1-praveenk-chintatejdeepreddy-code

March 28, 2023

Team 22 Name : Praveen K 211AI028 Chinta Tejdeep Reddy 211AI013

[121]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
sns.set()
```

[161]:
```
#Reading the csv file
df = pd.read_csv(r"C:\Users\Praveen\Desktop\ASSIGNMENTSEM-III\DS\2\mainDataset.
↪csv",encoding = "ISO-8859-1")
```

[123]:
```
#Getting the shape of the data frame
print(df.shape)
```

```
(961, 15)
```

[16]:
```
#Printing the dataframe
print(df)
```

```
                                                  Name  Rating Price Rs  RAM Gb  \
0              Realme Narzo 20 (Victory Blue, 128 GB)     4.5   11,499       4
1               Realme Narzo 20 (Victory Blue, 64 GB)     4.5   10,499       4
2              Realme Narzo 20 (Glory Sliver, 128 GB)     4.5   11,499       4
3               Realme Narzo 20 (Glory Sliver, 64 GB)     4.5   10,499       4
4                        POCO M2 (Pitch Black, 64 GB)     4.4   10,999       6
..                                                 …     …       …     …
956                    Vivo U20 (Racing Black, 64 GB)     4.4   14,499       4
957  Samsung Galaxy S7 Edge (Silver Titanium, 32 GB)     4.4   41,900       4
958               Asus Zenfone 2 Laser (Black, 16 GB)     4.1   12,999       3
959                           Homtom H5 (Gold, 32 GB)     3.5    6,499       3
960                    LG G7 ThinQ (Platinum, 64 GB)     4.4   53,000       4

     ROM Gb  Expandable GB  Size Cm  Size Inch  R1 Cam MP  R2 Cam MP  \
0       128          256.0    16.56       6.52       48.0        8.0
1        64          256.0    16.56       6.52       48.0        8.0
2       128          256.0    16.56       6.52       48.0        8.0
3        64          256.0    16.56       6.52       48.0        8.0
```

```
4        64          512.0    16.59       6.53       13.0        8.0
..      …            …      …         …        …         …
956      64            NaN    16.59       6.53       16.0        NaN
957      32          200.0    13.97       5.50       12.0        NaN
958      16          128.0    13.97       5.50       13.0        NaN
959      32            NaN    13.97       5.50       16.0        NaN
960      64            2.0    15.49       6.10       16.0       16.0

    R3 Cam MP  R4 Cam MP  Battery Mah  \
0           2        NaN         6000
1           2        NaN         6000
2           2        NaN         6000
3           2        NaN         6000
4           5        2.0         5000
..        …         …           …
956       NaN        NaN         5000
957       NaN        NaN         3600
958       NaN        NaN         3000
959       NaN        NaN         3300
960       NaN        NaN         3000

                                          Processor  \
0                       MediaTek Helio G85 Processor
1                       MediaTek Helio G85 Processor
2                       MediaTek Helio G85 Processor
3                       MediaTek Helio G85 Processor
4                       MediaTek Helio G80 Processor
..                                                 …
956                 Qualcomm Snapdragon 665 Processor
957                           Exynos 8890 Processor
958  Qualcomm Snapdragon 615 Octa Core 1.5GHz Proce…
959                               Quadcore Processor
960                 Qualcomm Snapdragon 845 Processor

                                                 Image
0    https://rukminim1.flixcart.com/image/312/312/k…
1    https://rukminim1.flixcart.com/image/312/312/k…
2    https://rukminim1.flixcart.com/image/312/312/k…
3    https://rukminim1.flixcart.com/image/312/312/k…
4    https://rukminim1.flixcart.com/image/312/312/k…
..                                                 …
956  https://img1a.flixcart.com/www/linchpin/fk-cp-…
957  https://img1a.flixcart.com/www/linchpin/fk-cp-…
958  https://img1a.flixcart.com/www/linchpin/fk-cp-…
959  https://img1a.flixcart.com/www/linchpin/fk-cp-…
960  https://img1a.flixcart.com/www/linchpin/fk-cp-…

[961 rows x 15 columns]
```

[17]:
```
#Print the first 5 Rows of the dataframe
df.head()
```

[17]:

	Name	Rating	Price Rs	RAM Gb	ROM Gb
0	Realme Narzo 20 (Victory Blue, 128 GB)	4.5	11,499	4	128
1	Realme Narzo 20 (Victory Blue, 64 GB)	4.5	10,499	4	64
2	Realme Narzo 20 (Glory Sliver, 128 GB)	4.5	11,499	4	128
3	Realme Narzo 20 (Glory Sliver, 64 GB)	4.5	10,499	4	64
4	POCO M2 (Pitch Black, 64 GB)	4.4	10,999	6	64

	Expandable GB	Size Cm	Size Inch	R1 Cam MP	R2 Cam MP	R3 Cam MP
0	256.0	16.56	6.52	48.0	8.0	2
1	256.0	16.56	6.52	48.0	8.0	2
2	256.0	16.56	6.52	48.0	8.0	2
3	256.0	16.56	6.52	48.0	8.0	2
4	512.0	16.59	6.53	13.0	8.0	5

	R4 Cam MP	Battery Mah	Processor
0	NaN	6000	MediaTek Helio G85 Processor
1	NaN	6000	MediaTek Helio G85 Processor
2	NaN	6000	MediaTek Helio G85 Processor
3	NaN	6000	MediaTek Helio G85 Processor
4	2.0	5000	MediaTek Helio G80 Processor

	Image
0	https://rukminim1.flixcart.com/image/312/312/k…
1	https://rukminim1.flixcart.com/image/312/312/k…
2	https://rukminim1.flixcart.com/image/312/312/k…
3	https://rukminim1.flixcart.com/image/312/312/k…
4	https://rukminim1.flixcart.com/image/312/312/k…

[36]:
```
#Describing the features of the dataset like count,mean,standard␣
 ↪deviation,min,max and quarantile ranges
df.describe()
```

[36]:

	Rating	RAM Gb	ROM Gb	Expandable GB	Size Cm
count	936.000000	961.000000	961.000000	715.000000	961.000000
mean	4.227137	4.500520	78.817898	298.641958	15.682352
std	0.368420	2.094201	65.408979	151.091947	1.162916
min	2.300000	2.000000	16.000000	1.000000	12.700000
25%	4.200000	3.000000	32.000000	256.000000	15.210000
50%	4.300000	4.000000	64.000000	256.000000	16.000000
75%	4.400000	6.000000	128.000000	512.000000	16.510000
max	5.000000	12.000000	512.000000	512.000000	17.780000

	Size Inch	R1 Cam MP	R2 Cam MP	R4 Cam MP	Battery Mah
count	961.000000	961.000000	619.000000	205.000000	961.000000

```
mean     6.173996   26.071176    6.707916    2.409756  4045.348595
std      0.457630   20.892832    5.038164    1.114918   864.281378
min      5.000000    5.000000    0.300000    2.000000  2200.000000
25%      5.990000   13.000000    2.000000    2.000000  3300.000000
50%      6.300000   13.000000    8.000000    2.000000  4000.000000
75%      6.500000   48.000000    8.000000    2.000000  5000.000000
max      7.000000  108.000000   48.000000    8.000000  6000.000000
```

[33]:
```
#Number of rows with rating greater than 4.8
dat = df[df['Rating']>=4.8]
```

[34]:
```
dat.head()
```

[34]:
```
                                    Name  Rating Price Rs  RAM Gb  ROM Gb  \
328   TECNO Spark 6 Air (Ocean Blue, 32 GB)     5.0    9,499       3      32
468  TECNO Spark 6 Air (Comet Black, 32 GB)     5.0    9,499       3      32

     Expandable GB  Size Cm  Size Inch  R1 Cam MP  R2 Cam MP R3 Cam MP  \
328            NaN    17.78        7.0       13.0        NaN       NaN
468            NaN    17.78        7.0       13.0        NaN       NaN

     R4 Cam MP  Battery Mah                    Processor  \
328        NaN         6000  MediaTek Helio A22 Processor
468        NaN         6000  MediaTek Helio A22 Processor

                                                 Image
328  https://img1a.flixcart.com/www/linchpin/fk-cp-…
468  https://img1a.flixcart.com/www/linchpin/fk-cp-…
```

[35]:
```
#Number of rows with rating greater than 4.6
dat = df[df['Rating']>=4.6]
dat.head()
```

[35]:
```
                                       Name  Rating Price Rs  RAM Gb  \
10        Realme Narzo 20 Pro (White Knight, 64 GB)     4.7   14,999       6
18         Realme Narzo 20 Pro (Black Ninja, 64 GB)     4.7   14,999       6
328           TECNO Spark 6 Air (Ocean Blue, 32 GB)     5.0    9,499       3
449  Samsung Galaxy Note 20 (Mystic Bronze, 256 GB)     4.6   77,999       8
464   Samsung Galaxy Note 20 (Mystic Green, 256 GB)     4.6   77,999       8

     ROM Gb  Expandable GB  Size Cm  Size Inch  R1 Cam MP  R2 Cam MP  \
10       64          256.0    16.51        6.5       48.0        8.0
18       64          256.0    16.51        6.5       48.0        8.0
328      32            NaN    17.78        7.0       13.0        NaN
449     256            NaN    17.02        6.7       64.0       12.0
464     256            NaN    17.02        6.7       64.0       12.0
```

```
    R3 Cam MP  R4 Cam MP  Battery Mah                    Processor  \
10          2        2.0         4500  MediaTek Helio G95 Processor
18          2        2.0         4500  MediaTek Helio G95 Processor
328       NaN        NaN         6000  MediaTek Helio A22 Processor
449        12        NaN         4300    Exynos Octa Core Processor
464        12        NaN         4300    Exynos Octa Core Processor

                                                  Image
10   https://img1a.flixcart.com/www/linchpin/fk-cp-…
18   https://img1a.flixcart.com/www/linchpin/fk-cp-…
328  https://img1a.flixcart.com/www/linchpin/fk-cp-…
449  https://img1a.flixcart.com/www/linchpin/fk-cp-…
464  https://img1a.flixcart.com/www/linchpin/fk-cp-…
```

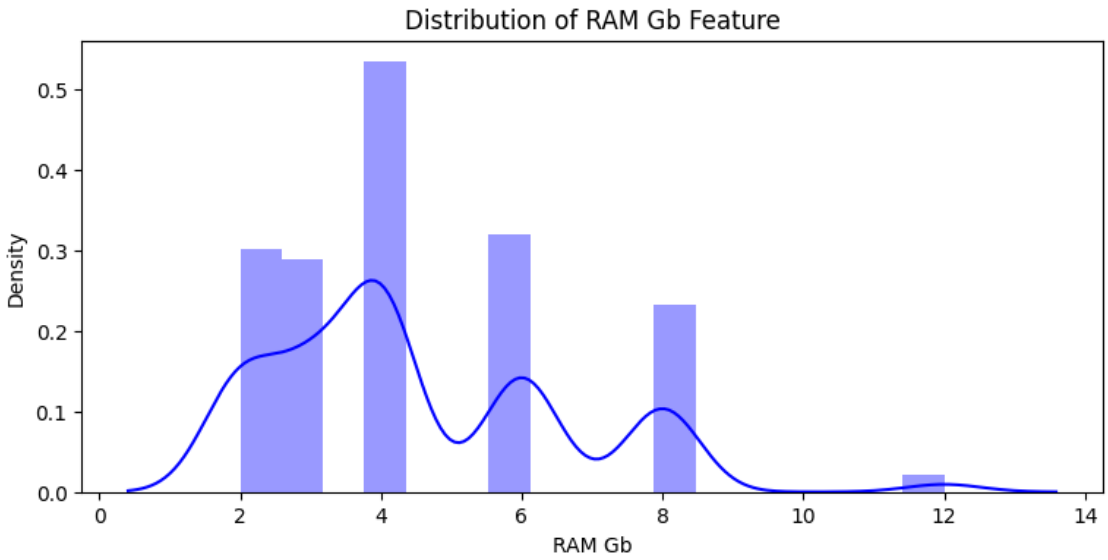
[22]:
```
#Counting the number of Null Values in the column feature
df.isnull().sum()
```

[22]:
```
Name              0
Rating           25
Price Rs          2
RAM Gb            0
ROM Gb            0
Expandable GB   246
Size Cm           0
Size Inch         0
R1 Cam MP         0
R2 Cam MP       342
R3 Cam MP       599
R4 Cam MP       756
Battery Mah       0
Processor         0
Image             0
dtype: int64
```

[27]:
```
#Density of Sales in Ram
f,ax = plt.subplots(figsize=(9,4))
sns.distplot(df["RAM Gb"],color="blue")
ax.set(xlabel="RAM Gb")
ax.set(title="Distribution of RAM Gb Feature")
plt.show()
```

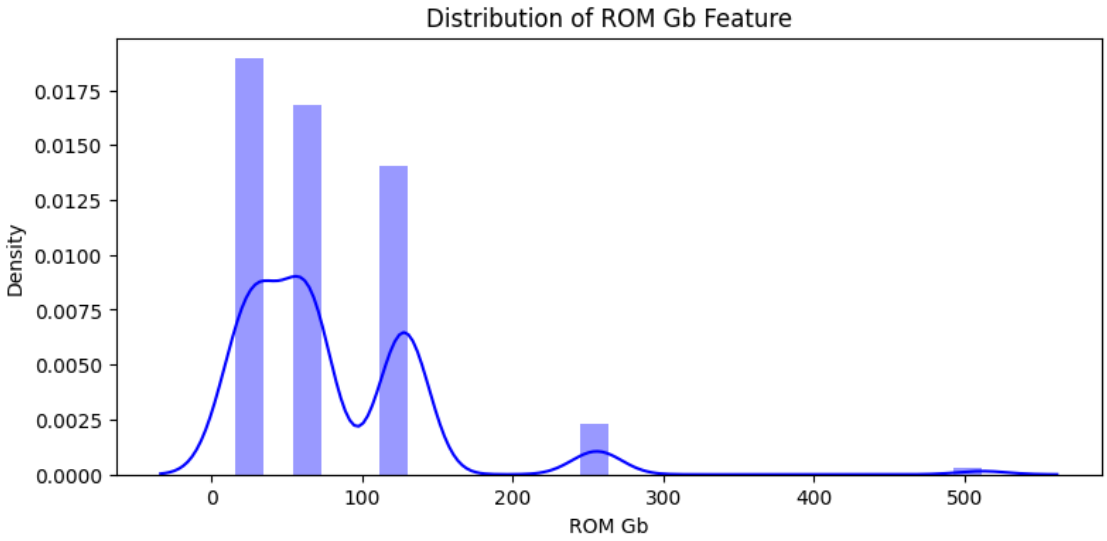
```
C:\Users\Praveen\AppData\Local\Temp\ipykernel_25236\2183096616.py:2:
UserWarning:

`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

  sns.distplot(df["RAM Gb"],color="blue")
```



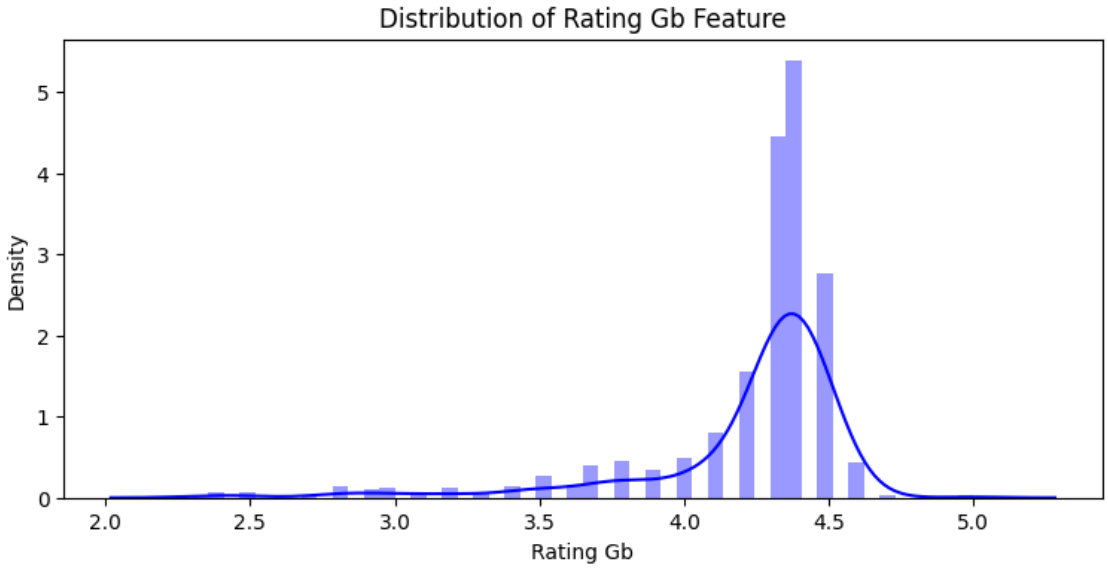
[28]:
```
#Density of Sales in Rom
f,ax = plt.subplots(figsize=(9,4))
sns.distplot(df["ROM Gb"],color="blue")
ax.set(xlabel="ROM Gb")
ax.set(title="Distribution of ROM Gb Feature")
plt.show()
```

