

## Bootcamp: Analista de Dados com ênfase para Mercado Financeiro

### Trabalho Prático

#### Módulo 2: Fundamentos em Ciências de Dados

#### Objetivos de Ensino

Exercitar os seguintes conceitos trabalhados no Módulo:

1. Coleta de dados.
2. Analisar e realizar tratamento de dados.
3. Analisar resultados obtidos.
4. Conhecimento teórico ministrado nas videoaulas.

#### Enunciado

Uma pesquisa abrangente foi conduzida em vários estados com o propósito de compreender as preferências dos entrevistados em relação à bebida favorita, clima, animal de estimação e hobbies. A riqueza desses dados oferece insights valiosos sobre as tendências e preferências culturais presentes em diferentes regiões do país.

Para extrair informações significativas, o próximo passo consistirá em submeter os dados a um cuidadoso tratamento. Será realizada uma análise minuciosa para identificar e abordar eventuais dados ausentes, garantindo que a integridade e a completude das informações sejam preservadas. Além disso, serão analisados e tratados possíveis registros duplicados, evitando distorções nos resultados finais.

A verificação e correção dos tipos de dados serão realizadas para garantir consistência e precisão ao longo da análise. Este processo incluirá a adequação de formatos, caso necessário, para promover a uniformidade na interpretação dos dados.

Como cientista de dados, você recebeu um conjunto de dados proveniente dessa pesquisa que investiga as preferências pessoais dos entrevistados em relação à bebida favorita, clima, animal de estimação e hobby. Sua tarefa é realizar uma análise abrangente desses dados, explorando padrões e tendências.

### ATENÇÃO PARA TRATAMENTO DE DADOS

Avalie se será necessário realizar tratamento de dados ausentes no datasets disponibilizado.

Instruções para correção de dados ausentes

1. Para os dados categóricos de forma geral, utilize a estratégia de alteração pela moda dos dados.
2. Para os dados numéricos, faça a correção dos dados com a média da coluna analisada.

### Atividades

Para esta atividade, os alunos deverão realizar análise de dados utilizando as bibliotecas abordadas no curso.

1. Criar uma virtualenv e instalar as bibliotecas necessárias.
2. Coletar os dados dos arquivos:

- a. dados\_pesquisa\_preferencias.csv
  - b. estado\_regiao.csv
3. Analisar os dados coletados.
  4. Avaliar a necessidade de tratamentos de dados ausentes.
  5. Manipulação e visualização dos resultados.
  6. Responder às questões teóricas e práticas do trabalho.

### Dicas do professor:

1. Analise os dados com cuidado.
2. Antes de enviar as respostas, verifique se o gabarito está correto.
3. Utilize as funções das bibliotecas abordadas nas videoaulas.
4. Atenção no momento de filtrar e corrigir dados (se necessário).
5. Tenha atenção no que pede cada questão.
6. Os dados disponibilizados no dataset são fictícios. Ou seja, não têm relação com o mundo real.
7. Siga fielmente todos os passos contidos no enunciado das questões
  - a. Não utilize outras estratégias, pois poderá impactar nas respostas da atividade.
8. O dataset utilizado no trabalho pode ser obtido no link:

[https://github.com/ProfLeandroLessa/classroom-datasets/raw/master/CDMF/TP/dados\\_pesquisa\\_preferencias.zip](https://github.com/ProfLeandroLessa/classroom-datasets/raw/master/CDMF/TP/dados_pesquisa_preferencias.zip)

9. Encontre neste link dicas sobre como realizar a coleta e extração automatizada de um arquivo ZIP.

<https://leandrolessa.com.br/tutoriais/automatizando-coleta-e-extracao-de-arquivos-zip-na-web-com-python/>

10. Tutorial para realizar o cálculo para conversão da data de nascimento em idades.

<https://leandrolessa.com.br/tutoriais/pandas-3-passos-para-converter-data-de-nascimento-em-idade/>

11. Atenção para as questões que solicitam a média de idades. Os resultados podem ser diferentes dependendo do dia que for realizado o cálculo do indicador. Mas não se preocupe! Esse comportamento já é esperado. Essa diferença pode acontecer devido à idade ser baseada entre a diferença do dia atual e a data de nascimento. Desta forma, ao realizar o cálculo em dias diferentes, as idades dos entrevistados podem ter variações. De qualquer modo, esse detalhe não invalida a questão. Obs.: geralmente as diferenças ocorrem nas casas decimais.