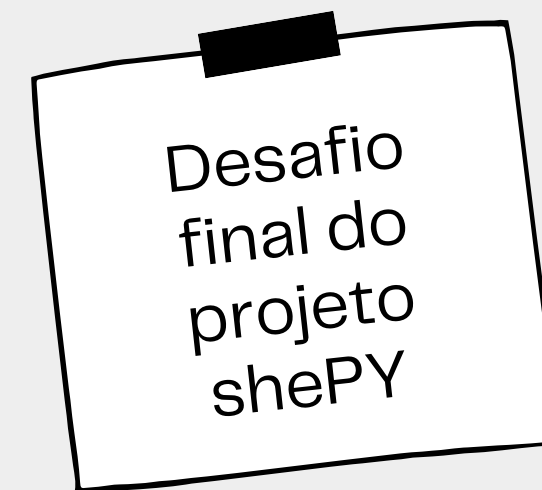


Análise dos 50 livros mais vendidos pela Amazon entre os anos de 2009 - 2019



Q Processos realizados

- 1 Análise dos dados presentes no dataset;
- 2 Livros de ficção e não ficção vendidos durante os anos de 2009 - 2019;
- 3 Valores e avaliação dos leitores;
- 4 Correlação de dados;
- 5 Correlação de todos os dados presentes no dataset.

Análise dos dados presentes no *dataset*

Explicação dos códigos Python selecionados e para que servem.

O que as funções realizadas nos revelaram?

Processos realizadas através de códigos no Google Colab

O repositório diz respeito aos 50 livros mais vendidos entre 2009 - 2019, totalizando 550 obras!

```
[ ] booklist.shape # Quantidade de informações disponíveis no dataset  
len(booklist)
```

550

```
[ ] len(booklist.Name.value_counts()) # Informações (livros) que não se repetem no dataset
```

351

```
[ ] # Quais podem ser os livros repetidos que se encontram no repositório?
```

```
booklist.columns  
print('Name:', '\n', booklist['Name'].value_counts())
```

```
Name:  
Publication Manual of the American Psychological Association, 6th Edition    10  
StrengthsFinder 2.0                                                         9  
Oh, the Places You'll Go!                                                  8  
The Very Hungry Caterpillar                                               7  
The 7 Habits of Highly Effective People: Powerful Lessons in Personal Change 7  
..                                                                           ..  
No Easy Day: The Autobiography of a Navy Seal: The Firsthand Account of the Mission That Killed Osama Bin Laden 1  
Little Bee: A Novel                                                        1  
George Washington's Sacred Fire                                           1  
The Racketeer                                                             1  
The Ugly Truth (Diary of a Wimpy Kid, Book 5)                            1  
Name: Name, Length: 351, dtype: int64
```

Resultados obtidos:

1. De 550 livros, 351 se repetem no dataset;
2. O livro que mais se repete é: Publication Manual of the American Psychological Association, 6th Edition.

Análise dos dados presentes no dataset

A biblioteca utilizada foi **seaborn**, inserida aqui, como **sns**

Q Elaboração do código para a realização do gráfico

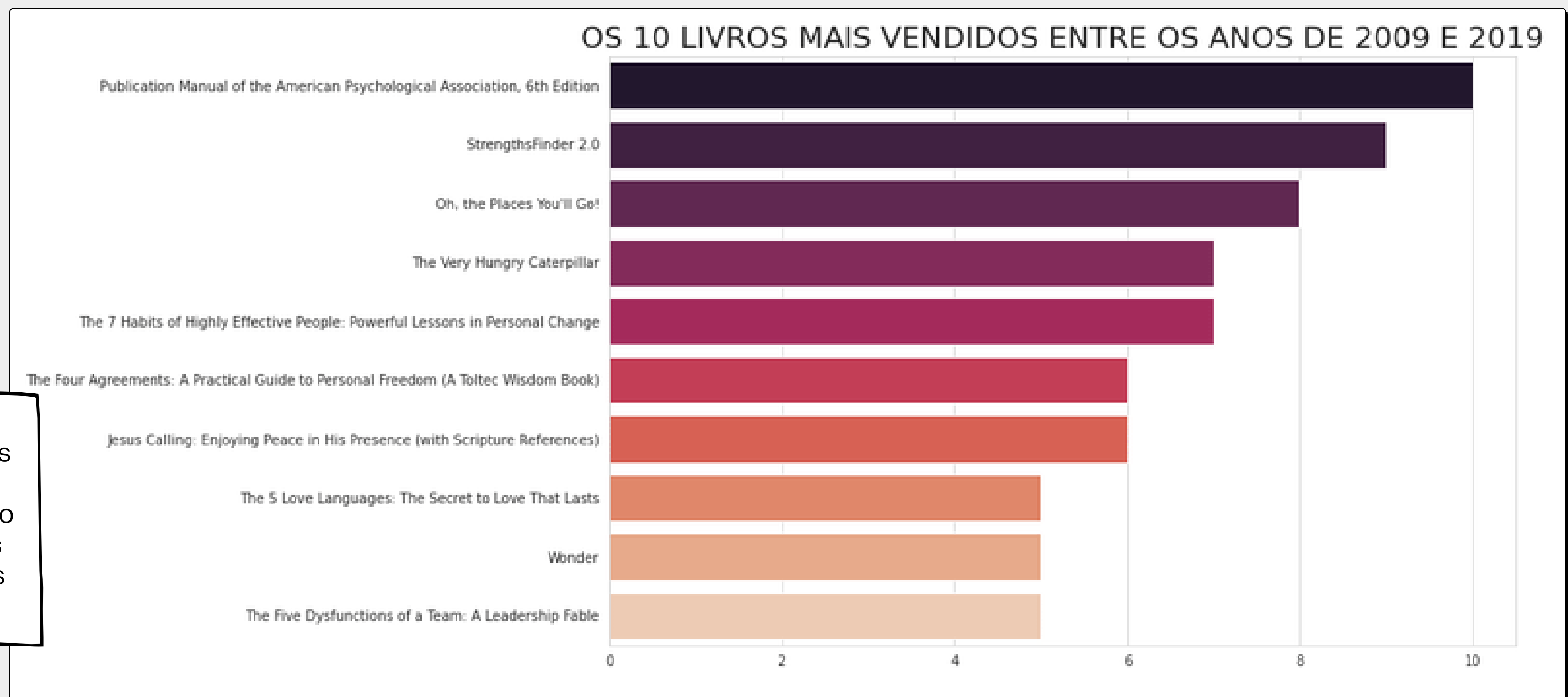
```
[ ] # Quais são os 10 livros que mais se repetem e consequentemente são os mais vendidos da Amazon entre os anos de 2009 e 2019?

tenbook = booklist ['Name'].value_counts()[:10]

fig, ax = plt.subplots(figsize = (12, 8))
sns.barplot(y = tenbook.index, x = tenbook, ax = ax, palette = 'rocket')
ax.set_xlabel(None)
ax.set_title('OS 10 LIVROS MAIS VENDIDOS ENTRE OS ANOS DE 2009 E 2019', fontdict = {'fontsize':22})

fig.show()
```

Observação: os livros com maior repetição na lista são os mais vendidos pela Amazon.



Análise dos dados presentes no dataset

Q Autores que mais se repetem no repositório

```
[ ] # Quanto aos autores, quais são os que mais se repetem no dataset?
```

```
booklist.columns
```

```
print('Author:', '\n', booklist['Author'].value_counts())
```

```
Author:
Jeff Kinney          12
Rick Riordan         11
Gary Chapman         11
Suzanne Collins      11
American Psychological Association 10
..
Brian Kilmeade        1
Michael Wolff          1
Atul Gawande           1
Raina Telgemeier       1
John Heilemann         1
Name: Author, Length: 248, dtype: int64
```

- 1 Diversos escritores possuem obras voltadas aos jovens, como Jeff Kinney, Rick Riordan, Stephenie Meyer, J.K. Rowling e Suzanne Collins;
- 2 Observa-se também obras voltadas a crianças e adultos.

Q Elaboração do código para a elaboração do gráfico

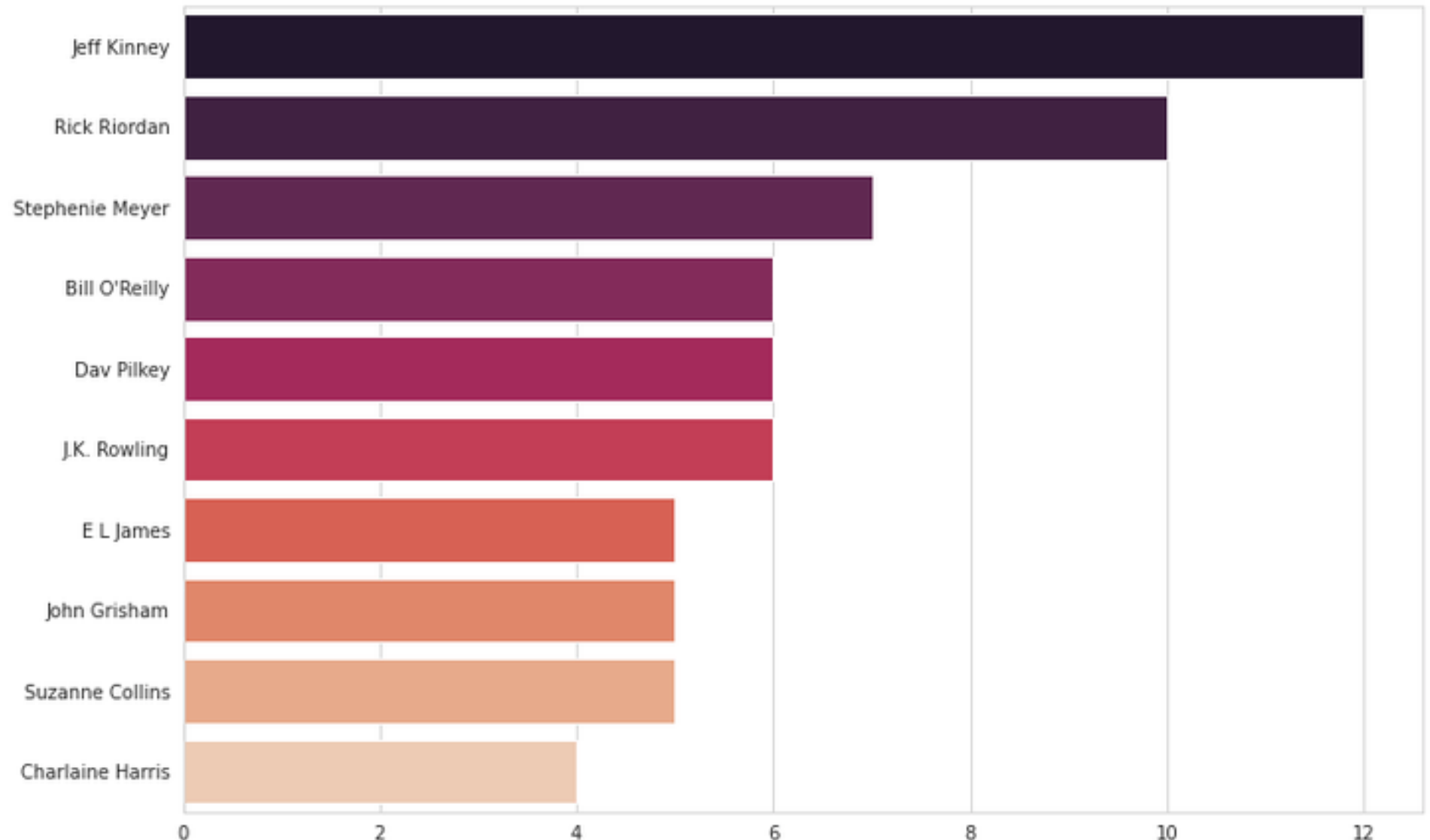
```
[ ] # Quais são os 10 autores que mais se repetem entre os livros vendidos da Amazon entre os anos de 2009 e 2019?

bestauthors = booklist[['Name', 'Author']].groupby('Author').agg('nunique').nlargest(10, columns = 'Name')

fig, ax = plt.subplots(figsize = (12,8))
sns.barplot(x = bestauthors.Name, y=bestauthors.index, ax = ax, palette = 'rocket')
ax.set_xlabel(None)
ax.set_ylabel(None)
ax.set_title('OS 10 AUTORES MAIS COMPRADOS DURANTE OS ANOS DE 2009 E 2019', fontdict = {'fontsize':22})

fig.show()
```

OS 10 AUTORES MAIS COMPRADOS DURANTE OS ANOS DE 2009 E 2019



Livros de ficção e não ficção vendidos durante os anos de 2009 - 2019

```
booklist.columns  
print('Genre:', '\n', booklist['Genre'].value_counts())
```

```
Genre:  
Non Fiction    310  
Fiction        240  
Name: Genre, dtype: int64
```

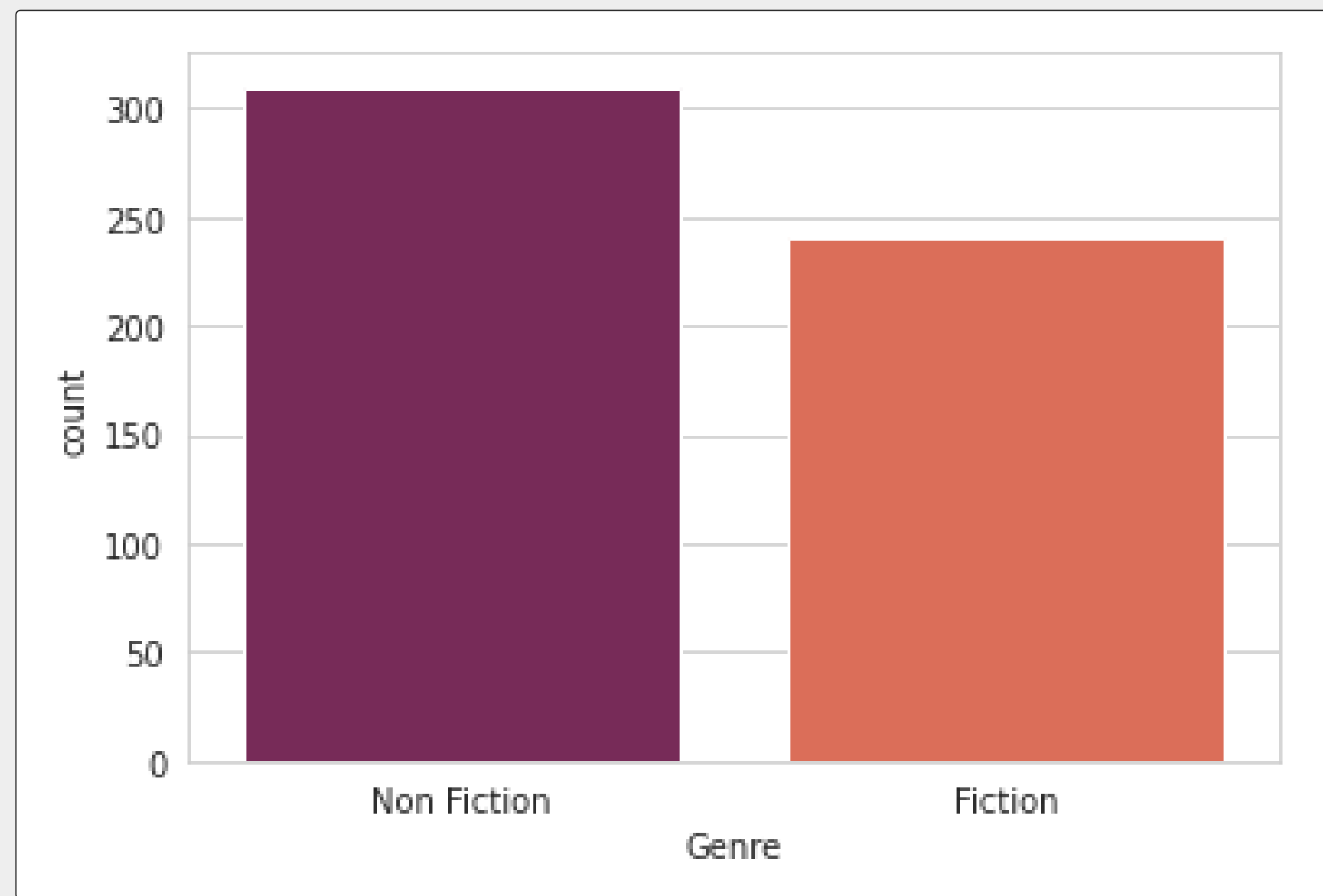
Verificou-se
que 310
livros são de
não ficção, e
240 são de
ficção!

Bibliotecas
importadas
seaborn (sns)
matplotlib.pyplot (plt).

```
[ ] # Podemos transformar esse resultado em um gráfico?
```

```
sns.countplot('Genre', data = booklist, palette = 'rocket')
```

```
plt.show()
```



Livros de ficção e não ficção vendidos durante os anos de 2009 - 2019

Maior venda de obras de não ficção: 2015.
2013 e 2017: aumento das vendas das obras de ficção.

```
[ ] # Qual foi a porcentagem de livros de ficção e não ficção vendidos durante cada ano do cenário selecionado para a análise?

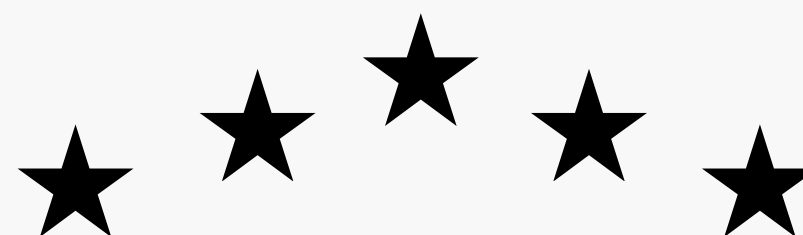
fig,ax = plt.subplots(int(np.ceil(booklist.Year.value_counts().shape[0]/3)),3,sharex = False,sharey = False,figsize = (15,15))
bookyear = booklist.Year.value_counts().sort_index(ascending=True).index
k = 0
for i in range(int(np.ceil(bookyear.value_counts().shape[0]/3))):
    for j in range(3):
        if k < 11:
            booklist[booklist.Year == bookyear[k]].Genre.value_counts().plot.bar(color = ['purple', 'pink'],ax = ax[i][j])
            ax[i][j].set_title(f"ANO {bookyear[k]}")
            k = k + 1
        else:
            break
    plt.tight_layout()

plt.show()
```

Utilização do
matplotlib.pyplot,
chamado
aqui de **plt**.



Valor e avaliação dos leitores



```
[13] # Maior e menor valor dos livros vendidos

booklist.columns
print('Price:', '\n', booklist['Price'].value_counts())
```

```
Price:
8      52
6      38
9      38
5      36
11     35
4      32
14     30
13     29
10     28
12     27
7      23
15     21
16     20
17     19
20     17
18     14
0      12
46     10
21      9
28      6
22      6
40      5
23      5
24      5
2       5
30      5
19      4
27      4
25      2
32      2
105     2
82      1
36      1
39      1
42      1
52      1
53      1
3       1
54      1
1       1
Name: Price, dtype: int64
```

Com esses dados
pode-se trabalhar
média, mediana e
a moda dos
preços presentes
na lista

```
[14] # Avaliações dadas pelos leitores e a respectiva quantidade
```

```
booklist.columns
print('User Rating:', '\n', booklist['User Rating'].value_counts())
```

```
User Rating:
4.8     127
4.7     108
4.6     105
4.5      60
4.9      52
4.4      38
4.3      25
4.0      14
4.2       8
4.1       6
3.9       3
3.8       2
3.3       1
3.6       1
Name: User Rating, dtype: int64
```

1 Também é possível observar que a maior quantidade de livros vendidos dizem respeito a obras que não ultrapassam 50,00. Como destaque para 52 livros que custam \$: 08,00.

Quando o assunto é referente a fração de livros vendidos que possuam um valor maior de \$: 100,00, é possível observar que apenas 2 obras possuem \$:105,00.

2 Tratando-se das avaliações dos leitores, considera-se que grande parte das obras apresentam valores acima de 4.5. Nos expondo, que em média os bestsellers apresentaram um valor significativo para os leitores.

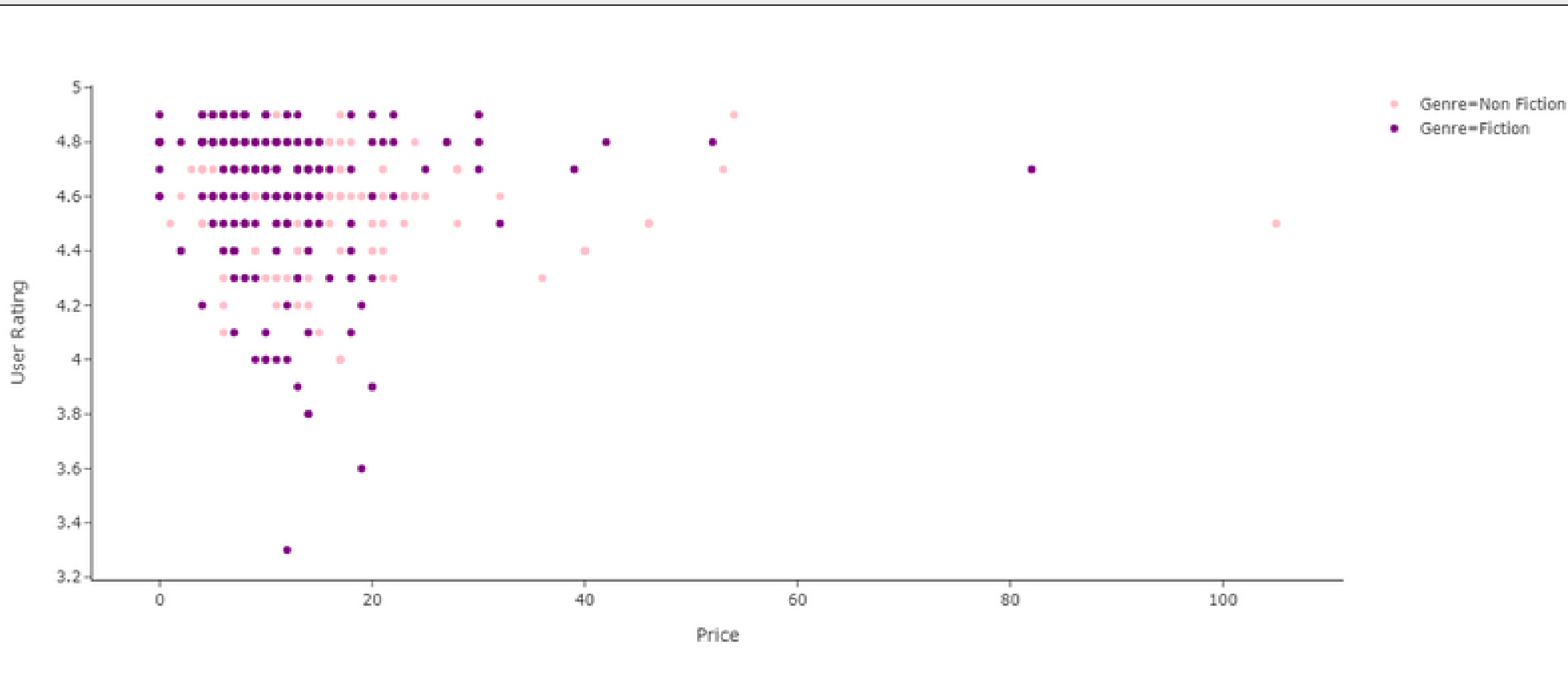
Correlação de dados

Q Elaboração do código para a realização da análise de correção

```
[ ] # Correlação entre o gênero dos livros (ficção e não ficção), o preço e a avaliação dos leitores
# Price: Preço
# Genre: Gênero
# User Rating: Média da Avaliação dos Leitores

fig = px.scatter(booklist, x = 'Price', y = 'User Rating', color = 'Genre', color_discrete_map = {'Fiction': 'purple', 'Non Fiction': 'pink'})

fig.update_layout(template = 'simple_white')
```



- 1 Os livros de maior avaliação são do gênero ficção, assim como os livros de menor avaliação;
- 2 As obras de não ficção não foram avaliadas com menos de 4,00 estrelas;
- 3 Em sua maioria, os livros de não ficção custam de 8,00 a 29,00;
- 4 A obra destaque em preço é uma obra de não ficção.

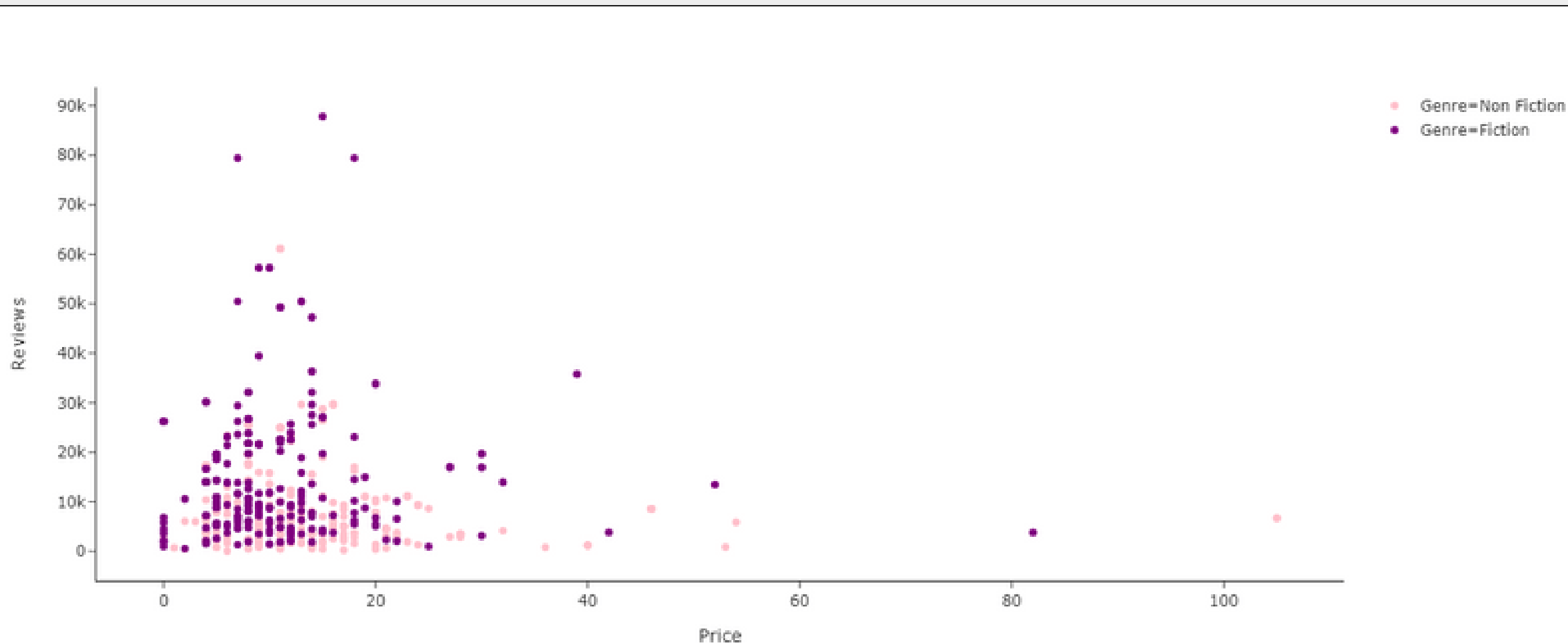
Correlação de dados

Q Elaboração do código para a realização da análise de correção

```
[ ] # Correlação entre o gênero dos livros (ficção e não ficção), avaliações (comentários) e preço
# Price: Preço
# Genre: Gênero
# Reviews: Avaliações (comentários)

fig = px.scatter(booklist, x = 'Price', y = 'Reviews', color = 'Genre', color_discrete_map = {'Fiction':'purple','Non Fiction':'pink'})

fig.update_layout(template = 'simple_white')
```



- 1 Os livros de ambos os gêneros com aproximadamente 20.000 comentários, custam no máximo \$: 25,00;
- 2 Os livros com maiores avaliações são os de ficção;
- 3 O livro com menores comentários diz respeito a um livro de não ficção, com aproximadamente 37 reviews.

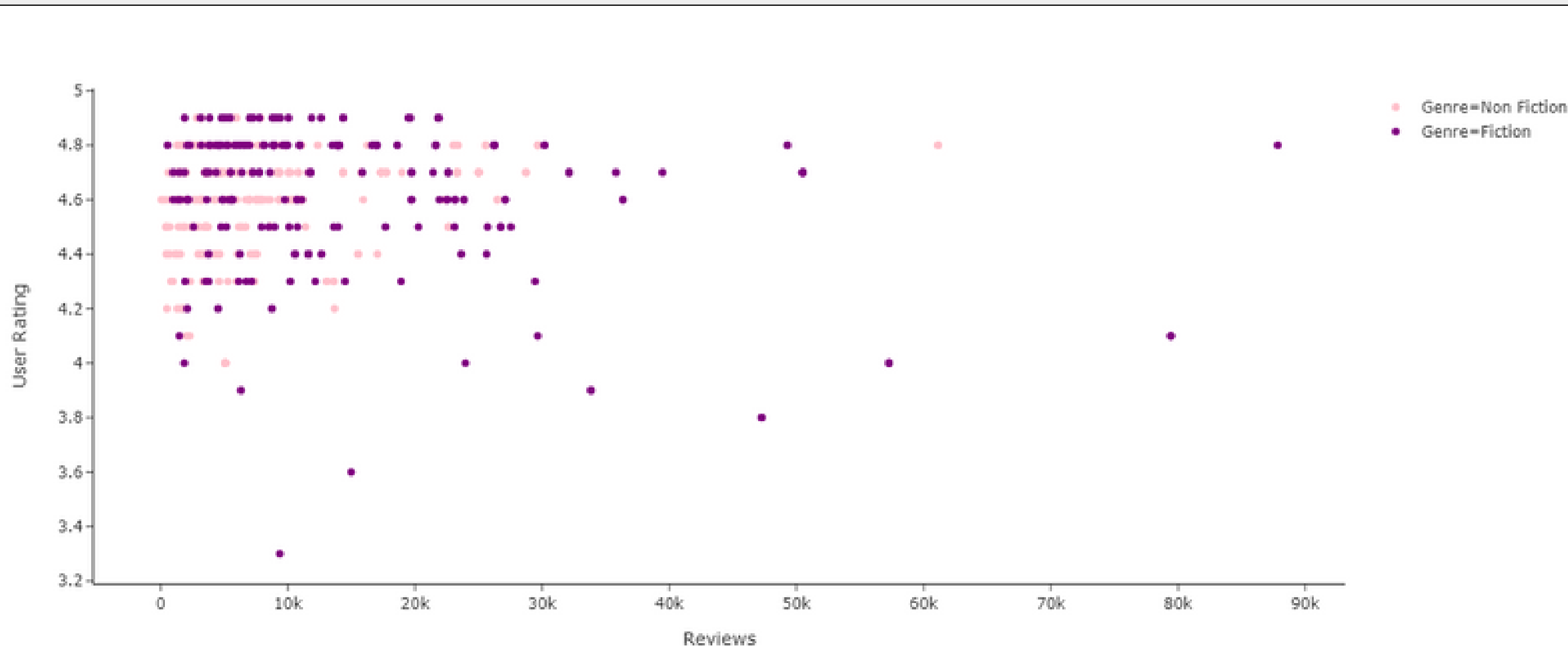
Correlação de dados

Q Elaboração do código para a realização da análise de correção

```
[ ] # Correlação entre o gênero dos livros (ficção e não ficção), avaliações (comentários) e preço
# User Rating: Média da Avaliação dos Leitores
# Genre: Gênero
# Reviews: Avaliações (comentários)

fig = px.scatter(booklist, x = 'Reviews', y = 'User Rating', color = 'Genre', color_discrete_map = {'Fiction':'purple','Non Fiction':'pink'})

fig.update_layout(template = 'simple_white')
```



- 1 Apenas 5 livros possuem mais de 50.000 reviews;
- 2 De 5 obras, apenas 1 possui mais de 50.000 comentários;
- 3 Uma considerável quantidade de livros com avaliação acima de 4.5 possui até 20.000 comentários ou estrelas;
- 4 O livro de menor avaliação diz respeito a uma ficção.

```
[16] # Corelação de todos os dados presentes no dataset

corr = booklist.corr().abs()
for i in corr.columns:
    corr.loc[i, i] = 0

sns.heatmap(corr)
```

Correlação de todos os dados presentes no *dataset*

Relação entre os itens presentes na lista e a significância.

<matplotlib.axes._subplots.AxesSubplot at 0x7f4f814e4510>

