

Profa. Solange Kanso

Disciplina: Ciência de Dados – Turma 24 - MM

1ª Avaliação

A. Instruções para a entrega

- ✓ **Entrega:** até às 23h59 do dia 02/04/2024
- ✓ Essa avaliação pode ser resolvida em grupo de no mínimo 5 alunos e no máximo 10 alunos (sem exceção). Trabalhos com número menor do que 5 integrantes ou com número maior do que 10 integrantes, não serão considerados
- ✓ Um dos integrantes irá enviar um e-mail para solange.kanso@uni9.pro.br com o código de extensão .ipynb. O nome do arquivo será “AVA1 - Turma 24 - nome e sobrenome de um integrante do grupo. ipynb”
- ✓ O título do e-mail será “AVA1 - Turma 24 - nome e sobrenome de um integrante do grupo”
- ✓ Enviar o código anexado no e-mail, não o link. Para salvar e anexar o código, vá em Arquivo > Fazer download > Fazer o download do . ipynb

B. Instruções da tabela de dados e o que deve conter no código

Com base no que trabalhamos em sala de aula, escreva um código em Python seguindo as solicitações.

- Utilize o arquivo “igm_modificado.csv” (disponível no Classroom dentro de tabela de dados) e o código “Ciência de Dados - Aula 03 04 - Churn - Análise exploratória.ipynb” como norte para desenvolver seu próprio código
- Todas as análises e resultados devem estar no código em formato de comentário ou texto
- **Não esqueça:** no início do código colocar o RA e o NOME COMPLETO de cada integrante do grupo. Nomes em mais de um trabalho serão penalizados
- Para iniciar, faça o upload da tabela em “csv” no seu drive para fazer a leitura da tabela em Python

C. Instruções para realizar a atividade:

1ª parte – utilize todos os municípios para as análises

1. Investigue o significado das colunas (1,0 ponto)
2. Faça uma checagem do arquivo (vimos diversos comandos, apresente pelo menos 3 deles) – (1,0 ponto)
3. Nesse arquivo, quais são os tipos de dados que encontramos? (1,0 ponto)
4. Escolha 2 variáveis e analise sua frequência absoluta e percentual (1,0 ponto)
5. Calcule o 1º quartil e o 3º quartil de alguma variável numérica e analise seus resultados (1,0 ponto)
6. Escolha uma variável numérica e por meio das estatísticas de tendência central (média, mediana e moda), identifique o tipo de assimetria da variável. Analise o seu resultado e justifique (1,0 ponto)

Sabendo que para isso é necessário comparar os valores da moda, mediana e média:

- Distribuição assimétrica positiva (ou assimétrica à direita): moda < mediana < média
- Distribuição simétrica ou ausente de assimetria: moda = mediana = média

- Distribuição assimétrica negativa (ou assimétrica à esquerda): média < mediana < moda
7. Há dados faltantes em algumas variáveis? Quais? Sugira formas de tratá-las e justifique. E em uma delas substitua o NaN pela mediana (1,0 ponto)
 8. Faça um gráfico de setores ou de pizza de alguma variável categórica e analise os resultados (1,0 ponto)
 9. Faça o boxplot de alguma variável numérica contínua por outra variável categórica e analise os resultados comparativos (1,0 ponto)

2ª parte – crie outro dataframe apenas com duas regiões.

10. Faça uma análise comparativa das estatísticas de 4 variáveis e analise as diferenças entre as regiões. Lembrando que é necessário definir uma hipótese sobre essas regiões (1,0 ponto)