



Aula 29 - Comandos da Propriedade Dataframe

☰ Ciclo	Ciclo 04: Manipulação de Dados
# Aula	29
🕒 Created	@November 12, 2022 9:38 AM
☑ Reviewed	<input type="checkbox"/>
📎 Material PDF	
☑ Finished	<input checked="" type="checkbox"/>

Objetivo da Aula:

1. Mostrar os principais comandos de análise de dados
2. Na próxima aula

Conteúdo:

▼ Comandos de propriedade do Dataframe

São comandos nativos da estrutura de dados Dataframe

▼ Principais comandos de inspeção

Comando	Comando em Python
Descobrir as dimensões da	<code>df.shape</code>

planilha de dados	
Mostrar as X primeiras linhas	<code>df.head(X)</code>
Mostrar as X últimas linhas	<code>df.tail(X)</code>
Descobrir o tipos da colunas	<code>df.dtypes</code>
Lista com valores únicos de uma coluna	<code>df[<COLUNA>].unique()</code>
Ordenar o dataframe baseado em uma coluna	<code>df.sort_values(<COLUNA>, ascending=True/False)</code>
Somar todas as linhas de uma coluna	<code>df[<COLUNA>].sum()</code>
Media todas as linhas de uma coluna	<code>df[<COLUNA>].mean()</code>
Converter os tipos de colunas	<code>df[<COLUNA>].astype(<NOVO TIPO>)</code>
Converter para tipos de datas	<code>pd.to_datetime(df[<COLUNA>], format='%d-%m-%Y')</code>

▼ Limpeza do conjunto de dados

```
# Lendo do Arquivo
df = pd.read_csv( '../dataset/train.csv' )

# Remover spaco da string
for i in range( len( df ) ):
    df.loc[i, 'ID'] = df.loc[i, 'ID'].strip()
    df.loc[i, 'Delivery_person_ID'] = df.loc[i, 'Delivery_person_ID'].strip()

# Excluir as linhas com a idade dos entregadores vazia
# ( Conceitos de seleção condicional )
linhas_vazias = df['Delivery_person_Age'] != 'NaN '
df = df.loc[linhas_vazias, :]

# Conversao de texto/categoria/string para numeros inteiros
df['Delivery_person_Age'] = df['Delivery_person_Age'].astype( int )
```

```
# Conversao de texto/categoria/strings para numeros decimais
df['Delivery_person_Ratings'] = df['Delivery_person_Ratings'].astype( float )

# Conversao de texto para data
df['Order_Date'] = pd.to_datetime( df['Order_Date'], format='%d-%m-%Y' )

#
linhas_vazias = df['multiple_deliveries'] != 'NaN '
df = df.loc[linhas_vazias, :]
df['multiple_deliveries'] = df['multiple_deliveries'].astype( int )

# Comando para remover o texto de números
df = df.reset_index( drop=True )
for i in range( len( df ) ):
    df.loc[i, 'Time_taken(min)'] = re.findall( r'\d+', df.loc[i, 'Time_taken(min)'] )
```

▼ Na próxima aula

Aula 30: As perguntas de negócio