

Aula 2: Manipulação Avançada de Dados com Pandas



Objetivo da Aula:

Aprender a realizar manipulações avançadas de dados com o Pandas, incluindo filtragem, agrupamento, tratamento de valores ausentes e combinações de DataFrames. Isso permitirá trabalhar com datasets reais de maneira mais eficiente.

1. Revisão da Estrutura de Dados no Pandas

- **Series:** Estruturas unidimensionais.
- **DataFrame:** Tabelas bidimensionais com colunas nomeadas e índices.

Se ainda não fez, lembre-se de importar o Pandas:

```
import pandas as pd
```

2. Seleção e Filtragem de Dados 🔍

Selecionando Colunas

Para acessar uma única coluna:

```
df['coluna']
```

Para acessar múltiplas colunas:

```
df[['coluna1', 'coluna2']]
```

Filtrando Linhas com Condições

Filtrar linhas com base em uma condição:

```
df[df['coluna'] > 10] # Apenas valores maiores que 10
```

Combinar múltiplas condições:

```
df[(df['coluna1'] > 10) & (df['coluna2'] == 'valor')]
```

3. Trabalhando com Valores Ausentes ?

Identificar Valores Ausentes

```
df.isnull() # Verifica onde há valores nulos  
df.isnull().sum() # Contagem de valores nulos por coluna
```

Lidando com Valores Ausentes

1. Remover Linhas/Colunas com Valores Ausentes

```
df.dropna() # Remove linhas com valores nulos  
df.dropna(axis=1) # Remove colunas com valores nulos
```

2. Substituir Valores Ausentes

```
df['coluna'].fillna(valor) # Substitui valores ausentes por um valor específico
```

4. Agrupamento e Agregação de Dados

Agrupar Dados com `groupby()`

Agrupar por uma ou mais colunas:

```
agrupado = df.groupby('coluna') # Cria um objeto agrupado
```

Executar operações em grupos:

```
agrupado['coluna2'].mean() # Calcula a média de cada grupo
```

Agregações Personalizadas

```
df.groupby('coluna')['coluna2'].agg(['mean', 'sum', 'max'])
```

5. Combinação de DataFrames

Concatenar DataFrames

Combinar DataFrames verticalmente:

```
df_novo = pd.concat([df1, df2], axis=0)
```

Combinar horizontalmente:

```
df_novo = pd.concat([df1, df2], axis=1)
```

Mesclar DataFrames com **merge**

Unir DataFrames com base em uma coluna comum:

```
df_unido = pd.merge(df1, df2, on='coluna_comum', how='inner') # Tipos: inner, outer, left, right
```

6. Conclusão e Próximos Passos 🚀

Hoje aprendemos a:

- Selecionar e filtrar dados no Pandas.
- Lidar com valores ausentes.
- Realizar agrupamentos e agregações.
- Combinar e mesclar DataFrames.

Próximo Passo: Na próxima aula, exploraremos a **análise exploratória de dados** com Pandas e bibliotecas de visualização como Matplotlib e Seaborn!