

Revisão P1

Estudos Avançados de Bancos de Dados

Sistemas de Informação

Pontifícia Universidade Católica de Campinas

Prof. Dr. Denis M. L. Martins

Series e DataFrames

1. O que é o Pandas e para que ele é utilizado na ciência de dados?
2. Explique a diferença entre um objeto Series e um DataFrame no Pandas.
3. Qual a diferença entre criar uma Series a partir de uma lista e a partir de um dicionário?
4. Descreva a diferença dos atributos index e values da Series.

Manipulação de Arquivos

1. O que é manipulação de arquivos em ciência de dados e por que é importante?
2. Qual a diferença entre arquivos estruturados e não estruturados? Dê exemplos.
3. Explique a função `read()`, `readline()` e `readlines()`. Qual a diferença entre elas?
4. O que é um arquivo CSV?
5. O que é um arquivo JSON e qual sua principal vantagem?

Análise Descritiva de Dados

1. O que é Análise Descritiva de Dados e qual seu principal objetivo?
2. Qual a diferença entre uma variável quantitativa e uma variável qualitativa? Dê exemplos.
3. O que é a mediana e quando seu uso é mais apropriado do que a média?
4. Explique o conceito de moda e como ela pode ser útil na análise de dados.
5. Defina o que são valores outliers.
6. Como a presença de outliers pode afetar a média, a mediana e a moda?
7. Explique a importância das medidas de dispersão na análise de dados.

Visualização de Dados

1. Por que gráficos de pizza/rosca devem ser usados com moderação?
2. O que um gráfico de dispersão pode revelar sobre a relação entre variáveis?
3. O que é um boxplot e quais informações ele fornece?
4. Como identificar outliers a partir do boxplot?
5. Explique o conceito de quartis e sua relação com o boxplot.