Yuna Chung A01709043

ACTIVIDAD 07 - VISUALIZACIÓN

```
[ ] # Carga las librerías necesarias.

    import numpy as np
    import pandas as pd
    import matplotlib.pyplot as plt
    import seaborn as sns
```

```
[ ] # Carga el conjunto de datos al ambiente de Google Colab y muestra los primeros
    # 6 renglones.

    from google.colab import files

    uploaded = files.upload()

    for fn in uploaded.keys():
      print('User uploaded file "{name}" with length {length} bytes'
      .format(name=fn, length=len(uploaded[fn])))

    df = pd.read_csv('bestsellers with categories.csv')
    df.head(6)
```

Elegir archivos Sin archivos seleccionados Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.

```
Saving bestsellers with categories.csv to bestsellers with categories (1).csv
User uploaded file "bestsellers with categories.csv" with length 51161 bytes
```

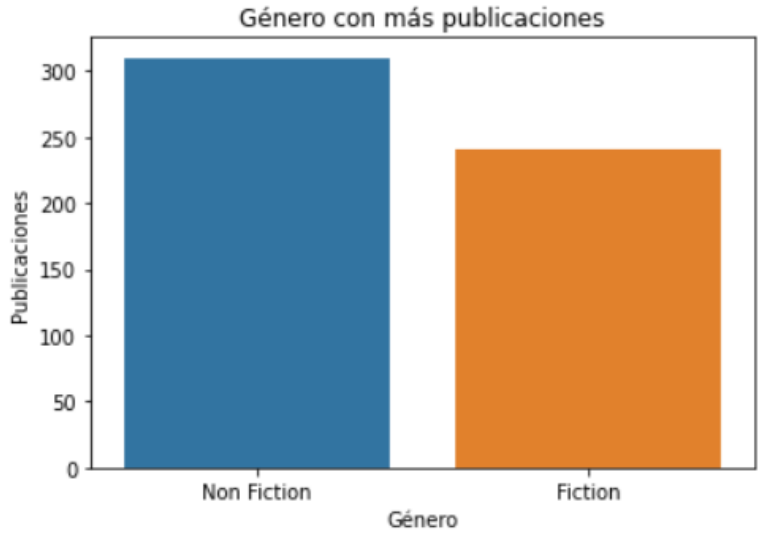
	Name	Author	User Rating	Reviews	Price	Year	Genre
0	10-Day Green Smoothie Cleanse	JJ Smith	4.7	17350	8	2016	Non Fiction
1	11/22/63: A Novel	Stephen King	4.6	2052	22	2011	Fiction

```
[ ] # Crea una tabla resumen con los estadísticas generales de las variables
    # numéricas.

    df.describe()
```

	User Rating	Reviews	Price	Year
count	550.000000	550.000000	550.000000	550.000000
mean	4.618364	11953.281818	13.100000	2014.000000
std	0.226980	11731.132017	10.842262	3.165156
min	3.300000	37.000000	0.000000	2009.000000
25%	4.500000	4058.000000	7.000000	2011.000000
50%	4.700000	8580.000000	11.000000	2014.000000
75%	4.800000	17253.250000	16.000000	2017.000000
max	4.900000	87841.000000	105.000000	2019.000000

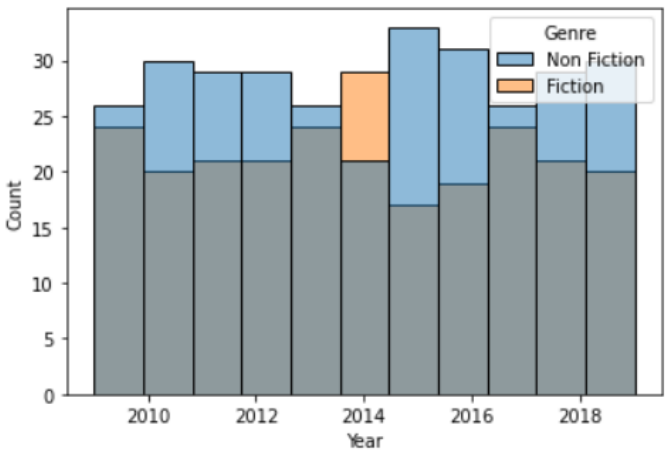
```
[ ]  ## ¿Cuál es el género con más publicaciones? Muéstralo en un gráfico.

     fig = plt.figure(figsize = (6, 4))
     sns.countplot(data = df, x = 'Genre')
     plt.title('Género con más publicaciones')
     plt.ylabel('Publicaciones')
     plt.xlabel('Género')
```

```
Text(0.5, 0, 'Género')
```



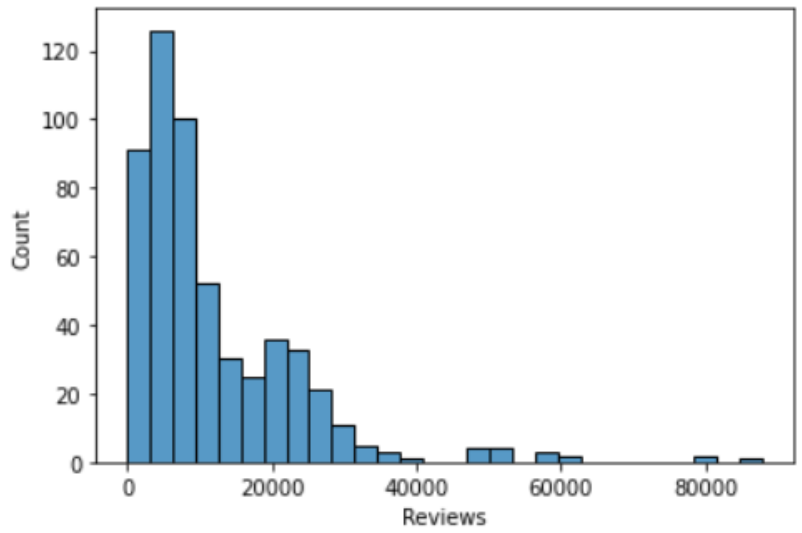

```
[ ]  # ¿Cuántos libros del top 50 se publicaron por género en cada año? ¿Hay algún
     # año donde hubo más libros de ficción en el top 50?. Muéstralo en un gráfico.

     fig = plt.figure(figsize = (6, 4))
     sns.histplot(data = df, x = 'Year', hue = 'Genre')
```

```
<Axes: xlabel='Year', ylabel='Count'>
```



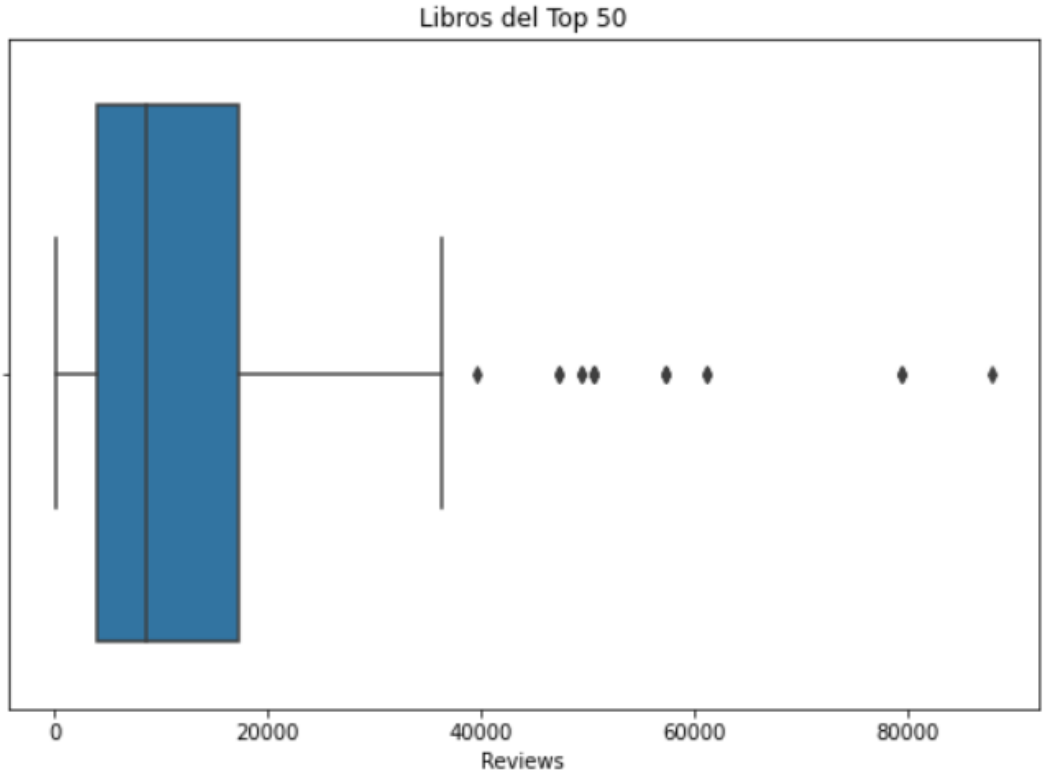
```
[ ]  # ¿Cómo se distribuye la variable Review? Muéstra el histografa.

     fig = plt.figure(figsize = (6, 4))
     sns.histplot(data = df, x = 'Reviews')
```

```
<Axes: xlabel='Reviews', ylabel='Count'>
```



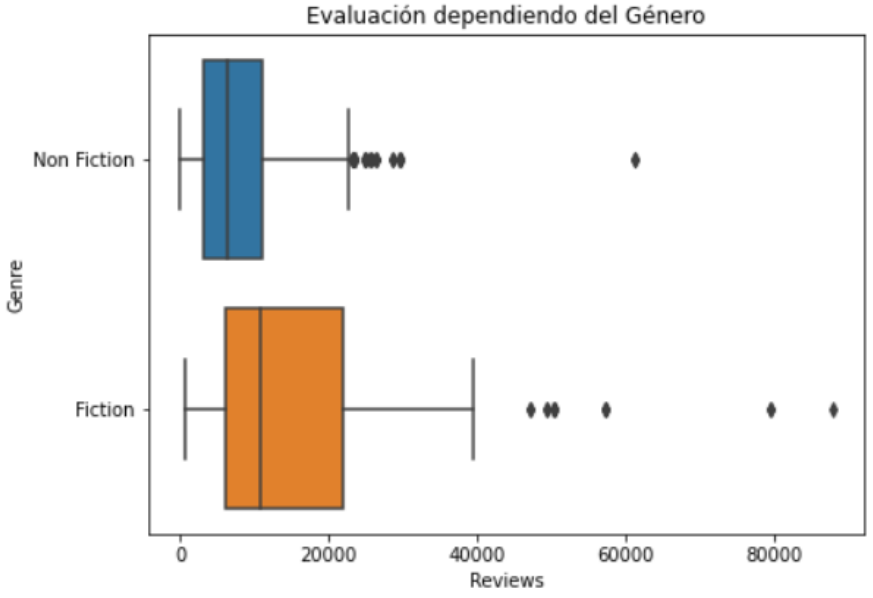
```
[ ]  # Ahora muéstralo en un gráfico de caja y bigote.

     fig = plt.figure(figsize = (9, 6))
     sns.boxplot(data = df, x = 'Reviews')
     plt.title('Libros del Top 50')
```

```
Text(0.5, 1.0, 'Libros del Top 50')
```



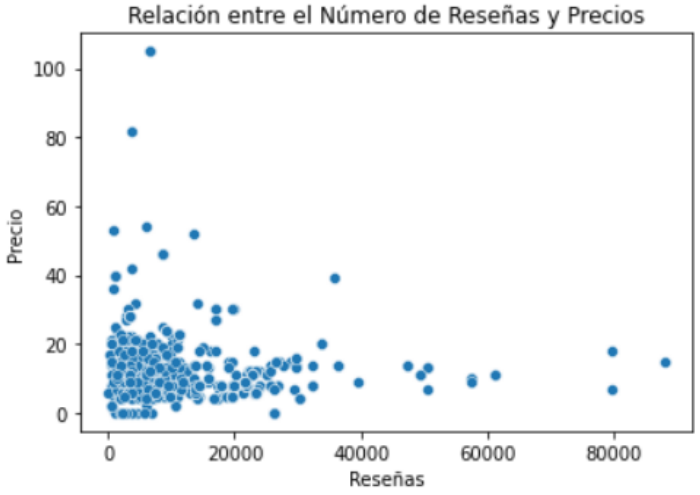
```
[ ]  # ¿Cómo se compara la evaluación del libro por género? ¿Qué genero es mejor
     # evaluado por los lectores? Muéstralo en un solo gráfico de caja y bigote.

     fig = plt.figure(figsize = (7, 5))
     sns.boxplot(data = df, x = 'Reviews', y = 'Genre')
     plt.title('Evaluación dependiendo del Género')
```

```
Text(0.5, 1.0, 'Evaluación dependiendo del Género')
```



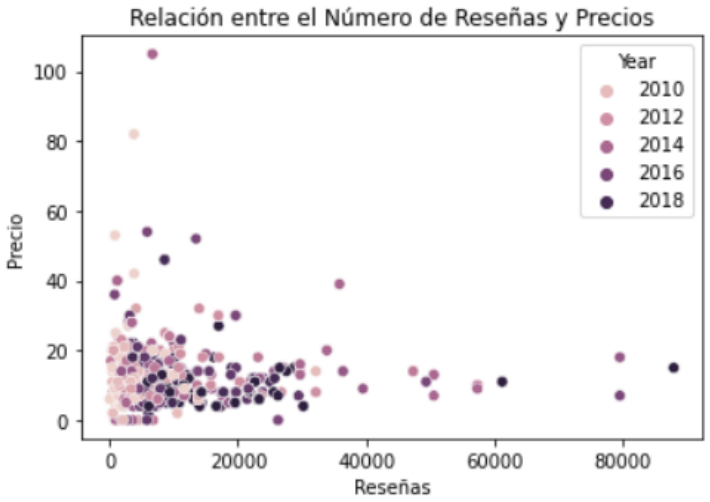
```
[ ]  # ¿Cuál es la relación entre el número de reseñas y precios? Muéstralo en un
     # gráfico de dispersión.

     fig = plt.figure(figsize = (6, 4))
     sns.scatterplot(data = df, x = 'Reviews', y = 'Price')
     plt.title('Relación entre el Número de Reseñas y Precios')
     plt.xlabel('Reseñas')
     plt.ylabel('Precio')
```

```
Text(0, 0.5, 'Precio')
```



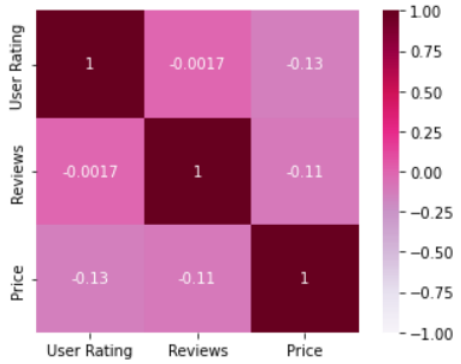
```
[ ] # De la pregunta anterior, ¿influye algo el año de publicación? ¿Cuál es la
    # relación entre el número de reseñar, el precio y el año de publicación?
    # IMPORTANTE: Selecciona una paleta de colores adeacuada.

    fig = plt.figure(figsize = (6, 4))
    sns.scatterplot(data = df, x = 'Reviews', y = 'Price', hue = 'Year')
    plt.title('Relación entre el Número de Reseñas y Precios')
    plt.xlabel('Reseñas')
    plt.ylabel('Precio')
```

```
Text(0, 0.5, 'Precio')
```



```
[ ]  # ¿Cuál es la correlación entre las variables numéricas? Muéstralo en un
     # gráfico. La variable año, a pesar de ser numérica, la vamos a considerar como
     # cualitativa, así que la eliminaremos del análisis.

     df2 = df[['Name', 'Author', 'User Rating', 'Reviews', 'Price', 'Genre']]
     correlacion = df2.corr()
     sns.heatmap(data = correlacion, vmin = -1, vmax = 1, cmap = 'PuRd', annot = True, square = True)
```

```
<Axes: >
```



,

¿Cuáles variables tiene una fuerte relación positiva entre sí y cuáles tienen una fuerte relación negativa? (Esta pregunta no es de código) Responde la pregunta en la siguiente celda de texto.

- Como se puede visualizar desde la gráfica anterior, no hay una correlación fuerte entre las variables, porque todos los valores son negativos y lejos del valor 1.

```
[ ]  # Haz una gráfica donde podemos comparar la relación entre las tres variables
     # numéricas (User Rating, Reviews y Price) y que, además, podamos ver el efecto
     # del libro. La variable año, a pesar de ser numérica, la vamos a considerar como
     # cualitativa, así que la eliminaremos del análisis.

     df3 = df[['User Rating', 'Reviews', 'Price', 'Genre']]
     sns.pairplot(data = df3, hue = 'Genre')
```

```
<seaborn.axisgrid.PairGrid at 0x7f6b417cda30>
```