

Python: análise de dados com Pandas II

Poliana N Ferreira

Filtros



Agregação

Forma de escolher apenas um pedaço do DF a partir de uma condição

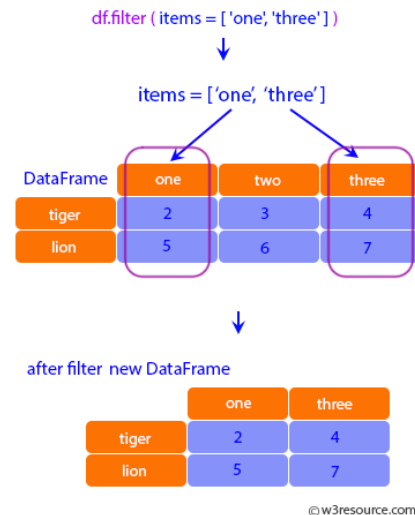
- Filtros comuns
- Query()
- Filtros compostos
- Isin()
- Uso de loc
- Negação
- Startswith()
- Lógica mais complexa
- filter() – colunas

Filter DataFrame on numeric value - greater than or equal to

	df					
	Year	Gender	Rank	Name	Count	
0	1982	Female	5	MICHELLE	2446	
1	2000	Female	24	DESTINY	1188	
2	1992	Female	6	AMANDA	2776	
3	1991	Female	23	JASMINE	1482	
4	1967	Male	19	THOMAS	2149	
5	2013	Female	1	SOPHIA	3447	
6	1980	Male	14	BRIAN	2737	
7	1983	Female	16	VANESSA	1517	
8	1995	Male	16	CHRISTIAN	2761	
9	1960	Male	21	PAUL	2136	

df [df['Rank'] >= 19]
df [df['Rank'].ge(19)]

	Year	Gender	Rank	Name	Count
1	2000	Female	24	DESTINY	1188
3	1991	Female	23	JASMINE	1482
4	1967	Male	19	THOMAS	2149
9	1960	Male	21	PAUL	2136



Funções Personalizadas



Funções personalizadas

Criamos funções para aplicar em elementos da base de dados

- Apply()
- Funções comuns
- Funções lambda
- Linhas e colunas
- Funções com texto
- Map()

```
def fahr_to_celsius(temp):  
    return ((temp - 32) * (5/9))
```

	Palavra-chave def
	Nome da Função
	Parâmetro
	Comando de Retorno
	Valor de Retorno

Agregações e Junções



Agregação

O que é?

- Operação essencial na análise de dados, permitindo resumir informações estatísticas de conjuntos de dados.
- No Pandas, a agregação pode ser realizada através de métodos como `groupby` e `agg`.
 - `groupby`: permite agrupar dados com base em uma ou mais colunas, formando um grupo para cada valor único da coluna especificada.
 - `agg`: utilizado para aplicar operações de agregação aos grupos formados, suportando funções como soma, média, desvio padrão, entre outras.

```
df.groupby('coluna').agg({'coluna_a': 'mean', 'coluna_b': 'sum'})
```

Junção

O que é?

- A junção de dados é uma técnica fundamental na análise de dados, permitindo combinar dados de diferentes fontes com base em chaves comuns.
- Pandas oferece métodos como `merge` e `join` para realizar junções, suportando operações equivalentes às junções SQL (INNER, LEFT, RIGHT, FULL OUTER).

```
pd.merge(df1, df2, on='coluna_chave', how='inner')
```


Exemplo



Agora vamos avaliar a aula?



Obrigada!