GOOGLE PLAY STORE project

July 21, 2022

[1]:
```python
#import all required libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

[2]:
```python
data=pd.read_csv('googleplaystore.csv')
```

[3]:
```python
data
```

[3]:

	App	Category \
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN
1	Coloring book moana	ART_AND_DESIGN
2	U Launcher Lite – FREE Live Cool Themes, Hide …	ART_AND_DESIGN
3	Sketch – Draw & Paint	ART_AND_DESIGN
4	Pixel Draw – Number Art Coloring Book	ART_AND_DESIGN
…	…	…
10836	Sya9a Maroc – FR	FAMILY
10837	Fr. Mike Schmitz Audio Teachings	FAMILY
10838	Parkinson Exercices FR	MEDICAL
10839	The SCP Foundation DB fr nn5n	BOOKS_AND_REFERENCE
10840	iHoroscope - 2018 Daily Horoscope & Astrology	LIFESTYLE

	Rating	Reviews	Size	Installs	Type	Price \
0	4.1	159	19M	10,000+	Free	0
1	3.9	967	14M	500,000+	Free	0
2	4.7	87510	8.7M	5,000,000+	Free	0
3	4.5	215644	25M	50,000,000+	Free	0
4	4.3	967	2.8M	100,000+	Free	0
…	…	…	…	…	…	…
10836	4.5	38	53M	5,000+	Free	0
10837	5.0	4	3.6M	100+	Free	0
10838	NaN	3	9.5M	1,000+	Free	0
10839	4.5	114	Varies with device	1,000+	Free	0
10840	4.5	398307	19M	10,000,000+	Free	0

```
        Content Rating                       Genres       Last Updated  \
0            Everyone                   Art & Design    January 7, 2018
1            Everyone   Art & Design;Pretend Play   January 15, 2018
2            Everyone                   Art & Design     August 1, 2018
3                Teen                   Art & Design       June 8, 2018
4            Everyone     Art & Design;Creativity      June 20, 2018
...               ...                           ...                ...
10836        Everyone                     Education      July 25, 2017
10837        Everyone                     Education       July 6, 2018
10838        Everyone                       Medical   January 20, 2017
10839       Mature 17+          Books & Reference   January 19, 2015
10840        Everyone                     Lifestyle      July 25, 2018

              Current Ver          Android Ver
0                   1.0.0          4.0.3 and up
1                   2.0.0          4.0.3 and up
2                   1.2.4          4.0.3 and up
3        Varies with device          4.2 and up
4                     1.1            4.4 and up
...                   ...                   ...
10836                1.48            4.1 and up
10837                 1.0            4.1 and up
10838                 1.0            2.2 and up
10839   Varies with device   Varies with device
10840   Varies with device   Varies with device

[10841 rows x 13 columns]
```

```
[4]: #Check for the null values
     data.isnull().sum()
```

```
[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
```

```
[5]: #Drop records with null values in any of the columns
     data.dropna(inplace=True)
```

```
[6]: data.isnull().sum()
```

```
[6]: App               0
     Category          0
     Rating            0
     Reviews           0
     Size              0
     Installs          0
     Type              0
     Price             0
     Content Rating    0
     Genres            0
     Last Updated      0
     Current Ver       0
     Android Ver       0
     dtype: int64
```

```
[7]: data.shape
```

```
[7]: (9360, 13)
```

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

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

Earlier there were 10841 rows now after dropping the null values it is 9630 rows

```
[9]: #Size column- Convert Mb to Kb
     #Convert to numeric
     data["Size"] = [ float(i.split('M')[0]) if 'M' in i else float(0) for i in↵
      ↪data["Size"]  ]
```

```
[11]: data.head()
```

```
[11]:                                                App         Category  Rating  \
      0        Photo Editor & Candy Camera & Grid & ScrapBook  ART_AND_DESIGN     4.1
      1                                   Coloring book moana  ART_AND_DESIGN     3.9
      2    U Launcher Lite – FREE Live Cool Themes, Hide …  ART_AND_DESIGN     4.7
      3                                  Sketch – Draw & Paint  ART_AND_DESIGN     4.5
      4              Pixel Draw – Number Art Coloring Book  ART_AND_DESIGN     4.3

         Reviews  Size      Installs  Type Price Content Rating  \
      0      159  19.0       10,000+  Free     0        Everyone
      1      967  14.0      500,000+  Free     0        Everyone
      2    87510   8.7    5,000,000+  Free     0        Everyone
      3   215644  25.0   50,000,000+  Free     0            Teen
      4      967   2.8      100,000+  Free     0        Everyone

                          Genres      Last Updated        Current Ver  \
      0              Art & Design   January 7, 2018              1.0.0
      1    Art & Design;Pretend Play  January 15, 2018              2.0.0
      2              Art & Design    August 1, 2018              1.2.4
      3              Art & Design      June 8, 2018  Varies with device
      4      Art & Design;Creativity    June 20, 2018                1.1

            Android Ver
      0   4.0.3 and up
      1   4.0.3 and up
      2   4.0.3 and up
      3     4.2 and up
      4     4.4 and up
```

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

```
[13]: data
```

```
[13]:                                                App                 Category  \
      0        Photo Editor & Candy Camera & Grid & ScrapBook           ART_AND_DESIGN
      1                                   Coloring book moana           ART_AND_DESIGN
      2        U Launcher Lite – FREE Live Cool Themes, Hide …         ART_AND_DESIGN
      3                                  Sketch – Draw & Paint           ART_AND_DESIGN
      4              Pixel Draw – Number Art Coloring Book           ART_AND_DESIGN
      …                                                 …                        …
      10834                                       FR Calculator                   FAMILY
      10836                                   Sya9a Maroc – FR                   FAMILY
      10837                    Fr. Mike Schmitz Audio Teachings                   FAMILY
      10839                      The SCP Foundation DB fr nn5n   BOOKS_AND_REFERENCE
      10840      iHoroscope – 2018 Daily Horoscope & Astrology                LIFESTYLE
```

```
        Rating Reviews     Size      Installs  Type Price Content Rating  \
0         4.1     159  19000.0      10,000+  Free     0        Everyone
1         3.9     967  14000.0     500,000+  Free     0        Everyone
2         4.7   87510   8700.0   5,000,000+  Free     0        Everyone
3         4.5  215644  25000.0  50,000,000+  Free     0            Teen
4         4.3     967   2800.0     100,000+  Free     0        Everyone
...       ...     ...      ...          ...   ...   ...             ...
10834     4.0       7   2600.0         500+  Free     0        Everyone
10836     4.5      38  53000.0       5,000+  Free     0        Everyone
10837     5.0       4   3600.0         100+  Free     0        Everyone
10839     4.5     114      0.0       1,000+  Free     0       Mature 17+
10840     4.5  398307  19000.0  10,000,000+  Free     0        Everyone

                         Genres     Last Updated         Current Ver  \
0                  Art & Design   January 7, 2018               1.0.0
1       Art & Design;Pretend Play  January 15, 2018               2.0.0
2                  Art & Design    August 1, 2018               1.2.4
3                  Art & Design     June 8, 2018  Varies with device
4           Art & Design;Creativity   June 20, 2018                 1.1
...                         ...              ...                 ...
10834                 Education     June 18, 2017               1.0.0
10836                 Education     July 25, 2017                1.48
10837                 Education      July 6, 2018                 1.0
10839         Books & Reference  January 19, 2015  Varies with device
10840                 Lifestyle     July 25, 2018  Varies with device

              Android Ver
0           4.0.3 and up
1           4.0.3 and up
2           4.0.3 and up
3             4.2 and up
4             4.4 and up
...                  ...
10834         4.1 and up
10836         4.1 and up
10837         4.1 and up
10839  Varies with device
10840  Varies with device

[9360 rows x 13 columns]
```

[14]: `data.isnull().sum()`

[14]:
```
App              0
Category         0
Rating           0
Reviews          0
```

```
Size              0
Installs          0
Type              0
Price             0
Content Rating    0
Genres            0
Last Updated      0
Current Ver       0
Android Ver       0
dtype: int64
```

[15]: `data.head()`

[15]:
```
                                                App        Category  Rating  \
0       Photo Editor & Candy Camera & Grid & ScrapBook  ART_AND_DESIGN     4.1
1                                  Coloring book moana  ART_AND_DESIGN     3.9
2  U Launcher Lite – FREE Live Cool Themes, Hide …  ART_AND_DESIGN     4.7
3                              Sketch – Draw & Paint  ART_AND_DESIGN     4.5
4                  Pixel Draw – Number Art Coloring Book  ART_AND_DESIGN     4.3

   Reviews     Size      Installs  Type Price Content Rating  \
0      159  19000.0       10,000+  Free     0       Everyone
1      967  14000.0      500,000+  Free     0       Everyone
2    87510   8700.0    5,000,000+  Free     0       Everyone
3   215644  25000.0   50,000,000+  Free     0           Teen
4      967   2800.0      100,000+  Free     0       Everyone

                      Genres      Last Updated       Current Ver  \
0             Art & Design   January 7, 2018             1.0.0
1  Art & Design;Pretend Play  January 15, 2018             2.0.0
2             Art & Design    August 1, 2018             1.2.4
3             Art & Design     June 8, 2018  Varies with device
4   Art & Design;Creativity    June 20, 2018               1.1

      Android Ver
0  4.0.3 and up
1  4.0.3 and up
2  4.0.3 and up
3    4.2 and up
4    4.4 and up
```

[16]:
```python
#Convert Reviews column to numeric (int/float)
data['Reviews']=data['Reviews'].astype('int')
```

[17]: `data.dtypes`

```
[17]: App                 object
      Category            object
      Rating             float64
      Reviews              int64
      Size               float64
      Installs            object
      Type                object
      Price               object
      Content Rating      object
      Genres              object
      Last Updated        object
      Current Ver         object
      Android Ver         object
      dtype: object
```
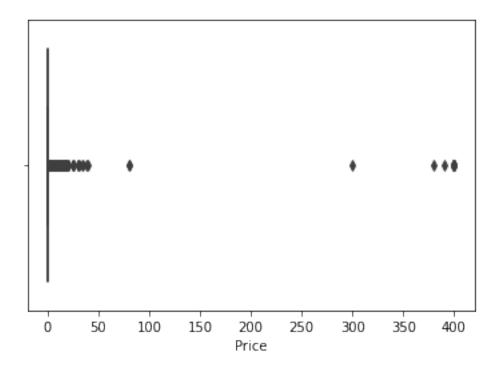
```
[18]: #Remove + and , from Installs field, Convert it to integer
      data['Installs']=data['Installs'].str.replace('[+,]','')
```

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

```
[19]:                                                   App      Category  Rating  \
      0       Photo Editor & Candy Camera & Grid & ScrapBook  ART_AND_DESIGN     4.1
      1                                  Coloring book moana  ART_AND_DESIGN     3.9
      2  U Launcher Lite - FREE Live Cool Themes, Hide …  ART_AND_DESIGN     4.7
      3                              Sketch - Draw & Paint  ART_AND_DESIGN     4.5
      4                  Pixel Draw - Number Art Coloring Book  ART_AND_DESIGN     4.3

         Reviews     Size  Installs  Type Price Content Rating  \
      0      159  19000.0     10000  Free     0       Everyone
      1      967  14000.0    500000  Free     0       Everyone
      2    87510   8700.0   5000000  Free     0       Everyone
      3   215644  25000.0  50000000  Free     0           Teen
      4      967   2800.0    100000  Free     0       Everyone

                           Genres      Last Updated         Current Ver  \
      0              Art & Design   January 7, 2018               1.0.0
      1  Art & Design;Pretend Play  January 15, 2018               2.0.0
      2              Art & Design    August 1, 2018               1.2.4
      3              Art & Design     June 8, 2018  Varies with device
      4    Art & Design;Creativity    June 20, 2018                 1.1

           Android Ver
      0  4.0.3 and up
      1  4.0.3 and up
      2  4.0.3 and up
      3    4.2 and up
      4    4.4 and up
```

```
[20]: data['Installs']=data['Installs'].astype('int')
```

```
[21]: data.dtypes
```

```
[21]: App               object
      Category          object
      Rating            float64
      Reviews           int64
      Size              float64
      Installs          int64
      Type              object
      Price             object
      Content Rating    object
      Genres            object
      Last Updated      object
      Current Ver       object
      Android Ver       object
      dtype: object
```

```
[22]: #Price field - Remove $ and convert it to numeric (int/float)
      data['Price']=data['Price'].str.replace('[$]','')
```

```
[23]: data.head()
```

```
[23]:                                                    App       Category  Rating  \
      0        Photo Editor & Candy Camera & Grid & ScrapBook  ART_AND_DESIGN     4.1
      1                                   Coloring book moana  ART_AND_DESIGN     3.9
      2  U Launcher Lite - FREE Live Cool Themes, Hide …  ART_AND_DESIGN     4.7
      3                              Sketch - Draw & Paint  ART_AND_DESIGN     4.5
      4              Pixel Draw - Number Art Coloring Book  ART_AND_DESIGN     4.3

         Reviews      Size   Installs  Type Price Content Rating  \
      0      159   19000.0      10000  Free     0       Everyone
      1      967   14000.0     500000  Free     0       Everyone
      2    87510    8700.0    5000000  Free     0       Everyone
      3   215644   25000.0   50000000  Free     0           Teen
      4      967    2800.0     100000  Free     0       Everyone

                           Genres     Last Updated        Current Ver  \
      0              Art & Design   January 7, 2018              1.0.0
      1  Art & Design;Pretend Play  January 15, 2018              2.0.0
      2              Art & Design     August 1, 2018              1.2.4
      3              Art & Design       June 8, 2018  Varies with device
      4    Art & Design;Creativity     June 20, 2018                1.1

            Android Ver
      0   4.0.3 and up
```

```
1    4.0.3 and up
2    4.0.3 and up
3      4.2 and up
4      4.4 and up
```

[24]: ```python
data['Price']=data['Price'].astype('float')
```

[25]: ```python
data.dtypes
```

[25]:
```
App                object
Category           object
Rating             float64
Reviews              int64
Size               float64
Installs             int64
Type               object
Price              float64
Content Rating     object
Genres             object
Last Updated       object
Current Ver        object
Android Ver        object
dtype: object
```

[26]: ```python
#Average rating should be within the range 1 to 5
#Drop records with rating value outside the range
data.shape
```

[26]: (9360, 13)

[27]: ```python
data.drop(data[(data['Rating']<1) & (data['Rating']>5)].index,inplace=True)
```

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

[28]: (9360, 13)

[29]: ```python
#Reviews should not be more than Installs.
#Drop them if there are any such records
data.drop(data.loc[data['Reviews']>data['Installs']].index,inplace=True)
```

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

[30]: (9353, 13)

[31]: ```python
#For free apps (Type=Free),the price should not be>0
#Drop any such records
```
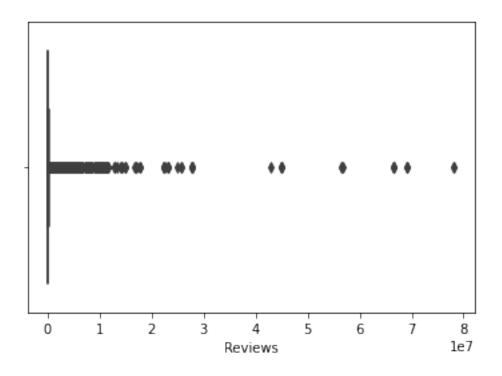
```
data.drop(data[(data['Type'] =='Free') & (data['Price'] > 0 )].index, inplace =␣
 ↪True)
```

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

[32]: (9353, 13)

[33]:
```
data.isnull().sum()
```

[33]:
```
App               0
Category          0
Rating            0
Reviews           0
Size              0
Installs          0
Type              0
Price             0
Content Rating    0
Genres            0
Last Updated      0
Current Ver       0
Android Ver       0
dtype: int64
```

Performing univarient analysis

[39]:
```
#boxplot for price
import warnings
warnings.filterwarnings("ignore")
sns.boxplot(data['Price'])
plt.figure(figsize=(12,8))
plt.show()
```

<Figure size 864x576 with 0 Axes>

There are outliers present where the price is within the range of 0−50 and at the range of 400$

```
[38]: #Boxplot for reviews
      sns.boxplot(data['Reviews'])
      plt.figure(figsize=(12,8))
      plt.show()
```
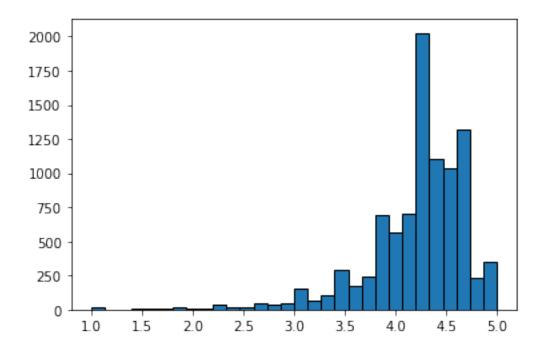
```
<Figure size 864x576 with 0 Axes>
```
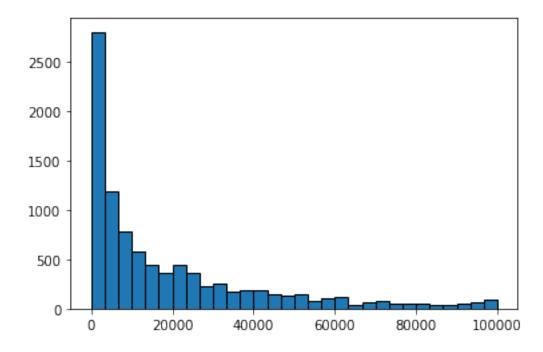
There are apps with very high reviews which is quiet not right

```
[40]:  #Histogram for Ratings
       plt.hist(data['Rating'],bins=30,edgecolor='black')
       plt.show()
```

This histogram shows high towards the higher ratings

```
[42]:  #Histogram for Size
       plt.hist(data['Size'],bins=30,edgecolor='black')
       plt.show()
```
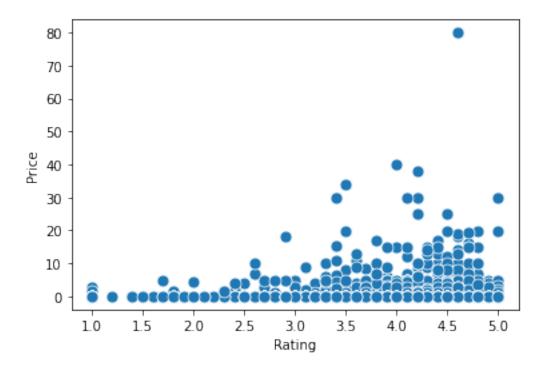
Outliers are present at the low size apps

Outliers Treatment

```
[43]:  #200$ per app is very high and suspicious
       #Check the apps above 200$ and drop the junk apps
       data.shape
```

```
[43]:  (9353, 13)
```

```
[46]:  data.drop(data.loc[data['Price']>200].index,inplace=True)
```

```
[47]:  data.shape
```

```
[47]:  (9338, 13)
```

```
[48]:  #Drop records more than 2 million reviews
       data.drop(data.loc[data['Reviews']>2000000].index,inplace=True)
```

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

```
[49]:  (8885, 13)
```

```
[50]:  #Installs:- Apps having very high no.of installs should be dropped
       #Different percentiles 10,25,50,70,90,95,99
       data.quantile([0.10,0.25,0.50,0.70,0.90,0.95,0.99])
```

```
[50]:        Rating      Reviews     Size      Installs  Price
      0.10     3.5        18.00      0.0        1000.0   0.00
      0.25     4.0       159.00   2600.0       10000.0   0.00
      0.50     4.3      4290.00   9500.0      500000.0   0.00
      0.70     4.5     35930.40  23000.0     1000000.0   0.00
      0.90     4.7    296771.00  50000.0    10000000.0   0.00
      0.95     4.8    637298.00  68000.0    10000000.0   1.99
      0.99     5.0   1462800.88  95000.0   100000000.0   7.49
```

```
[51]:  #The cut off range is install more than 10000000
       data.drop(data.loc[data['Installs']>10000000].index,inplace=True)
```

```
[52]:  data.shape
```

```
[52]:  (8496, 13)
```

Bivariate Analysis

```
[54]:  #Scatter Plot for Rating Vs Price
       sns.scatterplot(x='Rating',y='Price',data=data,s=80)
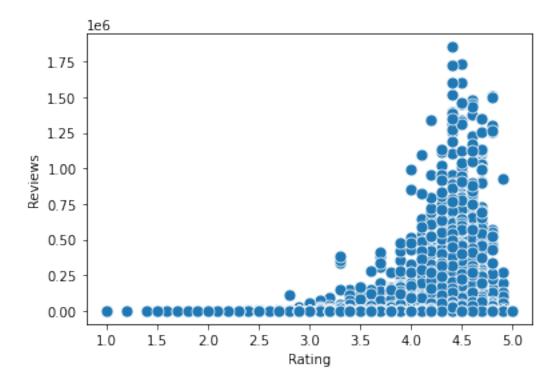```

[54]: <AxesSubplot:xlabel='Rating', ylabel='Price'>



Paid apps have higher ratings than free apps

[55]: ```
#Scatter Plot for Rating Vs Reviews
sns.scatterplot(x='Rating',y='Reviews',data=data,s=80)
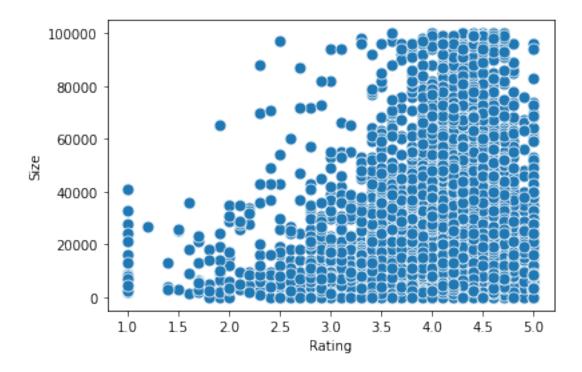```

[55]: <AxesSubplot:xlabel='Rating', ylabel='Reviews'>

Higher reviews have the high ratings.

```
[57]: #Scatter Plot for Rating Vs Size
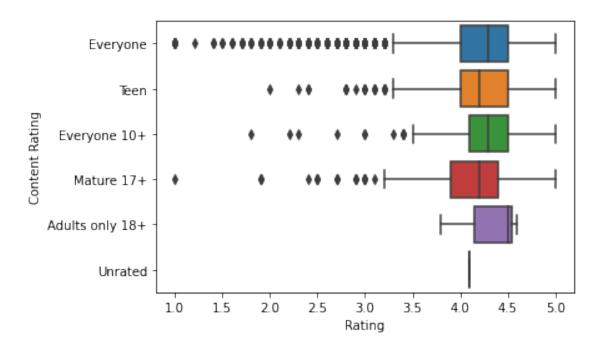      sns.scatterplot(x='Rating',y='Size',data=data,s=80)
```

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

Heavy apps are rated better

```
[58]:  #Boxplot for Rating Vs Content Rating
       sns.boxplot('Rating','Content Rating',data=data)
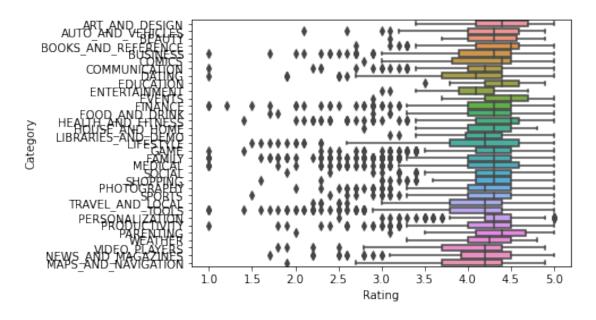       plt.figure(figsize=[30,25])
```

[58]: <Figure size 2160x1800 with 0 Axes>

<Figure size 2160x1800 with 0 Axes>

Apps for everyone have more outliers, 18+ apps have better rating in comparision to other apps

[59]: #Boxplot for Rating Vs Category
sns.boxplot('Rating','Category',data=data)

[59]: <AxesSubplot:xlabel='Rating', ylabel='Category'>

Apps for events category have better ratings

Data Preprocessing

```
[60]: #Apply log transformation (np.log 1p)to Reviews and Installs
      inp1=data
```

```
[61]: inp1.head()
```

```
[61]:                                                 App       Category  Rating  \
      0     Photo Editor & Candy Camera & Grid & ScrapBook  ART_AND_DESIGN     4.1
      1                                 Coloring book moana  ART_AND_DESIGN     3.9
      2  U Launcher Lite - FREE Live Cool Themes, Hide …  ART_AND_DESIGN     4.7
      4             Pixel Draw - Number Art Coloring Book  ART_AND_DESIGN     4.3
      5                          Paper flowers instructions  ART_AND_DESIGN     4.4

         Reviews     Size  Installs  Type  Price Content Rating  \
      0      159  19000.0     10000  Free    0.0       Everyone
      1      967  14000.0    500000  Free    0.0       Everyone
      2    87510   8700.0   5000000  Free    0.0       Everyone
      4      967   2800.0    100000  Free    0.0       Everyone
      5      167   5600.0     50000  Free    0.0       Everyone

                           Genres    Last Updated Current Ver    Android Ver
      0               Art & Design   January 7, 2018       1.0.0  4.0.3 and up
      1  Art & Design;Pretend Play  January 15, 2018       2.0.0  4.0.3 and up
      2               Art & Design    August 1, 2018       1.2.4  4.0.3 and up
      4     Art & Design;Creativity     June 20, 2018         1.1    4.4 and up
      5               Art & Design    March 26, 2017         1.0    2.3 and up
```

```
[62]: inp1.skew()
```

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

```
[63]: #skew for Reviews
      reviewskew=np.log1p(inp1['Reviews'])
      inp1['Reviews'] = reviewskew
```

```
[64]: reviewskew.skew()
```

```
[64]: -0.20039949659264134
```

```
[65]: #skew installs
      installsskew=np.log1p(inp1['Installs'])
      inp1['Installs']
```

```
[65]: 0              10000
      1             500000
      2            5000000
      4             100000
      5              50000
                    ...
      10834            500
      10836           5000
      10837            100
      10839           1000
      10840       10000000
      Name: Installs, Length: 8496, dtype: int64
```

```
[66]: installsskew.skew()
```

```
[66]: -0.5097286542754812
```

```
[67]: inp1.shape
```

```
[67]: (8496, 13)
```

```
[68]: inp1.head()
```

```
[68]:                                                  App        Category  Rating  \
      0        Photo Editor & Candy Camera & Grid & ScrapBook  ART_AND_DESIGN     4.1
      1                                   Coloring book moana  ART_AND_DESIGN     3.9
      2     U Launcher Lite – FREE Live Cool Themes, Hide …  ART_AND_DESIGN     4.7
      4                Pixel Draw – Number Art Coloring Book  ART_AND_DESIGN     4.3
      5                           Paper flowers instructions  ART_AND_DESIGN     4.4

            Reviews     Size  Installs  Type  Price Content Rating  \
      0     5.075174  19000.0     10000  Free    0.0        Everyone
      1     6.875232  14000.0    500000  Free    0.0        Everyone
      2    11.379520   8700.0   5000000  Free    0.0        Everyone
      4     6.875232   2800.0    100000  Free    0.0        Everyone
      5     5.123964   5600.0     50000  Free    0.0        Everyone

                          Genres     Last Updated Current Ver    Android Ver
      0             Art & Design   January 7, 2018       1.0.0  4.0.3 and up
      1   Art & Design;Pretend Play  January 15, 2018       2.0.0  4.0.3 and up
      2             Art & Design    August 1, 2018       1.2.4  4.0.3 and up
```

```
4      Art & Design;Creativity      June 20, 2018           1.1      4.4 and up
5               Art & Design      March 26, 2017           1.0      2.3 and up
```

[69]:
```python
#Drop the columns - Apps, Lastupdated, Current version, Android version
inp1.drop(["Last Updated","Current Ver","Android
 →Ver","App","Type"],axis=1,inplace=True)
```

[70]:
```python
inp1.head()
```

[70]:
```
        Category  Rating    Reviews     Size  Installs  Price Content Rating  \
0  ART_AND_DESIGN     4.1   5.075174  19000.0     10000    0.0       Everyone
1  ART_AND_DESIGN     3.9   6.875232  14000.0    500000    0.0       Everyone
2  ART_AND_DESIGN     4.7  11.379520   8700.0   5000000    0.0       Everyone
4  ART_AND_DESIGN     4.3   6.875232   2800.0    100000    0.0       Everyone
5  ART_AND_DESIGN     4.4   5.123964   5600.0     50000    0.0       Everyone

                    Genres
0              Art & Design
1   Art & Design;Pretend Play
2              Art & Design
4     Art & Design;Creativity
5              Art & Design
```

[73]:
```python
inp1.shape
```

[73]:
```
(8496, 8)
```

[74]:
```python
inp1.isnull().sum()
```

[74]:
```
Category          0
Rating            0
Reviews           0
Size              0
Installs          0
Price             0
Content Rating    0
Genres            0
dtype: int64
```

[75]:
```python
#Name the dataframe to inp2
inp2=inp1
```

[76]:
```python
inp2.head()
```

[76]:
```
        Category  Rating   Reviews     Size  Installs  Price Content Rating  \
0  ART_AND_DESIGN     4.1  5.075174  19000.0     10000    0.0       Everyone
1  ART_AND_DESIGN     3.9  6.875232  14000.0    500000    0.0       Everyone
```

```
2   ART_AND_DESIGN   4.7  11.379520   8700.0   5000000   0.0      Everyone
4   ART_AND_DESIGN   4.3   6.875232   2800.0    100000   0.0      Everyone
5   ART_AND_DESIGN   4.4   5.123964   5600.0     50000   0.0      Everyone

                       Genres
0               Art & Design
1   Art & Design;Pretend Play
2               Art & Design
4      Art & Design;Creativity
5               Art & Design
```

[77]: `inp2.shape`

[77]: `(8496, 8)`

[78]:
```python
#Dummy encoding for - Category, Genres, Content Rating
list1=['Category','Genres','Content Rating']
inp2=pd.get_dummies(inp2,columns=list1)
inp2.head()
```

[78]:
```
   Rating     Reviews      Size  Installs  Price  Category_ART_AND_DESIGN  \
0     4.1    5.075174  19000.0     10000    0.0                        1
1     3.9    6.875232  14000.0    500000    0.0                        1
2     4.7   11.379520   8700.0   5000000    0.0                        1
4     4.3    6.875232   2800.0    100000    0.0                        1
5     4.4    5.123964   5600.0     50000    0.0                        1

   Category_AUTO_AND_VEHICLES  Category_BEAUTY  Category_BOOKS_AND_REFERENCE  \
0                           0                0                             0
1                           0                0                             0
2                           0                0                             0
4                           0                0                             0
5                           0                0                             0

   Category_BUSINESS  …  Genres_Video Players & Editors;Creativity  \
0                  0  …                                          0
1                  0  …                                          0
2                  0  …                                          0
4                  0  …                                          0
5                  0  …                                          0

   Genres_Video Players & Editors;Music & Video  Genres_Weather  Genres_Word  \
0                                             0               0            0
1                                             0               0            0
2                                             0               0            0
4                                             0               0            0
5                                             0               0            0
```

```
      Content Rating_Adults only 18+   Content Rating_Everyone   \
0                                  0                          1
1                                  0                          1
2                                  0                          1
4                                  0                          1
5                                  0                          1

      Content Rating_Everyone 10+   Content Rating_Mature 17+   \
0                               0                            0
1                               0                            0
2                               0                            0
4                               0                            0
5                               0                            0

      Content Rating_Teen   Content Rating_Unrated
0                        0                        0
1                        0                        0
2                        0                        0
4                        0                        0
5                        0                        0

[5 rows x 159 columns]
```

[284]:
```python
#Train_test_split - Apply 70-30 split
#Name the new dataframe = df_train,df_test
from sklearn.model_selection import train_test_split
df_train,df_test=train_test_split(inp2,test_size=0.3,random_state=3)
```

[285]:
```python
df_train.shape
```

[285]: (5947, 159)

[286]:
```python
df_test.shape
```

[286]: (2549, 159)

[287]:
```python
#Separate dataframes into - X_train, y_train, X_test, y_test
y_train=df_train.pop("Rating")
X_train=df_train
```

[288]:
```python
y_test=df_test.pop("Rating")
X_test=df_test
```

[289]:
```python
X_train
```

```
[289]:          Reviews      Size   Installs  Price  Category_ART_AND_DESIGN  \
       4264    6.682109    6700.0      50000    0.0                        0
       7517    3.610918    3200.0        100    0.0                        0
       10315   4.077537    2700.0      10000    0.0                        0
       9259    2.079442    4200.0       1000    0.0                        0
       8129    4.110874   25000.0      10000    0.0                        0
       ...          ...       ...        ...                             ...
       9059    1.791759    1600.0        500    0.0                        0
       3290   11.621852       0.0   10000000    0.0                        0
       8100    5.209486    4900.0       5000    0.0                        0
       2055    9.784366   16000.0    1000000    0.0                        0
       7612    6.651572   55000.0     100000    0.0                        0

               Category_AUTO_AND_VEHICLES  Category_BEAUTY  \
       4264                             0                0
       7517                             0                0
       10315                            0                0
       9259                             0                0
       8129                             0                0
       ...                            ...              ...
       9059                             0                0
       3290                             0                0
       8100                             0                0
       2055                             0                0
       7612                             0                0

               Category_BOOKS_AND_REFERENCE  Category_BUSINESS  Category_COMICS  ...  \
       4264                               0                  0                0  ...
       7517                               0                  0                0  ...
       10315                              0                  0                0  ...
       9259                               0                  0                0  ...
       8129                               0                  0                0  ...
       ...                              ...                ...              ...  ...
       9059                               0                  0                0  ...
       3290                               0                  0                0  ...
       8100                               0                  0                0  ...
       2055                               0                  0                0  ...
       7612                               0                  0                0  ...

               Genres_Video Players & Editors;Creativity  \
       4264                                             0
       7517                                             0
       10315                                            0
       9259                                             0
       8129                                             0
       ...                                             ...
       9059                                             0
```

3290	0
8100	0
2055	0
7612	0

	Genres_Video Players & Editors;Music & Video	Genres_Weather \
4264	0	0
7517	0	0
10315	0	0
9259	0	0
8129	0	0
…	…	…
9059	0	0
3290	0	0
8100	0	0
2055	0	0
7612	0	0

	Genres_Word	Content Rating_Adults only 18+	Content Rating_Everyone \
4264	0	0	1
7517	0	0	0
10315	0	0	1
9259	0	0	1
8129	0	0	1
…	…	…	…
9059	0	0	1
3290	0	0	1
8100	0	0	1
2055	0	0	1
7612	0	0	1

	Content Rating_Everyone 10+	Content Rating_Mature 17+ \
4264	0	0
7517	0	0
10315	0	0
9259	0	0
8129	0	0
…	…	…
9059	0	0
3290	0	0
8100	0	0
2055	0	0
7612	0	0

	Content Rating_Teen	Content Rating_Unrated
4264	0	0
7517	1	0

```
10315                        0                        0
9259                         0                        0
8129                         0                        0
...                        ...                      ...
9059                         0                        0
3290                         0                        0
8100                         0                        0
2055                         0                        0
7612                         0                        0

[5947 rows x 158 columns]
```

[290]: `y_train`

```
[290]: 4264     4.1
       7517     5.0
       10315    4.5
       9259     4.1
       8129     3.3
                ...
       9059     4.0
       3290     4.3
       8100     4.2
       2055     4.4
       7612     4.5
       Name: Rating, Length: 5947, dtype: float64
```

[291]: `X_test`

```
[291]:          Reviews      Size   Installs   Price   Category_ART_AND_DESIGN  \
       9625   13.735117       0.0   10000000    0.00                         0
       5626    8.456806    1500.0    1000000    0.00                         0
       5005    8.427268   12000.0     100000    0.99                         0
       6934    2.890372       0.0       1000    0.00                         0
       8539    7.187657    4200.0     100000    0.00                         0
       ...          ...       ...        ...     ...                       ...
       8536    5.181784    1100.0      10000    0.00                         0
       2006   12.470941   51000.0   10000000    0.00                         0
       5720   13.261638   20000.0   10000000    0.00                         0
       10615   1.791759    9500.0        100    0.00                         0
       2633    8.210124   50000.0     100000    0.00                         0


              Category_AUTO_AND_VEHICLES   Category_BEAUTY  \
       9625                            0                 0
       5626                            0                 0
       5005                            0                 0
       6934                            0                 0
```

8539	0	0
…	…	…
8536	0	0
2006	0	0
5720	0	0
10615	0	0
2633	0	0

	Category_BOOKS_AND_REFERENCE	Category_BUSINESS	Category_COMICS	…
9625	1	0	0	…
5626	0	0	0	…
5005	0	0	0	…
6934	0	0	0	…
8539	0	0	0	…
…	…	…	… …	
8536	0	0	0	…
2006	0	0	0	…
5720	0	0	0	…
10615	0	0	0	…
2633	0	0	0	…

	Genres_Video Players & Editors;Creativity	\
9625	0	
5626	0	
5005	0	
6934	0	
8539	0	
…	…	
8536	0	
2006	0	
5720	0	
10615	0	
2633	0	

	Genres_Video Players & Editors;Music & Video	Genres_Weather	\
9625	0	0	
5626	0	0	
5005	0	0	
6934	0	0	
8539	0	0	
…	…	…	
8536	0	0	
2006	0	0	
5720	0	0	
10615	0	0	
2633	0	0	

```
         Genres_Word  Content Rating_Adults only 18+  Content Rating_Everyone  \
9625              0                              0                        1
5626              0                              0                        1
5005              0                              0                        1
6934              0                              0                        1
8539              0                              0                        1
...            ...                            ...                      ...
8536              0                              0                        1
2006              0                              0                        1
5720              0                              0                        1
10615             0                              0                        1
2633              0                              0                        0

         Content Rating_Everyone 10+  Content Rating_Mature 17+  \
9625                              0                          0
5626                              0                          0
5005                              0                          0
6934                              0                          0
8539                              0                          0
...                            ...                        ...
8536                              0                          0
2006                              0                          0
5720                              0                          0
10615                            0                          0
2633                             0                          1

         Content Rating_Teen  Content Rating_Unrated
9625                       0                       0
5626                       0                       0
5005                       0                       0
6934                       0                       0
8539                       0                       0
...                      ...                     ...
8536                       0                       0
2006                       0                       0
5720                       0                       0
10615                      0                       0
2633                       0                       0

[2549 rows x 158 columns]
```

[292]: `y_test`

[292]: 9625   4.9
       5626   4.4
       5005   4.2
       6934   4.2

```
8539    3.9
         …
8536    4.5
2006    4.6
5720    4.4
10615   3.2
2633    4.0
Name: Rating, Length: 2549, dtype: float64
```

Model Building

```
[293]:  #Using Linear Regression as technique
        #Report R2 on train set
        from sklearn.linear_model import LinearRegression
        lin_reg=LinearRegression()
```

```
[294]:  lin_reg.fit(X_train,y_train)
```

```
[294]:  LinearRegression()
```

```
[295]:  y_pred=lin_reg.predict(X_test)
```

```
[296]:  y_pred
```

```
[296]:  array([4.53258838, 3.99538896, 4.21270564, …, 4.33847685, 3.9695583 ,
               4.21528775])
```

```
[297]:  #Mean Square Error
        from sklearn.metrics import mean_squared_error,r2_score
```

```
[298]:  mse=mean_squared_error(y_pred,y_test)
```

```
[299]:  rmse=np.sqrt(mean_squared_error(y_test,y_pred))
        print(rmse)
```

```
0.5254494880832795
```

```
[302]:  #Reporting r2 for the model
        from sklearn.metrics import r2_score
        print(r2_score(y_pred,y_test))
```

```
-11.868176245645374
```

```
[301]:  from sklearn.metrics import r2_score
        print(r2_score(lin_reg.predict(X_train),y_train))
```

```
-10.849798204517723
```

[ ]: