

Processo Seletivo 01/2021

IEEE Computational Intelligence Society, UnB

Instruções gerais

1. Esse desafio tem caráter classificatório e eliminatório no Processo Seletivo para o CIS.
2. Sua solução deve ser entregue, via e-mail para ieeecisunb@gmail.com, até às **18h de domingo, 31/10**.
3. Essa etapa visa avaliar as habilidades dos candidatos em manipular, visualizar e interpretar grandes volumes de dados, de forma em que a classificação para a segunda etapa não depende obrigatoriamente de responder todas as perguntas de forma correta.
4. O envio é recomendado ainda que o candidato tenha desenvolvido apenas uma solução parcial, visto que **todo desenvolvimento será considerado**.
5. É esperado que o candidato apresente soluções originais e de própria autoria.
 - a. Caso seja constatada a influência de terceiros na solução apresentada, o candidato será sumariamente desclassificado do presente Processo Seletivo.
6. Soluções enviadas após o prazo limite serão desconsideradas.
7. Em caso de dúvidas, não hesite em nos contatar por meio do e-mail ieeecisunb@gmail.com

Introdução

É cada vez mais notória a importância da análise de dados nas mais variadas áreas, tanto que ela já representa uma das principais fontes consideradas na tomada de decisões. De modo particular, compilar e analisar dados sobre os diferentes países do mundo permite a comparação mais precisa e proporcional entre as regiões, levando em consideração suas particularidades e propiciando que decisões mais acertadas sejam tomadas.

Dataset e análise de dados

Neste contexto, o Desafio proposto utilizará quatro datasets diferentes contendo dados sobre mesmos países. São eles:

countries - Dados gerais sobre os países. Com dados sobre região, população, área do país etc.

Observação sobre as colunas **Agriculture**, **Industry** e **Service**: elas são calculadas de forma que somam 1 (100%).

human_freedom - Dados do Índice de Liberdade Humana. Esse índice tem o objetivo de pintar um quadro amplo, mas razoavelmente preciso, da extensão da liberdade geral no mundo. Cada país tem dados de 11 anos (de 2008 a 2018). O significado de cada coluna pode ser extraído [desse link](#).

happiness_report - Dados sobre o índice de felicidade. Atenção que os dados entregues informam sobre como a coluna afeta o cálculo do índice de felicidade. E não é a medida em si. Por exemplo, a coluna de PIB não mostra o PIB do país, e sim o quanto o PIB contribui para o índice de felicidade.

medals_olympics - Dados de quantas medalhas de ouro, prata e bronze cada país recebeu nas Olimpíadas de Tóquio 2021.

Sendo assim, responda, da forma mais completa possível, as seguintes questões:



1. No dataset **human_freedom**, existem alguns valores faltantes. Antes de começar a manipular os dados, trate essas informações e descreva sucintamente as alterações feitas;
2. Calcule o número total de medalhas recebido por cada país e adicione essa coluna no dataset;
3. Seguindo as regras oficiais das Olimpíadas, obtenha o ranking dos países. Visualize, também, os dez melhores países com base no total de medalhas;
4. Obtenha a média de alfabetização por região e a represente graficamente;
5. Liste os cinco países com maior número de celulares a cada mil habitantes. Faça uma representação gráfica;
6. Obtenha os dados do dataset **human_freedom** apenas para o ano de 2018;
7. Mostre a evolução do hf_score (human freedom score) do Brasil em função do tempo;
8. Mostre qual é a região com a maior média populacional;
9. Com base na análise dos dados, é possível afirmar que países com maior população tendem a ter mais pessoas saindo ou entrando no país (maior fluxo migratório)?
10. Caso o número de medalhas de bronze e de ouro fossem trocadas, qual seria o ranking dos países?
11. Normalize os valores dos datasets;
12. É possível observar correlação entre os índices de natalidade e de mortalidade? Justifique o resultado encontrado;
13. É possível observar correlação entre a generosidade da população e o índice de agricultura dos países? Justifique o resultado encontrado;
14. Monte grupos de forma similar a é feito na Copa do Mundo, isto é, um forte, dois intermediários e um fraco (deixe claro o parâmetro utilizado para definir as categorias);
15. Qual país apresentou maior aumento na taxa de homicídios ao longo do tempo?
16. Qual foi o impacto das eleições americanas de 2016 no ranking de liberdade do país? Explique;
17. É possível observar correlação forte entre a população de um país e a colocação dele nas Olimpíadas? Justifique;
18. Em quais características o Brasil precisa investir para melhorar sua colocação no ranking olímpico? Encontre a característica mais importante entre todos os datasets;
19. Suponha que a economia do país dependa apenas de agricultura e indústria. O fato de um país ser litorâneo faz com que ele aposte mais na agricultura do que na indústria? Explique;
20. Obtenha o ranking das regiões com maior número médio de medalhas.

Entrega das submissões

Os dados podem ser analisados utilizando quaisquer ferramentas de sua preferência. Por exemplo, **mas não limitados a:**

- Python;
 - Pandas;
 - Matplotlib;
 - Numpy;
 - Sklearn.
- Matlab;
- R;
- Excel.

Todos os gráficos utilizados para responder os questionamentos devem ser entregues, bem como todas as linhas de código produzidas. No intuito de facilitar a vida de vocês e agilizar nossa correção, pedimos que o relatório entregue seja preenchido de acordo com o fornecido [neste link](#) e entregue por e-mail para ieeecisunb@gmail.com até a data limite de domingo, 31/10, às 18h. Gráficos ilustrativos poderão acarretar em pontos extras.