

Nombre: Hilda Beltrán

Matrícula: A01251916

```
from google.colab import drive
drive.mount('/content/drive')
```

Mounted at /content/drive

```
# Carga las librerías necesarias.
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

```
# Carga el archivo bestsellers with categories.csv
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])))
```

 bestsellers ...tegories.csv

- **bestsellers with categories.csv**(text/csv) - 51161 bytes, last modified: 5/11/2022 - 100% done
Saving bestsellers with categories.csv to bestsellers with categories.csv
User uploaded file "bestsellers with categories.csv" with length 51161 bytes

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

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

| | 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 |
| 2 | 12 Rules for Life: An Antidote to Chaos | Jordan B. Peterson | 4.7 | 18979 | 15 | 2018 | Non Fiction |
| 3 | 1984 (Signet Classics) | George Orwell | 4.7 | 21424 | 6 | 2017 | Fiction |
| 4 | 5,000 Awesome Facts (About | National | 4.6 | 7005 | 10 | 2010 | Non |

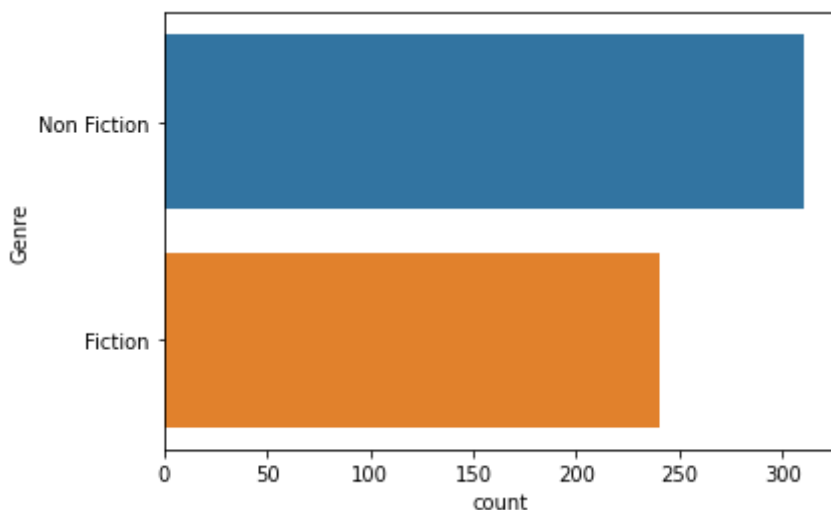
```
# Crea una tabla resumen con los estadísticas generales de las variables
# numéricas.
df.describe()
```

| | 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, y = 'Genre')
# El género con más publicaciones es Non Fiction
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f7309add990>



```
# ¿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.
df1 = pd.crosstab(df['Year'], df['Genre'])
df1
```

Genre Fiction Non Fiction 

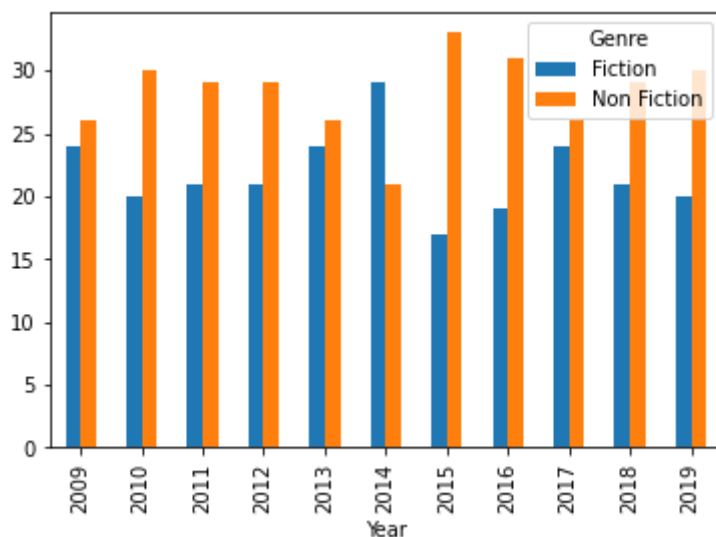
Year

| | | |
|------|----|----|
| 2009 | 24 | 26 |
| 2010 | 20 | 30 |
| 2011 | 21 | 29 |
| 2012 | 21 | 29 |
| 2013 | 24 | 26 |
| 2014 | 29 | 21 |
| 2015 | 17 | 33 |
| 2016 | 19 | 31 |
| 2017 | 24 | 26 |
| 2018 | 21 | 29 |

```
df1.plot(kind='bar')
```

```
# En el 2014 hubo más libros de ficción en el top 50
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f72f71ba390>
```



```
# ¿Cómo se distribuye la variable Review? Muéstrala el histograma.
```

```
sns.histplot(data=df, x='Reviews')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f72f74c43d0>
```

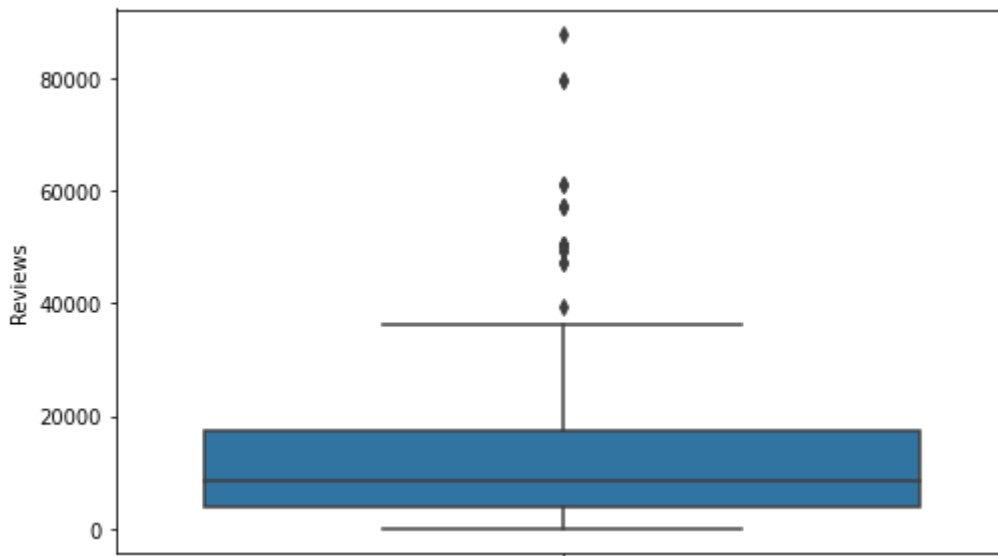


```
# Ahora muéstralo en un gráfico de caja y bigote.
```

```
fig = plt.figure(figsize=(8,5))
```

```
sns.boxplot(data=df, y = 'Reviews')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f72f747f250>
```



```
# ¿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.
```

```
fig1 = plt.figure(figsize=(8,5))
```

```
sns.boxplot(data=df, x = 'Genre', y = 'User Rating')
```

```
# El mejor evaluado es el de Fiction, tiene algunas evaluaciones dispersas que
```

```
# son los puntos que están por debajo
```

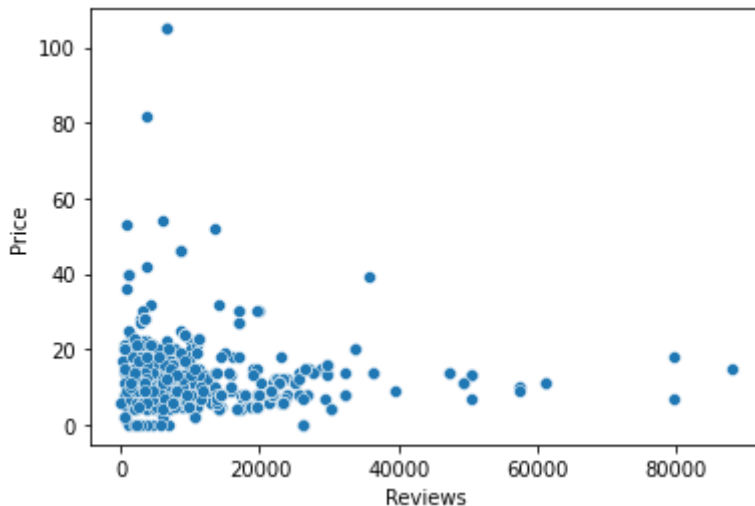
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f72f715f7d0>
```



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

```
fig2 = plt.figure(figsize=(6, 4))
sns.scatterplot(data=df, x='Reviews', y='Price')
```

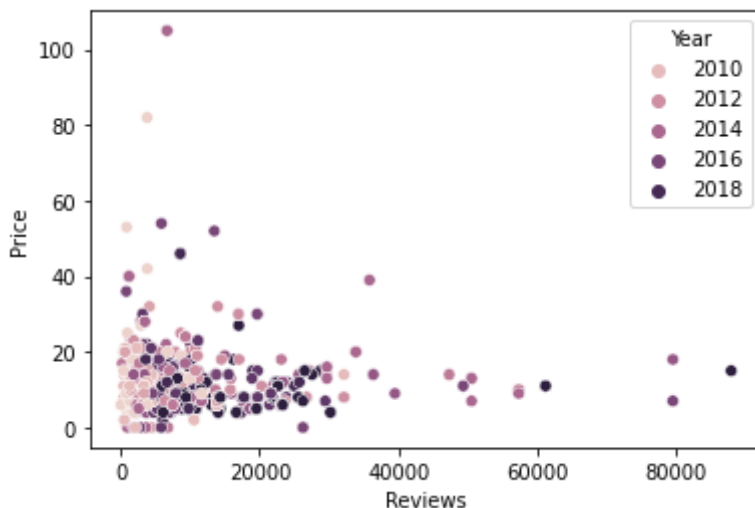
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f72f7478f50>
```



```
# 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 adecuada.
```

```
fig3 = plt.figure(figsize=(6, 4))
sns.scatterplot(data=df, x='Reviews', y='Price', hue='Year')
```

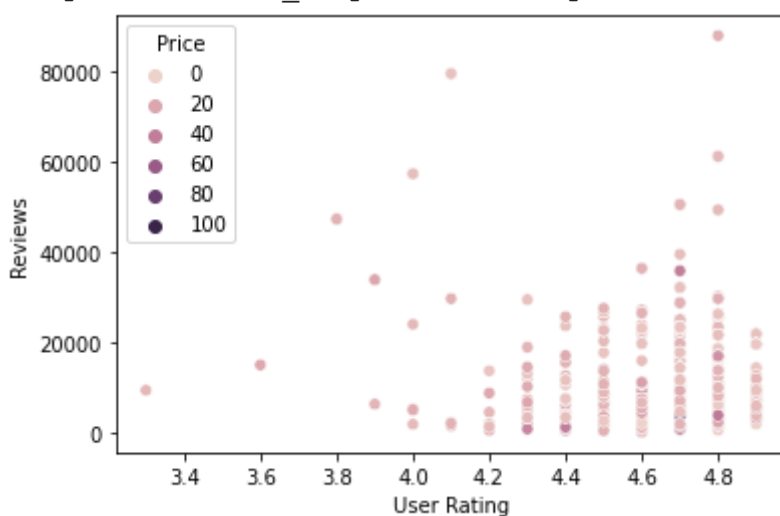
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f72f7402550>
```



```
# ¿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.
```

```
fig4 = plt.figure(figsize=(6, 4))  
sns.scatterplot(data=df, x='User Rating', y='Reviews', hue='Price')
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f72f720c910>



✓ 0s completed at 10:22 PM

