Amazon Bestselling Books

Dataset on Amazon's bestselling books from 2009 to 2019. Contains 550 books, data has been categorized into fiction and non-fiction using Goodreads

In [1]:

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

Loading the dataset :

In [2]:

```
path = (r"C:\Users\MY PC\Downloads\amazon bestsellers.csv")
```

In [12]:

```
df = pd.read_csv(path)
print(df)
```

```
                                                  Name             Author  \
0     Act Like a Lady, Think Like a Man: What Men Re...       Steve Harvey   
1     Arguing with Idiots: How to Stop Small Minds a...        Glenn Beck   
2             Breaking Dawn (The Twilight Saga, Book 4)   Stephenie Meyer   
3           Crazy Love: Overwhelmed by a Relentless God      Francis Chan   
4     Dead And Gone: A Sookie Stackhouse Novel (Sook...  Charlaine Harris   
..                                                  ...               ...   
545   Unicorn Coloring Book: For Kids Ages 4-8 (US E...        Silly Bear   
546   What Should Danny Do? (The Power to Choose Ser...         Adir Levy   
547                          Where the Crawdads Sing       Delia Owens   
548       Wrecking Ball (Diary of a Wimpy Kid Book 14)       Jeff Kinney   
549   You Are a Badass: How to Stop Doubting Your Gr...       Jen Sincero   

     User Rating  Reviews Price  Year        Genre  
0            4.6     5013   $17  2009  Non Fiction  
1            4.6      798    $5  2009  Non Fiction  
2            4.6     9769   $13  2009      Fiction  
3            4.7     1542   $14  2009  Non Fiction  
4            4.6     1541    $4  2009      Fiction  
..           ...      ...   ...   ...          ...  
545          4.8     6108    $4  2019  Non Fiction  
546          4.8     8170   $13  2019      Fiction  
547          4.8    87841   $15  2019      Fiction  
548          4.9     9413    $8  2019      Fiction  
549          4.7    14331    $8  2019  Non Fiction  

[550 rows x 7 columns]
```

In [13]:

```
head = df.head(200)
head
```

Out[13]:

	Name	Author	User Rating	Reviews	Price	Year	Genre
0	Act Like a Lady, Think Like a Man: What Men Re...	Steve Harvey	4.6	5013	$17	2009	Non Fiction
1	Arguing with Idiots: How to Stop Small Minds a...	Glenn Beck	4.6	798	$5	2009	Non Fiction
2	Breaking Dawn (The Twilight Saga, Book 4)	Stephenie Meyer	4.6	9769	$13	2009	Fiction
3	Crazy Love: Overwhelmed by a Relentless God	Francis Chan	4.7	1542	$14	2009	Non Fiction
4	Dead And Gone: A Sookie Stackhouse Novel (Sook...	Charlaine Harris	4.6	1541	$4	2009	Fiction
...	...	...	...	...	...	...	...
195	Thomas Jefferson: The Art of Power	Jon Meacham	4.5	1904	$23	2012	Non Fiction
196	Unbroken: A World War II Story of Survival, Re...	Laura Hillenbrand	4.8	29673	$16	2012	Non Fiction
197	Wheat Belly: Lose the Wheat, Lose the Weight, ...	William Davis	4.4	7497	$6	2012	Non Fiction
198	Wild: From Lost to Found on the Pacific Crest ...	Cheryl Strayed	4.4	17044	$18	2012	Non Fiction
199	Winter of the World: Book Two of the Century T...	Ken Follett	4.5	10760	$15	2012	Fiction

200 rows × 7 columns

In [14]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 550 entries, 0 to 549
Data columns (total 7 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   Name         550 non-null    object 
 1   Author       550 non-null    object 
 2   User Rating  550 non-null    float64
 3   Reviews      550 non-null    int64  
 4   Price        550 non-null    object 
 5   Year         550 non-null    int64  
 6   Genre        550 non-null    object 
dtypes: float64(1), int64(2), object(4)
memory usage: 30.2+ KB
```

In [15]:

```
rows , columns = df.shape
```

In [16]:

```
rows
```

Out[16]:

```
550
```

In [17]:

```
columns
```

Out[17]:

```
7
```

In [18]:

```
df.isnull().sum()
```

Out[18]:

```
Name           0
Author         0
User Rating    0
Reviews        0
Price          0
Year           0
Genre          0
dtype: int64
```

There are no null values in the dataset .

In [19]:

```
df.describe()
```

Out[19]:

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

Here we can se that : Maximum User Rating is "4.9", Minimum User rating is "3.3", Average User Rating is "4.6".

In [20]:

```
genre = df.groupby('Genre')
genre
```

Out[20]:

```
<pandas.core.groupby.generic.DataFrameGroupBy object at 0x000001F0E9C30E20>
```

In [26]:

```
genre.get_group('Fiction')
```

Out[26]:

	Name	Author	User Rating	Reviews	Price	Year	Genre
2	Breaking Dawn (The Twilight Saga, Book 4)	Stephenie Meyer	4.6	9769	$13	2009	Fiction
4	Dead And Gone: A Sookie Stackhouse Novel (Sook...	Charlaine Harris	4.6	1541	$4	2009	Fiction
5	Diary of a Wimpy Kid: The Last Straw (Book 3)	Jeff Kinney	4.8	3837	$15	2009	Fiction
7	Dog Days (Diary of a Wimpy Kid, Book 4) (Volum...	Jeff Kinney	4.8	3181	$12	2009	Fiction
10	Eclipse (Twilight Sagas)	Stephenie Meyer	4.7	5505	$7	2009	Fiction
...	...	...	...	...	...	...	...
542	The Wonky Donkey	Craig Smith	4.8	30183	$4	2019	Fiction
543	To Kill a Mockingbird	Harper Lee	4.8	26234	$7	2019	Fiction
546	What Should Danny Do? (The Power to Choose Ser...	Adir Levy	4.8	8170	$13	2019	Fiction
547	Where the Crawdads Sing	Delia Owens	4.8	87841	$15	2019	Fiction
548	Wrecking Ball (Diary of a Wimpy Kid Book 14)	Jeff Kinney	4.9	9413	$8	2019	Fiction

240 rows × 7 columns

Information about fiction genre.

In [27]:

```
genre.get_group('Non Fiction')
```

Out[27]:

	Name	Author	User Rating	Reviews	Price	Year	Genre
0	Act Like a Lady, Think Like a Man: What Men Re...	Steve Harvey	4.6	5013	$17	2009	Non Fiction
1	Arguing with Idiots: How to Stop Small Minds a...	Glenn Beck	4.6	798	$5	2009	Non Fiction
3	Crazy Love: Overwhelmed by a Relentless God	Francis Chan	4.7	1542	$14	2009	Non Fiction
6	Divine Soul Mind Body Healing and Transmission...	Zhi Gang Sha	4.6	37	$6	2009	Non Fiction
8	Eat This Not That! Supermarket Survival Guide:...	David Zinczenko	4.5	720	$1	2009	Non Fiction
...	...	...	...	...	...	...	...
538	The Total Money Makeover: Classic Edition: A P...	Dave Ramsey	4.7	11550	$10	2019	Non Fiction
539	The Unofficial Harry Potter Cookbook: From Cau...	Dinah Bucholz	4.7	9030	$10	2019	Non Fiction
544	Unfreedom of the Press	Mark R. Levin	4.9	5956	$11	2019	Non Fiction
545	Unicorn Coloring Book: For Kids Ages 4-8 (US E...	Silly Bear	4.8	6108	$4	2019	Non Fiction
549	You Are a Badass: How to Stop Doubting Your Gr...	Jen Sincero	4.7	14331	$8	2019	Non Fiction

310 rows × 7 columns

Information about non-fition genre

In [45]:

```
x = df.groupby(["Name"])["Reviews"].max().reset_index()
x.head(10)
```

Out[45]:

	Name	Reviews
0	10-Day Green Smoothie Cleanse	17350
1	11/22/63: A Novel	2052
2	12 Rules for Life: An Antidote to Chaos	18979
3	1984 (Signet Classics)	21424
4	5,000 Awesome Facts (About Everything!) (Natio...	7665
5	A Dance with Dragons (A Song of Ice and Fire)	12643
6	A Game of Thrones / A Clash of Kings / A Storm...	19735
7	A Gentleman in Moscow: A Novel	19699
8	A Higher Loyalty: Truth, Lies, and Leadership	5983
9	A Man Called Ove: A Novel	23848

MOST REVIEWED NAME OF THE BOOKS .

In [44]:

```
y = df.groupby(["Year"])["Reviews"].max().reset_index()
y.head(20)
```

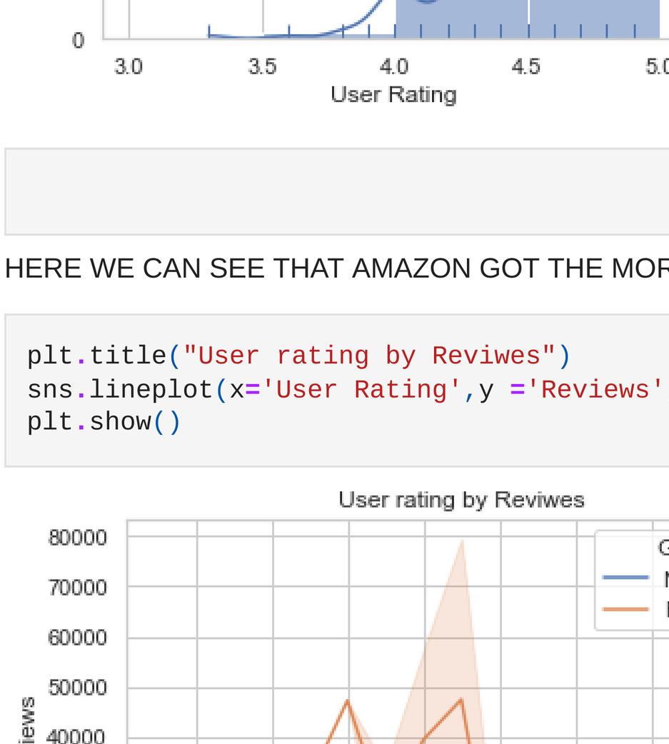
Out[44]:

	Year	Reviews
0	2009	19720
1	2010	32122
2	2011	32122
3	2012	57271
4	2013	57271
5	2014	57271
6	2015	79446
7	2016	79446
8	2017	29442
9	2018	61133
10	2019	87841

Reviews in all the years .

In [49]:

```
sns.displot(df["User Rating"],bins = [3.0 ,3.5,4.0,4.5,5.0],kde = True , rug = True)
plt.title("Count of User Rating")
plt.show()
```



In [ ]:

HERE WE CAN SEE THAT AMAZON GOT THE MORE NUMBER OF USER RATING ABOVE 4.5

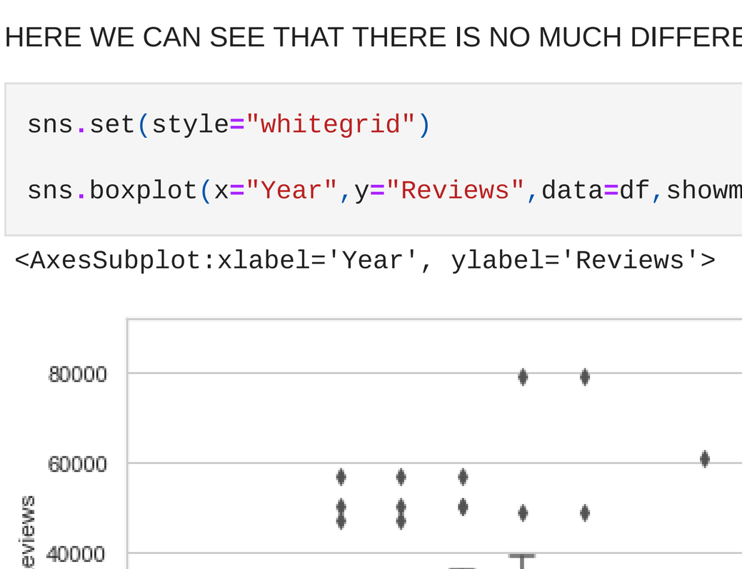
In [47]:

```
plt.title("User rating by Reviwes")
sns.lineplot(x='User Rating',y ='Reviews',data = df,hue='Genre',markers = ["o",">"])
plt.show()
```

In [32]:

```
sns.set(style = "whitegrid")

sns.barplot(x="Genre",y="User Rating",data = df,palette = "autumn_r")
plt.show()
```



HERE WE CAN SEE THAT THERE IS NO MUCH DIFFERENCE IN THE RATING OF FICTION AND NON FICTION .

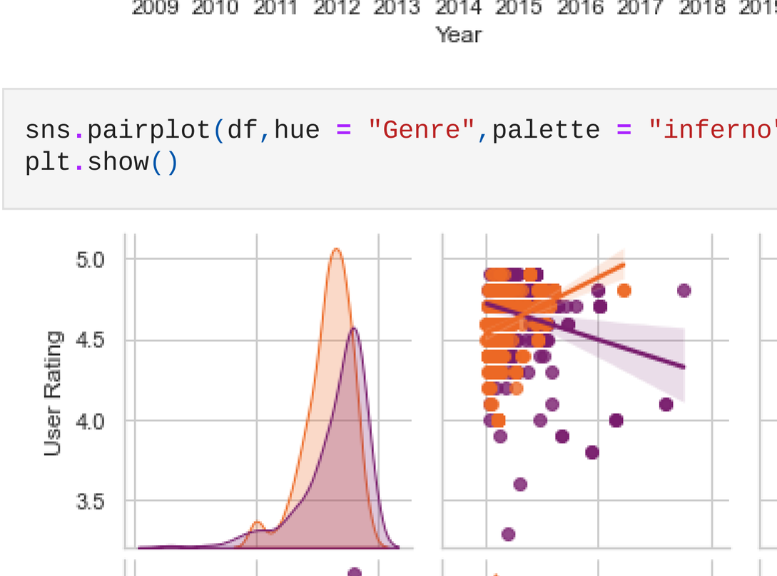
In [36]:

```
sns.set(style="whitegrid")

sns.boxplot(x="Year",y="Reviews",data=df,showmeans=True,palette ="spring_r")
```

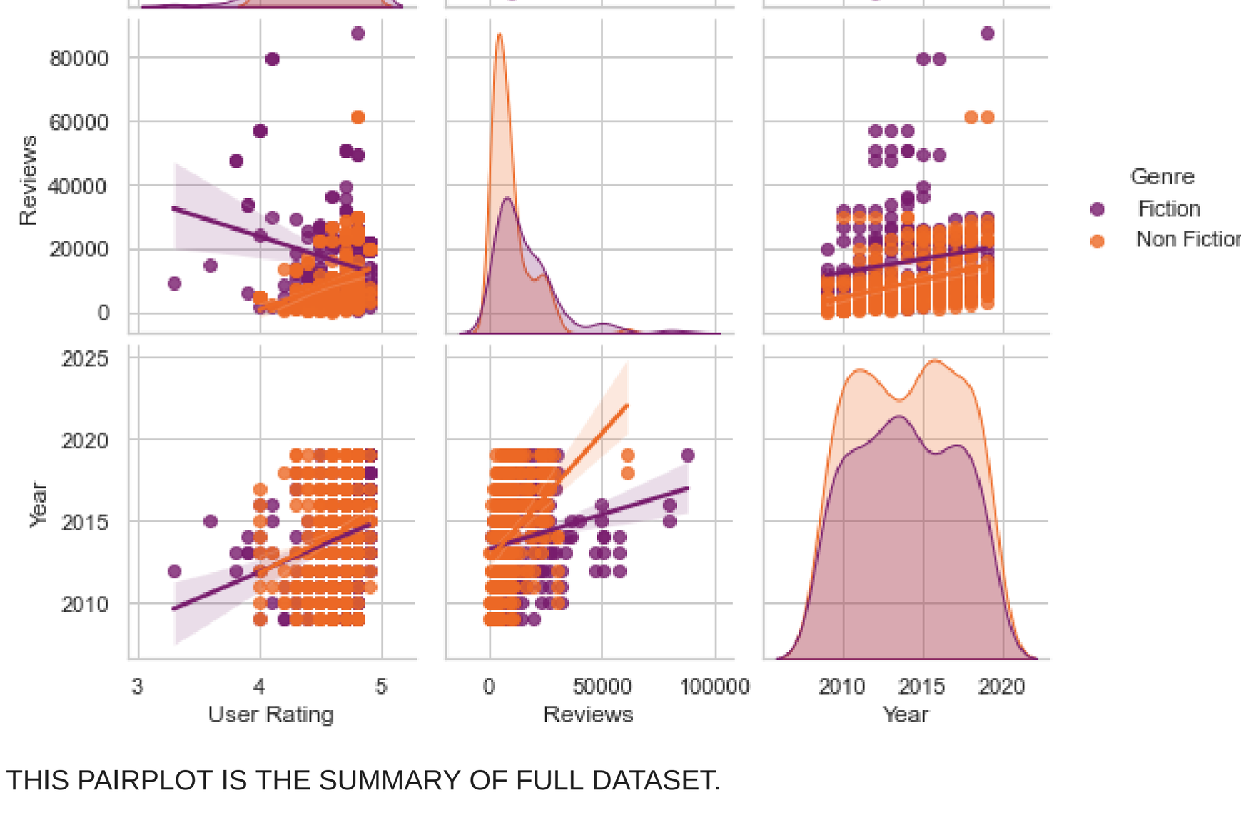
Out[36]:

```
<AxesSubplot:xlabel='Year', ylabel='Reviews'>
```



In [48]:

```
sns.pairplot(df,hue = "Genre",palette = "inferno",hue_order = ["Fiction","Non Fiction"],kind = "reg")
plt.show()
```



THIS PAIRPLOT IS THE SUMMARY OF FULL DATASET.

In [ ]: