

In [1]:

```
import pandas as pd
data=pd.read_csv("https://raw.githubusercontent.com/sunnysavita10/Statistics-With-Python")
```

In [2]:

```
data.head()
```

Out[2]:

	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 [3]:

```
data.shape
```

Out[3]:

(10841, 13)

In [4]:

```
data.describe()
```

Out[4]:

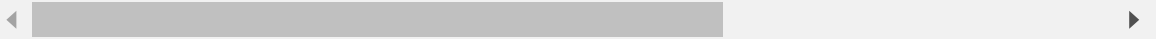
	Rating
count	9367.000000
mean	4.193338
std	0.537431
min	1.000000
25%	4.000000
50%	4.300000
75%	4.500000
max	19.000000

In [5]:

```
data.tail()
```

Out[5]:

	App	Category	Rating	Reviews	Size	Installs	Type	Pr
10836	Sya9a Maroc - FR	FAMILY	4.5	38	53M	5,000+	Free	
10837	Fr. Mike Schmitz Audio Teachings	FAMILY	5.0	4	3.6M	100+	Free	
10838	Parkinson Exercices FR	MEDICAL	NaN	3	9.5M	1,000+	Free	
10839	The SCP Foundation DB fr nn5n	BOOKS_AND_REFERENCE	4.5	114	Varies with device	1,000+	Free	
10840	iHoroscope - 2018 Daily Horoscope & Astrology	LIFESTYLE	4.5	398307	19M	10,000,000+	Free	



In [6]:

```
data.sample()
```

Out[6]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	U
4953	Ad Detector	TOOLS	3.8	208	730k	50,000+	Free	0	Everyone	Tools	J

In [7]:

```
data.sample(10)
```

Out[7]:

	App	Category	Rating	Reviews	Size	Installs	Type	Pri
10827	Fr Agnel Ambarnath	FAMILY	4.2	117	13M	5,000+	Free	
10820	Fr. Daoud Lamei	FAMILY	5.0	22	8.6M	1,000+	Free	
2671	Horn, free country requirements	SHOPPING	4.5	37186	7.9M	1,000,000+	Free	
10283	Bank FD Interest Calculator	FINANCE	4.5	166	3.5M	10,000+	Free	
4806	Gangster Town	GAME	4.1	74842	99M	5,000,000+	Free	
9801	OpenGL ES Extensions - The OpenGL Utility	TOOLS	4.2	513	2.0M	50,000+	Free	
9130	3G to 4G Converter - Simulator	TOOLS	4.2	268	2.4M	10,000+	Free	
10129	EZ Inspections	PRODUCTIVITY	2.8	160	7.6M	10,000+	Free	
7538	CM Launcher 3D Pro	PERSONALIZATION	4.7	23802	173k	100,000+	Paid	\$4.
10724	English To Shona Dictionary	BOOKS_AND_REFERENCE	4.6	60	3.7M	10,000+	Free	

In [8]:

```
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
```

In [9]:

```
data.describe(include='all').T
```

Out[9]:

	count	unique	top	freq	mean	std	min	25%	50%	75%	max
App	10841	9660	ROBLOX	9	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Category	10841	34	FAMILY	1972	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Rating	9367.0	NaN	NaN	NaN	4.193338	0.537431	1.0	4.0	4.3	4.5	19.0
Reviews	10841	6002	0	596	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Size	10841	462	Varies with device	1695	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Installs	10841	22	1,000,000+	1579	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Type	10840	3	Free	10039	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Price	10841	93	0	10040	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Content Rating	10840	6	Everyone	8714	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Genres	10841	120	Tools	842	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Last Updated	10841	1378	August 3, 2018	326	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Current Ver	10833	2832	Varies with device	1459	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Android Ver	10838	33	4.1 and up	2451	NaN	NaN	NaN	NaN	NaN	NaN	NaN

In [10]:

```
data.duplicated()
```

Out[10]:

```
0      False
1      False
2      False
3      False
4      False
...
10836   False
10837   False
10838   False
10839   False
10840   False
Length: 10841, dtype: bool
```

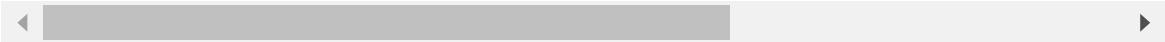
In [11]:

```
data[data.duplicated()]
```

Out[11]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Co Ri
229	Quick PDF Scanner + OCR FREE	BUSINESS	4.2	80805	Varies with device	5,000,000+	Free	0	Ever
236	Box	BUSINESS	4.2	159872	Varies with device	10,000,000+	Free	0	Ever
239	Google My Business	BUSINESS	4.4	70991	Varies with device	5,000,000+	Free	0	Ever
256	ZOOM Cloud Meetings	BUSINESS	4.4	31614	37M	10,000,000+	Free	0	Ever
261	join.me - Simple Meetings	BUSINESS	4.0	6989	Varies with device	1,000,000+	Free	0	Ever
...
8643	Wunderlist: To-Do List & Tasks	PRODUCTIVITY	4.6	404610	Varies with device	10,000,000+	Free	0	Ever
8654	TickTick: To Do List with Reminder, Day Planner	PRODUCTIVITY	4.6	25370	Varies with device	1,000,000+	Free	0	Ever
8658	ColorNote Notepad Notes	PRODUCTIVITY	4.6	2401017	Varies with device	100,000,000+	Free	0	Ever
10049	Airway Ex - Intubate. Anesthetize. Train.	MEDICAL	4.3	123	86M	10,000+	Free	0	Ever
10768	AAFP	MEDICAL	3.8	63	24M	10,000+	Free	0	Ever

