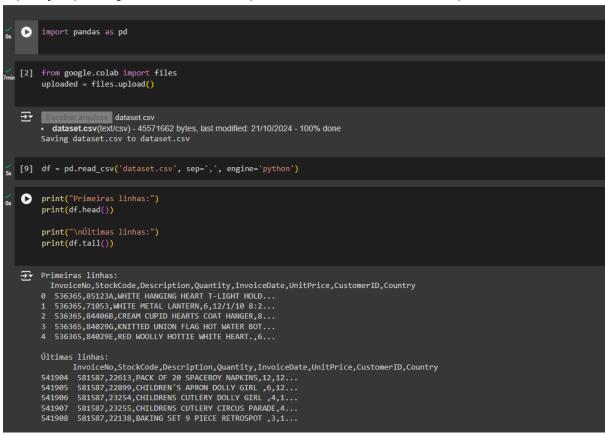
## Estácio

Nível 3 - Mundo 5 | Desenvolvimento FullStack 2024 Caique Thomaz Stopiglia - 202208292471

## Tratando a imensidão dos dados

## Realização de limpeza de dados utilizando comandos em Python com a biblioteca Pandas e Google Colab.

**1** - Carregamento do Pandas, upload do arquivo csv, leitura do arquivo csv utilizando separação por vírgula e exibindo as 5 primeiras e últimas linhas do arquivo.



2 - Verificando separador e dados.

```
[5] df.info()
 <class 'pandas.core.frame.DataFrame'>
     RangeIndex: 541909 entries, 0 to 541908
     Data columns (total 1 columns):
      # Column
                                                                                              Non-Null Count Dtype
      0 InvoiceNo,StockCode,Description,Quantity,InvoiceDate,UnitPrice,CustomerID,Country 541909 non-null object
     dtypes: object(1)
     memory usage: 4.1+ MB
[6] df_copy = df.copy()
[8] print(df.columns)
Triangle (['InvoiceNo,StockCode,Description,Quantity,InvoiceDate,UnitPrice,CustomerID,Country'], dtype='object')
[10] print(df.columns)
     print(df.head())
Description Quantity \
         536365 85123A WHITE HANGING HEART T-LIGHT HOLDER
536365 71053 WHITE METAL LANTERN
536365 84406B CREAM CUPID HEARTS COAT HANGER
536365 84029G KNITTED UNION FLAG HOT WATER BOTTLE
         536365 84029E
                              RED WOOLLY HOTTIE WHITE HEART.
         InvoiceDate UnitPrice CustomerID
                                                   Country
                                 17850.0 United Kingdom
     0 12/1/10 8:26 2,55
                                   17850.0 United Kingdom
       12/1/10 8:26
                                   17850.0 United Kingdom
       12/1/10 8:26
                                   17850.0 United Kingdom
        12/1/10 8:26
                          3.39
        12/1/10 8:26
                                   17850.0 United Kingdom
                          3.39
```

**3** - Substituindo valores nulos na coluna 'CustomerID' por 0 e na Coluna 'InvoiceDate' por '1900/01/01'.

```
print(of.hear())

fillna(c), implace=True)

fillna(c), i
```

**4** - Transformando a coluna 'InvoiceDate' em Datetime, resolvendo erros de formatação na coluna 'InvoiceData' e removendo registros com valores nulos.

```
| The content of the
```

5 - Verificando alterações.

