

# Analista BI - Dito

Esse desafio está dividido em três partes, sendo a primeira delas um problema que envolve SQL, a segunda Análise e Modelagem de dados e a terceira um Storytelling.

Em todas elas é esperado que você envie as soluções através de um repositório público no [Github](#).

## 01 - SQL

### Contexto

Na Dito utilizamos o [Google BigQuery](#) como fonte para diversas análises e os dados estão estruturados em uma única tabela.

As linhas representam eventos e possuem a coluna `type` que pode conter os valores `"track"` e `"identify"`.

### Eventos de "identify"

As linhas com `type = "identify"` são eventos que marcam a alteração de características das pessoas como nome, email e telefone. Para uma mesma pessoa podem existir vários eventos de identify ao longo do tempo, que correspondem às diversas vezes que essa pessoas foi identificada ou teve seus dados atualizados na Dito.

**Por exemplo:** uma pessoa pode mudar seu telefone ao longo do tempo, portanto o telefone atual de uma pessoa corresponde ao registro do tipo `"identify"` mais recente onde a coluna telefone esteja preenchida.

As características das pessoas estão guardadas no objeto `traits`. Então acessar nome e email seria: `traits.name` e `traits.email` respectivamente.

### Eventos de "track"

As linhas com `type = "track"` representam ações das pessoas. Um evento do tipo "track" pode ser uma compra, um login no site ou a navegação online em uma página de produto. As ações desses eventos são diferenciados pela coluna `properties.action`.

No contexto desse exercício, apenas o evento `"buy"` existe nos dados.

Os atributos dos eventos estão guardados no nó `properties`.

**Por exemplo:** a receita gerada a partir de uma compra pode ser acessada pelo campo `properties.revenue`.

### Timestamp dos eventos

Tanto os eventos `"identify"` quanto os eventos `"track"` possuem uma coluna chamada `timestamp`, que corresponde à data/hora que de fato aconteceu a identificação do usuário ou a ação. Ambos os tipos possuem a coluna `id` que identifica uma pessoa

### Preparação

Antes de tudo você deve criar uma conta [Google Cloud](#). O processo é bem rápido e gratuito, sendo necessário apenas estar logado com algum email Google.

O próximo passo é criar um projeto Google Cloud [nesse link](#).

Nesse momento você já consegue acessar o Dataset público criado pela Dito para esse desafio. Basta abrir o [Console do BigQuery](#) e fazer a seguinte consulta de teste, que lista os 5 produtos que mais geraram receita:

```
SELECT
properties.product,
SUM(properties.revenue) revenue
FROM `dito-data-scientist-challenge.tracking.dito`
WHERE type = 'track'
GROUP BY properties.product
ORDER BY revenue DESC
LIMIT 5
```

### Desafio

O desafio será responder duas perguntas a partir do Dataset fornecido e **usando apenas SQL**.

1. Qual o nome, email e telefone das 5 pessoas que mais geraram receita?
2. De quantos em quantos dias, em média, as pessoas compram? Use a mediana como média.

Você deve **entregar as consultas SQL** das duas perguntas (não apenas as respostas) e uma breve explicação de como você pensou.

## 02 - Análise e Modelagem de Dados

### Contexto

A Dito é uma ferramenta de automação de Marketing e CRM que ajuda as empresas comunicarem de forma personalizada e eficiente com seus clientes. Essas empresas utilizam nossa plataforma para segmentar seus clientes de acordo com perfis de atuação desejados, criar campanhas de Marketing, enviar notificações em vários canais e gerenciar o ciclo de vida dos clientes.

Nessa segunda parte do desafio, você é um Analista responsável por projetos de BI e precisa propor uma solução para um Data Mart de Campanhas, para que a Dito e seus clientes consigam acompanhar o retorno e efetividade das campanhas executadas.

Abaixo estão detalhados alguns campos encontrados em eventos relacionados a disparo de emails:

- **id:** identificador do usuário
- **timestamp:** data/hora que aconteceu o evento, no formato 2018-05-25 14:59:02 UTC
- **email\_id:** identificador que diferencia disparos de e-mail
- **action:** ação que representa o evento
  - **received:** usuário recebeu o e-mail
  - **open:** usuário abriu o e-mail
  - **click:** usuário acessou algum link presente no e-mail
  - **spamreport:** usuário reportou spam
  - **unsubscribe:** usuário se descadastrou

Além de emails, existem outros canais de comunicação como: Notificações em aplicativos (App push), Whatsapp, SMS, Notificações em Browser, etc.

### Desafio

1. Proponha um **modelo conceitual** para o Data Mart de Campanhas baseado na estrutura dos dados brutos detalhada nas questões anteriores, com a(s) tabela(s) fato(s), dimensão(ões) e métrica(s) que julgar necessário para gerar análises relevantes desse assunto.

2. Após isso, sugira **3 análises relevantes** que possam ser extraídas das métricas e modelo sugerido, além de uma breve explicação de como isso pode gerar valor para a Dito e seus clientes.

O intuito desse desafio é entender como você transformaria os dados brutos apresentados (e/ou outros dados não apresentados mas que julgaria necessário no domínio de Campanhas de Marketing Digital) em um Data Mart estruturado com métricas relevantes para o negócio =)

## 03 - Storytelling

### Desafio

#### **Transforme o seu currículo em um Storytelling!**

Como você apresentaria a sua trajetória profissional em um formato **Storytelling**?

Pode usar a ferramenta que quiser (Excel, PowerPoint, Google, ferramentas de visualização de dados, etc), o importante é conseguirmos avaliar habilidades de comunicação utilizando dados.

#dito/careers

#challenge