

# eda-of-play-store

September 14, 2024

```
[1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

import warnings
warnings.filterwarnings("ignore")
```

```
[3]: df=pd.read_csv("googleplaystore.csv")
df
```

```
[3]:
```

	App	Category \
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN
1	Coloring book moana	ART_AND_DESIGN
2	U Launcher Lite - FREE Live Cool Themes, Hide ...	ART_AND_DESIGN
3	Sketch - Draw & Paint	ART_AND_DESIGN
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN
...	...	...
10836	Sya9a Maroc - FR	FAMILY
10837	Fr. Mike Schmitz Audio Teachings	FAMILY
10838	Parkinson Exercices FR	MEDICAL
10839	The SCP Foundation DB fr nn5n	BOOKS_AND_REFERENCE
10840	iHoroscope - 2018 Daily Horoscope & Astrology	LIFESTYLE

	Rating	Reviews	Size	Installs	Type	Price \
0	4.1	159	19M	10,000+	Free	0
1	3.9	967	14M	500,000+	Free	0
2	4.7	87510	8.7M	5,000,000+	Free	0
3	4.5	215644	25M	50,000,000+	Free	0
4	4.3	967	2.8M	100,000+	Free	0
...	...	...	...	...	...	...
10836	4.5	38	53M	5,000+	Free	0
10837	5.0	4	3.6M	100+	Free	0
10838	NaN	3	9.5M	1,000+	Free	0
10839	4.5	114	Varies with device	1,000+	Free	0
10840	4.5	398307	19M	10,000,000+	Free	0

	Content Rating	Genres	Last Updated \
0	Everyone	Art & Design	January 7, 2018
1	Everyone	Art & Design;Pretend Play	January 15, 2018
2	Everyone	Art & Design	August 1, 2018
3	Teen	Art & Design	June 8, 2018
4	Everyone	Art & Design;Creativity	June 20, 2018
...	...	...	...
10836	Everyone	Education	July 25, 2017
10837	Everyone	Education	July 6, 2018
10838	Everyone	Medical	January 20, 2017
10839	Mature 17+	Books & Reference	January 19, 2015
10840	Everyone	Lifestyle	July 25, 2018

	Current Ver	Android Ver
0	1.0.0	4.0.3 and up
1	2.0.0	4.0.3 and up
2	1.2.4	4.0.3 and up
3	Varies with device	4.2 and up
4	1.1	4.4 and up
...	...	...
10836	1.48	4.1 and up
10837	1.0	4.1 and up
10838	1.0	2.2 and up
10839	Varies with device	Varies with device
10840	Varies with device	Varies with device

[10841 rows x 13 columns]

```
[25]: df.columns
```

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

```
[5]: df.head()
```

```
[5]:
```

	App	Category	Rating \
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1
1	Coloring book moana	ART_AND_DESIGN	3.9
2	U Launcher Lite - FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3

	Reviews	Size	Installs	Type	Price	Content Rating \
0	159	19M	10,000+	Free	0	Everyone
1	967	14M	500,000+	Free	0	Everyone

2	87510	8.7M	5,000,000+	Free	0	Everyone
3	215644	25M	50,000,000+	Free	0	Teen
4	967	2.8M	100,000+	Free	0	Everyone

	Genres	Last Updated	Current Ver	\
0	Art & Design	January 7, 2018	1.0.0	
1	Art & Design;Pretend Play	January 15, 2018	2.0.0	
2	Art & Design	August 1, 2018	1.2.4	
3	Art & Design	June 8, 2018	Varies with device	
4	Art & Design;Creativity	June 20, 2018	1.1	

	Android Ver
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

```
[7]: df.tail()
```

```
[7]:
```

	App	Category	\
10836	Sya9a Maroc - FR	FAMILY	
10837	Fr. Mike Schmitz Audio Teachings	FAMILY	
10838	Parkinson Exercices FR	MEDICAL	
10839	The SCP Foundation DB fr nn5n	BOOKS_AND_REFERENCE	
10840	iHoroscope - 2018 Daily Horoscope & Astrology	LIFESTYLE	

	Rating	Reviews	Size	Installs	Type	Price	\
10836	4.5	38	53M	5,000+	Free	0	
10837	5.0	4	3.6M	100+	Free	0	
10838	NaN	3	9.5M	1,000+	Free	0	
10839	4.5	114	Varies with device	1,000+	Free	0	
10840	4.5	398307	19M	10,000,000+	Free	0	

	Content Rating	Genres	Last Updated	Current Ver	\
10836	Everyone	Education	July 25, 2017	1.48	
10837	Everyone	Education	July 6, 2018	1.0	
10838	Everyone	Medical	January 20, 2017	1.0	
10839	Mature 17+	Books & Reference	January 19, 2015	Varies with device	
10840	Everyone	Lifestyle	July 25, 2018	Varies with device	

	Android Ver
10836	4.1 and up
10837	4.1 and up
10838	2.2 and up
10839	Varies with device
10840	Varies with device

```
[11]: df.sample(3)
```

```
[11]:
```

	App	Category	Rating	Reviews	Size	Installs	Type	\
8195	Knife Hit	GAME	4.5	462614	60M	10,000,000+	Free	
9908	EU VAT Checker	BUSINESS	3.1	7	39k	1,000+	Free	
6025	BD CRICKET LIVE	SPORTS	4.6	1238	9.5M	100,000+	Free	

	Price	Content Rating	Genres	Last Updated	Current Ver	\
8195	0	Everyone	Arcade	May 22, 2018	1.7.2	
9908	0	Everyone	Business	October 12, 2011	1	
6025	0	Everyone	Sports	January 15, 2018	3.0	

	Android Ver
8195	4.0.3 and up
9908	1.6 and up
6025	4.0.3 and up

```
[35]: df.shape
```

```
[35]: (10841, 13)
```

```
[29]: df.dtypes
```

```
[29]: App                object
Category            object
Rating              float64
Reviews             object
Size                object
Installs            object
Type                object
Price               object
Content Rating      object
Genres              object
Last Updated        object
Current Ver         object
Android Ver         object
dtype: object
```

```
[13]: df.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

```

```
[39]: df.duplicated().sum()
```

```
[39]: 483
```

```
[45]: df.drop_duplicates(inplace=True)
```

```
[47]: df
```

```
[47]:
```

	App	Category \
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN
1	Coloring book moana	ART_AND_DESIGN
2	U Launcher Lite - FREE Live Cool Themes, Hide ...	ART_AND_DESIGN
3	Sketch - Draw & Paint	ART_AND_DESIGN
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN
...	...	...
10836	Sya9a Maroc - FR	FAMILY
10837	Fr. Mike Schmitz Audio Teachings	FAMILY
10838	Parkinson Exercices FR	MEDICAL
10839	The SCP Foundation DB fr nn5n	BOOKS_AND_REFERENCE
10840	iHoroscope - 2018 Daily Horoscope & Astrology	LIFESTYLE

	Rating	Reviews	Size	Installs	Type	Price \
0	4.1	159	19M	10,000+	Free	0
1	3.9	967	14M	500,000+	Free	0
2	4.7	87510	8.7M	5,000,000+	Free	0
3	4.5	215644	25M	50,000,000+	Free	0
4	4.3	967	2.8M	100,000+	Free	0
...	...	...	...	...	...	...
10836	4.5	38	53M	5,000+	Free	0
10837	5.0	4	3.6M	100+	Free	0
10838	NaN	3	9.5M	1,000+	Free	0
10839	4.5	114	Varies with device	1,000+	Free	0
10840	4.5	398307	19M	10,000,000+	Free	0

	Content Rating	Genres	Last Updated \
0	Everyone	Art & Design	January 7, 2018
1	Everyone	Art & Design;Pretend Play	January 15, 2018
2	Everyone	Art & Design	August 1, 2018
3	Teen	Art & Design	June 8, 2018
4	Everyone	Art & Design;Creativity	June 20, 2018
...	...	...	...
10836	Everyone	Education	July 25, 2017
10837	Everyone	Education	July 6, 2018
10838	Everyone	Medical	January 20, 2017
10839	Mature 17+	Books & Reference	January 19, 2015
10840	Everyone	Lifestyle	July 25, 2018

	Current Ver	Android Ver
0	1.0.0	4.0.3 and up
1	2.0.0	4.0.3 and up
2	1.2.4	4.0.3 and up
3	Varies with device	4.2 and up
4	1.1	4.4 and up
...	...	...
10836	1.48	4.1 and up
10837	1.0	4.1 and up
10838	1.0	2.2 and up
10839	Varies with device	Varies with device
10840	Varies with device	Varies with device

[10358 rows x 13 columns]

```
[49]: df.describe()
```

```
[49]:
```

	Rating
count	8893.000000
mean	4.189542
std	0.545452
min	1.000000
25%	4.000000
50%	4.300000
75%	4.500000
max	19.000000

```
[101]: df_cp=df.copy()
```

```
[103]: df_cp
```

```
[103]:
```

	App	Category \
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN

1	Coloring book moana	ART_AND_DESIGN
2	U Launcher Lite - FREE Live Cool Themes, Hide ...	ART_AND_DESIGN
3	Sketch - Draw & Paint	ART_AND_DESIGN
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN
...	...	...
10836	Sya9a Maroc - FR	FAMILY
10837	Fr. Mike Schmitz Audio Teachings	FAMILY
10838	Parkinson Exercices FR	MEDICAL
10839	The SCP Foundation DB fr nn5n	BOOKS_AND_REFERENCE
10840	iHoroscope - 2018 Daily Horoscope & Astrology	LIFESTYLE

	Rating	Reviews	Size	Installs	Type	Price \
0	4.1	159	19M	10,000+	Free	0
1	3.9	967	14M	500,000+	Free	0
2	4.7	87510	8.7M	5,000,000+	Free	0
3	4.5	215644	25M	50,000,000+	Free	0
4	4.3	967	2.8M	100,000+	Free	0
...	...	...	...	...	...	...
10836	4.5	38	53M	5,000+	Free	0
10837	5.0	4	3.6M	100+	Free	0
10838	NaN	3	9.5M	1,000+	Free	0
10839	4.5	114	Varies with device	1,000+	Free	0
10840	4.5	398307	19M	10,000,000+	Free	0

	Content Rating	Genres	Last Updated \
0	Everyone	Art & Design	January 7, 2018
1	Everyone	Art & Design;Pretend Play	January 15, 2018
2	Everyone	Art & Design	August 1, 2018
3	Teen	Art & Design	June 8, 2018
4	Everyone	Art & Design;Creativity	June 20, 2018
...	...	...	...
10836	Everyone	Education	July 25, 2017
10837	Everyone	Education	July 6, 2018
10838	Everyone	Medical	January 20, 2017
10839	Mature 17+	Books & Reference	January 19, 2015
10840	Everyone	Lifestyle	July 25, 2018

	Current Ver	Android Ver
0	1.0.0	4.0.3 and up
1	2.0.0	4.0.3 and up
2	1.2.4	4.0.3 and up
3	Varies with device	4.2 and up
4	1.1	4.4 and up
...	...	...
10836	1.48	4.1 and up
10837	1.0	4.1 and up
10838	1.0	2.2 and up

```
10839 Varies with device Varies with device
10840 Varies with device Varies with device
```

```
[10358 rows x 13 columns]
```

```
[105]: df_cp.reset_index(drop=True,inplace=True)
```

```
[107]: df_cp
```

```
[107]:
```

	App	Category \
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN
1	Coloring book moana	ART_AND_DESIGN
2	U Launcher Lite - FREE Live Cool Themes, Hide ...	ART_AND_DESIGN
3	Sketch - Draw & Paint	ART_AND_DESIGN
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN
...	...	...
10353	Sya9a Maroc - FR	FAMILY
10354	Fr. Mike Schmitz Audio Teachings	FAMILY
10355	Parkinson Exercices FR	MEDICAL
10356	The SCP Foundation DB fr nn5n	BOOKS_AND_REFERENCE
10357	iHoroscope - 2018 Daily Horoscope & Astrology	LIFESTYLE

	Rating	Reviews	Size	Installs	Type	Price \
0	4.1	159	19M	10,000+	Free	0
1	3.9	967	14M	500,000+	Free	0
2	4.7	87510	8.7M	5,000,000+	Free	0
3	4.5	215644	25M	50,000,000+	Free	0
4	4.3	967	2.8M	100,000+	Free	0
...	...	...	...	...	...	...
10353	4.5	38	53M	5,000+	Free	0
10354	5.0	4	3.6M	100+	Free	0
10355	NaN	3	9.5M	1,000+	Free	0
10356	4.5	114	Varies with device	1,000+	Free	0
10357	4.5	398307	19M	10,000,000+	Free	0

	Content Rating	Genres	Last Updated \
0	Everyone	Art & Design	January 7, 2018
1	Everyone	Art & Design;Pretend Play	January 15, 2018
2	Everyone	Art & Design	August 1, 2018
3	Teen	Art & Design	June 8, 2018
4	Everyone	Art & Design;Creativity	June 20, 2018
...	...	...	...
10353	Everyone	Education	July 25, 2017
10354	Everyone	Education	July 6, 2018
10355	Everyone	Medical	January 20, 2017
10356	Mature 17+	Books & Reference	January 19, 2015
10357	Everyone	Lifestyle	July 25, 2018



	Current Ver	Android Ver
0	1.0.0	4.0.3 and up
1	2.0.0	4.0.3 and up
2	1.2.4	4.0.3 and up
3	Varies with device	4.2 and up
4	1.1	4.4 and up
...	...	...
10353	1.48	4.1 and up
10354	1.0	4.1 and up
10355	1.0	2.2 and up
10356	Varies with device	Varies with device
10357	Varies with device	Varies with device

[10358 rows x 13 columns]

```
[109]: df_cp.Reviews.dtype
```

```
[109]: dtype('O')
```

```
[115]: df_cp[~df_cp.Reviews.str.isnumeric()]
```

```
[115]:
```

	App	Category	Rating	Reviews	\
9990	Life Made WI-Fi Touchscreen Photo Frame	1.9	19.0	3.0M	

	Size	Installs	Type	Price	Content	Rating	Genres	\
9990	1,000+	Free	0	Everyone	NaN	February 11, 2018		

	Last Updated	Current Ver	Android Ver	Ver
9990	1.0.19	4.0 and up	NaN	

```
[129]: df_cp=df_cp.drop(df_cp.index[9990])
df_cp
```

```
[129]:
```

	App	Category	\
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	
1	Coloring book moana	ART_AND_DESIGN	
2	U Launcher Lite - FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	
3	Sketch - Draw & Paint	ART_AND_DESIGN	
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	
...	...	...	
10353	Sya9a Maroc - FR	FAMILY	
10354	Fr. Mike Schmitz Audio Teachings	FAMILY	
10355	Parkinson Exercices FR	MEDICAL	
10356	The SCP Foundation DB fr nn5n	BOOKS_AND_REFERENCE	
10357	iHoroscope - 2018 Daily Horoscope & Astrology	LIFESTYLE	

	Rating	Reviews		Size	Installs	Type	Price	\
0	4.1	159		19M	10,000+	Free	0	
1	3.9	967		14M	500,000+	Free	0	
2	4.7	87510		8.7M	5,000,000+	Free	0	
3	4.5	215644		25M	50,000,000+	Free	0	
4	4.3	967		2.8M	100,000+	Free	0	
...	...	...	...	...	...	...	...	...
10353	4.5	38		53M	5,000+	Free	0	
10354	5.0	4		3.6M	100+	Free	0	
10355	NaN	3		9.5M	1,000+	Free	0	
10356	4.5	114	Varies with device		1,000+	Free	0	
10357	4.5	398307		19M	10,000,000+	Free	0	

	Content	Rating		Genres	Last Updated	\
0	Everyone			Art & Design	January 7, 2018	
1	Everyone		Art & Design;Pretend Play		January 15, 2018	
2	Everyone			Art & Design	August 1, 2018	
3	Teen			Art & Design	June 8, 2018	
4	Everyone		Art & Design;Creativity		June 20, 2018	
...	...		...		...	
10353	Everyone			Education	July 25, 2017	
10354	Everyone			Education	July 6, 2018	
10355	Everyone			Medical	January 20, 2017	
10356	Mature 17+		Books & Reference		January 19, 2015	
10357	Everyone			Lifestyle	July 25, 2018	

	Current Ver	Android Ver
0	1.0.0	4.0.3 and up
1	2.0.0	4.0.3 and up
2	1.2.4	4.0.3 and up
3	Varies with device	4.2 and up
4	1.1	4.4 and up
...	...	...
10353	1.48	4.1 and up
10354	1.0	4.1 and up
10355	1.0	2.2 and up
10356	Varies with device	Varies with device
10357	Varies with device	Varies with device

[10356 rows x 13 columns]

```
[131]: df_cp[~df_cp.Reviews.str.isnumeric()]
```

[131]: Empty DataFrame

Columns: [App, Category, Rating, Reviews, Size, Installs, Type, Price, Content Rating, Genres, Last Updated, Current Ver, Android Ver]  
Index: []

```
[133]: df_cp["Reviews"]=df_cp.Reviews.astype(int)

[135]: df_cp.Reviews.dtype

[135]: dtype('int32')

[141]: df_cp.Size.dtype

[141]: dtype('O')

[143]: df_cp.Size.unique()

[143]: 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)

```

```
[145]: # This function is for showing the actual size of data in KB
```

```

def size_process(s):
    if str(s)[-1]=="M":
        res=float(str(s).replace("M",""))
        res=res*1024
        return res
    elif str(s)[-1]=="K":
        res=float(str(s).replace("K",""))
        return res
    else:
        return str(np.NaN)

```

```
[149]: df_cp.Size=df_cp.Size.apply(size_process)
```

```
[153]: df_cp.Size.dtype
```

```
[153]: dtype('O')
```

```
[155]: df_cp.Size=df_cp.Size.astype(float)
```

```
[157]: df_cp.Size.dtype
```

```
[157]: dtype('float64')
```

```
[159]: df_cp.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 10356 entries, 0 to 10357
Data columns (total 13 columns):
#   Column                Non-Null Count  Dtype
---  -
0   App                    10356 non-null  object
1   Category               10356 non-null  object
2   Rating                 8891 non-null   float64
3   Reviews                10356 non-null  int32
4   Size                   8515 non-null   float64
5   Installs               10356 non-null  object
6   Type                   10355 non-null  object
7   Price                  10356 non-null  object
8   Content Rating         10356 non-null  object
9   Genres                 10356 non-null  object
10  Last Updated           10356 non-null  object
11  Current Ver            10348 non-null  object
12  Android Ver            10354 non-null  object
dtypes: float64(2), int32(1), object(10)
memory usage: 1.1+ MB
```

```
[161]: df_cp.Installs
```

```
[161]: 0          10,000+
1          500,000+
2          5,000,000+
3          50,000,000+
4          100,000+
...
10353         5,000+
10354         100+
10355         1,000+
10356         1,000+
10357        10,000,000+
Name: Installs, Length: 10356, dtype: object
```

```
[163]: df_cp.Installs=df_cp.Installs.str.replace("+","").str.replace(",","")
```

```
[165]: df_cp.Installs
```

```
[165]: 0          10000
1         500000
2        5000000
3       50000000
4         100000
```

```

...
10353      5000
10354       100
10355      1000
10356      1000
10357    10000000
Name: Installs, Length: 10356, dtype: object

```

```
[167]: df_cp.Installs.astype(int)
```

```

[167]: 0      10000
      1    500000
      2   5000000
      3  50000000
      4   100000
...
10353      5000
10354       100
10355      1000
10356      1000
10357    10000000
Name: Installs, Length: 10356, dtype: int32

```

```
[171]: df_cp.Price.unique()
```

```

[171]: 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)

```

```
[173]: df_cp.Price=df_cp.Price.str.replace("$","")
```

```
[179]: df_cp.Price=df_cp.Price.astype(float)
```

```
[181]: df_cp.Price.dtype
```

```
[181]: dtype('float64')
```

```
[197]: df_cp["Last Updated"].dtype
```

```
[197]: dtype('O')
```

```
[201]: df_cp["Last Updated"]=pd.to_datetime(df_cp["Last Updated"])
```

```
[203]: df_cp["Last Updated"].dtype
```

```
[203]: dtype('<M8[ns]')
```

```
[209]: df_cp["Day"]=df_cp["Last Updated"].dt.day  
df_cp["Month"]=df_cp["Last Updated"].dt.month  
df_cp["Year"]=df_cp["Last Updated"].dt.year
```

```
[215]: df_cp.dtypes
```

```
[215]: App                object  
Category              object  
Rating                float64  
Reviews               int32  
Size                  float64  
Installs              object  
Type                  object  
Price                 float64  
Content Rating        object  
Genres                object  
Last Updated          datetime64[ns]  
Current Ver           object  
Android Ver           object  
Day                   int32  
Month                  int32  
Year                   int32  
dtype: object
```

```
[217]: df_cp.drop("Last Updated",axis=1,inplace=True)
```

```
[219]: df_cp
```

```
[219]:
```

	App	Category \
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN
1	Coloring book moana	ART_AND_DESIGN
2	U Launcher Lite - FREE Live Cool Themes, Hide ...	ART_AND_DESIGN
3	Sketch - Draw & Paint	ART_AND_DESIGN
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN
...	...	...
10353	Sya9a Maroc - FR	FAMILY
10354	Fr. Mike Schmitz Audio Teachings	FAMILY

10355	Parkinson Exercices FR	MEDICAL
10356	The SCP Foundation DB fr nn5n	BOOKS_AND_REFERENCE
10357	iHoroscope - 2018 Daily Horoscope & Astrology	LIFESTYLE

	Rating	Reviews	Size	Installs	Type	Price	Content Rating	\
0	4.1	159	19456.0	10000	Free	0.0	Everyone	
1	3.9	967	14336.0	500000	Free	0.0	Everyone	
2	4.7	87510	8908.8	5000000	Free	0.0	Everyone	
3	4.5	215644	25600.0	50000000	Free	0.0	Teen	
4	4.3	967	2867.2	100000	Free	0.0	Everyone	
...	...	...	...	...	...	...		
10353	4.5	38	54272.0	5000	Free	0.0	Everyone	
10354	5.0	4	3686.4	100	Free	0.0	Everyone	
10355	NaN	3	9728.0	1000	Free	0.0	Everyone	
10356	4.5	114	NaN	1000	Free	0.0	Mature 17+	
10357	4.5	398307	19456.0	10000000	Free	0.0	Everyone	

	Genres	Current Ver	Android Ver	Day	\
0	Art & Design	1.0.0	4.0.3 and up	7	
1	Art & Design;Pretend Play	2.0.0	4.0.3 and up	15	
2	Art & Design	1.2.4	4.0.3 and up	1	
3	Art & Design	Varies with device	4.2 and up	8	
4	Art & Design;Creativity	1.1	4.4 and up	20	
...	...	...	...	...	
10353	Education	1.48	4.1 and up	25	
10354	Education	1.0	4.1 and up	6	
10355	Medical	1.0	2.2 and up	20	
10356	Books & Reference	Varies with device	Varies with device	19	
10357	Lifestyle	Varies with device	Varies with device	25	

	Month	Year
0	1	2018
1	1	2018
2	8	2018
3	6	2018
4	6	2018
...	...	...
10353	7	2017
10354	7	2018
10355	1	2017
10356	1	2015
10357	7	2018

[10356 rows x 15 columns]

```
[221]: df_cp["Android Ver"].dtype
```



```
[221]: dtype('O')
```

```
[223]: df_cp["Android Ver"].unique()
```

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

```
[225]: df_cp["Android Ver"]=df_cp["Android Ver"].str.replace("and up","").str.  
        ↪replace("Varies with device","")
```

```
[233]: df_cp["App"].duplicated().sum()
```

```
[233]: 697
```

```
[235]: df_cp=df_cp.drop_duplicates(subset=["App"],keep='first')
```

```
[239]: df_cp["App"].duplicated().sum()
```

```
[239]: 0
```

```
[249]: df_cp.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
Index: 9659 entries, 0 to 10357  
Data columns (total 15 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   int32  
4   Size                  8118 non-null   float64  
5   Installs              9659 non-null   object  
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  Current Ver           9651 non-null   object  
11  Android Ver           9657 non-null   object  
12  Day                   9659 non-null   int32
```

```

13 Month          9659 non-null   int32
14 Year           9659 non-null   int32
dtypes: float64(3), int32(4), object(8)
memory usage: 1.0+ MB

```

```
[259]: categorical_features=[feature for feature in df_cp.columns if df_cp[feature].
    dtype == 'O']
```

```
[261]: categorical_features
```

```
[261]: ['App',
        'Category',
        'Installs',
        'Type',
        'Content Rating',
        'Genres',
        'Current Ver',
        'Android Ver']
```

```
[263]: Numerical_features=[feature for feature in df_cp.columns if df_cp[feature].
    dtype != 'O']
```

```
[265]: Numerical_features
```

```
[265]: ['Rating', 'Reviews', 'Size', 'Price', 'Day', 'Month', 'Year']
```

```
[267]: #EDA
        #First We count the categories of categorical Data

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

```

App: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: proportion, Length: 9659, dtype: float64
Category: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: proportion, dtype: float64

Installs:Installs

1000000	14.670256
100000	11.512579
10000	10.673983
100000000	9.700797
1000	9.193498
100	7.350657
50000000	6.284294
500000	5.228285
50000	4.855575
5000	4.845222
10	3.985920
500	3.395797
50	2.112020
500000000	2.091314
1000000000	1.946371

5	0.848949
1	0.693654
5000000000	0.248473
10000000000	0.207061
0	0.155296

Name: proportion, dtype: float64

Type:Type

Free	92.172292
Paid	7.827708

Name: proportion, dtype: float64

Content Rating: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: proportion, dtype: float64

Genres: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: proportion, Length: 118, dtype: float64

Current Ver: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: proportion, Length: 2817, dtype: float64

Android Ver:Android Ver

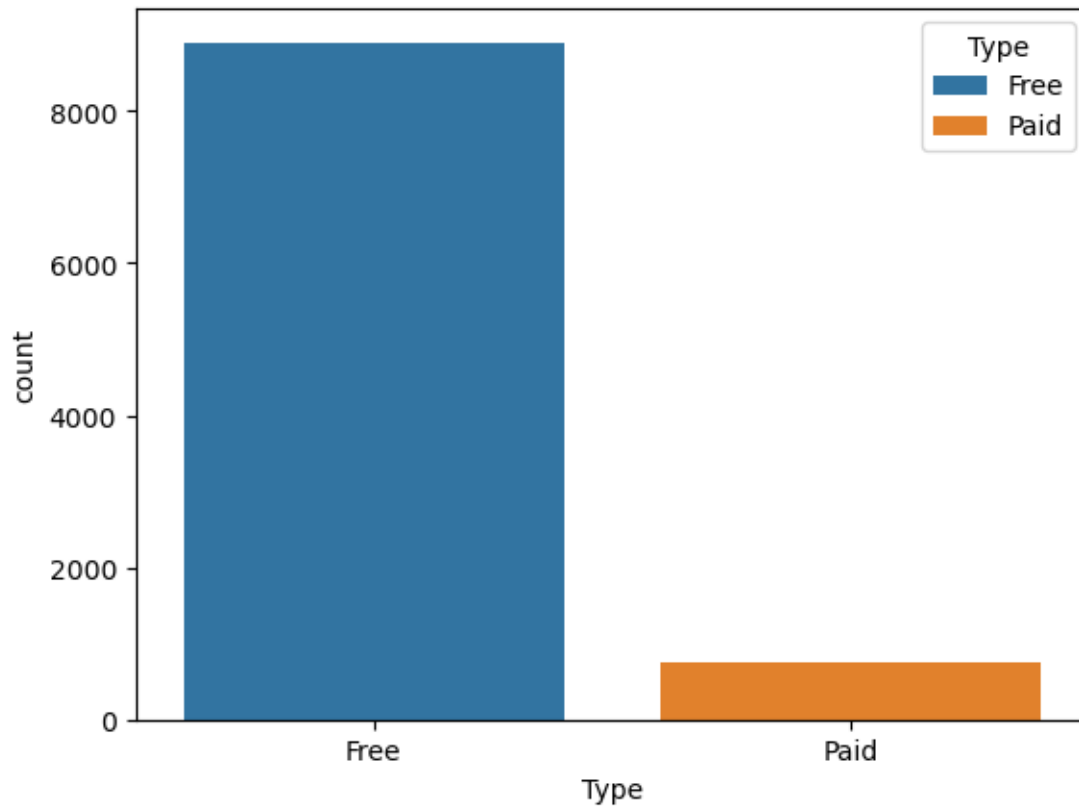
4.1	22.802112
4.0.3	14.445480
4.0	13.306410

	10.251631
4.4	8.470540
2.3	6.378793
5.0	5.301854
4.2	3.852128
2.3.3	2.826965
2.2	2.474889
3.0	2.392047
4.3	2.288495
2.1	1.377239
1.6	1.201201
6.0	0.559180
7.0	0.434918
3.2	0.372787
2.0	0.331366
5.1	0.227814
1.5	0.207104
4.4W	0.113907
3.1	0.103552
2.0.1	0.072486
8.0	0.062131
7.1	0.031066
4.0.3 - 7.1.1	0.020710
5.0 - 8.0	0.020710
1.0	0.020710
7.0 - 7.1.1	0.010355
4.1 - 7.1.1	0.010355
5.0 - 6.0	0.010355
2.2 - 7.1.1	0.010355
5.0 - 7.1.1	0.010355

Name: proportion, dtype: float64

```
[277]: sns.countplot(x=df_cp["Type"],hue=df_cp.Type)
```

```
[277]: <Axes: xlabel='Type', ylabel='count'>
```

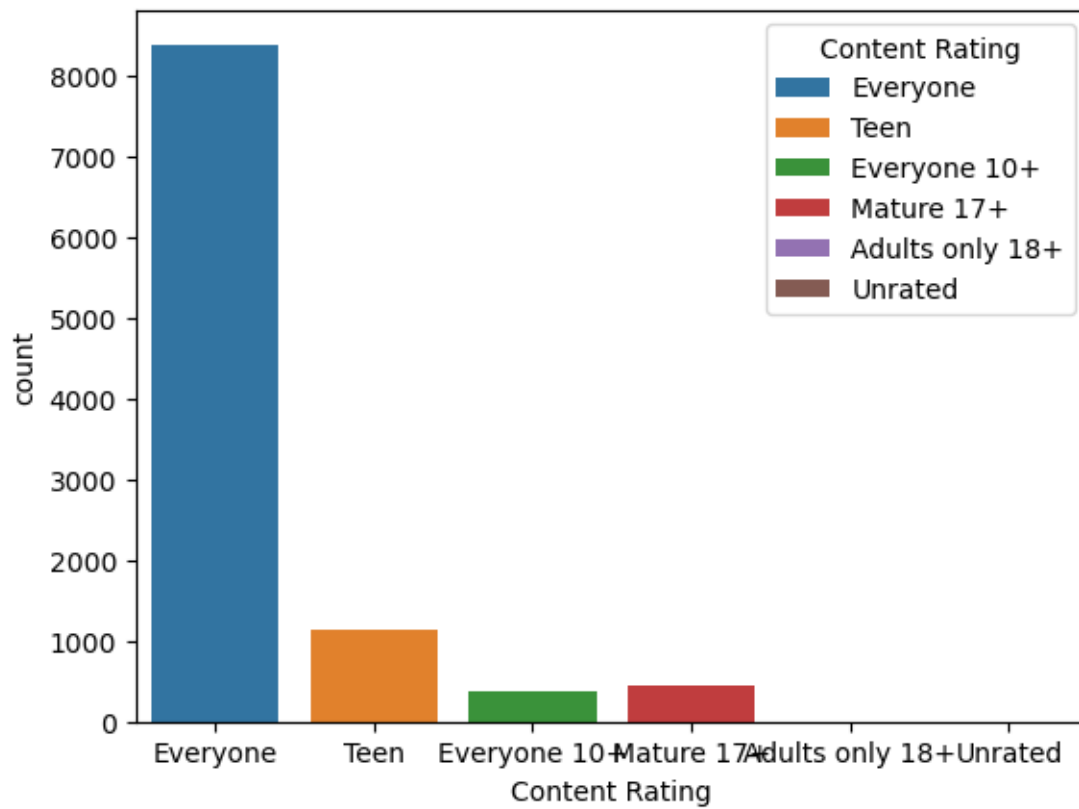


```
[281]: df_cp["Content Rating"].value_counts(normalize=True)*100
```

```
[281]: 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: proportion, dtype: float64
```

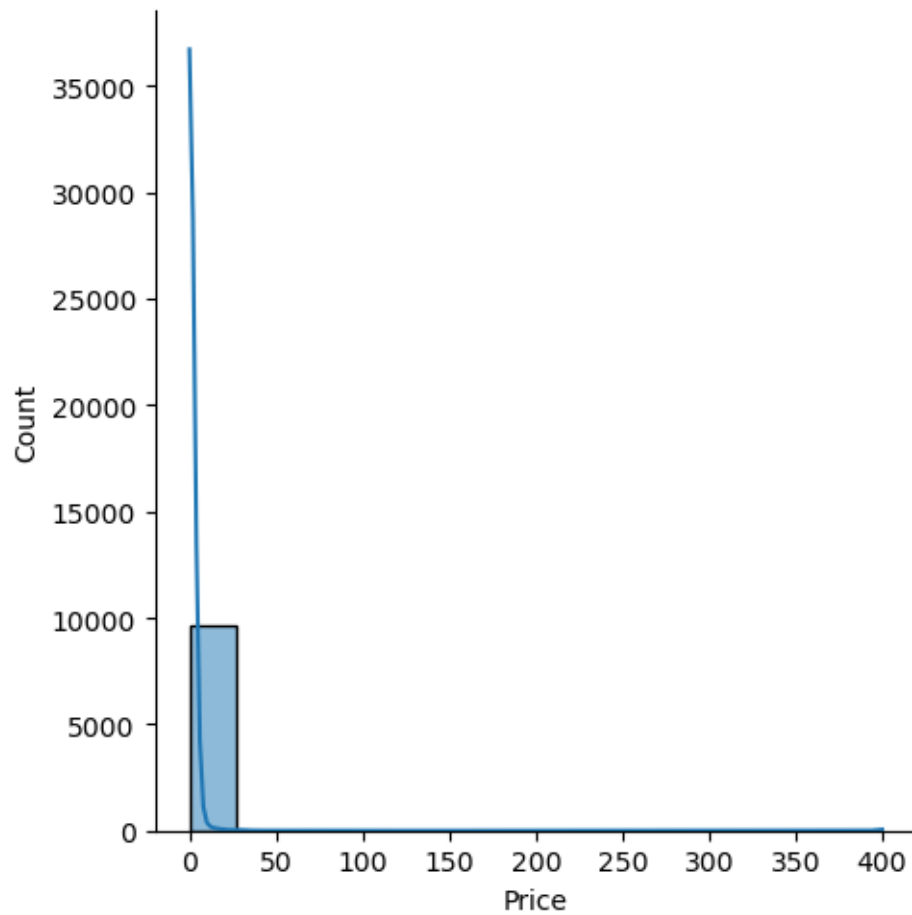
```
[295]: sns.countplot(x=df["Content Rating"],hue=df["Content Rating"])
```

```
[295]: <Axes: xlabel='Content Rating', ylabel='count'>
```



```
[299]: sns.displot(df_cp["Price"],kde=True)
```

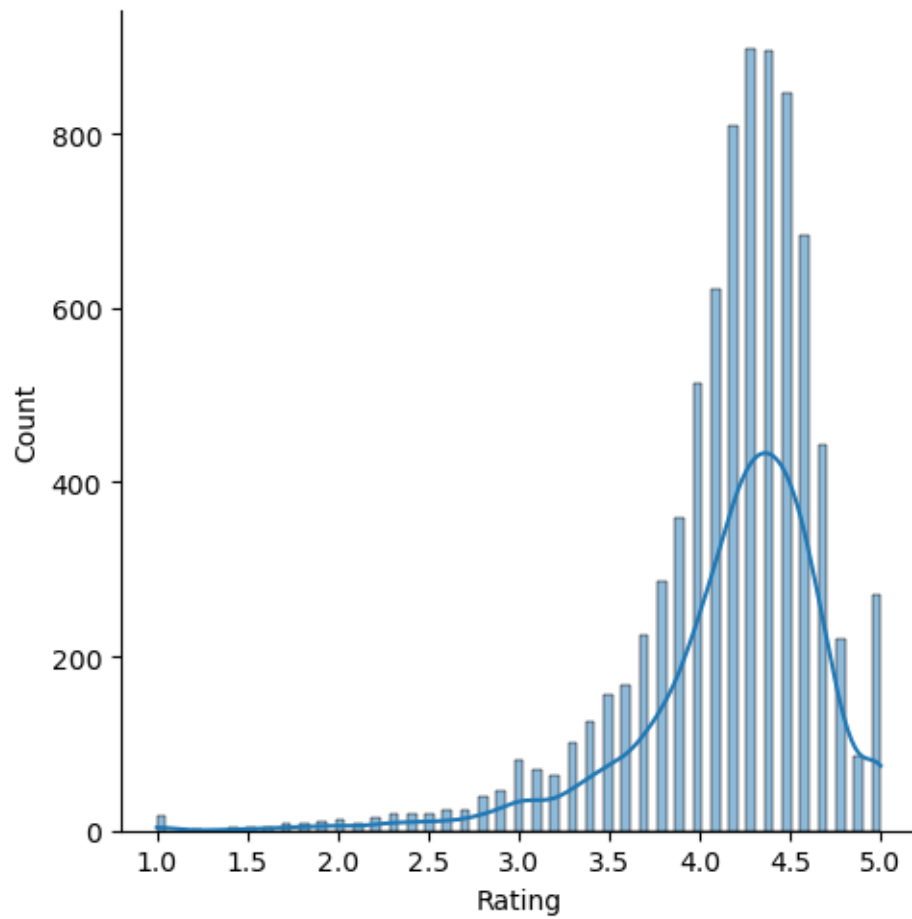
```
[299]: <seaborn.axisgrid.FacetGrid at 0x28f849c5cd0>
```



```
[303]: sns.displot(df_cp["Rating"],kde=True)
```

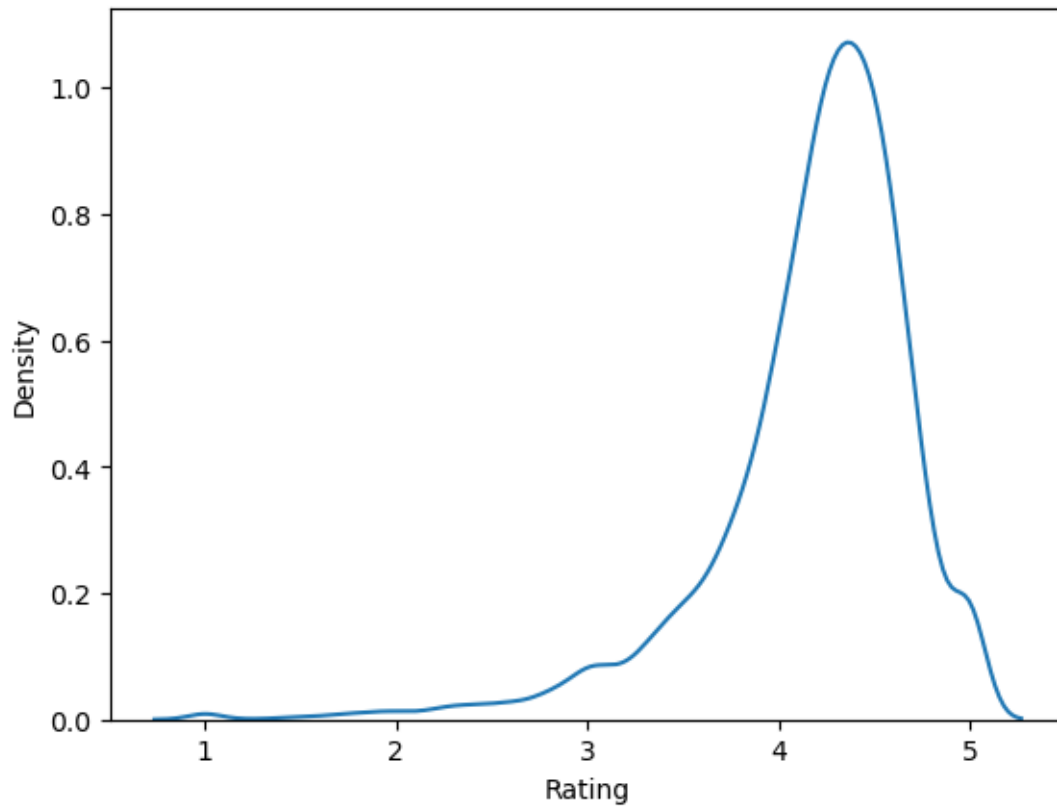
```
[303]: <seaborn.axisgrid.FacetGrid at 0x28f84ba52b0>
```



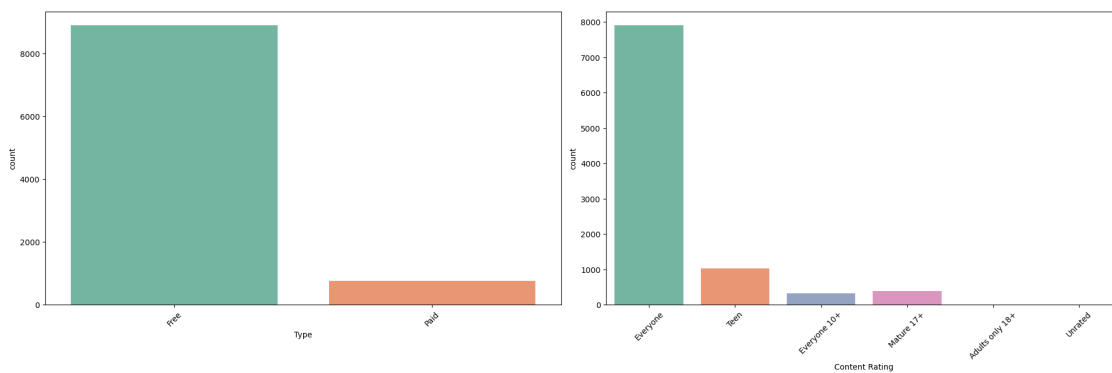


```
[305]: sns.kdeplot(df_cp["Rating"])
```

```
[305]: <Axes: xlabel='Rating', ylabel='Density'>
```

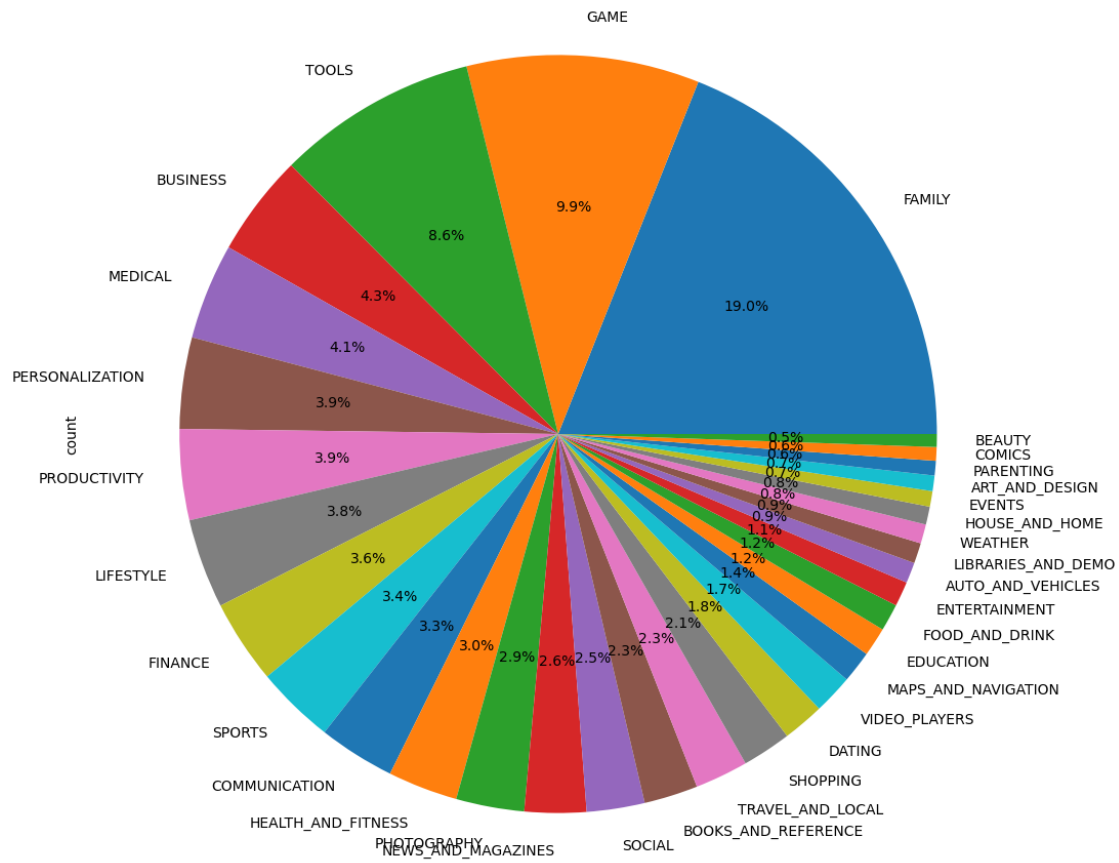


```
[311]: plt.figure(figsize=(20,12))
category=["Type","Content Rating"]
for i in range(0,len(category)):
    plt.subplot(2,2,i+1)
    sns.countplot(x=df_cp[category[i]],palette="Set2")
    plt.xlabel(category[i])
    plt.xticks(rotation=45)
    plt.tight_layout()
```



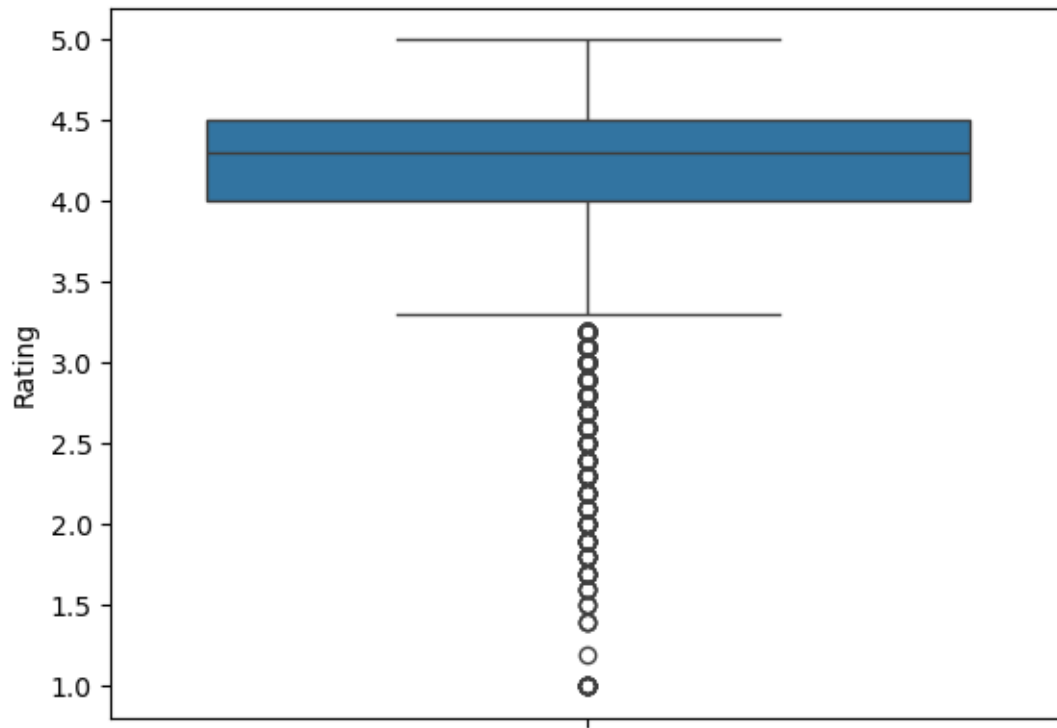
```
[329]: # Most Popular Category in the App
df_cp["Category"].value_counts().plot.
       pie(y=df_cp["Category"],figsize=(12,12),autopct="%1.1f%%")
```

```
[329]: <Axes: ylabel='count'>
```



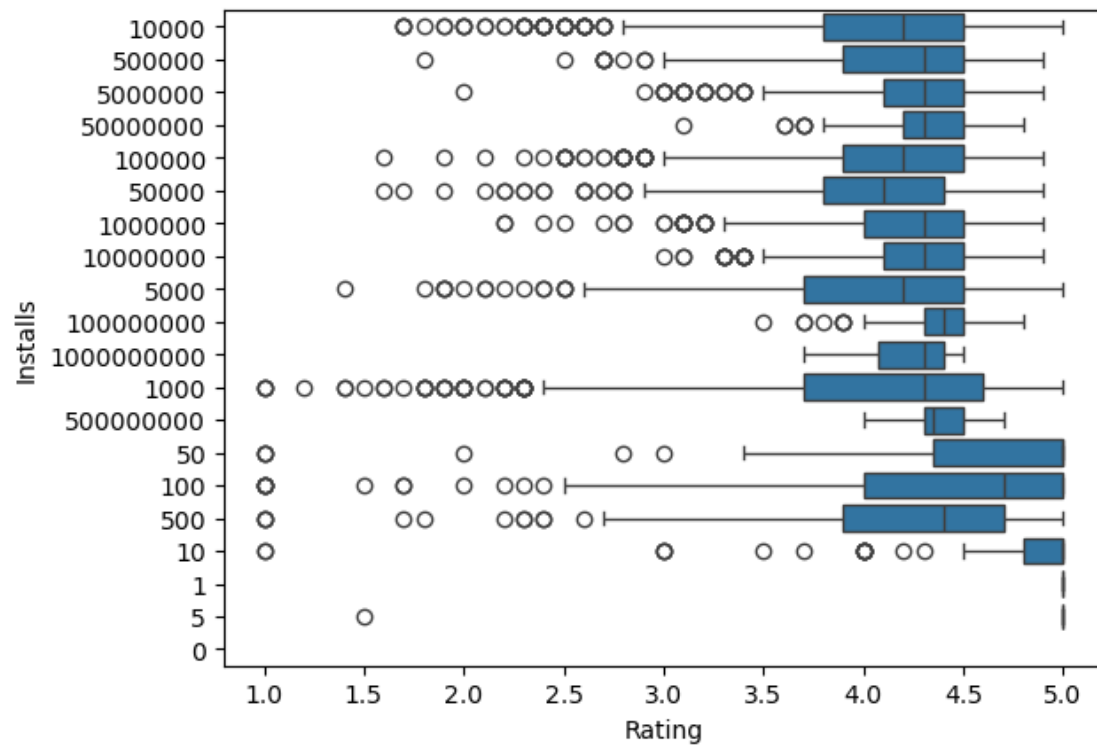
```
[362]: sns.boxplot(df_cp["Rating"])
```

```
[362]: <Axes: ylabel='Rating'>
```



```
[368]: sns.boxplot(x=df_cp["Rating"],y=df_cp["Installs"],data=df_cp)
```

```
[368]: <Axes: xlabel='Rating', ylabel='Installs'>
```



[ ]: