

Universidade Federal de Alagoas

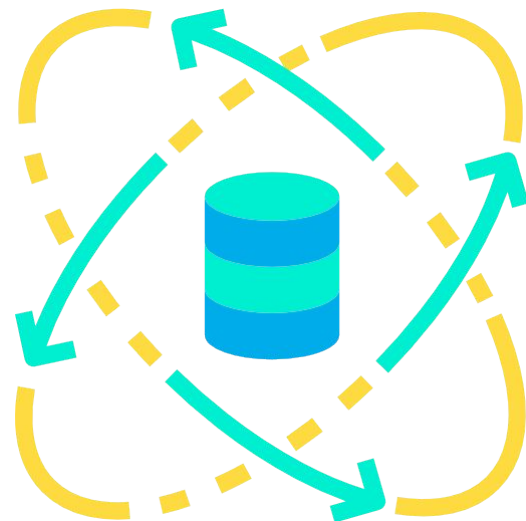
Ciência de Dados

Curso de Inverno



{kat;e}

28 de Julho 2022



Cronograma e Conteúdo

- 25/07/2022, 13:00h ~ 16:00h - Introdução à Ciência de Dados + Introdução ao Python para Ciência de Dados (Assíncrona);
- 26/07/2022, 13:00h ~ 16:00h - Inteligência Artificial e Aprendizado de Máquina + Introdução ao Python para Ciência de Dados (Assíncrona);
- 27/07/2022, 13:00h ~ 16:00h - Inteligência Artificial e Aprendizado de Máquina + Biblioteca Pandas e Numpy (Assíncrona);
- 28/07/2022, 13:00h ~ 16:00h - Visualização de Dados + Visualização de dados com Python (Assíncrona);
- 29/07/2022, 13:00h ~ 16:00h - Projeto Prático;
- 30/07/2022 até 02/08/2022 - Projeto Real na prática com Python (Assíncrona) + Entrega de atividades (Classroom)
- 03/08/2022, 13:00h ~ 16:00h - Apresentação do Projeto.

0 Time

Ministrantes

- Geymerson Ramos;
- John Omena.

Monitores

- Rebecca Brandão;
- Lilian Gisely.

Acesso à Sala de Aula

- Código do Google Sala de Aula: zzfsaug

Geymerson Ramos

FORMAÇÃO

- Mestrado em Informática, UFAL (2021)
- Bacharel em Engenharia de Computação, UFAL (2019)

ÁREAS DE INTERESSE

- AI/Machine Learning, Otimização, Internet das Coisas, Cidades Inteligentes.

ATIVIDADE ATUAL

- Gerência Técnica no Programa de Residência em Ciência de Dados LaCCAN/SECTI-AL.

INFORMAÇÕES ADICIONAIS

- E-mail: geymerson@laccan.ufal.br
- LinkedIn: <https://www.linkedin.com/in/geymerson-ramos-477267160/>
- Lattes: <http://lattes.cnpq.br/1615075725691676>

Conteúdo do Dia

- Introdução à Visualização de Dados
 - O Que é Visualização de Dados;
 - Ferramentas de Visualização;
 - Tipos de Visualização;
 - Os tipos de Gráfico;
- Boas e Más Práticas;
- Hands On!

{

Introdução à Visualização de Dados

}

O que é Visualização de Dados

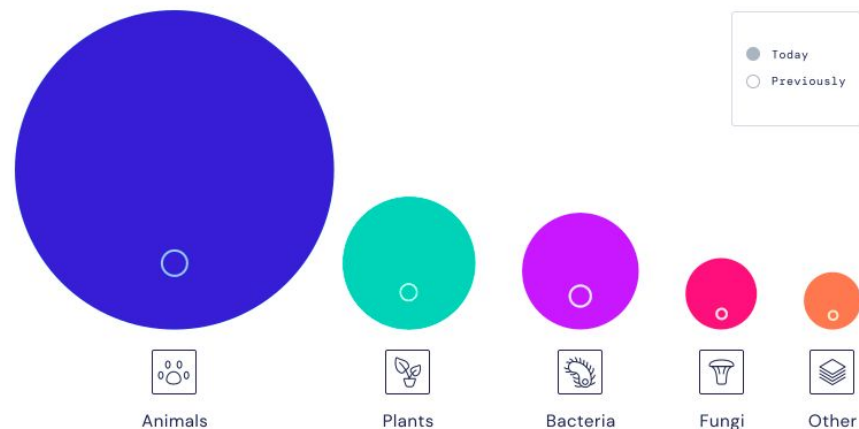
- Leitura e explicação de dados de maneira visual;
- Combinação de dados, criatividade e *Storytelling*;
 - *Storytelling* deve amplificar a compreensão e interpretação do dados;
- Auxilia na tomada de decisão.

Representação do impacto do projeto AlphaFold na base de dados proteínas

- Cientistas utilizam IA para construir a estrutura de quase todas as proteínas conhecidas e catalogadas pela ciência.
- A base de dados hoje contém mais de 200 milhões de estruturas.

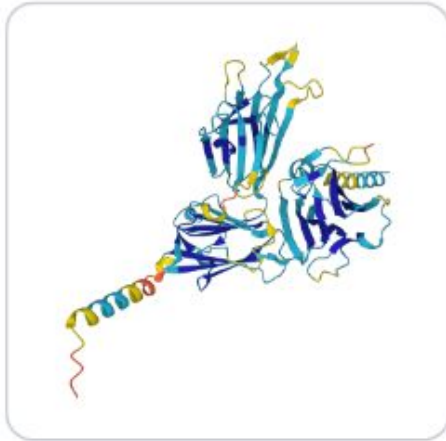
Number of species represented in AlphaFold DB

Total increase from ~10K to ~1M



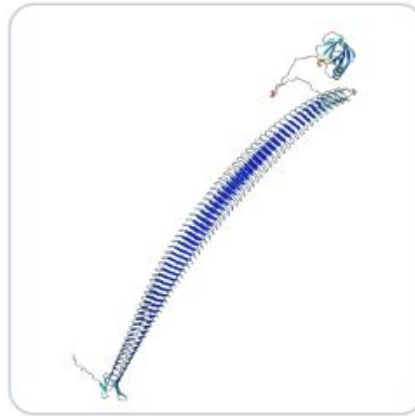
Fonte: [AlphaFold reveals the structure of the protein universe.](#)

Algumas Estruturas de Proteína Construídas pela IA por Meio de Predição.



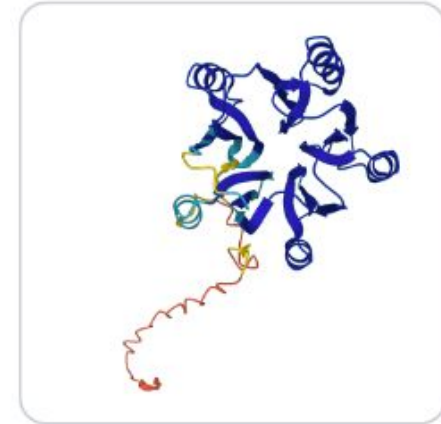
Gametocyte surface
protein P45/48

From the malaria parasite;
a candidate protein for
including in vaccines



Ice nucleation protein

Bacterial protein that can trigger
ice formation at relatively high
temperatures, causing frost
damage to plants



F20H23.2 protein

Plant protein; represents
a potential new structural
superfamily unlike anything
seen before

Ferramentas de Visualização

- Microsoft Power BI;
- Tableau;
- Qlik Sense;
- Klipfolio;
- Looker;
- Zoho Analytics;
- Domo.

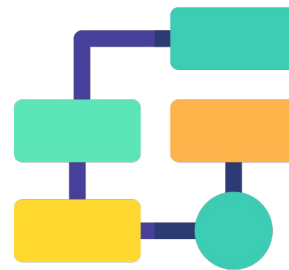
Tipos de Visualização



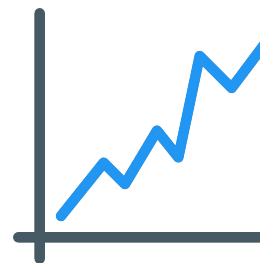
Infográficos

Title 1	Title 2	Title 3
Value 1	Value 2	Value 3
Value 4	Value 5	Value 6
Value 7	Value 8	Value 9
Value 10	Value 11	Value 12

Tabelas



Diagramas



Gráficos

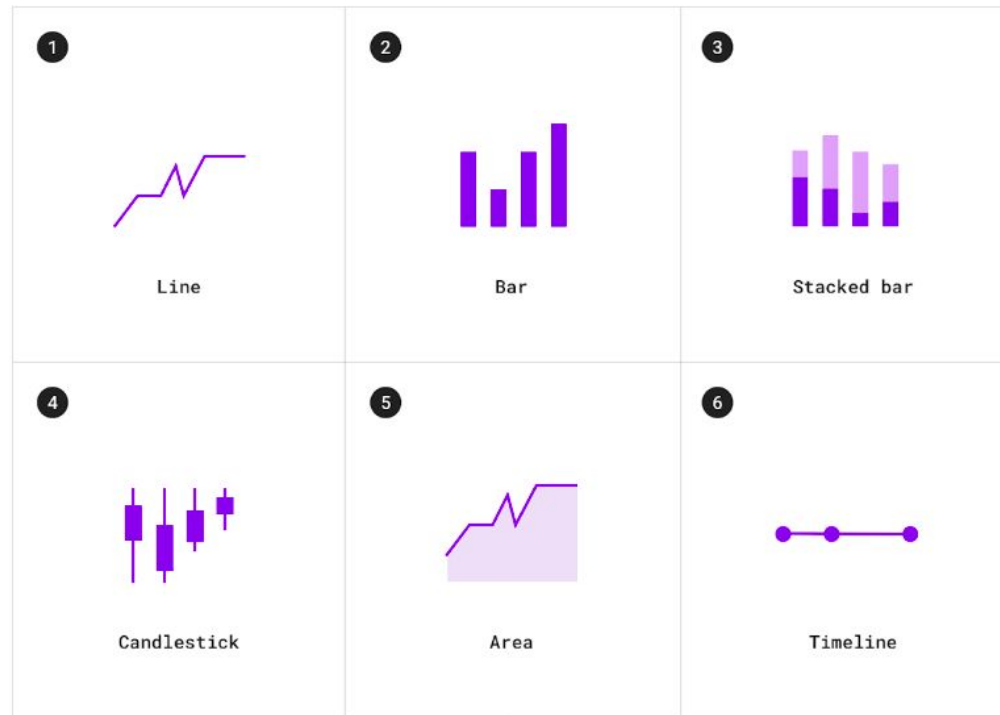
Tipos de Gráfico

- Variação no Tempo;
- Comparação de categorias;
- Ranqueamento;
- Proporção;
- Correlação;
- Distribuição;
- Fluxo;
- Relacionamento.

Gráficos de Variação no Tempo

Exemplo de uso

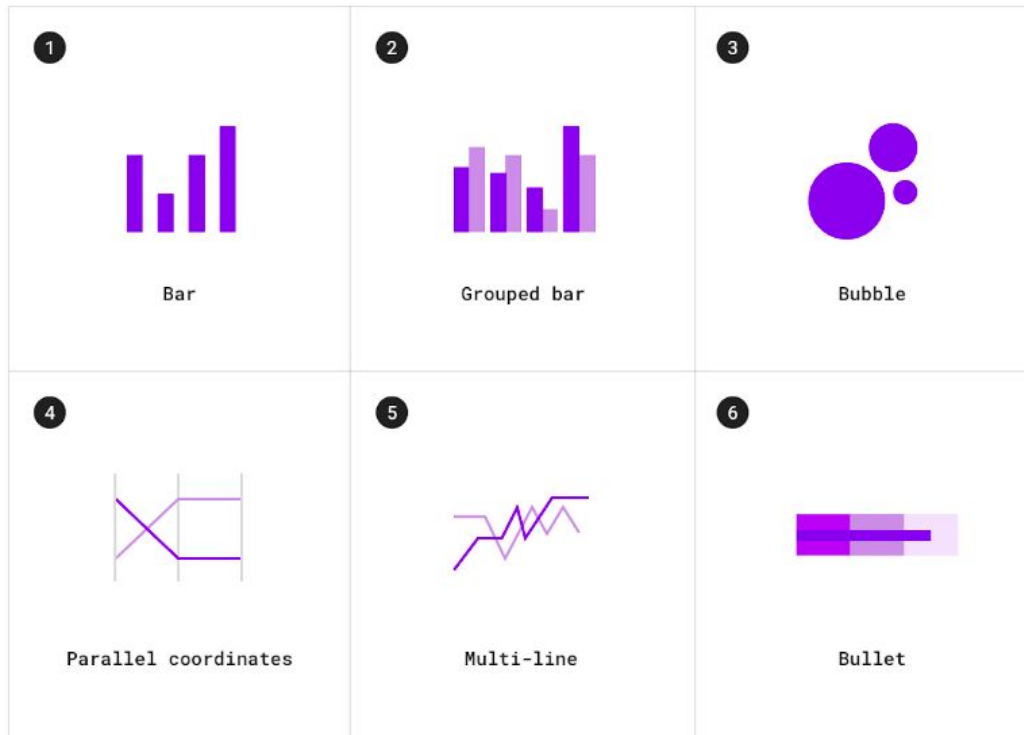
- Tendências de mercado
- Preço de ativos
- Cronologias



Gráficos de Comparação de Categorias

Exemplo de uso

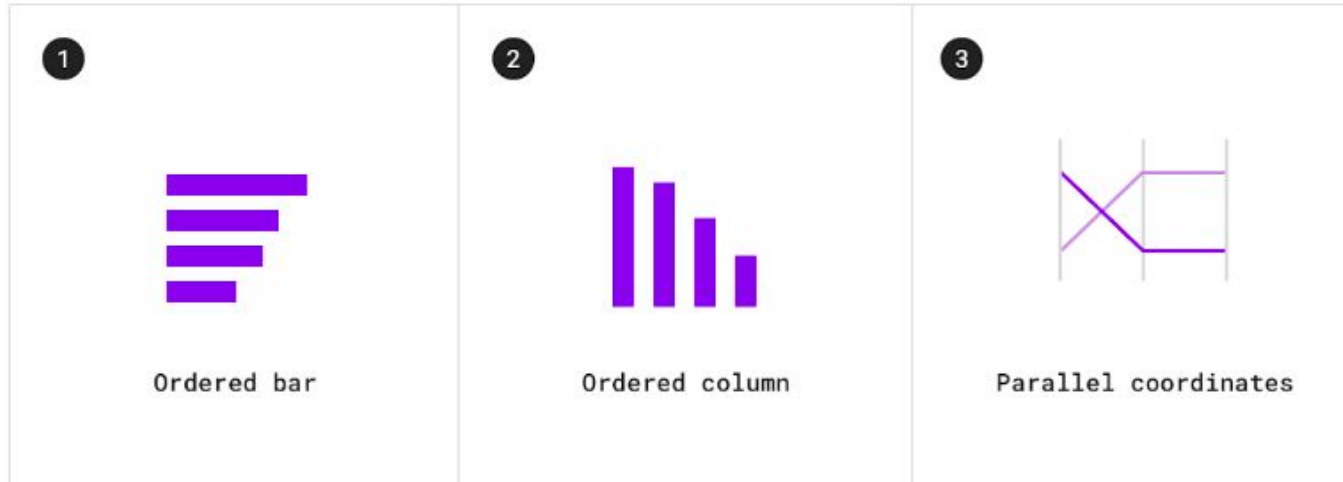
- Comparação de renda países
- Contagem de ocorrência.



Gráficos de Ranqueamento

Exemplo de uso

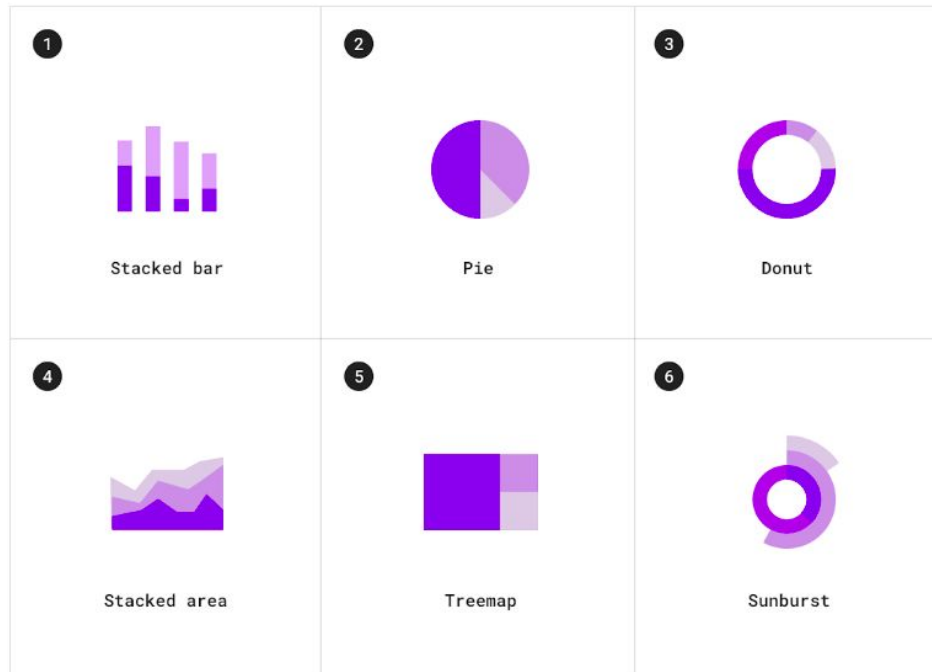
- Resultados de eleição
- Desempenho estatístico



Gráficos de Proporção

Exemplo de uso

- Despesas por categoria
- Alocação de recursos por categoria



Gráficos de Correlação

Exemplo de uso

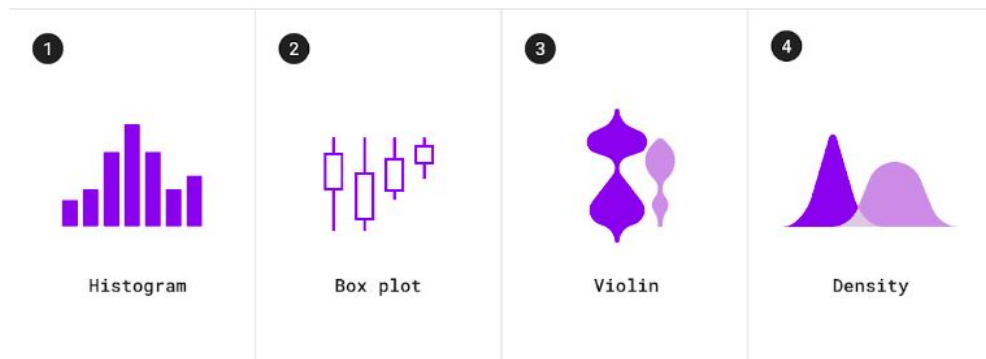
- Renda e expectativa de vida.



Gráficos de Distribuição

Exemplo de uso

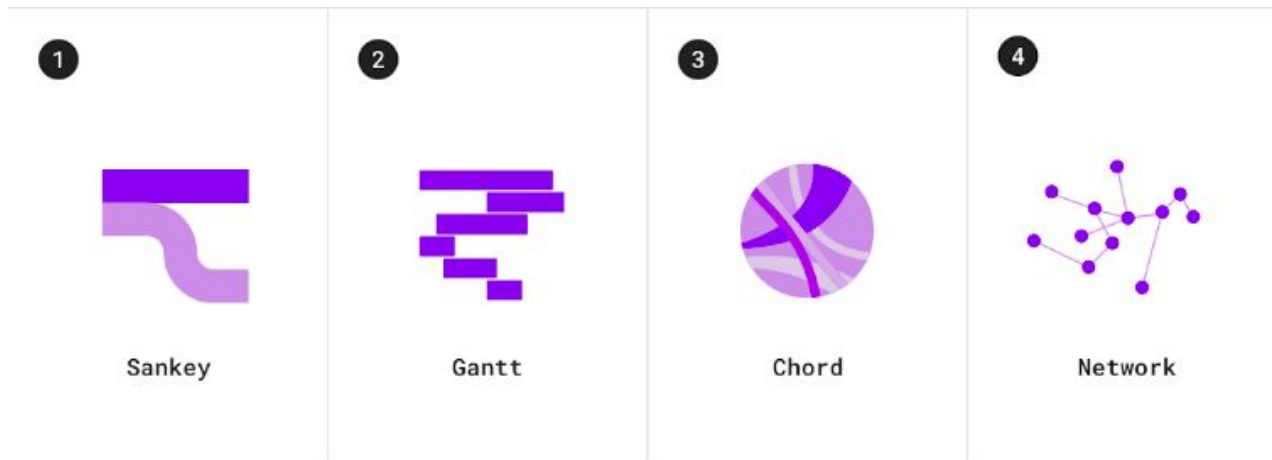
- Distribuição populacional
- Distribuição de renda;



Gráficos de Fluxo

Exemplo de uso

- Cronograma de projeto
- Transferências



Gráficos de Relacionamento

Exemplo de uso

- Redes sociais
- Nuvens de palavra



{

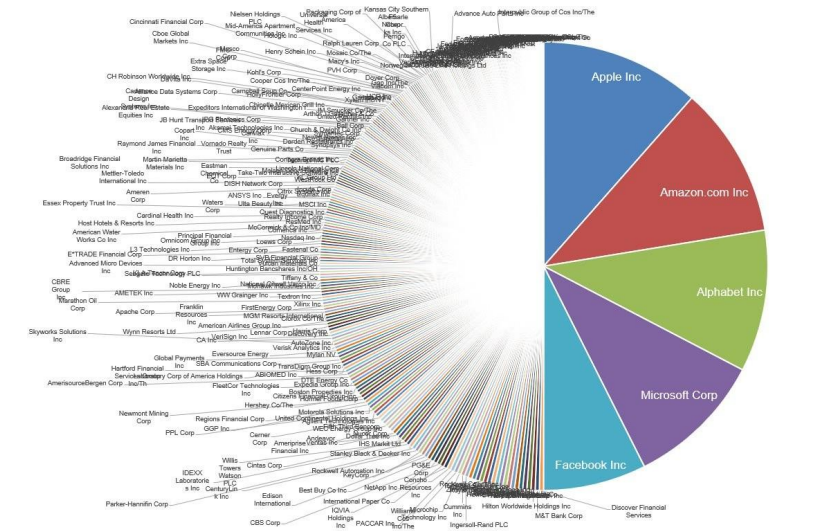
Boas e Más
Práticas

}

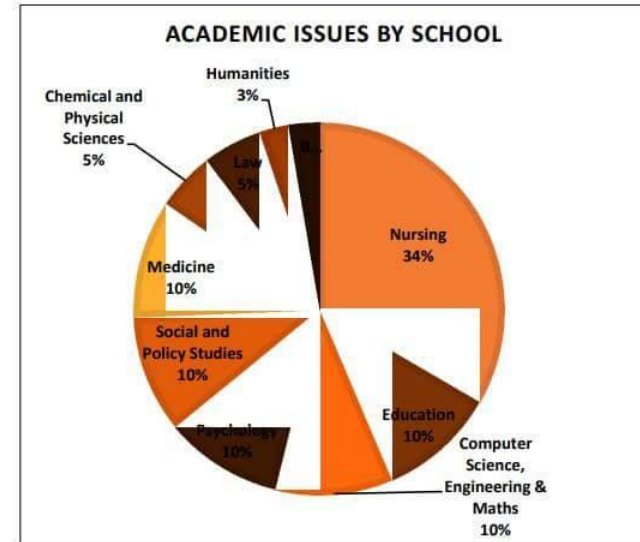
Boas Práticas

- Veracidade, **NUNCA** manipule maliciosamente seus dados;
- Prezar por simplicidade e máxima compreensão;
- Coerência de cores.

Más Práticas



Sobrecarga de dados



Design ruim, Difícil visualização

{ Hands on! }

Exemplos de Visualização de Dados em Python

Exemplo 1:

<https://colab.research.google.com/drive/1pZWH9Jbu0T8mo8NmXbu7-SIBdlU2ne8B?usp=sharing>

Exemplo 2:

<https://colab.research.google.com/drive/1eWddGix1Pmp1UONXtSFKaMNUYT0sE-nf?usp=sharing>

Dúvidas?

Descrição do Projeto

Entregáveis

- Relatório escrito (pelo menos 2 páginas);
- Código do Colab.

Requisitos do projeto

- Dupla ou individual;
- Leitura do conjunto de dados;
- Análise exploratória;
- Tratamento (opcional)
 - Remoção, transformação, preenchimento.
- Uso dos resultados da análise de estatística descritiva dos dados;
- Gráficos;
- Um fato interessante observado/percebido durante a análise dos dados;
- Entrega até 03/08/2022.

Descrição do Projeto

Recursos

- Monitoras
 - Lembrar de pedir ajuda!
- Consultas a conjunto de dados
 - <https://www.kaggle.com/>
 - Tomar cuidado com o tamanho do conjunto de dados;
 - Observar a pontuação de usabilidade do conjunto de dados;