

Contexto

Você foi contratado como consultor pela empresa IDC para construir um modelo de propensão a clique em campanhas de e-mail marketing.

O objetivo é gerar um scoring para cada cliente a partir de uma data base de execução do modelo, utilizando dados históricos calculados sobre uma janela móvel. Com esse resultado, a empresa pode filtrar os clientes por score e selecionar a quantidade que deseja trabalhar.

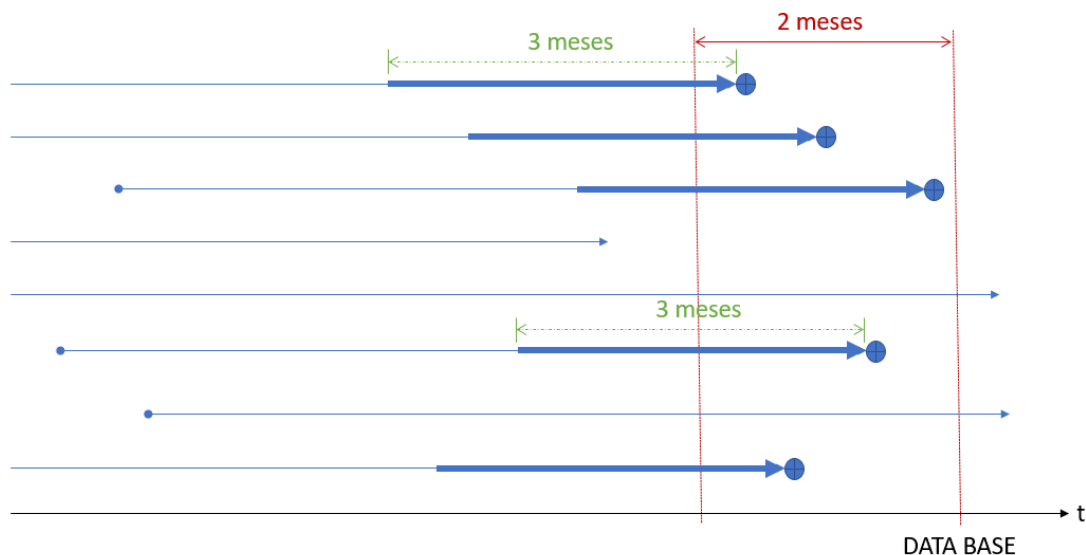
Você garantiu que o modelo é viável e que conseguirá fazer o trabalho :-p

Dados

A tabela “email_marketing.csv” foi gerada pelo pessoal de TI da empresa com todas as informações que eles têm sobre as campanhas já realizadas e sobre os clientes. Alguns dados podem estar faltando, outros podem estar errados. O pessoal de TI não consistiu estes dados; apenas gerou conforme sua solicitação.

De forma empírica, foram definidos, a partir de uma data base, uma janela dos últimos 2 meses para selecionar os clientes que foram alvos de alguma campanha, e uma janela de 3 meses, a partir desta última campanha, para avaliar o histórico de transações e interações com outras campanhas.

Esquema:



Cada linha é um cliente, sendo que as bolas azuis representam aqueles impactados por uma campanha (não precisa ser a mesma) nos últimos 2 meses. Os clientes que não foram objetos de campanha nesta janela de tempo estão representados pelas linhas sem bolas azuis. Também há clientes que deixaram a base antes destas datas analisadas.

Considerando este esquema, a tabela “email_marketing.csv” contém apenas clientes com bolas azuis. A partir desse filtro inicial, foram construídas variáveis considerando um histórico de 3 meses anteriores à última abordagem. Percebe-se que, apesar desta janela ser fixa, ela varia para cada cliente, dependendo desta data da última abordagem.

Descrição da tabela

A seguir estão detalhadas as variáveis da tabela recebida.

Variável	Tipo	Descrição
id_pess_unif	integer	id do cliente
fg_clik	smallint	flag indicando se o cliente clicou na última campanha
tpsexo	character varying(1)	sexo
tp_pess	character varying(1)	tipo do cliente
nr_idad	integer	idade
tp_estd_civil	smallint	estado civil
cd_cep5	character varying(5)	CEP do cliente com 5 dígitos
ds_uf	character varying(2)	UF do cliente
tp_resd	smallint	tipo de residência
vl_rend	numeric(11,2)	renda
vl_salr	numeric(11,2)	salário
tp_escl	smallint	escolaridade
ds_ocpc	character varying(40)	ocupação
qtd_dias_ult_camp	integer	quantidade de dias desde a última campanha enviada para o cliente
tp_camp	smallint	tipo da última campanha enviada para o cliente
tp_canl_camp	smallint	canal da última campanha enviada para o cliente
ds_orig	character varying(2)	origem da última campanha
tot_env	bigint	quantidade de campanhas enviadas para o cliente no período
tot_rec	bigint	quantidade de campanhas recebidas pelo cliente no período
tot_open	bigint	quantidade de campanhas abertas pelo cliente no período
tot_clik	bigint	quantidade de campanhas clicadas pelo cliente no período
tp_camp_open	smallint	tipo da última campanha enviada que foi aberta pelo cliente

tp_camp_clik	smallint	tipo da última campanha enviada que foi clicada pelo cliente
qt_parc	smallint	quantidade de parcelas da última transação finalizada
qtd_dias_ult_tran	integer	quantidade de dias desde a última transação finalizada
vl_tran	numeric(12,2)	valor da última transação finalizada
tp_form_pgto	smallint	tipo da forma de pagamento da última transação finalizada
tp_canl	smallint	tipo de canal da última transação finalizada
tot_per_id	bigint	quantidade total de transações finalizadas no período
sum_val	numeric	soma de todas as transações finalizadas no período

Entregas

Deverão ser entregues 2 arquivos:

1. Relatório contendo explicações OBJETIVAS do que foi feito na análise exploratória, no pré-processamento e na construção do modelo. Também devem constar os resultados obtidos (métricas) e uma análise destes resultados, mostrando, principalmente, que seu modelo funciona.
2. Um arquivo com os códigos em R (ou Python) que foram executados e utilizados para conseguir os resultados do item 1. A ideia é que, a partir da tabela fornecida, eu seja capaz de reproduzir os resultados mostrados. Desta forma, por exemplo, não é necessário incluir código de análise exploratória. Se uma variável foi eliminada por causa disso, precisa apenas constar no código a eliminação dela.