

ENTREGA 1

Entendimento do problema e dataset

A empresa Ana Health, que nos passou o problema, é uma empresa jovem que se difere das demais do ramo de saúde por sua proposta de flexibilizar o atendimento entre doutor e paciente, oferecendo consultas e cuidado proativo através de variadas plataformas digitais e horários de atendimento. Além disso, é uma empresa que utiliza análise de dados dos usuários para que eles recebam conselhos para se manterem saudáveis, oferecendo um dinamismo e um grau de personalização que planos de saúde convencionais não têm.

O problema proposto para nós resolvermos é o diagnóstico dos motivos pelos quais os clientes podem cancelar seu serviço, ou seja, minimizar o Churn, para que a empresa conseguisse melhorar a retenção de clientes. Sabendo e tendo uma visão clara dos motivos que podem levar clientes a cancelarem um plano, a empresa consegue trabalhar nesses motivos e ter um crescimento sustentável, uma vez que manter clientes é mais barato e mais rentável do que apenas conseguir novos.

Para desenvolver uma solução para esse problema, utilizaremos técnicas de machine learning e inteligência artificial, aprendidas ao longo do semestre. Através delas, poderemos filtrar e tratar dados, identificar padrões, fazer previsões, e disponibilizar as informações mais importantes através de uma aplicação, com fácil visualização.

O dataset *Ana Health_Tabela Modelo Previsão Churn* é dividido em 73 colunas que possuem 1204 clientes, sendo esses 1204 dividido em 289 pessoas físicas, 887 organizações (836 pessoas jurídicas) e 19 acolhimento desemprego. Além disso, os clientes são majoritariamente do estado de São Paulo.

Analisando o dataset vimos que, a data de início do primeiro contrato e de término do último contrato de assinatura do benefício da Ana Health foram:

- 2020: 4 aberturas e não possui informação de encerramentos.
- 2021: 176 aberturas e 34 encerramentos.
- 2022: 508 aberturas e 283 encerramentos.
- 2023: 514 aberturas e 247 encerramentos.

Sendo contrato mais antigo aberto: 16/12/2020 e o contrato mais recente aberto: 08/11/2023. Já o contrato mais antigo encerrado foi : 19/02/2021 e o contrato mais recente encerrado foi: 08/11/2023. Além disso, 347 pessoas deram continuidade ao serviço como pessoa física após o cancelamento por parte da empresa.

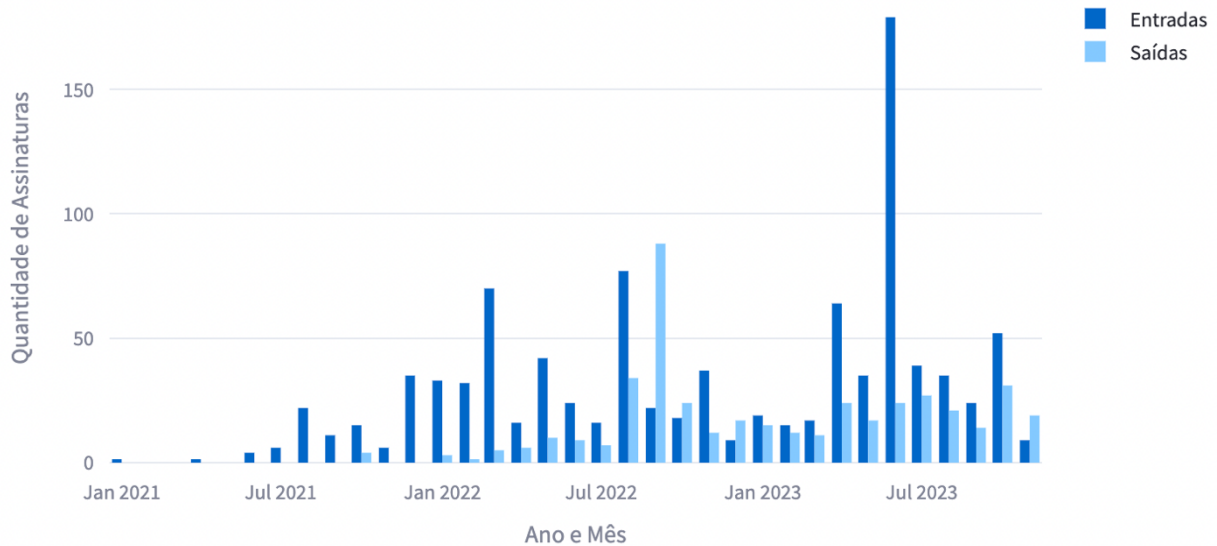
Ademais, em relação aos gêneros dos clientes, 602 clientes são do gênero masculino (64), 571 são do gênero feminino (63), 6 são outros gêneros e 24 estão sem informação (NULL). Também descobrimos que a idade média dos clientes é de 32 anos.

Sobre o funil de vendas, o status por assinatura: Cancelada (lost), Ativa (won), analisamos que a *Ana Health* teve 592 assinaturas ativas e 520 assinaturas canceladas. Sendo os motivos mais recorrentes de cancelamento: Cliente não quis seguir com a Ana, a empresa cancelou o benefício da Ana ou o cliente precisou cortar custos.

Gráficos da análise exploratória:

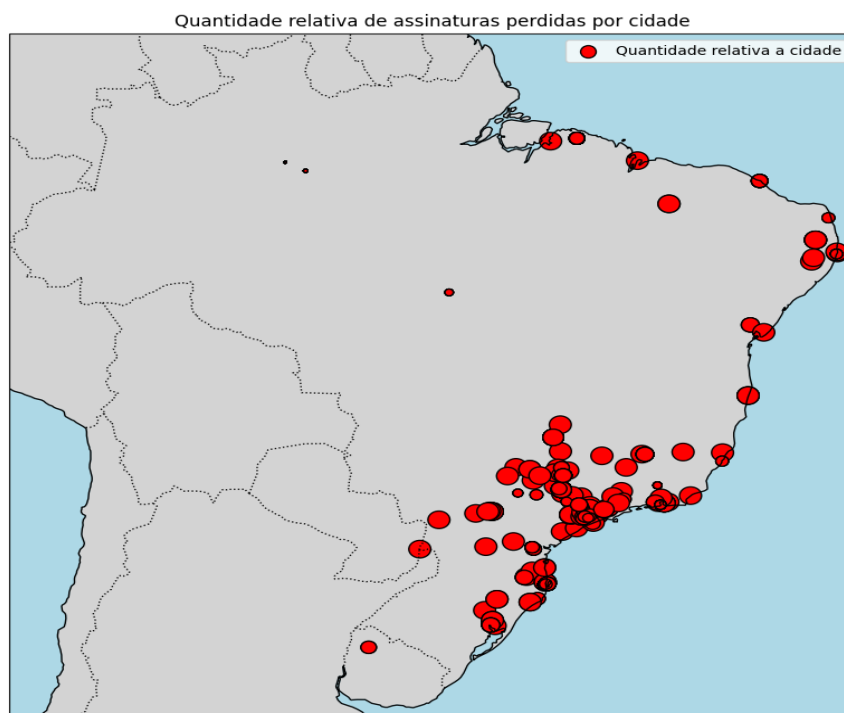
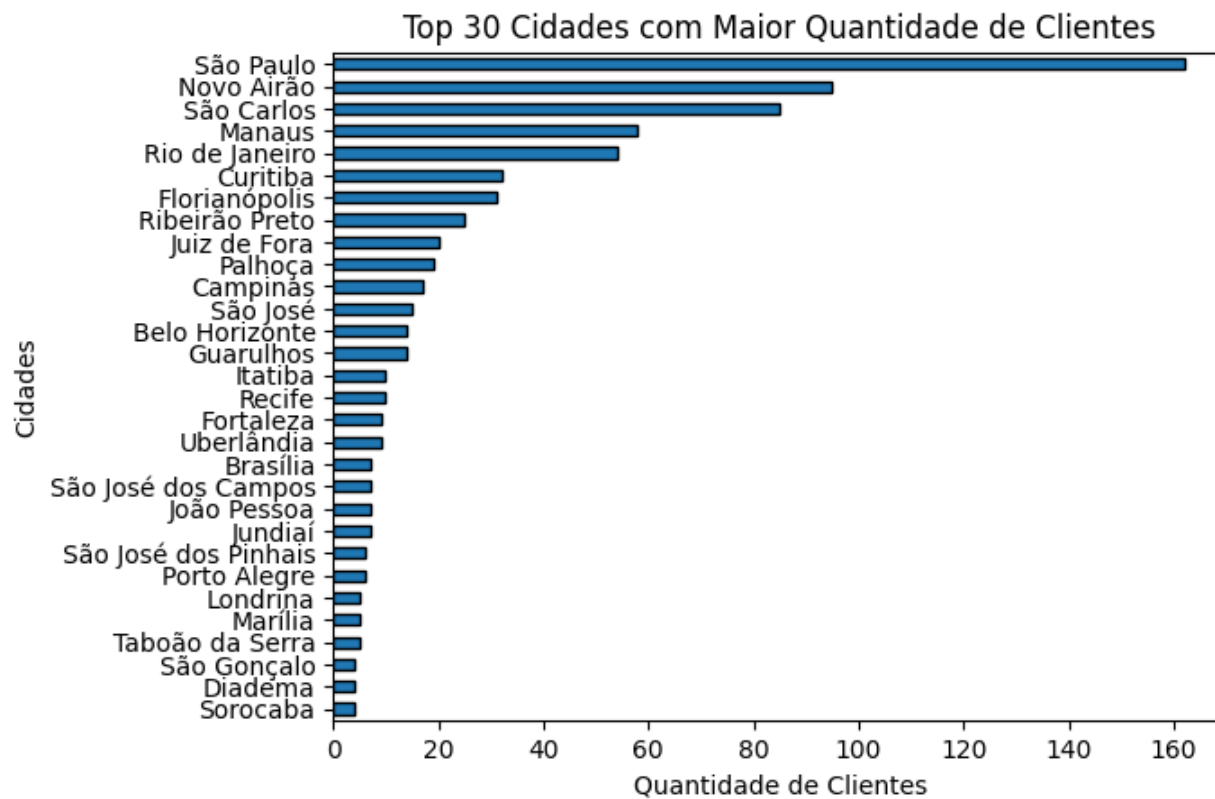
Gráfico 1:

Histograma de Entrada e Saída de Assinaturas por Mês



A imagem representa a quantidade de entradas e saídas ao longo dos meses. Sendo que Entradas significa: início do primeiro contrato de assinatura e Saídas: término do último contrato de assinatura. Dessa forma pelo histograma é possível analisar padrões interessantes, como uma grande entrada de novos contratos de assinatura em 2022-02, 2022-07, 2022-03, 2023-05, 2023-09. Assim como padrões de elevados de saída em datas específicas como em 2022-08. Uma possibilidade interessante seria comparar essas datas atípicas com eventos externos, como entrada ou saída de empresas com clientes.

Gráfico 2:



Este gráfico ilustra as perdas de assinatura em diferentes cidades do Brasil ao longo do período de 2021 a 2023. Os círculos vermelhos no gráfico representam indivíduos que cancelaram suas assinaturas. A dimensão dos círculos é proporcional ao tamanho populacional de cada cidade. Em outras palavras, se, por exemplo, houve 10 cancelamentos na cidade de São Paulo, o tamanho do círculo correspondente será

calculado como 10 dividido pela população de São Paulo. Vale ressaltar que a maioria dos cancelamentos ocorreu nas regiões Sul e Sudeste do país.