DATA VISUALTIZATION PROJECT

Import the Required Modules

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
```

Read Data

```
google_data = pd.read_csv('/content/apps.csv')
```

```
google_data.head()        #Inspecting the first 5 rows
```

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Con Ra
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0	Ever
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Ever
2	U Launcher Lite – FREE Live Cool Themes, Hide	ART_AND_DESIGN	4.7	87510	8.7M	5,000,000+	Free	0	Ever

```
google_data.shape
```

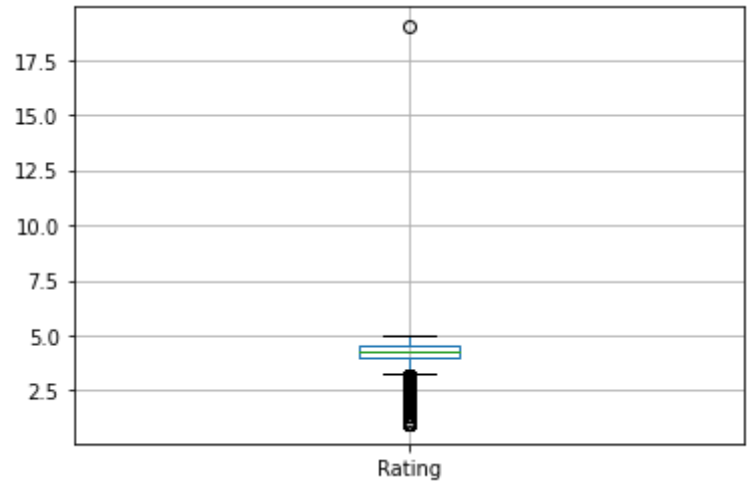
```
(10841, 13)
```

```
google_data.describe()       # Summary Statistics
```

	Rating
count	9367.000000
mean	4.193338
std	0.537431
min	1.000000
25%	4.000000
50%	4.300000

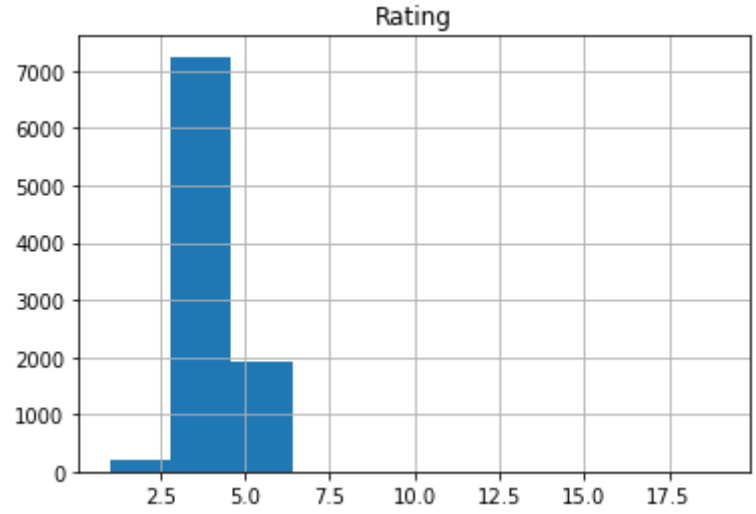
```
google_data.boxplot()
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7a2e873bd0>
```




```
google_data.hist()
```

```
array([[<matplotlib.axes._subplots.AxesSubplot object at 0x7f7a2e834ad0>]],
      dtype=object)
```




```
google_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10841 entries, 0 to 10840
Data columns (total 13 columns):
 #   Column          Non-Null Count  Dtype
```

```
---  ------          --------------  -----
 0   App             10841 non-null  object
 1   Category        10841 non-null  object
 2   Rating          9367 non-null   float64
 3   Reviews         10841 non-null  object
 4   Size            10841 non-null  object
 5   Installs        10841 non-null  object
 6   Type            10840 non-null  object
 7   Price           10841 non-null  object
 8   Content Rating  10840 non-null  object
 9   Genres          10841 non-null  object
 10  Last Updated    10841 non-null  object
 11  Current Ver     10833 non-null  object
 12  Android Ver     10838 non-null  object
dtypes: float64(1), object(12)
memory usage: 1.1+ MB
```

Data Cleaning

Count the number of missing values in the Dataframe

```
google_data.isnull()
```

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genre
0	False	False	False	False	False	False	False	False	False	Fals
1	False	False	False	False	False	False	False	False	False	Fals
2	False	False	False	False	False	False	False	False	False	Fals
3	False	False	False	False	False	False	False	False	False	Fals
4	False	False	False	False	False	False	False	False	False	Fals
...	...	...	...	...	...	...	...	...	...	.
10836	False	False	False	False	False	False	False	False	False	Fals
10837	False	False	False	False	False	False	False	False	False	Fals
10838	False	False	True	False	False	False	False	False	False	Fals
10839	False	False	False	False	False	False	False	False	False	Fals
10840	False	False	False	False	False	False	False	False	False	Fals

10841 rows × 13 columns

```
# Count the number of missing values in each column
google_data.isnull().sum()
```

```
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
```

Check how many ratings are more than 5 - Outliers

```
google_data[google_data.Rating > 5]
```

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Conte Rati
10472	Life Made WI-Fi Touchscreen	1.9	19.0	3.0M	1,000+	Free	0	Everyone	N

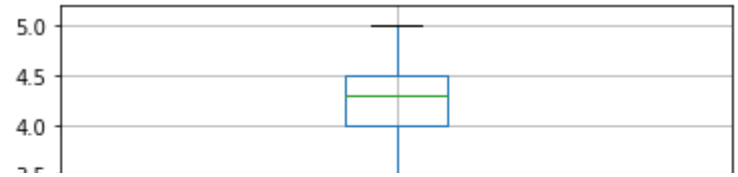
```
google_data.drop([10472],inplace=True)
```

```
google_data[10470:10475]
```

	App	Category	Rating	Reviews	Size	Installs	Type	Price	
10470	Jazz Wi-Fi	COMMUNICATION	3.4	49	4.0M	10,000+	Free	0	I
10471	Xposed Wi-Fi-Pwd	PERSONALIZATION	3.5	1042	404k	100,000+	Free	0	I
10473	osmino Wi-Fi: free WiFi	TOOLS	4.2	134203	4.1M	10,000,000+	Free	0	I
10474	Sat-Fi Voice	COMMUNICATION	3.4	37	14M	1,000+	Free	0	I

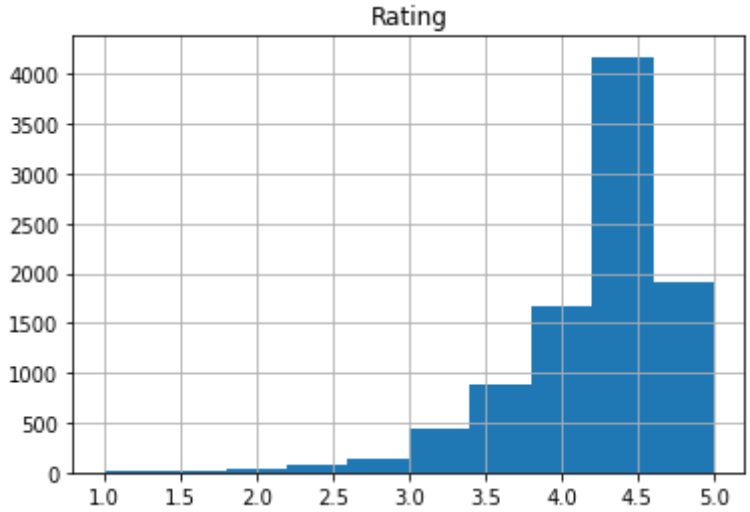
```
google_data.boxplot()
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7a2e15b190>
```




```
google_data.hist()
```

```
array([[<matplotlib.axes._subplots.AxesSubplot object at 0x7f7a4280dc50>]],
      dtype=object)
```



Remove columns that are 90% empty

```
threshold = len(google_data)* 0.1
threshold
```

```
1084.0
```

```
google_data.dropna(thresh=threshold, axis=1, inplace=True)
```

```
print(google_data.isnull().sum())
```

```
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
```

Data Imputation and Manipulation

Fill the null values with appropriate values using aggregate functions such as mean, median or mode.

[ ] *10 cells hidden*

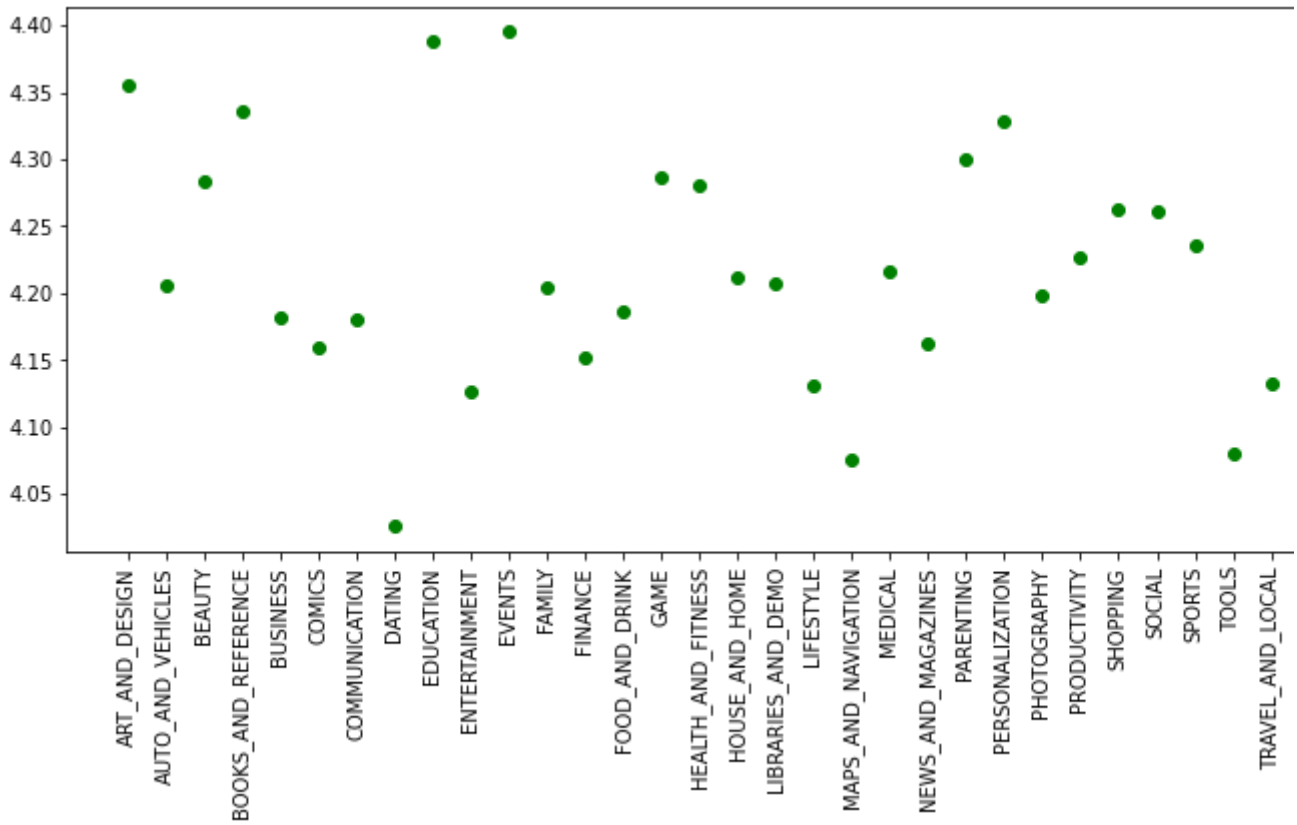
Data Visualization

```
grp = google_data.groupby('Category')
x = grp['Rating'].agg(np.mean)
y = grp['Price'].agg(np.sum)
z = grp['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
```

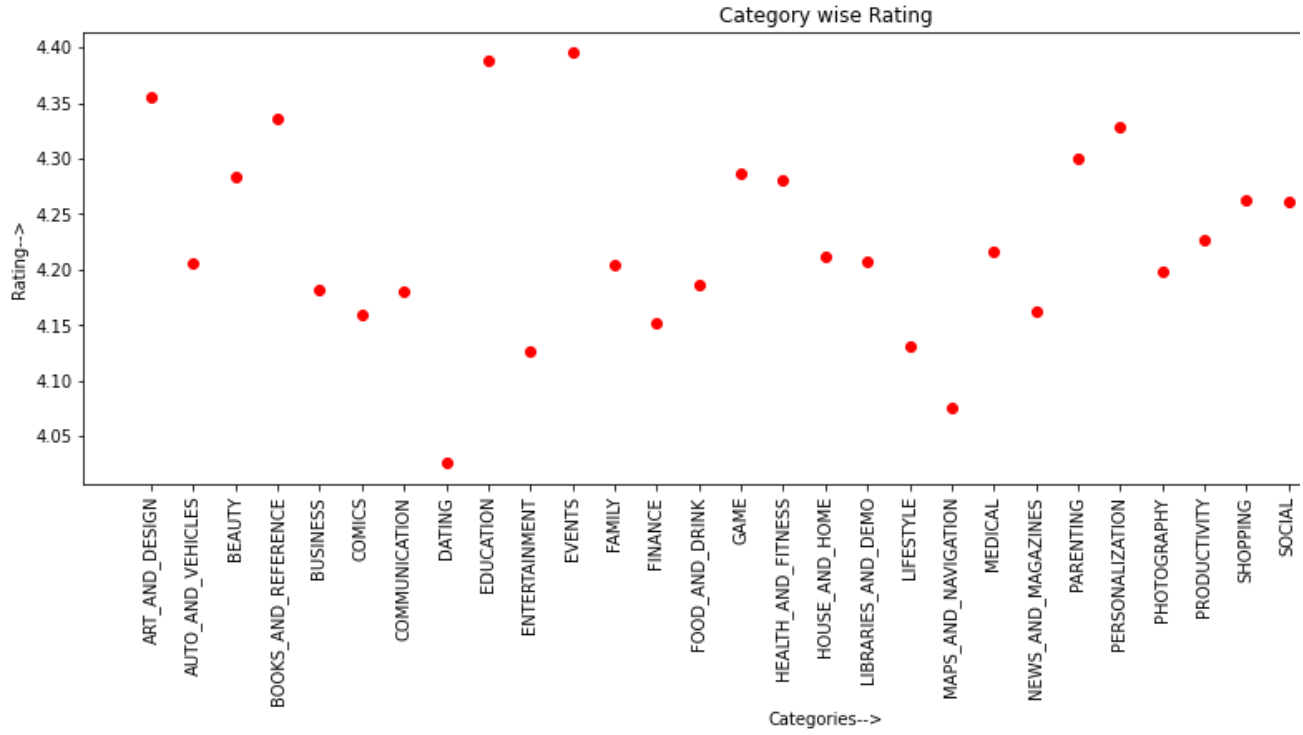
```
AUTO_AND_VEHICLES        13.47
BEAUTY                    0.00
BOOKS_AND_REFERENCE     119.77
BUSINESS                185.27
COMICS                    0.00
COMMUNICATION            83.14
DATING                   31.43
EDUCATION                17.96
ENTERTAINMENT             7.98
EVENTS                  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
```

```
plt.figure(figsize=(12,5))
plt.plot(x, "ro", color='g')
plt.xticks(rotation=90)
plt.show()
```

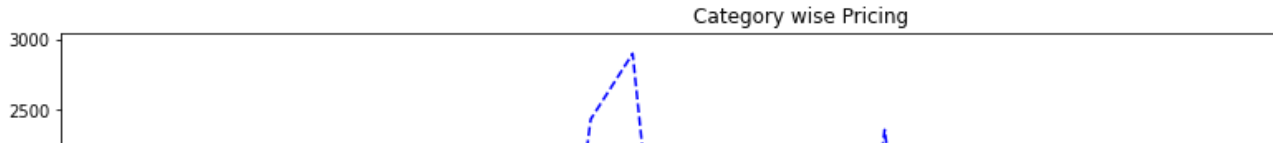



```
plt.figure(figsize=(16,5))
plt.plot(x,'ro', color='r')
plt.xticks(rotation=90)
plt.title('Category wise Rating')
plt.xlabel('Categories-->')
```

```
plt.ylabel('Rating-->')
plt.show()
```




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




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