

In [2]:

```
import pandas as pd
data=pd.read_csv("https://raw.githubusercontent.com/sunnysavital0/Statistics-With-Python-TheCompleteGuide/main/googleplaystore.csv")
```

## Dipak Mani

In [3]:

```
data.head()
```

Out[3]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Cu
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0	Everyone	Art & Design	January 7, 2018	
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone	Art & Design;Pretend Play	January 15, 2018	
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8.7M	5,000,000+	Free	0	Everyone	Art & Design	August 1, 2018	
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50,000,000+	Free	0	Teen	Art & Design	June 8, 2018	Vi
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100,000+	Free	0	Everyone	Art & Design;Creativity	June 20, 2018	de

In [4]:

```
data.shape
```

Out[4]:

(10841, 13)

In [5]:

```
#last 5 rows
data.tail()
```

Out[5]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	I
10836	Sya9a Maroc - FR	FAMILY	4.5	38	53M	5,000+	Free	0	Everyone	Education	July 2	
10837	Fr. Mike Schmitz Audio	FAMILY	5.0	4	3.6M	100+	Free	0	Everyone	Education	Ju	

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Updated
10838	Parkinson Exercises FR	MEDICAL	NaN	3	9.5M	1,000+	Free	0	Everyone	Medical	Jan 20, 2
10839	The SCP Foundation DB fr nn5n	BOOKS_AND_REFERENCE	4.5	114	Varies with device	1,000+	Free	0	Mature 17+	Books & Reference	Jan 19, 2
10840	iHoroscope - 2018 Daily Horoscope & Astrology	LIFESTYLE	4.5	398307	19M	10,000,000+	Free	0	Everyone	Lifestyle	July 2

--	--	--	--	--	--	--	--	--	--	--	--

In [6]:

```
data.sample()
```

Out[6]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver
3135	VZ Navigator for Tablets	TRAVEL_AND_LOCAL	3.6	285	39M	500,000+	Free	0	Everyone	Travel & Local	January 22, 2018	10.7.8.111

--	--	--	--	--	--	--	--	--	--	--	--	--

In [7]:

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

Out[7]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Updated
4452	P-Home for KLWP	PERSONALIZATION	5.0	4	12M	100+	Paid	\$0.99	Everyone	Personalization	M
5273	AK-47: Simulator and Shooting	FAMILY	4.0	345	14M	100,000+	Free	0	Everyone 10+	Simulation	Nov 11
8255	Marvel and DC Wallpapers	PERSONALIZATION	4.2	249	91M	10,000+	Free	0	Everyone	Personalization	
3936	Alto's Adventure	GAME	4.6	515240	63M	10,000,000+	Free	0	Everyone	Action	Ji
7041	BZ-Smart	NEWS_AND_MAGAZINES	4.3	828	8.2M	10,000+	Free	0	Everyone	News & Magazines	Oct 10
3175	Expedia Hotels, Flights & Car Rental Travel Deals	TRAVEL_AND_LOCAL	4.1	136626	14M	10,000,000+	Free	0	Everyone	Travel & Local	Aug
7517	CL Notifier	TOOLS	5.0	36	3.2M	100+	Free	0	Teen	Tools	Aug
6294	TED	FAMILY	4.6	181961	18M	10,000,000+	Free	0	Everyone 10+	Education	Ji
9948	HondaLink EV	TOOLS	2.8	44	41M	10,000+	Free	0	Everyone	Tools	Ju
7456	Company Kitchen	LIFESTYLE	2.8	81	7.7M	10,000+	Free	0	Everyone	Lifestyle	Ma

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 [13]:

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

Out[13]:

	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 [15]:

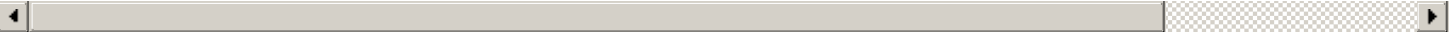
```
data[data.duplicated()]
```

Out[15]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated
229	Quick PDF Scanner + OCR FREE	BUSINESS	4.2	80805	Varies with device	5,000,000+	Free	0	Everyone	Business	February 26, 2018
236	Box	BUSINESS	4.2	159872	Varies with device	10,000,000+	Free	0	Everyone	Business	July 31, 2018

239	Google My Business	BUSINESS	4.4	70901	Varies with device	5,000,000+	Free	0	Everyone	Business	July 24, 2018
256	ZOOM Cloud Meetings	BUSINESS	4.4	31614	37M	10,000,000+	Free	0	Everyone	Business	July 20, 2018
261	join.me - Simple Meetings	BUSINESS	4.0	6989	Varies with device	1,000,000+	Free	0	Everyone	Business	July 16, 2018
...	...	...	...	...	...	...	...	...	...	...	...
8643	Wunderlist: To-Do List & Tasks	PRODUCTIVITY	4.6	404610	Varies with device	10,000,000+	Free	0	Everyone	Productivity	April 6, 2018
8654	TickTick: To Do List with Reminder, Day Planner	PRODUCTIVITY	4.6	25370	Varies with device	1,000,000+	Free	0	Everyone	Productivity	August 6, 2018
8658	ColorNote Notepad Notes	PRODUCTIVITY	4.6	2401017	Varies with device	100,000,000+	Free	0	Everyone	Productivity	June 27, 2018
10049	Airway Ex - Intubate. Anesthetize. Train.	MEDICAL	4.3	123	86M	10,000+	Free	0	Everyone	Medical	June 1, 2018
10768	AAFP	MEDICAL	3.8	63	24M	10,000+	Free	0	Everyone	Medical	June 22, 2018

483 rows x 13 columns



In [19]:

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

Out[19]:

```
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 [23]:

```
data['Reviews'].shape
```

Out[23]:

```
(10841,)
```

In [21]:

```
data['Reviews'].dtype
```

Out[21]:

dtype('O')

In [27]:

```
data.Reviews.str
```

Out[27]:

<pandas.core.strings.accessor.StringMethods at 0x7fa52c63b910>

In [25]:

```
data.Reviews.str.isnumeric()
```

Out[25]:

```
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 [31]:

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

Out[31]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Cu
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0	Everyone	Art & Design	January 7, 2018	
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone	Art & Design;Pretend Play	January 15, 2018	
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8.7M	5,000,000+	Free	0	Everyone	Art & Design	August 1, 2018	
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50,000,000+	Free	0	Teen	Art & Design	June 8, 2018	V
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100,000+	Free	0	Everyone	Art & Design;Creativity	June 20, 2018	de

In [29]:

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

Out[29]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Version	
10472	Life Made WI-Fi Touchscreen Photo Frame		1.9	19.0	3.0M	1,000+	Free	0	Everyone	NaN	February 11, 2018	1.0.19	4.0 and up	None

```
data.head()
```

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Cu
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0	Everyone	Art & Design	January 7, 2018	
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone	Art & Design;Pretend Play	January 15, 2018	
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8.7M	5,000,000+	Free	0	Everyone	Art & Design	August 1, 2018	
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50,000,000+	Free	0	Teen	Art & Design	June 8, 2018	Videodownload
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100,000+	Free	0	Everyone	Art & Design;Creativity	June 20, 2018	

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

```
data copy.head()
```

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Cumulative Rating
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0	Everyone	Art & Design	January 7, 2018	4.1
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone	Art & Design;Pretend Play	January 15, 2018	3.9
	U Launcher	ART_AND_DESIGN	3.9	10	14M	500,000+	Free	0	Everyone	Art & Design;Pretend Play	January 15, 2018	3.9

	Launcher	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Cu
2	FREE Live	ART_AND_DESIGN	4.7	87510	8.7M	5,000,000+	Free	0	Everyone	Art & Design	August 1, 2018		
	Cool Themes, Hide ...												
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50,000,000+	Free	0	Teen	Art & Design	June 8, 2018		Vi
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100,000+	Free	0	Everyone	Art & Design;Creativity	June 20, 2018		de

In [38]:

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

In [39]:

```
data_copy.shape
```

Out[39]:

```
(10840, 13)
```

In [40]:

```
data.shape
```

Out[40]:

```
(10841, 13)
```

In [42]:

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

Out[42]:

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

In [43]:

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

In [44]:

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

Out[44]:

```
dtype('int64')
```

In [45]:

```
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  int64
4   Size            10840 non-null  object
5   Installs        10840 non-null  object
6   Type            10839 non-null  object
7   Price           10840 non-null  object
```

```
7  Title 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), int64(1), object(11)
memory usage: 1.2+ MB
```

In [46]:

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

Out[46]:

```
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 [47]:

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

Out[47]:

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

In [49]:

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

Out[49]:

```
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 [51]:

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

In [52]:

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

In [54]:

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

Out[54]:

```

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', '238',

```

```

291', '350', '340', '313', '740', '203', '20', '314', '299',
'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 [56]:

```

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

```

In [57]:

```

data_copy["Size"].unique()

```

Out[57]:

```

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 [61]:

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

Out[61]:

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

In [62]:

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

In [63]:

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

Out[63]:

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

In [64]:

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

Out[64]:

```
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 [68]:

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

Out[68]:

```
'19M'
```

In [ ]:

```
19000
```

In [69]:

```
19*1000
```

Out[69]:

```
19000
```

In [66]:

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

Out[66]:

```
'8.7M'
```

In [ ]:

```
8.7000
```

In [67]:

```
8.7*1000
```

Out[67]:

```
8700.0
```

In [72]:

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

Out[72]:

```
8700.0
```

In [ ]:

```
17.7M
```

In [74]:

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

In [73]:

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

Out[73]:

```
0      19000.0
1      14000.0
2       8700.0
3      25000.0
4         2.8
...
10836   53000.0
10837     3.6
10838     9.5
10839     NaN
10840   19000.0
Name: Size, Length: 10840, dtype: float64
```

In [75]:

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

Out[75]:

```
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 [76]:

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

Out[76]:

```
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 [78]:

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

```
Out[78]:
```

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

```
In [79]:
```

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

```
Out[79]:
```

```
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 [81]:
```

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

Out [81]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0	Everyone	Art & Design	January 7, 2018	1.0.0
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone	Art & Design;Pretend Play	January 15, 2018	2.0.0

In [82]:

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

Out [82]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Cu
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19000.0	10,000+	Free	0	Everyone	Art & Design	January 7, 2018	
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14000.0	500,000+	Free	0	Everyone	Art & Design;Pretend Play	January 15, 2018	
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8700.0	5,000,000+	Free	0	Everyone	Art & Design	August 1, 2018	

In [ ]:

```
#written by soynddeep
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 [83]:

```
data_copy.columns
```

Out [83]:

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

In [87]:

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

```
/tmp/ipykernel_91/1518358678.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[87]:
```

```
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 [88]:
```

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

```
Out[88]:
```

```
<pandas.core.strings.accessor.StringMethods at 0x7fa52a0e25c0>
```

```
In [89]:
```

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

```
Out[89]:
```

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

```
In [90]:
```

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

```
Out[90]:
```

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

```
In [92]:
```

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

```
Out[92]:
```

```
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 [93]:
```

```
char_to_removes=["+", ",", "$"]
cols_to_clean=["Installs", "Price"]
for item in char_to_removes:
    for col in cols_to_clean:
```



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

```
/tmp/ipykernel_91/577029992.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 [95]:

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

Out[95]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Version
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19000.0	10000	Free	0	Everyone	Art & Design	January 7, 2018	1.0.0
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14000.0	500000	Free	0	Everyone	Art & Design;Pretend Play	January 15, 2018	2.0.0

In [97]:

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

Out[97]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Version
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0	Everyone	Art & Design	January 7, 2018	1.0.0
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone	Art & Design;Pretend Play	January 15, 2018	2.0.0
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8.7M	5,000,000+	Free	0	Everyone	Art & Design	August 1, 2018	1.2.0

In [98]:

```
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	int64
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), int64(1), object(10)
memory usage: 1.4+ MB
```

In [99]:

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

In [100]:

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

In [101]:

```
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  int64
4   Size             9145 non-null   float64
5   Installs         10840 non-null  int64
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), int64(2), object(8)
memory usage: 1.4+ MB
```

In [102]:

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

Out[102]:

```
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 [103]:

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

Out[103]:

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

In [107]:

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

In [109]:

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

Out[109]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Version
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19000.0	10000	Free	0.0	Everyone	Art & Design	2018-01-07	1.0.0
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14000.0	500000	Free	0.0	Everyone	Art & Design;Pretend Play	2018-01-15	2.0.0

In [110]:

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

Out[110]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Version
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0	Everyone	Art & Design	January 7, 2018	1.0.0
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone	Art & Design;Pretend Play	January 15, 2018	2.0.0

In [115]:

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

In [116]:

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

In [117]:

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

In [118]:

```
data_copy.head()
```

Out[118]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Version
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19000.0	10000	Free	0.0	Everyone	Art & Design	2018-01-07	1.0.0
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14000.0	500000	Free	0.0	Everyone	Art & Design;Pretend Play	2018-01-15	2.0.0

	App Launcher	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Cu
2	FREE Live Lite – Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8700.0	5000000	Free	0.0	Everyone	Art & Design	2018-08-01	
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25000.0	50000000	Free	0.0	Teen	Art & Design	2018-06-08	V d
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2800.0	100000	Free	0.0	Everyone	Art & Design;Creativity	2018-06-20	

In [119]:

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

Out[119]:

```
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
...
10836       4.1 and up
10837       4.1 and up
10838       2.2 and up
10839  Varies with device
10840  Varies with device
Name: Android Ver, Length: 10840, dtype: object
```

In [122]:

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

In [123]:

```
data_copy["Android Ver"]=data_copy['Android Ver'].str.replace("Varies with device", "NaN")
```

In [124]:

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

Out[124]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Curr Ver
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19000.0	10000	Free	0.0	Everyone	Art & Design	2018-01-07	1.0.
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14000.0	500000	Free	0.0	Everyone	Art & Design;Pretend Play	2018-01-15	2.0.

