playstore_data

December 31, 2021

```python
[43]: import pandas as pd
      import numpy as np
      import matplotlib.pyplot as plt
```

```python
[44]: playstore_data=pd.read_csv(r'C:\Users\BABU\Desktop\project\googleplaystore.csv')
```

```python
[45]: playstore_data
```

```
[45]:                                                      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
```

1

```
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]
```

[46]: `playstore_data.head()`

```
[46]:                                                    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   19M       10,000+  Free     0       Everyone
      1     967   14M      500,000+  Free     0       Everyone
      2   87510  8.7M    5,000,000+  Free     0       Everyone
      3  215644   25M   50,000,000+  Free     0           Teen
      4     967  2.8M      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
```

[5]: `playstore_data.shape # to know dimesnsions of your table`

[5]: (10841, 13)

[6]: `playstore_data.describe() # to know statistics of your data`
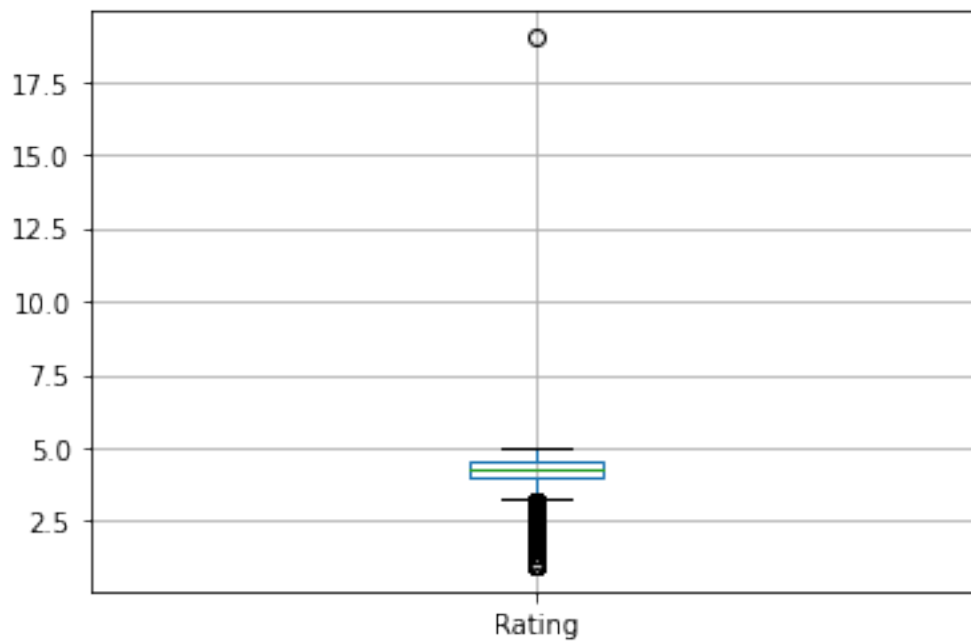
[6]:
```
            Rating
count  9367.000000
mean      4.193338
std       0.537431
min       1.000000
25%       4.000000
50%       4.300000
75%       4.500000
max      19.000000
```

[7]: `playstore_data.boxplot()`

[7]: <AxesSubplot:>

```
[8]: playstore_data.info() # to check how many null values are there
```

```
<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
```

```
[9]: playstore_data.isnull().sum() # no of null values in each column
```

```
[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
```

```
[10]: playstore_data[playstore_data.Rating>5] # outliner observed from box plot
```

```
[10]:                                             App Category  Rating Reviews  \
      10472  Life Made WI-Fi Touchscreen Photo Frame     1.9     19.0    3.0M

              Size Installs Type     Price Content Rating          Genres  \
```

```
10472   1,000+    Free    0  Everyone           NaN  February 11, 2018

        Last Updated Current Ver Android Ver
10472        1.0.19  4.0 and up          NaN
```

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

[12]: `playstore_data[10470:10475] # checking whether the row is deeleted or not`
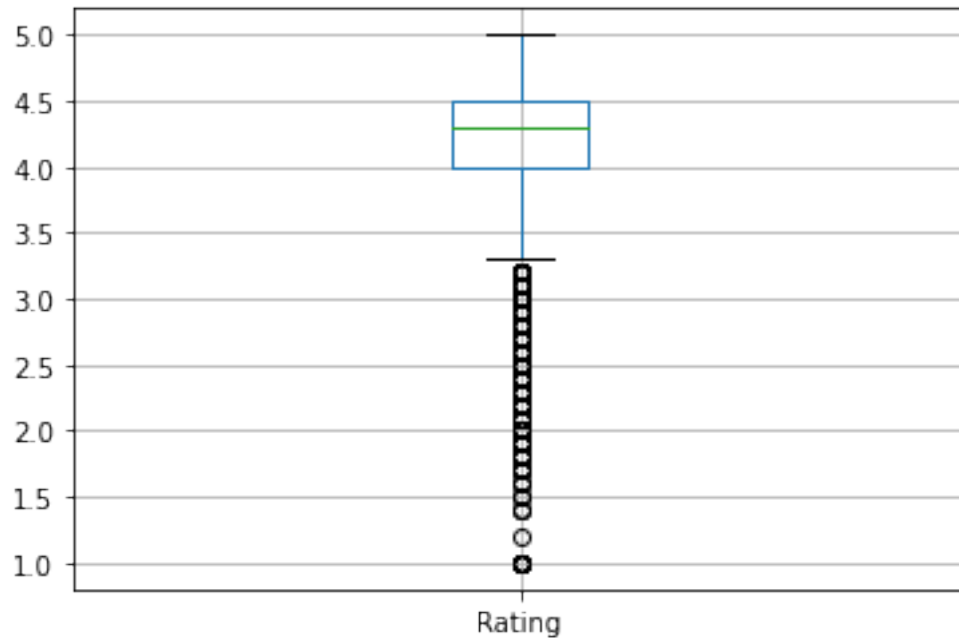
[12]:
```
                           App         Category  Rating Reviews  Size  \
10470               Jazz Wi-Fi    COMMUNICATION     3.4      49  4.0M
10471           Xposed Wi-Fi-Pwd  PERSONALIZATION     3.5    1042  404k
10473  osmino Wi-Fi: free WiFi            TOOLS     4.2  134203  4.1M
10474             Sat-Fi Voice    COMMUNICATION     3.4      37   14M
10475          Wi-Fi Visualizer            TOOLS     3.9     132  2.6M

          Installs  Type Price Content Rating         Genres  \
10470       10,000+  Free     0       Everyone    Communication
10471      100,000+  Free     0       Everyone  Personalization
10473  10,000,000+  Free     0       Everyone            Tools
10474        1,000+  Free     0       Everyone    Communication
10475       50,000+  Free     0       Everyone            Tools

           Last Updated Current Ver    Android Ver
10470  February 10, 2017        0.1    2.3 and up
10471     August 5, 2014      3.0.0  4.0.3 and up
10473     August 7, 2018    6.06.14    4.4 and up
10474  November 21, 2014    2.2.1.5    2.2 and up
10475       May 17, 2017      0.0.9    2.3 and up
```

[13]: `playstore_data.boxplot()`

[13]: `<AxesSubplot:>`
```


                            5
```

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

```
[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
```

```
[15]: def impute_median(series):
          return series.fillna(series.median())
```

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

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

```
[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
```

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

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

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

```
[20]: playstore_data.isnull().sum()
```

```
[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
```

[21]:
```python
playstore_data['Price'] = playstore_data['Price'].apply(lambda x: str(x).
 →replace('$','')if '$' in str(x) else str(x))
playstore_data['Price'] = playstore_data['Price'].apply(lambda x: float(x))
playstore_data['Reviews'] =pd.
 →to_numeric(playstore_data['Reviews'],errors='coerce')
```

[22]:
```python
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))
```

[23]:
```python
playstore_data.describe()
```

[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      1.027368
std        0.480342  2.927761e+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  5.477550e+04  5.000000e+06      0.000000
max        5.000000  7.815831e+07  1.000000e+09    400.000000
```

[28]:
```python
gby=playstore_data.groupby('Category')
x=gby['Rating'].agg(np.mean)
y=gby['Price'].agg(np.sum)
z=gby['Reviews'].agg(np.mean)
print(x)
print(y)
print(z)
```

```
Category
ART_AND_DESIGN        4.355385
AUTO_AND_VEHICLES     4.205882
BEAUTY                4.283019
BOOKS_AND_REFERENCE   4.335498
BUSINESS              4.182391
COMICS                4.160000
COMMUNICATION         4.180103
DATING                4.025641
EDUCATION             4.388462
ENTERTAINMENT         4.126174
```

```
EVENTS                4.395313
FAMILY                4.204564
FINANCE               4.151639
FOOD_AND_DRINK        4.185827
GAME                  4.286888
HEALTH_AND_FITNESS    4.280059
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             4.300000
PERSONALIZATION       4.328827
PHOTOGRAPHY           4.197910
PRODUCTIVITY          4.226651
SHOPPING              4.263077
SOCIAL                4.261017
SPORTS                4.236458
TOOLS                 4.080071
TRAVEL_AND_LOCAL      4.132946
VIDEO_PLAYERS         4.084000
WEATHER               4.248780
Name: Rating, dtype: float64
Category
ART_AND_DESIGN           5.97
AUTO_AND_VEHICLES       13.47
BEAUTY                   0.00
BOOKS_AND_REFERENCE    119.77
BUSINESS               185.27
COMICS                   0.00
COMMUNICATION           83.14
DATING                  31.43
EDUCATION               17.96
ENTERTAINMENT            7.98
EVENTS                 109.99
FAMILY                2434.78
FINANCE               2900.83
FOOD_AND_DRINK           8.48
GAME                   287.30
HEALTH_AND_FITNESS      67.34
HOUSE_AND_HOME           0.00
LIBRARIES_AND_DEMO       0.99
LIFESTYLE             2360.87
MAPS_AND_NAVIGATION     26.95
MEDICAL               1439.96
NEWS_AND_MAGAZINES       3.98
PARENTING                9.58
```

```
PERSONALIZATION          153.96
PHOTOGRAPHY              134.21
PRODUCTIVITY             250.93
SHOPPING                   5.48
SOCIAL                    15.97
SPORTS                   100.00
TOOLS                    267.25
TRAVEL_AND_LOCAL          49.95
VIDEO_PLAYERS             10.46
WEATHER                   32.42
Name: Price, dtype: float64
Category
ART_AND_DESIGN          2.637600e+04
AUTO_AND_VEHICLES       1.369019e+04
BEAUTY                  7.476226e+03
BOOKS_AND_REFERENCE     9.506090e+04
BUSINESS                3.033598e+04
COMICS                  5.638793e+04
COMMUNICATION           2.107138e+06
DATING                  3.115931e+04
EDUCATION               2.538191e+05
ENTERTAINMENT           3.971688e+05
EVENTS                  2.515906e+03
FAMILY                  2.080255e+05
FINANCE                 4.795281e+04
FOOD_AND_DRINK          6.994748e+04
GAME                    1.385859e+06
HEALTH_AND_FITNESS      1.111253e+05
HOUSE_AND_HOME          4.518619e+04
LIBRARIES_AND_DEMO      1.220139e+04
LIFESTYLE               3.372457e+04
MAPS_AND_NAVIGATION     2.237902e+05
MEDICAL                 3.425432e+03
NEWS_AND_MAGAZINES      1.922292e+05
PARENTING               1.597218e+04
PERSONALIZATION         2.279238e+05
PHOTOGRAPHY             6.373631e+05
PRODUCTIVITY            2.691438e+05
SHOPPING                4.424662e+05
SOCIAL                  2.105903e+06
SPORTS                  1.844536e+05
TOOLS                   3.240629e+05
TRAVEL_AND_LOCAL        2.427051e+05
VIDEO_PLAYERS           6.307439e+05
WEATHER                 1.781065e+05
Name: Reviews, dtype: float64
```
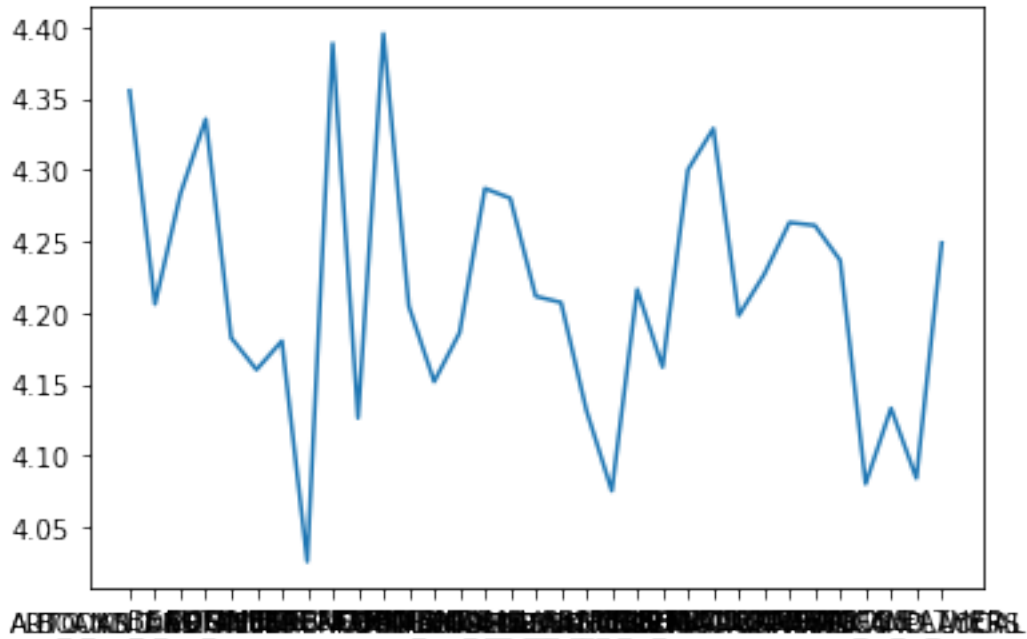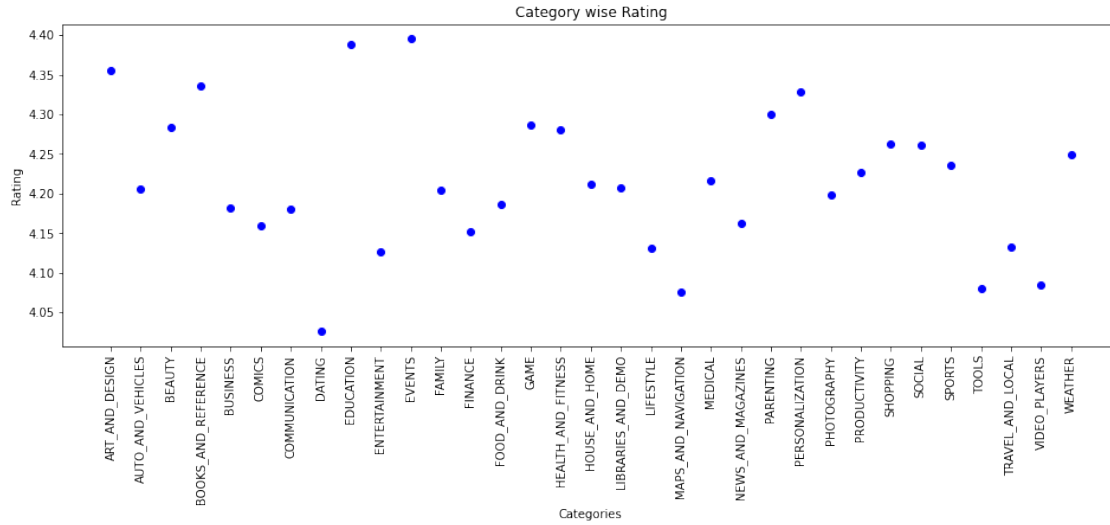
```
[42]: plt.plot(x)
```

```
[42]: [<matplotlib.lines.Line2D at 0x156d7a75160>]
```



```
[41]: plt.figure(figsize=(16,5))
      plt.plot(x,'ro',color='b')
      plt.xticks(rotation=90)
      plt.title('Category wise Rating')
      plt.xlabel('Categories')
      plt.ylabel('Rating')
      plt.show()
```
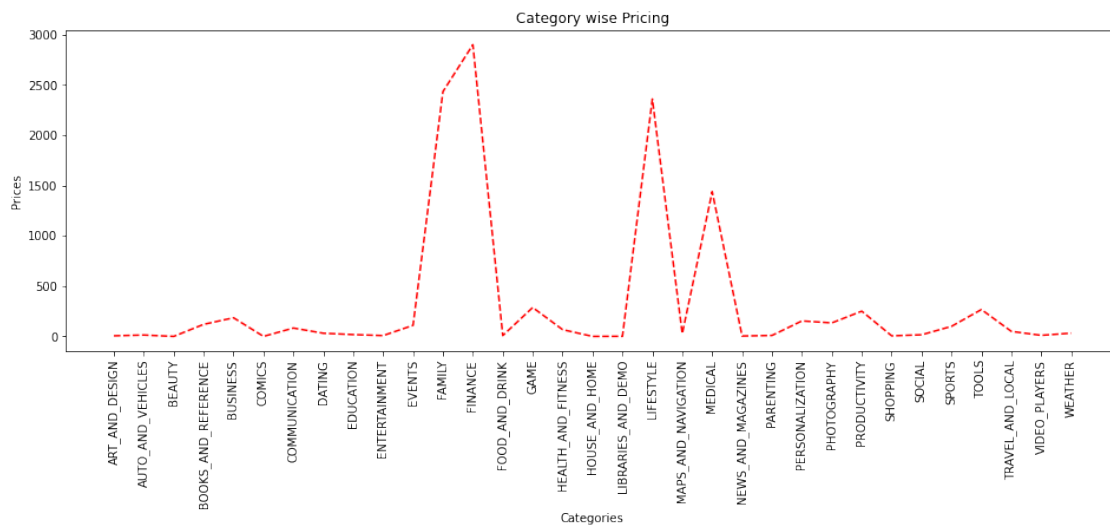
C:\Users\BABU\AppData\Local\Temp/ipykernel_2820/2414369624.py:2: UserWarning:
color is redundantly defined by the 'color' keyword argument and the fmt string
"ro" (-> color='r'). The keyword argument will take precedence.
  plt.plot(x,'ro',color='b')

Category wise Rating

```
[58]: plt.figure(figsize=(16,5))
      plt.plot(y,'r--',color='r')
      plt.xticks(rotation=90)
      plt.title('Category wise Pricing')
      plt.xlabel('Categories')
      plt.ylabel('Prices')
      plt.show()
```

C:\Users\BABU\AppData\Local\Temp/ipykernel_2820/4218611868.py:2: UserWarning:
color is redundantly defined by the 'color' keyword argument and the fmt string
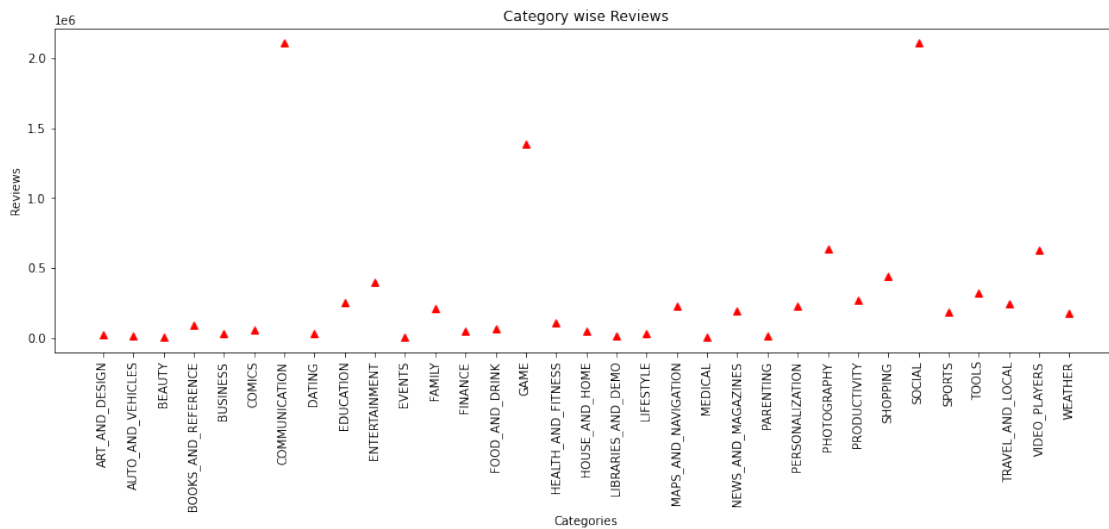"r--" (-> color='r'). The keyword argument will take precedence.
  plt.plot(y,'r--',color='r')

Category wise Pricing

```
[57]: plt.figure(figsize=(16,5))
      plt.plot(z,'r^',color='r')
      plt.xticks(rotation=90)
      plt.title('Category wise Reviews')
      plt.xlabel('Categories')
      plt.ylabel('Reviews')
      plt.show()
```

C:\Users\BABU\AppData\Local\Temp/ipykernel_2820/1754558929.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(z,'r^',color='r')

Category wise Reviews

```
[63]: plt.figure(figsize=(16,5))
      plt.plot(y, 'r*',color='b')
      plt.xticks(rotation=90)
      plt.title('category wise price')
      plt.xlabel('category names')
      plt.ylabel('price')
      plt.show()
```

C:\Users\BABU\AppData\Local\Temp/ipykernel_2820/2559667407.py:2: UserWarning:
color is redundantly defined by the 'color' keyword argument and the fmt string
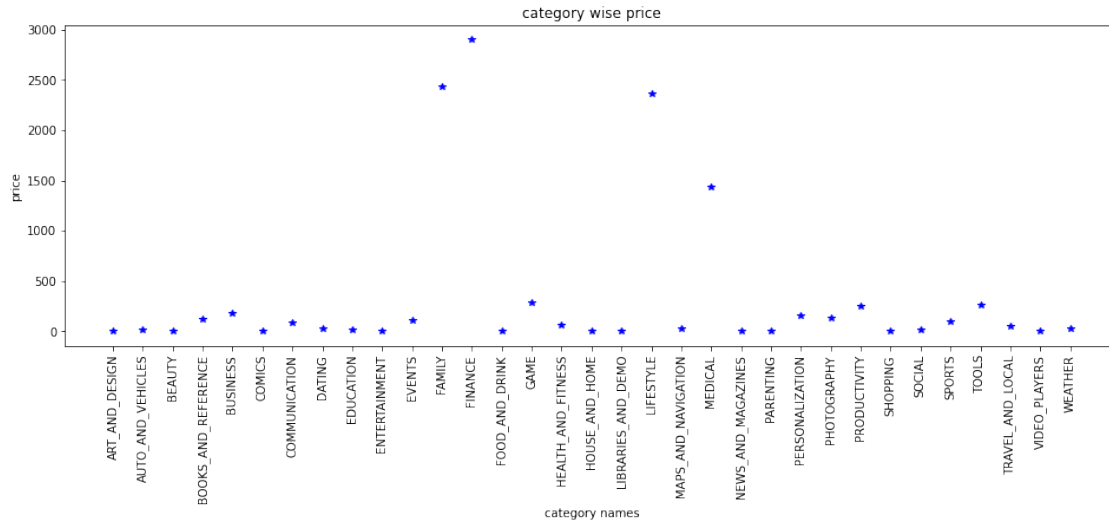"r*" (-> color='r'). The keyword argument will take precedence.
  plt.plot(y, 'r*',color='b')

category wise price

```
[64]: plt.figure(figsize=(16,5))
      plt.plot(z,'r^',color='g')
      plt.xticks(rotation=90)
      plt.title('Category wise Reviews')
      plt.xlabel('Categories')
      plt.ylabel('Reviews')
      plt.show()
```

C:\Users\BABU\AppData\Local\Temp/ipykernel_2820/2994671734.py:2: UserWarning:
color is redundantly defined by the 'color' keyword argument and the fmt string
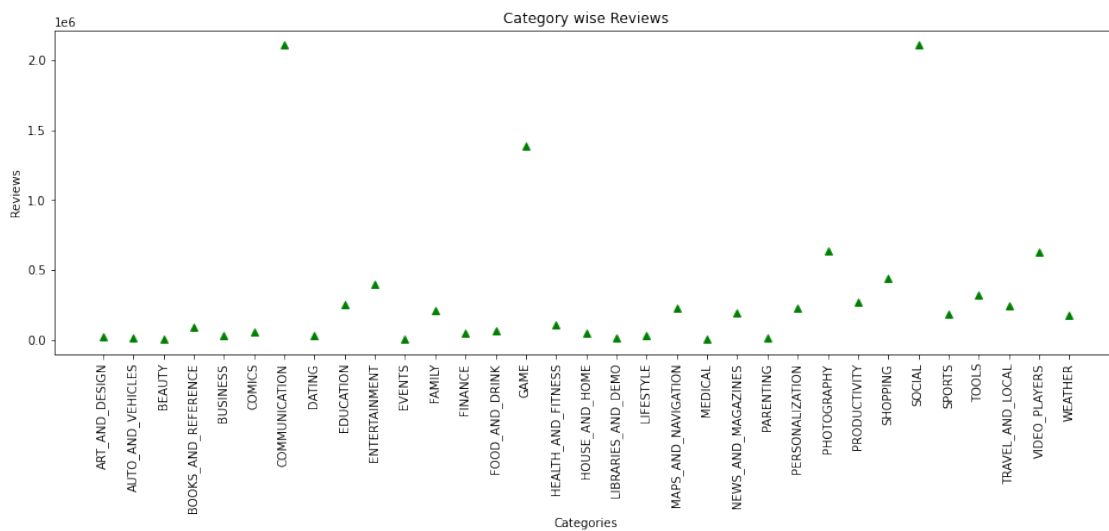"r^" (-> color='r'). The keyword argument will take precedence.
  plt.plot(z,'r^',color='g')

Category wise Reviews

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

```
[59]:                   App  Category  Rating  Reviews  Size  Installs  Type  Price  \
      Content Rating
      Adults only 18+     3         3       3        3     3         3     3      3
      Everyone         8714      8714    7420     8714  8714      8714  8714   8714
      Everyone 10+      414       414     397      414   414       414   413    414
      Mature 17+        499       499     461      499   499       499   499    499
      Teen             1208      1208    1084     1208  1208      1208  1208   1208
      Unrated             2         2       1        2     2         2     2      2

                       Genres  Last Updated  Current Ver  Android Ver
      Content Rating
      Adults only 18+       3             3            3            3
      Everyone           8714          8714         8706         8712
      Everyone 10+        414           414          414          414
      Mature 17+          499           499          499          499
      Teen               1208          1208         1208         1208
      Unrated               2             2            2            2
```

```python
[67]: playstore_data.groupby('Type').count()
```

```
[67]:         App  Category  Rating  Reviews   Size  Installs  Price  \
      Type
      0          1         1       1        1      1         1      1
      Free   10039     10039    8719    10039  10039     10039  10039
      Paid     800       800     647      800    800       800    800

             Content Rating  Genres  Last Updated  Current Ver  Android Ver
      Type
      0                   0       1             1            1            0
      Free            10039   10039         10039        10033        10038
      Paid              800     800           800          798          799
```

```python
[ ]:
```