

Aula_1_DTS_PLN_Exercício_1_corrigido

June 30, 2025

#Exercícios - Aula 1

0.1 1) Dado o dataset de produtos [1]:

[1] - <https://dados-ml-pln.s3-sa-east-1.amazonaws.com/produtos.csv>

```
[1]: import pandas as pd

df = pd.read_csv(
    "https://dados-ml-pln.s3-sa-east-1.amazonaws.com/produtos.csv",
    delimiter=";",
    encoding='utf-8' )
```

```
[2]: df.head()
```

```
[2]:
```

	nome \
0	O Hobbit - 7ª Ed. 2013
1	Livro - It A Coisa - Stephen King
2	Box As Crônicas De Gelo E Fogo Pocket 5 Li...
3	Box Harry Potter
4	Livro Origem - Dan Brown

	descricao	categoria
0	Produto NovoBilbo Bolseiro é um hobbit que lev...	livro
1	Produto NovoDurante as férias escolares de 195...	livro
2	Produto NovoTodo o reino de Westeros ao alcanc...	livro
3	Produto Novo e Físico A série Harry Potter ch...	livro
4	Produto NovoDe Onde Viemos? Para Onde Vamos? R...	livro

1.1. Analise o % de valores nulos no dataset

```
[3]: df.describe()
```

```
[3]:
```

	nome \
count	4080
unique	3696
top	Mais Escuro - Cinquenta Tons Mais Escuros Pel...
freq	20

	descricao	categoria
count	2916	4080
unique	2460	4
top	JOGO ORIGINAL. NOVO. LACRADO. PRONTA ENTREGA. ...	livro
freq	39	1020

```
[4]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4080 entries, 0 to 4079
Data columns (total 3 columns):
#   Column      Non-Null Count  Dtype
---  -
0   nome        4080 non-null   object
1   descricao   2916 non-null   object
2   categoria   4080 non-null   object
dtypes: object(3)
memory usage: 95.8+ KB
```

```
[5]: df.count()
```

```
[5]: nome        4080
descricao    2916
categoria    4080
dtype: int64
```

```
[6]: # dimensão do dataframe
df.shape[0]
```

```
[6]: 4080
```

```
[7]: print(df.shape[0], df.shape[1])
```

```
4080 3
```

```
[8]: # conta valores faltantes/nulos/missing
df.isnull().sum()
```

```
[8]: nome        0
descricao    1164
categoria    0
dtype: int64
```

```
[9]: # calcula o % de nulos no dataset
df.isnull().sum()/df.shape[0]
```

```
[9]: nome        0.000000
descricao    0.285294
categoria    0.000000
```

dtype: float64

1.2. Remova os registros/linhas com valores nulos, se houver

```
[10]: # dropna(inplace=True), inplace=True => modifica o próprio objeto.
      ↪ dropna(inplace=False), inplace=False => retorna uma cópia do objeto
df.dropna(inplace=True)
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 2916 entries, 0 to 4079
Data columns (total 3 columns):
#   Column      Non-Null Count  Dtype
---  -
0   nome        2916 non-null   object
1   descricao   2916 non-null   object
2   categoria   2916 non-null   object
dtypes: object(3)
memory usage: 91.1+ KB
```

```
[11]: df = df.dropna()
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 2916 entries, 0 to 4079
Data columns (total 3 columns):
#   Column      Non-Null Count  Dtype
---  -
0   nome        2916 non-null   object
1   descricao   2916 non-null   object
2   categoria   2916 non-null   object
dtypes: object(3)
memory usage: 91.1+ KB
```

1.3. Analise a distribuição das “categorias”

```
[12]: df.categoria
```

```
[12]: 0      livro
      1      livro
      2      livro
      3      livro
      4      livro
      ...
4073   game
4074   game
4075   game
4078   game
4079   game
```

Name: categoria, Length: 2916, dtype: object

```
[13]: df.categoria.value_counts()
```

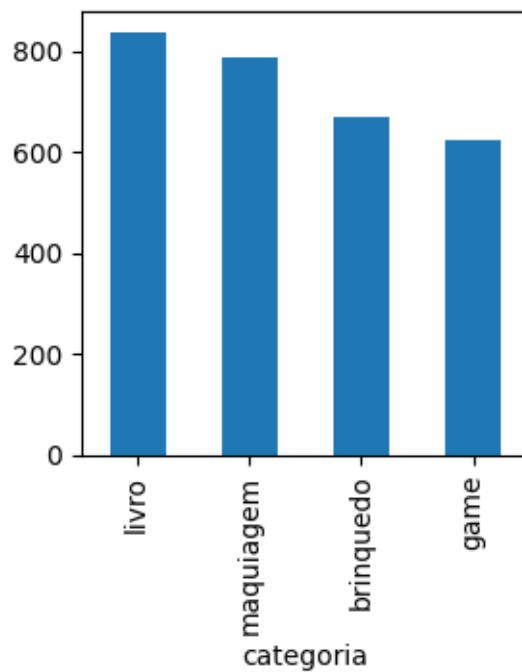
```
[13]: categoria  
livro      838  
maquiagem  788  
brinquedo  668  
game       622  
Name: count, dtype: int64
```

```
[14]: (df.categoria.value_counts()/df.shape[0])*100  
# df.categoria.value_counts()/df.categoria.count() # outra forma
```

```
[14]: categoria  
livro      28.737997  
maquiagem  27.023320  
brinquedo  22.908093  
game       21.330590  
Name: count, dtype: float64
```

