



Projeto Final Big Data Science

Como tarefa do time, vocês recebem a incumbência de fazer uma análise dos dados coletados no ENEM 2021. Como colher de chá, já que encontrar e vasculhar a fonte dos dados é seu trabalho, seu supervisor te envia o link de onde encontrar os dados:

<https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/enem>

Ele então combina com o time de entregar os relatórios 1 semana depois, como um jupyter notebook e um pdf/html no formato de relatório escrito, com título, introdução e contextualização histórica e político-social se necessário (com referências), introdução ao tema abordado, desenvolvimento de seus pontos com a ajuda da exploração dos dados, métodos estatísticos etc., hipóteses levantadas e possíveis direções para o órgão público que será dirigido. Ele diz também para se sentir livre para usar qualquer método estatístico ou de machine learning que você conhece, trazer previsões, gráficos, dentre outros métodos para suportar sua argumentação.

Lembre-se que você trabalha com dados, mas ficar somente olhando para os dados é geralmente uma má ideia. Então procure fontes em outras áreas especializadas nos problemas que vocês irão abordar, como ciências sociais, ciências políticas etc., mas sempre garanta com estatísticas e dados que o contexto previsto para determinada hipótese realmente se dá na situação específica que você está analisando. Ou seja, sempre case a teoria com o contexto que os dados trazem. Quando não é possível delimitar um contexto específico apenas com dados ou com os dados em mãos, deixe isso claro e traga argumentação de outras áreas do conhecimento, ou simplesmente ache outra linha para explorar.

O projeto deve estar no github.com, a forma de organizar o conteúdo é por sua conta, caso nunca tenha usado, este já é seu primeiro desafio.

Ao final do projeto, você deve preencher o formulário com o seu nome completo, e-mail utilizado no treinamento e o link do github do seu projeto.

Link do formulário para envio:

<https://forms.office.com/r/9qc1maw1sV>

O projeto deve ser enviado até dia **22/08/2022**.



Uma dica final: o poder dos dados somente aparece quando sabemos transportá-los para a realidade, dentro de um contexto estabelecido e muito bem estudado. Fora de contexto ele pode levar a conclusões enganosas, parciais, racistas, homofóbicas, xenófobas etc., ou seja, quando fora de contexto, a ciência de dados não é uma ciência de fato, e sim apenas um apanhado de números sem sentido ou simplesmente distorcidos! Em resumo, um bom cientista de dados sabe que antes de olhar para os dados ele precisa de suporte e conhecimento mínimo de área.

Bom projeto e estudos