

## Teste Prático

Junto com este descritivo, você está recebendo 3 arquivos com dados aleatórios de vendas de 2017 a 2019.

Para a execução deste teste, você pode utilizar as ferramentas que estiver mais familiarizado, seguindo apenas as seguintes premissas:

1. Os dados necessariamente devem ser armazenados em tabelas de banco de dados (MySQL, PostgreSQL, BigQuery, MS SQL, Oracle etc) e não em arquivos ou planilhas;
2. Você deve necessariamente utilizar as linguagens **SQL e Python** nos processos de carga, consulta e transformação dos dados;
3. Utilizar uma ferramenta que lhe permita criar os processos de ETL ou DAG's para ingestão e transformação de dados;
4. Você deve implementar um controle de versionamento para seus códigos.

Obs. Embora você possa utilizar ferramentas de sua escolha, estamos estruturando nossa plataforma de dados na nuvem do Google, se você realizar as implementações utilizando ferramentas do GCP, ficaremos mais felizes ainda! #ficaadica ;-)

### Você precisará:

1. Realizar a importação dos dados dos 3 arquivos em uma tabela criada por você no banco de dados de sua escolha;
2. Com os dados importados, modelar 4 novas tabelas e implementar processos que façam as transformações necessárias e insiram as seguintes visões nas tabelas:
  - a. Tabela1: Consolidado de vendas por ano e mês;
  - b. Tabela2: Consolidado de vendas por marca e linha;
  - c. Tabela3: Consolidado de vendas por marca, ano e mês;
  - d. Tabela4: Consolidado de vendas por linha, ano e mês;
3. Criar uma conta de acesso comum ao Twitter
4. Acessar <https://developer.twitter.com/en/apply-for-access> e criar uma conta de desenvolvedor;
5. Após criar a conta, acesse o menu App e crie um aplicativo. Para fins de testes, os dados não precisam ser precisos. A URL por exemplo, pode colocar <http://localhost.com>
6. Após criado o app, acesse o mesmo e gere o token e token secret
7. Criar um processo de captura de dados através da API do Twitter, que utilize os seguintes parâmetros:
  - a. Palavras a serem pesquisadas: "Boticário" e o nome da linha com mais vendas no mês 12 de 2019 (conforme item 2.d.);
  - b. Recuperar os 50 twitts mais recentes;
  - c. Recuperar apenas twitts que estejam em português.
8. Criar um processo que salve os nomes dos usuários e o texto dos twitts recuperados em uma tabela do banco de dados.

### O que esperamos de você?

1. Que consiga realizar todas as etapas acima;
2. Que gere um repositório com todos os scripts e nos disponibilize para consulta;
3. Que nos demonstre todos os processos rodando e as tabelas sendo carregadas. Marcaremos uma reunião virtual para que possa demonstrar;
4. Que nos apresente como se organizou para realizar as tarefas;