

Pandas no Jutsu

Preenchendo Lacuna como um Ninja

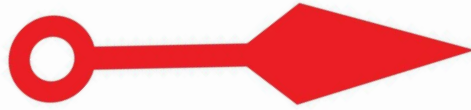


HEVANS VINICIUS PEREIRA

01

*Identificando
Datos Faltantes*

Identificando Dados Faltantes



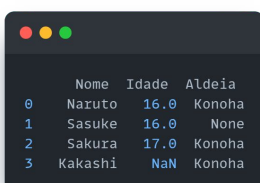
Antes de preencher os dados faltantes, precisamos identificá-los.

Exemplo:

Vamos criar um DataFrame simples com alguns dados faltantes:

```
data = {'Nome': ['Naruto', 'Sasuke', 'Sakura', 'Kakashi'],  
        'Idade': [16, 16, 17, None],  
        'Aldeia': ['Konoha', None, 'Konoha', 'Konoha']}  
df = pd.DataFrame(data)  
print(df)
```

Saída:



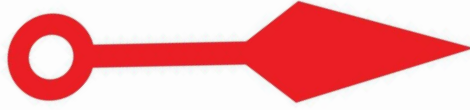
	Nome	Idade	Aldeia
0	Naruto	16.0	Konoha
1	Sasuke	16.0	None
2	Sakura	17.0	Konoha
3	Kakashi	NaN	Konoha

02

*Preenchendo com
Valor Constante*



Preenchendo com Valor Constante



Podemos preencher valores faltantes com um valor constante usando o método `fillna()`.

Exemplo:

Preenchendo a idade faltante com 18 e a aldeia faltante com 'Desconhecida':

```
df['Idade'].fillna(18, inplace=True)
df['Aldeia'].fillna('Desconhecida', inplace=True)
print(df)
```

Saída:

	Nome	Idade	Aldeia
0	Naruto	16.0	Konoha
1	Sasuke	16.0	Desconhecida
2	Sakura	17.0	Konoha
3	Kakashi	18.0	Konoha

03

*Preenchendo com
Moda, Média ou
Mediana*



Preenchendo com Moda, Média ou Mediana



Exemplo:

Vamos usar a média para preencher a idade faltante:

```
df['Idade'].fillna(df['Idade'].mean(), inplace=True)
print(df)
```

Saída:

snappify.com

	Nome	Idade	Aldeia
0	Naruto	16.0	Konoha
1	Sasuke	16.0	Desconhecida
2	Sakura	17.0	Konoha
3	Kakashi	16.33	Konoha

snappify.com

Preenchendo com Moda, Média ou Mediana



Para preencher com a mediana:

```
df['Idade'].fillna(df['Idade'].median(), inplace=True)
print(df)
```

Para preencher com a moda:

snappify.com

```
df['Aldeia'].fillna(df['Aldeia'].mode()[0], inplace=True)
print(df)
```

snappify.com

04

*Preenchendo com
Valores
Interpolados*



Preenchendo com Valores Interpolados



Interpolação é útil quando os dados estão relacionados em sequência, como séries temporais.

Exemplo:

```
data = {'Dia': [1, 2, 3, 4, 5],  
        'Vendas': [200, None, 250, None, 300]}  
df = pd.DataFrame(data)  
df['Vendas'] = df['Vendas'].interpolate()  
print(df)
```

Saída:

snappify.com

	Dia	Vendas
0	1	200.0
1	2	225.0
2	3	250.0
3	4	275.0
4	5	300.0

snappify.com

Conclusão



Conclusão



Assim como Naruto supera desafios para se tornar um Hokage, aprender a lidar com dados faltantes é essencial para se tornar um ninja dos dados.

Pratique com diferentes métodos e descubra qual é o melhor para o seu conjunto de dados.

Continue sua jornada ninja, explorando outras técnicas e aprofundando seus conhecimentos em ciência de dados!