



Bootcamp Data Science

Atividade Modular

Módulo 1: Introdução à Ciência de Dados e Python

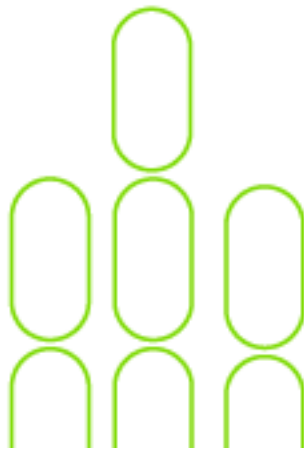
Objetivos de Ensino

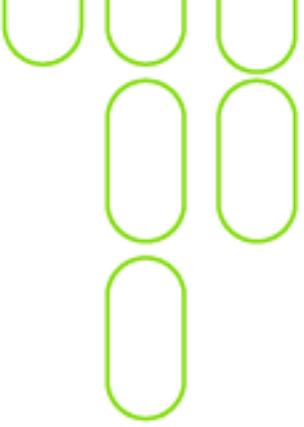
Exercitar os seguintes conceitos trabalhados no Módulo:

1. Coleta de dados
2. Analisar e explorar os conjuntos de dados
3. Analisar e realizar tratamento de dados se necessário
4. Criar visualização gráficas
5. Realizar agrupamentos de dados
6. Analisar resultados obtidos
7. Responder às questões objetivas do trabalho

Enunciado

Você foi contratado como cientista de dados para realizar uma análise detalhada dos dados de pacientes em um projeto de pesquisa na área da saúde. Os dados estão organizados em três conjuntos principais: `dados_clinicos`, `dados_pacientes` e `dados_estados`. Cada conjunto de dados contém as seguintes informações:

- **dados_clinicos:** Informações dos pacientes, como *id_cliente*, *peso*, *colesterol* e *gênero*.
 - **dados_pacientes:** Detalhes adicionais como *id_cliente*, *idade*, *classe_trabalho*, *escolaridade*, *id_estado*, *estado_civil*, *raça*, *quantidade de filhos* e *salário*.
- 

- 
- **dados_estados:** Dados sobre os estados, como *id_estado*, *sigla*, *estado*, *região* e *país*.

Importância da Análise de Dados em Saúde

Analisar dados de saúde é essencial para identificar padrões que podem melhorar os cuidados com os pacientes e a saúde pública. Com essa análise, você pode:

1. Identificar fatores de risco para a saúde de diferentes grupos de pessoas.
2. Prever resultados clínicos e ajudar os profissionais de saúde a tomarem decisões mais informadas.
3. Melhorar a distribuição de recursos de saúde, entendendo melhor as necessidades de cada região.
4. Apoiar políticas de saúde, fornecendo dados que ajudem a melhorar o acesso e a qualidade dos cuidados.

Benefícios


A análise de dados na saúde pode levar a tratamentos mais eficientes e personalizados, ajudando a melhorar a qualidade de vida das pessoas e a reduzir custos com saúde.

Atividades

1. Carregar e explorar o conjunto de dados para obter uma compreensão inicial dos dados.
2. Criar representações gráficas dos gráficos
3. Analisar e corrigir dados
4. Visualizar os clusters e analisar os padrões identificados para fornecer insights significativos para a equipe de pesquisa.

ATENÇÃO PARA TRATAMENTO DE DADOS





Avaliem se será necessário realizar tratamento de dados ausentes nos datasets disponibilizados.

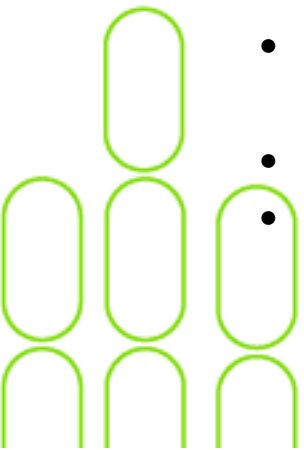
Instruções para correção de dados ausentes

- Moda para Variáveis Categóricas Preencher valores ausentes nas colunas categóricas com a moda.
- Mediana para Variáveis Numéricas: Preencher valores ausentes nas colunas numéricas com a mediana.

Dicas do professor:

- Leia atentamente todas as instruções da atividade.
- Analisem com cuidado os dados através da representação gráfica.
- Analisem bem o gráfico gerado e a disponibilização dos dados do agrupamento.
- Antes de enviar as respostas verifiquem se o gabarito está correto.
- Tenham atenção no que pede cada questão.
- Sigam exatamente o que o enunciado do trabalho solicita.
- Obs: Não existe pegadinhas nas alternativas.
- Os datasets utilizados no trabalho podem ser obtidos nesse link.

Dicas para integração dos dados.

- <https://leandrolessa.com.br/tutoriais/integracao-de-dados-descubra-os-4-tipos-de-joins-essenciais/>
 - Dicas para criar gráfico de dispersão:
 - <https://leandrolessa.com.br/tutoriais/grafico-de-dispersao-como-criar-e-analisar-na-pratica/>
 - Realizar download automático de bases na internet
 - <https://leandrolessa.com.br/tutoriais/automatizando-coleta-e-extracao-de-arquivos-zip-na-web-com-python/>
- 



Bibliotecas utilizadas

```
Pandas: 2.2.2  
seaborn: 0.13.2  
sklearn: 1.5.1  
matplotlib: 3.9.1
```

É crucial reconhecer que a linguagem de programação Python e suas bibliotecas associadas estão em constante evolução. Como resultado, pode ocorrer que funções ou métodos específicos, que costumavam estar disponíveis em versões anteriores, deixem de existir ou passem a ser implementados de maneira diferente em versões mais recentes.

Essas atualizações são realizadas para melhorar a eficiência, corrigir erros e fornecer novos recursos aos desenvolvedores. No entanto, essa dinâmica de mudança também pode criar desafios, especialmente quando se trabalha com código legado ou ao compartilhar código com outros membros da equipe. Portanto, é de extrema importância que os alunos estejam cientes dessas mudanças e estejam dispostos a se adaptar a elas.