483 rows × 13 columns



In [12]:

```
data["Reviews"].head
```

Out[12]:

```
<bound method NDFrame.head of 0          159
1           967
2        87510
3       215644
4           967
...
10836         38
10837         4
10838         3
10839        114
10840    398307
Name: Reviews, Length: 10841, dtype: object>
```

In [13]:

```
data["Reviews"].head(20)
```

Out[13]:

```
0         159
1         967
2        87510
3       215644
4         967
5         167
6         178
7        36815
8        13791
9         121
10       13880
11        8788
12       44829
13        4326
14        1518
15         55
16        3632
17         27
18       194216
19       224399
Name: Reviews, dtype: object
```

In [14]:

```
data["Reviews"].shape
```

Out[14]:

```
(10841,)
```

In [15]:

```
data["Reviews"].dtype
```

Out[15]:

```
dtype('O')
```

In [16]:

```
data["Reviews"].str
```

Out[16]:

```
<pandas.core.strings.accessor.StringMethods at 0x1bcefd56fa0>
```

In [17]:

```
data["Reviews"].str.isnumeric()
```

Out[17]:

```
0      True
1      True
2      True
3      True
4      True
...
10836  True
10837  True
10838  True
10839  True
10840  True
Name: Reviews, Length: 10841, dtype: bool
```


In [18]:

```
data[data["Reviews"].str.isnumeric()].head(10)
```

Out[18]:

	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
5	Paper flowers instructions	ART_AND_DESIGN	4.4	167	5.6M	50,000+	Free	0	Everyone
6	Smoke Effect Photo Maker - Smoke Editor	ART_AND_DESIGN	3.8	178	19M	50,000+	Free	0	Everyone
7	Infinite Painter	ART_AND_DESIGN	4.1	36815	29M	1,000,000+	Free	0	Everyone
8	Garden Coloring Book	ART_AND_DESIGN	4.4	13791	33M	1,000,000+	Free	0	Everyone
9	Kids Paint Free - Drawing Fun	ART_AND_DESIGN	4.7	121	3.1M	10,000+	Free	0	Everyone

In [19]:

```
data[~data.Reviews.str.isnumeric()]
```

Out[19]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	(
10472	Life Made WI-Fi Touchscreen Photo Frame		1.9	19.0	3.0M	1,000+	Free	0	Everyone	NaN

In [20]:

```
data.head()
```

Out[20]:

	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 [21]:

```
data_copy=data.copy()
```

In [22]:

```
data_copy.head()
```

Out[22]:

	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 [23]:

```
data_copy.index[10472]
```

Out[23]:

10472

In [24]:

```
data_copy=data_copy.drop(data_copy.index[10472])
```

In [25]:

```
data_copy.shape
```

Out[25]:

(10840, 13)

In [26]:

```
data.shape
```

Out[26]:

```
(10841, 13)
```

In [27]:

```
data_copy['Reviews'].dtype
```

Out[27]:

```
dtype('O')
```

In [28]:

```
data_copy['Reviews']=data_copy['Reviews'].astype('int')
```

In [29]:

```
data_copy['Reviews'].dtype
```

Out[29]:

```
dtype('int32')
```

In [30]:

```
data_copy.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 10840 entries, 0 to 10840
Data columns (total 13 columns):
 #   Column                Non-Null Count  Dtype  
---  -
 0   App                   10840 non-null  object 
 1   Category              10840 non-null  object 
 2   Rating                9366 non-null   float64
 3   Reviews               10840 non-null  int32  
 4   Size                  10840 non-null  object 
 5   Installs              10840 non-null  object 
 6   Type                  10839 non-null  object 
 7   Price                 10840 non-null  object 
 8   Content Rating        10840 non-null  object 
 9   Genres                10840 non-null  object 
10   Last Updated          10840 non-null  object 
11   Current Ver           10832 non-null  object 
12   Android Ver           10838 non-null  object 
dtypes: float64(1), int32(1), object(11)
memory usage: 1.1+ MB
```

In [31]:

```
data_copy["Size"]
```

Out[31]:

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

In [32]:

```
data_copy.size
```

Out[32]:

```
140920
```

In [33]:

```
data_copy["Size"].dtype
```

Out[33]:

```
dtype('O')
```

In [34]:

```
data_copy["Size"].unique()
```

Out[34]:

```

array(['19M', '14M', '8.7M', '25M', '2.8M', '5.6M', '29M', '33M', '3.1M',
      '28M', '12M', '20M', '21M', '37M', '2.7M', '5.5M', '17M', '39M',
      '31M', '4.2M', '7.0M', '23M', '6.0M', '6.1M', '4.6M', '9.2M',
      '5.2M', '11M', '24M', 'Varies with device', '9.4M', '15M', '10M',
      '1.2M', '26M', '8.0M', '7.9M', '56M', '57M', '35M', '54M', '201k',
      '3.6M', '5.7M', '8.6M', '2.4M', '27M', '2.5M', '16M', '3.4M',
      '8.9M', '3.9M', '2.9M', '38M', '32M', '5.4M', '18M', '1.1M',
      '2.2M', '4.5M', '9.8M', '52M', '9.0M', '6.7M', '30M', '2.6M',
      '7.1M', '3.7M', '22M', '7.4M', '6.4M', '3.2M', '8.2M', '9.9M',
      '4.9M', '9.5M', '5.0M', '5.9M', '13M', '73M', '6.8M', '3.5M',
      '4.0M', '2.3M', '7.2M', '2.1M', '42M', '7.3M', '9.1M', '55M',
      '23k', '6.5M', '1.5M', '7.5M', '51M', '41M', '48M', '8.5M', '46M',
      '8.3M', '4.3M', '4.7M', '3.3M', '40M', '7.8M', '8.8M', '6.6M',
      '5.1M', '61M', '66M', '79k', '8.4M', '118k', '44M', '695k', '1.6M',
      '6.2M', '18k', '53M', '1.4M', '3.0M', '5.8M', '3.8M', '9.6M',
      '45M', '63M', '49M', '77M', '4.4M', '4.8M', '70M', '6.9M', '9.3M',
      '10.0M', '8.1M', '36M', '84M', '97M', '2.0M', '1.9M', '1.8M',
      '5.3M', '47M', '556k', '526k', '76M', '7.6M', '59M', '9.7M', '78M',
      '72M', '43M', '7.7M', '6.3M', '334k', '34M', '93M', '65M', '79M',
      '100M', '58M', '50M', '68M', '64M', '67M', '60M', '94M', '232k',
      '99M', '624k', '95M', '8.5k', '41k', '292k', '11k', '80M', '1.7M',
      '74M', '62M', '69M', '75M', '98M', '85M', '82M', '96M', '87M',
      '71M', '86M', '91M', '81M', '92M', '83M', '88M', '704k', '862k',
      '899k', '378k', '266k', '375k', '1.3M', '975k', '980k', '4.1M',
      '89M', '696k', '544k', '525k', '920k', '779k', '853k', '720k',
      '713k', '772k', '318k', '58k', '241k', '196k', '857k', '51k',
      '953k', '865k', '251k', '930k', '540k', '313k', '746k', '203k',
      '26k', '314k', '239k', '371k', '220k', '730k', '756k', '91k',
      '293k', '17k', '74k', '14k', '317k', '78k', '924k', '902k', '818k',
      '81k', '939k', '169k', '45k', '475k', '965k', '90M', '545k', '61k',
      '283k', '655k', '714k', '93k', '872k', '121k', '322k', '1.0M',
      '976k', '172k', '238k', '549k', '206k', '954k', '444k', '717k',
      '210k', '609k', '308k', '705k', '306k', '904k', '473k', '175k',
      '350k', '383k', '454k', '421k', '70k', '812k', '442k', '842k',
      '417k', '412k', '459k', '478k', '335k', '782k', '721k', '430k',
      '429k', '192k', '200k', '460k', '728k', '496k', '816k', '414k',
      '506k', '887k', '613k', '243k', '569k', '778k', '683k', '592k',
      '319k', '186k', '840k', '647k', '191k', '373k', '437k', '598k',
      '716k', '585k', '982k', '222k', '219k', '55k', '948k', '323k',
      '691k', '511k', '951k', '963k', '25k', '554k', '351k', '27k',
      '82k', '208k', '913k', '514k', '551k', '29k', '103k', '898k',
      '743k', '116k', '153k', '209k', '353k', '499k', '173k', '597k',
      '809k', '122k', '411k', '400k', '801k', '787k', '237k', '50k',
      '643k', '986k', '97k', '516k', '837k', '780k', '961k', '269k',
      '20k', '498k', '600k', '749k', '642k', '881k', '72k', '656k',
      '601k', '221k', '228k', '108k', '940k', '176k', '33k', '663k',
      '34k', '942k', '259k', '164k', '458k', '245k', '629k', '28k',
      '288k', '775k', '785k', '636k', '916k', '994k', '309k', '485k',
      '914k', '903k', '608k', '500k', '54k', '562k', '847k', '957k',
      '688k', '811k', '270k', '48k', '329k', '523k', '921k', '874k',
      '981k', '784k', '280k', '24k', '518k', '754k', '892k', '154k',
      '860k', '364k', '387k', '626k', '161k', '879k', '39k', '970k',
      '170k', '141k', '160k', '144k', '143k', '190k', '376k', '193k',
      '246k', '73k', '658k', '992k', '253k', '420k', '404k', '470k',
      '226k', '240k', '89k', '234k', '257k', '861k', '467k', '157k',
      '44k', '676k', '67k', '552k', '885k', '1020k', '582k', '619k'],
      dtype=object)

```

In [35]:

```
data_copy["Size"]=data_copy["Size"].str.replace("M","000")
```

In [36]:

```
data_copy["Size"]=data_copy["Size"].str.replace("k","")
```


In [37]:

```
data_copy["Size"].unique()
```

Out[37]:

```

array(['19000', '14000', '8.7000', '25000', '2.8000', '5.6000', '29000',
      '33000', '3.1000', '28000', '12000', '20000', '21000', '37000',
      '2.7000', '5.5000', '17000', '39000', '31000', '4.2000', '7.0000',
      '23000', '6.0000', '6.1000', '4.6000', '9.2000', '5.2000', '11000',
      '24000', 'Varies with device', '9.4000', '15000', '10000',
      '1.2000', '26000', '8.0000', '7.9000', '56000', '57000', '35000',
      '54000', '201', '3.6000', '5.7000', '8.6000', '2.4000', '27000',
      '2.5000', '16000', '3.4000', '8.9000', '3.9000', '2.9000', '38000',
      '32000', '5.4000', '18000', '1.1000', '2.2000', '4.5000', '9.8000',
      '52000', '9.0000', '6.7000', '30000', '2.6000', '7.1000', '3.7000',
      '22000', '7.4000', '6.4000', '3.2000', '8.2000', '9.9000',
      '4.9000', '9.5000', '5.0000', '5.9000', '13000', '73000', '6.8000',
      '3.5000', '4.0000', '2.3000', '7.2000', '2.1000', '42000',
      '7.3000', '9.1000', '55000', '23', '6.5000', '1.5000', '7.5000',
      '51000', '41000', '48000', '8.5000', '46000', '8.3000', '4.3000',
      '4.7000', '3.3000', '40000', '7.8000', '8.8000', '6.6000',
      '5.1000', '61000', '66000', '79', '8.4000', '118', '44000', '695',
      '1.6000', '6.2000', '18', '53000', '1.4000', '3.0000', '5.8000',
      '3.8000', '9.6000', '45000', '63000', '49000', '77000', '4.4000',
      '4.8000', '70000', '6.9000', '9.3000', '10.0000', '8.1000',
      '36000', '84000', '97000', '2.0000', '1.9000', '1.8000', '5.3000',
      '47000', '556', '526', '76000', '7.6000', '59000', '9.7000',
      '78000', '72000', '43000', '7.7000', '6.3000', '334', '34000',
      '93000', '65000', '79000', '100000', '58000', '50000', '68000',
      '64000', '67000', '60000', '94000', '232', '99000', '624', '95000',
      '8.5', '41', '292', '11', '80000', '1.7000', '74000', '62000',
      '69000', '75000', '98000', '85000', '82000', '96000', '87000',
      '71000', '86000', '91000', '81000', '92000', '83000', '88000',
      '704', '862', '899', '378', '266', '375', '1.3000', '975', '980',
      '4.1000', '89000', '696', '544', '525', '920', '779', '853', '720',
      '713', '772', '318', '58', '241', '196', '857', '51', '953', '865',
      '251', '930', '540', '313', '746', '203', '26', '314', '239',
      '371', '220', '730', '756', '91', '293', '17', '74', '14', '317',
      '78', '924', '902', '818', '81', '939', '169', '45', '475', '965',
      '90000', '545', '61', '283', '655', '714', '93', '872', '121',
      '322', '1.0000', '976', '172', '238', '549', '206', '954', '444',
      '717', '210', '609', '308', '705', '306', '904', '473', '175',
      '350', '383', '454', '421', '70', '812', '442', '842', '417',
      '412', '459', '478', '335', '782', '721', '430', '429', '192',
      '200', '460', '728', '496', '816', '414', '506', '887', '613',
      '243', '569', '778', '683', '592', '319', '186', '840', '647',
      '191', '373', '437', '598', '716', '585', '982', '222', '219',
      '55', '948', '323', '691', '511', '951', '963', '25', '554', '351',
      '27', '82', '208', '913', '514', '551', '29', '103', '898', '743',
      '116', '153', '209', '353', '499', '173', '597', '809', '122',
      '411', '400', '801', '787', '237', '50', '643', '986', '97', '516',
      '837', '780', '961', '269', '20', '498', '600', '749', '642',
      '881', '72', '656', '601', '221', '228', '108', '940', '176', '33',
      '663', '34', '942', '259', '164', '458', '245', '629', '28', '288',
      '775', '785', '636', '916', '994', '309', '485', '914', '903',
      '608', '500', '54', '562', '847', '957', '688', '811', '270', '48',
      '329', '523', '921', '874', '981', '784', '280', '24', '518',
      '754', '892', '154', '860', '364', '387', '626', '161', '879',
      '39', '970', '170', '141', '160', '144', '143', '190', '376',
      '193', '246', '73', '658', '992', '253', '420', '404', '470',
      '226', '240', '89', '234', '257', '861', '467', '157', '44', '676',
      '67', '552', '885', '1020', '582', '619'], dtype=object)

```

In [38]:

```
import numpy as np
```

In [39]:

```
data_copy["Size"]=data_copy["Size"].str.replace('Varies with device',str(np.nan))
```

