

data

Introdução ao Kaggle

Gustavo Sutter
@suttergustavo

Kaggle

- <https://www.kaggle.com/>
- Maior site de ciência de dados
- Oferece diversos recursos
 - Competições
 - Datasets
 - Recurso computacional (GPUs <3)
 - Códigos
 - Cursos
 - Discussões



Competições

- O Kaggle surgiu como um site de competições de ciência de dados
- Empresas colocam problemas no site e a comunidade compete para resolvê-los
- Prêmios em dinheiro 🏆🏆🏆
- **Excelente fonte de prática e aprendizado**
- Ranking de usuários de acordo com resultados de competições



Competições: tipos de competições

- Featured
- Research
- Getting started
- Playground

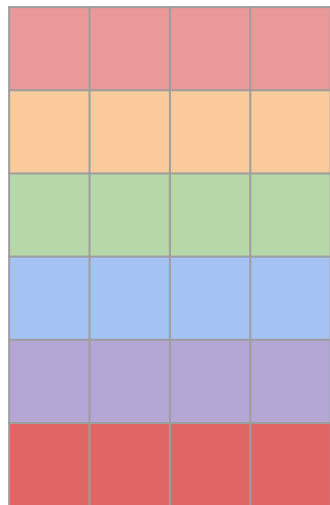


Competições: funcionamento

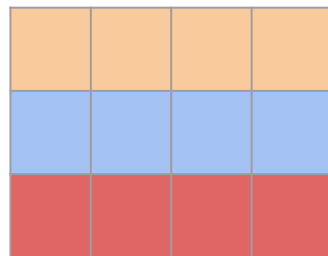
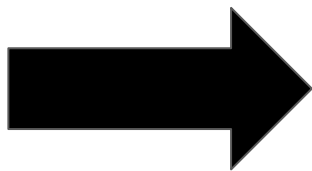
- Participantes recebem um conjunto de treino e um conjunto de teste
- Resultados são avaliados usando uma determinada métrica de avaliação
- Classificação dos participantes na tarefa em questão (**Leaderboard**)
 - Leaderboard público
 - Leaderboard privado



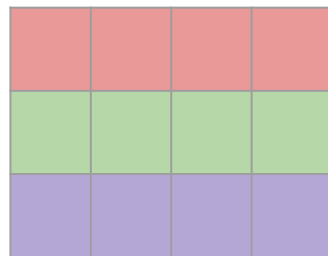
Competições: Leaderboard



Seu lado



**Public
LB**



**Private
LB**



Kaggle



Competição: ferramentas de aprendizado

- Notebooks abertos com códigos e explicações
- Fórum de discussões



Datasets

- Além de competições o Kaggle também tem um repositório com mais de 70 mil datasets
- Todo dataset também possui
 - Notebooks
 - Fórum de discussões



Recursos computacionais

- Kaggle tem ferramenta para executar notebooks gratuitamente
 - GPU
 - Memória RAM
- Semelhante ao Colab, mas com import de datasets já pronto



Cursos

- Cursos de diversas áreas
 - Intro ao Python
 - Pandas
 - Visão computacional
 - Feature Engineering
 - Aprendizado por reforço

