

```
In [1]: import pandas as pd
import numpy as np
import seaborn as sns
```

```
In [2]: data=pd.read_csv("googleplaystore.csv")
```

```
In [3]: data.head()
```

```
Out[3]:
```

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0	Everyone	Art & Design	January 7, 2018	1.0.0	4.0.3 and up
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone	Art & Design;Pretend Play	January 15, 2018	2.0.0	4.0.3 and up
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8.7M	5,000,000+	Free	0	Everyone	Art & Design	August 1, 2018	1.2.4	4.0.3 and up
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50,000,000+	Free	0	Teen	Art & Design	June 8, 2018	Varies with device	4.2 and up
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100,000+	Free	0	Everyone	Art & Design;Creativity	June 20, 2018	1.1	4.4 and up

```
In [4]: data.isnull().sum()
```

```
Out[4]: App                0
Category                0
Rating                1474
Reviews                0
Size                  0
Installs              0
Type                  1
```

```
Price          0
Content Rating 1
Genres         0
Last Updated   0
Current Ver    8
Android Ver    3
dtype: int64
```

In [5]:

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10841 entries, 0 to 10840
Data columns (total 13 columns):
#   Column          Non-Null Count  Dtype
---  -
0   App              10841 non-null  object
1   Category         10841 non-null  object
2   Rating           9367 non-null   float64
3   Reviews          10841 non-null  object
4   Size             10841 non-null  object
5   Installs         10841 non-null  object
6   Type             10840 non-null  object
7   Price            10841 non-null  object
8   Content Rating   10840 non-null  object
9   Genres           10841 non-null  object
10  Last Updated     10841 non-null  object
11  Current Ver      10833 non-null  object
12  Android Ver      10838 non-null  object
dtypes: float64(1), object(12)
memory usage: 1.1+ MB
```

In [6]:

```
data.shape
```

Out[6]: (10841, 13)

In [7]:

```
data.isnull().any()
```

Out[7]:

App	False
Category	False
Rating	True
Reviews	False

Size	False
Installs	False
Type	True
Price	False
Content Rating	True
Genres	False
Last Updated	False
Current Ver	True
Android Ver	True

dtype: bool

```
In [8]: data.isnull().sum()
```

```
Out[8]: App                0
Category                0
Rating                1474
Reviews                0
Size                  0
Installs              0
Type                  1
Price                 0
Content Rating        1
Genres                0
Last Updated          0
Current Ver           8
Android Ver           3
dtype: int64
```

```
In [9]: data=data.dropna()
```

```
In [10]: data.shape
```

```
Out[10]: (9360, 13)
```

```
In [11]: data["Size"] = [ float(i.split('M')[0]) if 'M' in i else float(0) for i in data["Size"]]
```

```
In [12]: data.head()
```

Out[12]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19.0	10,000+	Free	0	Everyone	Art & Design	January 7, 2018	1.0.0	4.0.3 and up
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14.0	500,000+	Free	0	Everyone	Art & Design;Pretend Play	January 15, 2018	2.0.0	4.0.3 and up
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8.7	5,000,000+	Free	0	Everyone	Art & Design	August 1, 2018	1.2.4	4.0.3 and up
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25.0	50,000,000+	Free	0	Teen	Art & Design	June 8, 2018	Varies with device	4.2 and up
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8	100,000+	Free	0	Everyone	Art & Design;Creativity	June 20, 2018	1.1	4.4 and up

In [13]:

```
data["Size"]=1000* data["Size"]
```

In [14]:

```
data
```

Out[14]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19000.0	10,000+	Free	0	Everyone	Art & Design	January 7, 2018	1.0.0	4.0.3 and up
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14000.0	500,000+	Free	0	Everyone	Art & Design;Pretend Play	January 15, 2018	2.0.0	4.0.3 and up

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8700.0	5,000,000+	Free	0	Everyone	Art & Design	August 1, 2018	1.2.4	4.0.3 and up
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25000.0	50,000,000+	Free	0	Teen	Art & Design	June 8, 2018	Varies with device	4.2 and up
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2800.0	100,000+	Free	0	Everyone	Art & Design;Creativity	June 20, 2018	1.1	4.4 and up
...
10834	FR Calculator	FAMILY	4.0	7	2600.0	500+	Free	0	Everyone	Education	June 18, 2017	1.0.0	4.1 and up
10836	Sya9a Maroc - FR	FAMILY	4.5	38	53000.0	5,000+	Free	0	Everyone	Education	July 25, 2017	1.48	4.1 and up
10837	Fr. Mike Schmitz Audio Teachings	FAMILY	5.0	4	3600.0	100+	Free	0	Everyone	Education	July 6, 2018	1.0	4.1 and up
10839	The SCP Foundation DB fr nn5n	BOOKS_AND_REFERENCE	4.5	114	0.0	1,000+	Free	0	Mature 17+	Books & Reference	January 19, 2015	Varies with device	Varies with device
10840	iHoroscope - 2018 Daily Horoscope & Astrology	LIFESTYLE	4.5	398307	19000.0	10,000,000+	Free	0	Everyone	Lifestyle	July 25, 2018	Varies with device	Varies with device

9360 rows × 13 columns

In [15]: `data.info()`

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 9360 entries, 0 to 10840
Data columns (total 13 columns):
#   Column                Non-Null Count  Dtype
---  -
0   App                    9360 non-null   object
1   Category               9360 non-null   object
2   Rating                 9360 non-null   float64
3   Reviews                9360 non-null   object
4   Size                   9360 non-null   float64
5   Installs               9360 non-null   object
6   Type                   9360 non-null   object
7   Price                  9360 non-null   object
8   Content Rating         9360 non-null   object
9   Genres                 9360 non-null   object
10  Last Updated           9360 non-null   object
11  Current Ver            9360 non-null   object
12  Android Ver            9360 non-null   object
dtypes: float64(2), object(11)
memory usage: 1023.8+ KB
```

In [16]: `data["Reviews"]=data["Reviews"].astype(float)`

In [17]: `data.info()`

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 9360 entries, 0 to 10840
Data columns (total 13 columns):
#   Column                Non-Null Count  Dtype
---  -
0   App                    9360 non-null   object
1   Category               9360 non-null   object
2   Rating                 9360 non-null   float64
3   Reviews                9360 non-null   float64
4   Size                   9360 non-null   float64
5   Installs               9360 non-null   object
6   Type                   9360 non-null   object
7   Price                  9360 non-null   object
8   Content Rating         9360 non-null   object
```

```

9   Genres          9360 non-null  object
10  Last Updated    9360 non-null  object
11  Current Ver     9360 non-null  object
12  Android Ver     9360 non-null  object
dtypes: float64(3), object(10)
memory usage: 1023.8+ KB

```

```
In [18]: data["Installs"] = [float(i.replace('+','').replace(',',' ')) if '+' in i or ',' in i else float(0) for i in data["Installs"]]
```

```
In [19]: data.head()
```

```
Out[19]:
```

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159.0	19000.0	10000.0	Free	0	Everyone	Art & Design	January 7, 2018	1.0.0	4.0.3 and up
1	Coloring book moana	ART_AND_DESIGN	3.9	967.0	14000.0	500000.0	Free	0	Everyone	Art & Design;Pretend Play	January 15, 2018	2.0.0	4.0.3 and up
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510.0	8700.0	5000000.0	Free	0	Everyone	Art & Design	August 1, 2018	1.2.4	4.0.3 and up
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644.0	25000.0	50000000.0	Free	0	Teen	Art & Design	June 8, 2018	Varies with device	4.2 and up
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967.0	2800.0	100000.0	Free	0	Everyone	Art & Design;Creativity	June 20, 2018	1.1	4.4 and up

```
In [20]: data.info()
```

```

<class 'pandas.core.frame.DataFrame'>
Int64Index: 9360 entries, 0 to 10840
Data columns (total 13 columns):
#   Column          Non-Null Count  Dtype
---  -
0   App              9360 non-null   object

```

```

1  Category      9360 non-null  object
2  Rating        9360 non-null  float64
3  Reviews       9360 non-null  float64
4  Size          9360 non-null  float64
5  Installs      9360 non-null  float64
6  Type          9360 non-null  object
7  Price         9360 non-null  object
8  Content Rating 9360 non-null  object
9  Genres        9360 non-null  object
10 Last Updated  9360 non-null  object
11 Current Ver   9360 non-null  object
12 Android Ver   9360 non-null  object
dtypes: float64(4), object(9)
memory usage: 1023.8+ KB

```

```
In [21]: data["Installs"]=data["Installs"].astype(int)
```

```
In [22]: data.info()
```

```

<class 'pandas.core.frame.DataFrame'>
Int64Index: 9360 entries, 0 to 10840
Data columns (total 13 columns):
#   Column          Non-Null Count  Dtype
---  -
0   App             9360 non-null  object
1   Category        9360 non-null  object
2   Rating          9360 non-null  float64
3   Reviews         9360 non-null  float64
4   Size            9360 non-null  float64
5   Installs        9360 non-null  int32
6   Type            9360 non-null  object
7   Price           9360 non-null  object
8   Content Rating  9360 non-null  object
9   Genres          9360 non-null  object
10  Last Updated    9360 non-null  object
11  Current Ver     9360 non-null  object
12  Android Ver     9360 non-null  object
dtypes: float64(3), int32(1), object(9)
memory usage: 987.2+ KB

```

```
In [23]: data['Price'] = [ float(i.split('$')[1]) if '$' in i else float(0) for i in data['Price']]
```


In [24]:

```
data.head()
```

Out[24]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159.0	19000.0	10000	Free	0.0	Everyone	Art & Design	January 7, 2018	1.0.0	4.0.3 and up
1	Coloring book moana	ART_AND_DESIGN	3.9	967.0	14000.0	500000	Free	0.0	Everyone	Art & Design;Pretend Play	January 15, 2018	2.0.0	4.0.3 and up
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510.0	8700.0	5000000	Free	0.0	Everyone	Art & Design	August 1, 2018	1.2.4	4.0.3 and up
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644.0	25000.0	50000000	Free	0.0	Teen	Art & Design	June 8, 2018	Varies with device	4.2 and up
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967.0	2800.0	100000	Free	0.0	Everyone	Art & Design;Creativity	June 20, 2018	1.1	4.4 and up

In [25]:

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 9360 entries, 0 to 10840
Data columns (total 13 columns):
#   Column          Non-Null Count  Dtype
---  -
0   App              9360 non-null   object
1   Category         9360 non-null   object
2   Rating           9360 non-null   float64
3   Reviews          9360 non-null   float64
4   Size             9360 non-null   float64
5   Installs         9360 non-null   int32
6   Type             9360 non-null   object
7   Price            9360 non-null   float64
8   Content Rating   9360 non-null   object
9   Genres           9360 non-null   object
10  Last Updated     9360 non-null   object
```

```
11 Current Ver      9360 non-null  object
12 Android Ver      9360 non-null  object
dtypes: float64(4), int32(1), object(8)
memory usage: 987.2+ KB
```

```
In [26]: data["Price"]=data["Price"].astype(int)
```

```
In [27]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 9360 entries, 0 to 10840
Data columns (total 13 columns):
#   Column          Non-Null Count  Dtype
---  -
0   App              9360 non-null   object
1   Category         9360 non-null   object
2   Rating           9360 non-null   float64
3   Reviews          9360 non-null   float64
4   Size             9360 non-null   float64
5   Installs         9360 non-null   int32
6   Type             9360 non-null   object
7   Price            9360 non-null   int32
8   Content Rating   9360 non-null   object
9   Genres           9360 non-null   object
10  Last Updated     9360 non-null   object
11  Current Ver      9360 non-null   object
12  Android Ver      9360 non-null   object
dtypes: float64(3), int32(2), object(8)
memory usage: 950.6+ KB
```

```
In [28]: data.shape
```

```
Out[28]: (9360, 13)
```

```
In [29]: data.drop(data[(data["Reviews"]<1) & (data["Reviews"]<5)].index,inplace=True)
```

```
In [30]: data.shape
```

Out[30]: (9360, 13)

```
In [31]: data.drop(data[data['Installs'] < data['Reviews'] ].index, inplace = True)
```

```
In [32]: data.shape
```

Out[32]: (9353, 13)

```
In [33]: data.drop(data[(data["Type"]=="Free") &(data["Price"]>0)].index,inplace=True)
```

```
In [34]: data.shape
```

Out[34]: (9353, 13)

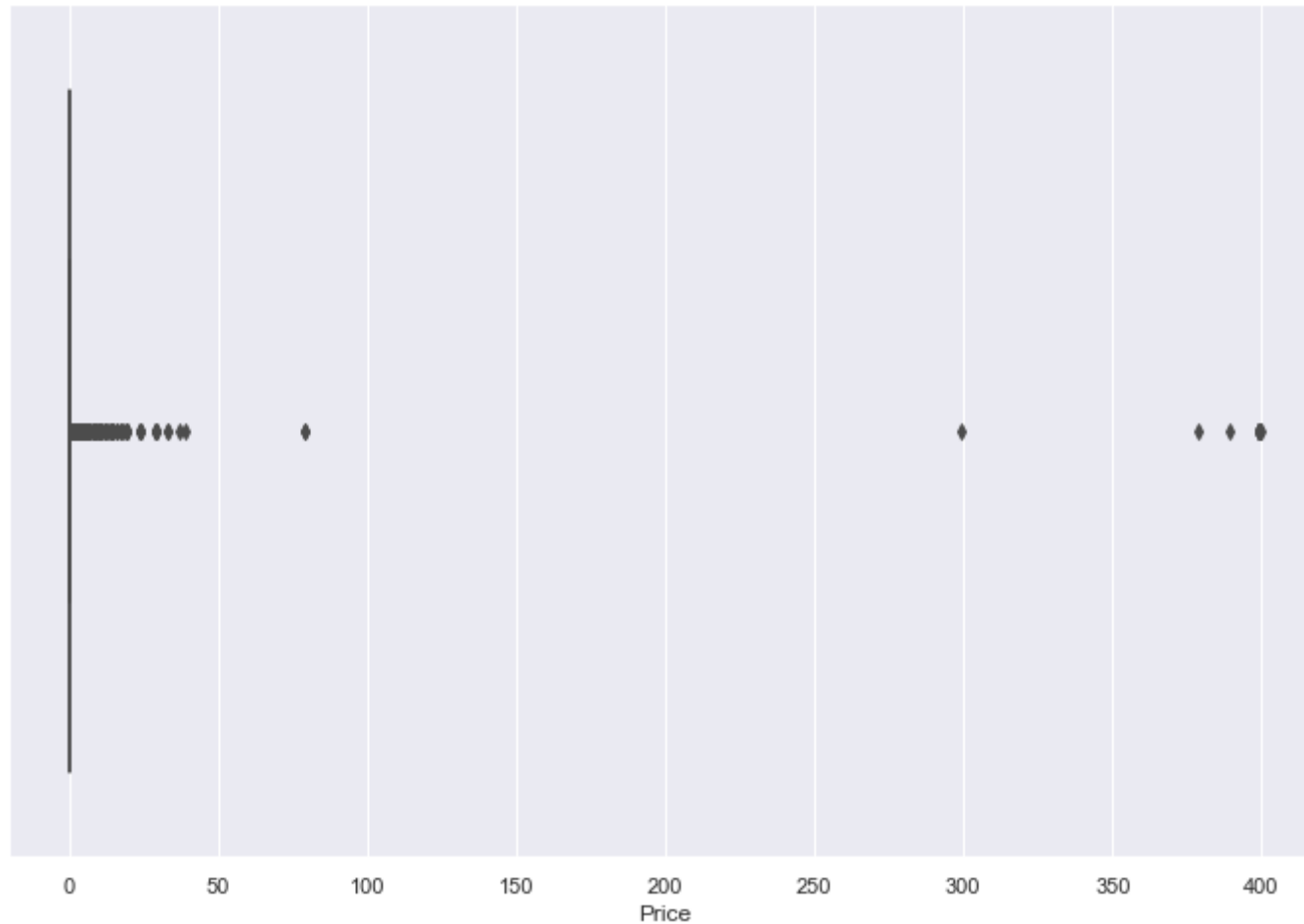
```
In [35]: sns.set(rc={"figure.figsize":(12,8)})
```

```
In [36]: sns.boxplot(data["Price"])
```

C:\Users\biswa\anaconda3\lib\site-packages\seaborn_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.

warnings.warn(

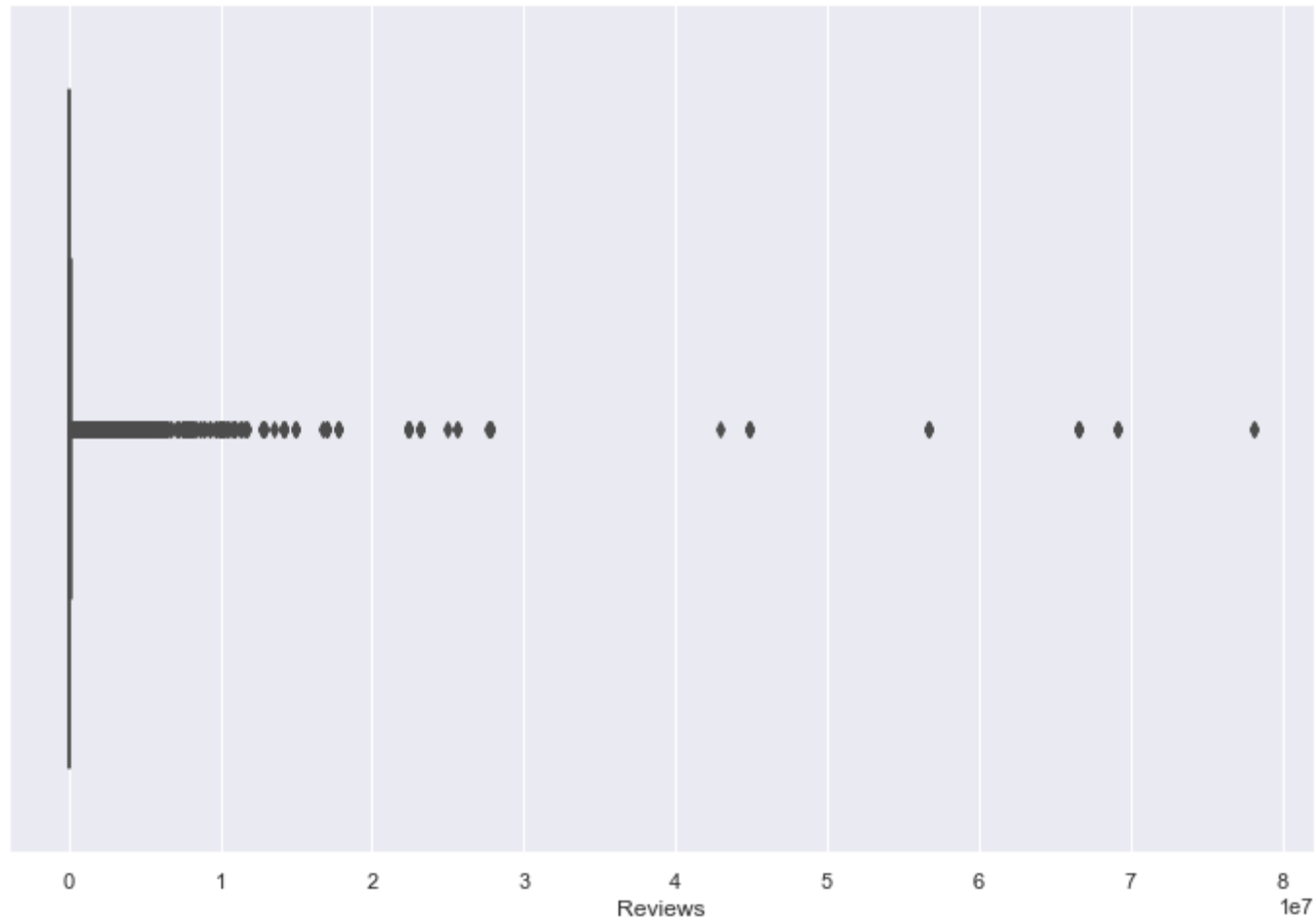
Out[36]: <AxesSubplot:xlabel='Price'>



```
In [37]: sns.boxplot(data["Reviews"])
```

C:\Users\biswa\anaconda3\lib\site-packages\seaborn_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.

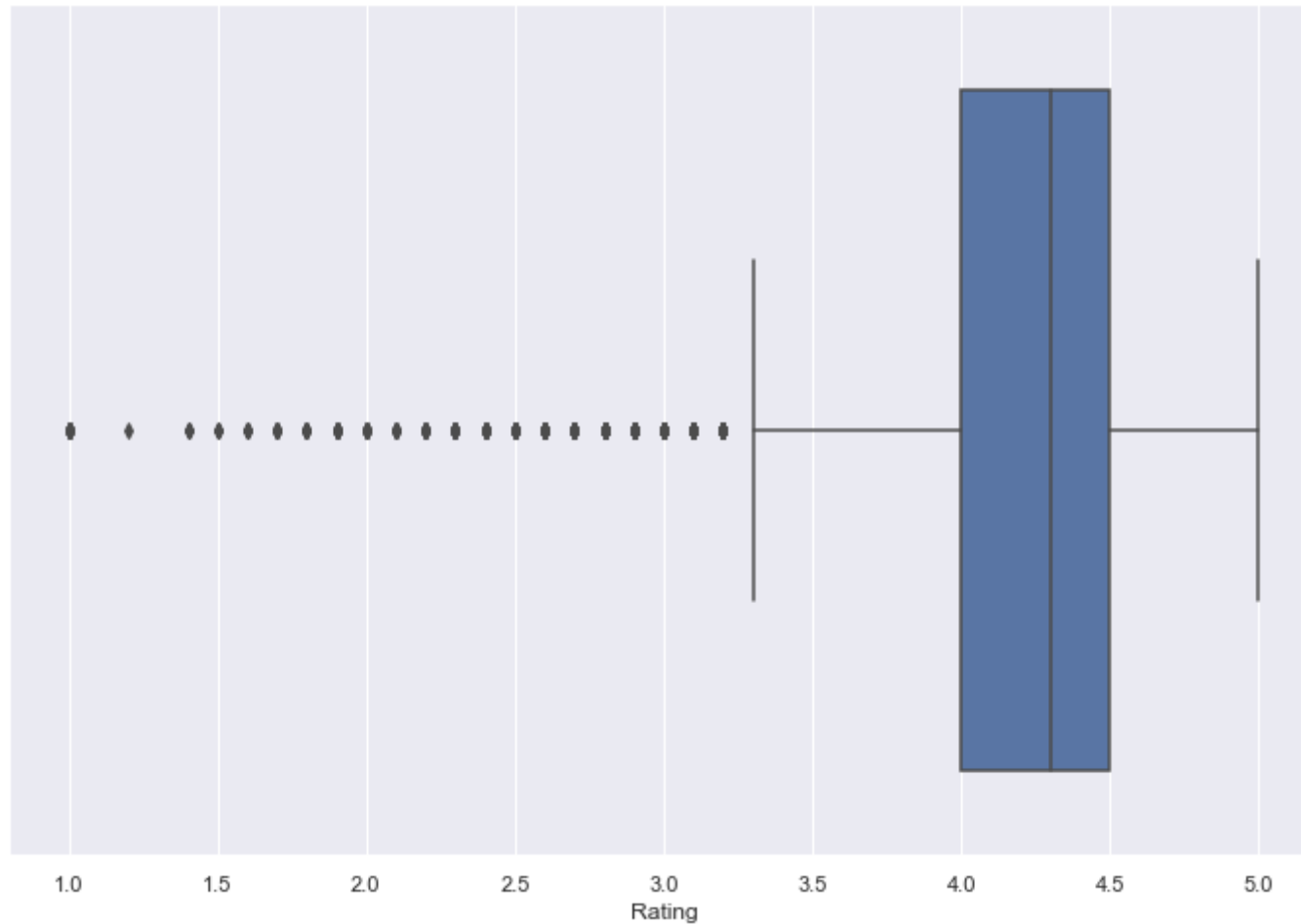
```
warnings.warn(  
Out[37]: <AxesSubplot:xlabel='Reviews'>
```



```
In [38]: sns.boxplot(data["Rating"])
```

C:\Users\biswa\anaconda3\lib\site-packages\seaborn_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.

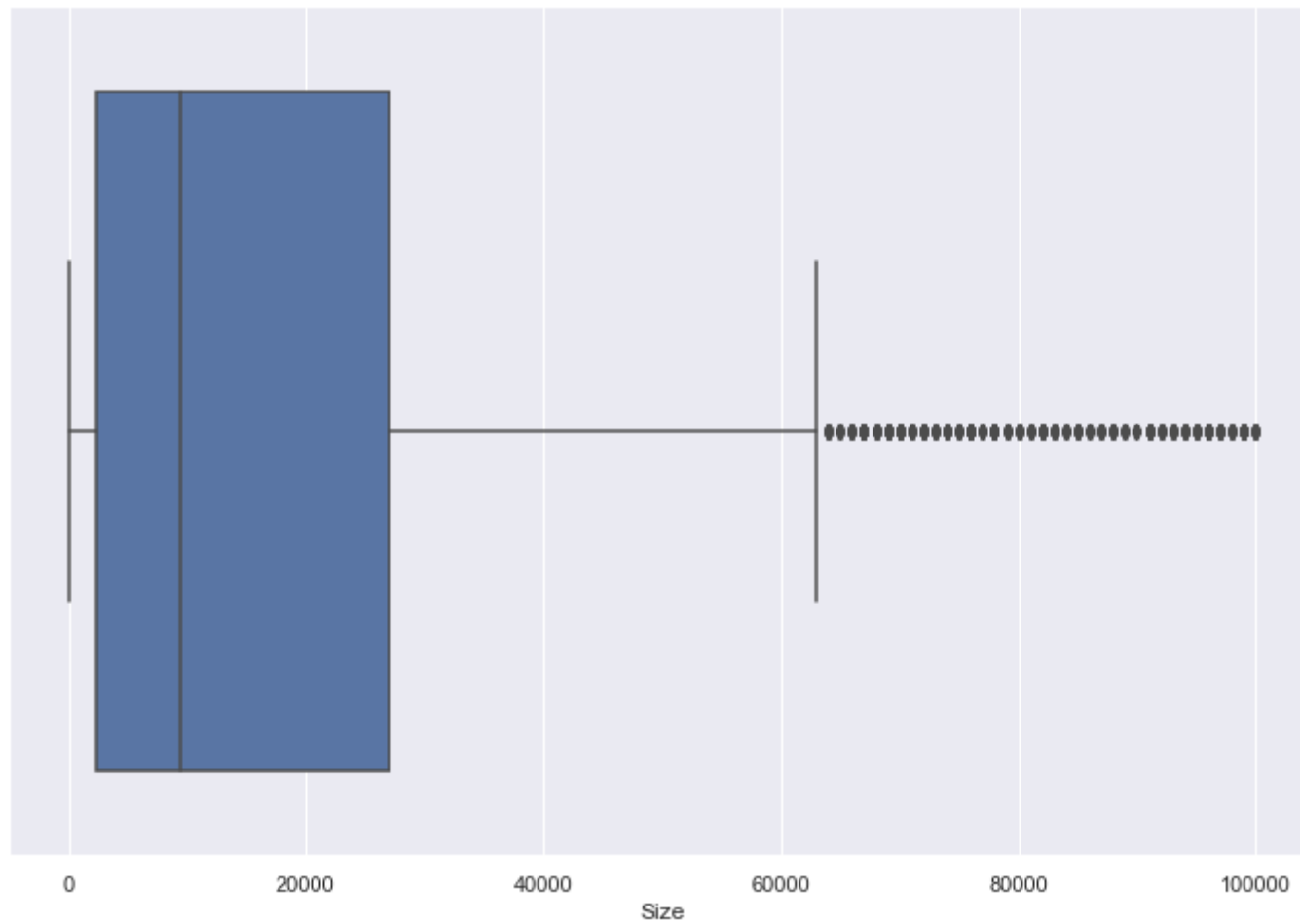
```
warnings.warn(  
Out[38]: <AxesSubplot:xlabel='Rating'>
```



```
In [39]: sns.boxplot(data['Size'])
```

C:\Users\biswa\anaconda3\lib\site-packages\seaborn_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.

```
warnings.warn(  
Out[39]: <AxesSubplot:xlabel='Size'>
```



```
In [40]: more=data.apply(lambda x: True if x["Price"] > 200 else False, axis=1)
```

```
In [41]: more_count=len(more[more==True].index)
```

```
In [42]: data.shape
```

```
Out[42]: (9353, 13)
```

```
In [43]:
```

```
data.drop(data[data["Price"]>200].index,inplace=True)
```

```
In [44]: data.shape
```

```
Out[44]: (9338, 13)
```

```
In [45]: data.drop(data[data["Reviews"]>2000000].index,inplace=True)
```

```
In [46]: data.shape
```

```
Out[46]: (8885, 13)
```

```
In [47]: data.quantile([.1, .25, .5, .70, .90, .95, .99], axis = 0)
```

```
Out[47]:
```

	Rating	Reviews	Size	Installs	Price
0.10	3.5	18.00	0.0	1000.0	0.0
0.25	4.0	159.00	2600.0	10000.0	0.0
0.50	4.3	4290.00	9500.0	500000.0	0.0
0.70	4.5	35930.40	23000.0	1000000.0	0.0
0.90	4.7	296771.00	50000.0	10000000.0	0.0
0.95	4.8	637298.00	68000.0	10000000.0	1.0
0.99	5.0	1462800.88	95000.0	100000000.0	7.0

```
In [48]: data.drop(data[data["Installs"]> 1000000].index,inplace=True)
```

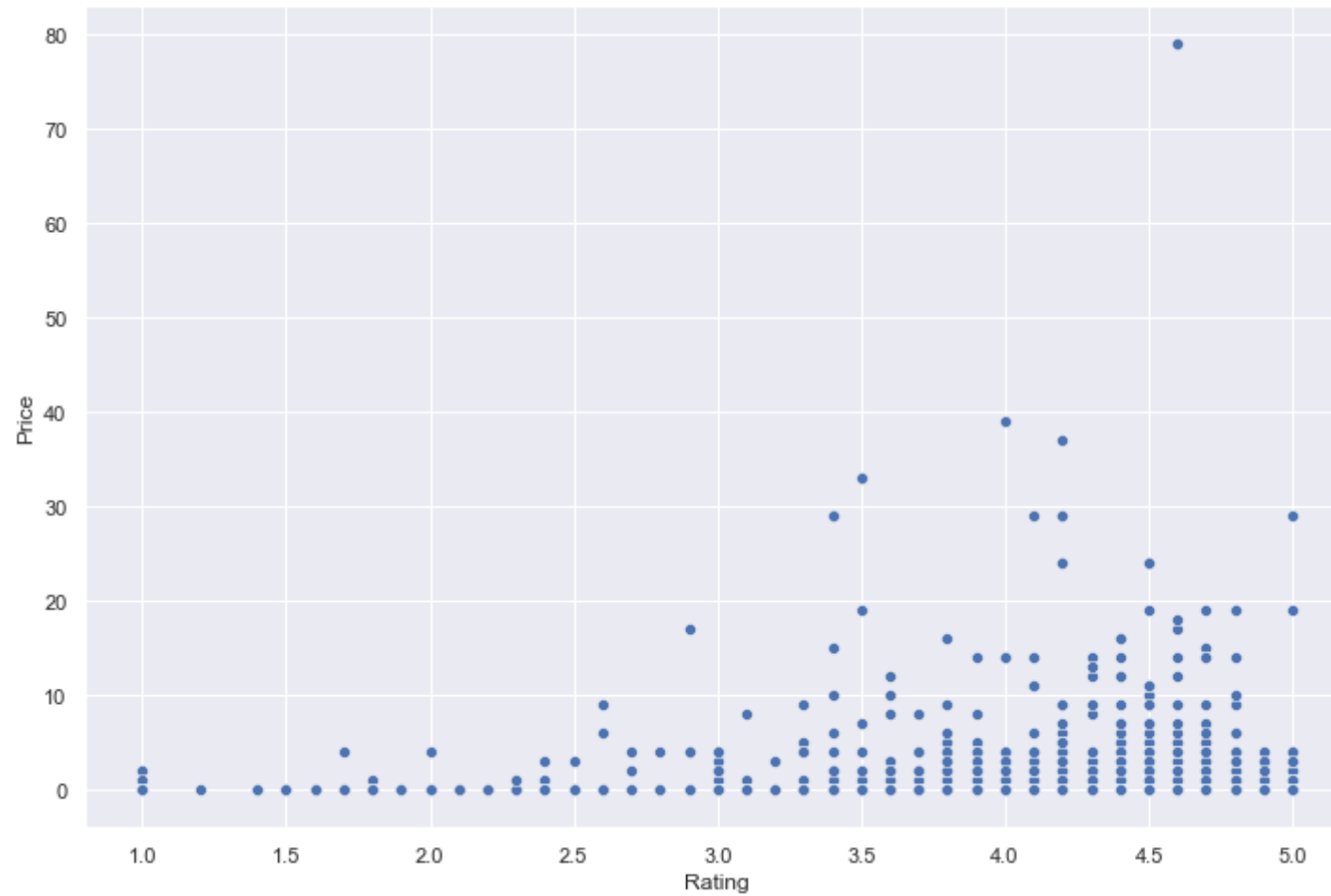
```
In [49]: data.shape
```

```
Out[49]: (6506, 13)
```



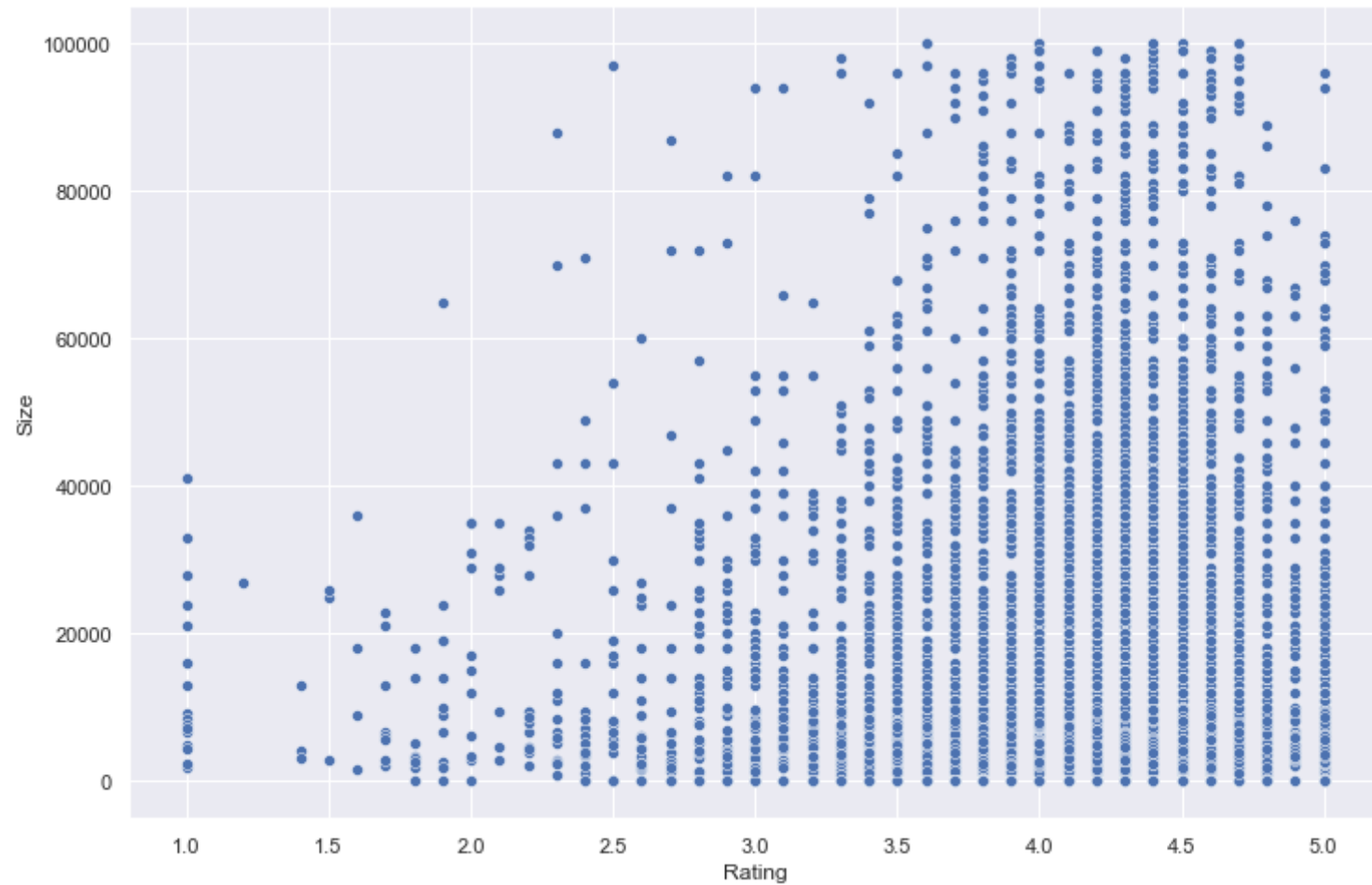
```
In [50]: sns.scatterplot(x="Rating",y="Price", data=data)
```

```
Out[50]: <AxesSubplot:xlabel='Rating', ylabel='Price'>
```



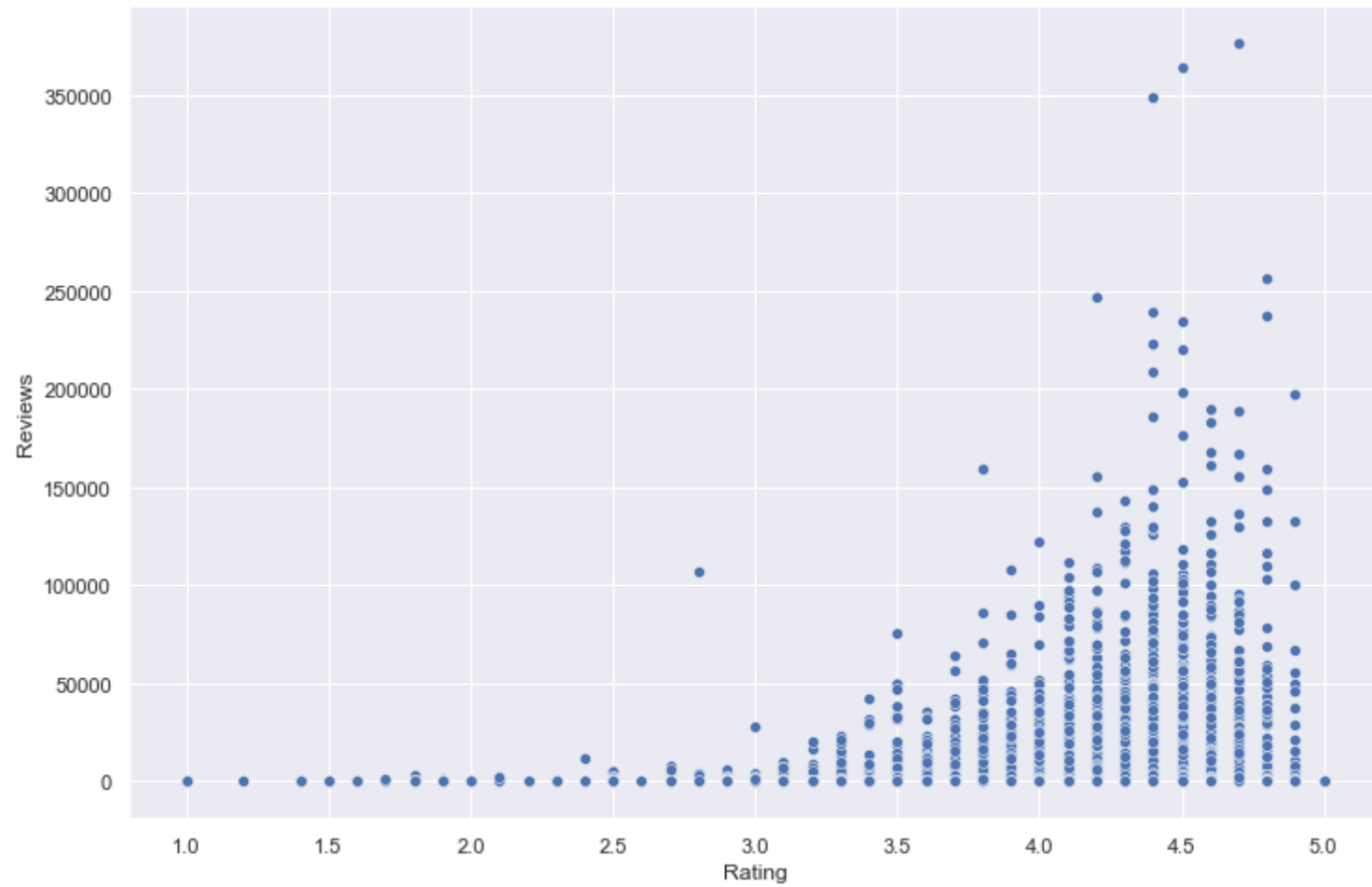
```
In [51]: sns.scatterplot(x="Rating", y="Size", data=data)
```

```
Out[51]: <AxesSubplot:xlabel='Rating', ylabel='Size'>
```



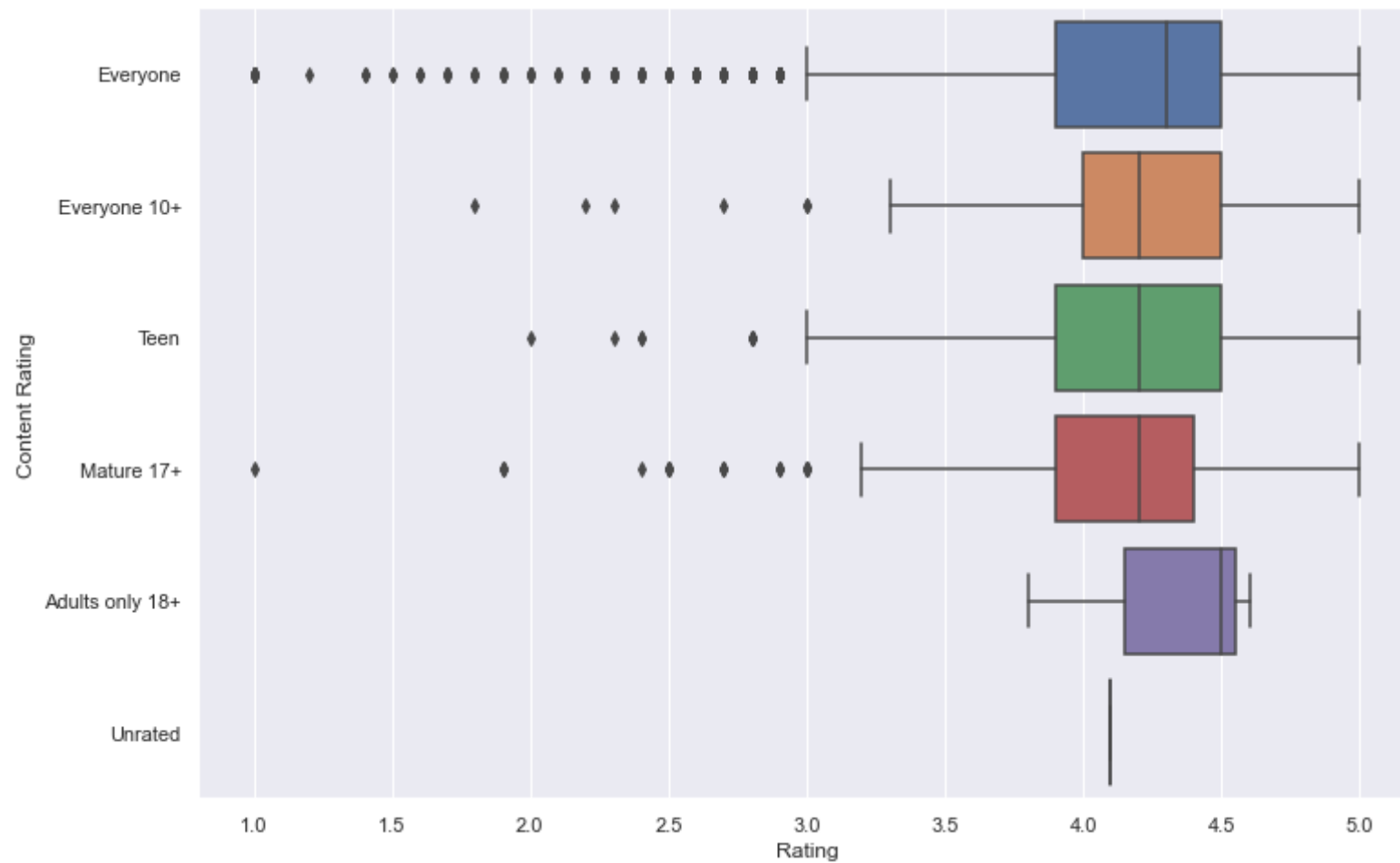
```
In [52]: sns.scatterplot(x="Rating",y="Reviews", data=data)
```

```
Out[52]: <AxesSubplot:xlabel='Rating', ylabel='Reviews'>
```



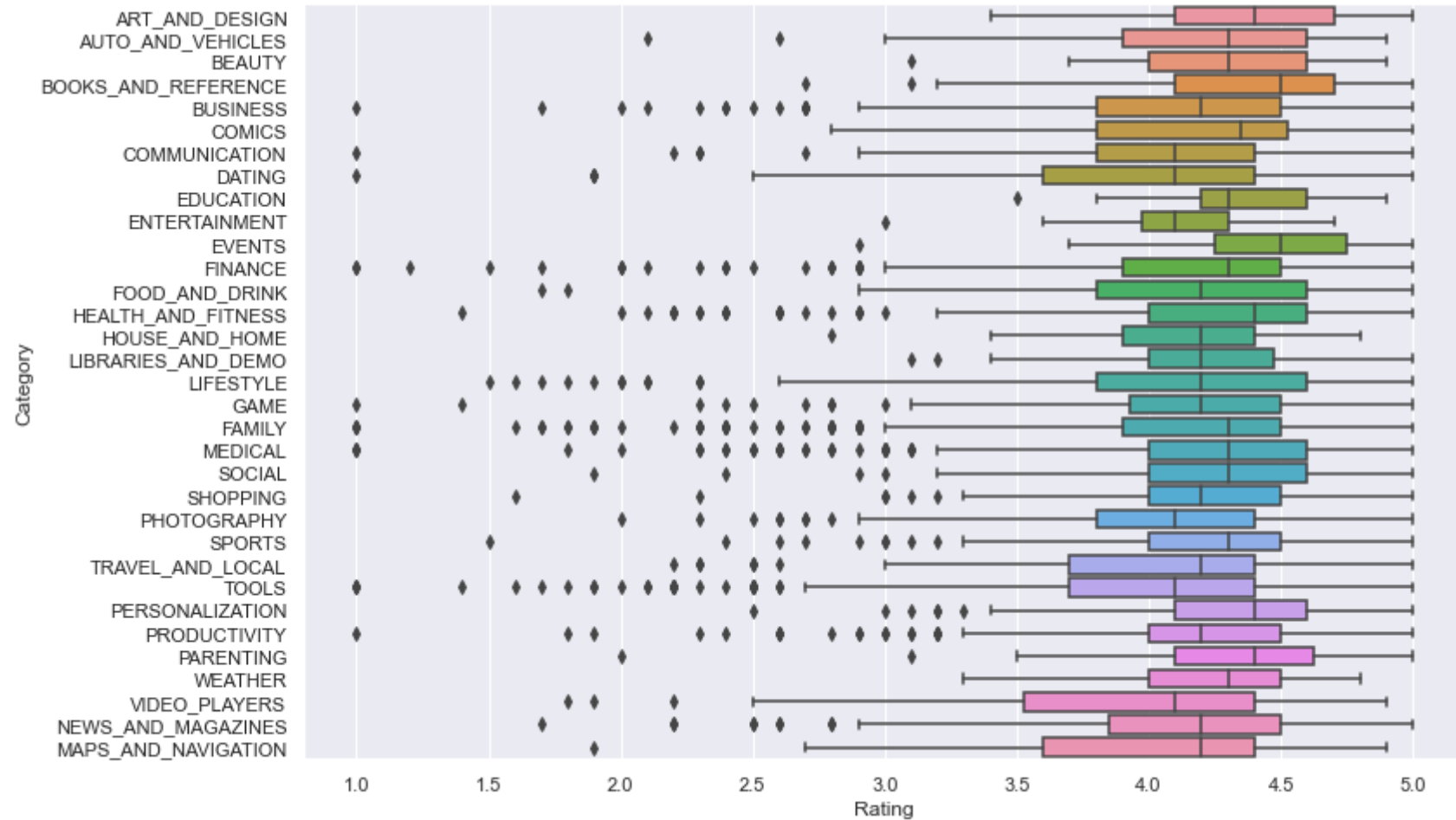
```
In [53]: sns.boxplot(x="Rating",y="Content Rating", data=data)
```

```
Out[53]: <AxesSubplot:xlabel='Rating', ylabel='Content Rating'>
```



```
In [54]: sns.boxplot(x="Rating",y="Category",data=data)
```

```
Out[54]: <AxesSubplot:xlabel='Rating', ylabel='Category'>
```



In [55]: `inp1=data`

In [56]: `inp1.head()`

Out[56]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159.0	19000.0	10000	Free	0	Everyone	Art & Design	January 7, 2018	1.0.0	4.0.3 and up

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
1	Coloring book moana	ART_AND_DESIGN	3.9	967.0	14000.0	500000	Free	0	Everyone	Design;Pretend Play	January 15, 2018	2.0.0	4.0.3 and up
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967.0	2800.0	100000	Free	0	Everyone	Art & Design;Creativity	June 20, 2018	1.1	4.4 and up
5	Paper flowers instructions	ART_AND_DESIGN	4.4	167.0	5600.0	50000	Free	0	Everyone	Art & Design	March 26, 2017	1.0	2.3 and up
6	Smoke Effect Photo Maker - Smoke Editor	ART_AND_DESIGN	3.8	178.0	19000.0	50000	Free	0	Everyone	Art & Design	April 26, 2018	1.1	4.0.3 and up

In [57]:

```
inp1.skew()
```

C:\Users\biswa\AppData\Local\Temp\ipykernel_8\3545313420.py:1: FutureWarning: Dropping of nuisance columns in DataFrame reductions (with 'numeric_only=None') is deprecated; in a future version this will raise TypeError. Select only valid columns before calling the reduction.

Out[57]:

```
inp1.skew()
Rating      -1.568598
Reviews      5.524059
Size         1.766327
Installs     0.946400
Price        15.848580
dtype: float64
```

In [58]:

```
reviewskew = np.log1p(inp1['Reviews'])
inp1['Reviews'] = reviewskew
```

In [59]:

```
reviewskew.skew()
```

Out[59]:

```
-0.12499523695350515
```

In [60]:

```
installsskew = np.log1p(inp1['Installs'])
inp1['Installs']
```

```
Out[60]: 0      10000
          1     500000
          4     100000
          5      50000
          6      50000
          ...
        10833     1000
        10834      500
        10836     5000
        10837      100
        10839     1000
        Name: Installs, Length: 6506, dtype: int32
```

```
In [61]: installsskew.skew()
```

```
Out[61]: -0.5715318924862697
```

```
In [62]: inp1.head()
```

```
Out[62]:
```

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	5.075174	19000.0	10000	Free	0	Everyone	Art & Design	January 7, 2018	1.0.0	4.0.3 and up
1	Coloring book moana	ART_AND_DESIGN	3.9	6.875232	14000.0	500000	Free	0	Everyone	Art & Design;Pretend Play	January 15, 2018	2.0.0	4.0.3 and up
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	6.875232	2800.0	100000	Free	0	Everyone	Art & Design;Creativity	June 20, 2018	1.1	4.4 and up
5	Paper flowers instructions	ART_AND_DESIGN	4.4	5.123964	5600.0	50000	Free	0	Everyone	Art & Design	March 26, 2017	1.0	2.3 and up
6	Smoke Effect Photo Maker - Smoke Editor	ART_AND_DESIGN	3.8	5.187386	19000.0	50000	Free	0	Everyone	Art & Design	April 26, 2018	1.1	4.0.3 and up

```
In [63]: inp1.drop(["Last Updated", "Current Ver", "Android Ver", "App", "Type"],axis=1,inplace=True)
```

In [64]:

```
inp1.head()
```

Out[64]:

	Category	Rating	Reviews	Size	Installs	Price	Content Rating	Genres
0	ART_AND_DESIGN	4.1	5.075174	19000.0	10000	0	Everyone	Art & Design
1	ART_AND_DESIGN	3.9	6.875232	14000.0	500000	0	Everyone	Art & Design;Pretend Play
4	ART_AND_DESIGN	4.3	6.875232	2800.0	100000	0	Everyone	Art & Design;Creativity
5	ART_AND_DESIGN	4.4	5.123964	5600.0	50000	0	Everyone	Art & Design
6	ART_AND_DESIGN	3.8	5.187386	19000.0	50000	0	Everyone	Art & Design

In [65]:

```
inp1.shape
```

Out[65]:

```
(6506, 8)
```

In [66]:

```
inp2=inp1
```

In [67]:

```
inp2.head()
```

Out[67]:

	Category	Rating	Reviews	Size	Installs	Price	Content Rating	Genres
0	ART_AND_DESIGN	4.1	5.075174	19000.0	10000	0	Everyone	Art & Design
1	ART_AND_DESIGN	3.9	6.875232	14000.0	500000	0	Everyone	Art & Design;Pretend Play
4	ART_AND_DESIGN	4.3	6.875232	2800.0	100000	0	Everyone	Art & Design;Creativity
5	ART_AND_DESIGN	4.4	5.123964	5600.0	50000	0	Everyone	Art & Design
6	ART_AND_DESIGN	3.8	5.187386	19000.0	50000	0	Everyone	Art & Design

In [68]:

```
inp2.Category.unique()
```

Out[68]:

```
array(['ART_AND_DESIGN', 'AUTO_AND_VEHICLES', 'BEAUTY',  
      'BOOKS_AND_REFERENCE', 'BUSINESS', 'COMICS', 'COMMUNICATION',
```



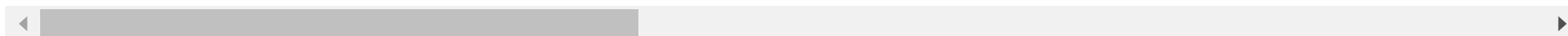
```
'DATING', 'EDUCATION', 'ENTERTAINMENT', 'EVENTS', 'FINANCE',
'FOOD_AND_DRINK', 'HEALTH_AND_FITNESS', 'HOUSE_AND_HOME',
'LIBRARIES_AND_DEMO', 'LIFESTYLE', 'GAME', 'FAMILY', 'MEDICAL',
'SOCIAL', 'SHOPPING', 'PHOTOGRAPHY', 'SPORTS', 'TRAVEL_AND_LOCAL',
'TOOLS', 'PERSONALIZATION', 'PRODUCTIVITY', 'PARENTING', 'WEATHER',
'VIDEO_PLAYERS', 'NEWS_AND_MAGAZINES', 'MAPS_AND_NAVIGATION'],
dtype=object)
```

```
In [69]: inp2.Category = pd.Categorical(inp2.Category)
x = inp2[['Category']]
del inp2['Category']
dummies = pd.get_dummies(x, prefix = 'Category')
inp2 = pd.concat([inp2,dummies], axis=1)
inp2.head()
```

```
Out[69]:
```

	Rating	Reviews	Size	Installs	Price	Content Rating	Genres	Category_ART_AND_DESIGN	Category_AUTO_AND_VEHICLES	Category_BEAUTY	...
0	4.1	5.075174	19000.0	10000	0	Everyone	Art & Design	1	0	0	...
1	3.9	6.875232	14000.0	500000	0	Everyone	Art & Design;Pretend Play	1	0	0	...
4	4.3	6.875232	2800.0	100000	0	Everyone	Art & Design;Creativity	1	0	0	...
5	4.4	5.123964	5600.0	50000	0	Everyone	Art & Design	1	0	0	...
6	3.8	5.187386	19000.0	50000	0	Everyone	Art & Design	1	0	0	...

5 rows × 40 columns



```
In [70]: inp2.shape
```

```
Out[70]: (6506, 40)
```

```
In [71]: inp2["Genres"].unique()
```

```

Out[71]: array(['Art & Design', 'Art & Design;Pretend Play',
               'Art & Design;Creativity', 'Auto & Vehicles', 'Beauty',
               'Books & Reference', 'Business', 'Comics', 'Comics;Creativity',
               'Communication', 'Dating', 'Education', 'Education;Education',
               'Education;Creativity', 'Education;Action & Adventure',
               'Education;Pretend Play', 'Education;Brain Games', 'Entertainment',
               'Entertainment;Brain Games', 'Entertainment;Creativity',
               'Entertainment;Music & Video', 'Events', 'Finance', 'Food & Drink',
               'Health & Fitness', 'House & Home', 'Libraries & Demo',
               'Lifestyle', 'Action', 'Arcade', 'Casual', 'Puzzle', 'Racing',
               'Sports', 'Simulation', 'Strategy', 'Role Playing', 'Board',
               'Simulation;Education', 'Casual;Brain Games',
               'Educational;Education', 'Card;Brain Games',
               'Educational;Brain Games', 'Entertainment;Education',
               'Casual;Education', 'Casual;Pretend Play', 'Music;Music & Video',
               'Educational;Creativity', 'Educational;Pretend Play',
               'Arcade;Pretend Play', 'Action;Action & Adventure',
               'Puzzle;Brain Games', 'Casual;Action & Adventure',
               'Adventure;Action & Adventure', 'Simulation;Pretend Play',
               'Puzzle;Creativity', 'Simulation;Action & Adventure',
               'Sports;Action & Adventure', 'Racing;Action & Adventure',
               'Educational;Action & Adventure', 'Arcade;Action & Adventure',
               'Entertainment;Action & Adventure', 'Strategy;Action & Adventure',
               'Music & Audio;Music & Video', 'Health & Fitness;Education',
               'Board;Action & Adventure', 'Board;Brain Games',
               'Board;Pretend Play', 'Education;Music & Video',
               'Role Playing;Pretend Play', 'Card;Action & Adventure', 'Medical',
               'Social', 'Shopping', 'Photography', 'Travel & Local',
               'Travel & Local;Action & Adventure', 'Tools', 'Personalization',
               'Productivity', 'Parenting', 'Parenting;Brain Games',
               'Parenting;Education', 'Parenting;Music & Video', 'Weather',
               'Video Players & Editors', 'News & Magazines', 'Maps & Navigation',
               'Adventure', 'Health & Fitness;Action & Adventure', 'Music',
               'Word', 'Educational', 'Card', 'Casino', 'Trivia',
               'Adventure;Brain Games', 'Role Playing;Action & Adventure',
               'Casual;Creativity', 'Lifestyle;Education',
               'Puzzle;Action & Adventure', 'Books & Reference;Education',
               'Puzzle;Education', 'Strategy;Education', 'Racing;Pretend Play',
               'Communication;Creativity', 'Entertainment;Pretend Play',
               'Strategy;Creativity'], dtype=object)

```

```

In [72]: lists = []
         for i in inp2.Genres.value_counts().index:

```

```

    if inp2.Genres.value_counts()[i]<20:
        lists.append(i)
inp2.Genres = ['Other' if i in lists else i for i in inp2.Genres]

```

```

In [73]: inp2["Genres"].unique()

```

```

Out[73]: array(['Art & Design', 'Other', 'Auto & Vehicles', 'Beauty',
        'Books & Reference', 'Business', 'Comics', 'Communication',
        'Dating', 'Education', 'Education;Education',
        'Education;Pretend Play', 'Entertainment', 'Events', 'Finance',
        'Food & Drink', 'Health & Fitness', 'House & Home',
        'Libraries & Demo', 'Lifestyle', 'Action', 'Arcade', 'Casual',
        'Puzzle', 'Racing', 'Sports', 'Simulation', 'Strategy',
        'Role Playing', 'Board', 'Educational;Education', 'Medical',
        'Social', 'Shopping', 'Photography', 'Travel & Local', 'Tools',
        'Personalization', 'Productivity', 'Parenting', 'Weather',
        'Video Players & Editors', 'News & Magazines', 'Maps & Navigation',
        'Adventure', 'Educational', 'Card', 'Casino', 'Trivia'],
        dtype=object)

```

```

In [74]: inp2.Genres = pd.Categorical(inp2['Genres'])
x = inp2[["Genres"]]
del inp2['Genres']
dummies = pd.get_dummies(x, prefix = 'Genres')
inp2 = pd.concat([inp2,dummies], axis=1)

```

```

In [75]: inp2.head()

```

```

Out[75]:

```

	Rating	Reviews	Size	Installs	Price	Content Rating	Category_ART_AND_DESIGN	Category_AUTO_AND_VEHICLES	Category_BEAUTY	Category_BOOKS_AN
0	4.1	5.075174	19000.0	10000	0	Everyone	1	0	0	
1	3.9	6.875232	14000.0	500000	0	Everyone	1	0	0	
4	4.3	6.875232	2800.0	100000	0	Everyone	1	0	0	
5	4.4	5.123964	5600.0	50000	0	Everyone	1	0	0	
6	3.8	5.187386	19000.0	50000	0	Everyone	1	0	0	

5 rows × 88 columns



```
In [76]: inp2.shape
```

```
Out[76]: (6506, 88)
```

```
In [77]: inp2["Content Rating"].unique()
```

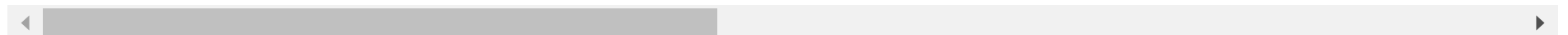
```
Out[77]: array(['Everyone', 'Everyone 10+', 'Teen', 'Mature 17+',  
              'Adults only 18+', 'Unrated'], dtype=object)
```

```
In [78]: inp2['Content Rating'] = pd.Categorical(inp2['Content Rating'])  
x = inp2[['Content Rating']]  
del inp2['Content Rating']  
dummies = pd.get_dummies(x, prefix = 'Content Rating')  
inp2 = pd.concat([inp2,dummies], axis=1)  
inp2.head()
```

```
Out[78]:
```

	Rating	Reviews	Size	Installs	Price	Category_ART_AND_DESIGN	Category_AUTO_AND_VEHICLES	Category_BEAUTY	Category_BOOKS_AND_REFERENCES
0	4.1	5.075174	19000.0	10000	0	1	0	0	
1	3.9	6.875232	14000.0	500000	0	1	0	0	
4	4.3	6.875232	2800.0	100000	0	1	0	0	
5	4.4	5.123964	5600.0	50000	0	1	0	0	
6	3.8	5.187386	19000.0	50000	0	1	0	0	

5 rows × 93 columns



```
In [79]: inp2.shape
```

Out[79]: (6506, 93)

```
In [80]: from sklearn.model_selection import train_test_split as tts
from sklearn.linear_model import LinearRegression as LR
from sklearn.metrics import mean_squared_error as mse
```

```
In [81]: d1 = inp2
X = d1.drop('Rating',axis=1)
y = d1['Rating']
Xtrain, Xtest, ytrain, ytest = tts(X,y, test_size=0.3, random_state=5)
```

```
In [82]: reg_all = LR()
reg_all.fit(Xtrain,ytrain)
```

Out[82]: LinearRegression()

```
In [83]: R2_train = round(reg_all.score(Xtrain,ytrain),3)
print("The R2 value of the Training Set is : {}".format(R2_train))
```

The R2 value of the Training Set is : 0.064

```
In [84]: R2_test = round(reg_all.score(Xtest,ytest),3)
print("The R2 value of the Testing Set is : {}".format(R2_test))
```

The R2 value of the Testing Set is : 0.027