

SEND TIME OPTIMIZATION

Desafio Dito

OBJETIVO

- O desafio tem o intuito de conseguir uma solução de *Send Time Optimization* para maximizar a taxa de abertura de e-mails escolhendo a melhor hora do dia para enviá-los.

CONTEXTUALIZAÇÃO

- O intuito do desafio então é ajudar os profissionais de marketing a fornecer mensagens instantâneas que gerem mais valor pras suas marcas e seus clientes.
- Alguns clientes preferem receber boletins informativos no início da manhã, enquanto outros analisam suas caixas de entrada à noite. Do ponto de vista do remetente, é desejável entregar sempre no momento certo. Porque assim o e-mail é colocado em cima de todos os outros e tem alta visibilidade.
- Se, no entanto, a entrega for muito tarde ou muito cedo, novos e-mails recebidos moverão gradualmente a mensagem para fora da área de foco. Isso porque as caixas de entrada são tradicionalmente classificadas cronologicamente. Visibilidade mais baixa significa uma diminuição nas taxas de resposta.

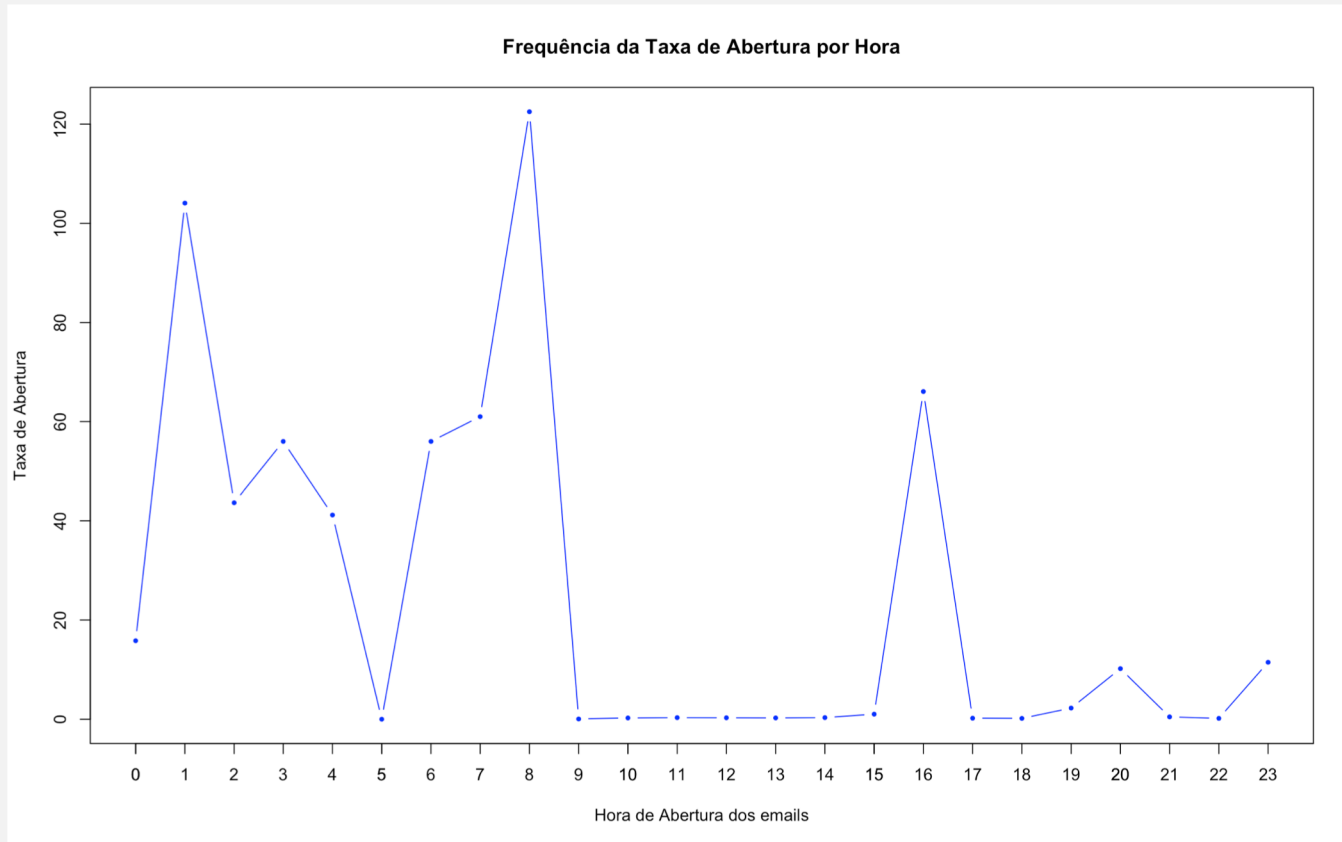
ANÁLISE DESCRITIVA

CLICOU		ABRIU		RECEBEU	
<i>Mínimo</i>	14,00	<i>Mínimo</i>	112,00	<i>Mínimo</i>	0,00
<i>1Q</i>	163,20	<i>1Q</i>	860,00	<i>1Q</i>	12,25
<i>Mediana</i>	428,00	<i>Mediana</i>	2214,00	<i>Mediana</i>	920,00
<i>Média</i>	439,30	<i>Média</i>	1966,00	<i>Média</i>	5707,12
<i>3Q</i>	737,00	<i>3Q</i>	2980,00	<i>3Q</i>	11644,25
<i>Máximo</i>	1013,00	<i>Máximo</i>	3818,00	<i>Máximo</i>	22733,00

REPORTOU SPAM		DESATIVOU		TAXA ABERTURA	
<i>Mínimo</i>	0,00	<i>Mínimo</i>	3,00	<i>Mínimo</i>	0,00
<i>1Q</i>	1,00	<i>1Q</i>	9,75	<i>1Q</i>	0,26
<i>Mediana</i>	3,00	<i>Mediana</i>	25,00	<i>Mediana</i>	-
<i>Média</i>	4,21	<i>Média</i>	33,08	<i>Média</i>	-
<i>3Q</i>	7,25	<i>3Q</i>	53,75	<i>3Q</i>	46,73
<i>Máximo</i>	11,00	<i>Máximo</i>	84,00	<i>Máximo</i>	122,50

- Pode-se observar pela análise descritiva que a média de pessoas que abriram e-mails é de quase 2 mil pessoas;
- Pode-se visualizar também que a Taxa de Abertura máxima de e-mails é de 123;
- Observa-se que a média de quem desativa os disparos é de 33 pessoas, dentro de um universo de mais de 24 mil presentes na base.

ANÁLISE DESCRITIVA



- Pode-se observar pelo gráfico que por frequência, entre 8:00 e 8:59, a taxa de abertura de e-mails, que é representada pelo cálculo de e-mails abertos dividido pela quantidade de recebidos, é a maior dentre todas as horas disponíveis, mostrando o tempo de envio ideal neste pico;
- Observa-se também que a taxa é quase irrelevante em relação aos disparos feitos entre 9:00 e 14:59 e 17:00 e 23:59.

MODELAGEM

Razão de Chances			
1hrs	1,4108	13hrs	0,07754
2hrs	1,2981	14hrs	0,09419
3hrs	1,442	15hrs	0,25231
4hrs	1,2832	16hrs	1,25108
5hrs	1,5991	17hrs	0,06042
6hrs	1,7456	18hrs	0,04964
7hrs	1,5053	19hrs	0,43873
8hrs	1,6123	20hrs	0,90645
9hrs	0,0146	21hrs	0,1301
10hrs	0,0745	22hrs	0,05104
11hrs	0,0923	23hrs	0,9953
12hrs	0,0852		

- Para se obter um modelo estatisticamente melhor, é necessário testar vários tipos de técnicas para escolher aquela que melhor se adequa aos dados e ao pedido do desafio. Como obtive um tempo reduzido, optei por trabalhar com regressão logística. Porém, há também outros métodos muito eficazes para se calcular isso, não só o modelo logístico;
- A razão de chances do modelo logístico identifica que nos horários entre 5:00 e 8:59 tem-se pelo menos 1,5 de chance a mais do cliente abrir os e-mails recebidos;
- O modelo ainda identifica que, entre 6:00 e 6:59, seria o horário ideal para o disparo dos e-mails.

CONCLUSÕES

- De acordo com o pedido do desafio, o modelo segue um método de *machine learning* o que significa que o mesmo aprende a partir do momento que é incluído novos dados;
- O método nos diz que a chance de um indivíduo abrir um e-mail que chega entre 6:00 e 6:59 da manhã é 1,75 maior que nos outros horários. Diante disso, o certo é explorarmos este horário a fim de maximizar a taxa de abertura, como sendo o ideal neste cenário. Obviamente, analisando de maneira mais profunda o horário poderá ser alterado visando, por exemplo, dias da semana, meses do ano, dentre outros.
- Com o modelo feito, há condições de verificar os preditores do mesmo, porém estes estão disponíveis no script;
- Pode-se identificar também outros tipos de modelo para teste a fim de aprimorar as análises;
- O modelo logístico usado, foi devidamente validado nas pressuposições que o diz respeito;
- Devo salientar a importância de dados bem estruturados, assim como foi entregue pela Dito, isto deixa a análise muito mais fácil com um resultado muito mais adequado.