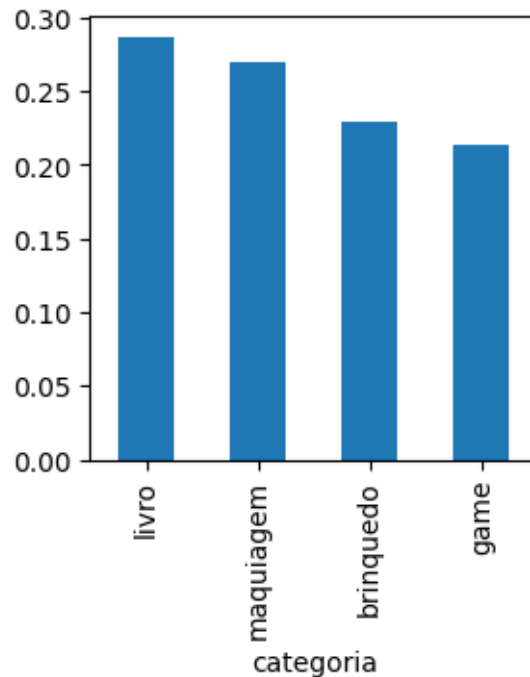
```
[15]: df.categoria.value_counts().plot(kind='bar', figsize=(3,3))
```

```
[15]: <Axes: xlabel='categoria'>
```



```
[16]: (df.categoria.value_counts()/df.shape[0]).plot(kind='bar', figsize=(3,3))
```

```
[16]: <Axes: xlabel='categoria'>
```



1.4 Crie uma nova coluna chamada “texto“, concatenando as colunas”nome” e “descricao”

```
[17]: # Exemplo de concatenação de informações  
print("Texto 1" + " " + "Texto 2")
```

Texto 1 Texto 2

```
[18]: df["texto"] = df['nome'] + " " + df['descricao'] # cria uma nova culuna com os  
↳ valores concatenados  
df.texto[0]
```

```
[18]: ' O Hobbit - 7ª Ed. 2013 Produto NovoBilbo Bolseiro é um hobbit que leva uma  
vida confortável e sem ambições. Mas seu contentamento é perturbado quando  
Gandalf, o mago, e uma companhia de anões batem à sua porta e levam-no para uma  
expedição. Eles têm um plano para roubar o tesouro guardado por Smaug, o  
Magnífico, um grande e perigoso dragão. Bilbo reluta muito em participar da  
aventura, mas acaba surpreendendo até a si mesmo com sua esperteza e sua  
habilidade como ladrão!CaracterísticasAutor: Tolkien, J. R. R.Peso:  
0.44I.S.B.N.: 9788578277109Altura: 20.000000Largura: 13.000000Profundidade:  
1.000000Número de Páginas: 328Idioma: PortuguêsAcabamento: BrochuraNúmero da  
edição: 7Ano da edição: 2013'
```

```
[19]: df.head()
```

```
[19]:
```

	nome	\
0	O Hobbit - 7ª Ed. 2013	
1	Livro - It A Coisa - Stephen King	
2	Box As Crônicas De Gelo E Fogo Pocket 5 Li...	
3	Box Harry Potter	
4	Livro Origem - Dan Brown	

	descricao	categoria	\
0	Produto NovoBilbo Bolseiro é um hobbit que lev...	livro	
1	Produto NovoDurante as férias escolares de 195...	livro	
2	Produto NovoTodo o reino de Westeros ao alcanc...	livro	
3	Produto Novo e Físico A série Harry Potter ch...	livro	
4	Produto NovoDe Onde Viemos? Para Onde Vamos? R...	livro	

	texto
0	O Hobbit - 7ª Ed. 2013 Produto NovoBilbo Bol...
1	Livro - It A Coisa - Stephen King Produto No...
2	Box As Crônicas De Gelo E Fogo Pocket 5 Li...
3	Box Harry Potter Produto Novo e Físico A sé...
4	Livro Origem - Dan Brown Produto NovoDe Onde...

1.5 Descubra as 10 palavras que mais ocorrem nessa nova coluna “texto”

```
[20]: from collections import Counter
counter = Counter()

counter.update('a')
counter.update('b')
print(counter)

counter['a'] += 1
print(counter)
counter.most_common(10)
```

```
Counter({'a': 1, 'b': 1})
Counter({'a': 2, 'b': 1})
```

```
[20]: [('a', 2), ('b', 1)]
```

```
[21]: "você, sabia?".split()
```

```
[21]: ['você,', 'sabia?']
```

```
[22]: df.texto.str.split() # default do split é o <espaço> como separador
#df.dtypes
```

```
[22]: 0      [0, Hobbit, -, 7ª, Ed., 2013, Produto, NovoBil...
      1      [Livro, -, It, A, Coisa, -, Stephen, King, Pro...
      2      [Box, As, Crônicas, De, Gelo, E, Fogo, Pocket,...
      3      [Box, Harry, Potter, Produto, Novo, e, Físico,...
      4      [Livro, Origem, -, Dan, Brown, Produto, NovoDe...

      ...

      4073     [Red, Dead, Redemption, Edição, Do, Ano, Goty,...
      4074     [Jogo, Gta, 5, Grand, Theft, Auto, V, Ps4, Míd...
      4075     [Zelda:, Breath, Of, The, Wild, -, Expansion, ...
      4078     [Gta, San, Andreas, Hd, Remastered, Ps3, Envio...
      4079     [Mini, Game, Nova, Portátil, 10mil, Jogos, Pla...
      Name: texto, Length: 2916, dtype: object
```

```
[23]: df.texto.str.lower().str.split() # default do split é o <espaço> como separador
      #df.dtypes
```

```
[23]: 0      [o, hobbit, -, 7ª, ed., 2013, produto, novobil...
      1      [livro, -, it, a, coisa, -, stephen, king, pro...
      2      [box, as, crônicas, de, gelo, e, fogo, pocket,...
      3      [box, harry, potter, produto, novo, e, físico,...
      4      [livro, origem, -, dan, brown, produto, novode...

      ...

      4073     [red, dead, redemption, edição, do, ano, goty,...
      4074     [jogo, gta, 5, grand, theft, auto, v, ps4, míd...
      4075     [zelda:, breath, of, the, wild, -, expansion, ...
      4078     [gta, san, andreas, hd, remastered, ps3, envio...
      4079     [mini, game, nova, portátil, 10mil, jogos, pla...
      Name: texto, Length: 2916, dtype: object
```

```
[24]: df.texto.str.lower().str.split(" ") # transforma a descrição em minúsculas e
      ↳ depois separa as palavras pelo <espaço> e retorna no formato de lista
```

```
[24]: 0      [, o, hobbit, -, 7ª, ed., 2013, , produto, nov...
      1      [, livro, -, it, a, coisa, -, stephen, king, ,...
      2      [, box, , as, crônicas, de, gelo, e, fogo, , p...
      3      [, box, harry, potter, , produto, novo, e, fís...
      4      [, livro, origem, -, dan, brown, , produto, no...

      ...

      4073     [, red, dead, redemption, edição, do, ano, got...
      4074     [, jogo, gta, 5, grand, theft, auto, v, ps4, m...
      4075     [, zelda:, breath, of, the, wild, -, expansion...
      4078     [, gta, san, andreas, hd, remastered, ps3, env...
      4079     [, mini, game, , nova, portátil, 10mil, jogos,...
      Name: texto, Length: 2916, dtype: object
```

```
[25]: from collections import Counter # implementa um container de tipo de dados. Ex:
      ↳ dict, list, set, and tuple
```

```
c = Counter() # subclasse dict para contar objetos
```

```
[26]: df.texto.str.lower().str.split().apply(c.update) # implementa a contagem

c.most_common(10)
```

```
[26]: [('de', 24300),
      ('e', 16042),
      ('o', 12424),
      ('a', 12152),
      ('do', 8791),
      ('para', 8032),
      ('-', 7413),
      ('em', 6900),
      ('que', 6797),
      ('com', 6247)]
```

```
[27]: # Outra forma de implementar
from collections import Counter

palavras = ' '.join(df['texto'].str.lower()).split()

contagem_palavras = Counter(palavras)
contagem_palavras.most_common(10)
```

```
[27]: [('de', 24300),
      ('e', 16042),
      ('o', 12424),
      ('a', 12152),
      ('do', 8791),
      ('para', 8032),
      ('-', 7413),
      ('em', 6900),
      ('que', 6797),
      ('com', 6247)]
```

Bônus! Monte uma nuvem de palavras

```
[28]: import matplotlib.pyplot as plt

try:
    from wordcloud import WordCloud
except:
    !pip install wordcloud
    from wordcloud import WordCloud

livro_words = ''
```



```
# Gera uma imagem da nuvem de palavras (Generate a word cloud image)
livro_wordcloud = WordCloud(width=600, height=400).generate(livro_words)

# Word cloud da categoria Livro
plt.figure( figsize=(4,4), facecolor='k')
plt.imshow(livro_wordcloud)
plt.axis("off")
plt.tight_layout(pad=0) # Preenchimento entre a borda e a figura
plt.show()
```