In [125]:

```
data_copy.to_csv("clead_google_playstore_data.csv", index=False)
```

In [129]:



4	- Number App Coloring Book	ART_AND_DESIGN	Category	Rating	Reviews	Size	Installs	Type	Price	Content Everyone Rating	Art & Genres Design;Creativity
...	...		...	...	...	...	...	...	...	...	...
10835	Sya9a Maroc - FR		FAMILY	4.5	38	53000.0	5000	Free	0.0	Everyone	Education
10836	Fr. Mike Schmitz Audio Teachings		FAMILY	5.0	4	3600.0	100	Free	0.0	Everyone	Education
10837	Parkinson Exercices FR		MEDICAL	NaN	3	9500.0	1000	Free	0.0	Everyone	Medical
10838	The SCP Foundation DB fr nn5n	BOOKS_AND_REFERENCE		4.5	114	NaN	1000	Free	0.0	Mature 17+	Books & Reference
10839	iHoroscope - 2018 Daily Horoscope & Astrology		LIFESTYLE	4.5	398307	19000.0	10000000	Free	0.0	Everyone	Lifestyle

10840 rows × 16 columns



In [6]:

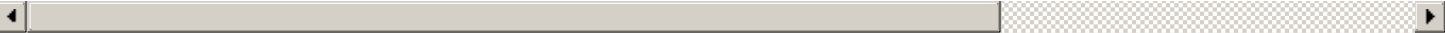
--

Out[6]:

	App		Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres
0	Photo Editor & Candy Camera & Grid & ScrapBook		ART_AND_DESIGN	4.1	159	19000.0	10000	Free	0.0	Everyone	Art & Design
1	Coloring book moana		ART_AND_DESIGN	3.9	967	14000.0	500000	Free	0.0	Everyone	Art & Design;Pretend Play
2	U Launcher Lite – FREE Live Cool Themes, Hide ...		ART_AND_DESIGN	4.7	87510	8700.0	5000000	Free	0.0	Everyone	Art & Design
3	Sketch - Draw & Paint		ART_AND_DESIGN	4.5	215644	25000.0	50000000	Free	0.0	Teen	Art & Design
4	Pixel Draw - Number Art Coloring Book		ART_AND_DESIGN	4.3	967	2800.0	100000	Free	0.0	Everyone	Art & Design;Creativity
...	...		...	...	...	...	...	...	...	...	...
10835	Sya9a Maroc - FR		FAMILY	4.5	38	53000.0	5000	Free	0.0	Everyone	Education
10836	Fr. Mike Schmitz Audio Teachings		FAMILY	5.0	4	3600.0	100	Free	0.0	Everyone	Education
10837	Parkinson Exercices FR		MEDICAL	NaN	3	9500.0	1000	Free	0.0	Everyone	Medical

App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres
<del>10838</del> The Foundation DB fr nn5n	BOOKS_AND_REFERENCE	4.5	114	NaN	1000	Free	0.0	17+	Books & Reference
10839 iHoroscope - 2018 Daily Horoscope & Astrology	LIFESTYLE	4.5	398307	19000.0	10000000	Free	0.0	Everyone	Lifestyle

10840 rows x 16 columns



In [1]:

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

warnings.filterwarnings("ignore")

%matplotlib inline
```

In [2]:

```
df=pd.read_csv("https://raw.githubusercontent.com/sunnysavital0/Statistics-With-Python-TheCompleteGuide/main/clead_google_playstore_data")
```

In [3]:

```
df.head()
```

Out[3]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Cu
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19000.0	10000	Free	0.0	Everyone	Art & Design	2018-01-07	
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14000.0	500000	Free	0.0	Everyone	Art & Design;Pretend Play	2018-01-15	
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8700.0	5000000	Free	0.0	Everyone	Art & Design	2018-08-01	
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25000.0	50000000	Free	0.0	Teen	Art & Design	2018-06-08	V
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2800.0	100000	Free	0.0	Everyone	Art & Design;Creativity	2018-06-20	d



In [4]:

```
df.tail()
```

Out[4]:

Out[4]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated
10835	Sya9a Maroc - FR	FAMILY	4.5	38	53000.0	5000	Free	0.0	Everyone	Education	2017-07-07
10836	Fr. Mike Schmitz Audio Teachings	FAMILY	5.0	4	3600.0	100	Free	0.0	Everyone	Education	2017-07-07
10837	Parkinson Exercices FR	MEDICAL	NaN	3	9500.0	1000	Free	0.0	Everyone	Medical	2017-01-01
10838	The SCP Foundation DB fr nn5n	BOOKS_AND_REFERENCE	4.5	114	NaN	1000	Free	0.0	Mature 17+	Books & Reference	2017-01-01
10839	iHoroscope - 2018 Daily Horoscope & Astrology	LIFESTYLE	4.5	398307	19000.0	10000000	Free	0.0	Everyone	Lifestyle	2017-07-07

In [5]:

```
df.sample(10)
```

Out[5]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated
6913	Bw Events	EVENTS	4.4	16	1700.0	100	Free	0.00	Everyone	Events	2016-08-09
3023	NBC Sports Gold	SPORTS	2.9	3017	24000.0	100000	Free	0.00	Everyone	Sports	2018-07-02
4419	Oddworld: New 'n' Tasty	GAME	4.6	1475	34000.0	10000	Paid	4.99	Teen	Action	2018-01-02
1431	Insight Timer - Free Meditation App	HEALTH_AND_FITNESS	4.6	20161	NaN	1000000	Free	0.00	Everyone	Health & Fitness	2018-08-01
2285	Anti Mosquito simulation	MEDICAL	4.2	74	2600.0	1000	Paid	1.50	Everyone	Medical	2014-09-06
2605	TwitCasting Live	SOCIAL	3.6	14835	NaN	1000000	Free	0.00	Teen	Social	2018-08-02
7212	MCQ CE IT	FAMILY	5.0	22	3600.0	1000	Free	0.00	Everyone	Education	2017-11-07
4626	Mũj T-Mobile Business	PRODUCTIVITY	3.7	10490	5800.0	1000000	Free	0.00	Everyone	Productivity	2018-07-24
7039	Evasion.bz	SOCIAL	4.1	8	4400.0	100	Free	0.00	Teen	Social	2017-07-14
10070	S.O.L : Stone of Life EX	FAMILY	4.3	171220	42000.0	5000000	Free	0.00	Everyone 10+	Role Playing	2015-07-28

In [6]:

```
df.shape
```



Out[6]:

(10840, 16)

In [7]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10840 entries, 0 to 10839
Data columns (total 16 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  int64
4   Size                   9145 non-null   float64
5   Installs               10840 non-null  int64
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            9476 non-null   object
13  day                    10840 non-null  int64
14  month                  10840 non-null  int64
15  year                   10840 non-null  int64
dtypes: float64(3), int64(5), object(8)
memory usage: 1.3+ MB
```

In [8]:

```
df.describe().T
```

Out[8]:

	count	mean	std	min	25%	50%	75%	max
Rating	9366.0	4.191757e+00	5.152189e-01	1.0	4.0	4.3	4.5	5.000000e+00
Reviews	10840.0	4.441529e+05	2.927761e+06	0.0	38.0	2094.0	54775.5	7.815831e+07
Size	9145.0	2.150653e+04	2.259602e+04	10.0	4900.0	13000.0	30000.0	1.000000e+05
Installs	10840.0	1.546434e+07	8.502936e+07	0.0	1000.0	100000.0	5000000.0	1.000000e+09
Price	10840.0	1.027368e+00	1.594970e+01	0.0	0.0	0.0	0.0	4.000000e+02
day	10840.0	1.560904e+01	9.561621e+00	1.0	6.0	16.0	24.0	3.100000e+01
month	10840.0	6.422325e+00	2.578388e+00	1.0	5.0	7.0	8.0	1.200000e+01
year	10840.0	2.017400e+03	1.100914e+00	2010.0	2017.0	2018.0	2018.0	2.018000e+03

In [9]:

```
df.head()
```

Out[9]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Cu
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19000.0	10000	Free	0.0	Everyone	Art & Design	2018-01-07	
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14000.0	500000	Free	0.0	Everyone	Art & Design;Pretend Play	2018-01-15	

	App Launcher	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver
2	FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8700.0	5000000	Free	0.0	Everyone	Art & Design	2018-08-01	
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25000.0	50000000	Free	0.0	Teen	Art & Design	2018-06-08	V
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2800.0	100000	Free	0.0	Everyone	Art & Design;Creativity	2018-06-20	d

In [10]:

```
df.columns
```

Out[10]:

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

In [11]:

```
df[df.duplicated()]
```

Out[11]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver
229	Quick PDF Scanner + OCR FREE	BUSINESS	4.2	80805	NaN	5000000	Free	0.0	Everyone	Business	2018-02-26	
236	Box	BUSINESS	4.2	159872	NaN	10000000	Free	0.0	Everyone	Business	2018-07-31	
239	Google My Business	BUSINESS	4.4	70991	NaN	5000000	Free	0.0	Everyone	Business	2018-07-24	2.19
256	ZOOM Cloud Meetings	BUSINESS	4.4	31614	37000.0	10000000	Free	0.0	Everyone	Business	2018-07-20	4
261	join.me - Simple Meetings	BUSINESS	4.0	6989	NaN	1000000	Free	0.0	Everyone	Business	2018-07-16	
...	...	...	...	...	...	...	...	...	...	...	...	
8643	Wunderlist: To-Do List & Tasks	PRODUCTIVITY	4.6	404610	NaN	10000000	Free	0.0	Everyone	Productivity	2018-04-06	
8654	TickTick: To Do List with Reminder, Day Planner	PRODUCTIVITY	4.6	25370	NaN	1000000	Free	0.0	Everyone	Productivity	2018-08-06	
8658	ColorNote Notepad Notes	PRODUCTIVITY	4.6	2401017	NaN	100000000	Free	0.0	Everyone	Productivity	2018-06-27	
10049	Airway Ex - Intubate. Anesthetize. Train.	MEDICAL	4.3	123	86000.0	10000	Free	0.0	Everyone	Medical	2018-06-01	
10767	AAFP	MEDICAL	3.8	63	24000.0	10000	Free	0.0	Everyone	Medical	2018-	

App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	06-22 Last Updated
-----	----------	--------	---------	------	----------	------	-------	----------------	--------	--------------------

483 rows x 16 columns

```
In [12]:
df[df.duplicated('Reviews')]

Out[12]:
```

App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres
Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2800.0	100000	Free	0.0	Everyone	Art & Design;Creativity
Extreme Rally Championship	AUTO_AND_VEHICLES	4.2	129	54000.0	100000	Free	0.0	Everyone	Auto & Vehicles
Skin Care and Natural Beauty	BEAUTY	NaN	654	7400.0	100000	Free	0.0	Teen	Beauty
Lady adviser (beauty, health)	BEAUTY	NaN	30	9900.0	10000	Free	0.0	Mature 17+	Beauty
Offline: English to Tagalog Dictionary	BOOKS_AND_REFERENCE	4.7	967	6700.0	500000	Free	0.0	Everyone	Books & Reference
...	...	...	...	...	...	...	...	...	...
FR Forms	BUSINESS	NaN	0	9600.0	10	Free	0.0	Everyone	Business
Sya9a Maroc - FR	FAMILY	4.5	38	53000.0	5000	Free	0.0	Everyone	Education
Fr. Mike Schmitz Audio Teachings	FAMILY	5.0	4	3600.0	100	Free	0.0	Everyone	Education
Parkinson Exercices FR	MEDICAL	NaN	3	9500.0	1000	Free	0.0	Everyone	Medical
The SCP Foundation DB fr nn5n	BOOKS_AND_REFERENCE	4.5	114	NaN	1000	Free	0.0	Mature 17+	Books & Reference

4839 rows x 16 columns

```
In [13]:
df=df.drop_duplicates(subset=['App'],keep='first')

In [14]:
df.shape

Out[14]:
(9659, 16)

In [15]:
df.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 9659 entries, 0 to 10839
Data columns (total 16 columns):
#    Column          Non-Null Count  Dtype
---
```

```
0 App 9659 non-null object
1 Category 9659 non-null object
2 Rating 8196 non-null float64
3 Reviews 9659 non-null int64
4 Size 8432 non-null float64
5 Installs 9659 non-null int64
6 Type 9658 non-null object
7 Price 9659 non-null float64
8 Content Rating 9659 non-null object
9 Genres 9659 non-null object
10 Last Updated 9659 non-null object
11 Current Ver 9651 non-null object
12 Android Ver 8667 non-null object
13 day 9659 non-null int64
14 month 9659 non-null int64
15 year 9659 non-null int64
dtypes: float64(3), int64(5), object(8)
memory usage: 1.3+ MB
```

In [16]:

```
numeric_features = [feature for feature in df.columns if df[feature].dtype != 'O']
```

In [17]:

```
numeric_features
```

Out[17]:

```
['Rating', 'Reviews', 'Size', 'Installs', 'Price', 'day', 'month', 'year']
```

In [18]:

```
df[numeric_features]
```

Out[18]:

	Rating	Reviews	Size	Installs	Price	day	month	year
0	4.1	159	19000.0	10000	0.0	7	1	2018
1	3.9	967	14000.0	500000	0.0	15	1	2018
2	4.7	87510	8700.0	5000000	0.0	1	8	2018
3	4.5	215644	25000.0	50000000	0.0	8	6	2018
4	4.3	967	2800.0	100000	0.0	20	6	2018
...	...	...	...	...	...	...	...	...
10835	4.5	38	53000.0	5000	0.0	25	7	2017
10836	5.0	4	3600.0	100	0.0	6	7	2018
10837	NaN	3	9500.0	1000	0.0	20	1	2017
10838	4.5	114	NaN	1000	0.0	19	1	2015
10839	4.5	398307	19000.0	10000000	0.0	25	7	2018

9659 rows x 8 columns

In [19]:

```
categorical_features = [feature for feature in df.columns if df[feature].dtype == 'O']
```

In [20]:

```
categorical_features
```

Out[20]:

```
['App',
 'Category',
 'Content Rating',
 'Genres',
 'Last Updated',
 'Type',
 'Current Ver',
 'Android Ver']
```

```
type',
'Content Rating',
'Genres',
'Last Updated',
'Current Ver',
'Android Ver']
```

In [21]:

```
df[categorical_features]
```

Out[21]:

	App	Category	Type	Content Rating	Genres	Last Updated	Current Ver	Android Ver
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	Free	Everyone	Art & Design	2018-01-07	1.0.0	4.0.3
1	Coloring book moana	ART_AND_DESIGN	Free	Everyone	Art & Design;Pretend Play	2018-01-15	2.0.0	4.0.3
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	Free	Everyone	Art & Design	2018-08-01	1.2.4	4.0.3
3	Sketch - Draw & Paint	ART_AND_DESIGN	Free	Teen	Art & Design	2018-06-08	Varies with device	4.2
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	Free	Everyone	Art & Design;Creativity	2018-06-20	1.1	4.4
...	...	...	...	...	...	...	...	...
10835	Sya9a Maroc - FR	FAMILY	Free	Everyone	Education	2017-07-25	1.48	4.1
10836	Fr. Mike Schmitz Audio Teachings	FAMILY	Free	Everyone	Education	2018-07-06	1.0	4.1
10837	Parkinson Exercices FR	MEDICAL	Free	Everyone	Medical	2017-01-20	1.0	2.2
10838	The SCP Foundation DB fr nn5n	BOOKS_AND_REFERENCE	Free	Mature 17+	Books & Reference	2015-01-19	Varies with device	NaN
10839	iHoroscope - 2018 Daily Horoscope & Astrology	LIFESTYLE	Free	Everyone	Lifestyle	2018-07-25	Varies with device	NaN

9659 rows × 8 columns

In [22]:

```
df.head()
```

Out[22]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Cu
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19000.0	10000	Free	0.0	Everyone	Art & Design	2018-01-07	
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14000.0	500000	Free	0.0	Everyone	Art & Design;Pretend Play	2018-01-15	
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	37510	37500.0	5000000	Free	0.0	Everyone	Art & Design	2018-	

2	FREE Live App Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8700.0	5000000	Free	0.0	Everyone	Art & Design	08-01-2018	Current Version
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25000.0	50000000	Free	0.0	Teen	Art & Design	2018-06-08	Version 1.0.0
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2800.0	100000	Free	0.0	Everyone	Art & Design; Creativity	2018-06-20	Version 1.0.0

In [23]:

```
categorical_features
```

Out[23]:

```
['App',  
 'Category',  
 'Type',  
 'Content Rating',  
 'Genres',  
 'Last Updated',  
 'Current Ver',  
 'Android Ver']
```

In [24]:

```
df["Type"].value_counts(normalize=True)
```

Out[24]:

```
Free      0.921723  
Paid      0.078277  
Name: Type, dtype: float64
```

In [25]:

```
categorical_features
```

Out[25]:

```
['App',  
 'Category',  
 'Type',  
 'Content Rating',  
 'Genres',  
 'Last Updated',  
 'Current Ver',  
 'Android Ver']
```

In [26]:

```
for col in categorical_features:  
    print(col)
```

```
App  
Category  
Type  
Content Rating  
Genres  
Last Updated  
Current Ver  
Android Ver
```

In [27]:

```
for col in categorical_features:  
    print(f"{col}:{df[col].value_counts(normalize=True)*100}")
```

```
print("=====")
```

```
App:Photo Editor & Candy Camera & Grid & ScrapBook      0.010353
Tic Tac CK                                              0.010353
INFAMY RO                                              0.010353
CK Call NEW                                            0.010353
Ck Coif                                              0.010353
...
Bike Race Free - Top Motorcycle Racing Games          0.010353
Dance School Stories - Dance Dreams Come True        0.010353
3D Bowling                                            0.010353
Mind Games Pro                                        0.010353
iHoroscope - 2018 Daily Horoscope & Astrology         0.010353
Name: App, Length: 9659, dtype: float64
=====
Category:FAMILY                                     18.966767
GAME                                                  9.928564
TOOLS                                                  8.561963
BUSINESS                                              4.348276
MEDICAL                                              4.089450
PERSONALIZATION                                     3.892743
PRODUCTIVITY                                         3.872036
LIFESTYLE                                            3.820271
FINANCE                                              3.571798
SPORTS                                               3.364738
COMMUNICATION                                       3.261207
HEALTH_AND_FITNESS                                 2.981675
PHOTOGRAPHY                                         2.909204
NEWS_AND_MAGAZINES                                2.629672
SOCIAL                                              2.474376
BOOKS_AND_REFERENCE                               2.298375
TRAVEL_AND_LOCAL                                  2.267315
SHOPPING                                            2.091314
DATING                                              1.770370
VIDEO_PLAYERS                                       1.687545
MAPS_AND_NAVIGATION                               1.356248
EDUCATION                                           1.232012
FOOD_AND_DRINK                                     1.159540
ENTERTAINMENT                                       1.056010
AUTO_AND_VEHICLES                                 0.880008
LIBRARIES_AND_DEMO                                0.869655
WEATHER                                             0.817890
HOUSE_AND_HOME                                    0.766125
EVENTS                                              0.662594
ART_AND_DESIGN                                    0.662594
PARENTING                                          0.621182
COMICS                                              0.579770
BEAUTY                                              0.548711
Name: Category, dtype: float64
=====
Type:Free      92.172292
Paid          7.827708
Name: Type, dtype: float64
=====
Content Rating:Everyone                             81.820064
Teen          10.725748
Mature 17+    4.068744
Everyone 10+  3.333678
Adults only 18+ 0.031059
Unrated      0.020706
Name: Content Rating, dtype: float64
=====
Genres:Tools                                     8.551610
Entertainment  5.808055
Education      5.280050
Business       4.348276
Medical        4.089450
...
Art & Design;Pretend Play 0.010353
Lifestyle;Pretend Play   0.010353
Comics;Creativity       0.010353
Art & Design;Action & Adventure 0.010353
.....
```

```

Strategy;Creativity                                0.010353
Name: Genres, Length: 118, dtype: float64
=====
Last Updated:2018-08-03      2.608966
2018-08-02      2.381199
2018-08-01      2.163785
2018-07-31      2.112020
2018-07-30      1.656486
...
2015-03-08      0.010353
2015-03-25      0.010353
2014-07-09      0.010353
2014-10-11      0.010353
2014-03-23      0.010353
Name: Last Updated, Length: 1377, dtype: float64
=====
Current Ver:Varies with device      10.931510
1.0      8.278935
1.1      2.694021
1.2      1.823645
2.0      1.543881
...
2.27.1      0.010362
1.6.11      0.010362
20180517      0.010362
3.1.00      0.010362
2.0.148.0      0.010362
Name: Current Ver, Length: 2817, dtype: float64
=====
Android Ver:4.1      25.406715
4.0.3      16.095535
4.0      14.826353
4.4      9.438099
2.3      7.107419
5.0      5.907465
4.2      4.292143
2.3.3      3.149879
2.2      2.757586
3.0      2.665282
4.3      2.549902
2.1      1.534556
1.6      1.338410
6.0      0.623053
7.0      0.484597
3.2      0.415369
2.0      0.369217
5.1      0.253836
1.5      0.230760
4.4W      0.126918
3.1      0.115380
2.0.1      0.080766
8.0      0.069228
7.1      0.034614
5.0 - 8.0      0.023076
4.0.3 - 7.1.1      0.023076
1.0      0.023076
7.0 - 7.1.1      0.011538
4.1 - 7.1.1      0.011538
5.0 - 6.0      0.011538
2.2 - 7.1.1      0.011538
5.0 - 7.1.1      0.011538
Name: Android Ver, dtype: float64
=====

```

In [28]:

```
sns.countplot(x=df["Type"])
```

Out[28]:

```
<AxesSubplot:xlabel='Type', ylabel='count'>
```



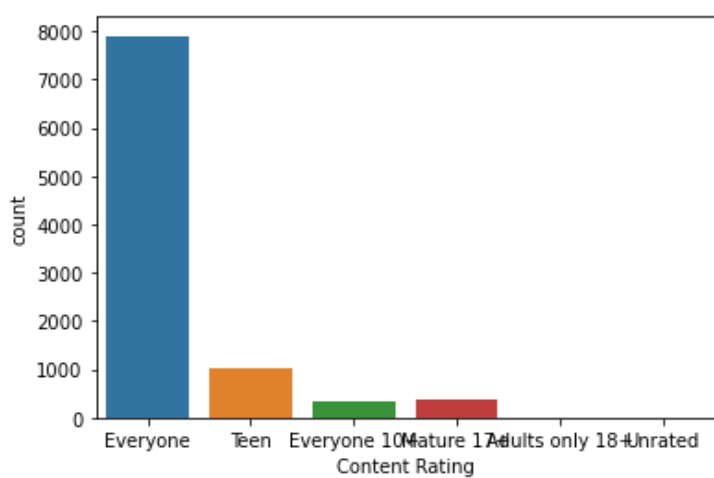


In [29]:

```
sns.countplot(x=df['Content Rating'])
```

Out[29]:

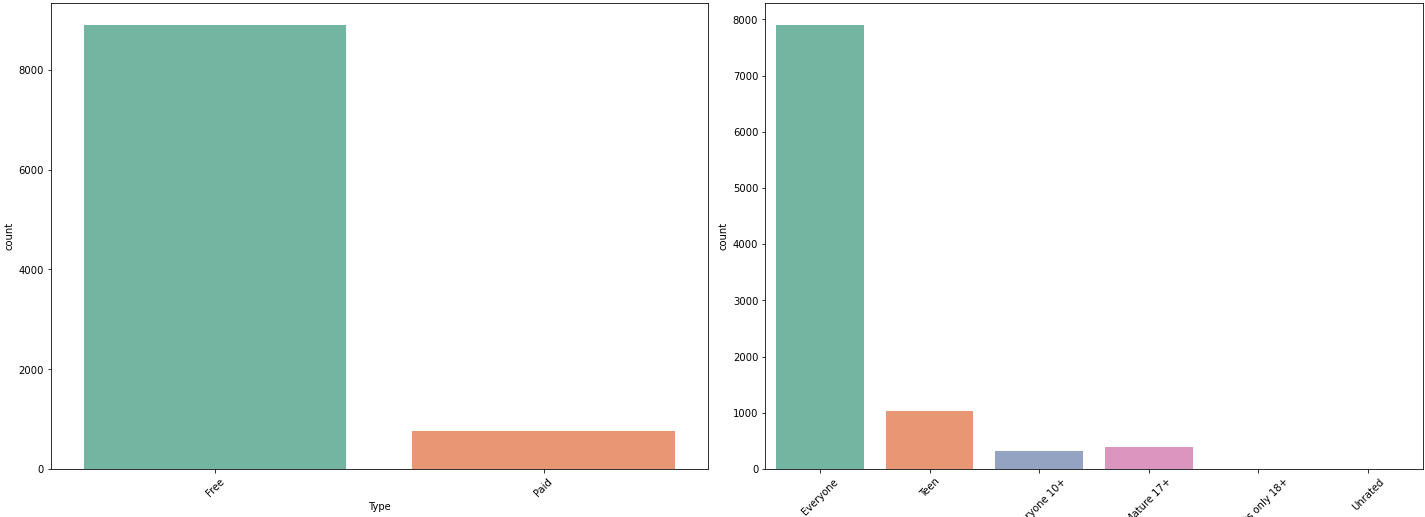
<AxesSubplot:xlabel='Content Rating', ylabel='count'>



In [30]:

```
# categorical columns
plt.figure(figsize=(20, 15))
plt.suptitle('Univariate Analysis of Categorical Features', fontsize=20, fontweight='bold', alpha=0.8, y=1.)
category = ['Type', 'Content Rating']
for i in range(0, len(category)):
    plt.subplot(2, 2, i+1)
    sns.countplot(x=df[category[i]],palette="Set2")
    plt.xlabel(category[i])
    plt.xticks(rotation=45)
plt.tight_layout()
```

Univariate Analysis of Categorical Features



In [31]:

```
numeric_features
```

Out[31]:

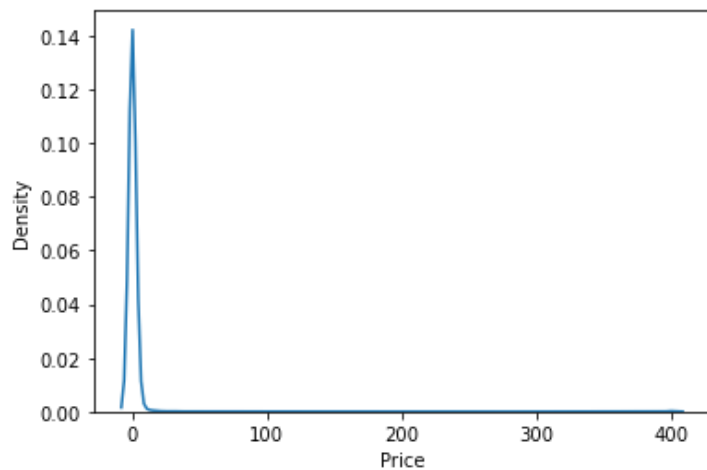
```
['Rating', 'Reviews', 'Size', 'Installs', 'Price', 'day', 'month', 'year']
```

In [32]:

```
sns.kdeplot(df['Price'])
```

Out[32]:

<AxesSubplot:xlabel='Price', ylabel='Density'>

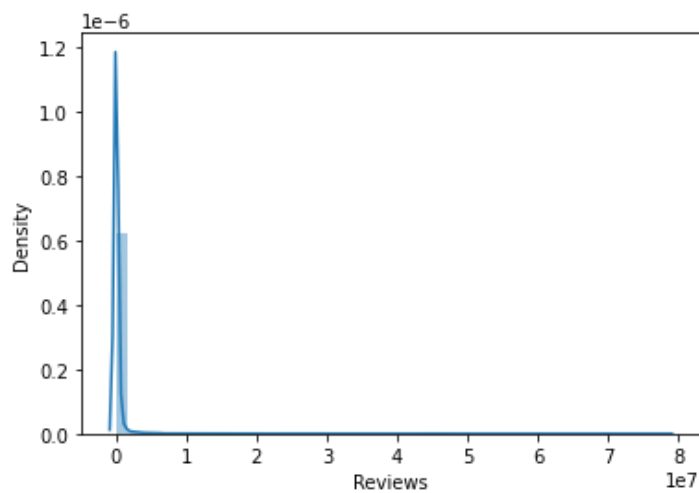


In [33]:

```
sns.distplot(df['Reviews'])
```

Out[33]:

<AxesSubplot:xlabel='Reviews', ylabel='Density'>



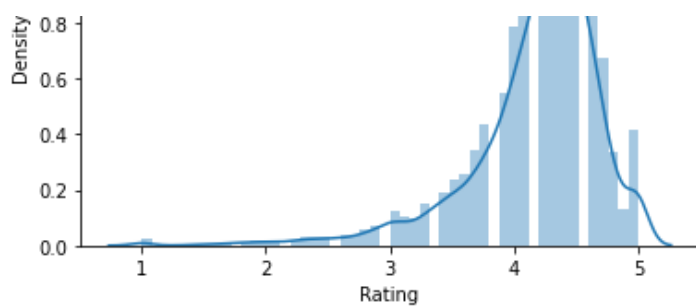
In [34]:

```
sns.distplot(df['Rating'])
```

Out[34]:

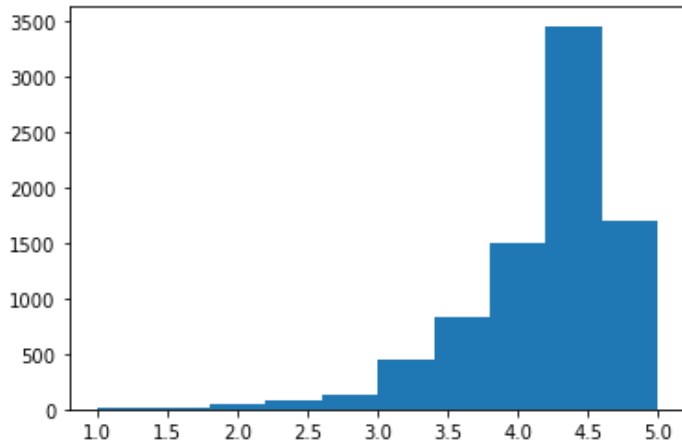
<AxesSubplot:xlabel='Rating', ylabel='Density'>





In [35]:

```
plt.hist(df["Rating"])
plt.show()
```

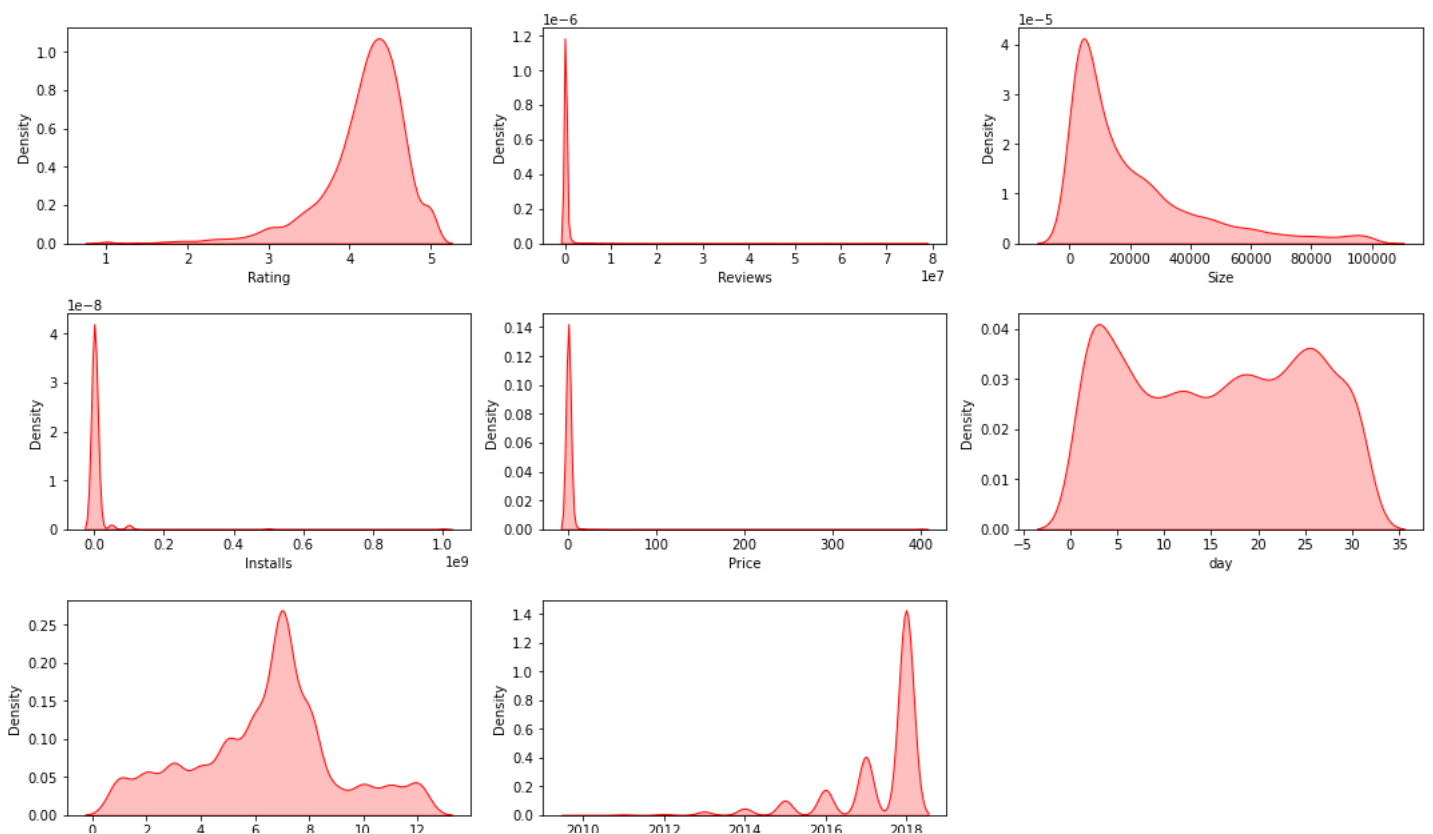


In [36]:

```
plt.figure(figsize=(15, 15))
plt.suptitle('Univariate Analysis of Numerical Features', fontsize=20, fontweight='bold',
alpha=0.8, y=1.)

for i in range(0, len(numeric_features)):
    plt.subplot(5, 3, i+1)
    sns.kdeplot(x=df[numeric_features[i]],shade=True, color='r')
    plt.xlabel(numeric_features[i])
    plt.tight_layout()
```

### Univariate Analysis of Numerical Features



In [37]:

```
df.head()
```

Out[37]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Cu
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19000.0	10000	Free	0.0	Everyone	Art & Design	2018-01-07	
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14000.0	500000	Free	0.0	Everyone	Art & Design;Pretend Play	2018-01-15	
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8700.0	5000000	Free	0.0	Everyone	Art & Design	2018-08-01	
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25000.0	50000000	Free	0.0	Teen	Art & Design	2018-06-08	V
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2800.0	100000	Free	0.0	Everyone	Art & Design;Creativity	2018-06-20	d

In [38]:

```
#univariate analysis
#most popular category in app

File "C:\Users\sunny\AppData\Local\Temp\ipykernel_38584\920096547.py", line 1
    univariate analysis
    ^
```

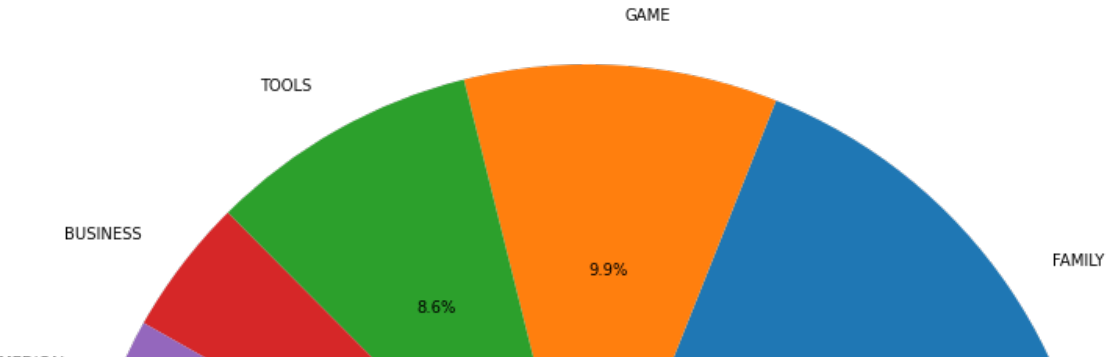
SyntaxError: invalid syntax

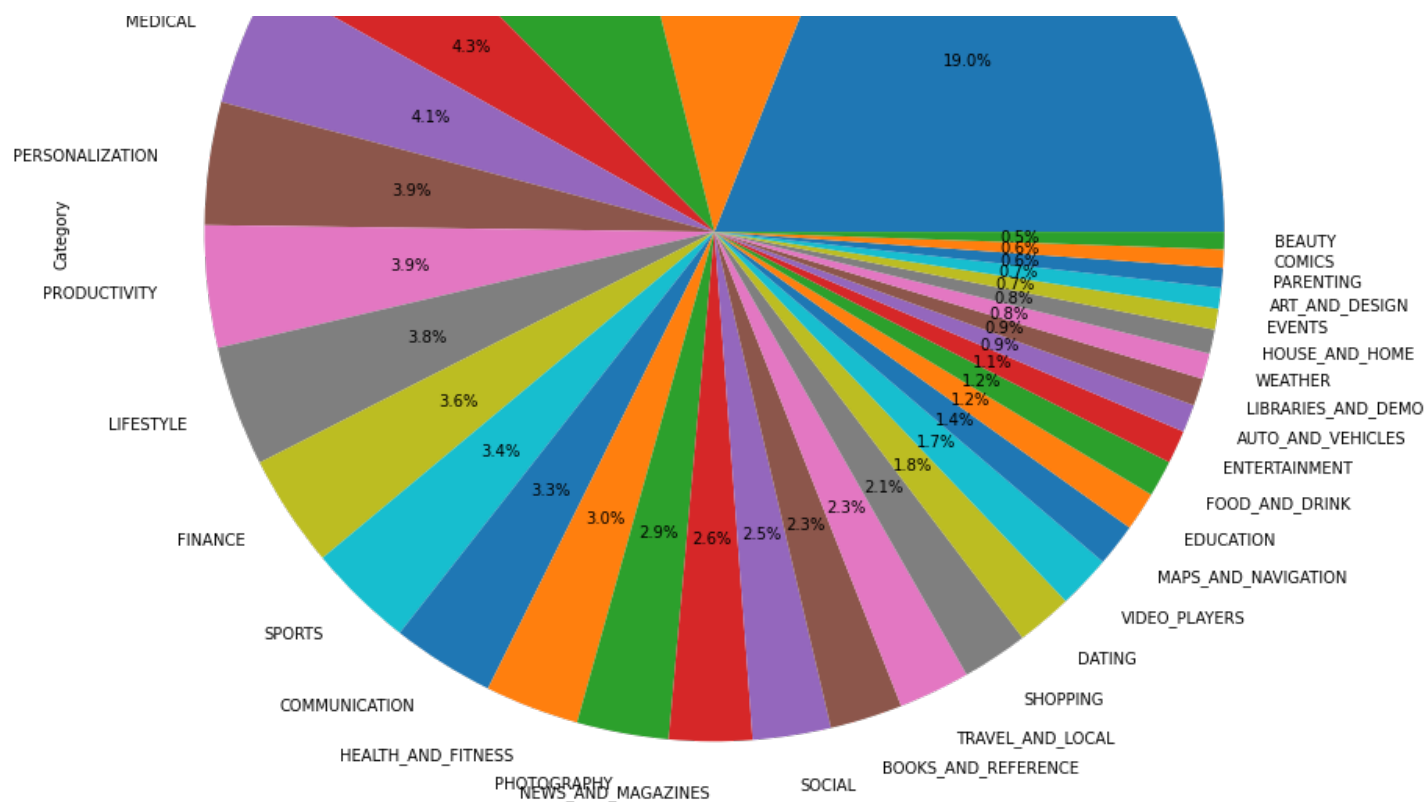
In [44]:

```
df["Category"].value_counts().plot.pie(y=df["Category"],figsize=(15,15),autopct='%1.1f%%')
```

Out[44]:

<AxesSubplot:ylabel='Category'>





```
In [48]:
```

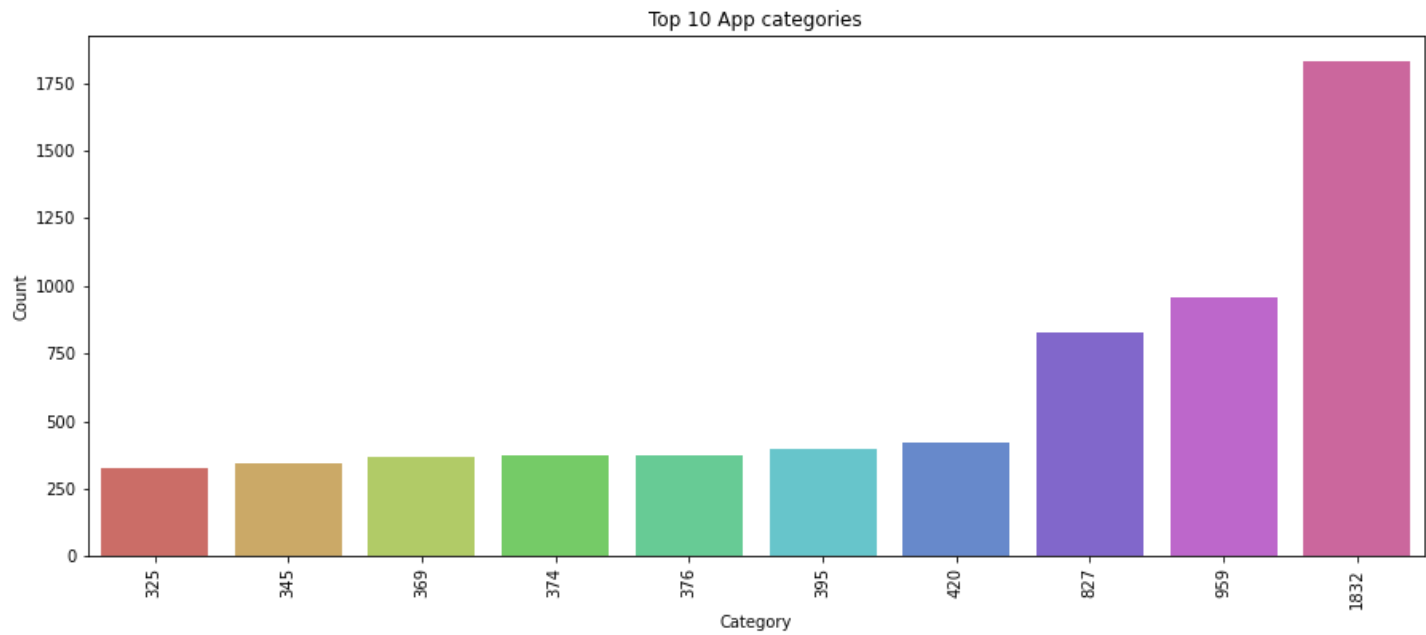
```
df_cat=df["Category"].value_counts()[ :10]
```

```
In [51]:
```

```
category = pd.DataFrame(df['Category'].value_counts()) #Dataframe of apps on the basis of category
category.rename(columns = {'Category':'Count'}, inplace=True)
```

```
In [52]:
```

```
plt.figure(figsize=(15,6))
sns.barplot(x=df_cat, y ='Count',data = category[:10],palette='hls')
plt.title('Top 10 App categories')
plt.xticks(rotation=90)
plt.show()
```



```
In [53]:
```

```
df.head()
```

Out [53]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Cu
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19000.0	10000	Free	0.0	Everyone	Art & Design	2018-01-07	
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14000.0	500000	Free	0.0	Everyone	Art & Design;Pretend Play	2018-01-15	
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8700.0	5000000	Free	0.0	Everyone	Art & Design	2018-08-01	
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25000.0	50000000	Free	0.0	Teen	Art & Design	2018-06-08	V
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2800.0	100000	Free	0.0	Everyone	Art & Design;Creativity	2018-06-20	d

In [ ]:

*#which category having largest number of installaion*

In [69]:

```
type(df.groupby(['Category'])['Installs'].sum().sort_values(ascending=False))
```

Out [69]:

pandas.core.series.Series

In [61]:

```
df_cat_installs=df.groupby(['Category'])['Installs'].sum().sort_values(ascending=False).reset_index()
```

In [64]:

```
df2=df_cat_installs.head(10)
```

In [65]:

df2

Out [65]:

	Category	Installs
0	GAME	13878924415
1	COMMUNICATION	11038276251
2	TOOLS	8001771915
3	PRODUCTIVITY	5793091369
4	SOCIAL	5487867902
5	PHOTOGRAPHY	4649147655

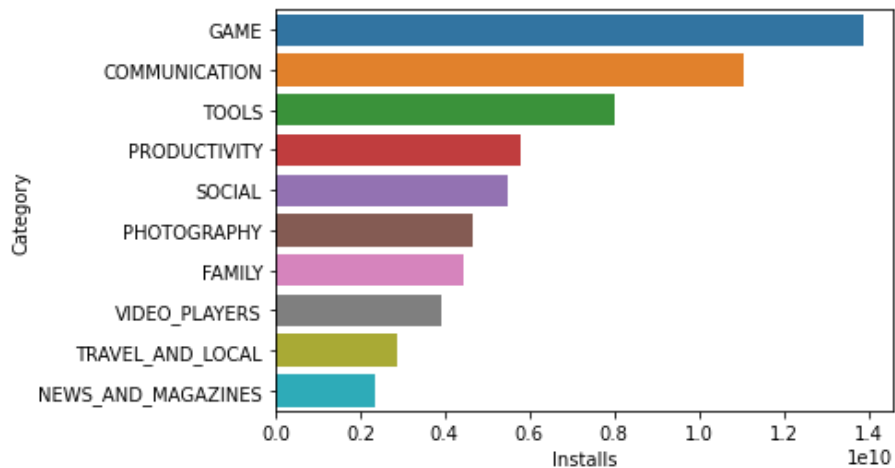
6	Category	Installs
	FAMILY	4427941505
7	VIDEO_PLAYERS	3926902720
8	TRAVEL_AND_LOCAL	2894887146
9	NEWS_AND_MAGAZINES	2369217760

In [66]:

```
sns.barplot(x="Installs",y="Category",data=df2)
```

Out[66]:

```
<AxesSubplot:xlabel='Installs', ylabel='Category'>
```



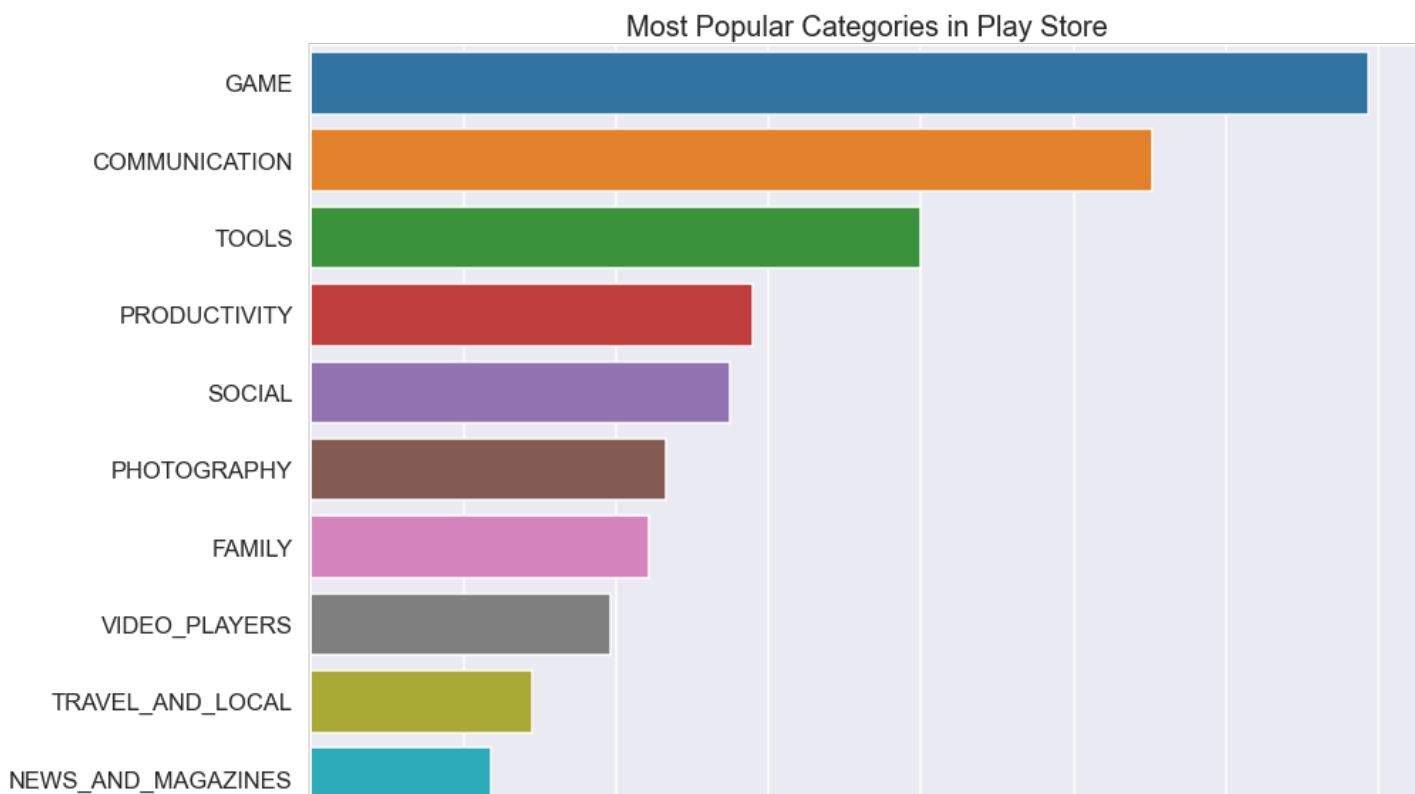
In [67]:

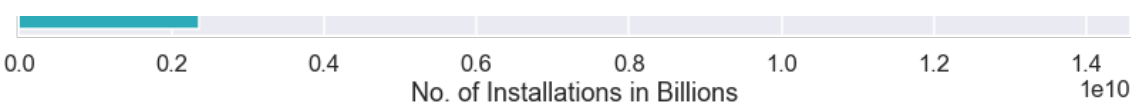
```
plt.figure(figsize = (14,10))
sns.set_context("talk")
sns.set_style("darkgrid")

ax = sns.barplot(x = 'Installs' , y = 'Category' , data = df2 )
ax.set_xlabel('No. of Installations in Billions')
ax.set_ylabel('')
ax.set_title("Most Popular Categories in Play Store", size = 20)
```

Out[67]:

```
Text(0.5, 1.0, 'Most Popular Categories in Play Store')
```





What are the Top 5 most installed Apps in Each popular Categories ??

In [70]:

```
df.head()
```

Out[70]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Cu
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19000.0	10000	Free	0.0	Everyone	Art & Design	2018-01-07	
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14000.0	500000	Free	0.0	Everyone	Art & Design;Pretend Play	2018-01-15	
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8700.0	5000000	Free	0.0	Everyone	Art & Design	2018-08-01	
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25000.0	50000000	Free	0.0	Teen	Art & Design	2018-06-08	V
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2800.0	100000	Free	0.0	Everyone	Art & Design;Creativity	2018-06-20	d

In [ ]:

```
install--app--category
```

In [74]:

```
dfa=df.groupby(['Category','App'])['Installs'].sum().reset_index()
```

In [75]:

```
dfa
```

Out[75]:

	Category	App	Installs
0	ART_AND_DESIGN	350 Diy Room Decor Ideas	10000
1	ART_AND_DESIGN	3D Color Pixel by Number - Sandbox Art Coloring	100000
2	ART_AND_DESIGN	AJ Styles HD Wallpapers	5000
3	ART_AND_DESIGN	AJ Styles Wallpaper 2018 - AJ Styles HD Wallpaper	1000
4	ART_AND_DESIGN	Ai illustrator viewer	100000
...	...	...	...
9654	WEATHER	Yahoo Weather	10000000



	Category	App	Installs
9655	WEATHER	Yahoo! Weather for SH Forecast for understandi...	1000000
9656	WEATHER	Yandex.Weather	10000000
9657	WEATHER	weather - weather forecast	1000000
9658	WEATHER	wetter.com - Weather and Radar	10000000

9659 rows × 3 columns

In [80]:

```
dfa=dfa.sort_values('Installs',ascending=False)
```

In [82]:

```
dfa[dfa.Category=='GAME'][:5]
```

Out[82]:

	Category	App	Installs
4773	GAME	Subway Surfers	1000000000
4167	GAME	Candy Crush Saga	500000000
4628	GAME	Pou	500000000
4802	GAME	Temple Run 2	500000000
4580	GAME	My Talking Tom	500000000

In [83]:

```
dfa[dfa.Category=='PRODUCTIVITY'][:5]
```

Out[83]:

	Category	App	Installs
7427	PRODUCTIVITY	Google Drive	1000000000
7472	PRODUCTIVITY	Microsoft Word	500000000
7425	PRODUCTIVITY	Google Calendar	500000000
7364	PRODUCTIVITY	Dropbox	500000000
7334	PRODUCTIVITY	Cloud Print	500000000

In [84]:

```
dfa[dfa.Category=='FAMILY'][:5]
```

Out[84]:

	Category	App	Installs
3333	FAMILY	Tom Loves Angela	100000000
3290	FAMILY	Talking Tom Cat 2	100000000
1976	FAMILY	Bitmoji – Your Personal Emoji	100000000
3289	FAMILY	Talking Tom Cat	100000000
3287	FAMILY	Talking Tom & Ben News	100000000

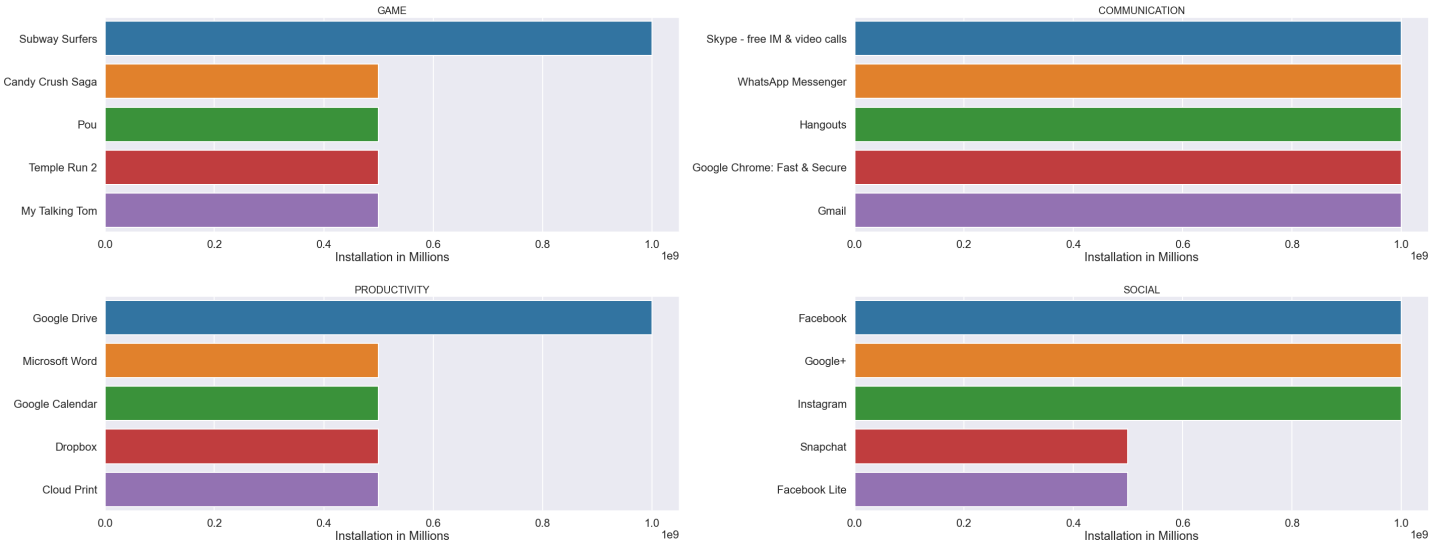
In [88]:

```
apps = ['GAME', 'COMMUNICATION', 'PRODUCTIVITY', 'SOCIAL' ]
sns.set_context("poster")
sns.set_style("darkgrid")
```

```
plt.figure(figsize=(40,30))

for i,app in enumerate(apps):
    df2 = dfa[dfa.Category == app]
    df3 = df2.head(5)
    plt.subplot(4,2,i+1)
    sns.barplot(data= df3,x= 'Installs' ,y='App' )
    plt.xlabel('Installation in Millions')
    plt.ylabel('')
    plt.title(app,size = 20)

plt.tight_layout()
plt.subplots_adjust(hspace= .3)
plt.show()
```



How many apps are there on Google Play Store which get 5 ratings??

In [89]:

```
df.head()
```

Out[89]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Cu
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19000.0	10000	Free	0.0	Everyone	Art & Design	2018-01-07	
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14000.0	500000	Free	0.0	Everyone	Art & Design;Pretend Play	2018-01-15	
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8700.0	5000000	Free	0.0	Everyone	Art & Design	2018-08-01	
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25000.0	50000000	Free	0.0	Teen	Art & Design	2018-06-08	V d
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2800.0	100000	Free	0.0	Everyone	Art & Design;Creativity	2018-06-20	

```
In [105]:
```

```
df.groupby(["App"])["Rating"].mean().sort_values(ascending=False).reset_index().shape
```

```
Out[105]:
```

```
(9659, 2)
```

```
In [96]:
```

```
rating=df.groupby(["App"])["Rating"].sum().sort_values(ascending=False).reset_index()
```

```
In [97]:
```

```
rating.Rating==5.0
```

```
Out[97]:
```

	App	Rating
0	Comunidad De Fe Minitries	5.0
1	Tablet Reminder	5.0
2	Noticias FC Barcelona	5.0
3	Superheroes, Marvel, DC, Comics, TV, Movies News	5.0
4	DC N COMPANY ENTERTAINMENT RADIO!	5.0
...	...	...
9654	DK Childcare Centers	0.0
9655	Pediatric Emergency Guide	0.0
9656	Pekalongan CJ	0.0
9657	Penn State Health OnDemand	0.0
9658	"i DT" Fútbol. Todos Somos Técnicos.	0.0

9659 rows x 2 columns

```
In [99]:
```

```
rating[rating.Rating==5.0]
```

```
Out[99]:
```

	App	Rating
0	Comunidad De Fe Minitries	5.0
1	Tablet Reminder	5.0
2	Noticias FC Barcelona	5.0
3	Superheroes, Marvel, DC, Comics, TV, Movies News	5.0
4	DC N COMPANY ENTERTAINMENT RADIO!	5.0
...	...	...
266	American Girls Mobile Numbers	5.0
267	daskal.eu	5.0
268	Free coupons and vouchers	5.0
269	Bk Usha behn	5.0
270	Asha Ek Hope - ALS/ MND	5.0

271 rows x 2 columns

In [100]:

```
rating2 = df.groupby(['Category', 'Installs', 'App'])['Rating'].sum().sort_values(ascending = False).reset_index()
```

In [101]:

```
rating2
```

Out[101]:

	Category	Installs	App	Rating
0	FAMILY	1000	CS & IT Interview Questions	5.0
1	DATING	100	Online Girls Chat Group	5.0
2	FAMILY	10	Chronolink DX	5.0
3	DATING	500	Spine- The dating app	5.0
4	MEDICAL	5	Clinic Doctor EHR	5.0
...	...	...	...	...
9654	TOOLS	100	CG Wizard	0.0
9655	TOOLS	100	CG 3D Model Viewer: view 3D models interactively	0.0
9656	BUSINESS	100	CS	0.0
9657	TOOLS	100	CD View Lite	0.0
9658	ART_AND_DESIGN	0	Test Application DT 02	0.0

9659 rows x 4 columns

In [106]:

```
df.columns
```

Out[106]:

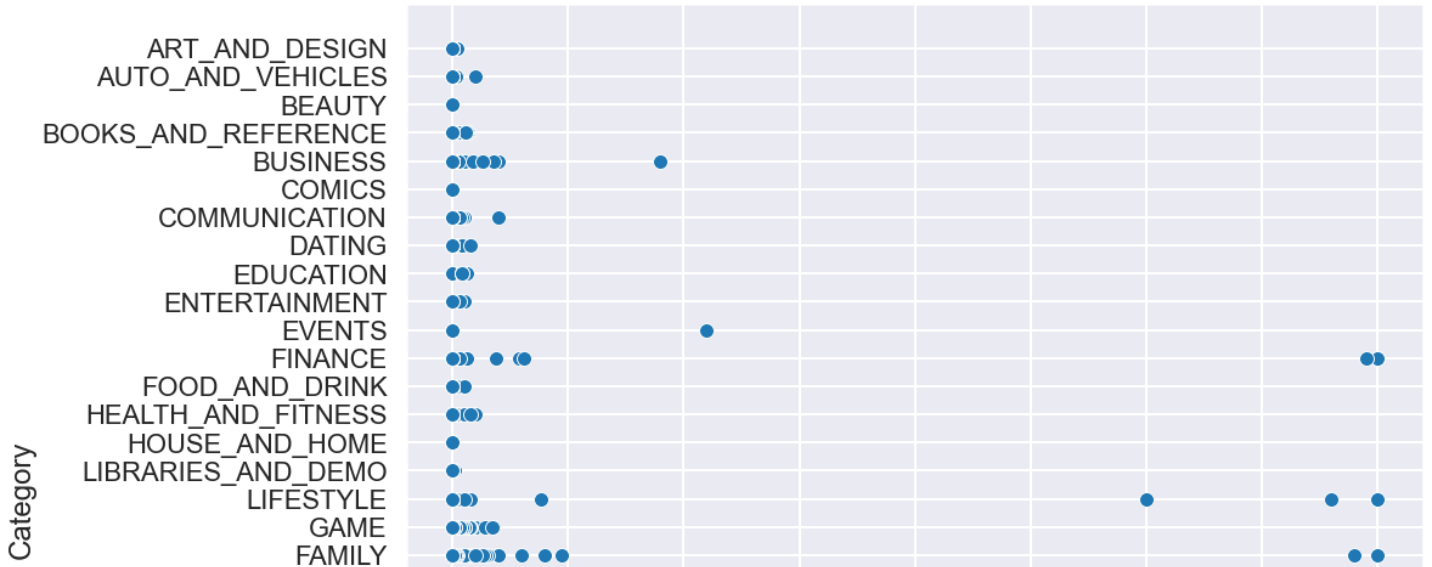
```
Index(['App', 'Category', 'Rating', 'Reviews', 'Size', 'Installs', 'Type',  
      'Price', 'Content Rating', 'Genres', 'Last Updated', 'Current Ver',  
      'Android Ver', 'day', 'month', 'year'],  
      dtype='object')
```

In [109]:

```
plt.figure(figsize=(15,15))  
sns.scatterplot(data=df,y="Category",x='Price')
```

Out[109]:

```
<AxesSubplot:xlabel='Price', ylabel='Category'>
```





## Which Category app users are reviewing the most ??

In [111]:

```
df.head(1)
```

Out[111]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Version
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19000.0	10000	Free	0.0	Everyone	Art & Design	2018-01-07	1.0.0	4.1

In [115]:

```
df.groupby(['Category', 'App'])["Reviews"].sum().reset_index()
```

Out[115]:

	Category	App	Reviews
0	ART_AND_DESIGN	350 Diy Room Decor Ideas	27
1	ART_AND_DESIGN	3D Color Pixel by Number - Sandbox Art Coloring	1518
2	ART_AND_DESIGN	AJ Styles HD Wallpapers	48
3	ART_AND_DESIGN	AJ Styles Wallpaper 2018 - AJ Styles HD Wallpaper	18
4	ART_AND_DESIGN	Ai illustrator viewer	486
...	...	...	...
9654	WEATHER	Yahoo Weather	1312037
9655	WEATHER	Yahoo! Weather for SH Forecast for understandi...	7457
9656	WEATHER	Yandex.Weather	309617
9657	WEATHER	weather - weather forecast	11118
9658	WEATHER	wetter.com - Weather and Radar	189313

9659 rows x 3 columns

In [119]:

```
df.groupby("Category").agg({"Installs":"sum", "Reviews":"sum"}).reset_index()
```

Out[119]:

Out[119]:

	Category	Installs	Reviews
0	ART_AND_DESIGN	114338100	1419203
1	AUTO_AND_VEHICLES	53130211	1163666
2	BEAUTY	27197050	396240
3	BOOKS_AND_REFERENCE	1665969576	16721314
4	BUSINESS	697164865	9890245
5	COMICS	44981150	2342071
6	COMMUNICATION	11038276251	285811368
7	DATING	140926107	3623544
8	EDUCATION	352952000	13364148
9	ENTERTAINMENT	2113660000	34762650
10	EVENTS	15973161	161018
11	FAMILY	4427941505	143825488
12	FINANCE	455348734	12662106
13	FOOD_AND_DRINK	211798751	6325028
14	GAME	13878924415	622298709
15	HEALTH_AND_FITNESS	1144022512	21361355
16	HOUSE_AND_HOME	97212461	1929847
17	LIBRARIES_AND_DEMO	52995910	906831
18	LIFESTYLE	503823539	11832671
19	MAPS_AND_NAVIGATION	503281890	17729148
20	MEDICAL	38193177	1182971
21	NEWS_AND_MAGAZINES	2369217760	23130228
22	PARENTING	31521110	958331
23	PERSONALIZATION	1532494782	53543080
24	PHOTOGRAPHY	4649147655	105351270
25	PRODUCTIVITY	5793091369	55590649
26	SHOPPING	1400348785	44551730
27	SOCIAL	5487867902	227927801
28	SPORTS	1096474498	35348813
29	TOOLS	8001771915	229356578
30	TRAVEL_AND_LOCAL	2894887146	26819741
31	VIDEO_PLAYERS	3926902720	67484568
32	WEATHER	361100520	12295164

In [125]:

```
df.groupby('Type').agg({'Installs':sum}).plot.pie(y='Installs')
```

Out[125]:

<AxesSubplot:ylabel='Installs'>



In [123]:

```
df.value_counts('Type').plot.pie(y='Type')
```

Out[123]:

<AxesSubplot:ylabel='None'>



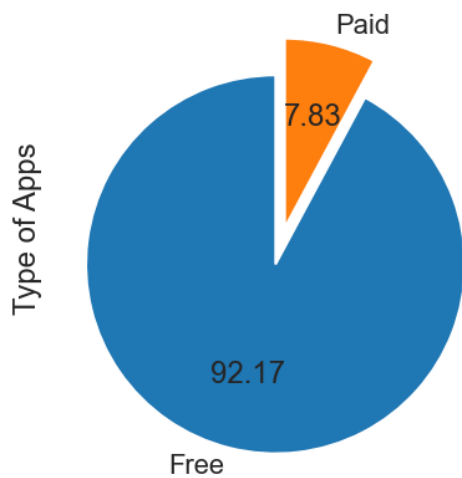
In [121]:

```
fig, ax = plt.subplots(1,2,figsize=(20,7))
df.value_counts('Type').plot.pie(y='Type',startangle=90, explode=(0.2,0), title='Percent
age of the Free App and Paid App', legend=False, autopct='%.2f', ax=ax[0])
ax[0].set(ylabel='Type of Apps')
df.groupby('Type').agg({'Installs':sum}).plot.pie(y='Installs', startangle=90, explode=(
0.2,0), title='Percentage of Installs Number for Free App and Paid App', legend=False, au
topct='%.2f', ax=ax[1])
```

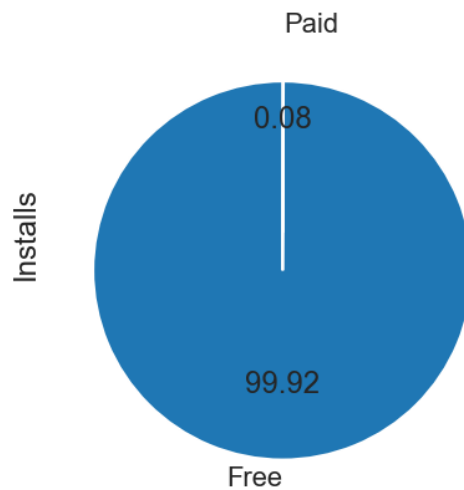
Out[121]:

<AxesSubplot:title={'center':'Percentage of Installs Number for Free App and Paid App'},  
ylabel='Installs'>

Percentage of the Free App and Paid App



Percentage of Installs Number for Free App and Paid App



In [126]:

```
df.head()
```

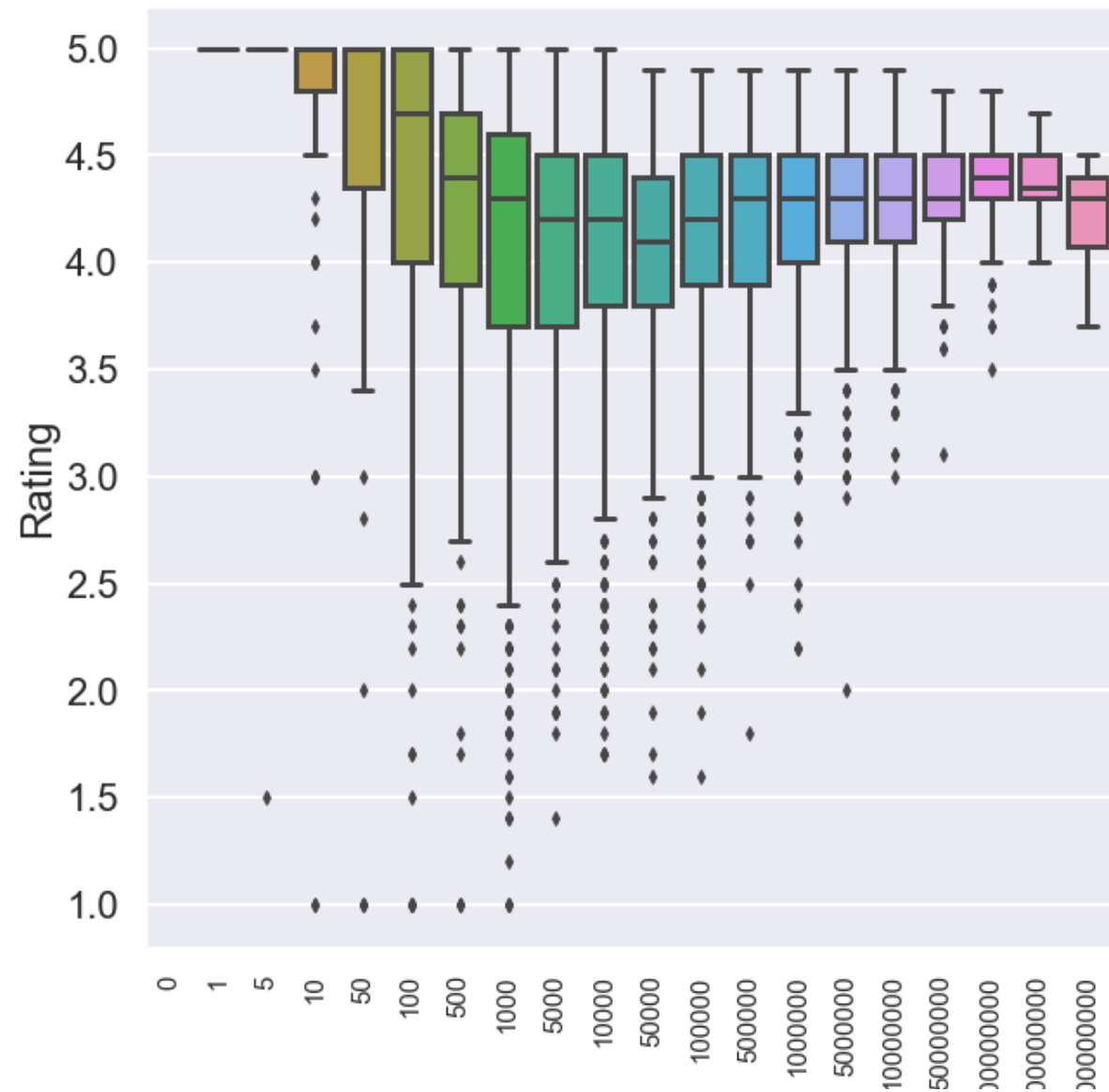
Out[126]:

App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Cu
-----	----------	--------	---------	------	----------	------	-------	-------------------	--------	-----------------	----

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Cu
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19000.0	10000	Free	0.0	Everyone	Art & Design	2018-01-07	
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14000.0	500000	Free	0.0	Everyone	Design;Pretend Play	2018-01-15	
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8700.0	5000000	Free	0.0	Everyone	Art & Design	2018-08-01	
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25000.0	50000000	Free	0.0	Teen	Art & Design	2018-06-08	V
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2800.0	100000	Free	0.0	Everyone	Design;Creativity	2018-06-20	d

In [132]:

```
plt.figure(figsize=(10,10))
sns.boxplot(x="Installs",y="Rating",data=df)
plt.xticks(size=15,rotation=90)
plt.show()
```

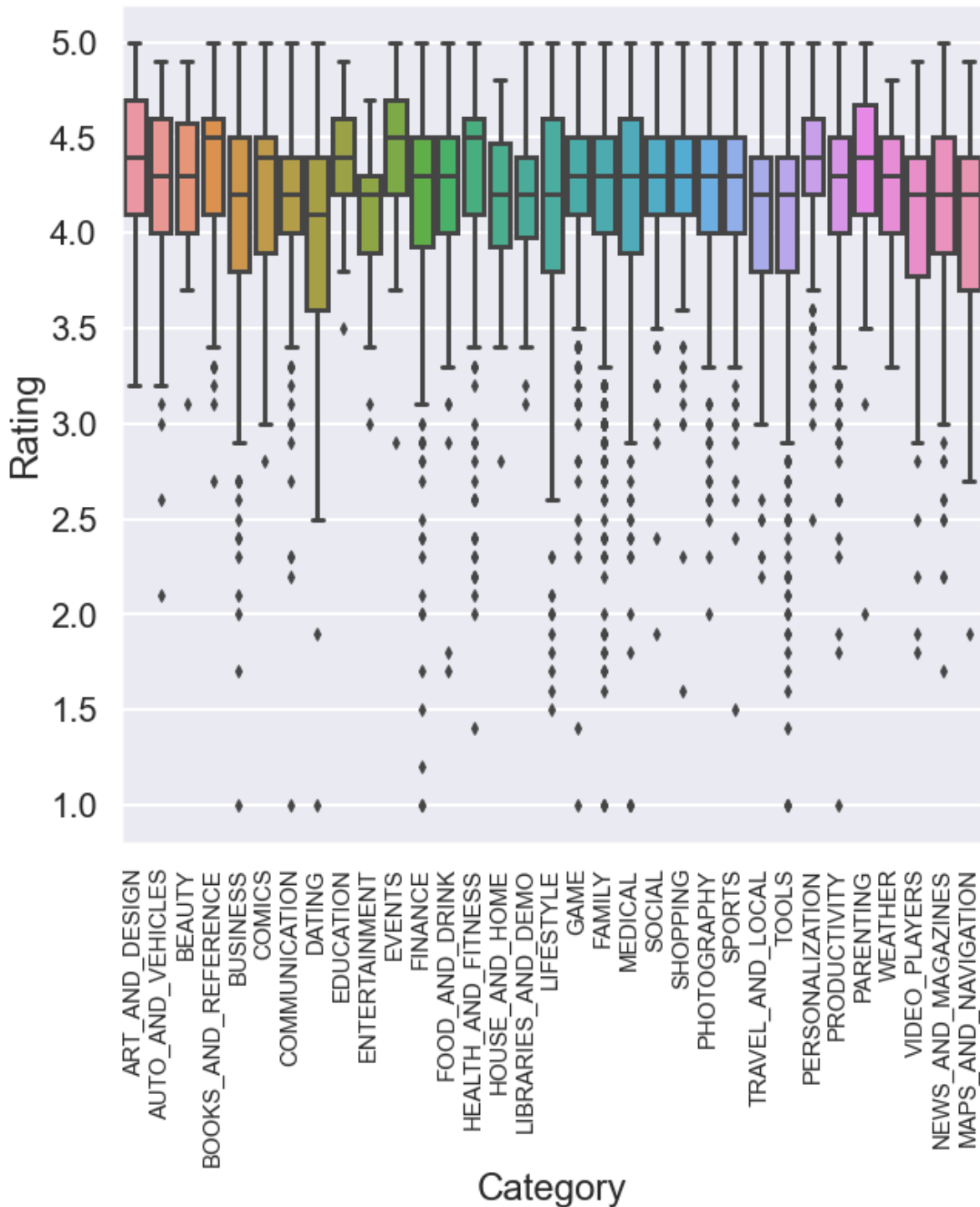




## Installs

In [134]:

```
plt.figure(figsize=(10,10))
sns.boxplot(x="Category",y="Rating",data=df)
plt.xticks(size=15,rotation=90)
plt.show()
```



In [135]:

```
df.head()
```

Out[135]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Cu
0	Photo Editor & Candy Camera & CamScanner	ART_AND_DESIGN	4.1	159	19000.0	10000	Free	0.0	Everyone	Art & Design	2018-01-07	

	Camera & App ScrapBook	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Version
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14000.0	500000	Free	0.0	Everyone	Design;Pretend Play	2018-01-15	
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8700.0	5000000	Free	0.0	Everyone	Art & Design	2018-08-01	
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25000.0	50000000	Free	0.0	Teen	Art & Design	2018-06-08	V
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2800.0	100000	Free	0.0	Everyone	Design;Creativity	2018-06-20	d

Which Category are of largest size ??

In [136]:

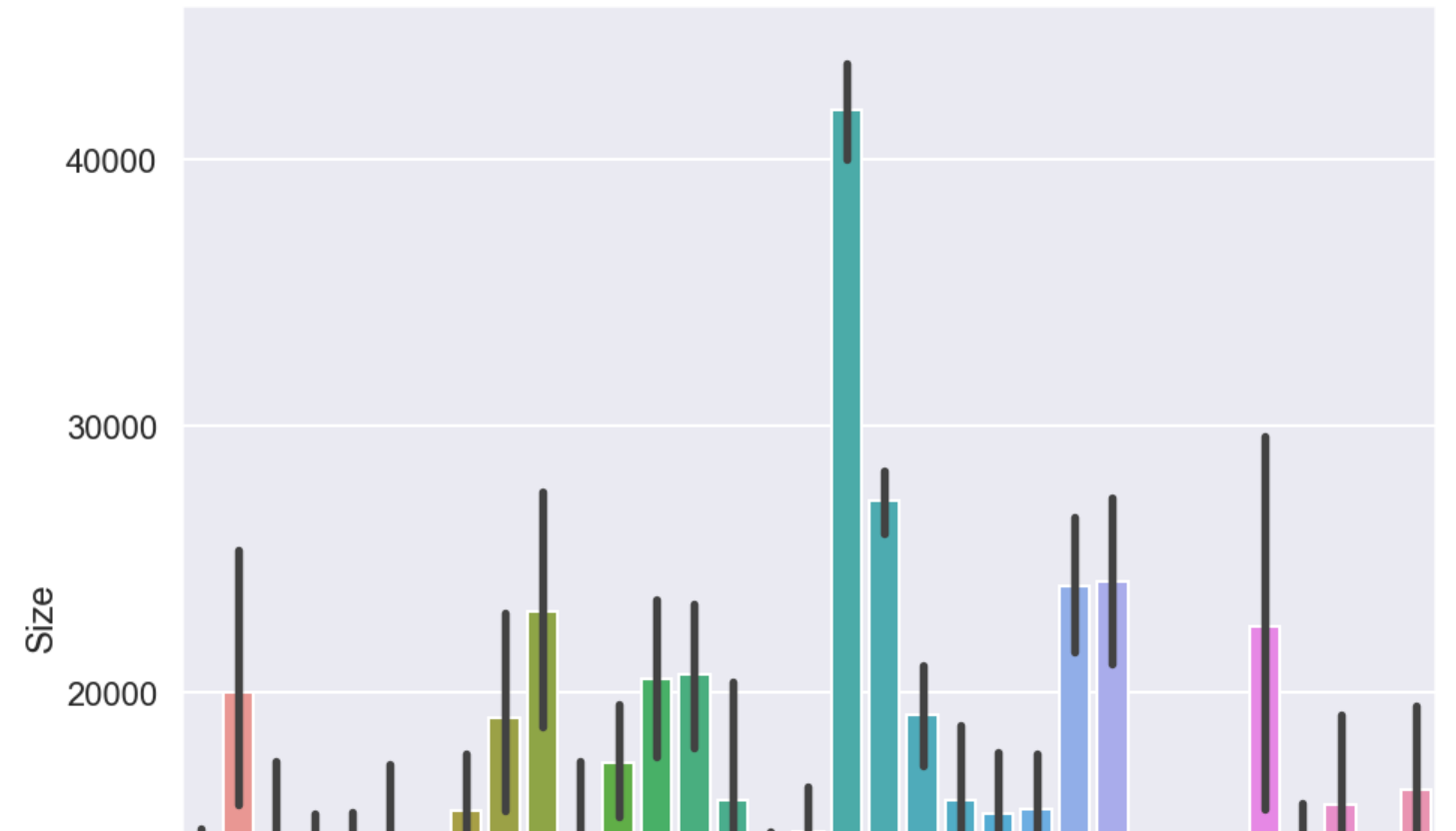
```
df.columns
```

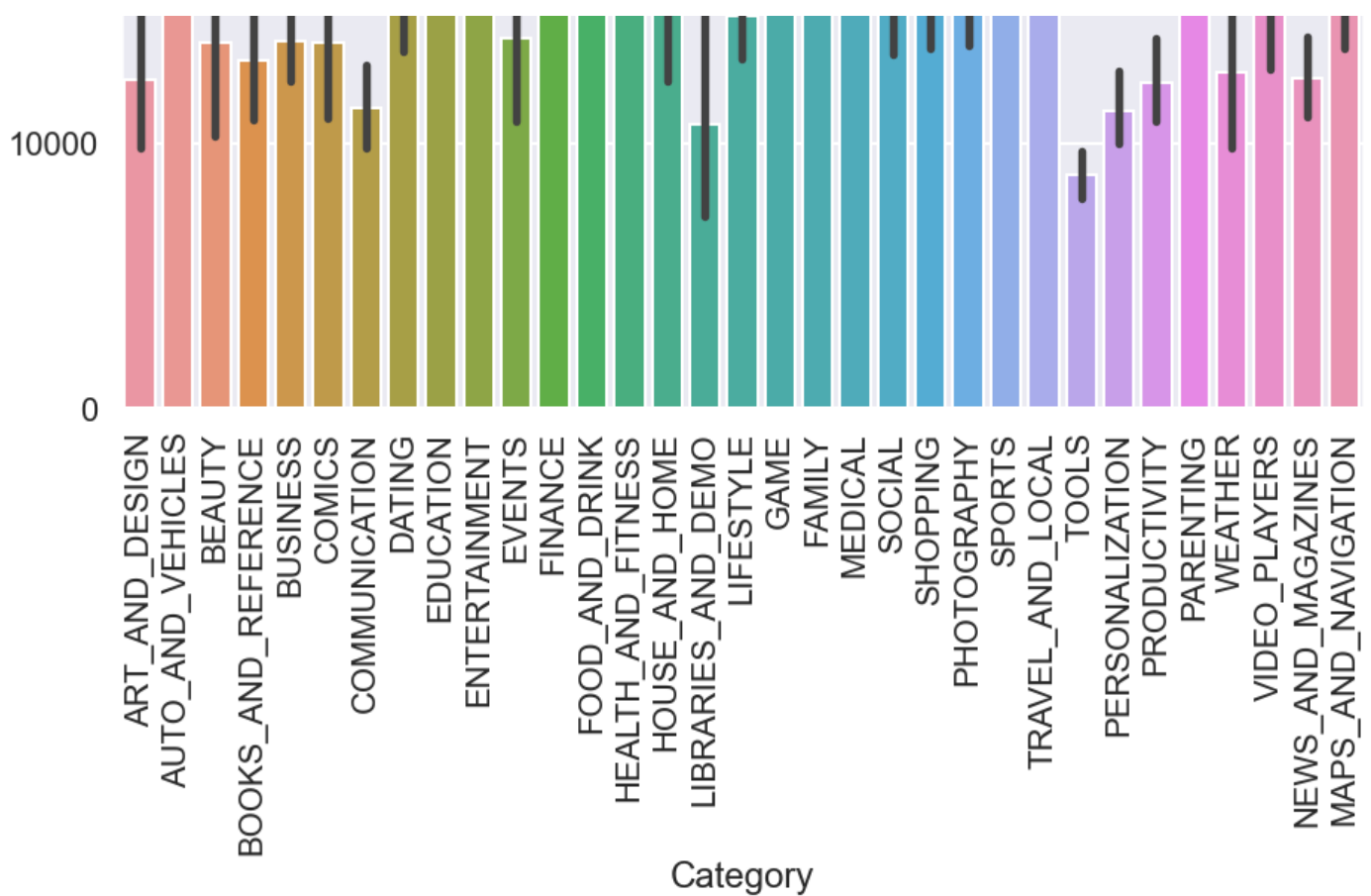
Out[136]:

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

In [139]:

```
plt.figure(figsize=(15,15))
plt.xticks(rotation = 90)
sns.barplot('Category','Size',data=df)
plt.show()
```



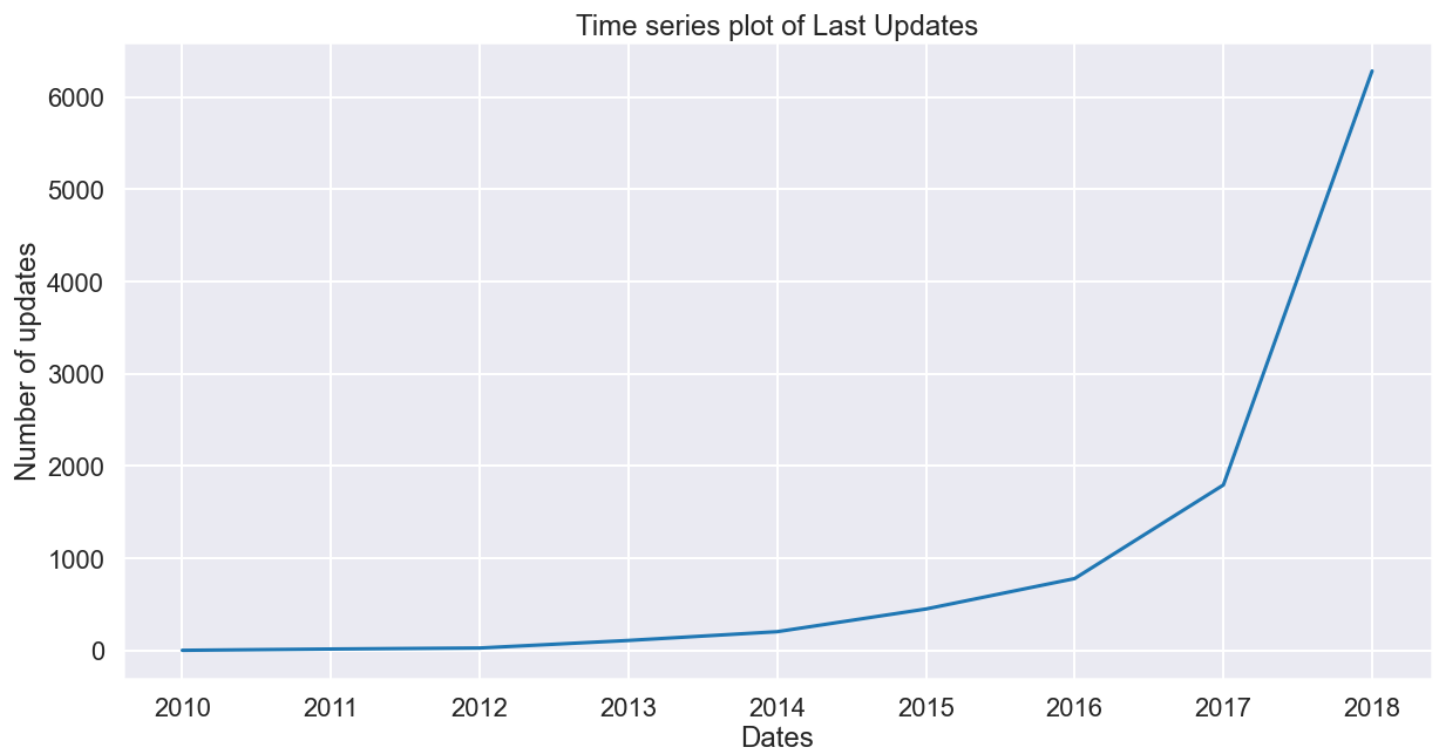


In [141]:

```
plt.subplots(figsize=(20,10))
freq= pd.Series()
freq=df['year'].value_counts()
freq.plot()
plt.xlabel("Dates")
plt.ylabel("Number of updates")
plt.title("Time series plot of Last Updates")
```

Out[141]:

Text(0.5, 1.0, 'Time series plot of Last Updates')



In [142]:

```
df.head()
```

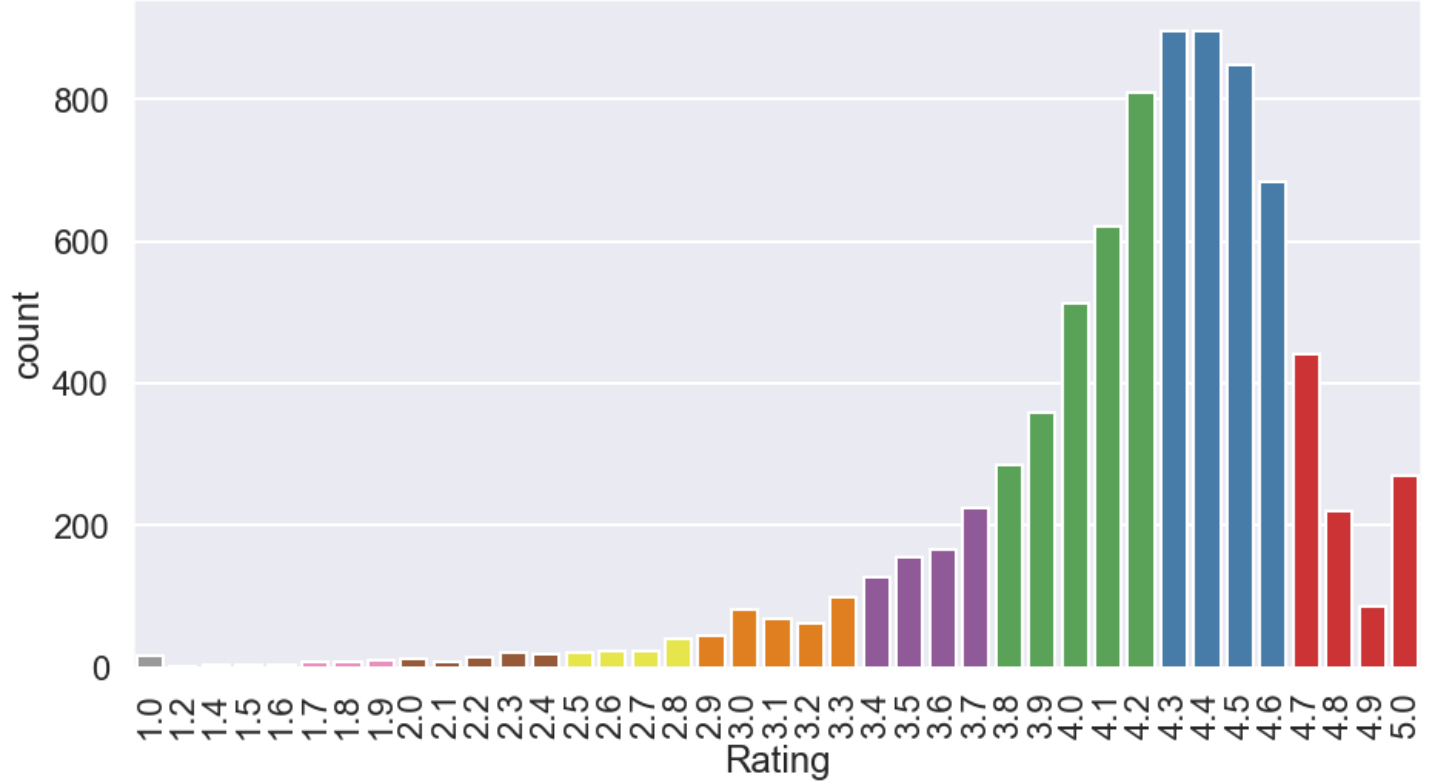
Out[142]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Cu
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19000.0	10000	Free	0.0	Everyone	Art & Design	2018-01-07	
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14000.0	500000	Free	0.0	Everyone	Art & Design;Pretend Play	2018-01-15	
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8700.0	5000000	Free	0.0	Everyone	Art & Design	2018-08-01	
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25000.0	50000000	Free	0.0	Teen	Art & Design	2018-06-08	V
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2800.0	100000	Free	0.0	Everyone	Art & Design;Creativity	2018-06-20	d

In [143]:

```
plt.figure(figsize=(15,8))
sns.countplot(x='Rating',data = df,palette="Set1_r")
plt.xticks(rotation =90)
plt.title('Countplot for ratings')
plt.show()
```

Countplot for ratings



In [ ]:

