

**Profª. Solange Kanso**

**Disciplina: Ciência de Dados – Turma 4 - MM**

**2ª Avaliação – 24/04/2025**

**Entrega:** até às 23h59 do dia 30/04/2025. Após esse horário não irei aceitar os trabalhos enviados

#### **A. Instruções para a entrega**

- ✓ Essa avaliação pode ser resolvida em grupo de no mínimo 3 alunos e no máximo 10 alunos (sem exceção). Trabalhos com número menor do que 3 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](mailto:solange.kanso@uni9.pro.br) com o código de extensão .ipynb. O nome do arquivo será “AVA2 - Turma 4 - nome e sobrenome do integrante do grupo que irá enviar o e-mail. ipynb”.
- ✓ O título do e-mail será “AVA2 - Turma 4 - nome e sobrenome do integrante do grupo que irá enviar o e-mail”
- ✓ Abrir o código **modelo** “AVA2 - Turma 4 - nome e sobrenome do integrante do grupo.ipynb”. Ir em **Arquivo > Salvar uma cópia no drive**. Assim que gerar a cópia, feche o original.
- ✓ **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
- ✓ **Deixar os resultados visíveis no código, nada deve estar oculto e todos resultados devem estar processados**

#### **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.

- Vocês irão ter como norte o notebook: Aprendizado de Máquina - aula 05 06 - Churn - Análise exploratória.ipynb
- **Não esqueça:** no início do código colocar o RA e o NOME COMPLETO de cada integrante do grupo.
  - a. Nomes em mais de um trabalho serão penalizados
  - b. Nomes solicitados para inserção após a data limite não serão considerados
- Para iniciar, faça o upload da tabela em “igm\_modificado.xlsx” no seu drive para fazer a leitura da tabela em Python
- Utilize o arquivo “igm\_modificado.xlsx” e o código das aulas como norte para desenvolver seu próprio código. **Essa tabela de dados é para ser utilizada em todas as questões da avaliação. Não há necessidade de gerar dados fictícios**
- Todas as análises e resultados devem estar no código em formato de comentário ou texto.
- Todas as análises devem ser com as **próprias palavras**

### C. Instruções para realizar a atividade:

O objetivo da atividade é fazer a análise exploratória dos municípios brasileiros segundo as variáveis socioeconômicas demográficas. Dessa forma, siga as orientações.

*Antes de iniciar, verifique se todas as variáveis escolhidas estão no formato correto e estão fazendo sentido.*

### D. Faça uma análise exploratória da tabela de dados

**Atenção com a sequência das questões e a organização da entrega do trabalho!**

1. Faça as checagens dos dados: utilize 3 comandos que vimos em sala de aula. Um deles, obrigatoriamente, deve ser a sinalização se há ou não dados faltantes em cada uma das colunas. Analise os resultados com suas próprias palavras (0,5 ponto)
2. Faça a frequência simples e relativa de 2 variáveis categóricas. Analise os resultados com suas próprias palavras (0,5 ponto)
3. Apresente as medidas de tendência central, uma medida de posição relativa e uma medida de dispersão de 2 variáveis numéricas. Analise os resultados com suas próprias palavras (1,0 ponto)
4. Das variáveis numéricas escolhidas no item anterior, o que se pode afirmar sobre a assimetria da distribuição. Analise os resultados com suas próprias palavras (1,0 ponto)
5. Faça um gráfico de setores ou pizza de uma variável categórica. Analise os resultados com suas próprias palavras (1,0 ponto)
6. Faça uma tabela com duas variáveis (cruzamento entre elas). Analise os resultados com suas próprias palavras (1,0 ponto)
7. Escolha uma variável numérica que tenha dado faltante. Veja qual a quantidade de dados faltantes e calcule a média e mediana dessa variável. Depois, substitua esses dados faltantes pela **estatística que melhor se adeque** e recalcule a média e mediana. Explique por que escolheu a estatística e analise as alterações nos resultados (1,0 ponto)
8. Faça um boxplot de uma variável numérica. Analise os resultados com suas próprias palavras (1,0 ponto)
9. Faça um boxplot da mesma variável numérica do item anterior por uma categórica. Analise os resultados com suas próprias palavras (1,0 ponto)
10. Faça o gráfico de dispersão da variável `pib_pc` com outras 2 variáveis numéricas (separadamente). Calcule as correlações entre elas. Analise os resultados com suas próprias palavras (1,0 ponto)
11. Escreva um parágrafo final, com suas próprias palavras, resumindo os principais achados (1,0 ponto)

## **Dicionário de dados do arquivo “igm\_modificado.xlsx”: nome das variáveis e o significado**

1. regioao: regiões em que os municípios estão
2. estado: estados em que os municípios estão
3. municipio: nome do município
4. codigo\_mun: código de cada município segundo o IBGE
5. capital: se é município de capital (igual a 1), se não é município de capital (igual a 0)
6. porte: porte de cada município conforme o tamanho da sua população: Grande porte, Médio porte, Pequeno porte 1 e Pequeno porte 2
7. populacao: total de habitantes no município
8. area: área do município por metro<sup>2</sup>
9. densidade\_dem: mede a relação entre a população absoluta e a extensão territorial, sendo expressa em habitantes/quilômetro quadrado
10. pib: produto interno bruto de cada município
11. pib\_pc: Produto Interno Bruto (PIB) do município dividido pela sua população.
12. participacao\_transf\_receita: proporção ou percentual da receita total que é transferida de uma entidade governamental para outra
13. servidores: total de servidores públicos no município
14. comissionados: total de comissionados no município (podem ser ocupadas por indivíduo sem vínculo com o órgão)
15. comissionados\_por\_servidor: número médio de funcionários em cargos comissionados em relação ao número total de servidores públicos no município
16. perc\_pop\_econ\_ativa: percentual da população economicamente ativa (PEA)
17. taxa\_empresendedorismo: proporção da população adulta que está envolvida em atividades empreendedoras no município
18. anos\_estudo\_empresendedor: número médio de anos de estudos dos empreendedores
19. jornada\_trabalho: número médio de horas trabalhadas
20. gasto\_pc\_saude: valor médio gasto por pessoa em relação saúde em cada município
21. hab\_p\_medico: número médio de habitantes por médico (medida utilizada para avaliar o acesso à assistência médica)
22. exp\_vida: expectativa de vida média ao nascer referente a população do município
23. gasto\_pc\_educacao: valor médio gasto por pessoa em relação a educação.
24. exp\_ano\_estudo: número médio de anos de estudo que se espera que um indivíduo complete durante sua vida
25. nota\_ciencias: nota média dos estudantes em ciências.
26. nota\_humanas: nota média dos estudantes em humanas.
27. nota\_linguagem: nota média dos estudantes em linguagens.
28. nota\_mat: nota média dos estudantes em matemática.
29. nota\_redacao: nota média dos estudantes em redação.
30. Idhm: Índice de Desenvolvimento Humano Municipal
31. ranking\_igm: Índice de Governança Municipal (IGM)
32. indice\_governanca: IGM - medida que avalia a eficiência e eficácia da gestão pública em nível municipal

**Boa avaliação!**