In [1]:
```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
from sklearn.linear_model import LinearRegression
```

In [2]:
```
data = pd.read_csv("C:/Users/pavan/Desktop/python project/googleplaystore.csv")
data.head()
```

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0	Everyone	Art &
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone	Art & Desig Play
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8.7M	5,000,000+	Free	0	Everyone	Art &
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50,000,000+	Free	0	Teen	Art &
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100,000+	Free	0	Everyone	Art & Desig

In [3]:
```
dummy = data
```

In [4]:
```
dummy.head()
```

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0	Everyone	Art &
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone	Art & Desig Play
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8.7M	5,000,000+	Free	0	Everyone	Art &
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50,000,000+	Free	0	Teen	Art &
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100,000+	Free	0	Everyone	Art & Desig

In [5]:
```
dummy.describe()
```

	Rating	Reviews
count	9366.000000	1.084000e+04
mean	4.191757	4.441529e+05
std	0.515219	2.927761e+06
min	1.000000	0.000000e+00
25%	4.000000	3.800000e+01
50%	4.300000	2.094000e+03
75%	4.500000	5.477550e+04
max	5.000000	7.815831e+07

In [6]:
```
dummy.dtypes
```

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

In [7]:
```
new_size = dummy['Size']
```

In [8]:
```
new_size
```

```
0                     19M
1                     14M
2                    8.7M
3                     25M
4                    2.8M
                ...
10835                 53M
10836                3.6M
10837                9.5M
10838    Varies with device
10839                 19M
Name: Size, Length: 10840, dtype: object
```

In [9]:
```
def case_1(value):
    if value[-1:] == 'k':
        #return float(value[:-2])*1
        return new_size.str.extract('(\d+.?\d+)')
    elif value[-1:] == 'M':
        #return float(value[:-2])*1000
        return new_size.str.extract('(\d+.?\d+)').astype('float')*1000
    else:
        return np.nan
p_df2 = new_size.map(lambda x: case_1(x))
```

In [10]:
```
p_df2[0]
```

	0
0	19000.0
1	14000.0
2	8700.0
3	25000.0
4	2800.0
...	...
10835	53000.0
10836	3600.0
10837	9500.0
10838	NaN
10839	19000.0

10840 rows × 1 columns

In [11]:
```
dummy['Size']=p_df2[0]
```

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

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19000.0	10,000+	Free	0	Everyone	Art
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14000.0	500,000+	Free	0	Everyone	Art De 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
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25000.0	50,000,000+	Free	0	Teen	Art
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2800.0	100,000+	Free	0	Everyone	Art De

In [13]:
```
dummy.isnull().sum()
```

```
App                0
Category           0
Rating          1474
Reviews            0
Size            1695
Installs           0
Type               1
Price              0
Content Rating     0
Genres             0
Last Updated       0
Current Ver        8
Android Ver        2
dtype: int64
```

In [14]:
```
dummy = dummy.dropna(how="any")
```

In [15]:
```
dummy
```

	App	Category	Rating	Reviews	Size	Installs	Type	Price
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19000.0	10,000+	Free	0
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14000.0	500,000+	Free	0
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8700.0	5,000,000+	Free	0
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25000.0	50,000,000+	Free	0
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2800.0	100,000+	Free	0
...	...	...	...	...	...	...	...	...
10832	Chemin (fr)	BOOKS_AND_REFERENCE	4.8	44	619000.0	1,000+	Free	0
10833	FR Calculator	FAMILY	4.0	7	2600.0	500+	Free	0
10835	Sya9a Maroc - FR	FAMILY	4.5	38	53000.0	5,000+	Free	0
10836	Fr. Mike Schmitz Audio Teachings	FAMILY	5.0	4	3600.0	100+	Free	0
10839	iHoroscope - 2018 Daily Horoscope & Astrology	LIFESTYLE	4.5	398307	19000.0	10,000,000+	Free	0

7723 rows × 13 columns

In [16]:
```
dummy.dtypes
```

```
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
```

In [17]:
```
install_a = dummy["Installs"]
```

In [18]:
```
install_a
```

```
0            10,000+
1           500,000+
2         5,000,000+
3        50,000,000+
4           100,000+
              ...
10832         1,000+
10833           500+
10835         5,000+
10836           100+
10839    10,000,000+
Name: Installs, Length: 7723, dtype: object
```

In [19]:
```
install_a = install_a.str.extract('(\d+,?\d+,?\d+,?\d+)').apply(lambda col: pd.to
```

In [20]:
```
install_a
```

	0
0	10000.0
1	500000.0
2	5000000.0
3	50000000.0
4	100000.0
...	...
10832	1000.0
10833	NaN
10835	5000.0
10836	NaN
10839	10000000.0

7723 rows × 1 columns

In [21]:
```
dummy['Installs'] = install_a
```

```
C:\Users\pavan\AppData\Local\Programs\Python\Python37\lib\site-packages\ipykernel_launcher.py:1: Settin
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_guide/indexing.html#returning
ta.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)
  """Entry point for launching an IPython kernel.
```

In [22]:
```
dummy
```

	App	Category	Rating	Reviews	Size	Installs	Type	Price
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19000.0	10000.0	Free	0
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14000.0	500000.0	Free	0
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8700.0	5000000.0	Free	0
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25000.0	50000000.0	Free	0
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2800.0	100000.0	Free	0
...	...	...	...	...	...	...	...	...
10832	Chemin (fr)	BOOKS_AND_REFERENCE	4.8	44	619000.0	1000.0	Free	0
10833	FR Calculator	FAMILY	4.0	7	2600.0	NaN	Free	0
10835	Sya9a Maroc - FR	FAMILY	4.5	38	53000.0	5000.0	Free	0
10836	Fr. Mike Schmitz Audio Teachings	FAMILY	5.0	4	3600.0	NaN	Free	0
10839	iHoroscope - 2018 Daily Horoscope & Astrology	LIFESTYLE	4.5	398307	19000.0	10000000.0	Free	0