In [40]:

```
data_copy["Size"].unique()
```

Out[40]:

```

array(['19000', '14000', '8.7000', '25000', '2.8000', '5.6000', '29000',
      '33000', '3.1000', '28000', '12000', '20000', '21000', '37000',
      '2.7000', '5.5000', '17000', '39000', '31000', '4.2000', '7.0000',
      '23000', '6.0000', '6.1000', '4.6000', '9.2000', '5.2000', '11000',
      '24000', 'nan', '9.4000', '15000', '10000', '1.2000', '26000',
      '8.0000', '7.9000', '56000', '57000', '35000', '54000', '201',
      '3.6000', '5.7000', '8.6000', '2.4000', '27000', '2.5000', '16000',
      '3.4000', '8.9000', '3.9000', '2.9000', '38000', '32000', '5.4000',
      '18000', '1.1000', '2.2000', '4.5000', '9.8000', '52000', '9.0000',
      '6.7000', '30000', '2.6000', '7.1000', '3.7000', '22000', '7.4000',
      '6.4000', '3.2000', '8.2000', '9.9000', '4.9000', '9.5000',
      '5.0000', '5.9000', '13000', '73000', '6.8000', '3.5000', '4.0000',
      '2.3000', '7.2000', '2.1000', '42000', '7.3000', '9.1000', '55000',
      '23', '6.5000', '1.5000', '7.5000', '51000', '41000', '48000',
      '8.5000', '46000', '8.3000', '4.3000', '4.7000', '3.3000', '40000',
      '7.8000', '8.8000', '6.6000', '5.1000', '61000', '66000', '79',
      '8.4000', '118', '44000', '695', '1.6000', '6.2000', '18', '53000',
      '1.4000', '3.0000', '5.8000', '3.8000', '9.6000', '45000', '63000',
      '49000', '77000', '4.4000', '4.8000', '70000', '6.9000', '9.3000',
      '10.0000', '8.1000', '36000', '84000', '97000', '2.0000', '1.9000',
      '1.8000', '5.3000', '47000', '556', '526', '76000', '7.6000',
      '59000', '9.7000', '78000', '72000', '43000', '7.7000', '6.3000',
      '334', '34000', '93000', '65000', '79000', '100000', '58000',
      '50000', '68000', '64000', '67000', '60000', '94000', '232',
      '99000', '624', '95000', '8.5', '41', '292', '11', '80000',
      '1.7000', '74000', '62000', '69000', '75000', '98000', '85000',
      '82000', '96000', '87000', '71000', '86000', '91000', '81000',
      '92000', '83000', '88000', '704', '862', '899', '378', '266',
      '375', '1.3000', '975', '980', '4.1000', '89000', '696', '544',
      '525', '920', '779', '853', '720', '713', '772', '318', '58',
      '241', '196', '857', '51', '953', '865', '251', '930', '540',
      '313', '746', '203', '26', '314', '239', '371', '220', '730',
      '756', '91', '293', '17', '74', '14', '317', '78', '924', '902',
      '818', '81', '939', '169', '45', '475', '965', '90000', '545',
      '61', '283', '655', '714', '93', '872', '121', '322', '1.0000',
      '976', '172', '238', '549', '206', '954', '444', '717', '210',
      '609', '308', '705', '306', '904', '473', '175', '350', '383',
      '454', '421', '70', '812', '442', '842', '417', '412', '459',
      '478', '335', '782', '721', '430', '429', '192', '200', '460',
      '728', '496', '816', '414', '506', '887', '613', '243', '569',
      '778', '683', '592', '319', '186', '840', '647', '191', '373',
      '437', '598', '716', '585', '982', '222', '219', '55', '948',
      '323', '691', '511', '951', '963', '25', '554', '351', '27', '82',
      '208', '913', '514', '551', '29', '103', '898', '743', '116',
      '153', '209', '353', '499', '173', '597', '809', '122', '411',
      '400', '801', '787', '237', '50', '643', '986', '97', '516', '837',
      '780', '961', '269', '20', '498', '600', '749', '642', '881', '72',
      '656', '601', '221', '228', '108', '940', '176', '33', '663', '34',
      '942', '259', '164', '458', '245', '629', '28', '288', '775',
      '785', '636', '916', '994', '309', '485', '914', '903', '608',
      '500', '54', '562', '847', '957', '688', '811', '270', '48', '329',
      '523', '921', '874', '981', '784', '280', '24', '518', '754',
      '892', '154', '860', '364', '387', '626', '161', '879', '39',
      '970', '170', '141', '160', '144', '143', '190', '376', '193',
      '246', '73', '658', '992', '253', '420', '404', '470', '226',
      '240', '89', '234', '257', '861', '467', '157', '44', '676', '67',
      '552', '885', '1020', '582', '619'], dtype=object)

```

In [41]:

```
data_copy["Size"].dtype
```

Out[41]:

```
dtype('O')
```

In [42]:

```
data_copy["Size"] = data_copy["Size"].astype("float")
```

In [43]:

```
data_copy["Size"].dtype
```

Out[43]:

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

In [44]:

```
data_copy["Size"].head(10)
```

Out[44]:

```
0    19000.0
1    14000.0
2         8.7
3    25000.0
4         2.8
5         5.6
6    19000.0
7    29000.0
8    33000.0
9         3.1
Name: Size, dtype: float64
```

In [45]:

```
data["Size"][0]
```

Out[45]:

```
'19M'
```

In [46]:

```
19000
```

Out[46]:

```
19000
```

In [47]:

```
19*1000
```

Out[47]:

```
19000
```

In [48]:

```
data["Size"][2]
```

Out[48]:

```
'8.7M'
```

In [49]:

```
8.7*1000
```

Out[49]:

```
8700.0
```

In [50]:

```
data_copy["Size"][2]*1000
```

Out[50]:

```
8700.0
```

In [51]:

```
for i in data_copy["Size"]:
    if i < 10:
        data_copy["Size"]=data_copy["Size"].replace(i,i*1000)
```

In [52]:

```
data_copy["Size"].replace(8.7,8.7*1000)
```

Out[52]:

```
0      19000.0
1      14000.0
2       8700.0
3      25000.0
4       2800.0
...
10836   53000.0
10837    3600.0
10838    9500.0
10839      NaN
10840   19000.0
Name: Size, Length: 10840, dtype: float64
```

In [53]:

```
data_copy["Size"].head(50)
```

Out[53]:

```
0    19000.0
1    14000.0
2     8700.0
3    25000.0
4     2800.0
5     5600.0
6    19000.0
7    29000.0
8    33000.0
9     3100.0
10   28000.0
11   12000.0
12   20000.0
13   21000.0
14   37000.0
15    2700.0
16    5500.0
17   17000.0
18   39000.0
19   31000.0
20   14000.0
21   12000.0
22    4200.0
23    7000.0
24   23000.0
25    6000.0
26   25000.0
27    6100.0
28    4600.0
29    4200.0
30    9200.0
31    5200.0
32   11000.0
33   11000.0
34    4200.0
35    9200.0
36   24000.0
37         NaN
38   11000.0
39    9400.0
40   15000.0
41   10000.0
42         NaN
43    1200.0
44   12000.0
45   24000.0
46   26000.0
47    8000.0
48    7900.0
49   25000.0
```

Name: Size, dtype: float64

In [54]:

```
data_copy["Size"]/1000
```

Out[54]:

```
0      19.0
1      14.0
2       8.7
3      25.0
4       2.8
...
10836   53.0
10837    3.6
10838    9.5
10839   NaN
10840   19.0
Name: Size, Length: 10840, dtype: float64
```

In [55]:

```
data_copy["Size"].dtype
```

Out[55]:

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

In [56]:

```
data["Size"].unique()
```

Out[56]:

```

array(['19M', '14M', '8.7M', '25M', '2.8M', '5.6M', '29M', '33M', '3.1M',
      '28M', '12M', '20M', '21M', '37M', '2.7M', '5.5M', '17M', '39M',
      '31M', '4.2M', '7.0M', '23M', '6.0M', '6.1M', '4.6M', '9.2M',
      '5.2M', '11M', '24M', 'Varies with device', '9.4M', '15M', '10M',
      '1.2M', '26M', '8.0M', '7.9M', '56M', '57M', '35M', '54M', '201k',
      '3.6M', '5.7M', '8.6M', '2.4M', '27M', '2.5M', '16M', '3.4M',
      '8.9M', '3.9M', '2.9M', '38M', '32M', '5.4M', '18M', '1.1M',
      '2.2M', '4.5M', '9.8M', '52M', '9.0M', '6.7M', '30M', '2.6M',
      '7.1M', '3.7M', '22M', '7.4M', '6.4M', '3.2M', '8.2M', '9.9M',
      '4.9M', '9.5M', '5.0M', '5.9M', '13M', '73M', '6.8M', '3.5M',
      '4.0M', '2.3M', '7.2M', '2.1M', '42M', '7.3M', '9.1M', '55M',
      '23k', '6.5M', '1.5M', '7.5M', '51M', '41M', '48M', '8.5M', '46M',
      '8.3M', '4.3M', '4.7M', '3.3M', '40M', '7.8M', '8.8M', '6.6M',
      '5.1M', '61M', '66M', '79k', '8.4M', '118k', '44M', '695k', '1.6M',
      '6.2M', '18k', '53M', '1.4M', '3.0M', '5.8M', '3.8M', '9.6M',
      '45M', '63M', '49M', '77M', '4.4M', '4.8M', '70M', '6.9M', '9.3M',
      '10.0M', '8.1M', '36M', '84M', '97M', '2.0M', '1.9M', '1.8M',
      '5.3M', '47M', '556k', '526k', '76M', '7.6M', '59M', '9.7M', '78M',
      '72M', '43M', '7.7M', '6.3M', '334k', '34M', '93M', '65M', '79M',
      '100M', '58M', '50M', '68M', '64M', '67M', '60M', '94M', '232k',
      '99M', '624k', '95M', '8.5k', '41k', '292k', '11k', '80M', '1.7M',
      '74M', '62M', '69M', '75M', '98M', '85M', '82M', '96M', '87M',
      '71M', '86M', '91M', '81M', '92M', '83M', '88M', '704k', '862k',
      '899k', '378k', '266k', '375k', '1.3M', '975k', '980k', '4.1M',
      '89M', '696k', '544k', '525k', '920k', '779k', '853k', '720k',
      '713k', '772k', '318k', '58k', '241k', '196k', '857k', '51k',
      '953k', '865k', '251k', '930k', '540k', '313k', '746k', '203k',
      '26k', '314k', '239k', '371k', '220k', '730k', '756k', '91k',
      '293k', '17k', '74k', '14k', '317k', '78k', '924k', '902k', '818k',
      '81k', '939k', '169k', '45k', '475k', '965k', '90M', '545k', '61k',
      '283k', '655k', '714k', '93k', '872k', '121k', '322k', '1.0M',
      '976k', '172k', '238k', '549k', '206k', '954k', '444k', '717k',
      '210k', '609k', '308k', '705k', '306k', '904k', '473k', '175k',
      '350k', '383k', '454k', '421k', '70k', '812k', '442k', '842k',
      '417k', '412k', '459k', '478k', '335k', '782k', '721k', '430k',
      '429k', '192k', '200k', '460k', '728k', '496k', '816k', '414k',
      '506k', '887k', '613k', '243k', '569k', '778k', '683k', '592k',
      '319k', '186k', '840k', '647k', '191k', '373k', '437k', '598k',
      '716k', '585k', '982k', '222k', '219k', '55k', '948k', '323k',
      '691k', '511k', '951k', '963k', '25k', '554k', '351k', '27k',
      '82k', '208k', '913k', '514k', '551k', '29k', '103k', '898k',
      '743k', '116k', '153k', '209k', '353k', '499k', '173k', '597k',
      '809k', '122k', '411k', '400k', '801k', '787k', '237k', '50k',
      '643k', '986k', '97k', '516k', '837k', '780k', '961k', '269k',
      '20k', '498k', '600k', '749k', '642k', '881k', '72k', '656k',
      '601k', '221k', '228k', '108k', '940k', '176k', '33k', '663k',
      '34k', '942k', '259k', '164k', '458k', '245k', '629k', '28k',
      '288k', '775k', '785k', '636k', '916k', '994k', '309k', '485k',
      '914k', '903k', '608k', '500k', '54k', '562k', '847k', '957k',
      '688k', '811k', '270k', '48k', '329k', '523k', '921k', '874k',
      '981k', '784k', '280k', '24k', '518k', '754k', '892k', '154k',
      '860k', '364k', '387k', '626k', '161k', '879k', '39k', '970k',
      '170k', '141k', '160k', '144k', '143k', '190k', '376k', '193k',
      '246k', '73k', '658k', '992k', '253k', '420k', '404k', '1,000+',
      '470k', '226k', '240k', '89k', '234k', '257k', '861k', '467k',
      '157k', '44k', '676k', '67k', '552k', '885k', '1020k', '582k',
      '619k'], dtype=object)

```

In [57]:

```
data.head(2)
```

Out[57]:

	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

In [58]:

```
data.head(3)
```

Out[58]:

	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

In [60]:

```
import numpy as np
def Process_Size(item):
    if str(item)[-1]=='M':
        res=float(str(item).replace('M',''))
        res=res*1000
        return str(res)
    elif str(item)[-1]=='k':
        res=str(item).replace('k','')
        return res
    else:
        return str(np.NaN)
```

In [61]:

```
data_copy.columns
```

Out[61]:

```
Index(['App', 'Category', 'Rating', 'Reviews', 'Size', 'Installs', 'Type',
       'Price', 'Content Rating', 'Genres', 'Last Updated', 'Current Ver',
       'Android Ver'],
      dtype='object')
```

In [62]:

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

C:\Users\97474\AppData\Local\Temp\ipykernel_6764\1334073603.py:1: FutureWarning: The default value of regex will change from True to False in a future version. In addition, single character regular expressions will *not* be treated as literal strings when regex=True.

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

Out[62]:

```
0          10,000
1         500,000
2        5,000,000
3       50,000,000
4         100,000
...
10836         5,000
10837          100
10838         1,000
10839         1,000
10840       10,000,000
Name: Installs, Length: 10840, dtype: object
```

In [63]:

```
data_copy["Installs"].str
```

Out[63]:

```
<pandas.core.strings.accessor.StringMethods at 0x1bcefd3b3d0>
```

In [64]:

```
data_copy["Installs"].dtype
```

Out[64]:

```
dtype('O')
```

In [65]:

```
data_copy["Installs"][0]
```

Out[65]:

```
'10,000+'
```

In [66]:

```
data_copy["Price"].unique()
```

Out[66]:

```
array(['0', '$4.99', '$3.99', '$6.99', '$1.49', '$2.99', '$7.99', '$5.99',  
      '$3.49', '$1.99', '$9.99', '$7.49', '$0.99', '$9.00', '$5.49',  
      '$10.00', '$24.99', '$11.99', '$79.99', '$16.99', '$14.99',  
      '$1.00', '$29.99', '$12.99', '$2.49', '$10.99', '$1.50', '$19.99',  
      '$15.99', '$33.99', '$74.99', '$39.99', '$3.95', '$4.49', '$1.70',  
      '$8.99', '$2.00', '$3.88', '$25.99', '$399.99', '$17.99',  
      '$400.00', '$3.02', '$1.76', '$4.84', '$4.77', '$1.61', '$2.50',  
      '$1.59', '$6.49', '$1.29', '$5.00', '$13.99', '$299.99', '$379.99',  
      '$37.99', '$18.99', '$389.99', '$19.90', '$8.49', '$1.75',  
      '$14.00', '$4.85', '$46.99', '$109.99', '$154.99', '$3.08',  
      '$2.59', '$4.80', '$1.96', '$19.40', '$3.90', '$4.59', '$15.46',  
      '$3.04', '$4.29', '$2.60', '$3.28', '$4.60', '$28.99', '$2.95',  
      '$2.90', '$1.97', '$200.00', '$89.99', '$2.56', '$30.99', '$3.61',  
      '$394.99', '$1.26', '$1.20', '$1.04'], dtype=object)
```

In [67]:

```
char_to_remove=["+", ",", "$"]  
cols_to_clean=["Installs", "Price"]  
for item in char_to_remove:  
    for col in cols_to_clean:  
        data_copy[col]=data_copy[col].str.replace(item, "")
```

C:\Users\97474\AppData\Local\Temp\ipykernel_6764\577029992.py:5: FutureWarning: The default value of regex will change from True to False in a future version. In addition, single character regular expressions will *not* be treated as literal strings when regex=True.

```
data_copy[col]=data_copy[col].str.replace(item, "")
```

In [68]:

```
data_copy.head(2)
```

Out[68]:

	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	10000	Free	0	Everyone
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14000.0	500000	Free	0	Everyone

In [69]:

```
data_copy.head(3)
```

Out[69]:

	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	10000	Free	0	Everyone
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14000.0	500000	Free	0	Everyone
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8700.0	5000000	Free	0	Everyone

In [70]:

```
data_copy.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 10840 entries, 0 to 10840
Data columns (total 13 columns):
#   Column                Non-Null Count  Dtype  
---  -
0   App                    10840 non-null  object 
1   Category               10840 non-null  object 
2   Rating                 9366 non-null   float64
3   Reviews                10840 non-null  int32  
4   Size                   9145 non-null   float64
5   Installs               10840 non-null  object 
6   Type                   10839 non-null  object 
7   Price                  10840 non-null  object 
8   Content Rating         10840 non-null  object 
9   Genres                 10840 non-null  object 
10  Last Updated           10840 non-null  object 
11  Current Ver            10832 non-null  object 
12  Android Ver            10838 non-null  object 
dtypes: float64(2), int32(1), object(10)
memory usage: 1.4+ MB
```

In [71]:

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

In [72]:

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


In [73]:

```
data_copy.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 10840 entries, 0 to 10840
Data columns (total 13 columns):
#   Column                Non-Null Count  Dtype
---  -
0   App                    10840 non-null  object
1   Category               10840 non-null  object
2   Rating                 9366 non-null   float64
3   Reviews                10840 non-null  int32
4   Size                   9145 non-null   float64
5   Installs               10840 non-null  int32
6   Type                   10839 non-null  object
7   Price                  10840 non-null  float64
8   Content Rating         10840 non-null  object
9   Genres                 10840 non-null  object
10  Last Updated           10840 non-null  object
11  Current Ver            10832 non-null  object
12  Android Ver            10838 non-null  object
dtypes: float64(3), int32(2), object(8)
memory usage: 1.3+ MB
```

In [74]:

```
data_copy["Last Updated"]
```

Out[74]:

```
0      January 7, 2018
1      January 15, 2018
2      August 1, 2018
3      June 8, 2018
4      June 20, 2018
...
10836   July 25, 2017
10837   July 6, 2018
10838   January 20, 2017
10839   January 19, 2015
10840   July 25, 2018
Name: Last Updated, Length: 10840, dtype: object
```

In [75]:

```
data_copy["Last Updated"].dtype
```

Out[75]:

```
dtype('O')
```

In [77]:

```
data_copy["Last Updated"]=pd.to_datetime(data_copy["Last Updated"])
```

In [78]:

```
data_copy.head(2)
```

Out[78]:

	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	10000	Free	0.0	Everyone
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14000.0	500000	Free	0.0	Everyone

In [79]:

```
data.head(2)
```

Out[79]:

	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

In [88]:

```
data_copy["Last Updated"].dt.day
```

Out[88]:

```
0         7
1        15
2         1
3         8
4        20
..
10836     25
10837      6
10838     20
10839     19
10840     25
Name: Last Updated, Length: 10840, dtype: int64
```

In [87]:

```
data_copy["day"]=data_copy["Last Updated"].dt.day
```

In [89]:

```
data_copy["Last Updated"].dt.month
```

Out[89]:

```
0         1
1         1
2         8
3         6
4         6
..
10836     7
10837     7
10838     1
10839     1
10840     7
Name: Last Updated, Length: 10840, dtype: int64
```

In [81]:

```
data_copy["day"]=data_copy["Last Updated"].dt.month
```

In [90]:

```
data_copy["Last Updated"].dt.year
```

Out[90]:

```
0      2018
1      2018
2      2018
3      2018
4      2018
...
10836   2017
10837   2018
10838   2017
10839   2015
10840   2018
Name: Last Updated, Length: 10840, dtype: int64
```

In [82]:

```
data_copy["day"]=data_copy["Last Updated"].dt.year
```

In [83]:

```
data_copy.head()
```

Out[83]:

	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	10000	Free	0.0	Everyone
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14000.0	500000	Free	0.0	Everyone
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8700.0	5000000	Free	0.0	Everyone
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25000.0	50000000	Free	0.0	Teen
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2800.0	100000	Free	0.0	Everyone



In [92]:

```
data_copy["Android Ver"]
```

Out[92]:

```
0          4.0.3
1          4.0.3
2          4.0.3
3          4.2
4          4.4
...
10836       4.1
10837       4.1
10838       2.2
10839    Varies with device
10840    Varies with device
Name: Android Ver, Length: 10840, dtype: object
```

In [94]:

```
data_copy["Android Ver"].unique()
```

Out[94]:

```
array(['4.0.3 ', '4.2 ', '4.4 ', '2.3 ', '3.0 ', '4.1 ', '4.0 ', '2.3.3 ',
      'Varies with device', '2.2 ', '5.0 ', '6.0 ', '1.6 ', '1.5 ',
      '2.1 ', '7.0 ', '5.1 ', '4.3 ', '4.0.3 - 7.1.1', '2.0 ', '3.2 ',
      '4.4W ', '7.1 ', '7.0 - 7.1.1', '8.0 ', '5.0 - 8.0', '3.1 ',
      '2.0.1 ', '4.1 - 7.1.1', nan, '5.0 - 6.0', '1.0 ', '2.2 - 7.1.1',
      '5.0 - 7.1.1'], dtype=object)
```

In [84]:

```
data_copy["Android Ver"]=data_copy["Android Ver"].str.replace("and up", "")
```

In [85]:

```
data_copy["Android Ver"]=data_copy["Android Ver"].str.replace("varies with device", "Nan")
```

In [86]:

```
data_copy.head(2)
```

Out[86]:

	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	10000	Free	0.0	Everyone
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14000.0	500000	Free	0.0	Everyone

In [91]:

```
data_copy
```

Out[91]:

	App	Category	Rating	Reviews	Size	Installs	Type	Pri
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19000.0	10000	Free	(
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14000.0	500000	Free	(
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8700.0	5000000	Free	(
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25000.0	50000000	Free	(
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2800.0	100000	Free	(
...
10836	Sya9a Maroc - FR	FAMILY	4.5	38	53000.0	5000	Free	(
10837	Fr. Mike Schmitz Audio Teachings	FAMILY	5.0	4	3600.0	100	Free	(
10838	Parkinson Exercices FR	MEDICAL	NaN	3	9500.0	1000	Free	(
10839	The SCP Foundation DB fr nn5n	BOOKS_AND_REFERENCE	4.5	114	NaN	1000	Free	(
10840	iHoroscope - 2018 Daily Horoscope & Astrology	LIFESTYLE	4.5	398307	19000.0	10000000	Free	(

10840 rows × 14 columns



In [95]:

```
data_copy.to_csv("cleaned _google_playstore_data",index= False)
```

In []: