

Apps final

April 3, 2019

1 App store data

Suppose that we are going to develop a new app, we are going to ignore staffing issues for the time being but address them at various points in the project. The standing assumption will be that whatever the app will be supported by in app ads. We are going to do some research into popular apps on both the Google Play and Apple Store as these are the two largest markets for smartphone users.

The data that is being used can be found at [Apple](#) and [Android](#)

It is worth noting that this project will probably be under construction for the foreseeable future as I learn and dabble with new techniques.

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import nbconvert

df_an = pd.read_csv('googleplaystore.csv', index_col = False)
df_ios = pd.read_csv('AppleStore.csv')

In [2]: print(df_an.dtypes)
print('\n')
print(df_ios.dtypes)
```

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

```

Unnamed: 0      int64
id              int64
track_name      object
size_bytes      int64
currency        object
price           float64
rating_count_tot int64
rating_count_ver int64
user_rating     float64
user_rating_ver float64
ver             object
cont_rating     object
prime_genre     object
sup_devices.num int64
ipadSc_urls.num int64
lang.num        int64
vpp_lic         int64
dtype: object

```

```
In [3]: df_an.head(5)
```

```

Out[3]:

```

	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 [4]: df_ios.head(5)
```

```

Out[4]:   Unnamed: 0      id      track_name \
0         1  281656475      PAC-MAN Premium
1         2  281796108      Evernote - stay organized
2         3  281940292  WeatherBug - Local Weather, Radar, Maps, Alerts
3         4  282614216  eBay: Best App to Buy, Sell, Save! Online Shop...
4         5  282935706      Bible

      size_bytes  currency  price  rating_count_tot  rating_count_ver \
0    100788224      USD    3.99             21292             26
1    158578688      USD    0.00             161065             26
2    100524032      USD    0.00             188583            2822
3    128512000      USD    0.00             262241             649
4     92774400      USD    0.00             985920            5320

      user_rating  user_rating_ver      ver  cont_rating  prime_genre \
0           4.0           4.5  6.3.5           4+      Games
1           4.0           3.5  8.2.2           4+  Productivity
2           3.5           4.5  5.0.0           4+      Weather
3           4.0           4.5  5.10.0          12+    Shopping
4           4.5           5.0  7.5.1           4+    Reference

      sup_devices.num  ipadSc_urls.num  lang.num  vpp_lic
0                38                5        10         1
1                37                5         23         1
2                37                5          3         1
3                37                5          9         1
4                37                5         45         1

```

2 Cleaning the data

We will proceed to do some basic cleaning for our data sets. First we take a reality check to make sure that our apps have the proper rating.

```

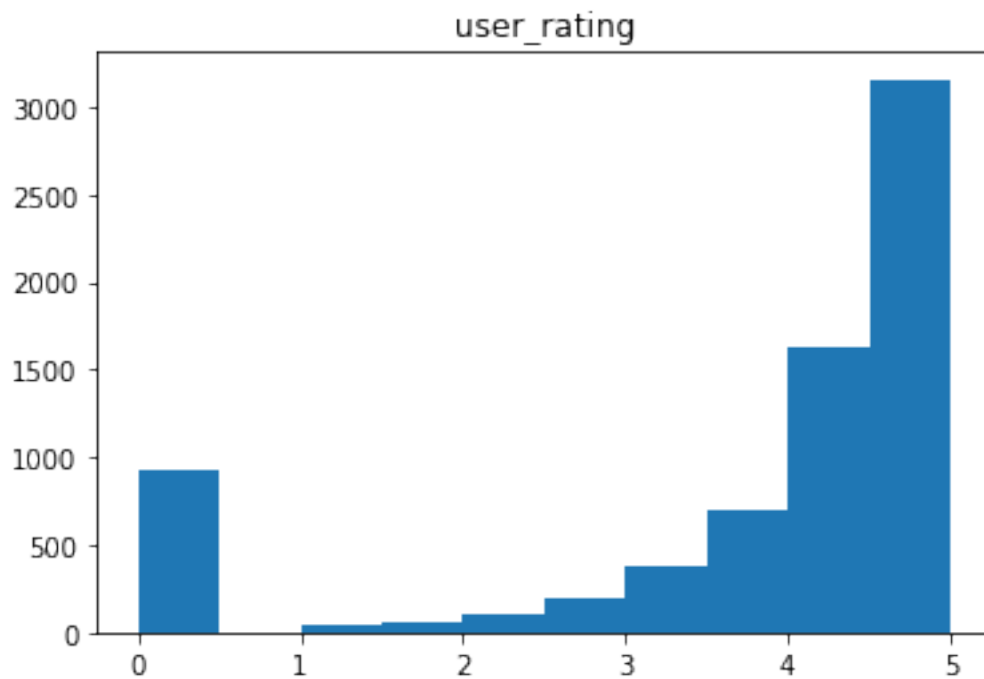
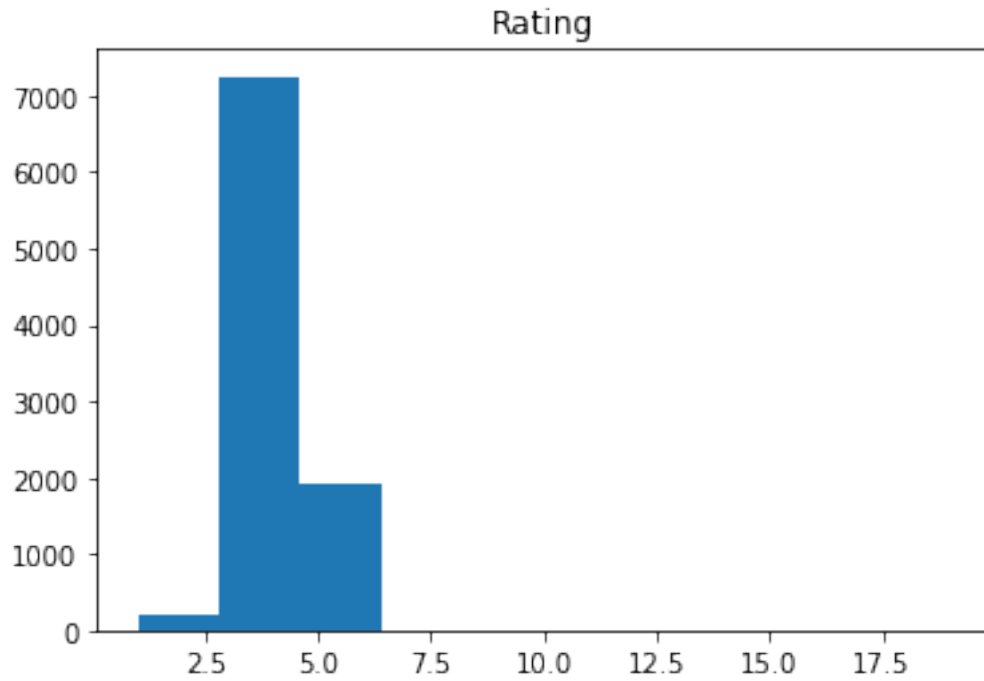
In [5]: df_an.hist(column = 'Rating', grid = False)
        df_ios.hist(column = 'user_rating', grid = False)

```

```

Out[5]: array([[<matplotlib.axes._subplots.AxesSubplot object at 0x000001C5E43E5F98>]],
              dtype=object)

```



Seeing as how the Android app histogram has a bin to the right of '5.0' and scales to at least 17.5, something must have gone wrong in the web scrape. We will find any such app in a second, but first we also want to check to see if any rating are not reported.

```
In [6]: df_an[df_an['Rating'].isnull() == True][['App', 'Installs', 'Reviews']].sort_values(by =
```

```
Out[6]:
```

	App	Installs	Reviews
1559	Young Speeches	500,000+	2221
5271	Aj Di Awaaz	500+	9
7277	Imperium CF	500+	9
7401	Mesure CI	500+	9
8322	DF	500+	9
8339	DF Wall Plus Droid Firewall	500+	9
10402	FH Calculator	500+	9
10411	FH Kufstein App	500+	9
10676	Calculator Fn	500+	9
6163	YourTube Video Views BG	500+	8
7161	CD Live - Club del Deportista	500+	8
5289	Ak-47 Wallpapers 2018	500+	7
6234	BH INVEST	500+	7
6482	BM Tahmin: Ücretsiz ddaa Tahminleri	500+	7
6928	IP address BW	500+	7
8329	DF BugMeNot	500+	7
9016	DW Cleaner	500+	7
10702	Theme fo Oppo A30 Wallpaper & Icon	500+	7
4130	F	500+	6
4269	J Balvin Piano Tiles	500+	6
6170	BG Cricket	500+	6
6480	BM Assets	500+	6
7171	CD JUANITO	500+	6
8055	CX-95WIFI	500+	6
8873	DT Practice	500+	6
9820	ES-1	500+	6
10236	EasyLive - FB Live Helper	500+	6
6857	Hilverda De Boer B.V. App	500+	5
7056	BZ Dealer	500+	5
7154	Design innovation CD Cassette	500+	5
...
9104	Dz kayas	1+	0
9139	DZ Register	1+	0
9201	EB Experience	1+	0
9262	EC-Contractors	1+	0
9264	EC Sampler	1+	0
9302	EF First	1+	0
9532	Ek-yatri: Travel where you Belong	1+	0
9664	Kenny e o Cupcake	1+	0
10101	Amleen Ey	1+	0
10417	FH Wallet	1+	0
10530	FK Sileks	1+	0
10541	FK Bregalnica tip	1+	0
10654	Wallpapers FN Herstal FNP 9	1+	0
10701	F-O-Meter	1+	0

10708	Custos F.O.	1+	0
4465	Popsicle Launcher for Android P 9.0 launcher	0+	0
5307	Ak Parti Yardm Toplama	0+	0
5486	AP Series Solution Pro	0+	0
5945	Ain Arabic Kids Alif Ba ta	0+	0
6692	cronometra-br	0+	0
7434	Pekalongan CJ	0+	0
8081	CX Network	0+	0
8614	Sweden Newspapers	0+	0
8871	Test Application DT 02	0+	0
9337	EG Explore Folegandros	0+	0
9719	EP Cook Book	0+	0
9905	Eu sou Rico	0+	0
9917	Eu Sou Rico	0+	0
9934	I'm Rich/Eu sou Rico/ /	0+	0
9148	Command & Conquer: Rivals	0	0

[1474 rows x 3 columns]

While we could have applied `.dropna()` to the data to remove these apps, we could have jumped the gun and deleted an app with a high number of downloads. As we can see, the app 'Young Speeches' claims to have at least half a million downloads and over two thousand reviews, but appears to have no rating in our data set. After a manual search the app does not readily appear on the Google Play store, so it can probably be ignored.

```
In [7]: df_ios[df_ios['user_rating'].isnull() == True][['track_name', 'rating_count_tot']].sort.
```

```
Out[7]: Empty DataFrame
Columns: [track_name, rating_count_tot]
Index: []
```

No apps on the Apple Store with rating NA! So now we turn our attention to finding the apps that skewed the Android histogram to the right.

```
In [8]: df_an[df_an['Rating'] > 5][['App', 'Installs', 'Reviews']]
```

```
Out[8]:
App Installs Reviews
10472 Life Made WI-Fi Touchscreen Photo Frame Free 3.0M
```

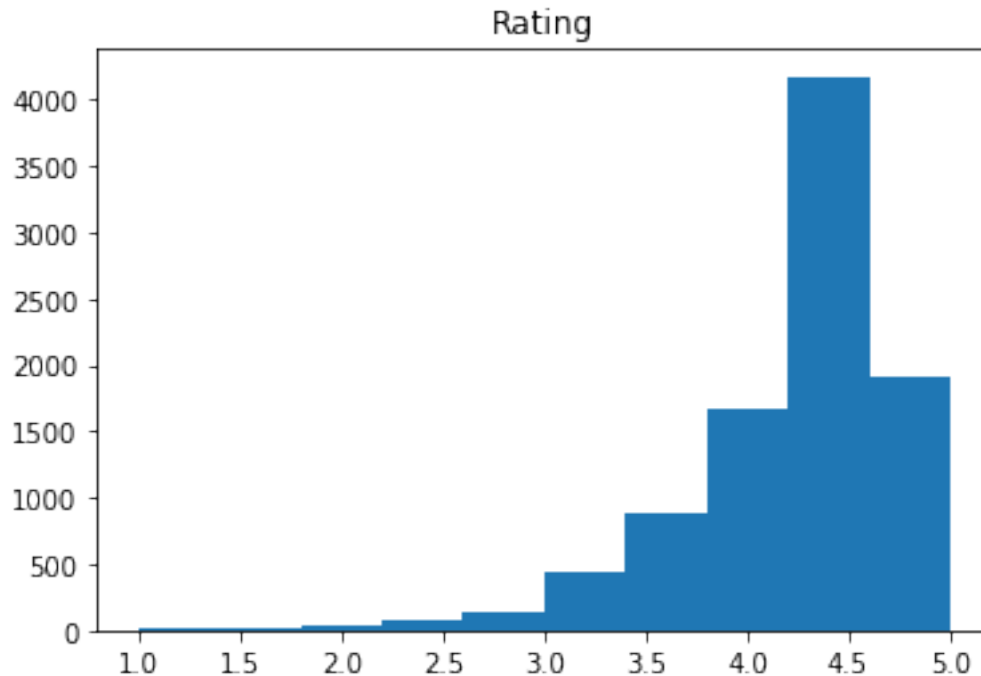
Since everything was fine with the Apple data, we now take only the Android apps that we are interested in.

```
In [9]: df_an1 = df_an[(df_an['Rating'].isnull() == False) & (df_an['Rating'] <= 5)]
```

And as a quick reality check, we plot the new Android data histogram.

```
In [10]: df_an1.hist(column = 'Rating', grid = False)
```

```
Out[10]: array([[<matplotlib.axes._subplots.AxesSubplot object at 0x000001C5E4809F28>]],
dtype=object)
```



Since this is supposed to be a near finished product, assume that I have magical foresight for the next bit of code to address that the Android data has multiple rows of the same information. We can see this below for when looking at the categories, 'GAME' and 'FAMILY'.

```
In [11]: game_bool = df_an1['Category'] == 'GAME'
df_an12 = df_an1[game_bool]
df_an12.sort_values(by = ['Installs', 'Rating'], ascending = False).head(10)
```

```
Out[11]:
```

	App	Category	Rating	Reviews	Size
1722	My Talking Tom	GAME	4.5	14891223	Varies with device
1908	My Talking Tom	GAME	4.5	14892469	Varies with device
1655	Candy Crush Saga	GAME	4.4	22426677	74M
1705	Candy Crush Saga	GAME	4.4	22428456	74M
1751	Candy Crush Saga	GAME	4.4	22428456	74M
1842	Candy Crush Saga	GAME	4.4	22429716	74M
1869	Candy Crush Saga	GAME	4.4	22430188	74M
1966	Candy Crush Saga	GAME	4.4	22430188	74M
1661	Temple Run 2	GAME	4.3	8118609	62M
1662	Pou	GAME	4.3	10485308	24M

	Installs	Type	Price	Content	Rating	Genres	Last Updated
1722	500,000,000+	Free	0	Everyone	Casual	July 19, 2018	
1908	500,000,000+	Free	0	Everyone	Casual	July 19, 2018	
1655	500,000,000+	Free	0	Everyone	Casual	July 5, 2018	
1705	500,000,000+	Free	0	Everyone	Casual	July 5, 2018	
1751	500,000,000+	Free	0	Everyone	Casual	July 5, 2018	

1842	500,000,000+	Free	0	Everyone	Casual	July 5, 2018
1869	500,000,000+	Free	0	Everyone	Casual	July 5, 2018
1966	500,000,000+	Free	0	Everyone	Casual	July 5, 2018
1661	500,000,000+	Free	0	Everyone	Action	July 5, 2018
1662	500,000,000+	Free	0	Everyone	Casual	May 25, 2018

	Current Ver	Android Ver
1722	4.8.0.132	4.1 and up
1908	4.8.0.132	4.1 and up
1655	1.129.0.2	4.1 and up
1705	1.129.0.2	4.1 and up
1751	1.129.0.2	4.1 and up
1842	1.129.0.2	4.1 and up
1869	1.129.0.2	4.1 and up
1966	1.129.0.2	4.1 and up
1661	1.49.1	4.0 and up
1662	1.4.77	4.0 and up

```
In [12]: fam_bool = df_an1['Category'] == 'FAMILY'
df_an13 = df_an1[fam_bool]
df_an13.sort_values(by = ['Installs', 'Rating'], ascending = False).head(10)
```

```
Out [12]:
```

	App	Category	Rating	\
3888	My Talking Tom	FAMILY	4.5	
3994	Candy Crush Saga	FAMILY	4.4	
3879	Pou	FAMILY	4.3	
3933	Hungry Hearts Diner: A Tale of Star-Crossed Souls	FAMILY	4.9	
3886	Find a Way: Addictive Puzzle	FAMILY	4.8	
7010	Pixyfy: coloring by number coloring book	FAMILY	4.8	
8995	German Listening	FAMILY	4.8	
2107	Ninja Dash - Ronin Jump RPG	FAMILY	4.7	
2141	Monica Toy TV	FAMILY	4.7	
2151	Toca Life: City	FAMILY	4.7	

	Reviews	Size	Installs	Type	Price	Content	Rating	\
3888	14885236	Varies with device	500,000,000+	Free	0		Everyone	
3994	22419455	74M	500,000,000+	Free	0		Everyone	
3879	10483141	24M	500,000,000+	Free	0		Everyone	
3933	46253	56M	500,000+	Free	0		Everyone 10+	
3886	39480	14M	500,000+	Free	0		Everyone	
7010	22290	13M	500,000+	Free	0		Everyone	
8995	18298	8.0M	500,000+	Free	0		Everyone	
2107	8508	54M	500,000+	Free	0		Everyone 10+	
2141	6188	49M	500,000+	Free	0		Everyone 10+	
2151	31100	24M	500,000+	Paid	\$3.99		Everyone	

	Genres	Last Updated	Current Ver	Android Ver
3888	Casual	July 19, 2018	4.8.0.132	4.1 and up

3994	Casual	July 5, 2018	1.129.0.2	4.1 and up
3879	Casual	May 25, 2018	1.4.77	4.0 and up
3933	Simulation	February 7, 2018	1.0.1	4.0 and up
3886	Puzzle	June 16, 2017	4.1.1	4.1 and up
7010	Puzzle	August 5, 2018	18.08.05	4.0 and up
8995	Education	July 23, 2018	2.94	2.3.3 and up
2107	Arcade;Action & Adventure	August 2, 2018	1.1.23	4.0.3 and up
2141	Entertainment;Music & Video	April 3, 2018	1.6.4	4.1 and up
2151	Education;Pretend Play	July 6, 2018	1.5-play	4.4 and up

As we can see from the above, there are multiple instances of Candy Crush Saga appearing in the category 'GAME', and it is worth noting that there is also a copy of Candy Crush Saga, with the same information, appearing in the category 'FAMILY' with a different index. So we are going to collapse all these rows down.

```
In [13]: df_an1.drop_duplicates(subset=['App'], keep = 'first', inplace = True)
```

```
C:\ProgramData\Anaconda3\lib\site-packages\ipykernel_launcher.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame
```

```
See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html
"""Entry point for launching an IPython kernel.
```

Once again a reality check to make sure that we did what we said we were going to do. It is also worth noting, that since we specified "keep = 'first'" the Candy Crush with index 3994 that appeared in 'FAMILY' is now gone because the first index of Candy Crush appeared in 'GAME'.

```
In [14]: df_an12 = df_an1[game_bool]
df_an12.sort_values(by = ['Installs', 'Rating'], ascending = False).head(10)
```

```
C:\ProgramData\Anaconda3\lib\site-packages\ipykernel_launcher.py:1: UserWarning: Boolean Series
"""Entry point for launching an IPython kernel.
```

```
Out[14]:
```

	App	Category	Rating	\
1722	My Talking Tom	GAME	4.5	
1655	Candy Crush Saga	GAME	4.4	
1661	Temple Run 2	GAME	4.3	
1662	Pou	GAME	4.3	
4975	Solitaire: Decked Out Ad Free	GAME	4.9	
1779	PLANK!	GAME	4.7	
5818	Zlax.io Zombs Luv Ax	GAME	4.6	
9607	PokerStars Play: Free Texas Holdem Poker Game	GAME	4.6	
3918	Law of Creation: A Playable Manga	GAME	4.5	
4777	Hero Fighter X	GAME	4.5	

	Reviews	Size	Installs	Type	Price	Content	Rating	\
1722	14891223	Varies with device	500,000,000+	Free	0		Everyone	

1655	22426677	74M	500,000,000+	Free	0	Everyone
1661	8118609	62M	500,000,000+	Free	0	Everyone
1662	10485308	24M	500,000,000+	Free	0	Everyone
4975	37302	35M	500,000+	Free	0	Everyone
1779	7196	38M	500,000+	Free	0	Everyone
5818	6344	32M	500,000+	Free	0	Everyone
9607	14989	78M	500,000+	Free	0	Teen
3918	13118	47M	500,000+	Free	0	Teen
4777	24210	45M	500,000+	Free	0	Teen

	Genres	Last Updated	Current Ver	Android Ver
1722	Casual	July 19, 2018	4.8.0.132	4.1 and up
1655	Casual	July 5, 2018	1.129.0.2	4.1 and up
1661	Action	July 5, 2018	1.49.1	4.0 and up
1662	Casual	May 25, 2018	1.4.77	4.0 and up
4975	Card	May 8, 2017	1.3.3	4.1 and up
1779	Arcade	July 17, 2018	1.0.2	4.3 and up
5818	Action	April 4, 2018	1.7	4.1 and up
9607	Card	July 6, 2018	1.9.2	4.0.3 and up
3918	Adventure	June 25, 2018	1.0.19	4.0.3 and up
4777	Action	November 13, 2015	1.091	2.2 and up

```
In [15]: df_an13 = df_an1[fam_bool]
df_an13.sort_values(by = ['Installs', 'Rating'], ascending = False).head(10)
```

```
C:\ProgramData\Anaconda3\lib\site-packages\ipykernel_launcher.py:1: UserWarning: Boolean Series
    ""Entry point for launching an IPython kernel.
```

```
Out[15]:
```

		App Category	Rating \
3933	Hungry Hearts Diner: A Tale of Star-Crossed Souls	FAMILY	4.9
3886	Find a Way: Addictive Puzzle	FAMILY	4.8
7010	Pixyfy: coloring by number coloring book	FAMILY	4.8
8995	German Listening	FAMILY	4.8
2107	Ninja Dash - Ronin Jump RPG	FAMILY	4.7
2141	Monica Toy TV	FAMILY	4.7
4404	Fantastic Chefs: Match 'n Cook	FAMILY	4.7
9814	Advanced calculator fx 991 es plus & 991 ms plus	FAMILY	4.7
2044	Garden Fruit Legend	FAMILY	4.6
2074	Fruit Cube Blast	FAMILY	4.6

	Reviews	Size	Installs	Type	Price	Content	Rating \
3933	46253	56M	500,000+	Free	0	Everyone	10+
3886	39480	14M	500,000+	Free	0	Everyone	
7010	22290	13M	500,000+	Free	0	Everyone	
8995	18298	8.0M	500,000+	Free	0	Everyone	
2107	8508	54M	500,000+	Free	0	Everyone	10+
2141	6188	49M	500,000+	Free	0	Everyone	10+

4404	8600	91M	500,000+	Free	0	Everyone
9814	7252	9.2M	500,000+	Free	0	Everyone
2044	4289	33M	500,000+	Free	0	Everyone
2074	9199	80M	500,000+	Free	0	Everyone

	Genres	Last Updated \
3933	Simulation	February 7, 2018
3886	Puzzle	June 16, 2017
7010	Puzzle	August 5, 2018
8995	Education	July 23, 2018
2107	Arcade;Action & Adventure	August 2, 2018
2141	Entertainment;Music & Video	April 3, 2018
4404	Puzzle	July 27, 2018
9814	Education	August 2, 2018
2044	Casual;Brain Games	July 31, 2018
2074	Puzzle;Brain Games	August 3, 2018

	Current Ver	Android Ver
3933	1.0.1	4.0 and up
3886	4.1.1	4.1 and up
7010	18.08.05	4.0 and up
8995	2.94	2.3.3 and up
2107	1.1.23	4.0.3 and up
2141	1.6.4	4.1 and up
4404	1.0.7	4.0.3 and up
9814	3.3.1-build-02-08-2018-19-release	4.0 and up
2044	3.5.3183	4.0.3 and up
2074	1.2.7	4.1 and up

Once again because of magical foresight, we will preemptively define the following function. The purpose of this function is to filter out any app that has more than 2 non-ascii characters in its name -- the 2 is subjective and could be whatever tolerance that we find acceptable.

In [16]: *# This function filters out apps with more than 3 nonascii characters in their name*

```
def in_english(string):
    non_ascii = 0

    for char in string:
        if ord(char) > 127:
            non_ascii += 1
    if non_ascii >= 3:
        return False
    else:
        return True
```

```
In [17]: english_bool_an = df_an1['App'].apply(in_english) == False
english_bool_ios = df_ios['track_name'].apply(in_english) == False
```

```
In [18]: df_an_tmp = df_an1[english_bool_an]
df_an_tmp
```

Out [18] :

	App	Category \
309	Truyen Vui Tý Quy	COMICS
710	Flame -	EDUCATION
1481	At home - rental ù real estate ù room finding ...	HOUSE_AND_HOME
1488	: Buying a house, selling a house, renting ...	HOUSE_AND_HOME
1592	Astrology - Min Thein Kha BayDin	LIFESTYLE
3750		NEWS_AND_MAGAZINES
3824	NAVITIMETimetable & Route Search in Japan T...	MAPS_AND_NAVIGATION
3960	MultiCraft Free Miner!	GAME
4193	H	ART_AND_DESIGN
4324	L.POINT - [, , , , ,	LIFESTYLE
4354	M3-X5-X6-M-3-8 RACNG 2018	GAME
5346	Al Quran Free - (Islam)	BOOKS_AND_REFERENCE
5698	AV	BOOKS_AND_REFERENCE
6006	BD Bangla TV	FAMILY
6201	CSCS BG ()	FAMILY
6406	BL	FAMILY
6415	SecondSecret BL	GAME
6417	BL	FAMILY
6418	BL	FAMILY
6420	BL	COMICS
6629	BQ-	GAME
6729	BS Calendar / Patro /	PRODUCTIVITY
7187	I Smart Remote MP3 CD Player	TOOLS
7290	Offline Jízdní ády CG Transit	MAPS_AND_NAVIGATION
7463	CK	LIFESTYLE
7559	Smart Ruler cm/inch measuring for homework!	TOOLS
7751	Conquerors	FAMILY
8394	DG / Digital Gram Panchayat	NEWS_AND_MAGAZINES
8456	Lep's World 3	GAME
8994	DW By dw-arab.com	NEWS_AND_MAGAZINES
9105	Météo Algérie DZ	WEATHER
9117	DZ	FAMILY
9200	flix	FAMILY
9211	RPG	FAMILY
9222	EC Dictionary	FAMILY
9248	ECE	LIFESTYLE
9308	NEWS_AND_MAGAZINES	
9317	EG SIM CARD (EGSIMCARD,)	BUSINESS
9828	Égalité et Réconciliation	NEWS_AND_MAGAZINES
10098	Ey Sey Storytime	FAMILY
10174	CNY Slots : Gong Xi Fa Cai	GAME
10264	My AEK - Official FC app	SPORTS
10719	Sona - Nær við allastaðni	LIFESTYLE
10763	FP	FAMILY

	Rating	Reviews	Size	Installs	Type	Price	\
309	4.5	144	4.7M	10,000+	Free	0	

710	4.6	56065		37M	1,000,000+	Free	0
1481	3.8	2496	Varies with device		500,000+	Free	0
1488	3.7	2248		15M	100,000+	Free	0
1592	4.7	2225		15M	100,000+	Free	0
3750	4.5	44274		8.0M	1,000,000+	Free	0
3824	4.4	50459	Varies with device		5,000,000+	Free	0
3960	4.3	1305050	Varies with device		50,000,000+	Free	0
4193	4.4	13		4.5M	1,000+	Free	0
4324	4.0	45224		49M	5,000,000+	Free	0
4354	4.4	596		44M	100,000+	Free	0
5346	4.7	1777		23M	50,000+	Free	0
5698	4.1	215		30M	10,000+	Free	0
6006	4.3	193		14M	10,000+	Free	0
6201	2.4	7		2.1M	100+	Paid	\$3.99
6406	4.2	3379		62M	100,000+	Free	0
6415	4.4	1563		40M	50,000+	Free	0
6417	4.2	1901		8.2M	100,000+	Free	0
6418	4.7	6073		8.5M	100,000+	Free	0
6420	4.4	190		34M	10,000+	Free	0
6629	4.7	191		7.2M	5,000+	Free	0
6729	4.2	218	Varies with device		50,000+	Free	0
7187	3.2	27		2.7M	5,000+	Free	0
7290	4.6	7314		7.0M	100,000+	Free	0
7463	4.0	294		153k	10,000+	Free	0
7559	4.0	19		3.2M	10,000+	Free	0
7751	4.5	108130		40M	5,000,000+	Free	0
8394	4.5	53		5.9M	10,000+	Free	0
8456	4.3	771001		52M	50,000,000+	Free	0
8994	4.7	11		4.4M	1,000+	Free	0
9105	4.1	1238		4.7M	100,000+	Free	0
9117	4.2	238		6.8M	10,000+	Free	0
9200	4.2	1111		7.3M	100,000+	Free	0
9211	3.4	159		3.1M	1,000+	Paid	\$10.99
9222	4.3	55408	Varies with device		1,000,000+	Free	0
9248	4.0	576		7.4M	50,000+	Free	0
9308	4.8	311		14M	5,000+	Free	0
9317	2.9	322		1.5M	50,000+	Free	0
9828	4.9	58		3.8M	500+	Paid	\$2.99
10098	4.7	1327		44M	10,000+	Free	0
10174	3.6	33		71M	5,000+	Free	0
10264	4.8	3346	Varies with device		50,000+	Free	0
10719	4.2	31		25M	1,000+	Free	0
10763	4.5	922		552k	50,000+	Free	0

	Content Rating	Genres	Last Updated \
309	Everyone	Comics	July 19, 2018
710	Everyone	Education	July 26, 2018
1481	Everyone	House & Home	July 5, 2018

1488	Everyone	House & Home	August 3, 2018
1592	Everyone	Lifestyle	July 26, 2018
3750	Everyone	News & Magazines	August 6, 2018
3824	Everyone	Maps & Navigation	July 26, 2018
3960	Everyone 10+	Adventure	July 29, 2018
4193	Everyone	Art & Design	March 27, 2018
4324	Everyone	Lifestyle	August 1, 2018
4354	Everyone	Arcade	February 12, 2017
5346	Everyone	Books & Reference	February 15, 2015
5698	Teen	Books & Reference	March 6, 2018
6006	Everyone	Entertainment	April 29, 2017
6201	Everyone	Education	March 23, 2017
6406	Mature 17+	Simulation	March 23, 2017
6415	Teen	Adventure	June 29, 2017
6417	Mature 17+	Simulation	July 7, 2016
6418	Mature 17+	Simulation	February 25, 2018
6420	Everyone	Comics	November 17, 2017
6629	Everyone	Trivia	July 31, 2018
6729	Everyone	Productivity	July 15, 2018
7187	Everyone	Tools	January 22, 2018
7290	Everyone	Maps & Navigation	September 7, 2017
7463	Everyone	Lifestyle	July 3, 2013
7559	Everyone	Tools	October 21, 2017
7751	Teen	Strategy	August 3, 2018
8394	Teen	News & Magazines	June 19, 2018
8456	Everyone 10+	Action	June 11, 2018
8994	Everyone	News & Magazines	March 26, 2017
9105	Everyone	Weather	December 17, 2014
9117	Everyone	Education	November 18, 2016
9200	Everyone	Entertainment	June 5, 2018
9211	Teen	Role Playing	April 26, 2018
9222	Everyone	Education	January 8, 2018
9248	Everyone	Lifestyle	June 7, 2018
9308	Everyone 10+	News & Magazines	July 15, 2018
9317	Everyone	Business	March 28, 2017
9828	Everyone	News & Magazines	May 26, 2018
10098	Everyone	Education;Education	May 18, 2017
10174	Teen	Casino	June 6, 2017
10264	Everyone	Sports	January 18, 2018
10719	Everyone	Lifestyle	August 2, 2018
10763	Everyone	Entertainment	September 20, 2014

	Current Ver	Android Ver
309	3.0	4.0.3 and up
710	3.3	4.1 and up
1481	Varies with device	Varies with device
1488	v3.1.1	4.0 and up
1592	4.2.1	4.0.3 and up

3750	4.0.6	4.4 and up
3824	Varies with device	Varies with device
3960	1.1.11.11	4.1 and up
4193	2.0	4.0.3 and up
4324	6.5.1	4.1 and up
4354	1,01	4.0 and up
5346	1.1	2.2 and up
5698	1.2	4.0 and up
6006	1.3	4.0 and up
6201	1.0	4.0.3 and up
6406	1.6.3	2.3.3 and up
6415	1.3.6	4.0 and up
6417	1.3.0	2.3.3 and up
6418	4.2.2	2.3 and up
6420	3.2.1	3.0 and up
6629	4.1.2	4.1 and up
6729	Varies with device	Varies with device
7187	1.2	4.0 and up
7290	3.1.4	4.0 and up
7463	1.0.12	2.1 and up
7559	1.0	4.2 and up
7751	1.8.0	4.0 and up
8394	1.7	4.0 and up
8456	2.0.1.3	4.1 and up
8994	1.0.0	4.1 and up
9105	1.1.6	2.3 and up
9117	6.0.0.0	4.1 and up
9200	3.6.1	4.1 and up
9211	2.12	2.3 and up
9222	Varies with device	Varies with device
9248	1.0.2	4.0 and up
9308	2.9.19a	4.2 and up
9317	2.13	2.2 and up
9828	1.1.1	5.0 and up
10098	1.2.2	4.0.3 and up
10174	1.2	4.0.3 and up
10264	2.0.2	Varies with device
10719	1.6.3	4.4 and up
10763	1.1	1.6 and up

```
In [19]: df_an1e = df_an1[english_bool_an == False]
df_an1e.head(10)
```

```
Out[19]:
```

	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
5	Paper flowers instructions	ART_AND_DESIGN	4.4
6	Smoke Effect Photo Maker - Smoke Editor	ART_AND_DESIGN	3.8
7	Infinite Painter	ART_AND_DESIGN	4.1
8	Garden Coloring Book	ART_AND_DESIGN	4.4
9	Kids Paint Free - Drawing Fun	ART_AND_DESIGN	4.7

	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	
5	167	5.6M	50,000+	Free	0	Everyone	
6	178	19M	50,000+	Free	0	Everyone	
7	36815	29M	1,000,000+	Free	0	Everyone	
8	13791	33M	1,000,000+	Free	0	Everyone	
9	121	3.1M	10,000+	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
5	Art & Design	March 26, 2017	1.0
6	Art & Design	April 26, 2018	1.1
7	Art & Design	June 14, 2018	6.1.61.1
8	Art & Design	September 20, 2017	2.9.2
9	Art & Design;Creativity	July 3, 2018	2.8

	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
5	2.3 and up
6	4.0.3 and up
7	4.2 and up
8	3.0 and up
9	4.0.3 and up

Magical foresight!

```
In [20]: df_ios = df_ios[english_bool_ios == False]
df_ios.head(10)
```

```
Out[20]: Unnamed: 0      id      track_name \
```


0	1	281656475	PAC-MAN Premium
1	2	281796108	Evernote - stay organized
2	3	281940292	WeatherBug - Local Weather, Radar, Maps, Alerts
3	4	282614216	eBay: Best App to Buy, Sell, Save! Online Shop...
4	5	282935706	Bible
5	6	283619399	Shanghai Mahjong
6	7	283646709	PayPal - Send and request money safely
7	8	284035177	Pandora - Music & Radio
8	9	284666222	PCalc - The Best Calculator
9	10	284736660	Ms. PAC-MAN

	size_bytes	currency	price	rating_count_tot	rating_count_ver	\
0	100788224	USD	3.99	21292	26	
1	158578688	USD	0.00	161065	26	
2	100524032	USD	0.00	188583	2822	
3	128512000	USD	0.00	262241	649	
4	92774400	USD	0.00	985920	5320	
5	10485713	USD	0.99	8253	5516	
6	227795968	USD	0.00	119487	879	
7	130242560	USD	0.00	1126879	3594	
8	49250304	USD	9.99	1117	4	
9	70023168	USD	3.99	7885	40	

	user_rating	user_rating_ver	ver	cont_rating	prime_genre	\
0	4.0	4.5	6.3.5	4+	Games	
1	4.0	3.5	8.2.2	4+	Productivity	
2	3.5	4.5	5.0.0	4+	Weather	
3	4.0	4.5	5.10.0	12+	Shopping	
4	4.5	5.0	7.5.1	4+	Reference	
5	4.0	4.0	1.8	4+	Games	
6	4.0	4.5	6.12.0	4+	Finance	
7	4.0	4.5	8.4.1	12+	Music	
8	4.5	5.0	3.6.6	4+	Utilities	
9	4.0	4.0	4.0.4	4+	Games	

	sup_devices.num	ipadSc_urls.num	lang.num	vpp_lic
0	38	5	10	1
1	37	5	23	1
2	37	5	3	1
3	37	5	9	1
4	37	5	45	1
5	47	5	1	1
6	37	0	19	1
7	37	4	1	1
8	37	5	1	1
9	38	0	10	1

Now that we have a list of valid apps, in latin characters. We return to our problem, we wanted

to make a free app. So let's find the free apps.

```
In [21]: # Technically we should check that we get the same answer when we use 'Type' == 'Free'
# In a previous scratch work this was checked.
# It is also worth noting that since Type or Price gives the same answer we could drop
df_anief = df_anie[df_anie['Type'] == 'Free']
df_anief.head(10)
```

```
Out [21]:
```

	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
5	Paper flowers instructions	ART_AND_DESIGN	4.4
6	Smoke Effect Photo Maker - Smoke Editor	ART_AND_DESIGN	3.8
7	Infinite Painter	ART_AND_DESIGN	4.1
8	Garden Coloring Book	ART_AND_DESIGN	4.4
9	Kids Paint Free - Drawing Fun	ART_AND_DESIGN	4.7

	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	
5	167	5.6M	50,000+	Free	0	Everyone	
6	178	19M	50,000+	Free	0	Everyone	
7	36815	29M	1,000,000+	Free	0	Everyone	
8	13791	33M	1,000,000+	Free	0	Everyone	
9	121	3.1M	10,000+	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
5	Art & Design	March 26, 2017	1.0
6	Art & Design	April 26, 2018	1.1
7	Art & Design	June 14, 2018	6.1.61.1
8	Art & Design	September 20, 2017	2.9.2
9	Art & Design;Creativity	July 3, 2018	2.8

	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
5    2.3 and up
6    4.0.3 and up
7    4.2 and up
8    3.0 and up
9    4.0.3 and up

```

```

In [22]: # The USD column can be dropped as we are working with free apps.
df_iosef = df_iose[df_iose['price'] == 0.00]
df_iosef.head(10)

```

```

Out[22]:      Unnamed: 0      id      track_name \
1          2  281796108      Evernote - stay organized
2          3  281940292      WeatherBug - Local Weather, Radar, Maps, Alerts
3          4  282614216      eBay: Best App to Buy, Sell, Save! Online Shop...
4          5  282935706      Bible
6          7  283646709      PayPal - Send and request money safely
7          8  284035177      Pandora - Music & Radio
12         13  284815942      Google Search made just for mobile
13         14  284847138      Bank of America - Mobile Banking
15         16  284876795      TripAdvisor Hotels Flights Restaurants
16         17  284882215      Facebook

```

```

      size_bytes  currency  price  rating_count_tot  rating_count_ver \
1    158578688      USD    0.0      161065          26
2    100524032      USD    0.0      188583         2822
3    128512000      USD    0.0      262241          649
4     92774400      USD    0.0      985920         5320
6    227795968      USD    0.0      119487          879
7    130242560      USD    0.0     1126879         3594
12   179979264      USD    0.0      479440          203
13   160925696      USD    0.0     119773         2336
15   207907840      USD    0.0      56194           87
16   389879808      USD    0.0     2974676         212

```

```

      user_rating  user_rating_ver      ver  cont_rating      prime_genre \
1          4.0          3.5    8.2.2          4+      Productivity
2          3.5          4.5    5.0.0          4+          Weather
3          4.0          4.5    5.10.0         12+      Shopping
4          4.5          5.0    7.5.1          4+      Reference
6          4.0          4.5    6.12.0          4+      Finance
7          4.0          4.5    8.4.1         12+          Music
12         3.5          4.0      27.0         17+      Utilities
13         3.5          4.5    7.3.8          4+      Finance
15         4.0          3.5     21.1          4+      Travel
16         3.5          3.5     95.0          4+  Social Networking

```

	sup_devices.num	ipadSc_urls.num	lang.num	vpp_lic
1	37	5	23	1
2	37	5	3	1
3	37	5	9	1
4	37	5	45	1
6	37	0	19	1
7	37	4	1	1
12	37	4	33	1
13	37	0	2	1
15	37	1	26	1
16	37	1	29	1

Now that we have isolated the free apps, it is now time to start looking at how they are distributed. If we start looking at the columns of the Android data to see what kind of app to make, we see that 'Category', 'Genres' and 'Content Rating' are probably the most helpful -- Content Rating probably won't be super helpful based on personal experience of looking at apps, but I wanted to see if I could do the following using a for loop, and looping over 2 things seems silly.

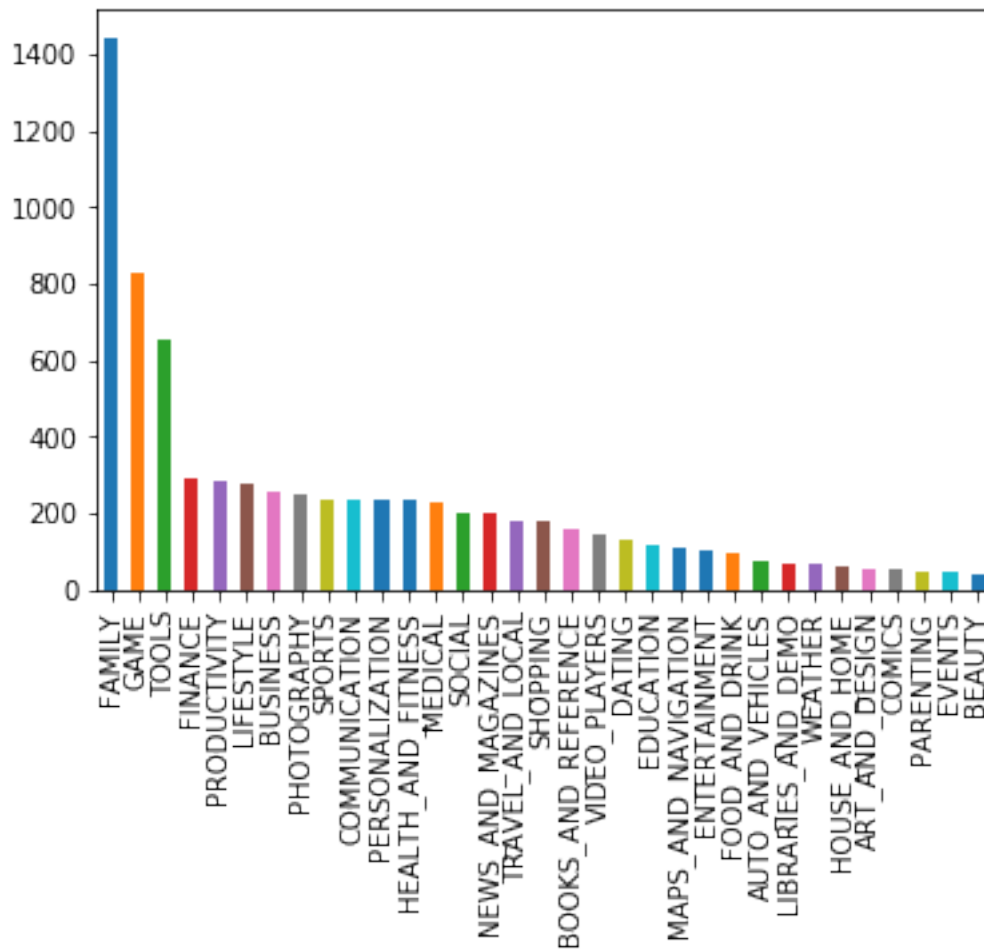
```
In [23]: an_list = ['Category', 'Genres', 'Content Rating']
```

```
for item in an_list:
    print(df_an1ef[item].value_counts())
    print('\n')
    df_an1ef[item].value_counts().plot.bar()
    plt.show()
```

FAMILY	1444
GAME	830
TOOLS	653
FINANCE	289
PRODUCTIVITY	282
LIFESTYLE	278
BUSINESS	253
PHOTOGRAPHY	248
SPORTS	237
COMMUNICATION	234
PERSONALIZATION	233
HEALTH_AND_FITNESS	233
MEDICAL	227
SOCIAL	201
NEWS_AND_MAGAZINES	198
TRAVEL_AND_LOCAL	179
SHOPPING	178
BOOKS_AND_REFERENCE	159
VIDEO_PLAYERS	144
DATING	131
EDUCATION	113
MAPS_AND_NAVIGATION	111

ENTERTAINMENT	100
FOOD_AND_DRINK	92
AUTO_AND_VEHICLES	72
LIBRARIES_AND_DEMO	64
WEATHER	64
HOUSE_AND_HOME	60
ART_AND_DESIGN	57
COMICS	52
PARENTING	48
EVENTS	45
BEAUTY	42

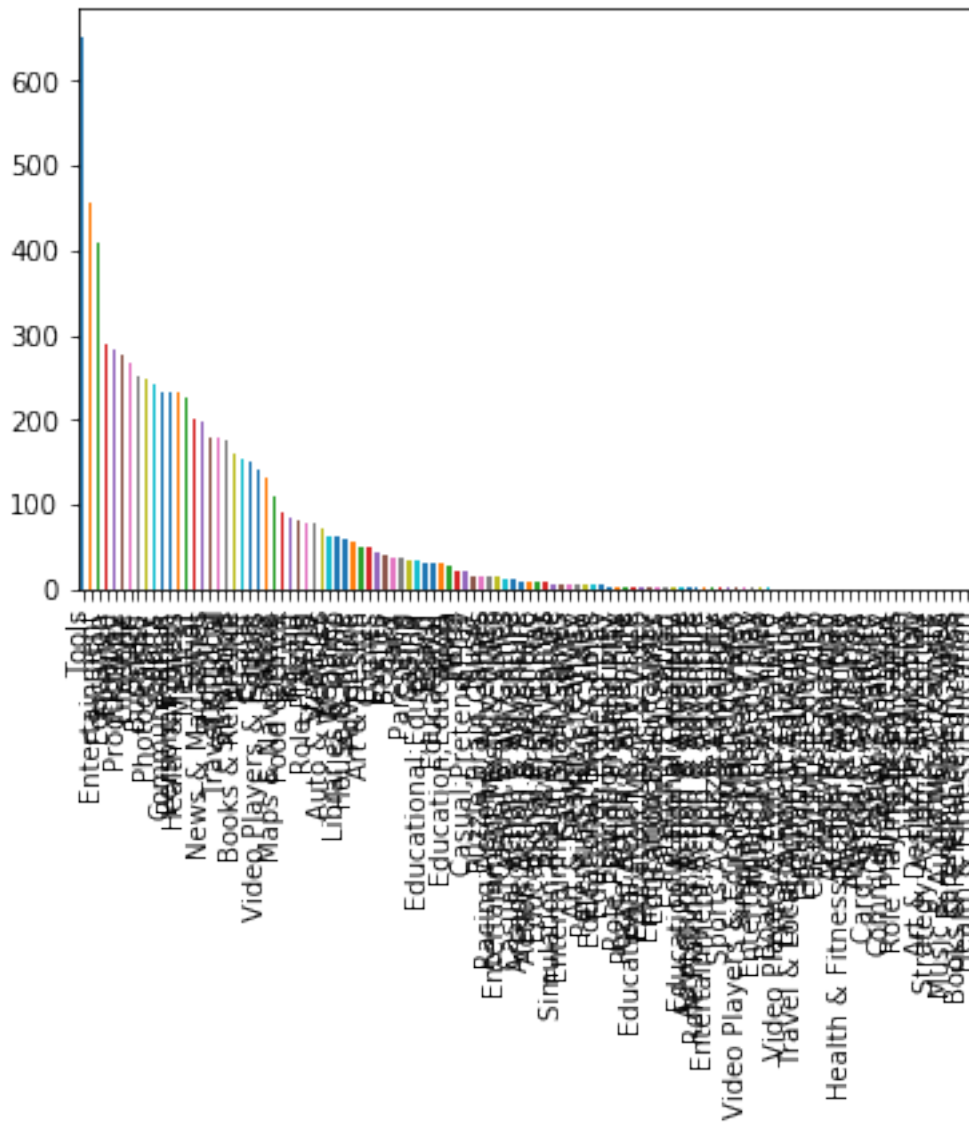
Name: Category, dtype: int64



Tools	652
Entertainment	455

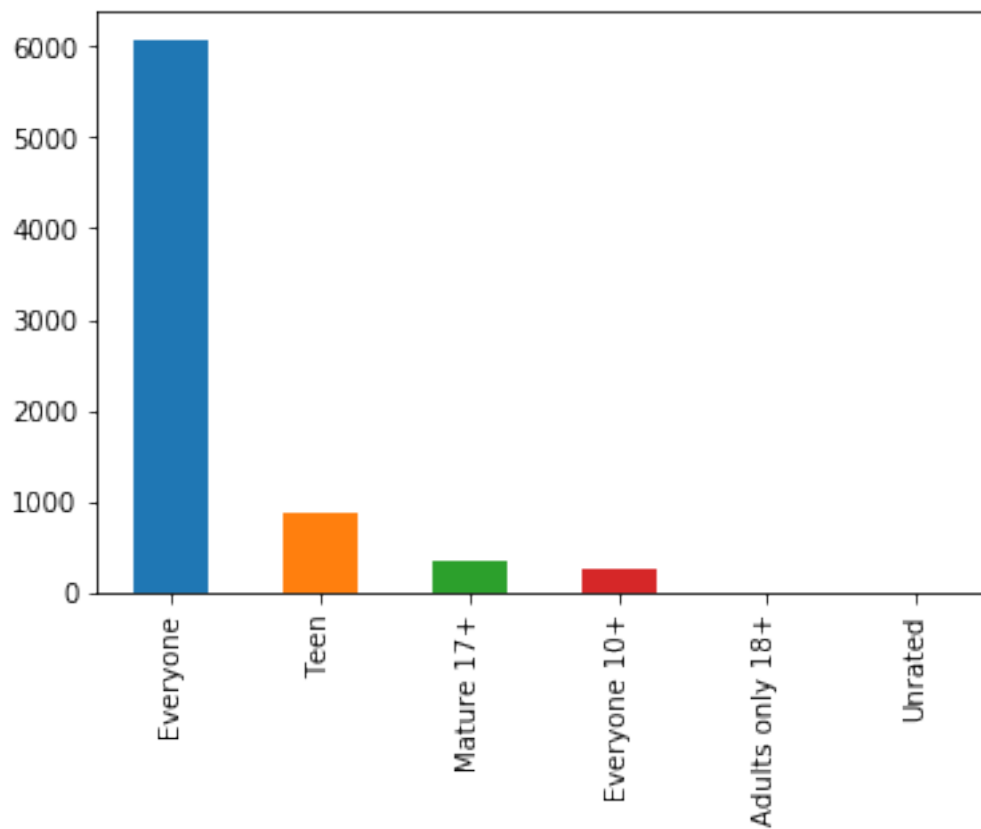
Education	408
Finance	289
Productivity	282
Lifestyle	277
Action	267
Business	253
Photography	248
Sports	243
Communication	234
Personalization	233
Health & Fitness	233
Medical	227
Social	201
News & Magazines	198
Shopping	178
Travel & Local	178
Simulation	175
Books & Reference	159
Arcade	153
Casual	150
Video Players & Editors	143
Dating	131
Maps & Navigation	111
Food & Drink	92
Racing	84
Puzzle	83
Role Playing	80
Strategy	77
...	
Video Players & Editors;Music & Video	2
Simulation;Pretend Play	2
Entertainment;Pretend Play	2
Music;Music & Video	2
Board;Action & Adventure	2
Video Players & Editors;Creativity	1
Entertainment;Education	1
Travel & Local;Action & Adventure	1
Simulation;Education	1
Lifestyle;Pretend Play	1
Casual;Music & Video	1
Lifestyle;Education	1
Arcade;Pretend Play	1
Health & Fitness;Action & Adventure	1
Racing;Pretend Play	1
Adventure;Education	1
Card;Action & Adventure	1
Strategy;Creativity	1
Communication;Creativity	1

Strategy;Education	1
Role Playing;Brain Games	1
Tools;Education	1
Puzzle;Education	1
Art & Design;Pretend Play	1
Strategy;Action & Adventure	1
Comics;Creativity	1
Music & Audio;Music & Video	1
Parenting;Brain Games	1
Books & Reference;Education	1
Health & Fitness;Education	1
Name: Genres, Length: 112, dtype: int64	



Everyone	6075
Teen	862
Mature 17+	339
Everyone 10+	271
Adults only 18+	3
Unrated	1

Name: Content Rating, dtype: int64



As we can see, Category seems to be dominated by the tags 'Family', 'Game' and 'Tools'. Genres is a complete mess and will need to be dealt with, and as stated above Content Rating wasn't super helpful. To deal with the problem of the Genres column, we note that it is largely due to things being double tagged. To deal with the double tag, we will simply just take the first tag. So something tagged "Adventure;Education" will be counted as "Adventure". Checking the list above does show that the tags are not in alphabetical order (Puzzle;Education exists) so this seems like a reasonable split.


```
In [24]: df_an1ef["Genres_split"] = df_an1ef["Genres"]
         df_an1ef["Genres_split"] = df_an1ef["Genres_split"].str.split(";", n = 1, expand = Tr
```

C:\ProgramData\Anaconda3\lib\site-packages\ipykernel_launcher.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/indexing.html>.
"""Entry point for launching an IPython kernel.

C:\ProgramData\Anaconda3\lib\site-packages\ipykernel_launcher.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

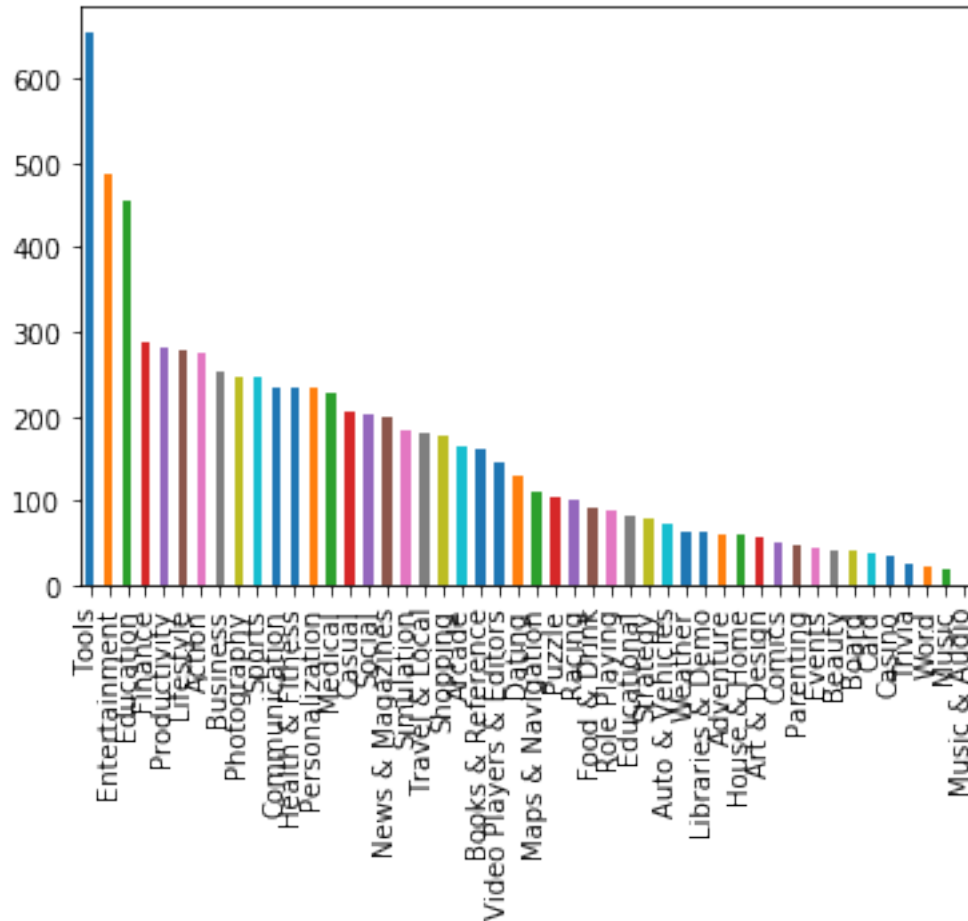
See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/indexing.html>.

```
In [25]: print(df_an1ef['Genres_split'].value_counts())
         print('\n')
         df_an1ef['Genres_split'].value_counts().plot.bar()
         plt.show()
```

Tools	653
Entertainment	486
Education	456
Finance	289
Productivity	282
Lifestyle	279
Action	276
Business	253
Photography	248
Sports	245
Communication	235
Health & Fitness	235
Personalization	233
Medical	227
Casual	204
Social	201
News & Magazines	198
Simulation	185
Travel & Local	179
Shopping	178
Arcade	164
Books & Reference	160
Video Players & Editors	146
Dating	131
Maps & Navigation	111
Puzzle	104

Racing	100
Food & Drink	92
Role Playing	88
Educational	83
Strategy	80
Auto & Vehicles	72
Weather	64
Libraries & Demo	64
Adventure	61
House & Home	60
Art & Design	58
Comics	52
Parenting	48
Events	45
Beauty	42
Board	41
Card	39
Casino	35
Trivia	27
Word	22
Music	19
Music & Audio	1

Name: Genres_split, dtype: int64



Cluttered but much more viewable. We can see that 'Tools', 'Entertainment' and 'Education' are the clear leaders with this lens.

We saw that the Category 'Family' has 1444 counted and in second is 'Game' with 830. So what is going on with this? We already partially addressed this, with apps like Candy Crush appearing as multiple rows, but maybe we can glean something else if we look at the genres that live in the 'Family' Category.

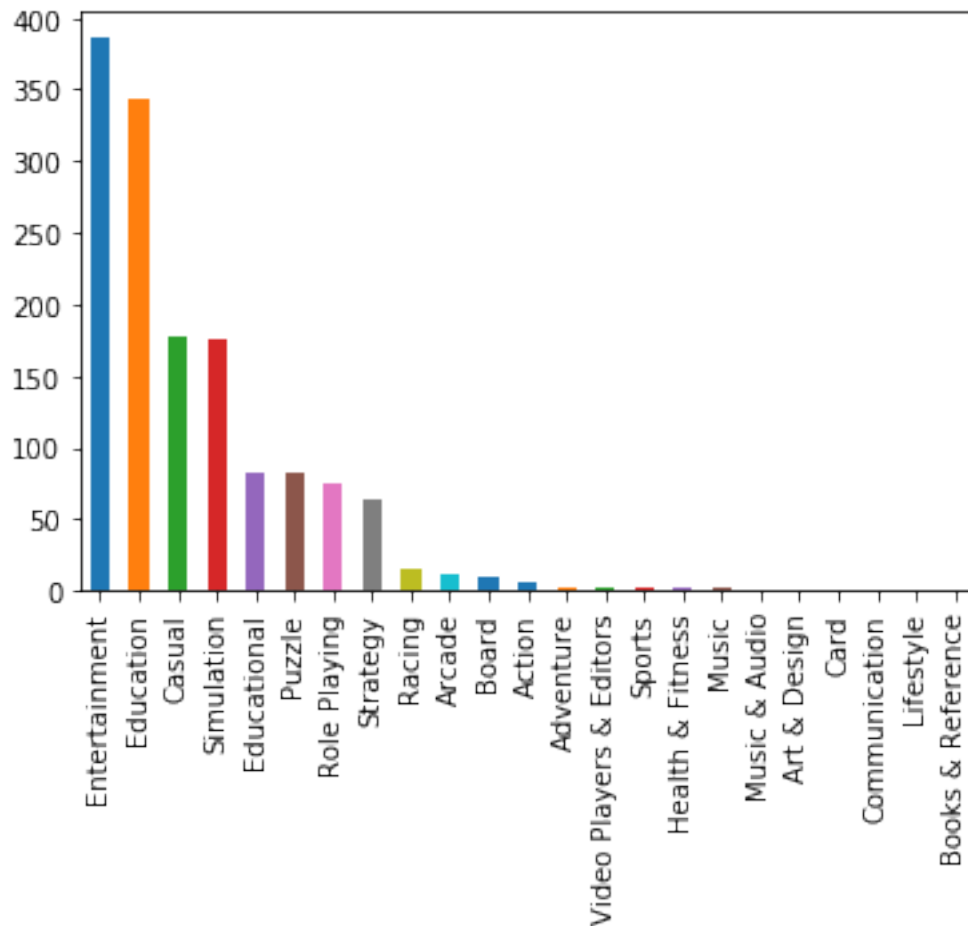
```
In [26]: an_fam = df_an1ef[df_an1ef['Category'] == 'FAMILY']

