

# Desenvolvimento de Regressão Logística em R

Score para priorização de Leads

Tibolão – Thiago P B Gerolamo

Mai/2024



## Thiago do Prado Barros Gerolamo - Tibolão



**Profissional com 19 anos em Analytics**, com aplicações em **Crédito, Risco e CRM**. Atuando como líder de pessoas a mais de 10 anos. Ampla experiência em modelagem de dados, análises estatísticas, criação e definição das **Políticas de Crédito**, estruturação de regras e melhorias nos fluxos de **Prevenção à Fraude**, projetos de **CRM e Business Intelligence (BI)**. Com habilidades multidisciplinares e ótimo relacionamento interpessoal tem facilidade em entender as causas dos problemas e gerar recomendações para as tomadas de decisões, sempre baseado em dados com **foco em redução de riscos, custos ou melhoria de processos**.



Trajetória nas Empresas: **Sky Banda Larga | Banco Santander | Serasa Experian | Kavak.com | Banco Cacique | Aymoré Financiamentos | CSU Cardsystem**





O **Grupo Tony Veículos** tem mais de **40 anos no Setor Automotivo**. Com mais de 7 lojas na região de São José dos Campos/SP é **líder nas vendas e compras de veículos no Vale do Paraíba**.

A **onCar**, uma das empresas do Grupo, vem com uma “pegada” de Inovação em Produtos e Serviços numa constante **Transformação Digital**. Sua missão é ser um facilitador entre as vendas e compras de veículos, **melhorando toda a jornada dos seus clientes**.



Atualmente, a onCar **não tem *capacity*** (capacidade) de atender todo o volume de clientes que realiza consultas/simulações no site da onCar. Existe a **necessidade de selecionar** clientes para o atendimento humano (via ligação). **\*Em média, conseguem atender 30% dos clientes que realizam simulação no site.**

## Algumas Soluções:

- ✓ Aumentar a quantidade de atendentes (alto custo, muitos “vales”)
- ✓ Priorizar os clientes com maiores chances de comprar um veículo (sem necessidade de mudar a estrutura de atendentes, fácil aplicação de regras de priorização)

## Decisão

*Desenvolver um modelo estatístico para identificar quais são os clientes que devem ser priorizados*



# Regressão Logística

Quando o objetivo da análise é responder perguntas do **tipo binário (Sim e Não, Bom e Mau)**, umas das técnicas estatísticas mais utilizadas é a **Regressão Logística**. O modelo tem poder de avaliar a probabilidade de um evento ocorrer, então se as perguntas são "**Qual é a probabilidade de um cliente adquirir um produto?**" ou "**Qual é a probabilidade de um cliente ser um bom pagador?**", este tipo de técnica tem um excelente resultado.

É muito usual a entrega ser um **Score, com pontuação entre 0 a 1000**. O Score é utilizado na estratégia da ação. Por exemplo, se o objetivo é atuar com clientes mais propenso a conversão utiliza os scores mais altos, até a capacidade de atuação dos atendentes. **O resultado para um ótimo modelo depende exclusivamente de um conjunto de ótimo conjunto de dados.**



# Passo a Passo da Modelagem

Para o **desenvolvimento** de um modelo de propensão, temos algumas **etapas** até chegar ao modelo final:

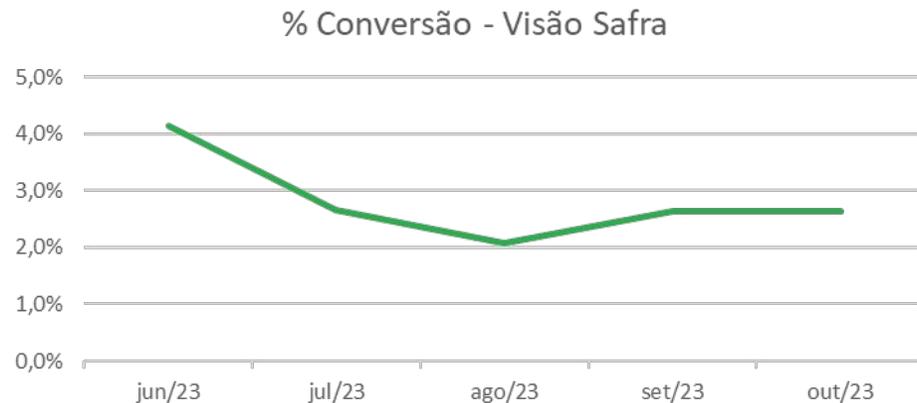
- ✓ **Kick-off com cliente:** Entendimento da necessidade e aplicabilidade do modelo.
- ✓ **Base de Dados:** Informações capturadas no momento do contato. Definição da variável resposta (*target*);
- ✓ **Entendimento da Base:** Aqui são realizados pré-análises (análise univariada, correlação, bivariada). Discussão sobre novas variáveis;
- ✓ **Amostragem:** Quebrar a base em uma amostra para Desenvolvimento (base de treino) e outra base para Validação do Modelo;
- ✓ **Modelagem:**
  - Pré-Seleção de Variáveis
  - Categorização e Interação de Variáveis
  - Ajuste de Modelo
  - Alinhamento do Score
  - Estabilidade
  - Medidas de Performance



# Variável Resposta

**O que é:** É um fator de resultado ou consequência de algo que **depende** do que foi manipulado. Não temos controle sobre esta variável, apenas observamos.

**Vendas:** No nosso modelo de propensão queremos identificar o perfil de clientes que tem maior chance de compra (% de conversão).



Safras	jun/23	jul/23	ago/23	set/23	out/23	Total
% Conversão	4,1%	2,6%	2,1%	2,6%	2,6%	2,6%

## % de Vendas – Quantidade de Dias Após a Criação

Dias até venda	jun/23	jul/23	ago/23	set/23	out/23	Total
Até 10 dias	85%	79%	74%	77%	71%	77%
Até 3 semanas	93%	88%	87%	92%	90%	90%
Até 30 dias	95%	91%	91%	98%	99%	94%

Em média, 90% das vendas ocorrem em até 3 semanas da data de criação. Como a última data de criação é do dia 14/10/2023, e a observação das vendas desses clientes ocorreu até 3 semanas, não vamos desconsiderar o mês de out/23 do desenvolvimento.

