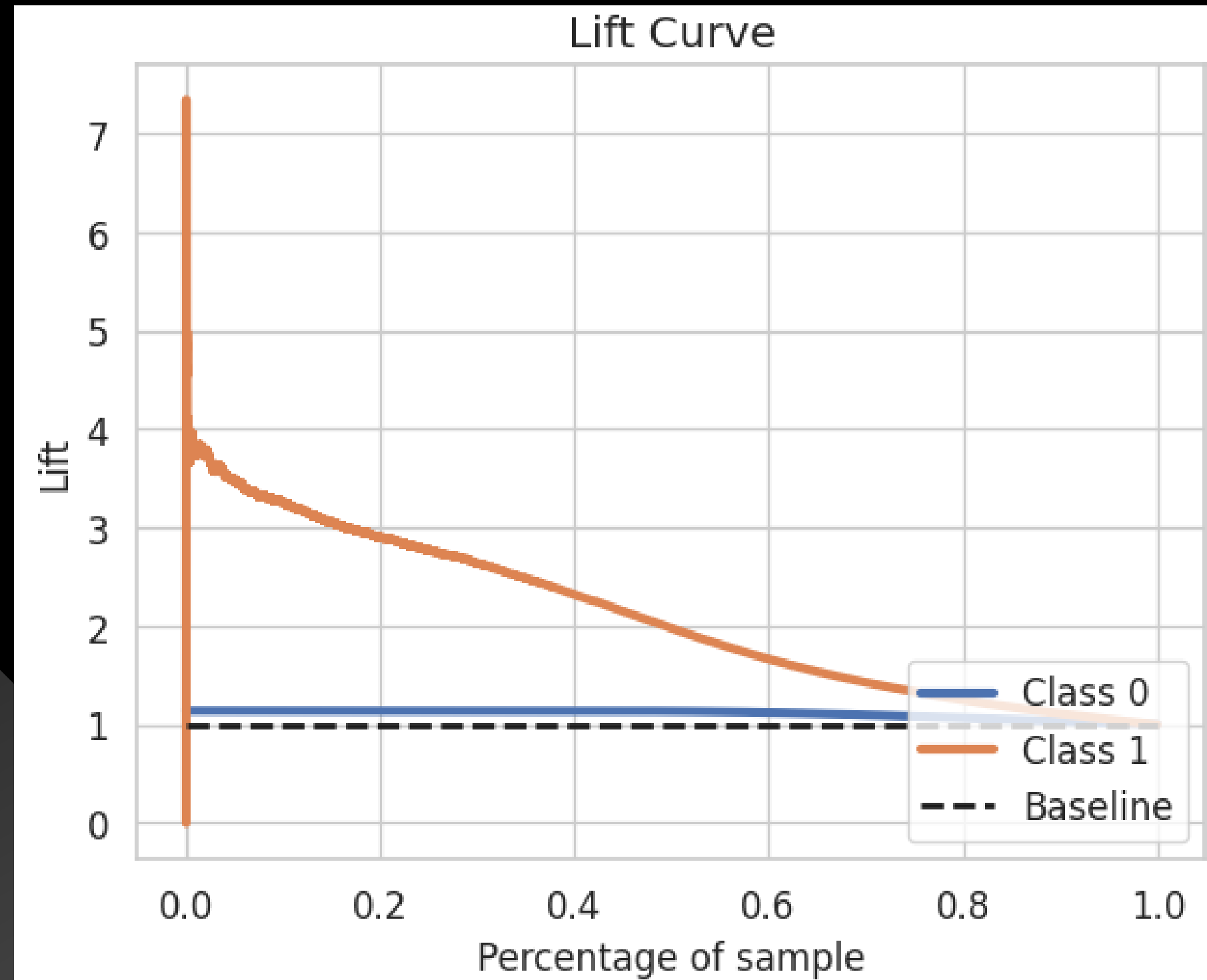
Machine Learning

Curva de Ganho Cumulativo:



Machine Learning

Curva de Ganho:



4.0 Conclusão & Demonstração

4.0 Conclusão & Demonstração

Tradução e Interpretação do Erro:

Considerando que a empresa esteja disposta a se comunicar com possíveis interessados e obtenha um lucro bruto de R\$ 600,00 para cada cliente que adquira o seguro de veículo, mas também incorra um custo de R\$ 50,00 a cada vez que entra em contato com um potencial cliente, podemos mensurar os seguintes planos de ação:

Tradução e Interpretação do Erro:

Plano de Ação 1:

Entrar em contato somente com pessoas que não tem seguro de veículo.

- Lucro ao entrar em contato com 68.595 clientes (54% da base): R\$ 5.871.610,00
- Lucro ao entrar em contato com toda a base de clientes: R\$ 2.842.391,00

Isso implica em um aumento de 2.06 vezes no lucro potencial se o plano for implementado em sua totalidade.

Tradução e Interpretação do Erro:

Plano de Ação 2:

Entrar em contato somente com pessoas que tiveram incidentes com seus veículos.

- Lucro ao entrar em contato com 63.835 clientes (50% da base): R\$ 5.934.596,00
- Lucro ao entrar em contato com toda a base de clientes: R\$ 2.842.391,00

Isso implica em um aumento de 2.09 vezes no lucro potencial se o plano for implementado em sua totalidade.

Tradução e Interpretação do Erro:

Plano de Ação 3:

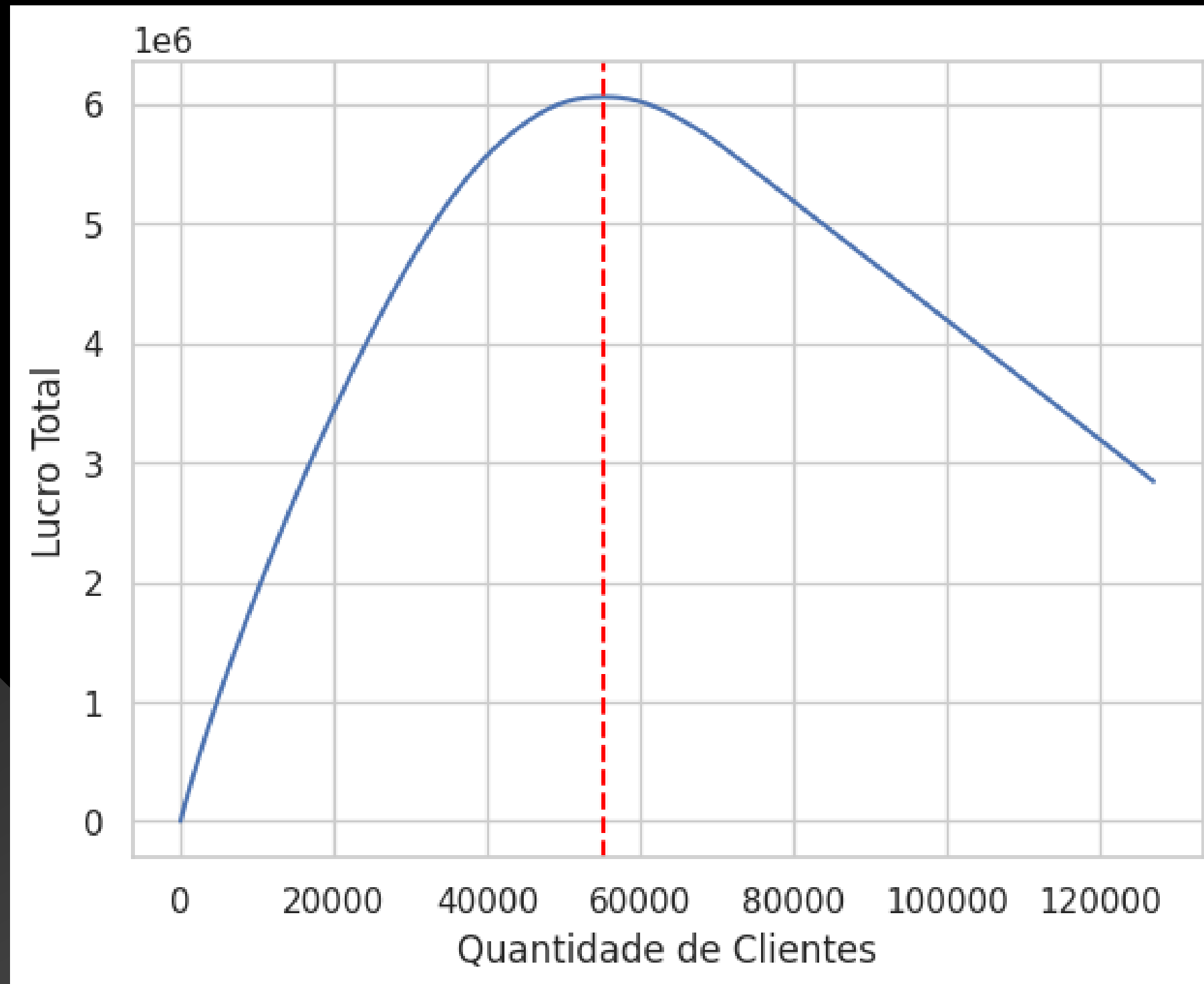
Entrar em contato somente com pessoas selecionadas pelo modelo.

- Lucro ao entrar em contato com os primeiros 55.100 clientes (43% da base): R\$ 6.056.576,00
- Lucro ao entrar em contato com toda a base de clientes: R\$ 2.842.391,00

Isso implica em um aumento de 2.13 vezes no lucro potencial se o plano for implementado em sua totalidade.

Tradução e Interpretação do Erro:

Performance Total:



Demonstração da Planilha:

Arquivo Editar Ver Inserir Formatar Dados Ferramentas Extensões Ajuda Health Insurance

Script concluído

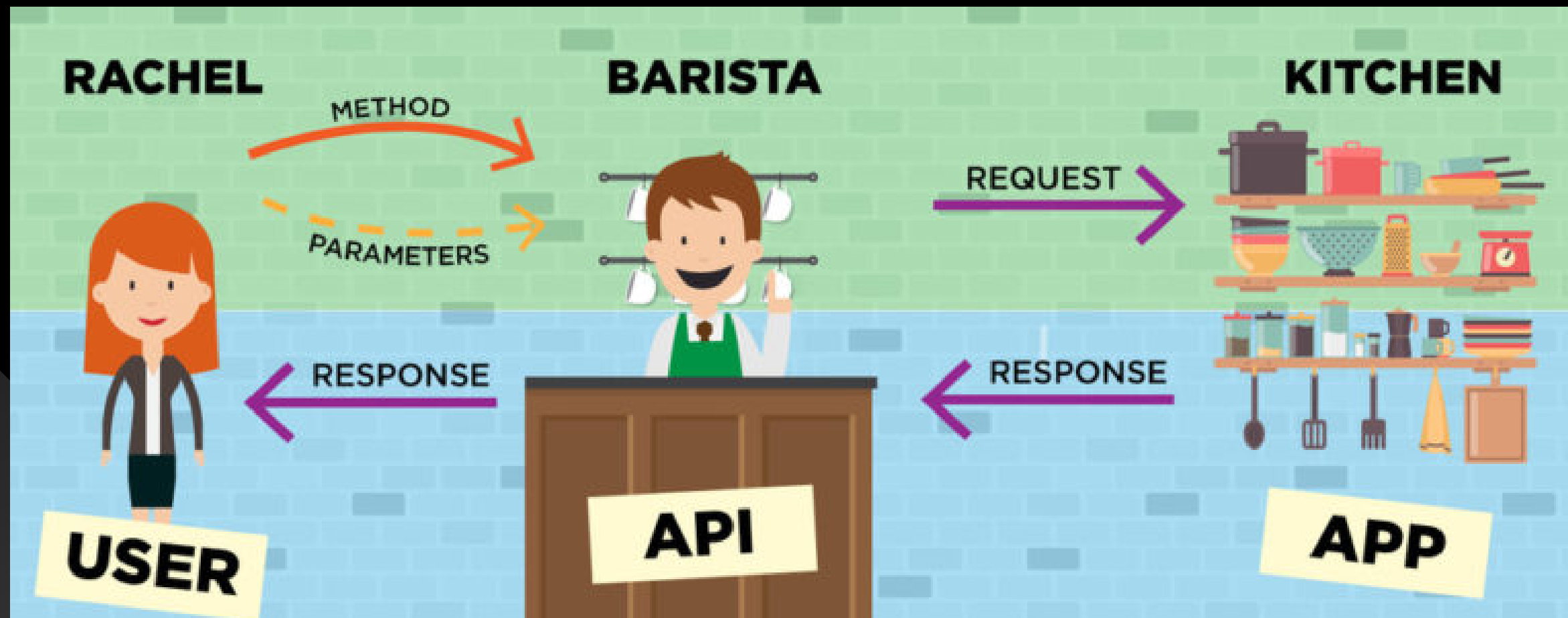
100% 123 Padrão 10 B I A

A9:K26 | fx 381313

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	id	Gender	Age	Driving_License	Region_Code	Previously_Insur	Vehicle_Age	Vehicle_Damage	Annual_Premium	Policy_Sales_Ct	Vintage	Score				
2	381278	Male	25	1 2.0		0	1-2 Year	Yes	2630.0	156.0	279	0,12310718				
3	381329	Male	66	1 2.0		0	1-2 Year	Yes	35386.0	124.0	129	0,1162946671				
4	381331	Male	56	1 8.0		0	1-2 Year	Yes	42395.0	154.0	44	0,115511708				
5	381308	Male	33	1 28.0		0	1-2 Year	Yes	2630.0	156.0	43	0,1062083915				
6	381281	Female	20	1 28.0		0	1-2 Year	Yes	41983.0	124.0	203	0,096777983				
7	381310	Male	28	1 28.0		0	1-2 Year	Yes	30267.0	124.0	161	0,096777983				
8	381279	Male	38	1 28.0		0	1-2 Year	Yes	33196.0	26.0	55	0,0792826265				
9	381313	Female	44	1 28.0		0	1-2 Year	Yes	46671.0	26.0	94	0,0792826265				
10	381270	Female	53	1 28.0		0	1-2 Year	Yes	2630.0	152.0	68	0,0544223674				
11	381320	Female	22	1 41.0		0	< 1 Year	Yes	52604.0	152.0	51	0,0481073298				
12	381157	Male	30	1 8.0		0	< 1 Year	No	24180.0	152.0	168	0,0122842053				
13	381335	Male	25	1 32.0		1	< 1 Year	No	25271.0	152.0	36	0,0003695015				
14	381268	Male	57	1 30.0		1	1-2 Year	No	26272.0	152.0	195	0,0003597092				
15	381311	Male	25	1 8.0		1	< 1 Year	No	61002.0	152.0	266	0,0003579371				
16	381319	Male	24	1 11.0		1	< 1 Year	No	29579.0	152.0	125	0,0001961116				
17	381337	Female	36	1 35.0		1	1-2 Year	No	25277.0	152.0	11	0,0001923362				
18	381274	Female	22	1 29.0		1	< 1 Year	No	33758.0	152.0	211	0,0001917916				