```
C:\Users\Praveen\AppData\Local\Temp\ipykernel_25236\95507032.py:2: UserWarning:

`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

  sns.distplot(df["ROM Gb"],color="blue")
```



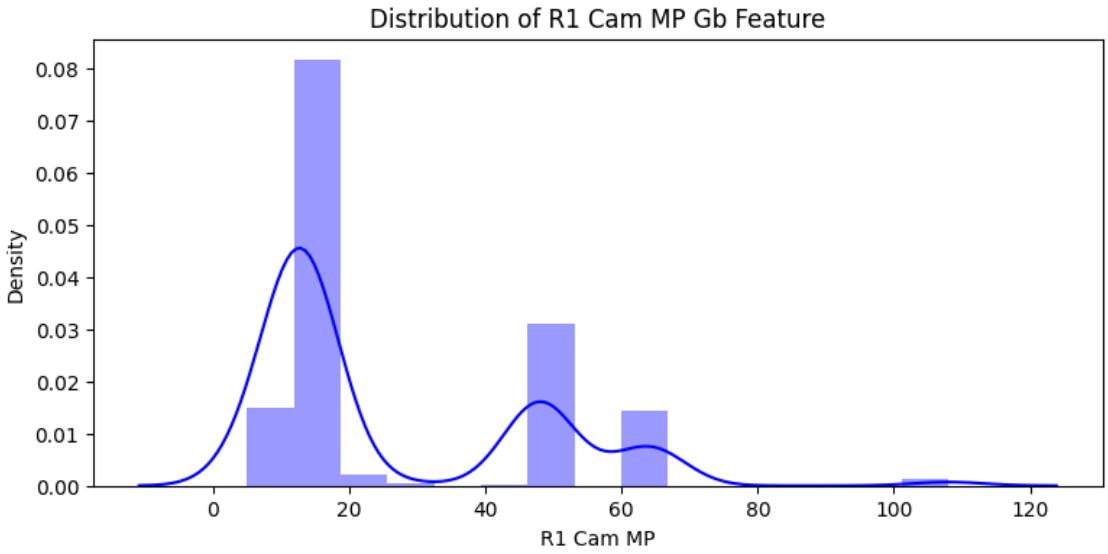
[29]:
```
#Density of Sales in Rating
f,ax = plt.subplots(figsize=(9,4))
sns.distplot(df["Rating"],color="blue")
ax.set(xlabel="Rating Gb")
ax.set(title="Distribution of Rating Gb Feature")
plt.show()
```

```
C:\Users\Praveen\AppData\Local\Temp\ipykernel_25236\1331219288.py:2:
UserWarning:

`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

  sns.distplot(df["Rating"],color="blue")
```



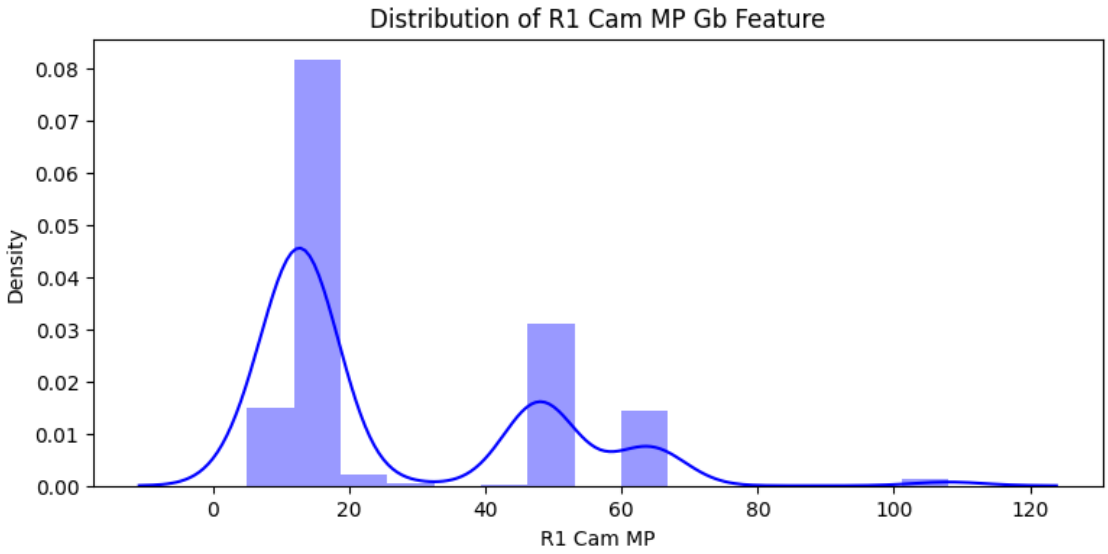
[30]:
```
#Density of Sales in R1 Cam MP
f,ax = plt.subplots(figsize=(9,4))
sns.distplot(df["R1 Cam MP"],color="blue")
ax.set(xlabel="R1 Cam MP")
ax.set(title="Distribution of R1 Cam MP Gb Feature")
plt.show()
```

```
C:\Users\Praveen\AppData\Local\Temp\ipykernel_25236\3219528217.py:2:
UserWarning:

`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

  sns.distplot(df["R1 Cam MP"],color="blue")
```



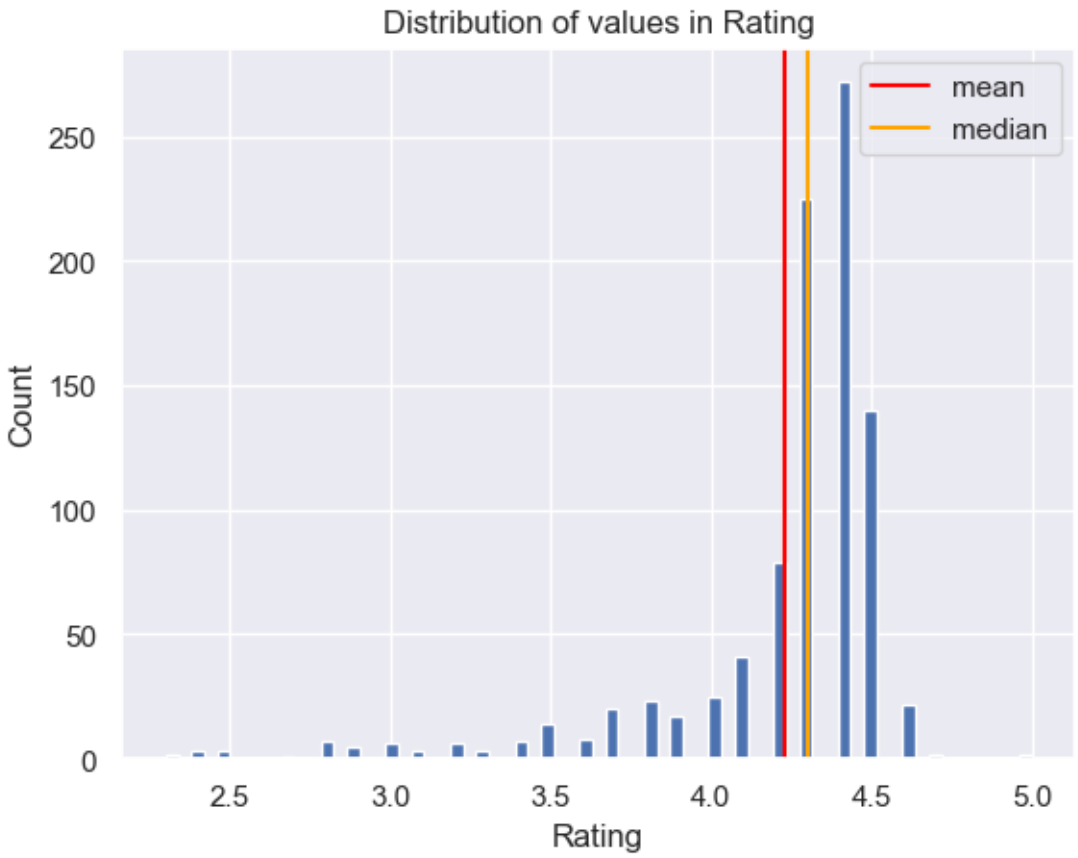
[31]:
```
#Density of Sales in R1 Cam MP
f,ax = plt.subplots(figsize=(9,4))
sns.distplot(df["R1 Cam MP"],color="blue")
ax.set(xlabel="R1 Cam MP")
ax.set(title="Distribution of R1 Cam MP Gb Feature")
plt.show()
```

```
C:\Users\Praveen\AppData\Local\Temp\ipykernel_25236\3219528217.py:2:
UserWarning:

`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

  sns.distplot(df["R1 Cam MP"],color="blue")
```

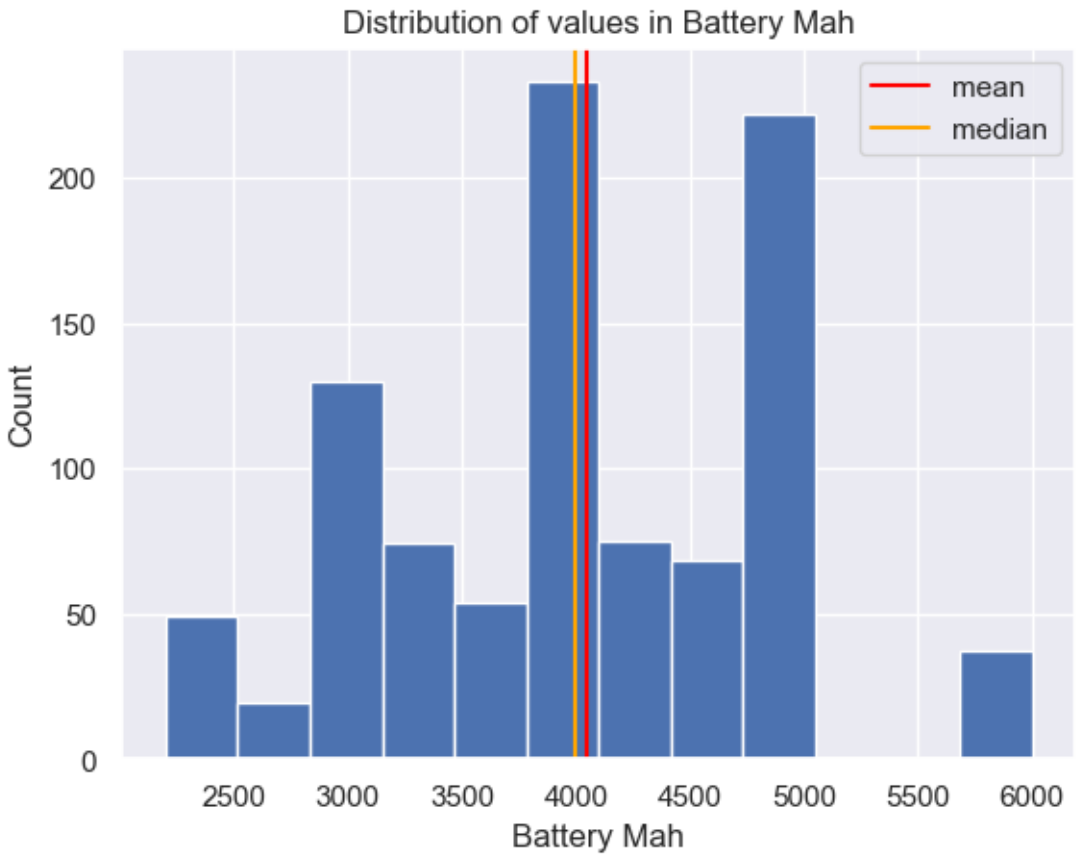


[48]:
```
#Creting a function for Histogram plot and a Mean, Median line for a given␣
↪feature and dataset
def plot_distribution(dataset, feature):
    plt.hist(dataset[feature], bins = "fd")

    plt.axvline(dataset[feature].mean(), color = "red", label = "mean")
    plt.axvline(dataset[feature].median(), color = "orange", label = "median")

    plt.xlabel(f"{feature}")
    plt.ylabel("Count")
    plt.legend()
    plt.title(f"Distribution of values in {feature}")
    plt.show()
```

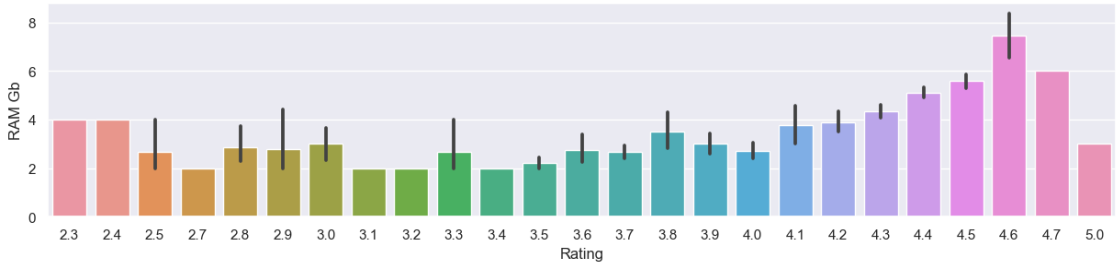
[49]:
```
#Plotting the the histogram plot of rating count and mean and meadian values
plot_distribution(df,"Rating")
```



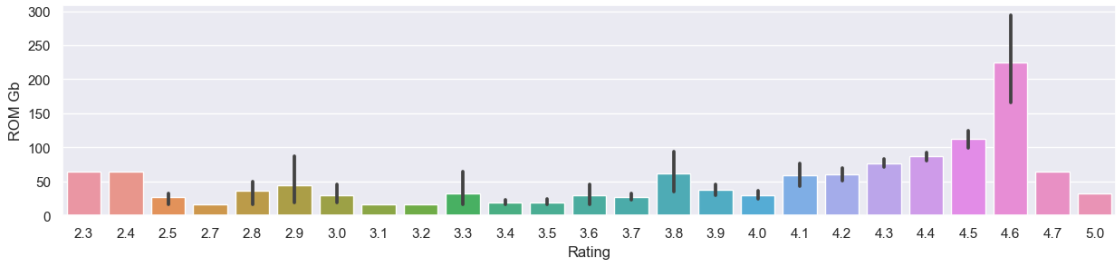
[50]:
```
#Plotting the the histogram plot of Battery MAH count and mean and meadian␣
↪values
plot_distribution(df,"Battery Mah")
```



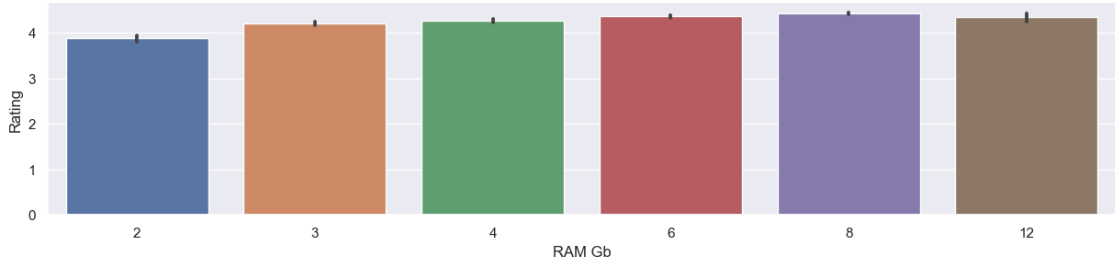
[57]:
```
#Plotting RAM vs Rating barplot
fig, ax = plt.subplots(figsize=(15,3))
ax=sns.barplot(x="Rating",y="RAM Gb",data=df)
```



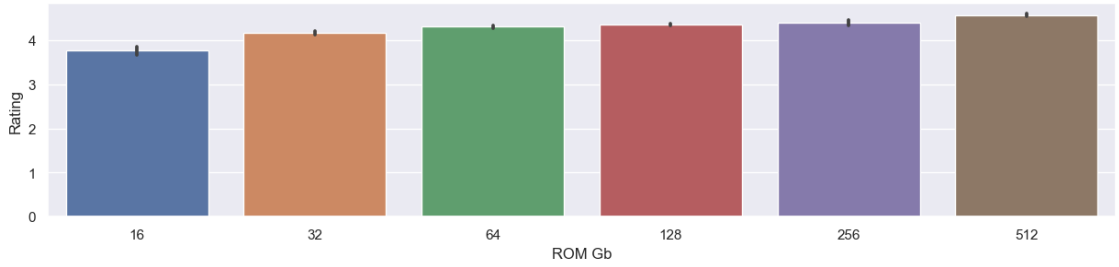
[60]:
```
#Plotting ROM vs Rating barplot
fig, ax = plt.subplots(figsize=(15,3))
ax=sns.barplot(x="Rating",y="ROM Gb",data=df)
```



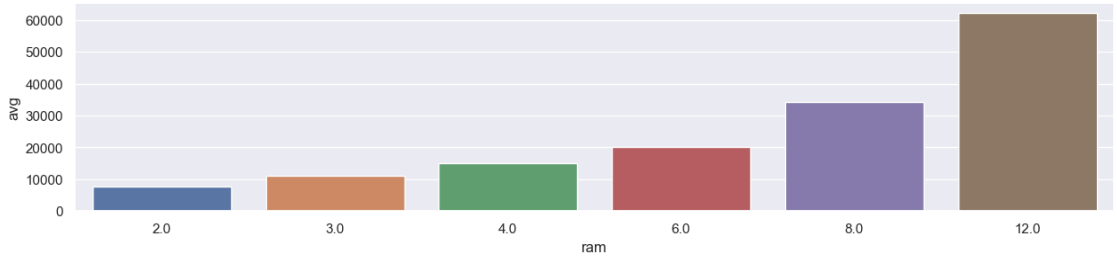
[61]:
```
#Plotting Rating VS RAM barplot
fig, ax = plt.subplots(figsize=(15,3))
ax=sns.barplot(x="RAM Gb",y="Rating",data=df)
```



[97]:
```
#Plotting Rating VS RAM barplot
fig, ax = plt.subplots(figsize=(15,3))
ax=sns.barplot(x="ROM Gb",y="Rating",data=df)
```

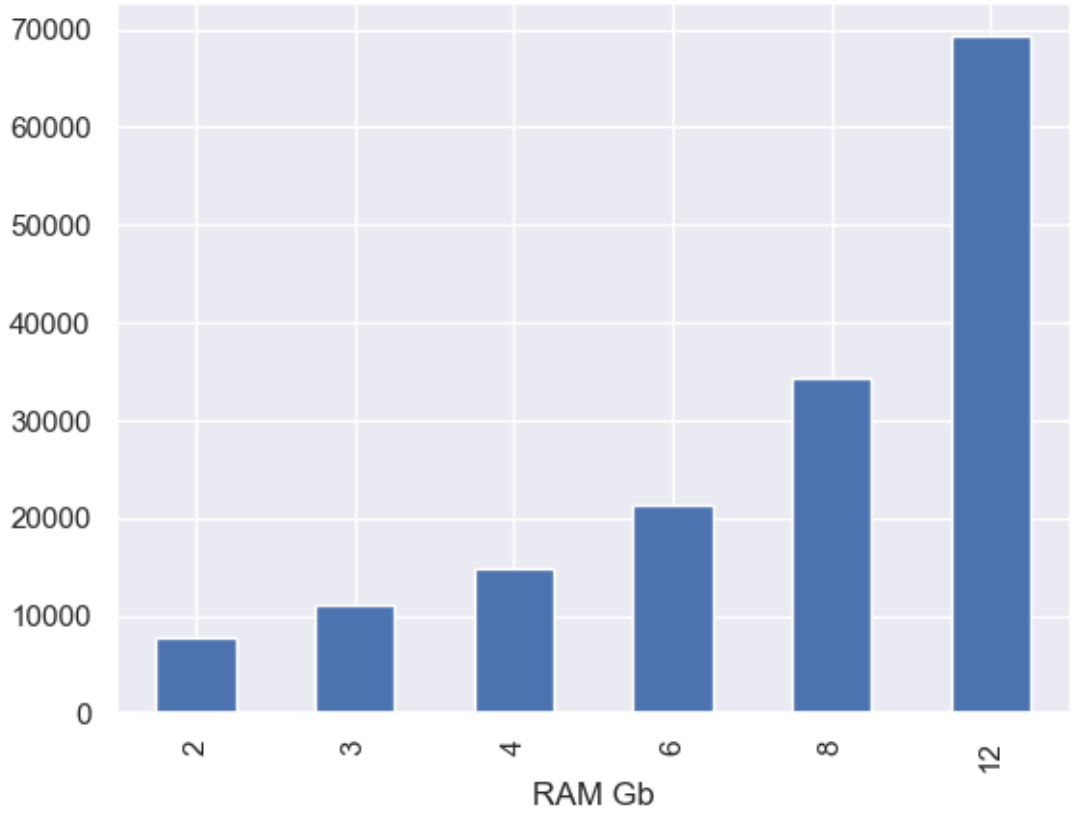


[92]:
```
#Plotting average prices and the RAM
fig, ax = plt.subplots(figsize=(15,3))
sns.barplot(x="ram",y="avg",data=df)
#ax.set_xlim(0,12000)
#ax.set_xticks(range(9000,12000,1000))
plt.show()
```

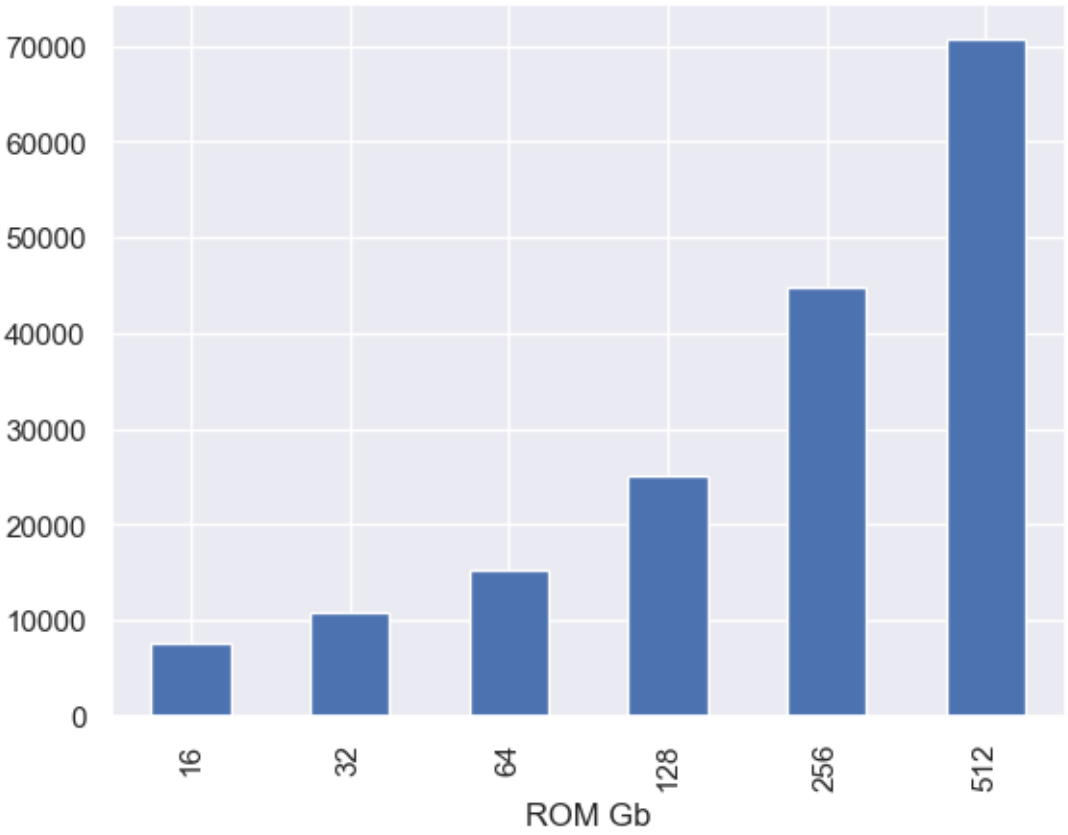


[151]:
```
#Plotting average price and RAM
fun = df.groupby(by="RAM Gb")["Price Rs"].mean().plot(kind="bar")

fun
#df
```

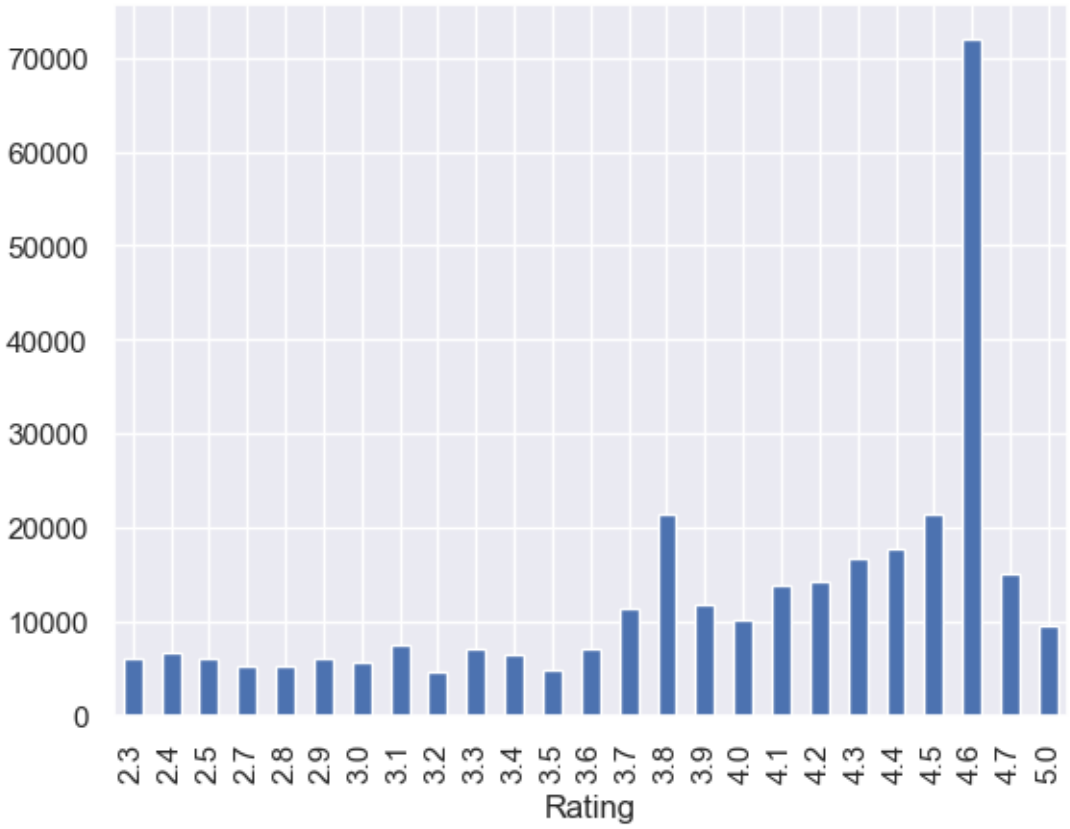
[151]: <AxesSubplot: xlabel='RAM Gb'>



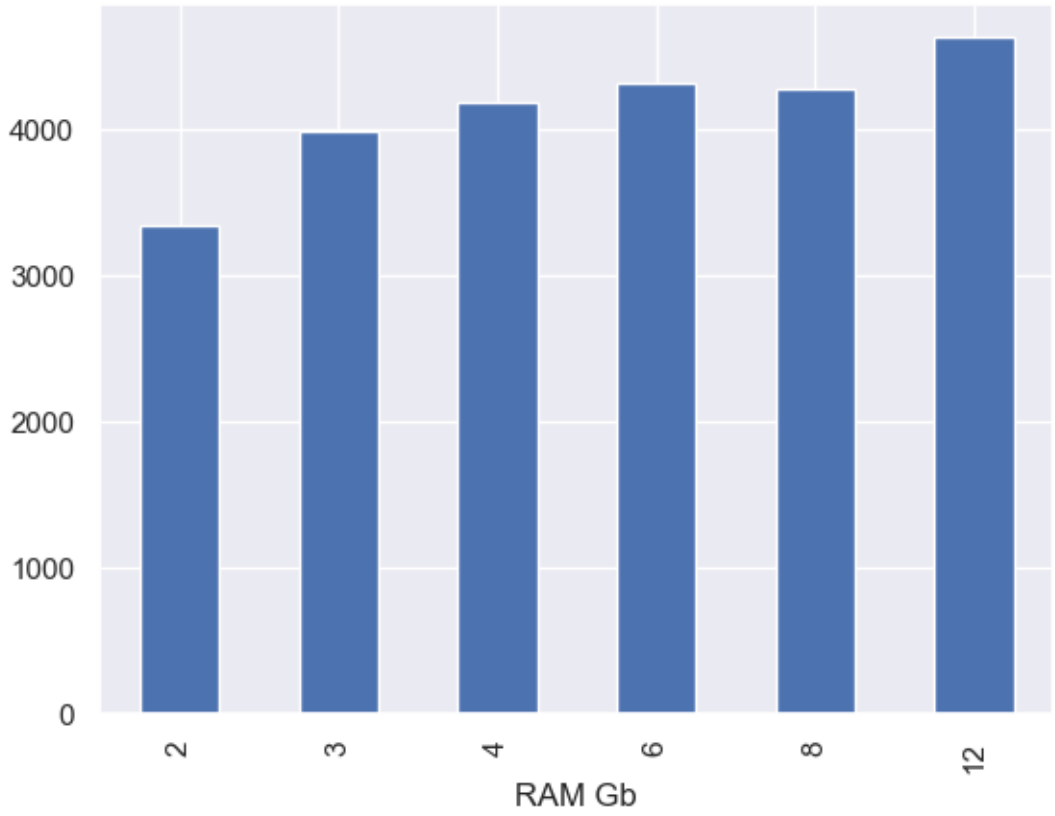
[155]:
```
#Plotting Prices vs ROM
fun = df.groupby(by="ROM Gb")["Price Rs"].mean().plot(kind="bar")
```



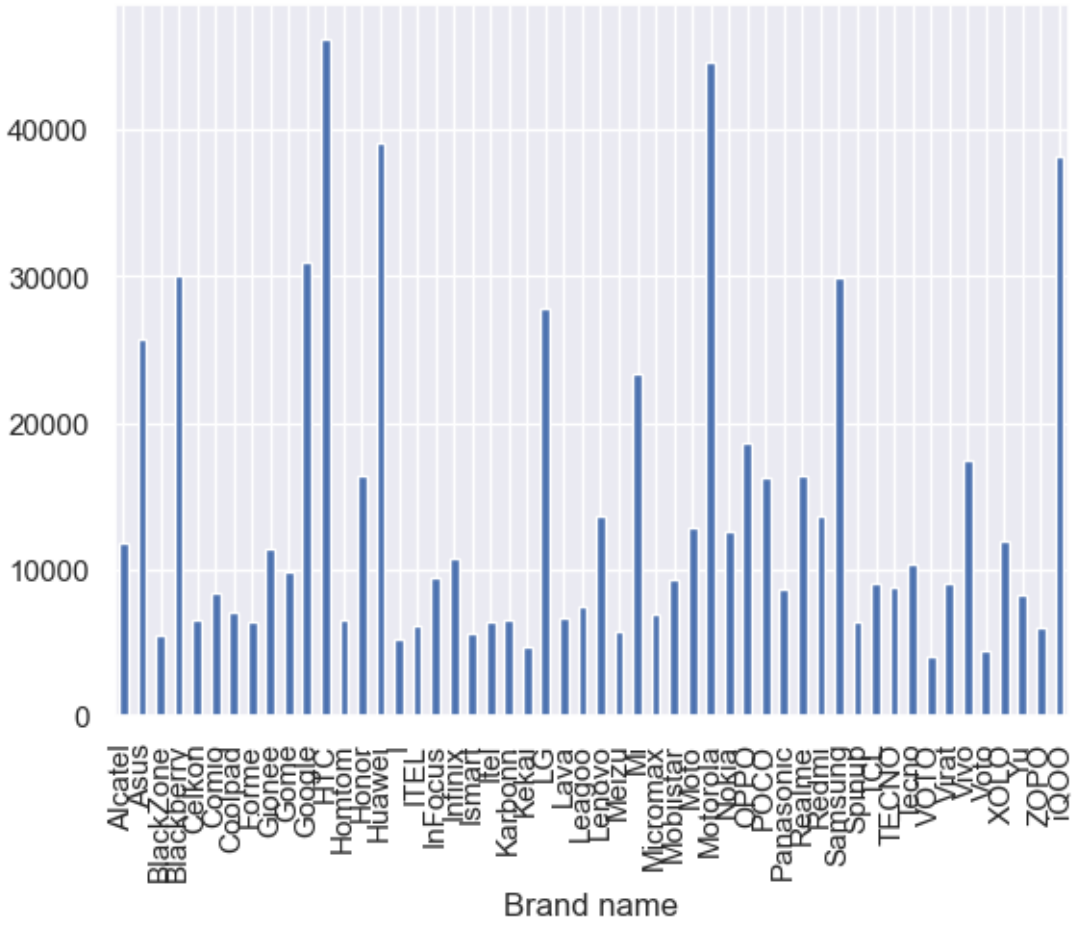
[157]:
```
#Plotting Prices vs Ratings
fun = df.groupby(by="Rating")["Price Rs"].mean().plot(kind="bar")
```



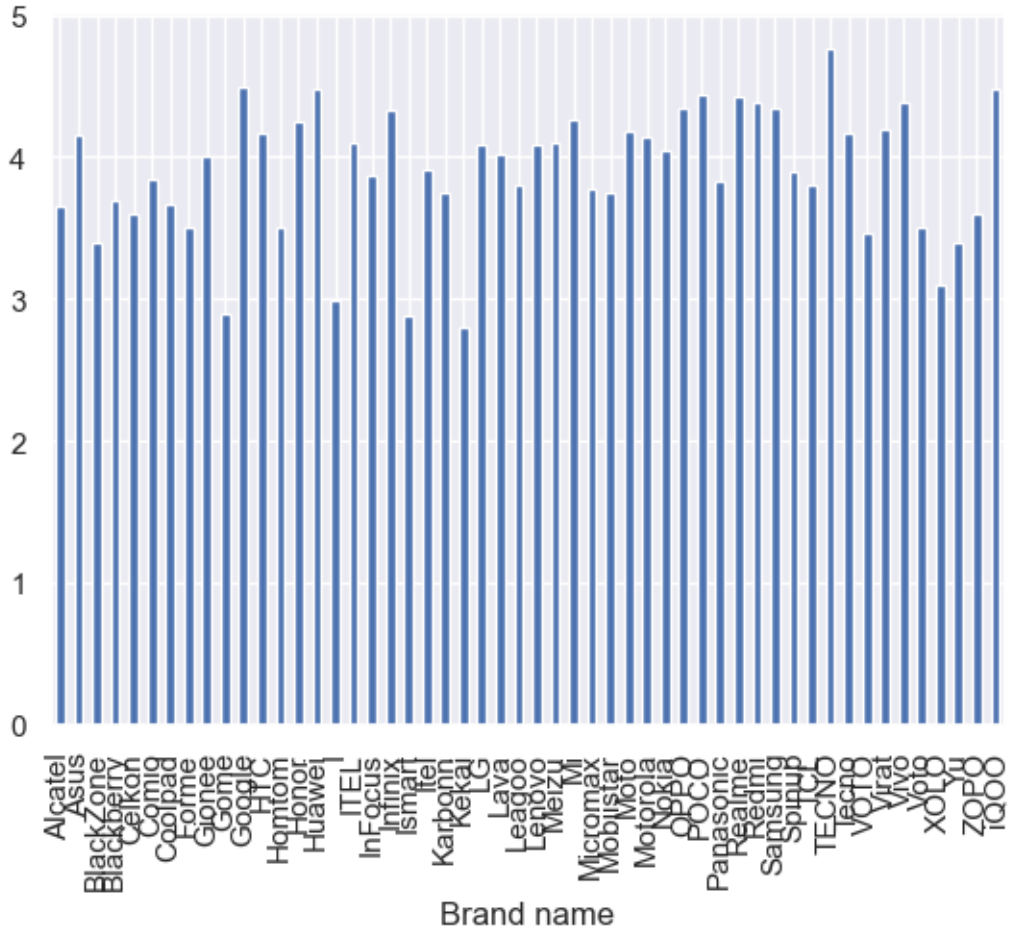
[160]:
```
#Plotting Batter MAH vs RAM for understanding the battery requirements for each␣
↪RAM
fun = df.groupby(by="RAM Gb")["Battery Mah"].mean().plot(kind="bar")
```



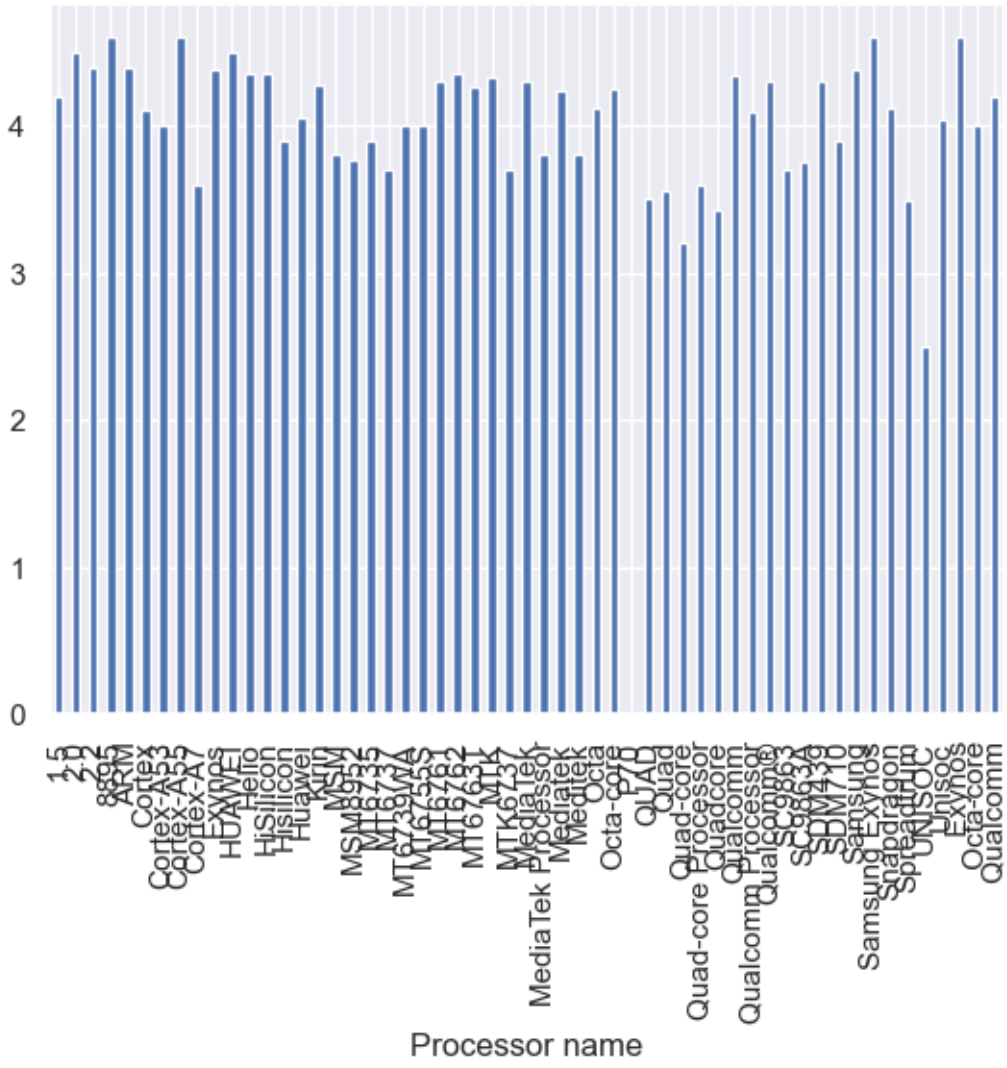
[163]:
```
#Plotting Average Price and Brand Name of phone
fun = df.groupby(by="Brand name")["Price Rs"].mean().plot(kind="bar")
```



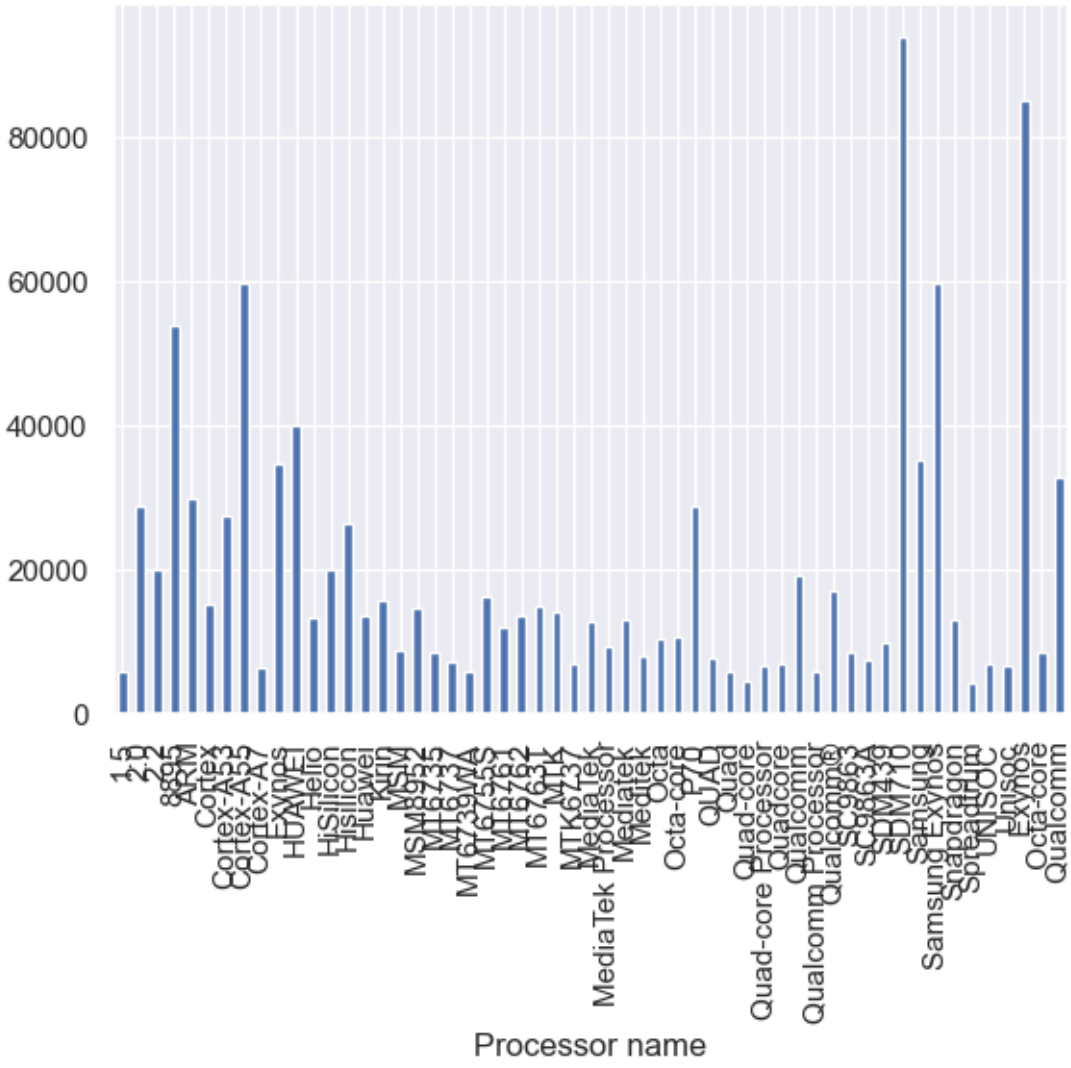
[165]:
```
#Plotting the Rating value and the brand name of the phone
fun = df.groupby(by="Brand name")["Rating"].mean().plot(kind="bar")
```



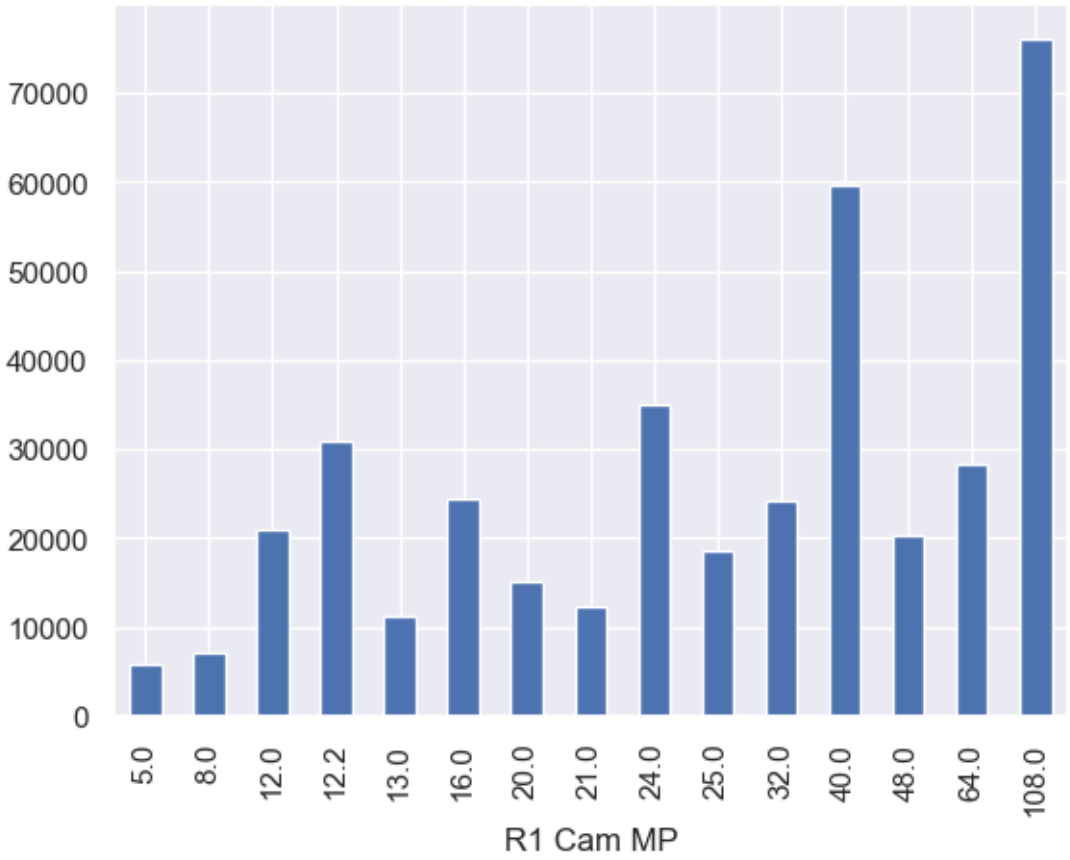
[166]:
```
#Plotting the Rating vs Processor Name
fun = df.groupby(by="Processor name")["Rating"].mean().plot(kind="bar")
```



[168]:
```
#Plotting the Price vs Processor Name
fun = df.groupby(by="Processor name")["Price Rs"].mean().plot(kind="bar")
```



[169]:
```
#Plotting the Cost vs R1 Cam MP
fun = df.groupby(by="R1 Cam MP")["Price Rs"].mean().plot(kind="bar")
```


70000
60000
50000
40000
30000
20000
10000
0
5.0
8.0
12.0
12.2
13.0
16.0
20.0
21.0
24.0
25.0
32.0
40.0
48.0
64.0
108.0
R1 Cam MP