7723 rows × 13 columns

In [23]:
```
dummy['Reviews'] = dummy['Reviews'].astype(int)
```

```
C:\Users\pavan\AppData\Local\Programs\Python\Python37\lib\site-packages\ipykernel_launcher.py:1: Settin
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_guide/indexing.html#returning
ta.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)
  """Entry point for launching an IPython kernel.
```

In [24]:
```
dummy.dtypes
```

```
App               object
Category          object
Rating           float64
Reviews            int32
Size             float64
Installs         float64
Type              object
Price             object
Content Rating    object
Genres            object
Last Updated      object
Current Ver       object
Android Ver       object
dtype: object
```

In [25]:
```
dummy['Category'].unique()
```

```
array(['ART_AND_DESIGN', 'AUTO_AND_VEHICLES', 'BEAUTY',
       'BOOKS_AND_REFERENCE', 'BUSINESS', 'COMICS', 'COMMUNICATION',
       'DATING', 'EDUCATION', 'ENTERTAINMENT', 'EVENTS', 'FINANCE',
       'FOOD_AND_DRINK', 'HEALTH_AND_FITNESS', 'HOUSE_AND_HOME',
       'LIBRARIES_AND_DEMO', 'LIFESTYLE', 'GAME', 'FAMILY', 'MEDICAL',
       'SOCIAL', 'SHOPPING', 'PHOTOGRAPHY', 'SPORTS', 'TRAVEL_AND_LOCAL',
       'TOOLS', 'PERSONALIZATION', 'PRODUCTIVITY', 'PARENTING', 'WEATHER',
       'VIDEO_PLAYERS', 'NEWS_AND_MAGAZINES', 'MAPS_AND_NAVIGATION'],
      dtype=object)
```

In [26]:
```
price = dummy['Price'].str.replace(',', '')
price = dummy['Price'].str.replace('$', '')
```

```
C:\Users\pavan\AppData\Local\Programs\Python\Python37\lib\site-packages\ipykernel_launcher.py:2: Future
will change from True to False in a future version. In addition, single character regular expressions w
ings when regex=True.
```

In [27]:
```
price
```

```
0        0
1        0
2        0
3        0
4        0
        ..
10832    0
10833    0
10835    0
10836    0
10839    0
Name: Price, Length: 7723, dtype: object
```

In [28]:
```
dummy['Price'] = price.astype(float)
```

```
C:\Users\pavan\AppData\Local\Programs\Python\Python37\lib\site-packages\ipykernel_launcher.py:1: Settin
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_guide/indexing.html#returning
ta.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)
  """Entry point for launching an IPython kernel.
```

In [29]:
```
dummy['Price'].dtypes
```

```
dtype('float64')
```

In [30]:
```
dummy['new'] = np.where(((dummy['Type']=='Free')==(dummy['Installs']==0)),dummy['
```

```
C:\Users\pavan\AppData\Local\Programs\Python\Python37\lib\site-packages\ipykernel_launcher.py:1: Settin
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_guide/indexing.html#returning
ta.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)
  """Entry point for launching an IPython kernel.
```

In [31]:
```
dummy['new']
```

```
0       NaN
1       NaN
2       NaN
3       NaN
4       NaN
        ...
10832   NaN
10833   NaN
10835   NaN
10836   NaN
10839   NaN
Name: new, Length: 7723, dtype: object
```

In [32]:
```
dummy['new'].value_counts()
```

```
Paid    577
Name: new, dtype: int64
```

In [33]:
```
dummy.drop(['new'], axis=1)
```

	App	Category	Rating	Reviews	Size	Installs	Type	Price
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19000.0	10000.0	Free	0.0
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14000.0	500000.0	Free	0.0
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8700.0	5000000.0	Free	0.0
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25000.0	50000000.0	Free	0.0
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2800.0	100000.0	Free	0.0
...	...	...	...	...	...	...	...	...
10832	Chemin (fr)	BOOKS_AND_REFERENCE	4.8	44	619000.0	1000.0	Free	0.0
10833	FR Calculator	FAMILY	4.0	7	2600.0	NaN	Free	0.0
10835	Sya9a Maroc - FR	FAMILY	4.5	38	53000.0	5000.0	Free	0.0
10836	Fr. Mike Schmitz Audio Teachings	FAMILY	5.0	4	3600.0	NaN	Free	0.0
10839	iHoroscope - 2018 Daily Horoscope & Astrology	LIFESTYLE	4.5	398307	19000.0	10000000.0	Free	0.0

7723 rows × 13 columns

In [34]:
```
dummy = dummy.drop(['Category'],axis=1)
```

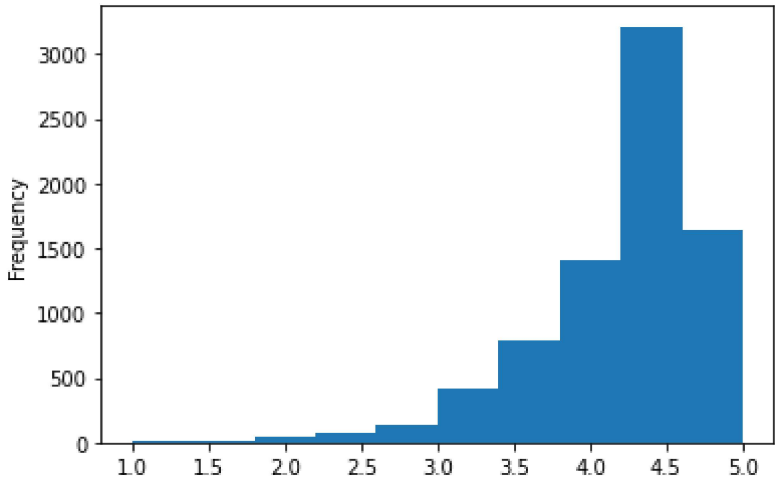
In [35]:
```
dummy
```

	App	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres
0	Photo Editor & Candy Camera & Grid & ScrapBook	4.1	159	19000.0	10000.0	Free	0.0	Everyone	Art & Design
1	Coloring book moana	3.9	967	14000.0	500000.0	Free	0.0	Everyone	Art & Design;Pretend Play
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	4.7	87510	8700.0	5000000.0	Free	0.0	Everyone	Art & Design
3	Sketch - Draw & Paint	4.5	215644	25000.0	50000000.0	Free	0.0	Teen	Art & Design
4	Pixel Draw - Number Art Coloring Book	4.3	967	2800.0	100000.0	Free	0.0	Everyone	Art & Design;Creativity
...	...	...	...	...	...	...	...	...	...
10832	Chemin (fr)	4.8	44	619000.0	1000.0	Free	0.0	Everyone	Books & Reference
10833	FR Calculator	4.0	7	2600.0	NaN	Free	0.0	Everyone	Education
10835	Sya9a Maroc - FR	4.5	38	53000.0	5000.0	Free	0.0	Everyone	Education
10836	Fr. Mike Schmitz Audio Teachings	5.0	4	3600.0	NaN	Free	0.0	Everyone	Education
10839	iHoroscope - 2018 Daily Horoscope & Astrology	4.5	398307	19000.0	10000000.0	Free	0.0	Everyone	Lifestyle

7723 rows × 13 columns

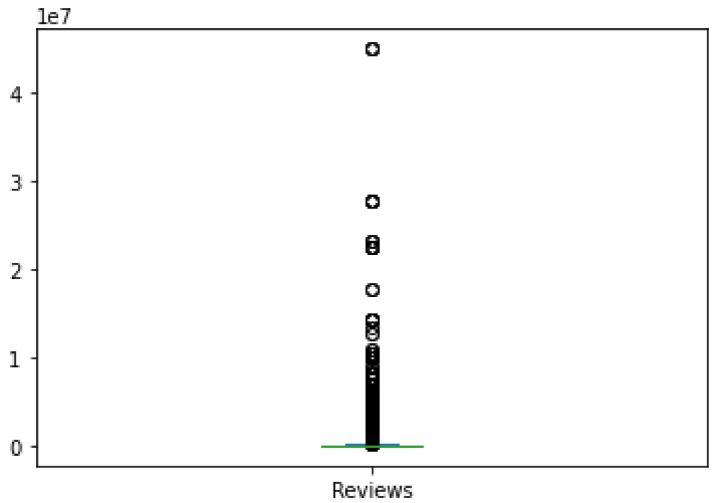
In [36]:
```
dummy['Rating'].plot.hist()
```

```
<AxesSubplot:ylabel='Frequency'>
```



In [37]:
```
dummy['Reviews'].plot.box()
```

```
<AxesSubplot:>
```



In [38]:
```
dummy['Reviews'].describe()
```

```
count    7.723000e+03
mean     2.948983e+05
std      1.863933e+06
min      1.000000e+00
25%      1.075000e+02
50%      2.332000e+03
75%      3.905300e+04
max      4.489389e+07
Name: Reviews, dtype: float64
```

In [39]:
```
dummy['Reviews'].quantile([0.1,0.35,0.5,0.7,0.9])
```

```
0.10        14.0
0.35       350.0
0.50      2332.0
0.70     24094.0
0.90    264207.0
Name: Reviews, dtype: float64
```

In [40]:
```
df = dummy[(dummy['Reviews']>=2332)&(dummy['Reviews']<=24094)]
```

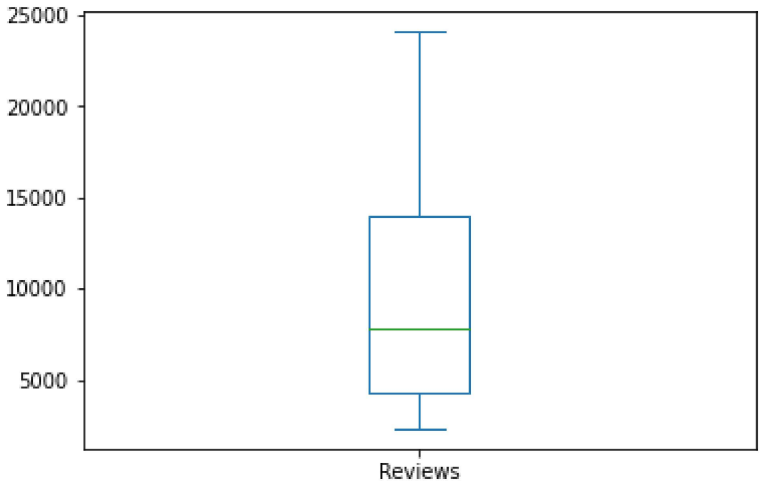
In [41]:
```
df
```

	App	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Up
8	Garden Coloring Book	4.4	13791	33000.0	1000000.0	Free	0.0	Everyone	Art & Design	Sept 20, 2
10	Text on Photo - Fonteee	4.4	13880	28000.0	1000000.0	Free	0.0	Everyone	Art & Design	Octo 27, 2
11	Name Art Photo Editor - Focus n Filters	4.4	8788	12000.0	1000000.0	Free	0.0	Everyone	Art & Design	July 2018
13	Mandala Coloring Book	4.6	4326	21000.0	100000.0	Free	0.0	Everyone	Art & Design	June 2018
16	Photo Designer - Write your name with shapes	4.7	3632	5500.0	500000.0	Free	0.0	Everyone	Art & Design	July 2018
...	...	...	...	...	...	...	...	...	...	...
10757	Fingerprint Lock Screen Prank	4.1	10786	4300.0	1000000.0	Free	0.0	Everyone	Tools	Dec 9, 20
10791	Soccer Clubs Logo Quiz	4.2	21661	16000.0	1000000.0	Free	0.0	Everyone	Trivia	May 2018
10794	Reindeer VPN - Proxy VPN	4.2	7339	4000.0	100000.0	Free	0.0	Everyone	Tools	May 2018
10803	Poker Pro.Fr	4.2	5442	17000.0	100000.0	Free	0.0	Teen	Card	May 2018
10814	Golden Dictionary (FR-AR)	4.2	5775	4900.0	500000.0	Free	0.0	Everyone	Books & Reference	July 2018

1546 rows × 13 columns

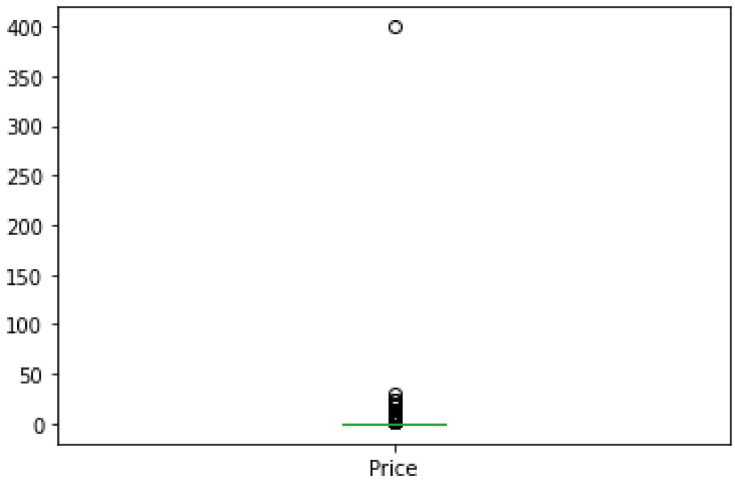
In [42]:
```
df['Reviews'].plot.box()
```

```
<AxesSubplot:>
```



In [43]:
```
df['Price'].plot.box()
```

```
<AxesSubplot:>
```



In [44]:
```
df['Price'].quantile([0.1,0.35,0.5,0.7,0.9])
```

```
0.10    0.0
0.35    0.0
0.50    0.0
0.70    0.0
0.90    0.0
Name: Price, dtype: float64
```

In [45]:
```
df['Installs'].quantile([0.1,0.35,0.5,0.55,0.9])
```

```
0.10     100000.0
0.35     500000.0
0.50     500000.0
0.55    1000000.0
0.90    1000000.0
Name: Installs, dtype: float64
```

In [46]:
```
df = df[(df['Installs']>=10000)&(df['Installs']<1000000)]
```

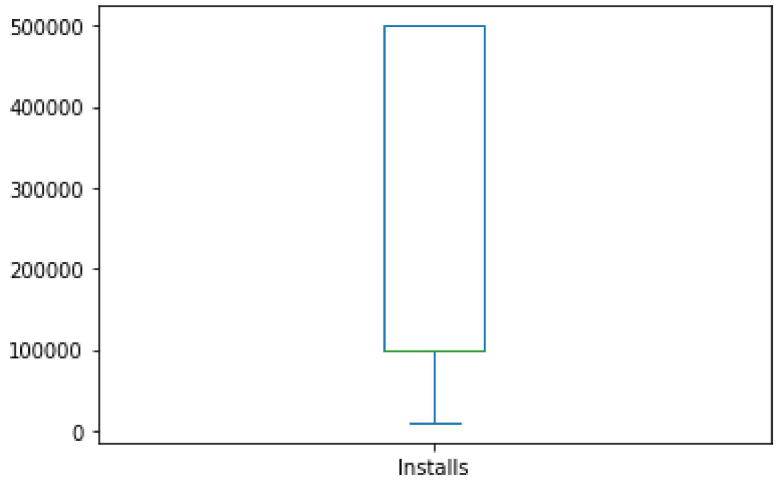
In [47]:
```
df
```

	App	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	U
13	Mandala Coloring Book	4.6	4326	21000.0	100000.0	Free	0.0	Everyone	Art & Design	Ju 20
16	Photo Designer - Write your name with shapes	4.7	3632	5500.0	500000.0	Free	0.0	Everyone	Art & Design	Ju 20
22	Superheroes Wallpapers | 4K Backgrounds	4.7	7699	4200.0	500000.0	Free	0.0	Everyone 10+	Art & Design	Ju 20
26	Colorfit - Drawing & Coloring	4.7	20260	25000.0	500000.0	Free	0.0	Everyone	Art & Design;Creativity	O 11
32	Anime Manga Coloring Book	4.5	5035	11000.0	100000.0	Free	0.0	Everyone	Art & Design	Ju 20
...	...	...	...	...	...	...	...	...	...	...
10738	FreedomPop Messaging Phone/SIM	3.6	9894	39000.0	500000.0	Free	0.0	Everyone	Communication	Ju 20
10749	Finger Scanner Gestures	4.2	2531	3300.0	100000.0	Free	0.0	Everyone	Tools	Ju 20
10794	Reindeer VPN - Proxy VPN	4.2	7339	4000.0	100000.0	Free	0.0	Everyone	Tools	M 20
10803	Poker Pro.Fr	4.2	5442	17000.0	100000.0	Free	0.0	Teen	Card	M 20
10814	Golden Dictionary (FR-AR)	4.2	5775	4900.0	500000.0	Free	0.0	Everyone	Books & Reference	Ju 20

790 rows × 13 columns

In [48]:
```
df['Installs'].plot.box()
```

```
<AxesSubplot:>
```



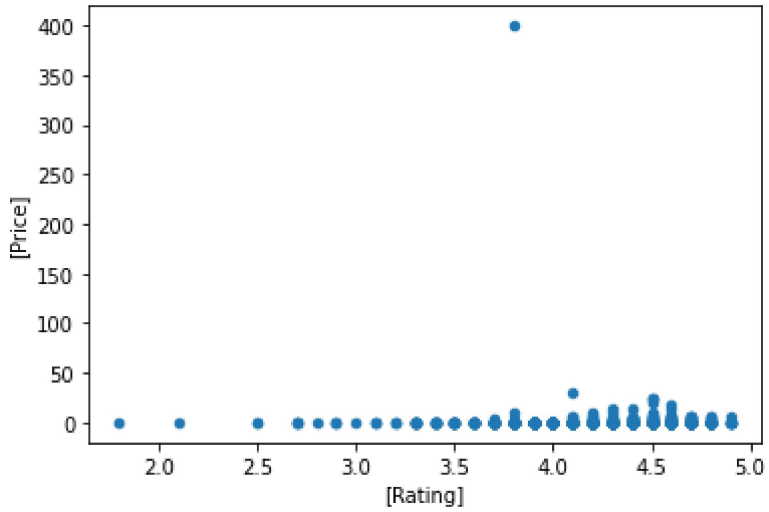
In [49]:
```
dummy = df
```

In [50]:
```
dummy.describe()
```

	Rating	Reviews	Size	Installs	Price
count	790.000000	790.000000	790.000000	790.000000	790.000000
mean	4.256582	6644.970886	30937.974684	273025.316456	1.113734
std	0.416973	4602.741818	64628.867156	203159.672024	14.443031
min	1.800000	2332.000000	1100.000000	10000.000000	0.000000
25%	4.100000	3364.250000	7125.000000	100000.000000	0.000000
50%	4.300000	4996.000000	17000.000000	100000.000000	0.000000
75%	4.600000	8032.000000	37000.000000	500000.000000	0.000000
max	4.900000	24005.000000	818000.000000	500000.000000	399.990000

In [51]:
```
dummy.plot.scatter(['Rating'],['Price'])
```

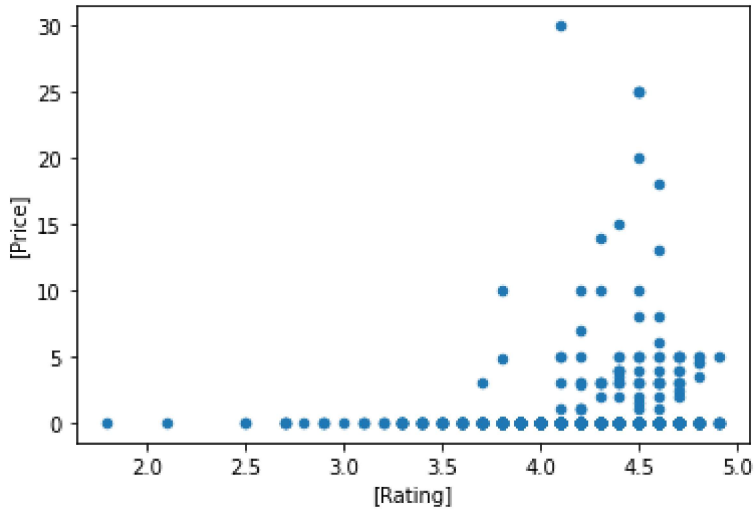
```
<AxesSubplot:xlabel='[Rating]', ylabel='[Price]'>
```



In [52]:
```
dummy = dummy[dummy['Price']<100]
```

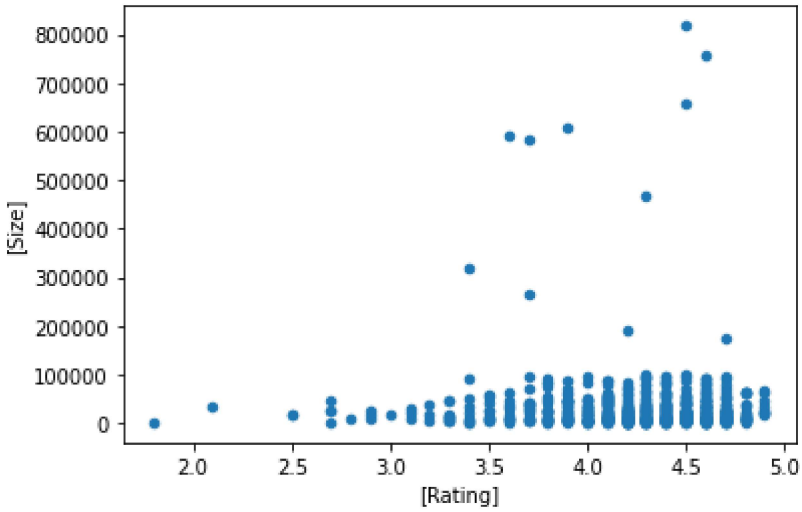
In [53]:
```
dummy.plot.scatter(['Rating'],['Price'])
```

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



In [54]:
```
dummy.plot.scatter(['Rating'],['Size'])
```

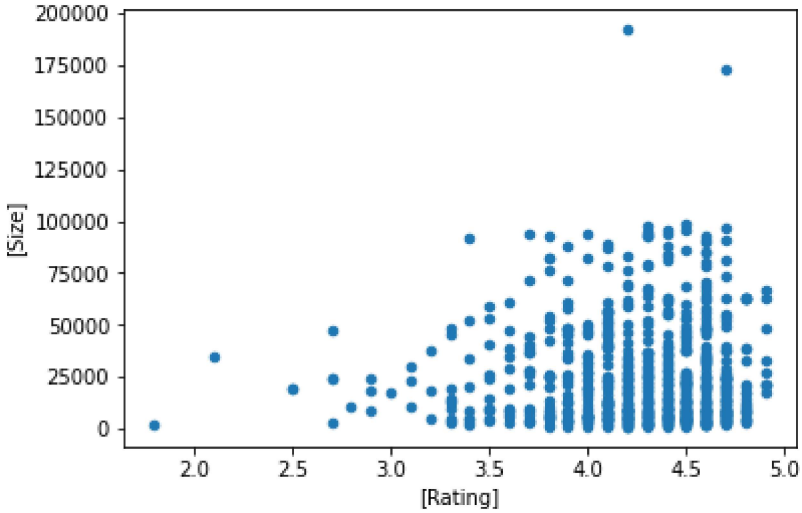
```
<AxesSubplot:xlabel='[Rating]', ylabel='[Size]'>
```



In [55]:
```
dummy = dummy[dummy['Size']<200000]
```

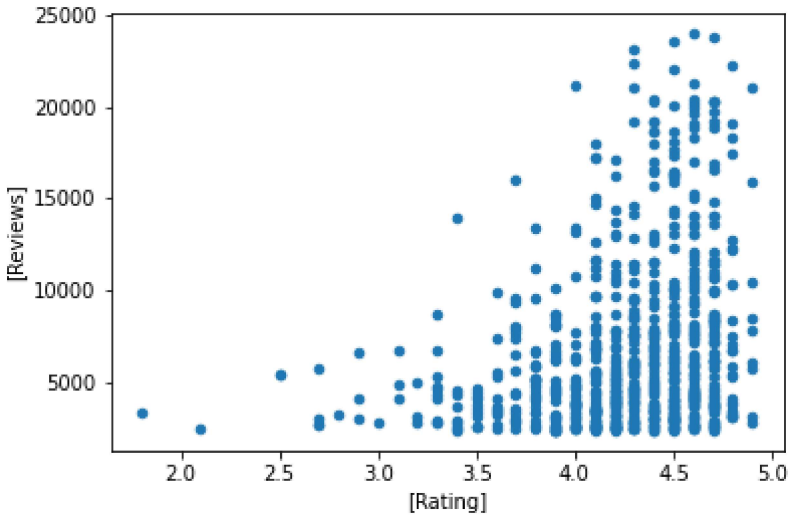
In [56]:
```
dummy.plot.scatter(['Rating'],['Size'])
```

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



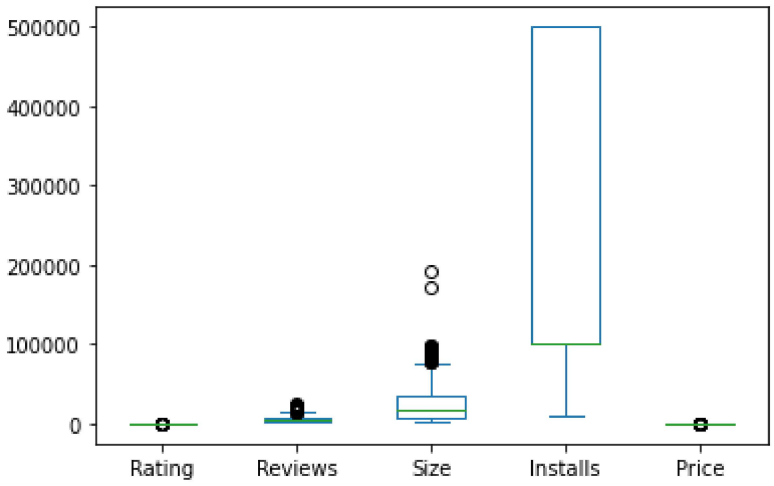
In [57]:
```
dummy.plot.scatter(['Rating'],['Reviews'])
```

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



In [58]:
```
dummy.plot.box(['Rating','Content Rating'])
```

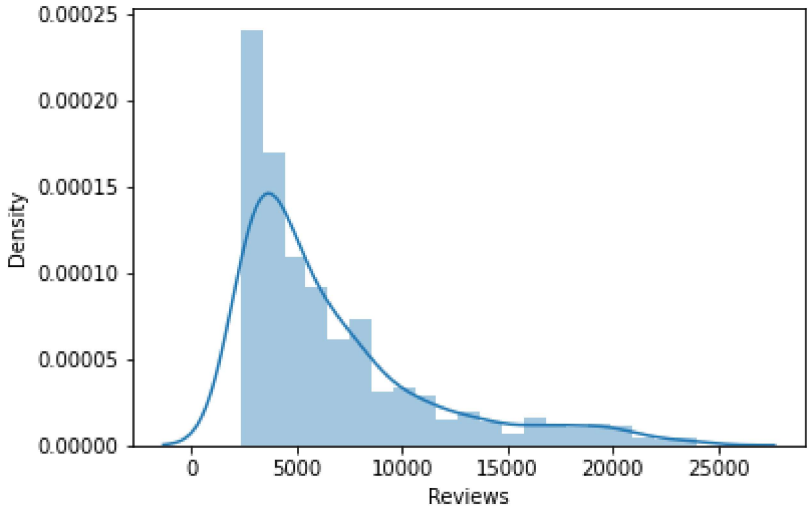
```
<AxesSubplot:>
```



In [59]:
```
import seaborn as sns
```

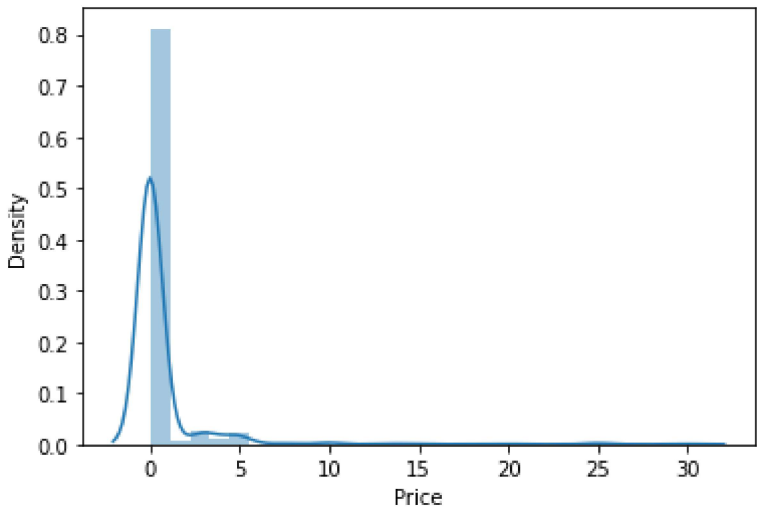
In [60]:
```
sns.distplot(dummy['Reviews'])
```

```
C:\Users\pavan\AppData\Local\Programs\Python\Python37\lib\site-packages\ipykernel_launcher.py:1: UserWa

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751 (https://gist.github.com/mwaskom/de44147ed2974457

  """Entry point for launching an IPython kernel.


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



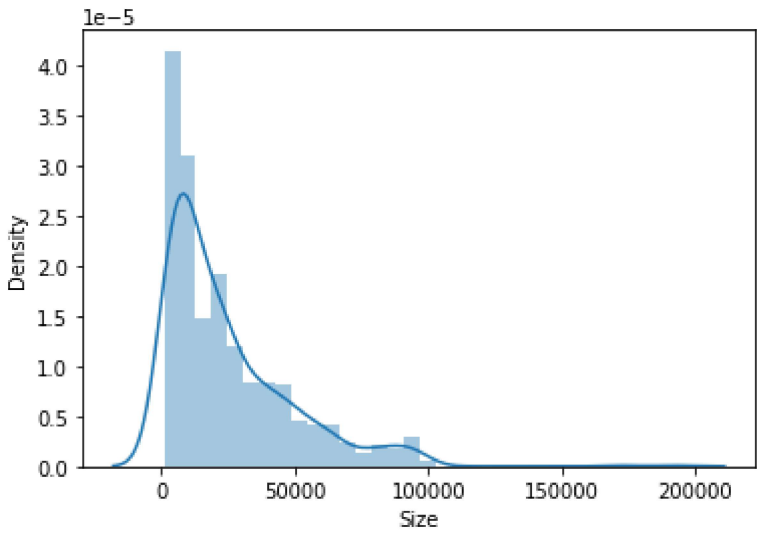
In [61]:
```
sns.distplot(dummy['Price'])
```

```
C:\Users\pavan\AppData\Local\Programs\Python\Python37\lib\site-packages\ipykernel_launcher.py:1: UserWa

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751 (https://gist.github.com/mwaskom/de44147ed2974457

  """Entry point for launching an IPython kernel.


<AxesSubplot:xlabel='Price', ylabel='Density'>
```



In [62]:
```
sns.distplot(dummy['Size'])
```

```
C:\Users\pavan\AppData\Local\Programs\Python\Python37\lib\site-packages\ipykernel_launcher.py:1: UserWa

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751 (https://gist.github.com/mwaskom/de44147ed2974457

  """Entry point for launching an IPython kernel.


<AxesSubplot:xlabel='Size', ylabel='Density'>
```



In [63]:
```
size_array = dummy['Size']
```

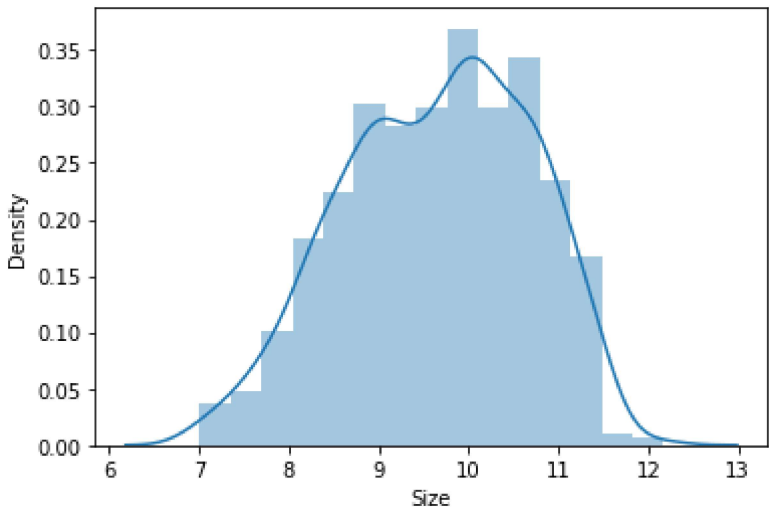
In [64]:
```
size_log = np.log1p(size_array)
```

In [65]:
```
size_log
```

```
13        9.952325
16        8.612685
22        8.343078
26       10.126671
32        9.305741
            ...
10738    10.571343
10749     8.101981
10794     8.294300
10803     9.741027
10814     8.497195
Name: Size, Length: 780, dtype: float64
```

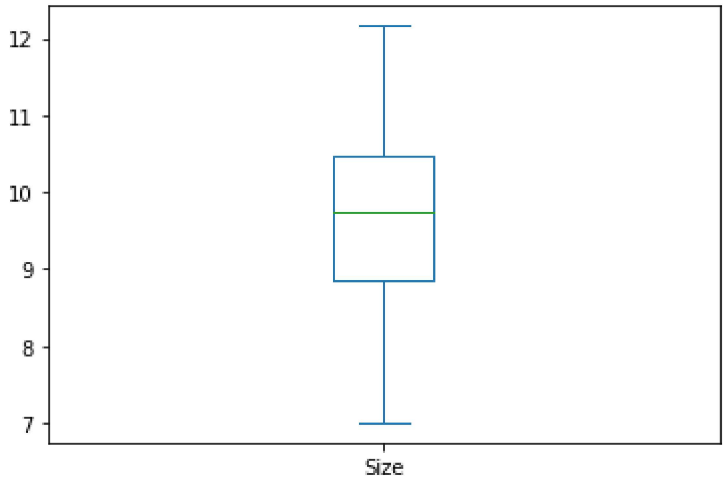
In [66]:
```
sns.distplot(size_log)
```

```
C:\Users\pavan\AppData\Local\Programs\Python\Python37\lib\site-packages\ipykernel_launcher.py:1: UserWa

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751 (https://gist.github.com/mwaskom/de44147ed2974457

  """Entry point for launching an IPython kernel.


<AxesSubplot:xlabel='Size', ylabel='Density'>
```



In [67]:
```
size_log.plot.box()
plt.show
```

```
<function matplotlib.pyplot.show(close=None, block=None)>
```



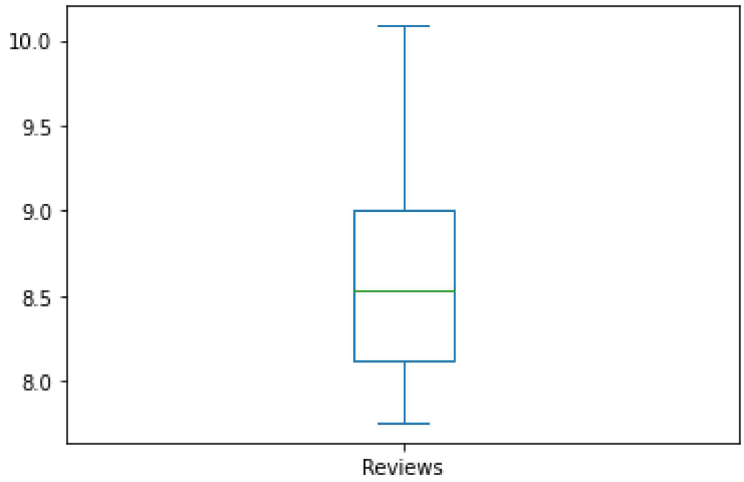
In [68]:
```
dummy['Size'] = size_log
```

In [69]:
```
reviews_array = dummy['Reviews']
```

In [70]:
```
reviews_log = np.log1p(reviews_array)
```

In [71]:
```
reviews_log.plot.box()
```

```
<AxesSubplot:>
```



In [72]:
```
dummy['Reviews'] = reviews_log
```

In [73]:
```
dummy.head()
```

	App	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Up
13	Mandala Coloring Book	4.6	8.372630	9.952325	100000.0	Free	0.0	Everyone	Art & Design	Jur 20
16	Photo Designer - Write your name with shapes	4.7	8.197814	8.612685	500000.0	Free	0.0	Everyone	Art & Design	Jul 20
22	Superheroes Wallpapers | 4K Backgrounds	4.7	8.948976	8.343078	500000.0	Free	0.0	Everyone 10+	Art & Design	Jul 20
26	Colorfit - Drawing & Coloring	4.7	9.916453	10.126671	500000.0	Free	0.0	Everyone	Art & Design;Creativity	Oc 11,
32	Anime Manga Coloring Book	4.5	8.524367	9.305741	100000.0	Free	0.0	Everyone	Art & Design	Jul 20

In [74]:
```
dummy
```

	App	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres
13	Mandala Coloring Book	4.6	8.372630	9.952325	100000.0	Free	0.0	Everyone	Art & Design
16	Photo Designer - Write your name with shapes	4.7	8.197814	8.612685	500000.0	Free	0.0	Everyone	Art & Design
22	Superheroes Wallpapers | 4K Backgrounds	4.7	8.948976	8.343078	500000.0	Free	0.0	Everyone 10+	Art & Design
26	Colorfit - Drawing & Coloring	4.7	9.916453	10.126671	500000.0	Free	0.0	Everyone	Art & Design;Creativity
32	Anime Manga Coloring Book	4.5	8.524367	9.305741	100000.0	Free	0.0	Everyone	Art & Design
...	...	...	...	...	...	...	...	...	...
10738	FreedomPop Messaging Phone/SIM	3.6	9.199785	10.571343	500000.0	Free	0.0	Everyone	Communication
10749	Finger Scanner Gestures	4.2	7.836765	8.101981	100000.0	Free	0.0	Everyone	Tools
10794	Reindeer VPN - Proxy VPN	4.2	8.901094	8.294300	100000.0	Free	0.0	Everyone	Tools
10803	Poker Pro.Fr	4.2	8.602086	9.741027	100000.0	Free	0.0	Teen	Card
10814	Golden Dictionary (FR-AR)	4.2	8.661467	8.497195	500000.0	Free	0.0	Everyone	Books & Reference

780 rows × 13 columns

In [80]:
```
data = dummy.drop(['Type'],axis=1)
```

In [76]:
```
dummy['Content Rating'].value_counts()
```

```
Everyone           557
Teen               111
Mature 17+          63
Everyone 10+        48
Adults only 18+      1
Name: Content Rating, dtype: int64
```

In [77]:
```
dummy['Genres'].value_counts()
```

```
Entertainment              41
Tools                      40
Health & Fitness           37
Dating                     35
Action                     35
                           ..
Simulation;Pretend Play     1
Music;Music & Video         1
Simulation;Education        1
Education;Pretend Play      1
Educational                 1
Name: Genres, Length: 70, dtype: int64
```

In [81]:
```
data = dummy.drop(['Genres'],axis=1)
```

In [85]:
```
dummy = dummy.drop(['Type','Content Rating','Genres','Last Updated','Current Ver'
                    'Android Ver','new'],axis=1)
```

In [86]:
```
dummy
```

	Rating	Reviews	Size	Installs	Price
13	4.6	8.372630	9.952325	100000.0	0.0
16	4.7	8.197814	8.612685	500000.0	0.0
22	4.7	8.948976	8.343078	500000.0	0.0
26	4.7	9.916453	10.126671	500000.0	0.0
32	4.5	8.524367	9.305741	100000.0	0.0
...	...	...	...	...	...
10738	3.6	9.199785	10.571343	500000.0	0.0
10749	4.2	7.836765	8.101981	100000.0	0.0
10794	4.2	8.901094	8.294300	100000.0	0.0
10803	4.2	8.602086	9.741027	100000.0	0.0
10814	4.2	8.661467	8.497195	500000.0	0.0

780 rows × 5 columns

In [93]:
```
from sklearn.model_selection import train_test_split
```

In [96]:
```
X = dummy[['Reviews','Size','Installs','Price']]
y = dummy['Rating']
```

In [97]:
```
X_train, X_test, y_train, y_test = train_test_split(X,y,test_size=0.3)
```

In [98]:
```
model = LinearRegression()
model.fit(X_train, y_train)
```

```
LinearRegression()
```

In [99]:
```
model.predict([[2345,1000,3456,0]])
```

```
C:\Users\pavan\AppData\Local\Programs\Python\Python37\lib\site-packages\sklearn\base.py:451: UserWarnin
mes, but LinearRegression was fitted with feature names
  "X does not have valid feature names, but"
```

```
array([561.45588867])
```

In [ ]: