

Cientes analisados:

561.158

Idade média de 45 anos

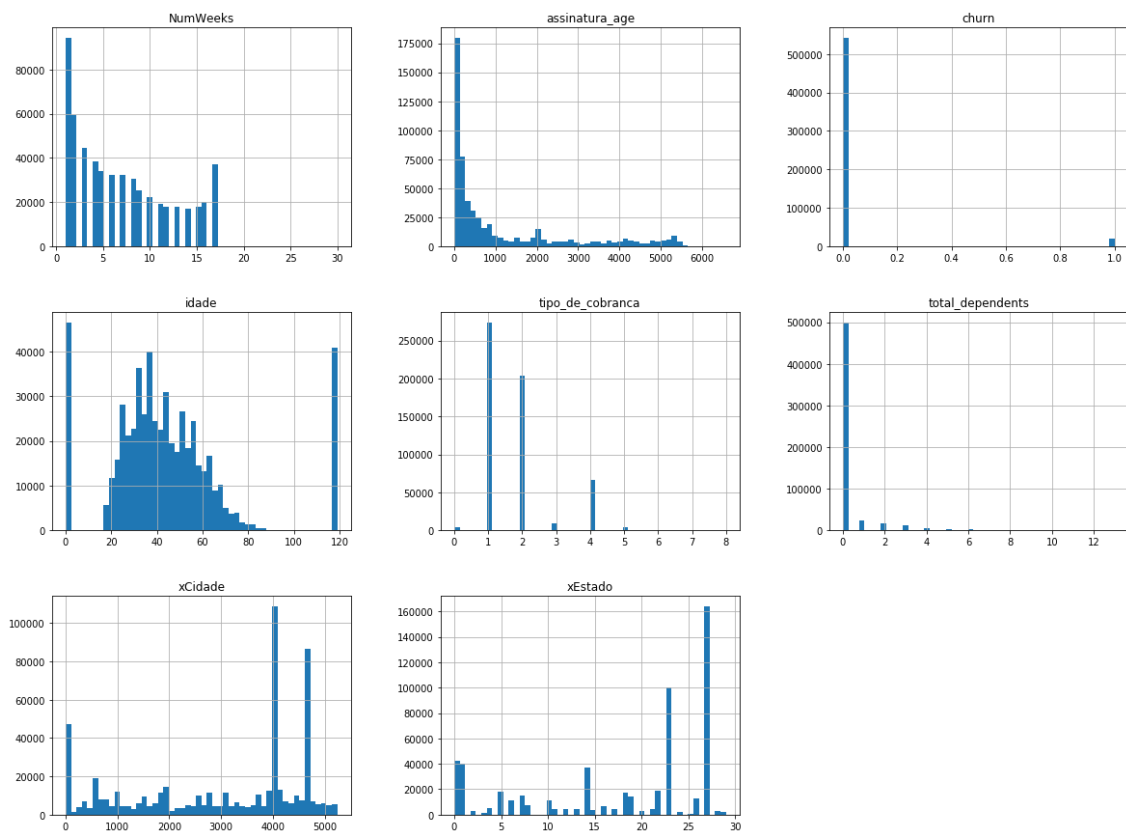
O Cartão de crédito é a forma mais utilizada de pagamento.

Na média permaneceram durante 7 semanas.

Sem dependentes.

Conforme pode ser observado no histograma para a variável Churn, existe uma grande predominância de clientes Ativos (churn=0) 96,4%

$$\text{Churn} = (19515/541643) * 100 = 3.6\%$$



Algumas técnicas para tratar datasets desbalanceados:

1- Data Sampling - Consiste na geração de novos exemplos positivos, ou na remoção de exemplos negativos, dada alguma heurística para balancear o conjunto de dados.

2- Cost Sensitive Learning - Atribui um custo para que o algoritmo aprenda que é muito pior perder um exemplo positivo do que um exemplo negativo.

3- Algorithm Adjustment - Modifica o Algoritmo de Aprendizado de Máquina para ser menos sensível a conjuntos de dados desbalanceados e sobrepostos a classes.

Foi aplicado a opção 1 de Data Sampling mas apresentou baixo rendimento.

Utilizando o dataset original, foi implementado o algoritmo de Árvore uma vez que este é o mais indicado para datasets desbalanceados, como é o caso deste trabalho.

Foi feita uma opção por totalizar as informações de clientes de tal maneira que só existisse uma única referência por cliente.

Os dados pessoais dos clientes foram desconsiderados para o cálculo do Churn.

De acordo com o algoritmo aplicado, as variáveis mais significativas para a saída de clientes foram “assinatura_age”, NumWeeks (número de semanas utilizadas) e time_spent_on_realities.

Isto indica que as primeiras semanas são decisivas para a permanência no Plano. Sendo assim, todo esforço de “encantar o cliente” deve ser feito neste momento.

Leonardo Moreira