

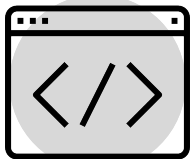
Profissão: Cientista de Dados



BOAS PRÁTICAS



Descritiva I



- **Conheça os tipos de variáveis**
- **Utilize a biblioteca Matplotlib**
- **Crie gráficos básicos**
- **Utilize boxplot e histograma**
- **Utilize Seborn e Bivariadas Discreta**
- **Analise Variáveis Bivariadas Discretas e Contínuas**



Conheça os tipos de variáveis

- **Use as ferramentas corretas para analisar cada tipo de variável:** variáveis discretas e contínuas podem ser analisadas através de cálculos de média e gráficos de caixa, enquanto variáveis qualitativas podem ser analisadas através de tabelas de frequência e gráficos de barras.
- **Entenda a associação entre variáveis:** a associação entre variáveis é um conceito importante na análise de dados. Para variáveis qualitativas, uma tabela de frequência e um gráfico de barras podem ser úteis, enquanto para variáveis contínuas, um gráfico de dispersão pode ser mais apropriado.



Conheça os tipos de variáveis

- **Evite fazer cálculos com variáveis nominais:** como as variáveis nominais não têm uma ordem, não faz sentido fazer cálculos com elas.
- **Sempre verifique a natureza dos seus dados antes de começar a análise:** saber se suas variáveis são discretas, contínuas, ordinais ou nominais pode ajudar a determinar a melhor maneira de analisá-las.



Utilize a biblioteca Matplotlib

- Ao criar gráficos com Matplotlib, comece com o gráfico básico e, em seguida, faça modificações incrementais para melhorar a apresentação dos dados.
- Use a funcionalidade de **subplot** para comparar diferentes conjuntos de dados ou diferentes aspectos do mesmo conjunto de dados.
- Sempre rotule seus eixos e adicione um título ao seu gráfico para que os leitores possam entender o que os dados representam.
- Ajuste as configurações do Matplotlib para atender às suas necessidades específicas, como alterar o tamanho da figura e das fontes usadas nos gráficos.



Crie gráficos básicos

- Embora o gráfico de pizza não seja recomendado por muitos especialistas devido à dificuldade de comparar categorias visualmente, ele ainda pode ser útil em algumas situações, como quando é necessário preencher um slide de PowerPoint com um gráfico colorido.
- O gráfico de Pareto pode ser uma ferramenta útil para mostrar as categorias mais relevantes de uma base de dados. Ao lidar com as categorias mais relevantes, a empresa pode atender a maioria das reclamações ou problemas.



Utilize boxplot e histograma

- Utilize a função 'describe' para fornecer uma visão geral rápida dos dados. Ela fornece informações como contagem, média, desvio padrão, mínimo, primeiro quartil, mediana, terceiro quartil e máximo.
- Ao analisar histogramas, esteja atento a valores que estão muito distantes dos demais. Esses valores extremos podem distorcer a análise e podem ser resultado de erros de entrada de dados ou situações incomuns.
- Utilize boxplots para visualizar a distribuição de uma variável. Eles podem ser criados usando tanto pandas quanto seaborn.
- Ao analisar boxplots e histogramas, considere adicionar mais dimensões ao gráfico para avaliar a relação entre as variáveis.



Utilize Seborn e Bivariadas Discreta

- Ao analisar dados, tente olhar para eles de diferentes perspectivas. Por exemplo, se você está analisando a frequência de um determinado grupo em diferentes dias da semana, também pode ser útil analisar qual dia da semana esse grupo prefere.
- Utilize tabelas de frequência cruzada para uma análise mais rigorosa. Elas permitem calcular os percentuais de linha e podem fornecer insights mais profundos sobre seus dados.
- Ao usar o Catplot para quebrar linhas por tempo e colunas por diferentes grupos, tenha cuidado para não quebrar demais a informação. Isso pode tornar a base de dados esparsa e dificultar a tomada de conclusões.



Analise Variáveis Bivariadas Discretas e Contínuas

- Utilize gráficos de dispersão para visualizar a relação entre variáveis contínuas. Isso pode ajudar a identificar correlações e padrões nos dados.
- Quando estiver lidando com um grande número de pontos de dados, considere o uso de uma matriz de dispersão. Isso pode tornar os dados mais fáceis de interpretar, dividindo o gráfico em uma grade e colorindo cada quadrado de acordo com a densidade de pontos nesse quadrado.
- Continue aprendendo e aprimorando suas habilidades. A ciência de dados é um campo em constante evolução, e é importante manter-se atualizado com as últimas técnicas e ferramentas.



Bons estudos!

