

Data Challenge

Cortex Intelligence

Realize o exercício número 1 e escolha pelo menos um dentre os demais.

Quanto mais melhor :)

1. Propagação de erro (Obrigatório)

Um cliente contrata um projeto para análise de sentimento e alcance em conteúdos que o envolvem (ex. notícias, blogs, tweets), onde a Cortex deverá classificar manualmente o material encontrado em *promotor* ou *detrator*. Como a mão de obra é limitada, não será possível classificar 100% das matérias que saem a cada semana envolvendo o cliente – somente 100 (cem) notícias por semana são classificadas pela equipe responsável. Com isso foi determinada a seguinte abordagem:

- Os conteúdos são rankeados de acordo com o seu *Alcance* (repercussão do conteúdo);
- No grupo com os 50 maiores alcances, todos são classificados manualmente;
- No grupo com o restante, 50 são selecionados aleatoriamente para classificação manual;
- A classificação determina se cada conteúdo é considerado *promotor*, *detrator* ou *neutro*;
- Calcula-se o *Alcance Promotor* e *Alcance Detrator* dentro de cada grupo, multiplicando-se o Alcance Total pelo percentual de notícias promotoras ou detratoras:
 $[ALC_{prom(i)} = ALC_{tot(i)} * \%_{prom(i)}]$, onde (i) é o grupo top 50 ou restante;
- O Alcance Promotor Total é a soma do Alcance Promotor dentro de cada grupo $[ALC_{prom} = ALC_{prom(top50)} + ALC_{prom(rest)}]$. De forma análoga, $[ALC_{det} = ALC_{det(top50)} + ALC_{det(rest)}]$.
- Calcula-se o Índice Promotor $[I_{pr}]$, que é a razão entre o Alcance Promotor Total e o a soma dos alcances totais promotor + detrator: $[I_{pr} = ALC_{prom} / (ALC_{prom} + ALC_{det})]$.

O gestor precisa informar ao cliente o nível de precisão do indicador apresentado e te pede ajuda para calcular a margem de erro do Índice Promotor de acordo com o método acima. Você tem acesso a uma amostra de 600 conteúdos, na qual a equipe classificou manualmente 100 conteúdos de acordo com a tabela abaixo:

	Top 50	Restante
Número de conteúdos	50	550
Classificadas manualmente	50	50
Conteúdos promotores	15	17
Conteúdos detratores	8	7
Conteúdos neutros	27	25
Alcance total	1,000,000	3,000,000

Calcule a margem de erro do Índice Promotor para comunicar ao gestor, utilizando as ferramentas que julgar necessárias. Em sua opinião, a metodologia utilizada é ideal para a redução da margem de erro? Ou seria melhor classificar as 100 notícias aleatoriamente, sem separar o top 50, por exemplo?

2. Marketing Mix Modeling

Uma grande rede de varejo brasileira deseja analisar a efetividade de seus investimentos em marketing, para um de seus produtos. Essa empresa costuma investir em 3 tipos de mídia: Jornal impresso, televisão e campanhas de email.

Para cada tipo de mídia há um gasto semanal recorrente para manutenção da visibilidade. Além disso, a empresa também faz campanhas esporádicas que aumentam esse gasto em determinados períodos, aumentando assim o alcance de suas campanhas.

Você foi contratado para um estudo de Marketing Mix Modeling, com o objetivo de explicar o retorno desses gastos com cada tipo de mídia. Você deverá analisar a elasticidade da receita em função do preço do produto e dos investimentos, fornecendo informações acerca de quais seriam os níveis ideais em cada canal de forma a maximizar o retorno. Suponha que a relação GRP x Investimento com propaganda seja perfeitamente linear onde:

Radio \rightarrow $GRP = 0.010 \times \text{Investimento [em Reais]}$

DM \rightarrow $GRP = 0.008 \times \text{Investimento [em Reais]}$

TV \rightarrow $GRP = 0.014 \times \text{Investimento [em Reais]}$

Para isso, foi disponibilizada uma base histórica ([mix_marketing_modeling.csv](#)), para uma das linhas de produto, com as seguintes informações (campos):

- Data (por semana)
- Preço do produto (em Reais)
- DM-GRP (Direct Message)
- Radio-GRP
- TV-GRP
- Receita (em 10k de Reais)

Faça uma análise exploratória da base e prepare uma breve apresentação, contemplando os seguintes pontos:

- 1) Explique seu entendimento sobre MMM, qual a importância, os prós e contras de um estudo como esse
- 2) Análise exploratória da base, com tendências e insights que julgar relevante
- 3) Elasticidade do investimento em cada canal, considerando o alcance da campanha em termos da audiência e em termos monetários
- 4) Análise da robustez dessas elasticidades, mostrando a significância estatística desses retornos
- 5) Recomendação final para o cliente, como os investimentos podem ser melhor aproveitados

Você pode utilizar as ferramentas que achar melhor para a análise dos dados (R, Python, etc).

3. Previsão de Vendas

Uma empresa do ramo de bens de consumo está com dificuldade de prever a venda de seus produtos para mês seguinte. Dado um histórico diário de vendas faça:

1. A previsão mensal para os próximos para 3 meses.
2. Bônus 1: Parametrizar para N meses à frente
3. Bônus 2: Ter a previsão no nível mais granular possível
4. Bônus 3: Ter a previsão dia a dia para do mês seguinte.
5. (Esses bônus podem ser feitos independentes e não estão em ordem de dificuldade. Faça os que você quiser)

A base de vendas está organizada em nível hierárquico (Distribuidor > Gerência > Vendedor). Podem existir campos em branco, fique à vontade para substituir os campos vazios por “-” mas se atente à hierarquia.

A categorias com a mesma letra (ex: A e AA) são semelhantes. Categorias com letras diferentes são diferentes (ex: A e B ou A e M).

Não existe diferenciação de importância (ex: AA é tão importante quanto A e quanto FFF)

4. Agrupamento de Eventos

Uma empresa do ramo financeiro contratou um projeto para captura de notícias relacionadas ao setor em mídias online e impressas. A Cortex fará a captura dessas notícias nas fontes disponíveis e as disponibilizará no feed do cliente.

Como muitas notícias publicadas se referem ao mesmo evento, a ideia é criar um algoritmo que possa ser capaz de agrupá-las dentro de um “Assunto específico”.

OBS: Nem sempre um evento é restrito a um dia. Mas ele também não fica em voga Ad infinitum. Use isso ao seu favor.

Exemplo abaixo:

Assunto específico	Publicação ↕
Itaú comprou empresa Recovery, do ...	7
BTG analisa influência do Coronaviru...	6
BTG analisa a empresa: Oi	4
Oi Móvel pode ser dividida entre Tim ...	4
BTG afirma que nova geração aposta ...	3
BTG analisa a empresa: Cielo	3
BTG coordena oferta da Compesa	3
BTG divulga dados dos maiores bancos	3
Matéria cita evento do BTG de 2019	3

A base contém quatro colunas ID da notícia (registro único), Data, Título da Notícia, Conteúdo.

O seu output será apenas a adição de mais uma coluna chamada “Assunto específico”. Esta coluna não precisa ter um descrição como acima. Ela pode ter apenas um identificador (1 ... N) do grupo que cada notícia pertence.

Bônus: Bolar um jeito de criar uma string descrevendo cada grupo.

**Caso encontre qualquer inconsistência remova-a da base e indique qual foi. Se não encontrar nenhuma, tudo bem, não fique tentando achar. (é mais provável que não tenha mesmo).*

5. Classificação de Notícias

Uma empresa da indústria de celulose está interessada em monitorar as novas publicações dentro de seu ramo. Dado o alto volume de publicações diárias, não é possível que o time atual dê conta de ler tudo. Para isso você foi escalado para criar um algoritmo que, baseado no conteúdo do texto, seja capaz de identificar se a matéria é de interesse ou não da empresa.

A empresa deixou claro que é preferível ter publicações irrelevantes classificadas como relevantes do que perder publicações de interesse para empresa.

Nenhuma informação prévia (como palavras chaves) nos foi fornecida. Apenas contamos com as marcações sobre a relevância do conteúdo, 1 para publicações que são de interesse da empresa e 0 para as que não são.

A base contém 3 colunas: ID (registro único), Conteúdo, Relevante.

**Caso encontre qualquer inconsistência remova-a da base e indique qual foi. Se não encontrar nenhuma, tudo bem, não fique tentando achar. (é mais provável que não tenha mesmo).*

Boa sorte, faça com calma, não se estresse, e aproveite o Carnaval.

;)

Fontes que podem ser úteis para estudo

<http://www.catalystinc.com/wp-content/uploads/2015/05/Marketing-Mix-Redesign.pdf>

<https://www.themarketingtechnologist.co/the-gam-approach-to-spend-your-money-more-efficiently/>