

Projeto de Ciência de Dados

Descrição

O objetivo do projeto de ciência de dados é que alunos usem o conteúdo visto em aula na prática, através de uma aplicação de interesse próprio. Cada pessoa pode escolher o tema, contanto que respeite os critérios descritos abaixo. O prazo para entrega do projeto é **dia 27/02 às 23:59h**.

Tecnologia e Entregáveis

- (Requisito) Usar as seguintes ferramentas
 - Github e template de repositório
 - Docker
 - R, Python, ou Julia
- (Opcional) Usar as seguintes ferramentas
 - SQL / Base dos Dados / DataZoom
 - Raspagem de dados
 - Computação de nuvem
- Entregáveis
 - Conjunto de dados (uma ou mais tabelas) no formato csv (via Dropbox, Drive, Box, etc).
 - Link para um repositório no GitHub.
 - Arquivo PDF descrevendo o projeto: objetivo, metodologia, dados utilizados, principais visualizações com explicações, etc.

Temas

O tema do projeto é livre. Abaixo estão algumas ideias para servir de inspiração:

1. Produzir um relatório sobre indicadores e tendências em educação no Brasil que servirá para pautar o planejamento estratégico de um grande *think-tank* no setor.
2. Fazer uma análise quantitativa da performance de diversos tipo de investimento no Brasil (e.g. ações, fundos de investimento, títulos, etc).
3. Produzir uma análise de sentimento baseada em textos de jornais ou Twitter e testar como esse indicador se correlaciona com eventos políticos relevantes.
4. Montar um modelo de previsão do clima em estações espalhadas pelo Brasil ou outros países.
5. Estudar ciclos políticos de contratações de funcionários públicos em cargos de confiança no Brasil.

Avaliação

O projeto será avaliado seguindo alguns critérios:

- Qualidade do código: clareza, limpeza, uso de funções.
- Estrutura dos dados: formato *tidy*, nomeação de colunas, eficiência.
- Estrutura e uso do repositório no GitHub.
- Qualidade das visualizações.