19	381277	Male	22	1 3.0		1	< 1 Year	No	26138.0	152.0	135	0,0001917916				
20	381265	Male	22	1 46.0		0	< 1 Year	Yes	32497.0	152.0	138					
21	381323	Male	76	1 28.0		0	1-2 Year	No	35996.0	26.0	175					
22	381269	Female	23	1 13.0		0	< 1 Year	Yes	28589.0	152.0	296					
23	381325	Male	36	1 14.0		0	1-2 Year	Yes	27610.0	152.0	35					
24	381276	Female	27	1 8.0		0	< 1 Year	Yes	49623.0	152.0	117					
25	381314	Female	61	1 8.0		0	1-2 Year	No	42327.0	26.0	59					
26	381317	Female	54	1 8.0		0	1-2 Year	No	31390.0	26.0	86					
27	381315	Female	26	1 41.0		1	< 1 Year	No	27868.0	152.0	246					
28	381328	Male	26	1 29.0		1	< 1 Year	No	71710.0	152.0	222					
29	381272	Male	23	1 3.0		1	< 1 Year	No	26341.0	152.0	219					
30	381333	Female	30	1 41.0		1	< 1 Year	No	47952.0	152.0	194					
31	381326	Male	25	1 28.0		1	< 1 Year	No	34003.0	152.0	61					
32	381332	Female	24	1 4.0		1	< 1 Year	No	27260.0	152.0	274					
33	381336	Male	24	1 28.0		1	< 1 Year	No	29925.0	152.0	135					
34	381266	Female	22	1 28.0		1	< 1 Year	No	28101.0	160.0	277					
35	381318	Male	28	1 11.0		0	< 1 Year	Yes	2630.0	156.0	201					
36	381321	Female	24	1 11.0		0	< 1 Year	Yes	40921.0	124.0	112					
37	381159	Male	61	1 48.0		0	1-2 Year	Yes	2630.0	124.0	132					
38	381158	Male	54	1 30.0		0	1-2 Year	Yes	23364.0	156.0	217					
39	381327	Male	51	1 0.0		0	1-2 Year	Yes	2630.0	12.0	218					
40	381322	Male	64	1 47.0		0	1-2 Year	Yes	29575.0	26.0	264					
41	381306	Male	49	1 8.0		0	1-2 Year	Yes	40088.0	26.0	25					
42	381316	Male	53	1 28.0		0	1-2 Year	Yes	61685.0	25.0	24					
43	381271	Female	67	1 28.0		0	1-2 Year	Yes	30477.0	26.0	290					
44	381305	Female	20	1 37.0		0	< 1 Year	Yes	55236.0	160.0	12					

+ test Tabela dinâmica 1

API



5.0 Próximos Passos

5.0 Próximos Passos

- Explorar novos modelos de Machine Learning.
- Integração com o Excel: Levar a praticidade da predição aos usuários do Excel, integrando o sistema de predição diretamente na plataforma.
- Formulação de Novas Hipóteses: Desenvolver e testar novas hipóteses para gerar insights adicionais para o negócio.

Q & A

Muito Obrigado!