

Unidade 01 – Introdução à Ciência de Dados

1. Identifique onde datasets (conjunto de dados) interessantes relevantes para os seguintes domínios podem ser encontrados na web:

- Livros

Kaggle: <https://www.kaggle.com/datasets/diegomariano/tabela-de-livros>

- Corrida de Cavalos

Horse Racing Datasets: <https://horseracingdatasets.com/>

- Preços de Ações no Mercado Financeiro

Banco Central do Brasil: <https://dadosabertos.bcb.gov.br/it/dataset/22910-sgs>

- Riscos de doenças

World Health Organization: <https://www.who.int/data/gho>

- Colégios e Universidades

<https://anonymousandata.inep.gov.br/analytics/saw.dll?Portal>

- Taxa de Criminalidade

<https://www.ipea.gov.br/atlasviolencia/>

Para cada uma destas fontes de dados acima, explique o que você poderia fazer para deixá-los em um formato utilizável no seu computador para análise.

2. Para cada uma das questões abaixo faça o seguinte:

- a) produza uma estimativa rápida com base apenas em sua compreensão de mundo e, em seguida,
- b) use o Google (ou Chat-GPT) para encontrar números aceitáveis para produzir uma estimativa mais baseada em princípios. Após, compare as duas estimativas em termos de quanto elas diferem entre si?

- Quantos afinadores de piano existem no mundo inteiro?

a) 100 mil

b) Estimativa de 30 mil

- Quantos postos de gasolina existem no Brasil?
 - a) 150 mil
 - b) 43 mil

- Quantas pessoas entram e saem do Aeroporto Internacional de Guarulhos todos os dias (antes da pandemia)?
 - a) 50 mil
 - b) Estimativa de 118 mil

- Quantos peixes existem em todos os oceanos do mundo?
 - a) 100 bilhões
 - b) 4 trilhões

- Quantas quilômetros de estradas pavimentadas existem no Brasil?
 - a) 100 mil km
 - b) Estimativa de 213 mil km

- Quantos gatos vivem na cidade de Joinville?
 - a) 100 mil
 - b) Estimativa de 37 mil gatos

- Quantas contas correntes existem no Brasil?
 - a) 140 mil
 - b) Estimativa 500 milhões