

# GooglePlayStore\_AppAnalysis

September 18, 2019

```
In [101]: import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib as mlp
from os import path
from PIL import Image
from wordcloud import WordCloud, STOPWORDS, ImageColorGenerator
import matplotlib.pyplot as plt
```

```
In [102]: gl=pd.read_csv("googleplaystore.csv")
```

```
In [103]: re=pd.read_csv("googleplaystore_user_reviews.csv")
```

```
In [105]: re.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 64295 entries, 0 to 64294
Data columns (total 5 columns):
App                64295 non-null object
Translated_Review  37427 non-null object
Sentiment          37432 non-null object
Sentiment_Polarity 37432 non-null float64
Sentiment_Subjectivity 37432 non-null float64
dtypes: float64(2), object(3)
memory usage: 2.5+ MB
```

```
In [106]: re.head()
```

```
Out[106]:
```

	App	Translated_Review	
0	10 Best Foods for You	I like eat delicious food. That's I'm cooking ...	
1	10 Best Foods for You	This help eating healthy exercise regular basis	
2	10 Best Foods for You		NaN
3	10 Best Foods for You	Works great especially going grocery store	
4	10 Best Foods for You	Best idea us	

  

	Sentiment	Sentiment_Polarity	Sentiment_Subjectivity
0	Positive	1.00	0.533333

1	Positive	0.25	0.288462
2	NaN	NaN	NaN
3	Positive	0.40	0.875000
4	Positive	1.00	0.300000

In [107]: gl.head()

Out[107]:

	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

In [108]: gl.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10841 entries, 0 to 10840
Data columns (total 13 columns):
App                10841 non-null object
Category           10841 non-null object
Rating             9367 non-null float64
Reviews            10841 non-null object
Size               10841 non-null object
Installs           10841 non-null object
Type               10840 non-null object
Price              10841 non-null object
Content Rating     10840 non-null object
```

```

Genres          10841 non-null object
Last Updated    10841 non-null object
Current Ver     10833 non-null object
Android Ver     10838 non-null object
dtypes: float64(1), object(12)
memory usage: 1.1+ MB

```

```

In [109]: cols = ['Reviews', 'Size', 'Price']
          gl[cols] = gl[cols].apply(pd.to_numeric, errors='coerce', axis=1)

```

```

In [110]: gl.info()

```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10841 entries, 0 to 10840
Data columns (total 13 columns):
App          10841 non-null object
Category     10841 non-null object
Rating       9367 non-null float64
Reviews      10840 non-null float64
Size         0 non-null float64
Installs     10841 non-null object
Type         10840 non-null object
Price        10040 non-null float64
Content Rating 10840 non-null object
Genres       10841 non-null object
Last Updated 10841 non-null object
Current Ver  10833 non-null object
Android Ver  10838 non-null object
dtypes: float64(4), object(9)
memory usage: 1.1+ MB

```

```

In [111]: # Groupby by country
          category= gl.groupby("Category")
          care=re.groupby("App")

```

```

In [112]: # Drop all rows with missing values
          clre=re.dropna()

```

```

In [113]: clre.head()

```

```

Out[113]:
           App  Translated_Review \
0  10 Best Foods for You  I like eat delicious food. That's I'm cooking ...
1  10 Best Foods for You  This help eating healthy exercise regular basis
3  10 Best Foods for You  Works great especially going grocery store
4  10 Best Foods for You  Best idea us
5  10 Best Foods for You  Best way

```

	Sentiment	Sentiment_Polarity	Sentiment_Subjectivity
0	Positive	1.00	0.533333
1	Positive	0.25	0.288462
3	Positive	0.40	0.875000
4	Positive	1.00	0.300000
5	Positive	1.00	0.300000

```
In [114]: # Start with one review
text=clre.Translated_Review[0]
```

```
In [115]: wordcloud=WordCloud().generate(text)
```

```
In [116]: # Display the image
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()
```



```
In [117]: # lower max_font_size, change the maximum number of word and lighten the background:
wordcloud = WordCloud(max_font_size=50, max_words=100, background_color="white").generate(text)
plt.figure()
plt.imshow(wordcloud, interpolation="bilinear")
plt.axis("off")
plt.show()
```



```
In [118]: text = " ".join(review for review in clre.Translated_Review)
          print ("There are {} words in the combination of all review.".format(len(text)))
```

There are 4274608 words in the combination of all review.

```
In [119]: # Create stopword list:
          stopwords = set(STOPWORDS)
          stopwords.update(["app", "good", "phone", "make", "still", "give", "even", "need", "want"])

          # Generate a word cloud image
          wordcloud = WordCloud(stopwords=stopwords, background_color="white").generate(text)

          # Display the generated image:
          # the matplotlib way:
          plt.imshow(wordcloud, interpolation='bilinear')
          plt.axis("off")
          plt.show()
```



PERSONALIZATION	309.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	314.0
PHOTOGRAPHY	313.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	317.0
PRODUCTIVITY	396.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	351.0
SHOPPING	258.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	238.0
SOCIAL	292.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	259.0
SPORTS	360.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	319.0
TOOLS	765.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	734.0
TRAVEL_AND_LOCAL	246.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	226.0
VIDEO_PLAYERS	171.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	160.0
WEATHER	74.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	75.0

Category	mean	...	Reviews		Size					\
			75%		max	count	mean	std	min	
Category		...								
1.9	19.000000	...	NaN	NaN	NaN	0.0	NaN	NaN	NaN	
ART_AND_DESIGN	4.358065	...	5754.00	295237.0	0.0	NaN	NaN	NaN	NaN	
AUTO_AND_VEHICLES	4.190411	...	4140.00	271920.0	0.0	NaN	NaN	NaN	NaN	
BEAUTY	4.278571	...	3595.00	113715.0	0.0	NaN	NaN	NaN	NaN	
BOOKS_AND_REFERENCE	4.346067	...	19085.00	2915189.0	0.0	NaN	NaN	NaN	NaN	
BUSINESS	4.121452	...	3810.75	1279800.0	0.0	NaN	NaN	NaN	NaN	
COMICS	4.155172	...	19748.25	1013944.0	0.0	NaN	NaN	NaN	NaN	
COMMUNICATION	4.158537	...	447478.50	69119316.0	0.0	NaN	NaN	NaN	NaN	
DATING	3.970769	...	11762.75	516917.0	0.0	NaN	NaN	NaN	NaN	
EDUCATION	4.389032	...	144111.25	6290507.0	0.0	NaN	NaN	NaN	NaN	
ENTERTAINMENT	4.126174	...	288150.00	7165362.0	0.0	NaN	NaN	NaN	NaN	
EVENTS	4.435556	...	874.50	40113.0	0.0	NaN	NaN	NaN	NaN	
FAMILY	4.192272	...	26455.75	44881447.0	0.0	NaN	NaN	NaN	NaN	
FINANCE	4.131889	...	19848.50	1374549.0	0.0	NaN	NaN	NaN	NaN	
FOOD_AND_DRINK	4.166972	...	45076.50	1032935.0	0.0	NaN	NaN	NaN	NaN	
GAME	4.286326	...	643259.00	44893888.0	0.0	NaN	NaN	NaN	NaN	
HEALTH_AND_FITNESS	4.277104	...	69395.00	4559407.0	0.0	NaN	NaN	NaN	NaN	
HOUSE_AND_HOME	4.197368	...	36857.00	417907.0	0.0	NaN	NaN	NaN	NaN	
LIBRARIES_AND_DEMO	4.178462	...	2087.00	332083.0	0.0	NaN	NaN	NaN	NaN	
LIFESTYLE	4.094904	...	6608.75	2789775.0	0.0	NaN	NaN	NaN	NaN	
MAPS_AND_NAVIGATION	4.051613	...	41225.00	7232629.0	0.0	NaN	NaN	NaN	NaN	
MEDICAL	4.189143	...	490.00	156410.0	0.0	NaN	NaN	NaN	NaN	
NEWS_AND_MAGAZINES	4.132189	...	41232.50	11667403.0	0.0	NaN	NaN	NaN	NaN	
PARENTING	4.300000	...	3657.75	658087.0	0.0	NaN	NaN	NaN	NaN	
PERSONALIZATION	4.335987	...	28905.25	7464996.0	0.0	NaN	NaN	NaN	NaN	
PHOTOGRAPHY	4.192114	...	351254.50	10859051.0	0.0	NaN	NaN	NaN	NaN	
PRODUCTIVITY	4.211396	...	118547.25	5383985.0	0.0	NaN	NaN	NaN	NaN	
SHOPPING	4.259664	...	144657.75	6212081.0	0.0	NaN	NaN	NaN	NaN	
SOCIAL	4.255598	...	363520.50	78158306.0	0.0	NaN	NaN	NaN	NaN	
SPORTS	4.223511	...	91031.00	14184910.0	0.0	NaN	NaN	NaN	NaN	
TOOLS	4.047411	...	23347.50	42916526.0	0.0	NaN	NaN	NaN	NaN	
TRAVEL_AND_LOCAL	4.109292	...	48930.00	9235373.0	0.0	NaN	NaN	NaN	NaN	
VIDEO_PLAYERS	4.063750	...	139926.00	25655305.0	0.0	NaN	NaN	NaN	NaN	
WEATHER	4.244000	...	64351.50	2371543.0	0.0	NaN	NaN	NaN	NaN	

