

Introdução ao Pandas e Fundamentos de Estatística



Objetivo da Aula:

Introduzir a biblioteca **Pandas** e os conceitos básicos de **estatística descritiva**, essenciais para análise de dados. Vamos aprender como trabalhar com dados tabulares e realizar cálculos simples para extrair informações úteis.

1. O que é o Pandas? 🐼

Pandas é uma biblioteca Python usada para manipulação e análise de dados. Ela permite trabalhar com:

- **Dados tabulares** (linhas e colunas) usando o `DataFrame` .
- **Séries temporais** e outros tipos de dados estruturados.

Principais vantagens:

- Fácil leitura e manipulação de dados.
- Compatibilidade com diversos formatos de dados: CSV, Excel, JSON, etc.
- Integração com outras bibliotecas como NumPy e Matplotlib.

2. Instalando e Importando o Pandas 🛠️

Instalação:

Se ainda não tiver o Pandas instalado, use o comando:

```
pip install pandas
```

Importando:

Para começar, importe a biblioteca:

```
import pandas as pd
```

3. Estruturas de Dados no Pandas

1. Series

Uma **Series** é uma lista de dados com um índice associado.

```
import pandas as pd

# Criando uma Series
dados = pd.Series([10, 20, 30, 40])
print(dados)
```

Saída:

```
0    10
1    20
2    30
3    40
dtype: int64
```

2. DataFrame

Um **DataFrame** é uma tabela bidimensional com linhas e colunas.

```
# Criando um DataFrame
dados = {
    'Nome': ['Ana', 'João', 'Maria'],
    'Idade': [20, 25, 22]
}
df = pd.DataFrame(dados)
print(df)
```

Saída:

| | Nome | Idade |
|---|-------|-------|
| 0 | Ana | 20 |
| 1 | João | 25 |
| 2 | Maria | 22 |

4. Operações Básicas com Pandas

Carregar Dados de um Arquivo CSV:

```
df = pd.read_csv('arquivo.csv')
```

Visualizar os Dados:

- Exibir as primeiras linhas: `df.head()`
- Exibir as últimas linhas: `df.tail()`

Informações sobre o DataFrame:

- `df.info()` - Mostra informações gerais (colunas, tipos, valores ausentes).
- `df.describe()` - Estatísticas descritivas (média, mediana, etc.).

5. Estatística Descritiva

Principais Métricas:

1. **Média (mean):** Valor médio dos dados.

```
media = df['coluna'].mean()
```

2. **Mediana (median):** Valor central dos dados ordenados.

```
mediana = df['coluna'].median()
```

3. **Moda (mode):** Valor mais frequente nos dados.

```
moda = df['coluna'].mode()
```

4. **Desvio Padrão (std):** Mede a dispersão dos dados.

```
desvio_padrao = df['coluna'].std()
```

Exemplo Prático:

```
import pandas as pd

dados = {'Idades': [22, 25, 30, 22, 25, 35]}
df = pd.DataFrame(dados)

print("Média:", df['Idades'].mean())
print("Mediana:", df['Idades'].median())
print("Moda:", df['Idades'].mode()[0])
print("Desvio Padrão:", df['Idades'].std())
```


6. Conclusão e Próximos Passos 🚀

Resumo: Hoje aprendemos a:

- Introduzir o Pandas.
- Criar e manipular `Series` e `DataFrames`.
- Realizar cálculos estatísticos básicos.

Próximo Passo: Na próxima aula, exploraremos funcionalidades avançadas do Pandas, como filtragem, agrupamento e manipulação de dados!