

In [43]:

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

In [44]:

```
playstore_data=pd.read_csv("C:\Users\BABU\Desktop\project\googleplaystore.csv")
```

In [45]:

```
playstore_data
```

Out [45]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
0	Photo Editor & Candy Camera & Glit & 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 miana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone	Art & Design;Presnd 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	Pixol 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
...
10836	Syaña Marco - FR	FAMILY	4.5	38	53M	5,000+	Free	0	Everyone	Education	July 25, 2017	1.48	4.1 and up
10837	Pt. Mike Schmitz Audio Teachings	FAMILY	5.0	4	3.6M	1,000+	Free	0	Everyone	Education	July 6, 2018	1.0	4.1 and up
10838	Parkinson Exercises FR	MEDICAL	NaN	3	9.5M	1,000+	Free	0	Everyone	Medical	January 20, 2017	1.0	2.2 and up
10839	The SCP Foundation DB & m5n6	BOOKS_AND_REFERENCE	4.5	114	Varies with device	1,000+	Free	0	Mature 17+	Books & Reference	January 19, 2018	Varies with device	Varies with device
10840	Homescope - 2018 Daily Homescope & Astrology	LIFESTYLE	4.5	398307	19M	10,000,000+	Free	0	Everyone	Lifestyle	July 25, 2018	Varies with device	Varies with device

10841 rows × 13 columns

In [46]:

```
playstore_data.head()
```

Out [46]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
0	Photo Editor & Candy Camera & Glit & 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 miana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone	Art & Design;Presnd 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	Pixol 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 [5]:

```
playstore_data.shape # to know dimensions of your table
```

Out [5]:

```
(10841, 13)
```

In [6]:

```
playstore_data.describe() # to know statistics of your data
```

Out [6]:

	Rating
count	9367.000000
mean	4.193338
std	0.537431
min	1.000000
25%	4.000000
50%	4.300000
75%	4.500000
max	18.000000

In [7]:

```
playstore_data.boxplot()
```

Out [7]:

<AxesSubplot:~>

In [8]:

```
playstore_data.info() # to check how many null values are there
```

Out [8]:

```
<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          10839 non-null object
12   Android Ver          10838 non-null object
dtypes: float64(1), object(12)
memory usage: 1.1+ MB
```

In [9]:

```
playstore_data.isnull().sum() # no of null values in each column
```

Out [9]:

```
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 [10]:

```
playstore_data[playstore_data.Rating<5] # outlier observed from box plot
```

Out [10]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
10472	Life Made Wi-Fi Touchscreen Photo Frame	ART_AND_DESIGN	1.0	19.0	3.0M	1,000+	Free	0	Everyone	NaN	February 11, 2018	1.0.19	4.0 and up

In [11]:

```
playstore_data.drop([10472],inplace=True) # dropping the wrong rating row
```

In [12]:

```
playstore_data[10470:10475] # checking whether the row is deleted or not
```

Out [12]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
10470	Jazz Wi-Fi	COMMUNICATION	3.4	49	4.0M	10,000+	Free	0	Everyone	Communication	February 16, 2017	0.1	2.3 and up
10471	Xposed Wi-Fi Free WP	PERSONALIZATION	3.5	1042	40M	100,000+	Free	0	Everyone	Personalization	August 5, 2014	3.0.0	4.0.3 and up
10472	ommo Wi-Fi Free WP	TOOLS	4.2	134203	4.3M	10,000,000+	Free	0	Everyone	Tools	August 7, 2018	6.06.14	4.4 and up
10473	Sa-Fi Voice	COMMUNICATION	3.4	97	14M	1,000+	Free	0	Everyone	Communication	November 21, 2014	2.2.1.5	2.2 and up
10474	Wi-Fi Visualizer	TOOLS	3.9	132	2.8M	50,000+	Free	0	Everyone	Tools	May 17, 2017	0.0.0	2.3 and up

In [13]:

```
playstore_data.boxplot()
```

Out [13]:

<AxesSubplot:~>

In [14]:

```
playstore_data.isnull().sum()
```

Out [14]:

```
App                  0
Category             0
Rating              1474
Reviews              0
Size                 0
Installs             0
Type                 1
Price                0
Content Rating       0
Genres               0
Last Updated         0
Current Ver          8
Android Ver          2
dtype: int64
```

In [15]:

```
def impute_median(series):
    return series.fillna(series.median())
```

In [16]:

```
playstore_data.Rating=playstore_data['Rating'].transform(impute_median)
```

In [17]:

```
playstore_data.isnull().sum() # all the values as been filled in ratings column
```

Out [17]:

```
App                  0
Category             0
Rating              0
Reviews              0
Size                 0
Installs             0
Type                 1
Price                0
Content Rating       0
Genres               0
Last Updated         0
Current Ver          8
Android Ver          2
dtype: int64
```

In [18]:

```
print(playstore_data['Type'].mode())
print(playstore_data['Current Ver'].mode())
print(playstore_data['Android Ver'].mode())
```

Out [18]:

```
0    Free
dtype: object
0    Varies with device
dtype: object
0    4.1 and up
dtype: object
```

In [19]:

```
playstore_data['Type'].fillna(str(playstore_data['Type'].mode().values[0]),inplace=True)
playstore_data['Price'].fillna(str(playstore_data['Price'].mode().values[0]),inplace=True)
playstore_data['Android Ver'].fillna(str(playstore_data['Android Ver'].mode().values[0]),inplace=True)
```

In [20]:

```
playstore_data.isnull().sum()
```

Out [20]:

```
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
```

In [21]:

```
playstore_data['Price'] = playstore_data['Price'].apply(lambda x: str(x).replace('$','')if 's' in str(x) else str(x))
playstore_data['Reviews'] = pd.to_numeric(playstore_data['Reviews'],errors='coerce')
```

In [22]:

```
playstore_data['Installs']=playstore_data['Installs'].apply(lambda x: str(x).replace('+','') if '+' in str(x) else str(x))
playstore_data['Installs']=playstore_data['Installs'].apply(lambda x: str(x).replace(',','') if ',' in str(x) else str(x))
playstore_data['Installs']=playstore_data['Installs'].apply(lambda x: float(x))
```

In [23]:

```
playstore_data.describe()
```

Out [23]:

	Rating	Reviews	Installs	Price
count	10840.000000	1.084000e+04	1.084000e+04	10840.000000
mean	4.206476	4.441529e+05	1.546434e+07	1.027386
std	0.480342	2.927701e+06	8.502936e+07	15.949703
min	1.000000	0.000000e+00	0.000000e+00	0.000000
25%	4.100000	3.800000e+01	1.000000e+03	0.000000
50%	4.300000	2.094000e+03	1.000000e+05	0.000000
75%	4.500000	5.477500e+04	5.000000e+06	0.000000
max	5.000000	7.815801e+07	1.000000e+08	400.000000

In [26]:

```
gby=playstore_data.groupby('Category')
zigby['Rating'].agg(np.mean)
zigby['Price'].agg(np.sum)
zigby['Reviews'].agg(np.mean)
print(x)
print(y)
print(z)
```

Out [26]:

```
Category
ART_AND_DESIGN      4.255385
AUTO_AND_VEHICLES   4.295882
BEAUTY               4.283919
BOOKS_AND_REFERENCE 4.354988
BUSINESS             4.182391
COMICS               4.169809
COMMUNICATION        4.180103
DATING               4.628641
EDUCATION            4.384602
ENTERTAINMENT        4.126174
EVENTS               4.393123
FAMILY               4.204564
FINANCE              4.151939
FOOD_AND_DRINK       4.189827
GAME                 4.286888
HEALTH_AND_FITNESS   4.289959
HOUSE_AND_HOME       4.211364
LIBRARIES_AND_DEMO   4.207059
LIFESTYLE             4.131414
MAPS_AND_NAVIGATION  4.075182
MEDICAL              4.216199
NEWS_AND_MAGAZINES   4.161837
PARENTING            4.300009
PERSONALIZATION      4.218827
PHOTOGRAPHY          4.197918
PRODUCTIVITY         4.226651
SHOPPING             4.263977
SOCIAL               4.261917
SPORTS              4.236458
TOOLS                4.488071
TRAVEL_AND_LOCAL     4.132446
VIDEO_PLAYERS        4.884089
WEATHER              4.248788
Name: Rating, dtype: float64

Category
ART_AND_DESIGN      5.47
AUTO_AND_VEHICLES   13.47
BEAUTY               0.99
BOOKS_AND_REFERENCE 119.77
BUSINESS             185.27
COMICS               0.99
COMMUNICATION        83.14
DATING               31.43
EDUCATION            17.96
ENTERTAINMENT        7.98
EVENTS               189.39
FAMILY               2434.78
FINANCE              2998.83
FOOD_AND_DRINK       8.48
GAME                 287.30
HEALTH_AND_FITNESS   67.34
HOUSE_AND_HOME       0.90
LIBRARIES_AND_DEMO   0.99
LIFESTYLE             2360.87
MAPS_AND_NAVIGATION  26.95
MEDICAL              1429.86
NEWS_AND_MAGAZINES   3.98
PARENTING            9.58
PERSONALIZATION      153.86
PHOTOGRAPHY          134.21
PRODUCTIVITY         250.80
SHOPPING             9.48
SOCIAL               15.97
SPORTS               189.89
TOOLS                267.25
TRAVEL_AND_LOCAL     6.97
VIDEO_PLAYERS        18.46
WEATHER              32.42
Name: Price, dtype: float64

Category
ART_AND_DESIGN      2.637699e+04
AUTO_AND_VEHICLES   1.369919e+04
BEAUTY               7.476226e+03
BOOKS_AND_REFERENCE 5.569098e+04
BUSINESS             8.935998e+04
COMICS               9.594748e+04
COMMUNICATION        2.107138e+06
DATING               1.159313e+04
EDUCATION            2.538191e+05
ENTERTAINMENT        3.971888e+05
EVENTS               2.518908e+09
FAMILY               2.889255e+05
FINANCE              4.795281e+04
FOOD_AND_DRINK       5.594748e+04
GAME                 1.385959e+06
HEALTH_AND_FITNESS   1.112334e+09
HOUSE_AND_HOME       4.518619e+04
LIBRARIES_AND_DEMO   1.226139e+04
LIFESTYLE             3.172457e+04
MAPS_AND_NAVIGATION  2.237026e+05
MEDICAL              3.426432e+09
NEWS_AND_MAGAZINES   1.822292e+05
PARENTING            2.79218e+04
PERSONALIZATION      6.73931e+05
PHOTOGRAPHY          2.793238e+05
PRODUCTIVITY         4.424662e+05
SHOPPING             2.105903e+06
SOCIAL               1.544036e+05
SPORTS               2.40629e+05
TOOLS                3.740629e+05
TRAVEL_AND_LOCAL     2.47015e+05
VIDEO_PLAYERS        6.387439e+05
WEATHER              1.781865e+05
Name: Reviews, dtype: float64
```

In [42]:

```
[<matplotlib.lines.Line2D at 0x156d7a75108>]
```

Out [42]:

In [41]:

```
plt.figure(figsize=(16,5))
plt.plot(x,'r',color='b')
plt.xticks(rotation=90)
plt.title('Category wise Rating')
plt.xlabel('Categories')
plt.ylabel('Rating')
plt.show()
```

Out [41]:

```
C:\Users\BABU\AppData\Local\Temp\ipykernel_2828\241488624.py:2: UserWarning: color is redundantly defined by the 'color' keyword argument and the fmt string 'r' (-> color='r'). The keyword argument will take precedence.
  plt.plot(x,'r',color='b')
```

In [56]:

```
plt.figure(figsize=(16,5))
plt.plot(x,'r',color='r')
plt.xticks(rotation=90)
plt.title('Category wise Pricing')
plt.xlabel('Categories')
plt.ylabel('Prices')
plt.show()
```

Out [56]:

```
C:\Users\BABU\AppData\Local\Temp\ipykernel_2828\2754558929.py:2: UserWarning: color is redundantly defined by the 'color' keyword argument and the fmt string 'r' (-> color='r'). The keyword argument will take precedence.
  plt.plot(x,'r',color='r')
```

In [57]:

```
plt.figure(figsize=(16,5))
plt.plot(x,'r',color='r')
plt.xticks(rotation=90)
plt.title('Category wise Reviews')
plt.xlabel('Categories')
plt.ylabel('Reviews')
plt.show()
```

Out [57]:

```
C:\Users\BABU\AppData\Local\Temp\ipykernel_2828\1784558929.py:2: UserWarning: color is redundantly defined by the 'color' keyword argument and the fmt string 'r' (-> color='r'). The keyword argument will take precedence.
  plt.plot(x,'r',color='r')
```

In [63]:

```
plt.figure(figsize=(16,5))
plt.plot(x,'r',color='b')
plt.xticks(rotation=90)
plt.title('category wise price')
plt.xlabel('category names')
plt.ylabel('price')
plt.show()
```

Out [63]:

```
C:\Users\BABU\AppData\Local\Temp\ipykernel_2828\2595667487.py:2: UserWarning: color is redundantly defined by the 'color' keyword argument and the fmt string 'r' (-> color='r'). The keyword argument will take precedence.
  plt.plot(y,'r',color='b')
```

In [64]:

```
plt.figure(figsize=(16,5))
plt.plot(x,'r',color='g')
plt.xticks(rotation=90)
plt.title('Category wise Reviews')
plt.xlabel('Categories')
plt.ylabel('Reviews')
plt.show()
```

Out [64]:

```
C:\Users\BABU\AppData\Local\Temp\ipykernel_2828\2994671734.py:2: UserWarning: color is redundantly defined by the 'color' keyword argument and the fmt string 'r' (-> color='r'). The keyword argument will take precedence.
  plt.plot(x,'r',color='g')
```

In [59]:

```
playstore_data.groupby(['Content Rating']).count()
```

Out [59]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Genres	Last Updated	Current Ver	Android Ver
Content Rating												
Adults only 18+												
Everyone	8714	8714	7420	8714	8714	8714	8714	8714	8714	8714	8714	8712
Everyone 14+	414	414	397	414	414	414	414	414	414	414	414	414
Everyone 17+	499	499	461	499	499	499	499	499	499	499	499	499
Teen	1208	1208	1204	1208	1208	1208	1208	1208	1208	1208	1208	1208
Unrated	2	2	1	2	2	2	2	2	2	2	2	2

In [67]:

```
playstore_data.groupby('Type').count()
```

Out [67]:

	App	Category	Rating	Reviews	Size	Installs	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
Type	0	1	1	1	1	1	1	0	1	1	1	0
Free	10039	10039	8719	10039	10039	10039	10039	10039	10039	10039	10033	10038
Paid	800	800	647	800	800	800	800	800	800	798	799	799

In []: