



ARA0051 – Tópicos em Big Data

3º Exercício de Nota Final (NF)

1. Obtenha o *dataset* do endereço https://www.w3resource.com/python-exercises/pandas/filter/world_alcohol.csv e realize as seguintes análises (3,0 pontos):
 - a. Agrupe os dados por tipo de bebidas;
 - b. Agrupe os dados por Região e por Ano;
 - c. Seção de Contagens: Contar a ocorrência de Regiões, de Países e a soma da coluna de valores por Bebida.
 - d. Realize análises estatísticas da coluna dos valores: Média, Moda, Mediana, Estatística Descritiva e Gráfico de comparação dos valores agrupados por tipo de bebida.
 - e. Mostre resultados de acordo com alguns critérios:
 - i. Mostrar a coluna de bebidas do ano de 1985.
 - ii. Mostrar a coluna de Região com valores acima de 4.

***Observação:** Cada ponto acima, é uma instrução, basta enviar a instrução/código relativo ao que se pede, devidamente identificado para avaliação.

2. Seja o *dataset* chamado **cursos-prouni.csv** que se encontra no endereço [Cursos e notas de corte do PROUNI 2018 - Datasets - Brasil.IO](#). Baixe o arquivo, armazene no seu *github*** e obtenha os dados a partir desse endereço do *github* e realize as seguintes análises (3,0 pontos):
 - a. Efetuar a limpeza das colunas de notas: onde tiver NaN (*Not a Number*), substituir por 0,0.
 - b. Agrupe os dados pelo grau (Bacharelado, Licenciatura, etc).
 - c. Agrupe os dados pelos cursos de Matemática, Medicina e Pedagogia.
 - d. Agrupe os dados por Estado.
 - e. Agrupe os dados pelos cursos Tecnológicos.
 - f. Elimine a coluna “*cidade_filtro*” do *dataframe*.
 - g. Apresente a média das mensalidades dos cursos de Medicina.
 - h. Média das notas de corte dos cursos de tempo integral.
 - i. Estatística Descritiva das Notas Integral Ampla dos cursos de Bacharelado.
 - j. Gráfico comparativo entre o grau dos cursos (Bacharelado, Licenciatura, Tecnologia, etc) pelas Notas Integral de Cotas.

***Observação 1:** Cada ponto acima, é uma instrução, basta enviar a instrução/código relativo ao que se pede, devidamente identificado para avaliação.

****Observação 2:** O seu código deverá estar em pleno funcionamento para a correção, portanto, carregar o *dataset* o seu *github* é **obrigatório** e necessário p/ devida correção.



ARA0051 – Tópicos em Big Data

3º Exercício de Nota Final (NF)

Observações:

1. Este é o 3º Exercício que irá compor a Nota Final da disciplina ARA0168 e vale 6,0 pontos.
2. Exercício realizado individual ou em dupla.
3. O exercício deve ser respondido em arquivo formato pdf (ou no formato .ipynb) e suas respostas enviadas para o email exercicios.profsimone@gmail.com com o nome da dupla no título do email, no formato <NF-BIG-DATA: Nome-da-Dupla>, até o dia 25/11/2022, às 23:59. Não será aceito após a data especificada de entrega.
4. **A interpretação e organização das questões faz parte do processo avaliativo.** Se as questões não estiverem identificadas, não serão corrigidas.
5. Os códigos podem ser feitos na mesma célula do Notebook, porém, cada questão deve estar **devidamente identificada** (seja no documento ou no código, por comentários).
6. Os códigos devem estar funcionando / executando corretamente, sem erros e sem retorno de exceções para ser considerado **correto**. Portanto, verifique seu endereço do *github* está correto, se os dados estão carregados corretamente, se estão **identados** corretamente etc.
7. A Nota Final (NF) será composta pela soma das notas obtidas nos exercícios já realizados, sendo esta:
 - a. **Exercício Nota Final 1** – 19/10 (já realizado): 2,0 pontos
 - b. **Exercício Nota Final 2** – 09/11 (já realizado): 2,0 pontos
 - c. **Exercício Nota Final 3** – Até 25/11/2022: 6,0 pontos.
8. Exercícios com codificação claramente iguais serão penalizados tendo a nota dividida pela metade.
9. Lembrando que a disciplina é de nota final (NF) e a média é de 6,0 pontos. A soma das notas obtidas sendo menor que 6 leva a reprovação.

Bom trabalho

Prof. Simone Gama