

In [5]:

```
import numpy as np
import pandas as pd
import warnings
warnings.filterwarnings("ignore")
```

In [6]:

```
df = pd.read_csv("gplay.csv")
df
```

Out[6]:

	Unnamed: 0	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating
0	0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0	Everyone
1	1	Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone
2	2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8.7M	5,000,000+	Free	0	Everyone
3	3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50,000,000+	Free	0	Teen
4	4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100,000+	Free	0	Everyone
...
10836	10836	Sya9a Maroc - FR	FAMILY	4.5	38	53M	5,000+	Free	0	Everyone
10837	10837	Fr. Mike Schmitz Audio Teachings	FAMILY	5.0	4	3.6M	100+	Free	0	Everyone
10838	10838	Parkinson Exercices FR	MEDICAL	NaN	3	9.5M	1,000+	Free	0	Everyone
10839	10839	The SCP Foundation DB fr nn5n	BOOKS_AND_REFERENCE	4.5	114	Varies with device	1,000+	Free	0	Mature 17+
10840	10840	iHoroscope - 2018 Daily Horoscope & Astrology	LIFESTYLE	4.5	398307	19M	10,000,000+	Free	0	Everyone

10841 rows × 10 columns

In [7]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10841 entries, 0 to 10840
Data columns (total 10 columns):
#   Column              Non-Null Count  Dtype
---  -
0   Unnamed: 0          10841 non-null  int64
1   App                 10841 non-null  object
2   Category            10841 non-null  object
3   Rating              9367 non-null   float64
4   Reviews             10841 non-null  object
5   Size                10841 non-null  object
6   Installs            10841 non-null  object
7   Type                10840 non-null  object
8   Price               10841 non-null  object
9   Content Rating      10840 non-null  object
dtypes: float64(1), int64(1), object(8)
memory usage: 847.1+ KB
```

In [8]:

```
df.columns
```

Out[8]:

```
Index(['Unnamed: 0', 'App', 'Category', 'Rating', 'Reviews', 'Size',
      'Installs', 'Type', 'Price', 'Content Rating'],
      dtype='object')
```

In [9]:

```
df.drop("Unnamed: 0",axis=1,inplace=True)
```

In [10]:

```
df.head()
```

Out[10]:

	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
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8.7M	5,000,000+	Free	0	Everyone
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50,000,000+	Free	0	Teen
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100,000+	Free	0	Everyone

In [11]:

```
df.isnull().sum()
```

Out[11]:

App 0
Category 0
Rating 1474
Reviews 0
Size 0
Installs 0
Type 1
Price 0
Content Rating 1
dtype: int64

In [12]:

```
df.isna().sum()
```

Out[12]:

App 0
Category 0
Rating 1474
Reviews 0
Size 0
Installs 0
Type 1
Price 0
Content Rating 1
dtype: int64

In [13]:

```
rmean = df["Rating"].mean()
```

In [14]:

```
df["Rating"].fillna(rmean,inplace=True)
```

In [15]:

```
df.isnull().sum()
```

Out[15]:

App 0
Category 0
Rating 0
Reviews 0
Size 0
Installs 0
Type 1
Price 0
Content Rating 1
dtype: int64

In [16]:

```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10841 entries, 0 to 10840
Data columns (total 9 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   App              10841 non-null  object
1   Category         10841 non-null  object
2   Rating           10841 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
dtypes: float64(1), object(8)
memory usage: 762.4+ KB
```

In [17]:

```
df.dropna(inplace=True)
```

In [18]:

```
df.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 10839 entries, 0 to 10840
Data columns (total 9 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   App              10839 non-null  object
1   Category         10839 non-null  object
2   Rating           10839 non-null  float64
3   Reviews          10839 non-null  object
4   Size             10839 non-null  object
5   Installs         10839 non-null  object
6   Type             10839 non-null  object
7   Price            10839 non-null  object
8   Content Rating   10839 non-null  object
dtypes: float64(1), object(8)
memory usage: 846.8+ KB
```

In [19]:

```
df["Reviews"]=df["Reviews"].astype(float)
```

In [20]:

```
df.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 10839 entries, 0 to 10840
Data columns (total 9 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   App              10839 non-null  object
1   Category         10839 non-null  object
2   Rating           10839 non-null  float64
3   Reviews          10839 non-null  float64
4   Size             10839 non-null  object
5   Installs         10839 non-null  object
6   Type             10839 non-null  object
7   Price            10839 non-null  object
8   Content Rating   10839 non-null  object
dtypes: float64(2), object(7)
memory usage: 846.8+ KB
```

In [21]:

```
df.head()
```

Out[21]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159.0	19M	10,000+	Free	0	Everyone
1	Coloring book moana	ART_AND_DESIGN	3.9	967.0	14M	500,000+	Free	0	Everyone
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510.0	8.7M	5,000,000+	Free	0	Everyone
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644.0	25M	50,000,000+	Free	0	Teen
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967.0	2.8M	100,000+	Free	0	Everyone

In [22]:

```
df["Reviews"].value_counts()
```

Out[22]:

```
0.0      595
1.0      272
2.0      214
3.0      175
4.0      137
...
4272.0     1
5517.0     1
4057.0     1
459797.0   1
398307.0   1
Name: Reviews, Length: 6001, dtype: int64
```

In [23]:

```
df["Size"].value_counts()
```

Out[23]:

```
Varies with device    1694
11M                    198
12M                    196
14M                    194
13M                    191
...
430k                    1
429k                    1
200k                    1
460k                    1
619k                    1
Name: Size, Length: 461, dtype: int64
```

In [24]:

```
def clean_size(x):
    if "k" in x:
        x=x.replace("k", "")
        return float(x)/1024

    elif "M" in x:
        x=x.replace("M", "")
        return float(x)

    elif "Varies with device" in x:
        return np.nan

    else:
        return float(x)
```

In [25]:

```
df["Size"]=df["Size"].map(clean_size)
```

In [26]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 10839 entries, 0 to 10840
Data columns (total 9 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   App              10839 non-null  object
1   Category         10839 non-null  object
2   Rating           10839 non-null  float64
3   Reviews          10839 non-null  float64
4   Size             9145 non-null   float64
5   Installs         10839 non-null  object
6   Type             10839 non-null  object
7   Price            10839 non-null  object
8   Content Rating   10839 non-null  object
dtypes: float64(3), object(6)
memory usage: 846.8+ KB
```

In [27]:

```
df.head()
```

Out[27]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159.0	19.0	10,000+	Free	0	Everyone
1	Coloring book moana	ART_AND_DESIGN	3.9	967.0	14.0	500,000+	Free	0	Everyone
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510.0	8.7	5,000,000+	Free	0	Everyone
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644.0	25.0	50,000,000+	Free	0	Teen
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967.0	2.8	100,000+	Free	0	Everyone

In [29]:

```
df["Size"].value_counts()
```

Out[29]:

```
11.000000    198
12.000000    196
14.000000    194
13.000000    191
15.000000    184
...
0.419922      1
0.418945      1
0.195312      1
0.449219      1
0.604492      1
Name: Size, Length: 459, dtype: int64
```

In [30]:

```
smean =df['Size'].mean()
```

In [31]:

```
df["Size"].fillna(smean,inplace=True)
```

In [32]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 10839 entries, 0 to 10840
Data columns (total 9 columns):
#   Column          Non-Null Count  Dtype
---  -
0   App              10839 non-null  object
1   Category         10839 non-null  object
2   Rating           10839 non-null  float64
3   Reviews          10839 non-null  float64
4   Size             10839 non-null  float64
5   Installs         10839 non-null  object
6   Type             10839 non-null  object
7   Price            10839 non-null  object
8   Content Rating   10839 non-null  object
dtypes: float64(3), object(6)
memory usage: 846.8+ KB
```

In [33]:

```
df["Installs"]=df["Installs"].str.replace(",","")
df["Installs"]=df["Installs"].str.replace("+","")
```

In [44]:

```
df.head()
```

Out[44]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159.0	19.0	10000	Free	0	Everyone
1	Coloring book moana	ART_AND_DESIGN	3.9	967.0	14.0	500000	Free	0	Everyone
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510.0	8.7	5000000	Free	0	Everyone
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644.0	25.0	50000000	Free	0	Teen
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967.0	2.8	100000	Free	0	Everyone

In [45]:

```
df["Installs"]=df["Installs"].astype(int)
```

In [46]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 10839 entries, 0 to 10840
Data columns (total 9 columns):
#   Column          Non-Null Count  Dtype
---  -
0   App              10839 non-null  object
1   Category         10839 non-null  object
2   Rating           10839 non-null  float64
3   Reviews          10839 non-null  float64
4   Size             10839 non-null  float64
5   Installs         10839 non-null  int32
6   Type             10839 non-null  object
7   Price            10839 non-null  object
8   Content Rating   10839 non-null  object
dtypes: float64(3), int32(1), object(5)
memory usage: 804.5+ KB
```

In [47]:

```
df.head()
```

Out[47]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159.0	19.0	10000	Free	0	Everyone
1	Coloring book moana	ART_AND_DESIGN	3.9	967.0	14.0	500000	Free	0	Everyone
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510.0	8.7	5000000	Free	0	Everyone
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644.0	25.0	50000000	Free	0	Teen
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967.0	2.8	100000	Free	0	Everyone

In [41]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 10839 entries, 0 to 10840
Data columns (total 9 columns):
#   Column          Non-Null Count  Dtype
---  -
0   App              10839 non-null  object
1   Category         10839 non-null  object
2   Rating           10839 non-null  float64
3   Reviews          10839 non-null  float64
4   Size             10839 non-null  float64
5   Installs         10839 non-null  int32
6   Type             10839 non-null  object
7   Price            10839 non-null  object
8   Content Rating   10839 non-null  object
dtypes: float64(3), int32(1), object(5)
memory usage: 804.5+ KB
```

In [48]:

```
df["Type"].value_counts()
```

Out[48]:

```
Free    10039
Paid      800
Name: Type, dtype: int64
```

In [49]:

```
df["Price"].value_counts()
```

Out[49]:

```
0          10039
$0.99       148
$2.99       129
$1.99        73
$4.99        72
...
$19.90         1
$1.75          1
$14.00          1
$4.85           1
$1.04           1
Name: Price, Length: 92, dtype: int64
```

In [55]:

```
df["Price"]=df["Price"].str.replace("$", "")
```

In [56]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 10839 entries, 0 to 10840
Data columns (total 9 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   App              10839 non-null  object
1   Category         10839 non-null  object
2   Rating           10839 non-null  float64
3   Reviews          10839 non-null  float64
4   Size             10839 non-null  float64
5   Installs         10839 non-null  int32
6   Type             10839 non-null  object
7   Price            10839 non-null  object
8   Content Rating   10839 non-null  object
dtypes: float64(3), int32(1), object(5)
memory usage: 804.5+ KB
```

In [57]:

```
df["Price"]=df["Price"].astype("float")
```

In [58]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 10839 entries, 0 to 10840
Data columns (total 9 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   App              10839 non-null  object
1   Category         10839 non-null  object
2   Rating           10839 non-null  float64
3   Reviews          10839 non-null  float64
4   Size             10839 non-null  float64
5   Installs         10839 non-null  int32
6   Type             10839 non-null  object
7   Price            10839 non-null  float64
8   Content Rating   10839 non-null  object
dtypes: float64(4), int32(1), object(4)
memory usage: 804.5+ KB
```

In [59]:

```
df.head()
```

Out[59]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159.0	19.0	10000	Free	0.0	Everyone
1	Coloring book moana	ART_AND_DESIGN	3.9	967.0	14.0	500000	Free	0.0	Everyone
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510.0	8.7	5000000	Free	0.0	Everyone
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644.0	25.0	50000000	Free	0.0	Teen
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967.0	2.8	100000	Free	0.0	Everyone

In [54]:

```
#Find the Highest Rating App
df["Rating"].max()
```

Out[54]:

5.0

In [47]:

```
df[df.Rating==5.0]
```

Out[47]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating
329	Hojiboy Tojiboyev Life Hacks	COMICS	5.0	15.0	37.0	1000	Free	0.0	Everyone
612	American Girls Mobile Numbers	DATING	5.0	5.0	4.4	1000	Free	0.0	Mature 17+
615	Awake Dating	DATING	5.0	2.0	70.0	100	Free	0.0	Mature 17+
633	Spine- The dating app	DATING	5.0	5.0	9.3	500	Free	0.0	Teen
636	Girls Live Talk - Free Text and Video Chat	DATING	5.0	6.0	5.0	100	Free	0.0	Mature 17+
...
10721	Mad Dash Fo' Cash	GAME	5.0	14.0	16.0	100	Free	0.0	Everyone
10742	GKPB FP Online Church	LIFESTYLE	5.0	32.0	7.9	1000	Free	0.0	Everyone
10776	Monster Ride Pro	GAME	5.0	1.0	24.0	10	Free	0.0	Everyone
10820	Fr. Daoud Lamei	FAMILY	5.0	22.0	8.6	1000	Free	0.0	Teen
10837	Fr. Mike Schmitz Audio Teachings	FAMILY	5.0	4.0	3.6	100	Free	0.0	Everyone

274 rows × 9 columns

In []:

```
#Display all The Records having Max Rating...

#Display only App Name,Category and Rating whose Rating value Is 5.0

#Display ALL records of Free App

#Display ALL records of Free App
```

In [68]:

```
#Display all The Records having Max Rating...
df[df.Rating==5.0]
```

Out[68]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating
329	Hojiboy Tojiboyev Life Hacks	COMICS	5.0	15.0	37.0	1000	Free	0.0	Everyone
612	American Girls Mobile Numbers	DATING	5.0	5.0	4.4	1000	Free	0.0	Mature 17+
615	Awake Dating	DATING	5.0	2.0	70.0	100	Free	0.0	Mature 17+
633	Spine- The dating app	DATING	5.0	5.0	9.3	500	Free	0.0	Teen
636	Girls Live Talk - Free Text and Video Chat	DATING	5.0	6.0	5.0	100	Free	0.0	Mature 17+
...
10721	Mad Dash Fo' Cash	GAME	5.0	14.0	16.0	100	Free	0.0	Everyone
10742	GKPB FP Online Church	LIFESTYLE	5.0	32.0	7.9	1000	Free	0.0	Everyone
10776	Monster Ride Pro	GAME	5.0	1.0	24.0	10	Free	0.0	Everyone
10820	Fr. Daoud Lamei	FAMILY	5.0	22.0	8.6	1000	Free	0.0	Teen
10837	Fr. Mike Schmitz Audio Teachings	FAMILY	5.0	4.0	3.6	100	Free	0.0	Everyone

274 rows × 9 columns

In [64]:

```
#Display only App Name,Category and Rating whose Rating value Is 5.0
df[["App","Category","Rating"]][df.Rating==5.0]
```

Out[64]:

	App	Category	Rating
329	Hojiboy Tojiboyev Life Hacks	COMICS	5.0
612	American Girls Mobile Numbers	DATING	5.0
615	Awake Dating	DATING	5.0
633	Spine- The dating app	DATING	5.0
636	Girls Live Talk - Free Text and Video Chat	DATING	5.0
...
10721	Mad Dash Fo' Cash	GAME	5.0
10742	GKPB FP Online Church	LIFESTYLE	5.0
10776	Monster Ride Pro	GAME	5.0
10820	Fr. Daoud Lamei	FAMILY	5.0
10837	Fr. Mike Schmitz Audio Teachings	FAMILY	5.0

274 rows × 3 columns

In [66]:

```
#Display ALL records of Free App
df[df.Type=="Free"]
```

Out[66]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.100000	159.0	19.000000	10000	Free	0.0	Everyone
1	Coloring book moana	ART_AND_DESIGN	3.900000	967.0	14.000000	500000	Free	0.0	Everyone
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.700000	87510.0	8.700000	5000000	Free	0.0	Everyone
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.500000	215644.0	25.000000	50000000	Free	0.0	Teen
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.300000	967.0	2.800000	100000	Free	0.0	Everyone
...
10836	Sya9a Maroc - FR	FAMILY	4.500000	38.0	53.000000	5000	Free	0.0	Everyone
10837	Fr. Mike Schmitz Audio Teachings	FAMILY	5.000000	4.0	3.600000	100	Free	0.0	Everyone
10838	Parkinson Exercices FR	MEDICAL	4.193338	3.0	9.500000	1000	Free	0.0	Everyone
10839	The SCP Foundation DB fr nn5n	BOOKS_AND_REFERENCE	4.500000	114.0	21.516165	1000	Free	0.0	Mature 17+
10840	iHoroscope - 2018 Daily Horoscope & Astrology	LIFESTYLE	4.500000	398307.0	19.000000	10000000	Free	0.0	Everyone

10039 rows × 9 columns

In [67]:

```
#Display ALL records of Free App
df[df.Type=="Paid"]
```

Out[67]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating
234	TurboScan: scan documents and receipts in PDF	BUSINESS	4.700000	11442.0	6.800000	100000	Paid	4.99	Everyone
235	Tiny Scanner Pro: PDF Doc Scan	BUSINESS	4.800000	10295.0	39.000000	100000	Paid	4.99	Everyone
290	TurboScan: scan documents and receipts in PDF	BUSINESS	4.700000	11442.0	6.800000	100000	Paid	4.99	Everyone
291	Tiny Scanner Pro: PDF Doc Scan	BUSINESS	4.800000	10295.0	39.000000	100000	Paid	4.99	Everyone
427	Puffin Browser Pro	COMMUNICATION	4.000000	18247.0	21.516165	100000	Paid	3.99	Everyone
...
10735	FP VoiceBot	FAMILY	4.193338	17.0	0.153320	100	Paid	0.99	Mature 17+
10760	Fast Tract Diet	HEALTH_AND_FITNESS	4.400000	35.0	2.400000	1000	Paid	7.99	Everyone
10782	Trine 2: Complete Story	GAME	3.800000	252.0	11.000000	10000	Paid	16.99	Teen
10785	sugar, sugar	FAMILY	4.200000	1405.0	9.500000	10000	Paid	1.20	Everyone
10798	Word Search Tab 1 FR	FAMILY	4.193338	0.0	0.996094	50	Paid	1.04	Everyone

800 rows × 9 columns

In []: