1 Data Analysis and Visulisation - Project

1.0.1 Submitted By -

Bhavyaa Garg

College Roll No. : 8166

University Roll No. : 19025570012

B.Sc. (H) Computer Science, Sem V

Hansraj College,DU

1.1 (a) Dataset Description

The dataset on which this project is based is "Googleplaystore.csv", which consists of various details like rating, installs, reviews etc. for around 10K android apps.This dataset is formed by the creator in order to help analyse the android app market. The dataset was last updated in August 2018

1.1.1 Download link

https://www.kaggle.com/lava18/google-play-store-apps
(https://www.kaggle.com/lava18/google-play-store-apps)

In [1]:
```
import numpy as np
import pandas as pd
from pandas import Series,DataFrame
import seaborn as sns
import matplotlib.pyplot as plt
from dateutil.parser import parse
from matplotlib import rcParams
```

In [2]:
```
apps_df=pd.read_csv("googleplaystore.csv")
apps_df
```

Out[2]:

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

10841 rows × 13 columns

In [3]:
```
apps_df.shape
```

Out[3]:
```
(10841, 13)
```

In [4]:
```
apps_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
```

In [5]:
```
apps_df.describe(include=object)
```

Out[5]:

	App	Category	Reviews	Size	Installs	Type	Price	Content Rating	Genres
count	10841	10841	10841	10841	10841	10840	10841	10840	10841
unique	9660	34	6002	462	22	3	93	6	120
top	ROBLOX	FAMILY	0	Varies with device	1,000,000+	Free	0	Everyone	Tools
freq	9	1972	596	1695	1579	10039	10040	8714	842

In [6]:
```
apps_df.describe()
```

Out[6]:

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

1.2 (b) Data Cleaning

1.3 (i) Detecting Outlier

In [7]:
```
apps_df[apps_df['Rating']>5.0]
```

Out[7]:

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

1.3.0.1 Detected an outlier in the dataset where 'Rating' is greater than 5.0. So, dropping that row

In [8]:
```
apps_df.drop([10472],inplace=True)
```

1.4 (ii) Changing datatype of columns

1.4.1 1. Changing datatype of 'Reviews' from object to int

In [9]:
```
apps_df['Reviews']=apps_df['Reviews'].astype(int)
```

1.4.2 2. Removing unnecessary symbols and changing datatype of 'Installs' from object to int

In [10]:
```
apps_df['Installs']=apps_df['Installs'].replace({'\,' : '', '\+' : ''},reg
```

1.4.3 3. Changing the column "Last updated" to datetime object

In [11]:
```
apps_df['Last Updated']=pd.to_datetime(apps_df['Last Updated'])
```

1.4.3.1 Dataset info after changing datatypes :

In [12]:
```
apps_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 10840 entries, 0 to 10840
Data columns (total 13 columns):
 #   Column          Non-Null Count  Dtype         
---  ------          --------------  -----         
 0   App             10840 non-null  object        
 1   Category        10840 non-null  object        
 2   Rating          9366 non-null   float64       
 3   Reviews         10840 non-null  int32         
 4   Size            10840 non-null  object        
 5   Installs        10840 non-null  int32         
 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  datetime64[ns]
 11  Current Ver     10832 non-null  object        
 12  Android Ver     10838 non-null  object        
dtypes: datetime64[ns](1), float64(1), int32(2), object(9)
memory usage: 1.1+ MB
```

1.5 (iii) Manipulating values and dropping unnecessary columns

1.5.1 Changing all 0 values in 'Price' to value 'Free' and dropping the 'Type' column which is not required

In [13]:
```
apps_df['Price'].replace(to_replace='0',value='Free',inplace=True)
```

In [14]:
```
apps_df.drop(['Type'],axis=1,inplace=True)
```

In [15]:
```
apps_df.head()
```

Out[15]:

	App	Category	Rating	Reviews	Size	Installs	Price	Content Rating	
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10000	Free	Everyone	Ar
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500000	Free	Everyone	Desig
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8.7M	5000000	Free	Everyone	Ar
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50000000	Free	Teen	Ar
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100000	Free	Everyone	Design

1.5.1.1 So, values in 'Price' column have successfully changed and column 'Type' has been dropped

1.6 (iv) Removing Duplicate values

1.6.0.1 Checking for duplicate values according to column 'App'

In [16]:
```
apps_df[apps_df.duplicated(subset='App')].sort_values(by="App")
```

Out[16]:

	App	Category	Rating	Reviews	Size	Installs	Price	Conte Rati
1407	10 Best Foods for You	HEALTH_AND_FITNESS	4.0	2490	3.8M	500000	Free	Everyc 1
2543	1800 Contacts - Lens Store	MEDICAL	4.7	23160	26M	1000000	Free	Everyc
2385	2017 EMRA Antibiotic Guide	MEDICAL	4.4	12	3.8M	1000	$16.99	Everyc
1434	21-Day Meditation Experience	HEALTH_AND_FITNESS	4.4	11506	15M	100000	Free	Everyc
5415	365Scores - Live Scores	SPORTS	4.6	666246	25M	10000000	Free	Everyc
...	...	...	...	...	...	...	...	
3055	theScore: Live Sports Scores, News, Stats & Vi...	SPORTS	4.4	133833	34M	10000000	Free	Everyc 1
3014	theScore: Live Sports Scores, News, Stats & Vi...	SPORTS	4.4	133825	34M	10000000	Free	Everyc 1
3202	trivago: Hotels & Travel	TRAVEL_AND_LOCAL	4.2	219848	Varies with device	50000000	Free	Everyc
3118	trivago: Hotels & Travel	TRAVEL_AND_LOCAL	4.2	219848	Varies with device	50000000	Free	Everyc
8291	wetter.com - Weather and Radar	WEATHER	4.2	189310	38M	10000000	Free	Everyc

1181 rows × 12 columns

1.6.0.2 Found 1181 duplicate rows. So, dropping the duplicate rows and keeping only the first row with unique 'App' value

In [17]:
```
apps_df.drop_duplicates(['App'],inplace=True)
```

In [18]:
```
apps_df[apps_df.duplicated(subset='App')].sort_values(by="App")
```

Out[18]:

App	Category	Rating	Reviews	Size	Installs	Price	Content Rating	Genres	Last Updated	Current Ver

In [19]:
```
apps_df.shape
```

Out[19]:
```
(9659, 12)
```

1.6.0.3 The number of rows reduced to 9659 after removing 1181 duplicate rows

1.7 (v) Handling missing data

In [20]:
```
apps_df.isnull().sum()
```

Out[20]:
```
App                0
Category           0
Rating          1463
Reviews            0
Size               0
Installs           0
Price              0
Content Rating     0
Genres             0
Last Updated       0
Current Ver        8
Android Ver        2
dtype: int64
```

1.7.1 1. Filling the missing 'Rating' values with the median of the column.

In [21]:
```
medianr=apps_df['Rating'].median()
medianr
```

Out[21]:
```
4.3
```

In [22]:
```
apps_df['Rating']=apps_df['Rating'].fillna(medianr)
```

In [23]:
```
apps_df.isnull().sum()
```

Out[23]:
```
App             0
Category        0
Rating          0
Reviews         0
Size            0
Installs        0
Price           0
Content Rating  0
Genres          0
Last Updated    0
Current Ver     8
Android Ver     2
dtype: int64
```

1.7.2 2. Filling the missing values in 'Current Ver' and 'Android Ver' with the mode of the columns

In [24]:
```
modecv=apps_df['Current Ver'].mode().values[0]
modecv
```

Out[24]:
```
'Varies with device'
```

In [25]:
```
modeav=apps_df['Android Ver'].mode().values[0]
modeav
```

Out[25]:
```
'4.1 and up'
```

In [26]:
```
apps_df['Current Ver']=apps_df['Current Ver'].fillna(modecv)
apps_df['Android Ver']=apps_df['Android Ver'].fillna(modeav)
```

In [27]:
```
apps_df.isnull().sum()
```

Out[27]:
```
App               0
Category          0
Rating            0
Reviews           0
Size              0
Installs          0
Price             0
Content Rating    0
Genres            0
Last Updated      0
Current Ver       0
Android Ver       0
dtype: int64
```

1.7.2.1 All missing values have been handled

1.7.3 Final info and description of the dataset after data cleaning :

In [28]:
```
apps_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 9659 entries, 0 to 10840
Data columns (total 12 columns):
 #   Column          Non-Null Count  Dtype         
---  ------          --------------  -----         
 0   App             9659 non-null   object        
 1   Category        9659 non-null   object        
 2   Rating          9659 non-null   float64       
 3   Reviews         9659 non-null   int32         
 4   Size            9659 non-null   object        
 5   Installs        9659 non-null   int32         
 6   Price           9659 non-null   object        
 7   Content Rating  9659 non-null   object        
 8   Genres          9659 non-null   object        
 9   Last Updated    9659 non-null   datetime64[ns]
 10  Current Ver     9659 non-null   object        
 11  Android Ver     9659 non-null   object        
dtypes: datetime64[ns](1), float64(1), int32(2), object(8)
memory usage: 905.5+ KB
```

In [29]:
```
apps_df.describe()
```

Out[29]:

	Rating	Reviews	Installs
count	9659.000000	9.659000e+03	9.659000e+03
mean	4.192442	2.165926e+05	7.777507e+06
std	0.496397	1.831320e+06	5.375828e+07
min	1.000000	0.000000e+00	0.000000e+00
25%	4.000000	2.500000e+01	1.000000e+03
50%	4.300000	9.670000e+02	1.000000e+05
75%	4.500000	2.940100e+04	1.000000e+06
max	5.000000	7.815831e+07	1.000000e+09

In [30]:
```
apps_df.describe(include=object)
```

Out[30]:

	App	Category	Size	Price	Content Rating	Genres	Current Ver	Android Ver
count	9659	9659	9659	9659	9659	9659	9659	9659
unique	9659	33	461	92	6	118	2769	33
top	Color by Disney	FAMILY	Varies with device	Free	Everyone	Tools	Varies with device	4.1 and up
freq	1	1832	1227	8903	7903	826	1063	2204

1.7.4 Recording the new cleaned dataset in a new file :

In [31]:
```
apps_df.to_csv('Googleplaystore_cleaned.csv')
```

1.8 (c) EDA Queries

1.8.1 (i) Query 1 :

1.8.2 Recommend top apps according to the conditions : Rating >= 4.5 , Installs >= 100,000,000 , Reviews >= 2,500,000

In [33]:
```
apps_df.loc[(apps_df['Rating']>=4.5) & (apps_df['Installs']>=100000000) &
```

Out[33]:

	App	Category	Rating	Reviews	Size	Installs	Price	C
139	Wattpad 📖 Free Books	BOOKS_AND_REFERENCE	4.6	2914724	Varies with device	100000000	Free	
351	Opera Mini - fast web browser	COMMUNICATION	4.5	5149854	Varies with device	100000000	Free	Ev
378	UC Browser - Fast Download Private & Secure	COMMUNICATION	4.5	17712922	40M	500000000	Free	
449	Truecaller: Caller ID, SMS spam blocking & Dialer	COMMUNICATION	4.5	7820209	Varies with device	100000000	Free	Ev
699	Duolingo: Learn Languages Free	EDUCATION	4.7	6289924	Varies with device	100000000	Free	Ev
...	...	...	...	...	...	...	...	
4812	GO Launcher - 3D parallax Themes & HD Wallpapers	PERSONALIZATION	4.5	7464996	Varies with device	100000000	Free	Ev
5695	AVG AntiVirus 2018 for Android Security	TOOLS	4.5	6207063	Varies with device	100000000	Free	Ev
7536	Security Master - Antivirus, VPN, AppLock, Boo...	TOOLS	4.7	24900999	Varies with device	500000000	Free	Ev
7550	Battery Doctor- Battery Life Saver & Battery Co...	TOOLS	4.5	8190074	17M	100000000	Free	Ev
8896	DU Battery Saver - Battery Charger & Battery Life	TOOLS	4.5	13479633	14M	100000000	Free	Ev

67 rows × 12 columns

1.8.2.1 So, these 73 apps are the recommended popular apps according to their rating, installs and reviews.

1.8.3 (ii) Query 2 :

1.8.4 Find all the 'Google' apps and construct bar graph for the number of reviews. Also construct a scatter plot for their rating and installs to see if they are related

In [35]:
```
Google_apps=apps_df[apps_df['App'].str.contains('Google')]
Google_apps
```

Out[35]:

	App	Category	Rating	Reviews	Size	Installs	Price
152	Google Play Books	BOOKS_AND_REFERENCE	3.9	1433233	Varies with device	1000000000	Free
193	Google My Business	BUSINESS	4.4	70991	Varies with device	5000000	Free
198	Google Primer	BUSINESS	4.4	62272	18M	10000000	Free
238	Google Ads	BUSINESS	4.3	29313	20M	5000000	Free
249	Google Analytics	BUSINESS	4.5	78662	22M	1000000	Free
338	Google Chrome: Fast & Secure	COMMUNICATION	4.3	9642995	Varies with device	1000000000	Free
359	Google Voice	COMMUNICATION	4.2	171031	Varies with device	10000000	Free
371	Google Duo - High Quality Video Calls	COMMUNICATION	4.6	2083237	Varies with device	500000000	Free
401	Google Allo	COMMUNICATION	4.3	346982	Varies with device	10000000	Free
432	PHONE for Google Voice & GTalk	COMMUNICATION	4.3	72065	13M	1000000	Free
791	Google Classroom	EDUCATION	4.2	69493	Varies with device	10000000	Free
865	Google Play Games	ENTERTAINMENT	4.3	7165362	Varies with device	1000000000	Free
1083	Google Pay	FINANCE	4.2	347838	Varies with device	100000000	Free
1317	Google Fit - Fitness Tracking	HEALTH_AND_FITNESS	3.9	249855	Varies with device	10000000	Free
2554	Google+	SOCIAL	4.2	4831125	Varies with device	1000000000	Free
2808	Google Photos	PHOTOGRAPHY	4.5	10858556	Varies with device	1000000000	Free
2855	QuickPic - Photo Gallery with Google Drive Sup...	PHOTOGRAPHY	4.6	847159	4.2M	10000000	Free

	App	Category	Rating	Reviews	Size	Installs	Price
3114	Google Trips - Travel Planner	TRAVEL_AND_LOCAL	4.1	26871	Varies with device	5000000	Free
3121	Google Earth	TRAVEL_AND_LOCAL	4.3	2338655	Varies with device	100000000	Free
3127	Google Street View	TRAVEL_AND_LOCAL	4.2	2129689	Varies with device	1000000000	Free
3234	Google	TOOLS	4.4	8033493	Varies with device	1000000000	Free
3235	Google Translate	TOOLS	4.4	5745093	Varies with device	500000000	Free
3257	Files Go by Google: Free up space on your phone	TOOLS	4.6	315585	8.5M	10000000	Free
3265	Gboard - the Google Keyboard	TOOLS	4.2	1859115	Varies with device	500000000	Free
3266	Google Korean Input	TOOLS	3.5	74819	Varies with device	100000000	Free
3268	Google app for Android TV	TOOLS	3.0	66	Varies with device	10000000	Free
3275	Google Assistant Go	TOOLS	3.7	315	4.6M	500000	Free
3330	Google Handwriting Input	TOOLS	4.3	94427	Varies with device	10000000	Free
3454	Google Drive	PRODUCTIVITY	4.4	2731171	Varies with device	1000000000	Free
3458	Google PDF Viewer	PRODUCTIVITY	4.2	226456	Varies with device	10000000	Free
3462	Google Assistant	PRODUCTIVITY	4.2	58675	1.3M	10000000	Free
3467	Google Keep	PRODUCTIVITY	4.4	691474	Varies with device	100000000	Free
3476	Google Calendar	PRODUCTIVITY	4.2	858208	Varies with device	500000000	Free
3477	Google Docs	PRODUCTIVITY	4.3	815981	Varies with device	100000000	Free
3526	Google Sheets	PRODUCTIVITY	4.3	496399	Varies with device	100000000	Free

	App	Category	Rating	Reviews	Size	Installs	Price
3534	Google Slides	PRODUCTIVITY	4.2	244567	Varies with device	100000000	Free
3687	Google Play Movies & TV	VIDEO_PLAYERS	3.7	906384	Varies with device	1000000000	Free
3736	Google News	NEWS_AND_MAGAZINES	3.9	877635	13M	1000000000	Free
4151	Google Now Launcher	TOOLS	4.2	857215	7.9M	100000000	Free
4436	Google I/O 2018	BOOKS_AND_REFERENCE	4.3	22401	4.6M	500000	Free
4965	Google AdSense	PRODUCTIVITY	4.3	52677	2.9M	1000000	Free
4969	Local Services ads by Google	BUSINESS	4.1	7	12M	1000	Free
5073	PhotoScan by Google Photos	PHOTOGRAPHY	4.3	61990	Varies with device	10000000	Free
6995	One Today by Google	LIFESTYLE	4.5	2586	9.4M	100000	Free
7012	Project Fi by Google	TOOLS	4.6	7342	Varies with device	1000000	Free
8353	Fix Error Google Playstore	BOOKS_AND_REFERENCE	4.3	18	5.7M	1000	Free
8659	Cal - Google Calendar + Widget	PRODUCTIVITY	4.2	86172	13M	1000000	Free
9838	Google Arts & Culture	FAMILY	3.7	24137	Varies with device	5000000	Free

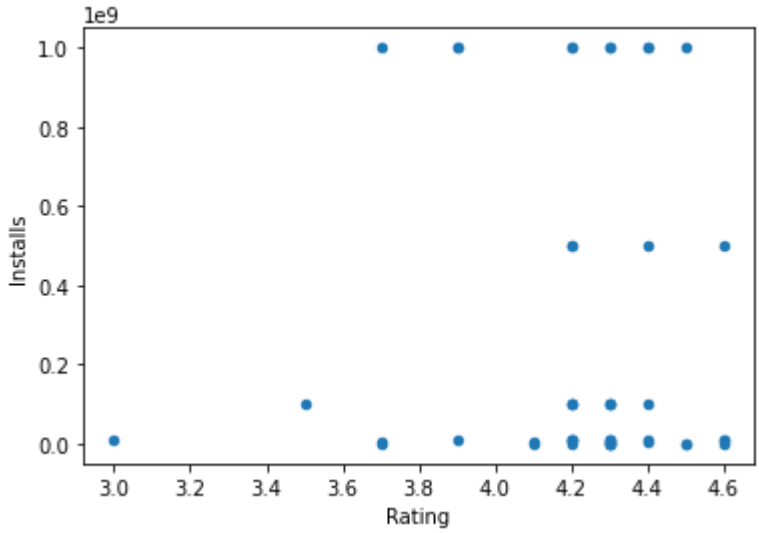
In [36]:
```
Google_apps.shape
```

Out[36]: (48, 12)

1.8.4.1 So, in total we found 48 apps from Google

In [37]:
```
Google_apps.plot.scatter(x='Rating',y='Installs')
```

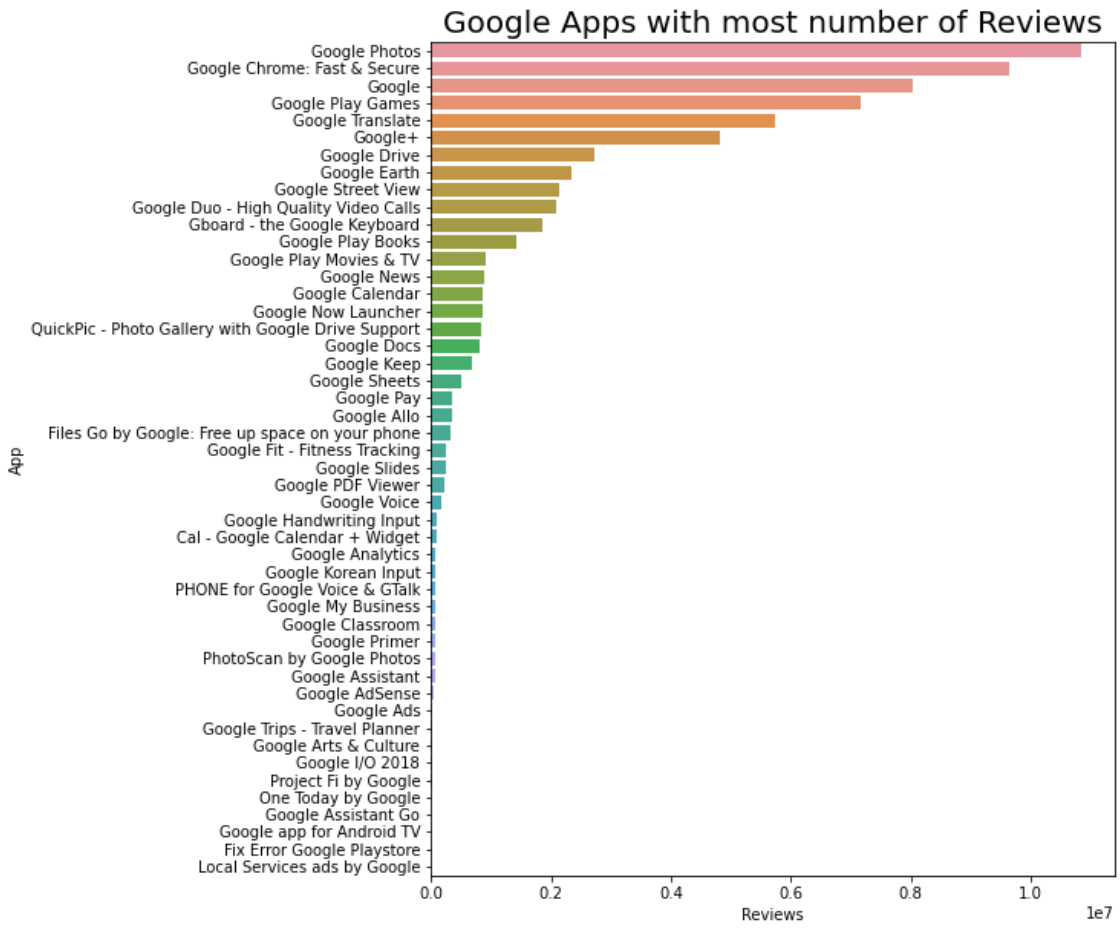
Out[37]: <AxesSubplot:xlabel='Rating', ylabel='Installs'>



1.8.4.2 As we can see from the scatter plot, there is relation between the rating and installs for Google apps. The data is quite scattered.

In [38]:
```
rcParams['figure.figsize']=8,10
sortga=Google_apps.sort_values(by=['Reviews'],ascending=0)
ax=sns.barplot(x='Reviews',y='App',data=sortga)
ax.set_xlabel('Reviews')
ax.set_ylabel('App')
ax.set_title("Google Apps with most number of Reviews",size=20)
```

Out[38]:
```
Text(0.5, 1.0, 'Google Apps with most number of Reviews')
```



1.8.4.3 From this graph we could see that 'Google Photos' has most number of reviews for Google apps, and can be considered as most popular Google app if we consider number of reviews as the condition

1.8.5 (iii) Query 3 :

1.8.6 For each 'Category', find the app with maximum 'Installs' and 'Rating'

In [40]:
```
sortbycir=apps_df.sort_values(by=['Category','Installs','Rating'],ascendin
topappbycat=sortbycir.drop_duplicates('Category').sort_values(by='Category
topappbycat
```

Out[40]:

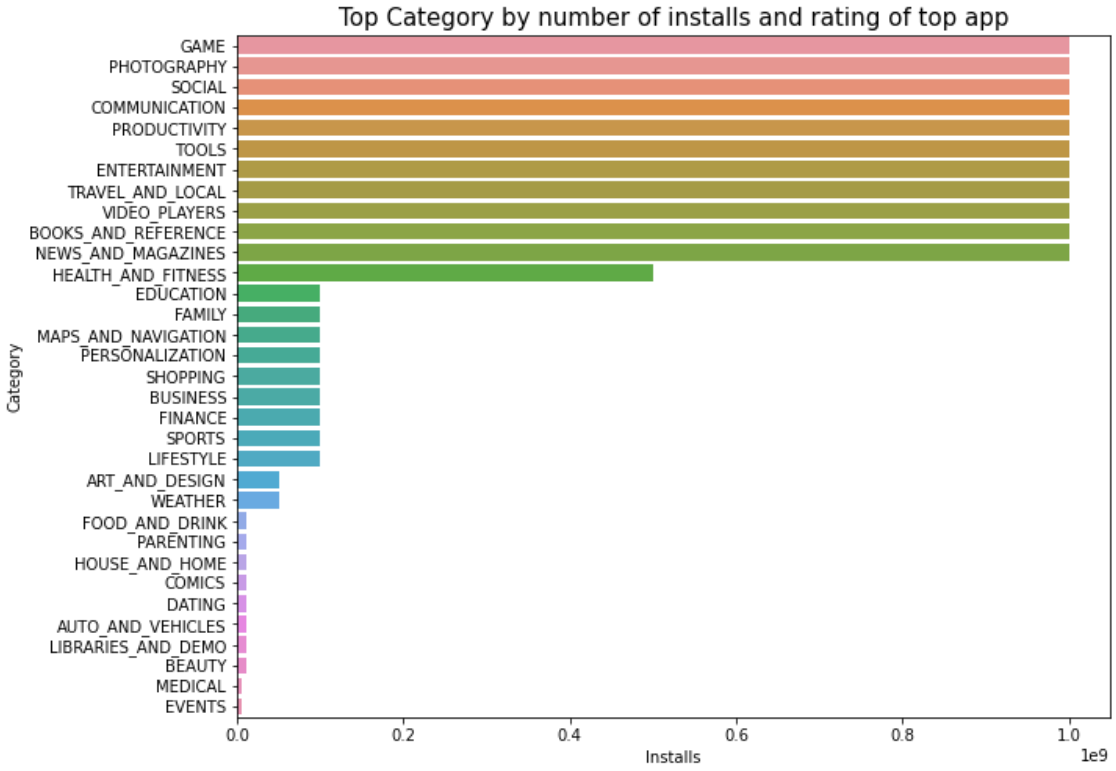
	App	Category	Rating	Reviews	Size	Installs	Price
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50000000	Free
8289	AutoScout24 - used car finder	AUTO_AND_VEHICLES	4.4	186648	42M	10000000	Free
117	Beauty Camera - Selfie Camera	BEAUTY	4.0	113715	Varies with device	10000000	Free
152	Google Play Books	BOOKS_AND_REFERENCE	3.9	1433233	Varies with device	1000000000	Free
194	OfficeSuite : Free Office + PDF Editor	BUSINESS	4.3	1002861	35M	100000000	Free
297	LINE WEBTOON - Free Comics	COMICS	4.5	1013635	Varies with device	10000000	Free
336	WhatsApp Messenger	COMMUNICATION	4.4	69119316	Varies with device	1000000000	Free
502	Find Real Love — YouLove Premium Dating	DATING	4.5	212626	11M	10000000	Free
699	Duolingo: Learn Languages Free	EDUCATION	4.7	6289924	Varies with device	100000000	Free
865	Google Play Games	ENTERTAINMENT	4.3	7165362	Varies with device	1000000000	Free
1005	Ticketmaster Event Tickets	EVENTS	4.0	40113	36M	5000000	Free
6269	Bitmoji – Your Personal Emoji	FAMILY	4.6	2312084	Varies with device	100000000	Free
1083	Google Pay	FINANCE	4.2	347838	Varies with device	100000000	Free
1183	Tastely	FOOD_AND_DRINK	4.7	611136	19M	10000000	Free
1654	Subway Surfers	GAME	4.5	27722264	76M	1000000000	Free
5596	Samsung Health	HEALTH_AND_FITNESS	4.3	480208	70M	500000000	Free
1456	Houzz Interior Design Ideas	HOUSE_AND_HOME	4.6	353799	Varies with device	10000000	Free

	App	Category	Rating	Reviews	Size	Installs	Price
10729	MX Player Codec (ARMv7)	LIBRARIES_AND_DEMO	4.3	332083	6.3M	10000000	Free
4587	Tinder	LIFESTYLE	4.0	2789775	68M	100000000	Free
3820	Waze - GPS, Maps, Traffic Alerts & Live Naviga...	MAPS_AND_NAVIGATION	4.6	7232629	Varies with device	100000000	Free
2319	My Calendar - Period Tracker	MEDICAL	4.7	156410	14M	5000000	Free
3736	Google News	NEWS_AND_MAGAZINES	3.9	877635	13M	1000000000	Free
7229	Pregnancy Tracker & Countdown to Baby Due Date	PARENTING	4.7	658087	62M	10000000	Free
3354	ZEDGE™ Ringtones & Wallpapers	PERSONALIZATION	4.6	6466641	Varies with device	100000000	Free
2808	Google Photos	PHOTOGRAPHY	4.5	10858556	Varies with device	1000000000	Free
3454	Google Drive	PRODUCTIVITY	4.4	2731171	Varies with device	1000000000	Free
2660	AliExpress - Smarter Shopping, Better Living	SHOPPING	4.6	5916606	Varies with device	100000000	Free
2545	Instagram	SOCIAL	4.5	66577313	Varies with device	1000000000	Free
8445	FIFA Soccer	SPORTS	4.2	3909032	51M	100000000	Free
3234	Google	TOOLS	4.4	8033493	Varies with device	1000000000	Free
3117	Maps - Navigate & Explore	TRAVEL_AND_LOCAL	4.3	9235155	Varies with device	1000000000	Free
3665	YouTube	VIDEO_PLAYERS	4.3	25655305	Varies with device	1000000000	Free
3649	GO Weather - Widget, Theme, Wallpaper, Efficient	WEATHER	4.5	1422858	Varies with device	50000000	Free

1.8.6.1 So, these are the top app for each category by most number of installs and

In [41]:
```
topappbycat.sort_values(by=['Installs','Rating'],ascending=False,inplace=T
rcParams['figure.figsize']=10,8
ax=sns.barplot(x='Installs',y='Category',data=topappbycat)
ax.set_xlabel('Installs')
ax.set_ylabel('Category')
ax.set_title("Top Category by number of installs and rating of top app",si
```

Out[41]:
```
Text(0.5, 1.0, 'Top Category by number of installs and rating of top ap
p')
```



1.8.6.2 From the graph we could see that 'Photography', 'Social' and 'Game' are the top three categories if seeing the rating and number of installs of the top app in that category

1.8.7 (iv) Query 4 :

1.8.8 Get the top 3 paid and free games, by number of installs and rating

In [42]:
```
paid_games=apps_df[(apps_df['Category']=='GAME') & (apps_df['Price']!='Fre
```

In [43]:
```
paid_games.sort_values(by=['Installs','Rating'],ascending=False)[:3]
```

Out[43]:

	App	Category	Rating	Reviews	Size	Installs	Price	Content Rating	Genres	L Updat
4034	Hitman Sniper	GAME	4.6	408292	29M	10000000	$0.99	Mature 17+	Action	20 07
5631	Five Nights at Freddy's	GAME	4.6	100805	50M	1000000	$2.99	Teen	Action	20 12
8804	DraStic DS Emulator	GAME	4.6	87766	12M	1000000	$4.99	Everyone	Action	20 07

1.8.8.1 So, top 3 paid games are "Hitman Sniper", "Five Nights and Freddy's" and "DraStic DS Emulator"

In [44]:
```
free_games=apps_df[(apps_df['Category']=='GAME') & (apps_df['Price']=='Fre
```

In [45]:
```
free_games.sort_values(by=['Installs','Rating'],ascending=False)[:3]
```

Out[45]:

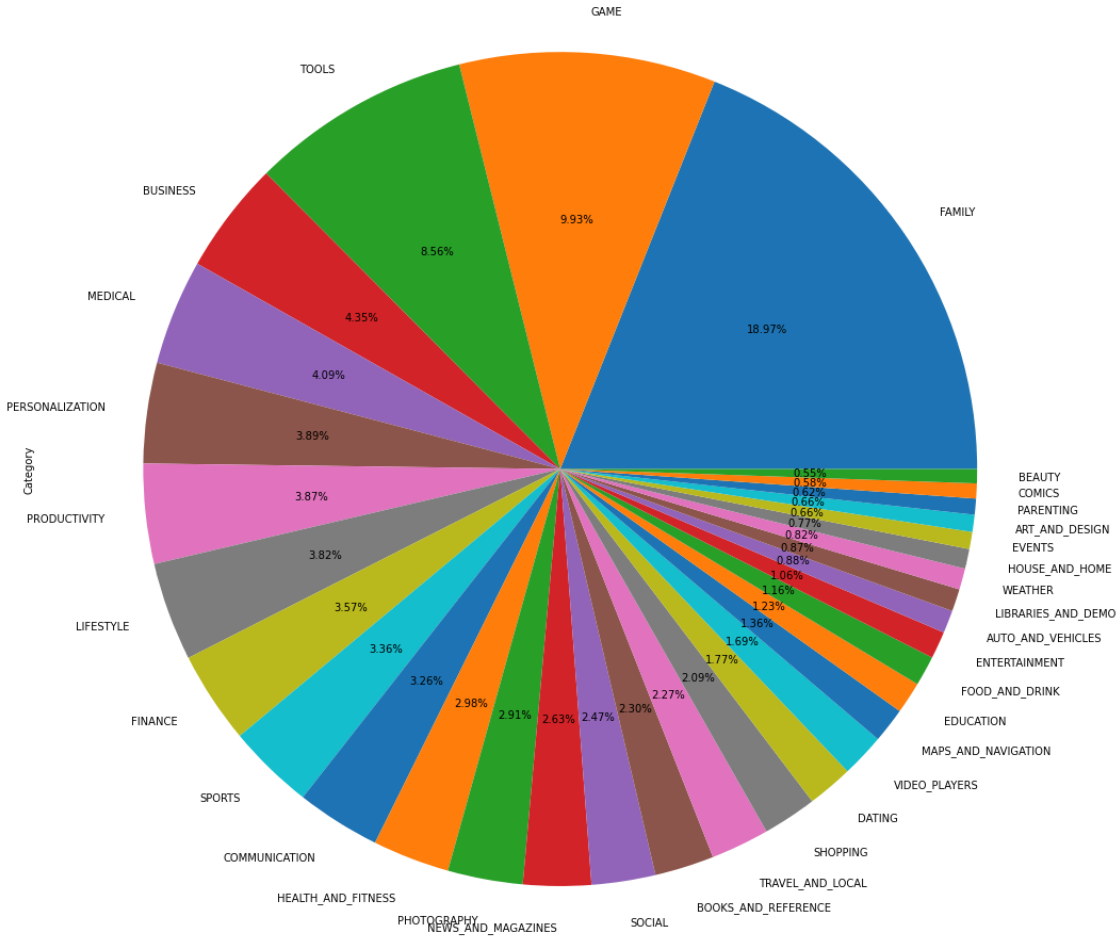
	App	Category	Rating	Reviews	Size	Installs	Price	Content Rating	Genres	U
1654	Subway Surfers	GAME	4.5	27722264	76M	1000000000	Free	Everyone 10+	Arcade	
1722	My Talking Tom	GAME	4.5	14891223	Varies with device	500000000	Free	Everyone	Casual	
1655	Candy Crush Saga	GAME	4.4	22426677	74M	500000000	Free	Everyone	Casual	

1.8.8.2 So, top 3 free games are "Subway Surfers", "My Talking Tom" and "Candy Crush Saga"

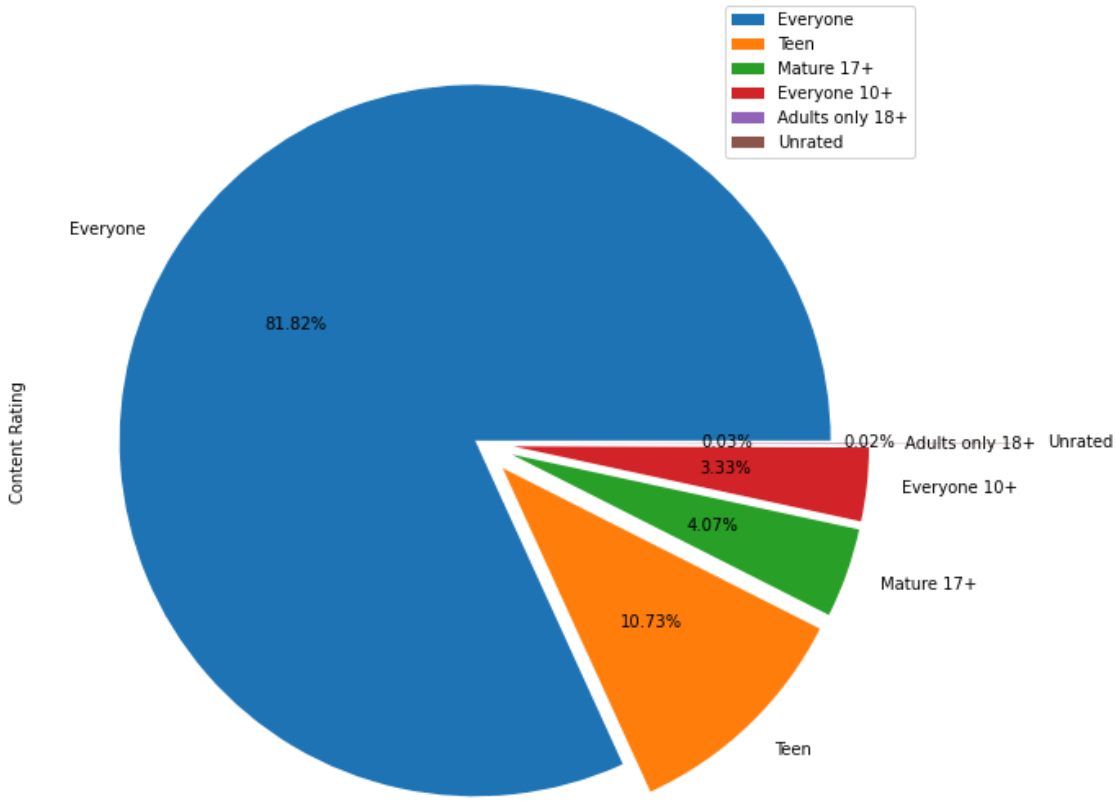
1.8.9 (v) Query 5 :

1.8.10 Make pie charts showing the percentage of apps for diiferent 'Category' and 'Content Rating' labels

In [46]:
```
plt.figure(figsize=(18,18))
apps_df['Category'].value_counts().plot(kind='pie',autopct="%.2f%%")
plt.show()
```



In [47]:
```
plt.figure(figsize=(10,10))
explode=[0.01,0.1,0.1,0.1,0.1,0.5]
apps_df['Content Rating'].value_counts().plot(kind='pie',autopct="%.2f%%",
plt.legend()
plt.show()
```



1.8.10.1 So, there are maximum apps in 'FAMILY' category and least in 'BEAUTY'

1.8.10.2 And there are maximum apps with 'Everyone' content rating and least 'Unrated' apps

1.8.11 (vi) Query 6 :

1.8.12 Find out all the apps last updated in or before 2014, and installs less than or equal to 100

In [48]:
```
obsolete=apps_df.loc[(apps_df['Last Updated']<='2014-12-31') & (apps_df['I
obsolete
```

Out[48]:

	App	Category	Rating	Reviews	Size	Installs	Price
4166	G-Force Driving Assistant	SPORTS	4.6	10	6.1M	100	$3.88
4973	Ad Removal: thereisonlywe	PRODUCTIVITY	3.4	16	293k	100	$6.49
5051	AF-STROKE	MEDICAL	4.3	5	902k	100	$5.00
5144	AH Alarm Panel	TOOLS	3.9	7	81k	100	$4.99
5475	500 AP World History Questions	FAMILY	4.7	7	1.2M	100	$9.99
5482	meStudying: AP English Lit	FAMILY	5.0	1	655k	10	$4.99

In [49]:
```
obsolete.shape
```

Out[49]: (51, 12)

1.8.12.1 So found these 51 apps last updated in or before 2014 and installs less than 100. These apps can be considered as obsolete apps, and the creator of the app can be asked to update the app to increase the functionality and thus number of installs of the app

1.8.13 (vii) Query 7 :

1.8.14 Get all apps with rating less than or equal to 3, and number of installs greater than or equal to 500,000

In [52]:
```
rating3=apps_df[(apps_df['Rating']<=3) & (apps_df['Installs']>=500000)]
rating3
```

Out[52]:

	App	Category	Rating	Reviews	Size	Installs	Price	C
520	EliteSingles – Dating for Single Professionals	DATING	2.5	5377	19M	500000	Free	
1576	Telstra	LIFESTYLE	3.0	4260	6.3M	5000000	Free	Ev
2317	Anthem Anywhere	MEDICAL	2.7	2657	24M	500000	Free	Ev
3139	VZ Navigator for Galaxy S4	TRAVEL_AND_LOCAL	3.0	2750	39M	5000000	Free	Ev
3268	Google app for Android TV	TOOLS	3.0	66	Varies with device	10000000	Free	Ev

In [53]:
```
rating3.shape
```

Out[53]: (26, 12)

1.8.14.1 These 26 apps with high number of installs but low rating should be looked into more by the creators. The number of installs shows a potential for the success of the app, but the low rating indicates some errors or lack of functionality in the app which can be improved on

1.8.15 (viii) Query 8 :

1.8.16 Find out all the paid apps with rating greater than or equal to 4.5, and installs greater than or equal to 1 million

In [54]:
```
paidapps=apps_df[(apps_df['Installs']>=1000000) & (apps_df['Price']!='Free
paidapps
```

Out[54]:

	App	Category	Rating	Reviews	Size	Installs	Price	Content Rating
2241	Minecraft	FAMILY	4.5	2376564	Varies with device	10000000	$6.99	Everyone 10+
4034	Hitman Sniper	GAME	4.6	408292	29M	10000000	$0.99	Mature 17+
4260	Cut the Rope GOLD	FAMILY	4.6	61264	43M	1000000	$0.99	Everyone
5578	Sleep as Android Unlock	LIFESTYLE	4.5	23966	872k	1000000	$5.99	Everyone
5631	Five Nights at Freddy's	GAME	4.6	100805	50M	1000000	$2.99	Teen
7355	Threema	COMMUNICATION	4.5	51110	Varies with	1000000	$2.99	Everyone

In [55]:
```
paidapps.shape
```

Out[55]: (11, 12)

1.8.16.1 These 11 apps have high number of installs and rating despite being paid. Maximum of them being family and game apps show that people are willing to pay for games if it is worth

1.8.17 (ix) Query 9 :

1.8.18 Get all the communication apps for everyone with number of installs greater than or equal to 100 million

In [56]:
```
commapps=apps_df[(apps_df['Category']=="COMMUNICATION") & (apps_df['Conten
commapps
```

Out[56]:

	App	Category	Rating	Reviews	Size	Installs	Price	Conte Ratin
335	Messenger – Text and Video Chat for Free	COMMUNICATION	4.0	56642847	Varies with device	1000000000	Free	Everyor
336	WhatsApp Messenger	COMMUNICATION	4.4	69119316	Varies with device	1000000000	Free	Everyor
338	Google Chrome: Fast & Secure	COMMUNICATION	4.3	9642995	Varies with device	1000000000	Free	Everyor
339	Messenger Lite: Free Calls & Messages	COMMUNICATION	4.4	1429035	Varies with device	100000000	Free	Everyor

In [57]:
```
commapps.shape
```

Out[57]: (22, 12)

1.8.18.1 Online Communication being a major need of us in today's time, these 22 apps seems to lead this market ahead, making our everyday life better.

1.8.19 (x) Query 10 :

1.8.19.1 Get the 10 most popular apps of all time, regardless of their category

In [58]:
```
apps_df.sort_values(by=['Installs','Rating'],ascending=False)[:10]
```

Out[58]:

	App	Category	Rating	Reviews	Size	Installs	Price	Conte Rati
1654	Subway Surfers	GAME	4.5	27722264	76M	1000000000	Free	Everyc 1
2545	Instagram	SOCIAL	4.5	66577313	Varies with device	1000000000	Free	Te
2808	Google Photos	PHOTOGRAPHY	4.5	10858556	Varies with device	1000000000	Free	Everyc
336	WhatsApp Messenger	COMMUNICATION	4.4	69119316	Varies with device	1000000000	Free	Everyc
3234	Google	TOOLS	4.4	8033493	Varies with device	1000000000	Free	Everyc
3454	Google Drive	PRODUCTIVITY	4.4	2731171	Varies with device	1000000000	Free	Everyc
338	Google Chrome: Fast & Secure	COMMUNICATION	4.3	9642995	Varies with device	1000000000	Free	Everyc
340	Gmail	COMMUNICATION	4.3	4604324	Varies with device	1000000000	Free	Everyc
865	Google Play Games	ENTERTAINMENT	4.3	7165362	Varies with device	1000000000	Free	Te
3117	Maps - Navigate & Explore	TRAVEL_AND_LOCAL	4.3	9235155	Varies with device	1000000000	Free	Everyc

1.8.19.2 So these 10 apps are the most popular apps of all time on Google play store by number of Installs and Rating

1.8.20 (xi) Query 11 :

1.8.21 Get the top 3 category of apps with most number of apps with rating greater than or equal to 4.5 and top 3 category of apps with most number of apps with installs greater than or equal to 100 million

In [59]:
```
ratinghigh=apps_df.loc[apps_df['Rating']>=4.5]
topapps=ratinghigh.groupby('Category').count()
topapps[['App']].sort_values(by=['App'],ascending=False)[:3]
```

Out[59]:

Category	App
FAMILY	500
GAME	281
TOOLS	169

1.8.21.1 So, most apps with high rating of greater than or equal to 4.5 belong to 'FAMILY', 'GAME' and 'TOOLS' category

In [60]:
```
ratinghigh=apps_df.loc[apps_df['Installs']>=100000000]
topapps=ratinghigh.groupby('Category').count()
topapps[['App']].sort_values(by=['App'],ascending=False)[:3]
```

Out[60]:

Category	App
GAME	62
TOOLS	28
COMMUNICATION	27

1.8.21.2 So, most apps with high number of installs of greater than or equal to 100 million belongs to 'GAME', 'TOOLS' and 'COMMUNICATION' categories

1.8.22 (xii) Query 12 :

1.8.23 Find all apps with creativity in it's genre, and group by their category

In [61]:
```
creativity=apps_df[apps_df['Genres'].str.contains('Creativity')]
creativity=creativity.set_index(['Category','App'])
creativity.sort_index(level=0)
```

Out[61]:

Category	App	Rating	Reviews	Size	Installs	Price	Content Rating	
ART_AND_DESIGN	Colorfit - Drawing & Coloring	4.7	20260	25M	500000	Free	Everyone	Art
	Kids Paint Free - Drawing Fun	4.7	121	3.1M	10000	Free	Everyone	Art
	Paint Splash!	3.8	2206	1.2M	100000	Free	Everyone	Art
	Pixel Draw - Number Art Coloring Book	4.3	967	2.8M	100000	Free	Everyone	Art
	UNICORN - Color By Number & Pixel Art Coloring	4.7	8145	24M	500000	Free	Everyone	Art
COMICS	Unicorn Pokez - Color By Number	4.8	516	12M	50000	Free	Everyone	
	Cars Coloring Pages	4.4	1090	49M	1000000	Free	Everyone	I
EDUCATION	Mermaids	4.2	14286	Varies with device	5000000	Free	Everyone	I
	Princess Coloring Book	4.5	9770	39M	5000000	Free	Everyone	I
ENTERTAINMENT	Adult Glitter Color by Number Book - Sandbox Pages	4.3	8918	Varies with device	1000000	Free	Everyone	Ente

Category	App	Rating	Reviews	Size	Installs	Price	Content Rating	
FAMILY	Barbie Magical Fashion	4.0	328619	15M	10000000	Free	Everyone	
	Beauty and the Beast	4.4	70883	31M	1000000	Free	Everyone	
	Color By Number - Sandbox Pixel Coloring Book	4.7	24557	24M	1000000	Free	Everyone	
	Coloring & Learn	4.4	12753	51M	5000000	Free	Everyone	Ec
	Cutie Cubies	4.4	6356	83M	500000	Free	Everyone	
	Dolphin and fish coloring book	3.9	2249	Varies with device	500000	Free	Everyone	Art
	Dr. Panda Art Class	4.2	1013	58M	50000	$2.99	Everyone	I
	Draw Color by Number - Sandbox Pixel Art	4.7	34279	24M	1000000	Free	Everyone	
	Fairy Kingdom: World of Magic and Farming	4.4	129542	63M	1000000	Free	Everyone	
	Hello Kitty Lunchbox	4.2	51838	15M	5000000	Free	Everyone	
	Messenger Kids – Safer Messaging and Video Chat	4.2	3478	Varies with device	500000	Free	Everyone	Comm
	No.Color – Color by Number	4.4	23474	8.8M	5000000	Free	Everyone	
	No.Diamond – Colors by Number	4.6	9016	24M	1000000	Free	Everyone	Ente
	Pet Beauty Salon	3.8	20292	47M	1000000	Free	Everyone	Ec
	Pinkalicious Party	3.0	2	82M	500	$2.99	Everyone	Ec
	Princess Adventures Puzzles	4.4	382	44M	500000	Free	Everyone	I
	Princess Palace: Royal Pony	4.4	11442	66M	1000000	Free	Everyone	Ec

Category	App	Rating	Reviews	Size	Installs	Price	Content Rating	
	Sandbox - Color by Number Coloring Pages	4.7	412744	10M	10000000	Free	Everyone	Ente
	Toca Builders	4.2	3328	Varies with device	100000	$3.99	Everyone	I
	Toca Mystery House	4.2	96	79M	5000	$3.99	Everyone	Ec
	Video Editor	4.1	159619	23M	5000000	Free	Everyone	
	Wuwu & Co.	4.3	9	77M	100	$2.99	Everyone	F
GAME	**Barbie™ Fashion Closet**	4.1	68057	85M	10000000	Free	Everyone	

1.8.23.1 From the grouped table we can see most apps with 'Creativity' genre are in 'FAMILY' category

1.8.24 (xiii) Query 13 :

1.8.25 Get all the apps with android ver "5.0 and up", rating greater than 4.5 and installs greater than 10 million

In [62]:
```
apps_df[(apps_df['Android Ver']=="5.0 and up") & (apps_df['Rating']>=4.5)
```

Out[62]:

	App	Category	Rating	Reviews	Size	Installs	Price	
1077	Capital One® Mobile	FINANCE	4.6	510392	79M	10000000	Free	E
1173	Chase Mobile	FINANCE	4.6	1374549	32M	10000000	Free	E
1296	8fit Workouts & Meal Planner	HEALTH_AND_FITNESS	4.6	115721	67M	10000000	Free	E
1311	Freeletics: Personal Trainer & Fitness Workouts	HEALTH_AND_FITNESS	4.5	130104	25M	10000000	Free	E
1312	Nike Training Club - Workouts & Fitness	HEALTH_AND_FITNESS	4.6	251534	93M	10000000	Free	E

1.8.25.1 So these apps are the apps of higher ver with most installs and ratings

1.8.26 In conclusion, android app industry plays a major part in everybody's life. Be it Entertainment, Education, Games, Business, anything, appropriate apps are always required for the same. There is always scope for numerous oppurtunities and improvement in this industry and generous attention should be paid to the same

In [ ]: