In [1]:

```
#Loading the data file using pandas.
import pandas as pd
import numpy as np
import matplotlib.pyplot as pyp
import seaborn as sns
```

In [2]:

```
ds = pd.read_csv('googleplaystore.csv')
```

In [3]:

```
ds.head()
```

Out[3]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Cu
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	Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone	Art & Design;Pretend Play	January 15, 2018	
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	
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50,000,000+	Free	0	Teen	Art & Design	June 8, 2018	V de
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	

In [4]:

```
ds.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 [5]:

```
ds.shape
```

Out[5]:

```
(10841, 13)
```

In [6]:

```
# 2 Checkig for null values for each column in the data.
ds.isnull().any()
```

Out[6]:

```
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 [7]:

```
ds.isnull().sum()
```

Out[7]:

```
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 [8]:

```
# 3 Dropping records with nulls in any of the columns.
ds=ds.dropna()
```

In [9]:

```
ds.isnull().any()
```

Out[9]:

```
App               False
Category          False
Rating            False
Reviews           False
Size              False
Installs          False
```

```
Type              False
Price             False
Content Rating    False
Genres            False
Last Updated      False
Current Ver       False
Android Ver       False
dtype: bool
```

In [10]:

```
ds.shape
```

Out[10]:

```
(9360, 13)
```

In [11]:

```
# 4 Fixing the incorrect type and inconsistent formatting Variables.
ds["Size"] = [ float(i.split('M')[0]) if 'M' in i else float(0) for i in ds["Size"]  ]
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_g
uide/indexing.html#returning-a-view-versus-a-copy
```

In [12]:

```
ds.head()
```

Out[12]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Cur
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	Coloring book moana	ART_AND_DESIGN	3.9	967	14.0	500,000+	Free	0	Everyone	Art & Design;Pretend Play	January 15, 2018	
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	
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25.0	50,000,000+	Free	0	Teen	Art & Design	June 8, 2018	Va de
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	

In [13]:

```
# 4-1(1) Multiplying the value by 1,000, if size is mentioned in Mb
ds["Size"] = 1000 * ds["Size"]
ds
```

Out[13]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19000.0	10,000+	Free	0	Everyone	Art & Desig
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14000.0	500,000+	Free	0	Everyone	Art & Design;Preten Pla
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8700.0	5,000,000+	Free	0	Everyone	Art & Desig
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25000.0	50,000,000+	Free	0	Teen	Art & Desig
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2800.0	100,000+	Free	0	Everyone	Art & Design;Creativit
...	...	...	...	...	...	...	...	...	...	.
10834	FR Calculator	FAMILY	4.0	7	2600.0	500+	Free	0	Everyone	Educatio
10836	Sya9a Maroc - FR	FAMILY	4.5	38	53000.0	5,000+	Free	0	Everyone	Educatio
10837	Fr. Mike Schmitz Audio Teachings	FAMILY	5.0	4	3600.0	100+	Free	0	Everyone	Educatio
10839	The SCP Foundation DB fr nn5n	BOOKS_AND_REFERENCE	4.5	114	0.0	1,000+	Free	0	Mature 17+	Books Referenc
10840	iHoroscope - 2018 Daily Horoscope & Astrology	LIFESTYLE	4.5	398307	19000.0	10,000,000+	Free	0	Everyone	Lifestyl

9360 rows × 13 columns

In [14]:

```
ds.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 
```

```
 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 [15]:

```
# 4-2 Reviews is a numeric field that is loaded as a string field. Convert it to numeric (int/float).
ds["Reviews"] = ds["Reviews"].astype(int)
ds.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   int64
 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), int64(1), object(10)
memory usage: 1023.8+ KB
```

In [16]:

```
# 4- 3 Installs field is currently stored as string and has values like 1,000,000+.
ds['Installs']=ds['Installs'].map(lambda x:str(x).replace('+',''))
ds['Installs']=ds['Installs'].map(lambda x:str(x).replace(',',''))
ds['Installs']=ds['Installs'].astype(int)
ds.head()
```

Out[16]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Cu
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19000.0	10000	Free	0	Everyone	Art & Design	January 7, 2018	
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14000.0	500000	Free	0	Everyone	Art & Design;Pretend Play	January 15, 2018	
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8700.0	5000000	Free	0	Everyone	Art & Design	August 1, 2018	
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25000.0	50000000	Free	0	Teen	Art & Design	June 8, 2018	V d
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2800.0	100000	Free	0	Everyone	Art & Design;Creativity	June 20, 2018	

In [17]:

```
# 4-4 Price field is a string and has $ symbol. Remove '$' sign, and convert it to numeri
c.
ds['Price'] = [ float(i.split('$')[1]) if '$' in i else float(0) for i in ds['Price'] ]
ds.head()
```

Out[17]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Cu
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19000.0	10000	Free	0.0	Everyone	Art & Design	January 7, 2018	
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14000.0	500000	Free	0.0	Everyone	Art & Design;Pretend Play	January 15, 2018	
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8700.0	5000000	Free	0.0	Everyone	Art & Design	August 1, 2018	
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25000.0	50000000	Free	0.0	Teen	Art & Design	June 8, 2018	V d
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2800.0	100000	Free	0.0	Everyone	Art & Design;Creativity	June 20, 2018	

In [18]:

```
# 4-5 Sanity checks:
ds.drop(ds[(ds['Reviews'] < 1) & (ds['Reviews'] > 5 )].index, inplace = True)
ds
```

Out[18]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19000.0	10000	Free	0.0	Everyone	Art & Design
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14000.0	500000	Free	0.0	Everyone	Art & Design;Pretend Play
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8700.0	5000000	Free	0.0	Everyone	Art & Design
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25000.0	50000000	Free	0.0	Teen	Art & Design
	Pixel Draw - Number									

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2800.0	100000	Free	0.0	Everyone	Art & Design;Creativity
...	...	...	...	...	...	...	...	...	...	...
10834	FR Calculator	FAMILY	4.0	7	2600.0	500	Free	0.0	Everyone	Education
10836	Sya9a Maroc - FR	FAMILY	4.5	38	53000.0	5000	Free	0.0	Everyone	Education
10837	Fr. Mike Schmitz Audio Teachings	FAMILY	5.0	4	3600.0	100	Free	0.0	Everyone	Education
10839	The SCP Foundation DB fr nn5n	BOOKS_AND_REFERENCE	4.5	114	0.0	1000	Free	0.0	Mature 17+	Books & Reference
10840	iHoroscope - 2018 Daily Horoscope & Astrology	LIFESTYLE	4.5	398307	19000.0	10000000	Free	0.0	Everyone	Lifestyle

9360 rows × 13 columns

In [19]:

```
ds=ds.reset_index()
ds.drop('index', inplace=True, axis=1)
ds
```

Out[19]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19000.0	10000	Free	0.0	Everyone	Art & Design
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14000.0	500000	Free	0.0	Everyone	Art & Design;Pretend Play
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8700.0	5000000	Free	0.0	Everyone	Art & Design
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25000.0	50000000	Free	0.0	Teen	Art & Design
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2800.0	100000	Free	0.0	Everyone	Art & Design;Creativity
...	...	...	...	...	...	...	...	...	...	...
9355	FR Calculator	FAMILY	4.0	7	2600.0	500	Free	0.0	Everyone	Education
9356	Sya9a Maroc - FR	FAMILY	4.5	38	53000.0	5000	Free	0.0	Everyone	Education
9357	Fr. Mike Schmitz Audio Teachings	FAMILY	5.0	4	3600.0	100	Free	0.0	Everyone	Education

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	U
9358	The SCP Foundation DB fr nn5n	BOOKS_AND_REFERENCE	4.5	114	0.0	1000	Free	0.0	Mature 17+	Books & Reference	
9359	iHoroscope - 2018 Daily Horoscope & Astrology	LIFESTYLE	4.5	398307	19000.0	10000000	Free	0.0	Everyone	Lifestyle	

9360 rows × 13 columns

In [20]:

```
ds.drop(ds[ds['Installs'] < ds['Reviews'] ].index, inplace = True)
```

In [21]:

```
ds.head()
```

Out[21]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Cu
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19000.0	10000	Free	0.0	Everyone	Art & Design	January 7, 2018	
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14000.0	500000	Free	0.0	Everyone	Art & Design;Pretend Play	January 15, 2018	
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8700.0	5000000	Free	0.0	Everyone	Art & Design	August 1, 2018	
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25000.0	50000000	Free	0.0	Teen	Art & Design	June 8, 2018	V d
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2800.0	100000	Free	0.0	Everyone	Art & Design;Creativity	June 20, 2018	

In [22]:

```
ds.shape
```

Out[22]:

```
(9353, 13)
```

In [23]:

```
ds.drop(ds[(ds['Type'] =='Free') & (ds['Price'] > 0 )].index, inplace = True)
ds
```

Out[23]:

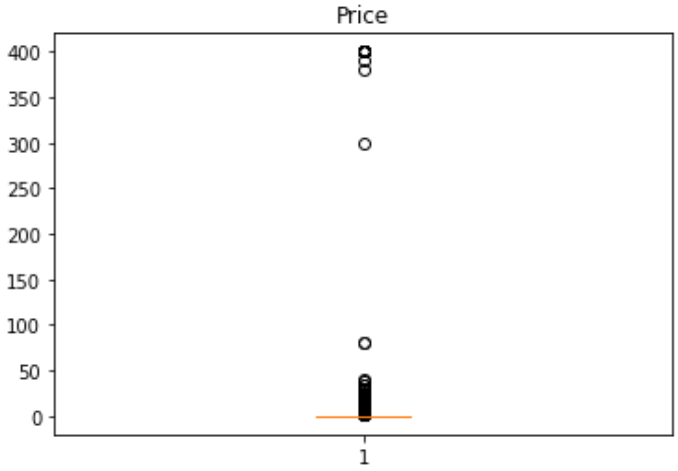
	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	U
	Photo										

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres
0	Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19000.0	10000	Free	0.0	Everyone	Art & Design
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14000.0	500000	Free	0.0	Everyone	Art & Design;Pretend Play
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8700.0	5000000	Free	0.0	Everyone	Art & Design
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25000.0	50000000	Free	0.0	Teen	Art & Design
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2800.0	100000	Free	0.0	Everyone	Art & Design;Creativity
...	...	...	...	...	...	...	...	...	...	...
9355	FR Calculator	FAMILY	4.0	7	2600.0	500	Free	0.0	Everyone	Education
9356	Sya9a Maroc - FR	FAMILY	4.5	38	53000.0	5000	Free	0.0	Everyone	Education
9357	Fr. Mike Schmitz Audio Teachings	FAMILY	5.0	4	3600.0	100	Free	0.0	Everyone	Education
9358	The SCP Foundation DB fr nn5n	BOOKS_AND_REFERENCE	4.5	114	0.0	1000	Free	0.0	Mature 17+	Books & Reference
9359	iHoroscope - 2018 Daily Horoscope & Astrology	LIFESTYLE	4.5	398307	19000.0	10000000	Free	0.0	Everyone	Lifestyle

9353 rows × 13 columns

In [24]:

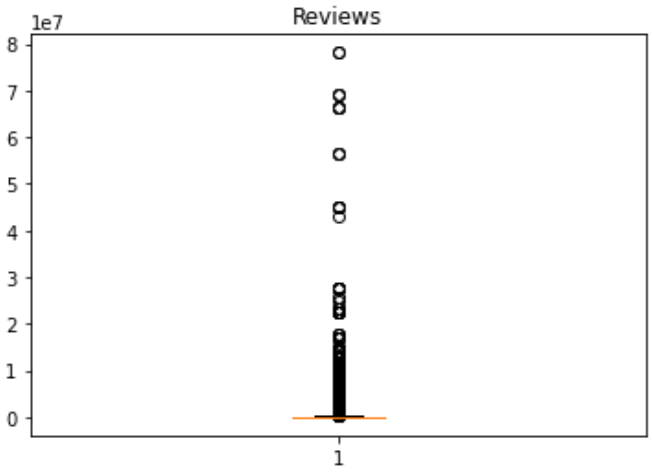
```
# 4-5 (1)Boxplot for Price _Are there any outliers? Think about the price of usual apps o
n Play Store.
pyp.boxplot(ds['Price']);
pyp.title ('Price');
```



In [25]:

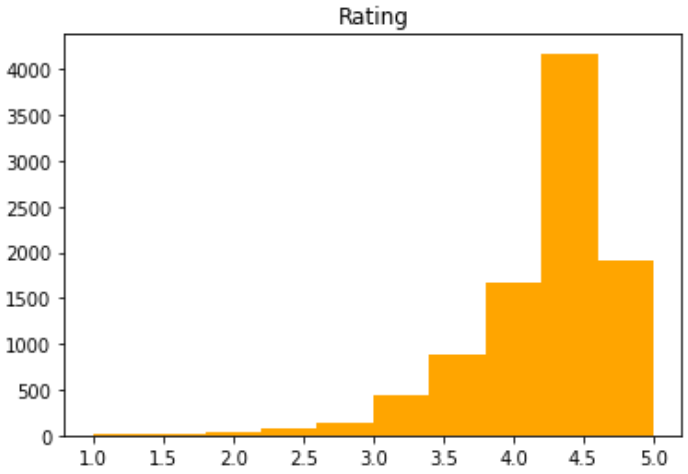
```
# 4-5 (2) Boxplot for Reviews Are there any apps with very high number of reviews? Do the
```

```
values seem right?
pyp.boxplot(ds['Reviews']);
pyp.title ('Reviews');
```



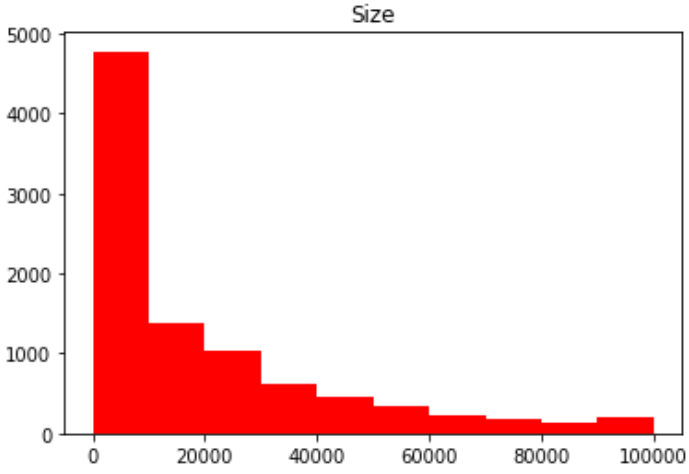
In [26]:

```
# 4-5(3) Histogram for Rating_How are the ratings distributed? Is it more toward higher r
atings?
pyp.hist(ds['Rating'],color='Orange');
pyp.title('Rating');
```



In [27]:

```
pyp.hist(ds['Size'],color='red');
pyp.title('Size');
```



In [28]:

```
# 6. Outlier treatment: Price: From the box plot, it seems like there are some apps with
very high price. A price of $200 for an application on the Play Store is very high and su
spicious!
```

```
more = ds.apply(lambda x : True
            if x['Price'] > 200 else False, axis = 1)
more_count = len(more[more == True].index)
ds
```

Out[28]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19000.0	10000	Free	0.0	Everyone	Art & Design
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14000.0	500000	Free	0.0	Everyone	Art & Design;Pretend Play
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8700.0	5000000	Free	0.0	Everyone	Art & Design
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25000.0	50000000	Free	0.0	Teen	Art & Design
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2800.0	100000	Free	0.0	Everyone	Art & Design;Creativity
...	...	...	...	...	...	...	...	...	...	...
9355	FR Calculator	FAMILY	4.0	7	2600.0	500	Free	0.0	Everyone	Education
9356	Sya9a Maroc - FR	FAMILY	4.5	38	53000.0	5000	Free	0.0	Everyone	Education
9357	Fr. Mike Schmitz Audio Teachings	FAMILY	5.0	4	3600.0	100	Free	0.0	Everyone	Education
9358	The SCP Foundation DB fr nn5n	BOOKS_AND_REFERENCE	4.5	114	0.0	1000	Free	0.0	Mature 17+	Books & Reference
9359	iHoroscope - 2018 Daily Horoscope & Astrology	LIFESTYLE	4.5	398307	19000.0	10000000	Free	0.0	Everyone	Lifestyle

9353 rows × 13 columns

In [29]:

```
ds.shape
```

Out[29]:

```
(9353, 13)
```

In [30]:

```
ds.drop(ds[ds['Price'] > 200].index, inplace = True)
ds
```

Out[30]:

	App App	Category Category	Rating Rating	Reviews Reviews	Size Size	Installs Installs	Type Type	Price Price	Content Rating Content Rating	Genres Genres
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19000.0	10000	Free	0.0	Everyone	Art & Design
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14000.0	500000	Free	0.0	Everyone	Art & Design;Pretend Play
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8700.0	5000000	Free	0.0	Everyone	Art & Design
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25000.0	50000000	Free	0.0	Teen	Art & Design
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2800.0	100000	Free	0.0	Everyone	Art & Design;Creativity
...	...	...	...	...	...	...	...	...	...	...
9355	FR Calculator	FAMILY	4.0	7	2600.0	500	Free	0.0	Everyone	Education
9356	Sya9a Maroc - FR	FAMILY	4.5	38	53000.0	5000	Free	0.0	Everyone	Education
9357	Fr. Mike Schmitz Audio Teachings	FAMILY	5.0	4	3600.0	100	Free	0.0	Everyone	Education
9358	The SCP Foundation DB fr nn5n	BOOKS_AND_REFERENCE	4.5	114	0.0	1000	Free	0.0	Mature 17+	Books & Reference
9359	iHoroscope - 2018 Daily Horoscope & Astrology	LIFESTYLE	4.5	398307	19000.0	10000000	Free	0.0	Everyone	Lifestyle

9338 rows × 13 columns

In [31]:

```
ds.shape
```

Out[31]:

```
(9338, 13)
```

In [32]:

```
ds.drop(ds[ds['Reviews'] > 2000000].index, inplace = True)
ds
```

Out[32]:

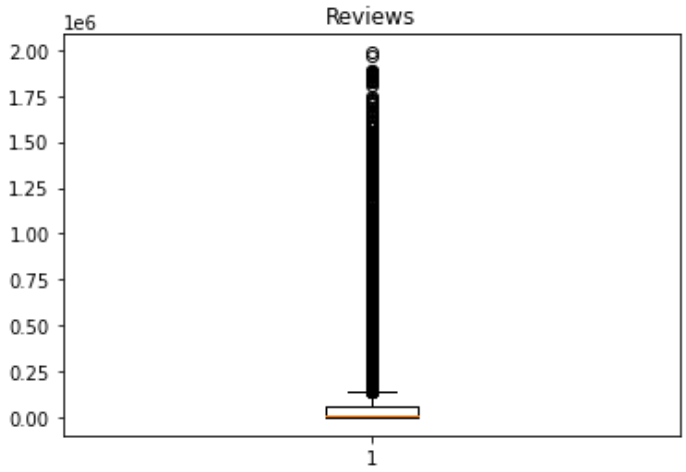
	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19000.0	10000	Free	0.0	Everyone	Art & Design

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14000.0	500000	Free	0.0	Everyone	Art & Design;Pretend Play
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8700.0	5000000	Free	0.0	Everyone	Art & Design
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25000.0	50000000	Free	0.0	Teen	Art & Design
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2800.0	100000	Free	0.0	Everyone	Art & Design;Creativity
...	...	...	...	...	...	...	...	...	...	...
9355	FR Calculator	FAMILY	4.0	7	2600.0	500	Free	0.0	Everyone	Education
9356	Sya9a Maroc - FR	FAMILY	4.5	38	53000.0	5000	Free	0.0	Everyone	Education
9357	Fr. Mike Schmitz Audio Teachings	FAMILY	5.0	4	3600.0	100	Free	0.0	Everyone	Education
9358	The SCP Foundation DB fr nn5n	BOOKS_AND_REFERENCE	4.5	114	0.0	1000	Free	0.0	Mature 17+	Books & Reference
9359	iHoroscope - 2018 Daily Horoscope & Astrology	LIFESTYLE	4.5	398307	19000.0	10000000	Free	0.0	Everyone	Lifestyle

8885 rows × 13 columns

In [33]:

```
pyp.boxplot(ds['Reviews']);
pyp.title ('Reviews');
```



In [34]:

```
# dropping more than 10000000 Installs value
ds.drop(ds[ds['Installs'] > 10000000].index, inplace = True)
ds
```

Out[34]:

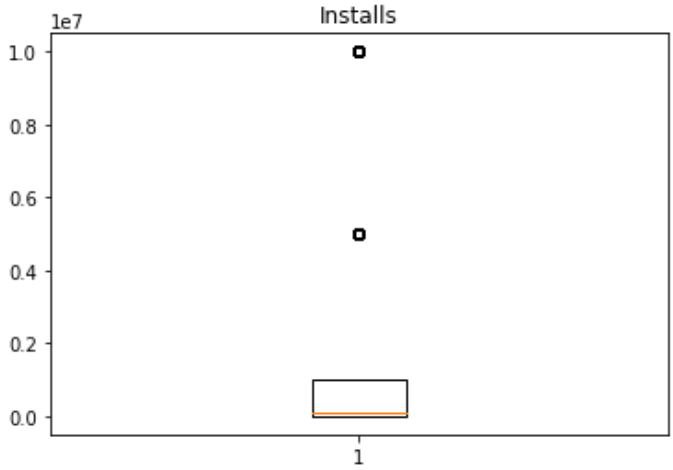
	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content

	App App	Category Category	Rating Rating	Reviews Reviews	Size Size	Installs Installs	Type Type	Price Price	Content Rating Rating	Genres Genres
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19000.0	10000	Free	0.0	Everyone	Art & Design
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14000.0	500000	Free	0.0	Everyone	Art & Design;Pretend Play
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8700.0	5000000	Free	0.0	Everyone	Art & Design
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2800.0	100000	Free	0.0	Everyone	Art & Design;Creativity
5	Paper flowers instructions	ART_AND_DESIGN	4.4	167	5600.0	50000	Free	0.0	Everyone	Art & Design
...	...	...	...	...	...	...	...	...	...	...
9355	FR Calculator	FAMILY	4.0	7	2600.0	500	Free	0.0	Everyone	Education
9356	Sya9a Maroc - FR	FAMILY	4.5	38	53000.0	5000	Free	0.0	Everyone	Education
9357	Fr. Mike Schmitz Audio Teachings	FAMILY	5.0	4	3600.0	100	Free	0.0	Everyone	Education
9358	The SCP Foundation DB fr nn5n	BOOKS_AND_REFERENCE	4.5	114	0.0	1000	Free	0.0	Mature 17+	Books & Reference
9359	iHoroscope - 2018 Daily Horoscope & Astrology	LIFESTYLE	4.5	398307	19000.0	10000000	Free	0.0	Everyone	Lifestyle

8496 rows × 13 columns

In [35]:

```
pyp.boxplot(ds['Installs']);
pyp.title ('Installs');
```



In [36]:

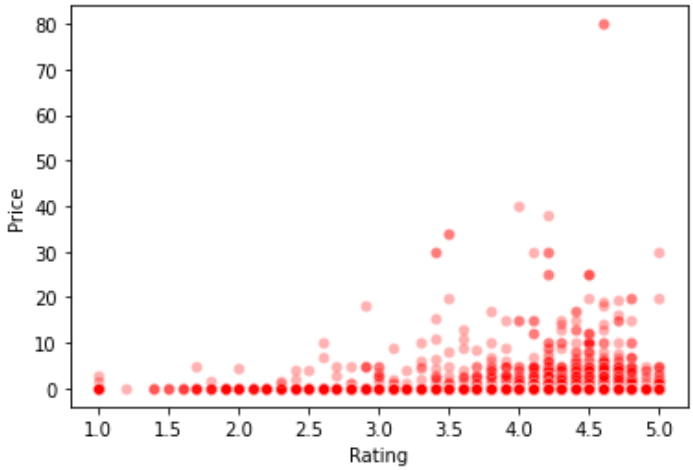
```
# 6-3 Installs:  There seems to be some outliers in this field too. Apps having very high number of installs should be dropped from the analysis.
ds.quantile([.1, .25, .5, .70, .90, .95, .99], axis = 0)
```

Out[36]:

	Rating	Reviews	Size	Installs	Price
0.10	3.5	16.0	0.0	1000.0	0.00
0.25	4.0	134.0	2900.0	10000.0	0.00
0.50	4.3	3325.5	9800.0	100000.0	0.00
0.70	4.5	27560.0	23000.0	1000000.0	0.00
0.90	4.7	192651.0	50000.0	10000000.0	0.00
0.95	4.8	355858.0	68250.0	10000000.0	1.99
0.99	5.0	899046.9	95000.0	10000000.0	7.99

In [37]:

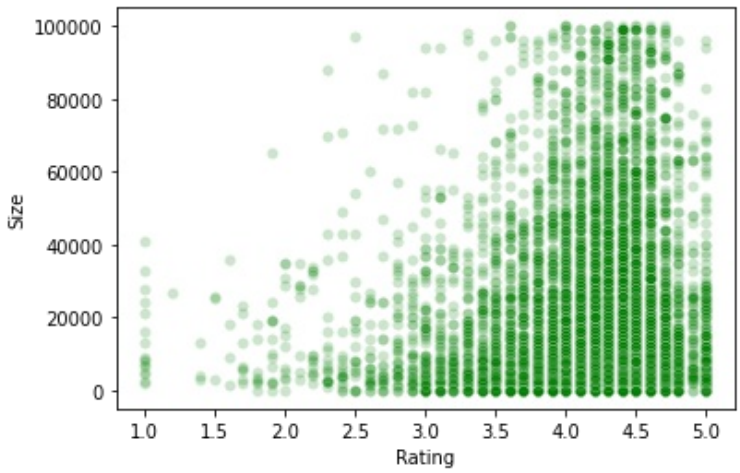
```
# 7 Bivariate analysis: Let's look at how the available predictors relate to the variable of interest, i.e., our target variable rating.
# Make scatter plots (for numeric features) and box plots (for character features) to assess the relations between rating and the other features.
# 7-1 Make scatter plot/joinplot for Rating vs. Price

sns.scatterplot(x='Rating',color='red',alpha=0.3,y='Price',data=ds);
```



YES, Rating does Increase with Price

In [38]:

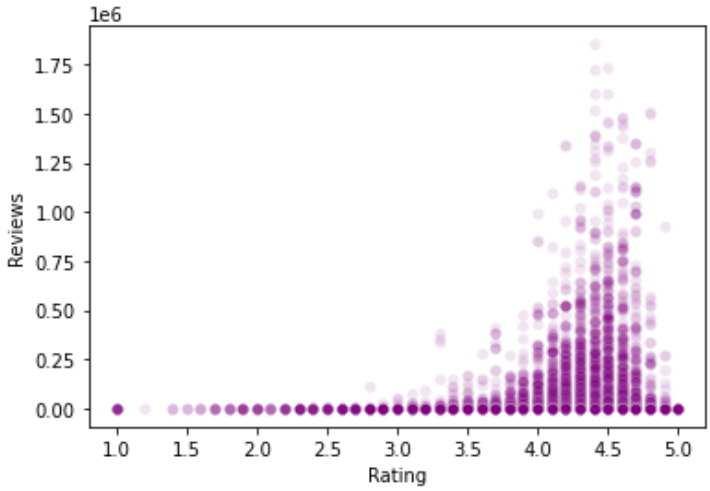
```
#7-2 Make scatter plot/joinplot for Rating vs. Size
sns.scatterplot(x='Rating',color='green',alpha= 0.2,y='Size',data=ds);
```



Yes, heavier apps are rated better

In [39]:

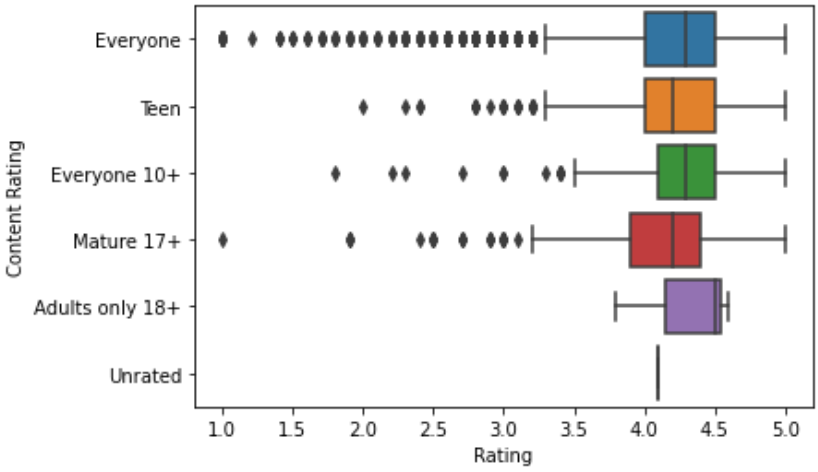
```
# 7-3 Make scatter plot/joinplot for Rating vs. Reviews
sns.scatterplot(x='Rating',color='purple',alpha=0.1,y='Reviews',data=ds);
```



As shown,more review mean a better rating

In [40]:

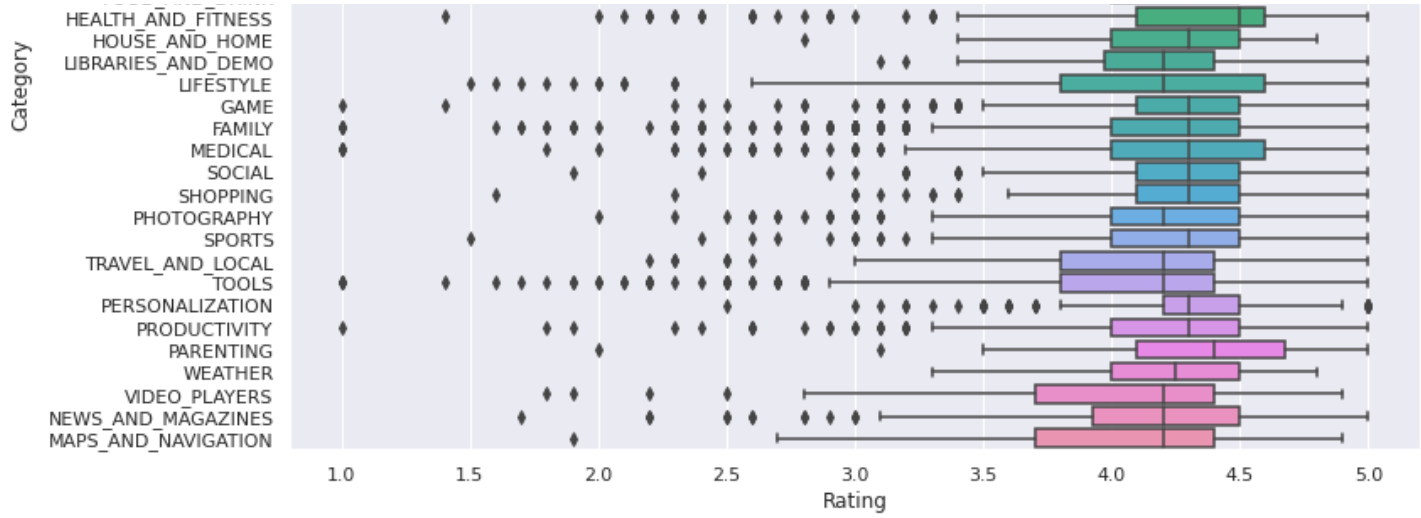
```
# 7- 4 Make boxplot for Rating vs. Content Rating
sns.boxplot(x="Rating", y="Content Rating", data=ds);
```



18+ apps have better ratings as compare to content open for all

In [41]:

```
# 7-5 Make boxplot for Ratings vs. Category
sns.set(rc={'figure.figsize':(12,8)})
sns.boxplot(x="Rating", y="Category", data=ds);
```



All category are better than other app:

In [42]:

```
# 8 Data preprocessing
#8 -1 Reviews and Install have some values that are still relatively very high.
#Before building a linear regression model, you need to reduce the skew. Apply log transf
ormation (np.log1p) to Reviews and Installs.
inp1 = ds
inp1.head()
```

Out[42]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Cu
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19000.0	10000	Free	0.0	Everyone	Art & Design	January 7, 2018	
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14000.0	500000	Free	0.0	Everyone	Art & Design;Pretend Play	January 15, 2018	
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8700.0	5000000	Free	0.0	Everyone	Art & Design	August 1, 2018	
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2800.0	100000	Free	0.0	Everyone	Art & Design;Creativity	June 20, 2018	
5	Paper flowers instructions	ART_AND_DESIGN	4.4	167	5600.0	50000	Free	0.0	Everyone	Art & Design	March 26, 2017	

In [43]:

```
inp1.skew()
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:1: FutureWarning: Dropping o
f 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 r
eduction.
  """Entry point for launching an IPython kernel.
```

Out[43]:

```
Rating      -1.749753
Reviews      4.576494
Size         1.655917
Installs     1.543697
Price       16.264811
dtype: float64
```

In [44]:

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

Out[44]:

```
-0.20039949659264134
```

In [45]:

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

Out[45]:

```
0          10000
1         500000
2        5000000
4         100000
5          50000
          ...
9355         500
9356        5000
9357         100
9358        1000
9359    10000000
Name: Installs, Length: 8496, dtype: int64
```

In [46]:

```
installsskew.skew()
```

Out[46]:

```
-0.5097286542754812
```

In [47]:

```
inp1.head()
```

Out[47]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	C
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	5.075174	19000.0	10000	Free	0.0	Everyone	Art & Design	January 7, 2018	
1	Coloring book moana	ART_AND_DESIGN	3.9	6.875232	14000.0	500000	Free	0.0	Everyone	Art & Design;Pretend Play	January 15, 2018	
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	11.379520	8700.0	5000000	Free	0.0	Everyone	Art & Design	August 1, 2018	
4	Pixel Draw - Number Art	ART_AND_DESIGN	4.3	6.875232	2800.0	100000	Free	0.0	Everyone	Art & Design;Creativity	June 20, 2018	

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	C
5	Paper flowers instructions	ART_AND_DESIGN	4.4	5.123964	5600.0	50000	Free	0.0	Everyone	Art & Design	March 26, 2017	

In [48]:

```
# 8-2 Drop columns App, Last Updated, Current Ver, and Android Ver. These variables are n
ot useful for our task
inp1.drop(["Last Updated","Current Ver","Android Ver","App","Type"],axis=1,inplace=True)
```

In [49]:

```
inp1.head()
```

Out[49]:

	Category	Rating	Reviews	Size	Installs	Price	Content Rating	Genres
0	ART_AND_DESIGN	4.1	5.075174	19000.0	10000	0.0	Everyone	Art & Design
1	ART_AND_DESIGN	3.9	6.875232	14000.0	500000	0.0	Everyone	Art & Design;Pretend Play
2	ART_AND_DESIGN	4.7	11.379520	8700.0	5000000	0.0	Everyone	Art & Design
4	ART_AND_DESIGN	4.3	6.875232	2800.0	100000	0.0	Everyone	Art & Design;Creativity
5	ART_AND_DESIGN	4.4	5.123964	5600.0	50000	0.0	Everyone	Art & Design

In [50]:

```
inp1.shape
```

Out[50]:

```
(8496, 8)
```

In [51]:

```
# 8 -3 Get dummy columns for Category, Genres, and Content Rating. This needs to be done
as the models do not understand categorical data, and all data should be numeric.
#Dummy encoding is one way to convert character fields to numeric. Name of dataframe shou
ld be inp2.
inp2 = inp1
inp2.head()
```

Out[51]:

	Category	Rating	Reviews	Size	Installs	Price	Content Rating	Genres
0	ART_AND_DESIGN	4.1	5.075174	19000.0	10000	0.0	Everyone	Art & Design
1	ART_AND_DESIGN	3.9	6.875232	14000.0	500000	0.0	Everyone	Art & Design;Pretend Play
2	ART_AND_DESIGN	4.7	11.379520	8700.0	5000000	0.0	Everyone	Art & Design
4	ART_AND_DESIGN	4.3	6.875232	2800.0	100000	0.0	Everyone	Art & Design;Creativity
5	ART_AND_DESIGN	4.4	5.123964	5600.0	50000	0.0	Everyone	Art & Design

Applying Dummy EnCoding on Column "Category"

In [52]:

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

Out[52]:

```
array(['ART_AND_DESIGN', 'AUTO_AND_VEHICLES', 'BEAUTY',
       'BOOKS_AND_REFERENCE', 'BUSINESS', 'COMICS', 'COMMUNICATION',
       'DATING', 'EDUCATION', 'ENTERTAINMENT', 'EVENTS', 'FINANCE',
```

```
       '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 [53]:

```
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[53]:

	Rating	Reviews	Size	Installs	Price	Content Rating	Genres	Category_ART_AND_DESIGN	Category_AUTO_AND_
0	4.1	5.075174	19000.0	10000	0.0	Everyone	Art & Design	1	
1	3.9	6.875232	14000.0	500000	0.0	Everyone	Art & Design;Pretend Play	1	
2	4.7	11.379520	8700.0	5000000	0.0	Everyone	Art & Design	1	
4	4.3	6.875232	2800.0	100000	0.0	Everyone	Art & Design;Creativity	1	
5	4.4	5.123964	5600.0	50000	0.0	Everyone	Art & Design	1	

5 rows × 40 columns

Applying Dummy EnCoding on Column "Genres"

In [54]:

```
inp2["Genres"].unique()
```

Out[54]:

```
array(['Art & Design', 'Art & Design;Pretend Play',
       'Art & Design;Creativity', 'Auto & Vehicles', 'Beauty',
       'Books & Reference', 'Business', 'Comics', 'Comics;Creativity',
       'Communication', 'Dating', 'Education', 'Education;Creativity',
       'Education;Education', 'Education;Music & Video',
       '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', 'Lifestyle;Pretend Play', 'Card', 'Casual', 'Puzzle',
       'Action', 'Arcade', 'Word', 'Racing', 'Casual;Creativity',
       'Sports', 'Board', 'Simulation', 'Role Playing', 'Adventure',
       'Strategy', 'Simulation;Education', 'Action;Action & Adventure',
       'Trivia', 'Casual;Brain Games', 'Simulation;Action & Adventure',
       'Educational;Creativity', 'Puzzle;Brain Games',
       'Educational;Education', 'Card;Brain Games',
       'Educational;Brain Games', 'Educational;Pretend Play',
       'Casual;Action & Adventure', 'Entertainment;Education',
       'Casual;Education', 'Casual;Pretend Play', 'Music;Music & Video',
       'Racing;Action & Adventure', 'Arcade;Pretend Play',
       'Adventure;Action & Adventure', 'Role Playing;Action & Adventure',
       'Simulation;Pretend Play', 'Puzzle;Creativity',
       'Sports;Action & Adventure', 'Educational;Action & Adventure',
```

```
       'Arcade;Action & Adventure', 'Entertainment;Action & Adventure',
       'Puzzle;Action & Adventure', 'Strategy;Action & Adventure',
       'Music & Audio;Music & Video', 'Health & Fitness;Education',
       'Adventure;Education', 'Board;Brain Games',
       'Board;Action & Adventure', 'Board;Pretend Play',
       'Casual;Music & Video', 'Role Playing;Pretend Play',
       'Entertainment;Pretend Play', 'Video Players & Editors;Creativity',
       'Card;Action & Adventure', 'Medical', 'Social', 'Shopping',
       'Photography', 'Travel & Local',
       'Travel & Local;Action & Adventure', 'Tools', 'Tools;Education',
       'Personalization', 'Productivity', 'Parenting',
       'Parenting;Music & Video', 'Parenting;Brain Games',
       'Parenting;Education', 'Weather', 'Video Players & Editors',
       'Video Players & Editors;Music & Video', 'News & Magazines',
       'Maps & Navigation', 'Health & Fitness;Action & Adventure',
       'Music', 'Educational', 'Casino', 'Adventure;Brain Games',
       'Lifestyle;Education', 'Books & Reference;Education',
       'Puzzle;Education', 'Role Playing;Brain Games',
       'Strategy;Education', 'Racing;Pretend Play',
       'Communication;Creativity', 'Strategy;Creativity'], dtype=object)
```

In [55]:

```
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]
inp2["Genres"].unique()
```

Out[55]:

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

In [56]:

```
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 [57]:

```
inp2.head()
```

Out[57]:

	Rating	Reviews	Size	Installs	Price	Content Rating	Category_ART_AND_DESIGN	Category_AUTO_AND_VEHICLES	Categ
0	4.1	5.075174	19000.0	10000	0.0	Everyone	1	0	
1	3.9	6.875232	14000.0	500000	0.0	Everyone	1	0	
2	4.7	11.379520	8700.0	5000000	0.0	Everyone	1	0	
4	4.3	6.875232	2800.0	100000	0.0	Everyone	1	0	
5	4.4	5.123964	5600.0	50000	0.0	Everyone	1	0	

Rating	Reviews	Size	Installs	Price	Content Rating	Category_ART_AND_DESIGN	Category_AUTO_AND_VEHICLES	Categ

5 rows × 91 columns

In [58]:

```
inp2.shape
```

Out[58]:

```
(8496, 91)
```

Applying Dummy EnCoding on Column "Content Rating"

In [59]:

```
inp2["Content Rating"].unique()
```

Out[59]:

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

In [60]:

```
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[60]:

	Rating	Reviews	Size	Installs	Price	Category_ART_AND_DESIGN	Category_AUTO_AND_VEHICLES	Category_BEAUT
0	4.1	5.075174	19000.0	10000	0.0	1	0	
1	3.9	6.875232	14000.0	500000	0.0	1	0	
2	4.7	11.379520	8700.0	5000000	0.0	1	0	
4	4.3	6.875232	2800.0	100000	0.0	1	0	
5	4.4	5.123964	5600.0	50000	0.0	1	0	

5 rows × 96 columns

In [61]:

```
inp2.shape
```

Out[61]:

```
(8496, 96)
```

In [62]:

```
# 9  Train test split  and apply 70-30 split. Name the new dataframes df_train and df_tes
t.
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 [63]:

```
# 10 Separating the dataframes into X_train, y_train, X_test, and y_test
```

```
# To separating the dataframes into X_train, y_train, X_test, and y_test.
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)
Xtest
```

Out[63]:

	Reviews	Size	Installs	Price	Category_ART_AND_DESIGN	Category_AUTO_AND_VEHICLES	Category_BEAUTY
9274	10.518187	4000.0	1000000	0.00	0	0	0
7750	12.828137	21000.0	10000000	0.00	0	0	0
585	9.066701	9500.0	1000000	0.00	0	0	0
4811	7.859027	26000.0	100000	0.00	0	0	0
2276	4.532599	32000.0	1000	79.99	0	0	0
...	...	...	...	...	...	...	...
9079	7.978654	22000.0	100000	0.00	0	0	0
4034	3.850148	26000.0	10000	0.00	0	0	0
5771	12.234811	63000.0	10000000	0.00	0	0	0
8034	8.190632	17000.0	100000	0.00	0	0	0
7860	12.986809	10000.0	10000000	0.00	0	0	0

2549 rows × 95 columns

In [64]:

```
ytest
```

Out[64]:

```
9274    4.4
7750    4.4
585     3.9
4811    3.9
2276    4.6
       ...
9079    4.5
4034    4.2
5771    4.3
8034    4.2
7860    4.6
Name: Rating, Length: 2549, dtype: float64
```

In [65]:

```
Xtrain
```

Out[65]:

	Reviews	Size	Installs	Price	Category_ART_AND_DESIGN	Category_AUTO_AND_VEHICLES	Category_BEAUTY
8571	8.677269	4400.0	100000	0.00	0	0	0
733	11.412906	0.0	5000000	0.00	0	0	0
7846	11.830208	51000.0	5000000	0.00	0	0	0
1755	9.147933	33000.0	100000	4.99	0	0	0
2021	7.686621	99000.0	500000	0.00	0	0	0

	...	...	...	...	...	...	...
8121	2.397895	3000.0	5000	0.00	0	0	0
	Reviews	Size	Installs	Price	Category_ART_AND_DESIGN	Category_AUTO_AND_VEHICLES	Category_BEAUTY
3573	10.542258	9100.0	1000000	0.00	0	0	0
4816	12.381957	35000.0	10000000	0.00	0	0	0
2595	8.881003	7300.0	500000	0.00	0	0	0
3425	8.832150	11000.0	100000	0.00	0	0	0

5947 rows × 95 columns

In [66]:

```
ytrain
```

Out[66]:

```
8571    4.7
733     4.4
7846    4.6
1755    4.6
2021    4.2
       ...
8121    3.5
3573    4.3
4816    4.4
2595    4.1
3425    4.1
Name: Rating, Length: 5947, dtype: float64
```

In [67]:

```
# 11 -Model building (Use linear regression as the technique)
reg_all = LR()
reg_all.fit(Xtrain,ytrain)
```

Out[67]:

```
LinearRegression()
```

In [68]:

```
#11 -Report the R2 on the train set
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 : 0.074
```

In [69]:

```
#12 -Make predictions on test set and report R2.
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 : 0.063
```

In [3]:

```
!wget -nc https://raw.githubusercontent.com/brpy/colab-pdf/master/colab_pdf.py
from colab_pdf import colab_pdf
colab_pdf('pandas-assignment.ipynb')
```

```
File ‘colab_pdf.py’ already there; not retrieving.
```

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
<ipython-input-3-a77988038c96> in <module>
      1 get_ipython().system('wget -nc https://raw.githubusercontent.com/brpy/colab-pdf/master/colab_pdf.py')
      2 from colab_pdf import colab_pdf
```

```
----> 3 colab_pdf('pandas-assignment.ipynb')

/content/colab_pdf.py in colab_pdf(file_name, notebookpath)
     20     # Check if the notebook exists in the Drive.
     21     if not os.path.isfile(os.path.join(notebookpath, file_name)):
---> 22         raise ValueError(f"file '{file_name}' not found in path '{notebookpath}'.
")
     23
     24     # Installing all the recommended packages.

ValueError: file 'pandas-assignment.ipynb' not found in path '/content/drive/MyDrive/Cola
b Notebooks/'.
```