



**escola
britânica de
artes criativas
& tecnologia**

Módulo | SQL: Projeto Final

Caderno de Aula

Professor [Mariane Neiva](#)

▼ Tópicos

1. Introdução aos Kaggle;
2. Exploração de dados e análise de dados (EDA)
3. Storytelling

1. Plataforma Kaggle

História

O [Kaggle](#) foi fundada em 2010 por [Anthony Goldbloom](#) e adquirida em 2017 pelo Google.

O que é?

É uma comunidade de programadores (principalmente voltada pra Data Science) que possui diversas bases de dados, competições, códigos e troca de informações entre desenvolvedores. Existem inclusive competições com prêmios e patrocinadas por grandes empresas.

Por que utilizar?

Plataforma mundialmente conhecida, você pode construir seu portfólio por meio de notebooks e descobrir novos datasets. Você também pode utilizar o Kaggle para obter inspirações e ver como outros cientistas de dados/analistas desenvolver suas análises e insights

Prática:

Vamos explorar o Kaggle? Faça sua conta [aqui!](#)

2. Exploração e análise de dados

Por quê?

- Para entender o que temos na mão, precisamos explorar os dados.
- Para ter insights sobre os dados, precisamos analisá-los.

Muitas vezes, é a partir da análise exploratória inicial que conseguimos desenvolver as perguntas de negócio, o ponto chave para a análise de dados!

O que é?

A exploração de dados é uma investigação das principais características dos dados. Nessa fase, você pode entender:

- variáveis individuais
- relação entre variáveis
- quais variáveis são mais importantes
- quais variáveis têm valores nulos
- como minimizar possíveis erros
- busca de outliers

Na análise de dados, fazemos as perguntas aos dados e utilizamos ferramentas de matemática/estatística para respondê-las. Você pode utilizar funções como:

- análise de distribuição dos dados
- valor médio/desvio
- padrões (análise preditiva)
- maiores/menos vendas dado categorias/condições

Pontos chaves para não esquecer

- Você pode utilizar o Kaggle para ter ideias de como fazer a EDA e a análise de dados
- *Garbage in, garbage out* -
- Sempre que puder, visualize graficamente os dados (Você pode utilizar uma linguagem de programação ou o CSV com o retorno do SELECT no EXCEL)
- Não há necessidade de usar todos os dados que você tem! Explore as informações, separe as perguntas e use as colunas certas

- As vezes é preciso coletar mais dados.
- 80% exploração, limpeza e tratamento, 20% análise
- Não deixe de documentar tudo (falaremos na próxima aula)

Prática: vamos para nosso projeto de EDA + Análise?

3. Storytelling

Não basta apenas codificar, é importante entender, documentar e contar os resultados.

O processo de comunicação dos resultados é chamada de **storytelling**.

Mas não é apenas falar os comandos, é sobre **contar uma história**.

É como se você tivesse contando uma história para uma criança mesmo: como problemas, emoção, com descobertas e a mágica no final. [A jornada do herói!](#)

Em análise de dados, você pode contar com as ferramentas que você tiver: slides, texto, animações, dashboards, etc.

Alguns pontos importantes para lembrar no storytelling que tem tudo a ver com a exploração e análise de dados. Apesar das histórias de contos de fada serem floreadas com as cores e detalhes de cada planta na floresta, nem tudo aqui é relevante.

Escolher o que mostrar e como mostrar pode fazer seu insight ser ou não convincente para quem está ouvindo: ou seja, encontrar o equilíbrio entre o que colocar e o que não colocar na sua história de dados é importante!

Para aprofundar, o livro [Storytelling with data](#) nos auxilia a escolher os melhores gráficos para contar os resultados e exibi-los da forma mais apropriada!

Na prática: Depois de toda EDA e Análise de dados feita com os dados da parte prática do módulo, conte cada um dos passos em um notebook Kaggle! Você pode inserir gráficos, figuras, explicar o passo-a-passo e o que for necessário para encantar os leitores (ou os recrutadores!).

Link para o projeto feito na aula [aqui](#)