escola britânica de artes criativas & tecnologia

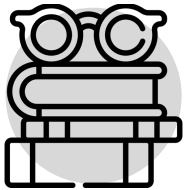
Profissão: Cientista de Dados

escola
britânica de
artes criativas
& tecnologia

GLOSSÁRIO

Limpeza e preparação de dados

- **Identifique e trate dados ausentes**
- **Renomeie índices e colunas**
- **Categorização e dummies**
- **Compreenda a amostragem**
- **Junte tabelas**



Dica: para encontrar rapidamente a palavra que procura aperte o comando CTRL+F e digite o termo que deseja achar.

escola britânica de artes criativas & tecnologia

Identifique e trate dados ausentes

Identifique e trate dados ausentes

- **bfill**
  Método em Python que preenche os dados ausentes com o valor seguinte na coluna ou linha.

- **drop_duplicates**
  Método em Python que remove linhas duplicadas de um DataFrame.

- **dropna**
  Método em Python que remove linhas ou colunas com dados ausentes de um DataFrame.

Identifique e trate dados ausentes

- **dropna**
  Método em Python que remove linhas ou colunas com dados ausentes de um DataFrame.

- **ffill**
  Método em Python que preenche os dados ausentes com o valor anterior na coluna ou linha.

- **fillna**
  Método em Python que preenche os dados ausentes com um valor específico.

- **isna e isnull**
  Métodos em Python que retornam um DataFrame booleano onde os valores True indicam a presença de um valor ausente.

Identifique e trate dados ausentes

- **map**
  Método em Python que mapeia valores em uma coluna para outros valores.

- **Preenchimento com a média ou mediana**
  Técnica de tratamento de dados ausentes que substitui os valores ausentes pela média ou mediana dos dados existentes.

- **NaN (Not a Number)**
  Representação de um valor ausente ou indefinido em Python.

Renomeie índices e colunas

Renomeie índices e colunas

- **Método 'columns'**
  Este método é usado para visualizar ou alterar os nomes das colunas em um DataFrame.

- **Método 'set_index**
  Este método é usado para transformar uma coluna em um índice em um DataFrame.

- **Método 'reset_index**
  Este método é usado para redefinir os índices de um DataFrame para seus valores padrão, que é uma sequência de números inteiros começando do zero.

Renomeie índices e colunas

Parâmetro 'drop'

Este parâmetro é usado com o método 'reset_index' para descartar a coluna de índice atual e evitar que ela se torne uma coluna no DataFrame.

Parâmetro 'inplace

Este parâmetro é usado para aplicar alterações diretamente ao DataFrame original, em vez de retornar uma cópia do DataFrame com as alterações.

Categorização e dummies

Categorização e dummies

Concatenar

É o processo de combinar duas ou mais estruturas de dados em uma. Neste contexto, foi usado para combinar os dados categorizados e as variáveis dummy em um único DataFrame.

Quartis

São valores que dividem uma distribuição de dados em quatro partes iguais. Na aula, foram usados para definir categorias com base na distribuição dos dados.

Compreenda a amostragem

Compreenda a amostragem

Método 'head'
Método em Python usado para selecionar as primeiras linhas de um DataFrame.

Método 'tail'
Método em Python usado para selecionar as últimas linhas de um DataFrame.

Método 'sample'
Método em Python que permite selecionar uma amostra aleatória de linhas de um DataFrame. Este método pode ser usado para selecionar um número específico de linhas ou uma fração do total de linhas.

Parâmetro 'random_state'
Parâmetro no método 'sample' que permite definir uma semente para o gerador de números aleatórios, garantindo que você obtenha os mesmos resultados cada vez que você executa o código.

Junte tabelas

Junte tabelas

Append

É um método em Python que serve como atalho para o método 'concat' ao longo do eixo das linhas. Ele junta as tabelas colocando uma embaixo da outra.

Inner Join

É um método de junção de tabelas que traz apenas os registros que estão presentes em ambas as tabelas.

Left Join

É um método de junção de tabelas que preenche primeiro com os dados da tabela da esquerda e depois com os dados da tabela da direita.

Merge

É um comando em Python usado para juntar duas tabelas em uma só, com base em uma chave comum.

Junte tabelas

- **Outer Join**
  É um método de junção de tabelas que traz todos os registros de ambas as tabelas, preenchendo com dados faltantes quando necessário.

- **Right Join**
  É um método de junção de tabelas que preenche primeiro com os dados da tabela da direita e depois com os dados da tabela da esquerda.

escola britânica de artes criativas & tecnologia

Bons estudos!