	25%	50%	75%	max
Category				
1.9	NaN	NaN	NaN	NaN
ART_AND_DESIGN	NaN	NaN	NaN	NaN
AUTO_AND_VEHICLES	NaN	NaN	NaN	NaN
BEAUTY	NaN	NaN	NaN	NaN
BOOKS_AND_REFERENCE	NaN	NaN	NaN	NaN
BUSINESS	NaN	NaN	NaN	NaN
COMICS	NaN	NaN	NaN	NaN
COMMUNICATION	NaN	NaN	NaN	NaN
DATING	NaN	NaN	NaN	NaN
EDUCATION	NaN	NaN	NaN	NaN
ENTERTAINMENT	NaN	NaN	NaN	NaN
EVENTS	NaN	NaN	NaN	NaN
FAMILY	NaN	NaN	NaN	NaN
FINANCE	NaN	NaN	NaN	NaN
FOOD_AND_DRINK	NaN	NaN	NaN	NaN
GAME	NaN	NaN	NaN	NaN
HEALTH_AND_FITNESS	NaN	NaN	NaN	NaN
HOUSE_AND_HOME	NaN	NaN	NaN	NaN
LIBRARIES_AND_DEMO	NaN	NaN	NaN	NaN
LIFESTYLE	NaN	NaN	NaN	NaN
MAPS_AND_NAVIGATION	NaN	NaN	NaN	NaN
MEDICAL	NaN	NaN	NaN	NaN
NEWS_AND_MAGAZINES	NaN	NaN	NaN	NaN
PARENTING	NaN	NaN	NaN	NaN
PERSONALIZATION	NaN	NaN	NaN	NaN
PHOTOGRAPHY	NaN	NaN	NaN	NaN
PRODUCTIVITY	NaN	NaN	NaN	NaN
SHOPPING	NaN	NaN	NaN	NaN
SOCIAL	NaN	NaN	NaN	NaN
SPORTS	NaN	NaN	NaN	NaN
TOOLS	NaN	NaN	NaN	NaN
TRAVEL_AND_LOCAL	NaN	NaN	NaN	NaN
VIDEO_PLAYERS	NaN	NaN	NaN	NaN
WEATHER	NaN	NaN	NaN	NaN

[34 rows x 32 columns]

```
In [121]: com=gl.loc[gl['Category'] == "COMMUNICATION"]
          com.head()
```

```
Out[121]:
```

	App	Category	Rating	\
335	Messenger Text and Video Chat for Free	COMMUNICATION	4.0	
336	WhatsApp Messenger	COMMUNICATION	4.4	
337	Messenger for SMS	COMMUNICATION	4.3	



338	Google Chrome: Fast & Secure	COMMUNICATION	4.3
339	Messenger Lite: Free Calls & Messages	COMMUNICATION	4.4

	Reviews	Size	Installs	Type	Price	Content Rating	\
335	56642847.0	NaN	1,000,000,000+	Free	0.0	Everyone	
336	69119316.0	NaN	1,000,000,000+	Free	0.0	Everyone	
337	125257.0	NaN	10,000,000+	Free	0.0	Teen	
338	9642995.0	NaN	1,000,000,000+	Free	0.0	Everyone	
339	1429035.0	NaN	100,000,000+	Free	0.0	Everyone	

	Genres	Last Updated	Current Ver	Android Ver
335	Communication	August 1, 2018	Varies with device	Varies with device
336	Communication	August 3, 2018	Varies with device	Varies with device
337	Communication	June 6, 2018	1.8.9	4.1 and up
338	Communication	August 1, 2018	Varies with device	Varies with device
339	Communication	July 25, 2018	37.0.0.7.163	2.3 and up

```
In [122]: com.to_csv('com.csv')
```

```
In [123]: game=gl.loc[gl['Category'] == "GAME"]
game.head()
```

```
Out[123]:
```

	App	Category	Rating	Reviews	Size	Installs	\
1653	ROBLOX	GAME	4.5	4447388.0	NaN	100,000,000+	
1654	Subway Surfers	GAME	4.5	27722264.0	NaN	1,000,000,000+	
1655	Candy Crush Saga	GAME	4.4	22426677.0	NaN	500,000,000+	
1656	Solitaire	GAME	4.7	254258.0	NaN	10,000,000+	
1657	Bubble Shooter	GAME	4.5	148897.0	NaN	10,000,000+	

	Type	Price	Content Rating	Genres	\
1653	Free	0.0	Everyone 10+	Adventure;Action &	Adventure
1654	Free	0.0	Everyone 10+		Arcade
1655	Free	0.0	Everyone		Casual
1656	Free	0.0	Everyone		Card
1657	Free	0.0	Everyone		Casual

	Last Updated	Current Ver	Android Ver
1653	July 31, 2018	2.347.225742	4.1 and up
1654	July 12, 2018	1.90.0	4.1 and up
1655	July 5, 2018	1.129.0.2	4.1 and up
1656	August 1, 2018	2.137.0	4.1 and up
1657	July 17, 2018	1.20.1	4.0.3 and up

```
In [124]: game.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1144 entries, 1653 to 10804
Data columns (total 13 columns):
App                1144 non-null object
```

```

Category      1144 non-null object
Rating        1097 non-null float64
Reviews       1144 non-null float64
Size          0 non-null float64
Installs      1144 non-null object
Type          1144 non-null object
Price         1061 non-null float64
Content Rating 1144 non-null object
Genres        1144 non-null object
Last Updated  1144 non-null object
Current Ver   1144 non-null object
Android Ver   1144 non-null object
dtypes: float64(4), object(9)
memory usage: 125.1+ KB

```

```
In [125]: game.to_csv('game.csv')
```

```
In [126]: top2=game.append(com)
          top2.head()
```

```
Out[126]:
```

	App	Category	Rating	Reviews	Size	Installs \
1653	ROBLOX	GAME	4.5	4447388.0	NaN	100,000,000+
1654	Subway Surfers	GAME	4.5	27722264.0	NaN	1,000,000,000+
1655	Candy Crush Saga	GAME	4.4	22426677.0	NaN	500,000,000+
1656	Solitaire	GAME	4.7	254258.0	NaN	10,000,000+
1657	Bubble Shooter	GAME	4.5	148897.0	NaN	10,000,000+

	Type	Price	Content Rating	Genres \
1653	Free	0.0	Everyone 10+	Adventure;Action & Adventure
1654	Free	0.0	Everyone 10+	Arcade
1655	Free	0.0	Everyone	Casual
1656	Free	0.0	Everyone	Card
1657	Free	0.0	Everyone	Casual

	Last Updated	Current Ver	Android Ver
1653	July 31, 2018	2.347.225742	4.1 and up
1654	July 12, 2018	1.90.0	4.1 and up
1655	July 5, 2018	1.129.0.2	4.1 and up
1656	August 1, 2018	2.137.0	4.1 and up
1657	July 17, 2018	1.20.1	4.0.3 and up

```
In [127]: top2.to_csv('top2.csv')
```

```
In [128]: ap=pd.read_csv("AppleStore.csv")
```

```
In [129]: ap.dropna()
```

```
Out[129]:
```

	Unnamed: 0	id \
0	1	281656475

1	2	281796108
2	3	281940292
3	4	282614216
4	5	282935706
5	6	283619399
6	7	283646709
7	8	284035177
8	9	284666222
9	10	284736660
10	11	284791396
11	12	284815117
12	13	284815942
13	14	284847138
14	15	284862767
15	16	284876795
16	17	284882215
17	18	284910350
18	20	284993459
19	21	285005463
20	22	285946052
21	23	285994151
22	24	286058814
23	25	286070473
24	27	286799607
25	28	286906691
26	29	286911400
27	31	288113403
28	32	288120394
29	33	288419283
...	...	...
7167	10995	1182265441
7168	10998	1182331762
7169	11002	1182568288
7170	11010	1183234072
7171	11013	1183260922
7172	11016	1183548754
7173	11019	1183709176
7174	11022	1183856228
7175	11024	1183986102
7176	11027	1184711626
7177	11031	1184800011
7178	11033	1185209084
7179	11035	1185328193
7180	11036	1185365336
7181	11038	1185428381
7182	11040	1185538497
7183	11041	1185580782
7184	11042	1185731859

7185	11043	1185777521
7186	11050	1186108496
7187	11051	1186126548
7188	11060	1186384912
7189	11074	1187128255
7190	11077	1187279979
7191	11079	1187282363
7192	11081	1187617475
7193	11082	1187682390
7194	11087	1187779532
7195	11089	1187838770
7196	11097	1188375727

	track_name	size_bytes	currency	\
0	PAC-MAN Premium	100788224	USD	
1	Evernote - stay organized	158578688	USD	
2	WeatherBug - Local Weather, Radar, Maps, Alerts	100524032	USD	
3	eBay: Best App to Buy, Sell, Save! Online Shop...	128512000	USD	
4	Bible	92774400	USD	
5	Shanghai Mahjong	10485713	USD	
6	PayPal - Send and request money safely	227795968	USD	
7	Pandora - Music & Radio	130242560	USD	
8	PCalc - The Best Calculator	49250304	USD	
9	Ms. PAC-MAN	70023168	USD	
10	Solitaire by MobilityWare	49618944	USD	
11	SCRABBLE Premium	227547136	USD	
12	Google Search made just for mobile	179979264	USD	
13	Bank of America - Mobile Banking	160925696	USD	
14	FreeCell	55153664	USD	
15	TripAdvisor Hotels Flights Restaurants	207907840	USD	
16	Facebook	389879808	USD	
17	Yelp - Nearby Restaurants, Shopping & Services	167407616	USD	
18	Shazam - Discover music, artists, videos & lyrics	147093504	USD	
19	Crash Bandicoot Nitro Kart 3D	10735026	USD	
20	iQuran	70707916	USD	
21	:) Sudoku +	6169600	USD	
22	Yahoo Sports - Teams, Scores, News & Highlights	130583552	USD	
23	Mileage Log   Fahrtenbuch	71203840	USD	
24	ClearTune - Chromatic Tuner	11423008	USD	
25	Lifesum Inspiring healthy lifestyle app	188017664	USD	
26	Hangman.	4765696	USD	
27	iTranslate - Language Translator & Dictionary	287933440	USD	
28	TouchOSC	4263936	USD	
29	RadarScope	172772352	USD	
...	...	...	...	...
7167	-- 177498112	USD		
7168	Escape from the frigid Igloo.	89188352	USD	
7169	Talking Santa - Video santa claus calls you	32685056	USD	

7170	CTFxCmoji	26077184	USD
7171	Room Escape Game - Santa's Room	143346688	USD
7172	Rescue the Enchanter	242505728	USD
7173	My Diary -	18164736	USD
7174	VR Thrills: Roller Coaster 360 (Google Cardboard)	169535488	USD
7175	Santa Kids Hair Salon - Christmas Makeover Games	64244736	USD
7176	Human Juggling Cup	184324096	USD
7177	Again - room escape game	33946624	USD
7178	Saloons Unleashed	327731200	USD
7179	Fam Group video calling for iMessage	113382400	USD
7180	Laurie Hernandez the Human Emoji	94008320	USD
7181		178160640	USD
7182	camera for filter	9362432	USD
7183	Survivalcraft 2	57349120	USD
7184	-	171944960	USD
7185	-	208026624	USD
7186	- - ""	85580800	USD
7187	Escape Game: illumination	52342784	USD
7188	Demolition Derby Virtual Reality (VR) Racing	168774656	USD
7189	-PK	537462784	USD
7190	Add-Ons Studio for Minecraft	22999040	USD
7191	Plead the Fifth - The Game	27853824	USD
7192	Kubik	126644224	USD
7193	VR Roller-Coaster	120760320	USD
7194	Bret Michaels Emojis + Lyric Keyboard	111322112	USD
7195	VR Roller Coaster World - Virtual Reality	97235968	USD
7196	Escape the Sweet Shop Series	90898432	USD

	price	rating_count_tot	rating_count_ver	user_rating	user_rating_ver	\
0	3.99	21292	26	4.0	4.5	
1	0.00	161065	26	4.0	3.5	
2	0.00	188583	2822	3.5	4.5	
3	0.00	262241	649	4.0	4.5	
4	0.00	985920	5320	4.5	5.0	
5	0.99	8253	5516	4.0	4.0	
6	0.00	119487	879	4.0	4.5	
7	0.00	1126879	3594	4.0	4.5	
8	9.99	1117	4	4.5	5.0	
9	3.99	7885	40	4.0	4.0	
10	4.99	76720	4017	4.5	4.5	
11	7.99	105776	166	3.5	2.5	
12	0.00	479440	203	3.5	4.0	
13	0.00	119773	2336	3.5	4.5	
14	4.99	6340	668	4.5	4.5	
15	0.00	56194	87	4.0	3.5	
16	0.00	2974676	212	3.5	3.5	
17	0.00	223885	3726	4.0	4.5	
18	0.00	402925	136	4.0	4.5	

19	2.99	31456	4178	4.0	3.5
20	1.99	2929	966	4.5	4.5
21	2.99	11447	781	5.0	5.0
22	0.00	137951	131	4.0	4.5
23	5.99	8	0	4.5	0.0
24	3.99	3241	297	4.0	4.0
25	0.00	5795	12	3.5	4.0
26	0.00	42316	248	3.0	3.5
27	0.00	123215	25	3.5	5.0
28	4.99	782	7	4.0	3.5
29	9.99	3449	23	4.0	4.5
...	...	...	...	...	...
7167	0.00	1	1	1.0	1.0
7168	0.00	3	3	4.0	4.0
7169	2.99	9	3	3.0	3.5
7170	0.00	39	4	5.0	5.0
7171	0.00	10	10	5.0	5.0
7172	3.99	55	29	4.5	4.5
7173	0.99	0	0	0.0	0.0
7174	0.00	14	4	4.0	3.5
7175	0.00	41	19	4.5	4.5
7176	0.00	0	0	0.0	0.0
7177	0.00	11	8	4.0	4.0
7178	0.99	0	0	0.0	0.0
7179	0.00	279	5	3.5	3.0
7180	0.00	26	3	5.0	5.0
7181	0.99	0	0	0.0	0.0
7182	0.00	0	0	0.0	0.0
7183	3.99	292	292	4.0	4.0
7184	0.00	0	0	0.0	0.0
7185	0.99	0	0	0.0	0.0
7186	0.00	1	1	2.0	2.0
7187	0.00	23	23	4.5	4.5
7188	0.00	18	18	4.0	4.0
7189	0.99	0	0	0.0	0.0
7190	2.99	97	97	3.0	3.0
7191	2.99	11	0	4.0	0.0
7192	0.00	142	75	4.5	4.5
7193	0.00	30	30	4.5	4.5
7194	1.99	15	0	4.5	0.0
7195	0.00	85	32	4.5	4.5
7196	0.00	3	3	5.0	5.0

	ver	cont_rating	prime_genre	sup_devices.num \
0	6.3.5	4+	Games	38
1	8.2.2	4+	Productivity	37
2	5.0.0	4+	Weather	37
3	5.10.0	12+	Shopping	37

4	7.5.1	4+	Reference	37
5	1.8	4+	Games	47
6	6.12.0	4+	Finance	37
7	8.4.1	12+	Music	37
8	3.6.6	4+	Utilities	37
9	4.0.4	4+	Games	38
10	4.10.1	4+	Games	38
11	5.19.0	4+	Games	37
12	27	17+	Utilities	37
13	7.3.8	4+	Finance	37
14	4.0.3	4+	Games	38
15	21.1	4+	Travel	37
16	95	4+	Social	37
17	11.15.0	12+	Travel	37
18	11.0.3	12+	Music	37
19	1.0.0	4+	Games	47
20	3.3	4+	Reference	43
21	5.2.6	4+	Games	40
22	6.9	4+	Sports	37
23	9.0.5	4+	Business	37
24	2.1.3	4+	Music	43
25	8.4.1	4+	Health & Fitness	37
26	2.0.6	9+	Games	38
27	10.5.4	4+	Productivity	37
28	1.9.8	4+	Music	43
29	3.4.1	4+	Weather	37
...	...	...	...	...
7167	1.0.1	4+	Games	38
7168	1.0.1	4+	Games	38
7169	1.0.1	4+	Entertainment	37
7170	1.3	9+	Social	37
7171	1	4+	Games	40
7172	1.4	4+	Games	40
7173	1.1	4+	Utilities	37
7174	1.3.0	4+	Games	37
7175	1.3	4+	Games	37
7176	1.0.1	4+	Games	40
7177	1.1.0	4+	Games	37
7178	1.1	4+	Games	37
7179	0.6.41	4+	Social	37
7180	1.0.5	9+	Utilities	37
7181	1	9+	Games	40
7182	1.0.3	4+	Photo & Video	37
7183	2.0.20.1	9+	Games	40
7184	1	9+	Games	40
7185	1	9+	Games	38
7186	1.0.1	4+	Games	38
7187	1	4+	Games	37

7188	1.0.0	12+	Games	38
7189	2.1.0	9+	Games	38
7190	1	4+	Games	37
7191	1.1.1	17+	Games	37
7192	1.3	4+	Games	38
7193	0.9	4+	Games	38
7194	1.0.2	9+	Utilities	37
7195	1.0.15	12+	Games	38
7196	1	4+	Games	40

	ipadSc_urls.num	lang.num	vpp_lic
0	5	10	1
1	5	23	1
2	5	3	1
3	5	9	1
4	5	45	1
5	5	1	1
6	0	19	1
7	4	1	1
8	5	1	1
9	0	10	1
10	4	11	1
11	0	6	1
12	4	33	1
13	0	2	1
14	5	2	1
15	1	26	1
16	1	29	1
17	5	18	1
18	3	16	1
19	0	1	1
20	0	2	1
21	5	1	1
22	2	6	1
23	5	3	1
24	2	10	1
25	5	11	1
26	0	1	1
27	5	23	1
28	1	1	1
29	5	1	1
...	...	...	...
7167	4	1	1
7168	5	1	1
7169	0	31	1
7170	1	1	1
7171	0	1	1
7172	5	1	1



7173	0	1	1
7174	5	1	1
7175	5	25	1
7176	4	1	1
7177	5	1	1
7178	5	1	1
7179	4	1	1
7180	4	1	1
7181	5	0	1
7182	0	1	1
7183	5	1	1
7184	5	0	1
7185	5	1	1
7186	5	1	1
7187	5	2	1
7188	4	1	1
7189	5	1	1
7190	5	3	1
7191	0	1	1
7192	5	1	1
7193	0	1	1
7194	1	1	1
7195	0	2	1
7196	0	2	1

[7197 rows x 17 columns]

```
In [130]: apgame=ap.loc[ap['prime_genre'] == "Games"]
```

```
In [131]: apcom=ap.loc[ap['prime_genre']=='Social']
```

```
In [132]: aptop2=apgame.append(apcom)
```

```
In [133]: aptop2.to_csv('aptop2.csv')
```

```
In [16]: top_cat=gl.groupby('Category').size().reset_index(name='Count').nlargest(6,'Count')
display(top_cat)
```

	Category	Count
12	FAMILY	1972
15	GAME	1144
30	TOOLS	843
21	MEDICAL	463
5	BUSINESS	460
26	PRODUCTIVITY	424

```
In [11]: gl['Reviews'].max()
```

```
Out[11]: '9992'
```

```

In [10]: intdata=blk.drop(columns=['User_ID', 'Gender', 'Age', 'Product_ID', 'City_Category', 'Stay_
intdata

Out[10]: Empty DataFrame
Columns: [App, Category, Rating, Reviews, Size, Installs, Type, Price, Content Rating
Index: []

In [5]: Installs = np.unique(gl['Installs']) # Age segmentation
Installs

Out[5]: array(['0', '0+', '1+', '1,000+', '1,000,000+', '1,000,000,000+', '10+',
'10,000+', '10,000,000+', '100+', '100,000+', '100,000,000+', '5+',
'5,000+', '5,000,000+', '50+', '50,000+', '50,000,000+', '500+',
'500,000+', '500,000,000+', 'Free'], dtype=object)

In [13]: price=np.unique(gl['Price'])
price

Out[13]: array(['$0.99', '$1.00', '$1.04', '$1.20', '$1.26', '$1.29', '$1.49',
'$1.50', '$1.59', '$1.61', '$1.70', '$1.75', '$1.76', '$1.96',
'$1.97', '$1.99', '$10.00', '$10.99', '$109.99', '$11.99',
'$12.99', '$13.99', '$14.00', '$14.99', '$15.46', '$15.99',
'$154.99', '$16.99', '$17.99', '$18.99', '$19.40', '$19.90',
'$19.99', '$2.00', '$2.49', '$2.50', '$2.56', '$2.59', '$2.60',
'$2.90', '$2.95', '$2.99', '$200.00', '$24.99', '$25.99', '$28.99',
'$29.99', '$299.99', '$3.02', '$3.04', '$3.08', '$3.28', '$3.49',
'$3.61', '$3.88', '$3.90', '$3.95', '$3.99', '$30.99', '$33.99',
'$37.99', '$379.99', '$389.99', '$39.99', '$394.99', '$399.99',
'$4.29', '$4.49', '$4.59', '$4.60', '$4.77', '$4.80', '$4.84',
'$4.85', '$4.99', '$400.00', '$46.99', '$5.00', '$5.49', '$5.99',
'$6.49', '$6.99', '$7.49', '$7.99', '$74.99', '$79.99', '$8.49',
'$8.99', '$89.99', '$9.00', '$9.99', '0', 'Everyone'], dtype=object)

In [15]: free=gl.loc[gl['Price'] == 0]
free

Out[15]: Empty DataFrame
Columns: [App, Category, Rating, Reviews, Size, Installs, Type, Price, Content Rating
Index: []

In [7]: sns.countplot(gl['Installs'],hue=gl['Category'])

Out[7]: <matplotlib.axes._subplots.AxesSubplot at 0x1a16951828>

```



```

Out[12]: Installs
          0          1
0+         14
1+         67
1,000+     907
1,000,000+ 1579
1,000,000,000+ 58
10+        386
10,000+    1054
10,000,000+ 1252
100+       719
100,000+   1169
100,000,000+ 409
5+         82
5,000+     477
5,000,000+ 752
50+        205
50,000+    479
50,000,000+ 289
500+       330
500,000+   539
500,000,000+ 72
Free        1
Name: Category, dtype: int64

```