

### APS 1 – Como prever o *Churn* de clientes?



<sup>1</sup>Saber por que um cliente abandona ou deixa de comprar seu produto é fundamental para garantir um crescimento sustentável a longo prazo para a sua empresa. No entanto, e em meio a tantas métricas que compõe esse tipo de análise, o *churn* talvez seja a mais básica – e, também, a mais fundamental como ponto de partida, por representar o número total de cliente cancelados.

#### O que é *churn*?

*Churn* diz respeito aos casos em que um cliente ou usuário para de comprar ou usar o produto ou serviço de uma empresa. Ou seja, representa o número total de clientes cancelados.

Na prática, o *churn* de clientes atrasa o crescimento da empresa de forma direta e, por isso, é uma métrica muito importante. Para que você consiga identificar potenciais causas de *churn* antes do cliente de fato desistir do seu negócio, é essencial que você entenda quais variáveis parecem estar mais relacionadas com os clientes que deixam de usar seu produto ou serviço ou, em outras palavras, identificar qual tipo de cliente tem maior potencial de *churn*.

Diante deste cenário, seu grupo tem a tarefa de analisar os dados de *churn* de clientes de uma empresa de telecomunicações que forneceu produtos de telefone residencial e internet para 7043 clientes na Califórnia durante o terceiro trimestre de 2021.

<sup>1</sup> Texto retirado e adaptado de <https://mindminers.com/blog/o-que-e-churn/>

O arquivo<sup>2</sup> `churn.xlsx` contém as seguintes informações:

- **customer\_ID**: código único de identificação do usuário
- **gender**: sexo do cliente (Female ou Male)
- **SeniorCitizen**: indica se o cliente é um cidadão sênior (1) ou não (0)
- **tenure**: número de meses que o cliente ficou na empresa
- **PhoneService**: indica se o cliente possui serviços de telefone (Yes) ou não (No)
- **OnlineBackup**: indica se o cliente possui serviço de backup online (Yes) ou não (No)
- **StreamingMovies**: indica se o cliente possui serviço de Streaming de Filmes (Yes) ou não (No)
- **QtyServices**: quantidade total de serviços contratados pelo cliente
- **Contract**: tipo de contrato do cliente (Month-to-month, One year, Two year)
- **MonthlyCharges**: valor mensal pago pelo cliente
- **Churn**: indica se o cliente deixou a empresa (Yes) ou não (No)

---

<sup>2</sup> Dados obtidos em <https://www.kaggle.com/datasets/blastchar/telco-customer-churn>

Para guiar sua análise, responda:

- a) Calcule e organize em forma de tabela a proporção de clientes que realizaram *churn* na base de dados. Comente.

## VARIÁVEIS DEMOGRÁFICAS

- b) Utilizando as variáveis demográficas a seguir presentes na base construa dois gráficos adequados para comparar a proporção de clientes que realizaram *churn*:

I) dentre os homens vs dentre as mulheres (variável `gender`)

II) dentre os cidadãos sêniores vs cidadãos não sêniores (variável `SeniorCitizen`)

As duas variáveis parecem ser importantes para explicar o *churn* de clientes? Justifique sua resposta. Resposta sem justificativa será desconsiderada.

## VARIÁVEIS DE COMPORTAMENTO

- c) A partir de medidas numéricas adequadas, descreva o comportamento da variável `tenure` para os clientes que realizaram *churn* em comparação com os clientes não realizaram. Há diferenças? O resultado era esperado? Justifique calculando e interpretando tais resultados.
- d) Analisando as três variáveis que indicam quais serviços o cliente contratou (variáveis `PhoneService`, `OnlineBackup`, `StreamingMovies`) há indícios de que clientes que realizam *churn* tenham comportamentos diferentes de consumo comparados aos clientes que se mantém na empresa? Justifique sua resposta com base em tabelas ou gráficos adequados para cada uma das variáveis.

## VARIÁVEIS DE PAGAMENTO

- e) O valor mensal pago pelo usuário pode depender da quantidade de serviços contratados. Construa uma nova coluna na base de dados que contenha o valor pago mensal por serviço contratado, ou seja, a nova coluna deve conter o valor da divisão das colunas `MonthlyCharges` por `QtyServices`.
- f) Há diferenças no valor mensal pago por serviço (utilize a variável criada no item e) dos clientes que realizaram *churn* em relação aos que não realizaram? Justifique utilizando medidas numéricas e gráficos adequados.
- g) A variável `Contract` possui o tipo de contrato que o cliente adquiriu. Há diferenças no valor pago mensalmente por serviço (utilize a variável criada no item e) pelos clientes de cada tipo de contrato? Justifique sua resposta calculando medidas numéricas e gráficos adequados.
- h) Analisando conjuntamente as conclusões obtidas nos itens f) e g), qual variável parece justificar o valor pago mensalmente por serviço pelos clientes que realizam *churn*?

## PERFIL DO CLIENTE QUE REALIZA *CHURN*

- i) Com base nas análises realizadas nos itens anteriores, trace um perfil do cliente dessa empresa que realiza *churn*, ou seja, identifique as variáveis que parecem ser mais importantes para representar o grupo dos clientes que deixaram a empresa. Justifique indicando os cálculos e/ou gráficos envolvidos na sua resposta.
- j) Há outra análise exploratória com as variáveis presentes no conjunto de dados, além das feitas anteriormente, que ajudaria a empresa prever um possível cliente a ser *churn*? Se sim, justifique claramente sua resposta.

## INSTRUÇÕES GERAIS:

- **GRUPO:** máximo de 4 alunos.
- **DEADLINE:** 26/09/2022.
- **ENTREGA:** via link dentro do Blackboard, anexar um arquivo com extensão docx (**não enviar PDF**) e o script construído para realização da análise necessária. Para construção do relatório, atente-se as seguintes informações:
  - O trabalho deverá ser entregue no formato de relatório ou texto acadêmico, não no formato de lista de exercício. Assim, clareza, organização, formatação, fluidez no texto e ortografia serão considerados neste trabalho. Figuras, tabelas e gráficos devem ser numerados e corretamente citados no texto. Elementos não textuais que não forem analisados serão ignorados para fins de avaliação.
  - Para realizar a atividade apresente tabelas, gráficos e demais técnicas estatísticas vistas em aula, que julgarem necessários.
  - O relatório deve ser autocontido e profissional, sendo importante apresentar e analisar as informações geradas, deixando códigos e saídas coladas do R (que não sejam tabelas ou gráficos) restritas ao script (que deve ser entregue obrigatoriamente), os elementos não textuais devem estar formatados adequadamente no relatório, evitando prints de códigos e console do R (exceto de gráficos e tabelas formatadas construídas no R).
  - Gráficos, tabelas e medidas numéricas que estejam no *script* mas não estejam no relatório serão desconsiderados. Ademais, códigos de gráficos, tabelas, medidas ou outros cujo resultado não apareça no relatório também serão ignorados.
  - Atente-se para o envio correto dos DOIS documentos necessários dentro do prazo e pelo canal de envio correto (link no Blackboard). Trabalhos enviados fora de prazo e por outros canais estão sujeitos a penalização da nota.