print(an_fam['Genres_split'].value_counts())
print('\n')
an_fam['Genres_split'].value_counts().plot.bar()
plt.show()
```

Entertainment	386
Education	343
Casual	177
Simulation	175
Educational	83

Puzzle	82
Role Playing	74
Strategy	64
Racing	16
Arcade	11
Board	10
Action	6
Adventure	3
Video Players & Editors	2
Sports	2
Health & Fitness	2
Music	2
Music & Audio	1
Art & Design	1
Card	1
Communication	1
Lifestyle	1
Books & Reference	1

Name: Genres_split, dtype: int64



As kind of expected, it appears to be that the apps in the 'Family' Category are split between Games and Education, possibly educational games. But this double indexing seems to be common and should be taken into account when making an app. Adding the tag 'Family' might be necessary to make sure that your app is seen.

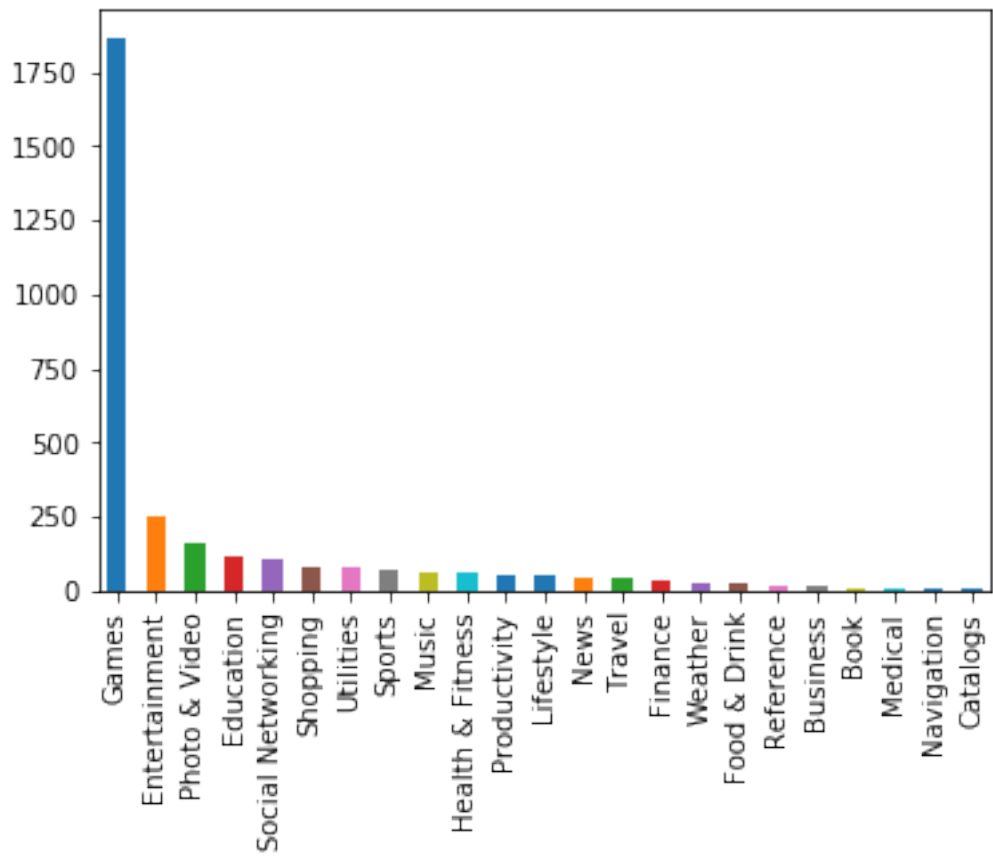
All of the talk so far has been about the Android store. Let's turn our attention to the Apple store. The only real descriptive column we have is 'prime_genre', so let's go with that -- no for loop this time.

```
In [27]: print(df_iosef['prime_genre'].value_counts())
print('\n')
df_iosef['prime_genre'].value_counts().plot.bar()
plt.show()
```

Games	1866
Entertainment	251
Photo & Video	160
Education	118
Social Networking	106
Shopping	83

Utilities	79
Sports	69
Music	66
Health & Fitness	65
Productivity	56
Lifestyle	50
News	43
Travel	40
Finance	35
Weather	28
Food & Drink	26
Reference	17
Business	17
Book	12
Medical	6
Navigation	6
Catalogs	4

Name: prime_genre, dtype: int64



As we can see almost all free apps on the Apple store are games. So if we want to make one app and send it to both stores, we should probably make a game of some kind. However, there are the funding considerations. A lot of games tend to be free but opt for the so called 'Freemium' model -- more commonly known as 'pay to not wait'.

Let's suppose that we have no real interest in making a game. As Educational apps tend to be popular on the Android store, we could make one of those, but we could run into the problem of lack of specialized knowledge. For example, second language apps, like Duolingo, tend to be quite popular but if we don't have people who can speak another language we are out of luck.

Finally we since 'Tools' is fairly popular on the Android store, maybe we could do something there (see below). But looking at the apps, it looks like a vast majority of the apps are related to either the telecom companies or the company the made the phone. Also apps like "Google Translate" fall into the category. While I tend to be optimistic and ambitious it might be hard to beat out Google. However, calculators also fall into this category. While it seem like not much can be done with calculators at this point, we could attempt to overthrow Texas Instruments and aim to make a graphing calculator app. We could have banner ads and make it so high school students no longer have to buy a graphing calculator for school. This also assume that the schools are trusting enough of their students to allow them to use their phones in class...

```
In [28]: df_an1ef[df_an1ef['Category'] == 'TOOLS'].head(25)
```

```
Out[28]:
```

	App	Category	Rating	\
3233	Moto File Manager	TOOLS	4.1	
3234	Google	TOOLS	4.4	
3235	Google Translate	TOOLS	4.4	
3236	Moto Display	TOOLS	4.2	
3237	Motorola Alert	TOOLS	4.2	
3238	Motorola Assist	TOOLS	4.1	
3239	Cache Cleaner-DU Speed Booster (booster & clea...	TOOLS	4.5	
3240	Moto Suggestions	TOOLS	4.6	
3241	Moto Voice	TOOLS	4.1	
3243	Device Help	TOOLS	4.2	
3244	Account Manager	TOOLS	4.1	
3245	myMetro	TOOLS	4.0	
3246	File Manager	TOOLS	4.6	
3247	My Telcel	TOOLS	3.1	
3248	Calculator - free calculator, multi calculator...	TOOLS	4.6	
3249	ASUS Sound Recorder	TOOLS	4.5	
3250	iWnn IME for Nexus	TOOLS	3.2	
3251	Samsung Max - Data Savings & Privacy Protection	TOOLS	4.3	
3253	ZenUI Help	TOOLS	4.6	
3254	Calculator - free calculator ,multi calculator...	TOOLS	4.4	
3255	SHAREit - Transfer & Share	TOOLS	4.6	
3256	ZenUI Keyboard Emoji, Theme	TOOLS	4.5	
3257	Files Go by Google: Free up space on your phone	TOOLS	4.6	
3259	Nokia mobile support	TOOLS	4.4	
3260	File Manager -- Take Command of Your Files Easily	TOOLS	4.2	

	Reviews	Size	Installs	Type	Price	Content	Rating	\
3233	38655	5.9M	10,000,000+	Free	0		Everyone	
3234	8033493	Varies with device	1,000,000,000+	Free	0		Everyone	
3235	5745093	Varies with device	500,000,000+	Free	0		Everyone	
3236	18239	Varies with device	10,000,000+	Free	0		Everyone	
3237	24199	3.9M	50,000,000+	Free	0		Everyone	
3238	37333	Varies with device	50,000,000+	Free	0		Everyone	
3239	12759663	15M	100,000,000+	Free	0		Everyone	
3240	308	4.3M	1,000,000+	Free	0		Everyone	
3241	33216	Varies with device	10,000,000+	Free	0		Everyone	
3243	28860	Varies with device	100,000,000+	Free	0		Everyone	
3244	76604	Varies with device	100,000,000+	Free	0		Everyone	
3245	26189	5.3M	10,000,000+	Free	0		Everyone	
3246	739329	Varies with device	50,000,000+	Free	0		Everyone	
3247	45838	16M	50,000,000+	Free	0		Everyone	
3248	25592	7.5M	10,000,000+	Free	0		Everyone	
3249	34126	Varies with device	10,000,000+	Free	0		Everyone	
3250	2394	Varies with device	5,000,000+	Free	0		Everyone	
3251	330468	Varies with device	10,000,000+	Free	0		Everyone	
3253	136874	Varies with device	10,000,000+	Free	0		Everyone	
3254	1236	5.8M	100,000+	Free	0		Everyone	
3255	7790693	17M	500,000,000+	Free	0		Everyone	
3256	537554	Varies with device	10,000,000+	Free	0		Everyone	
3257	315585	8.5M	10,000,000+	Free	0		Everyone	
3259	12215	9.1M	5,000,000+	Free	0		Everyone	
3260	127223	7.6M	10,000,000+	Free	0		Everyone	

	Genres	Last Updated	Current Ver	Android Ver	\
3233	Tools	February 1, 2018	v3.7.93	5.0 and up	
3234	Tools	August 3, 2018	Varies with device	Varies with device	
3235	Tools	August 4, 2018	Varies with device	Varies with device	
3236	Tools	August 6, 2018	Varies with device	Varies with device	
3237	Tools	November 21, 2014	1.02.53	4.4 and up	
3238	Tools	January 17, 2016	Varies with device	Varies with device	
3239	Tools	July 25, 2018	3.1.2	4.0 and up	
3240	Tools	June 8, 2018	0.2.32	8.0 and up	
3241	Tools	June 5, 2018	Varies with device	Varies with device	
3243	Tools	July 18, 2018	Varies with device	Varies with device	
3244	Tools	July 27, 2016	Varies with device	Varies with device	
3245	Tools	June 27, 2018	MyMetro_HTML_2.0_430029	4.0 and up	
3246	Tools	April 17, 2018	Varies with device	Varies with device	
3247	Tools	July 25, 2018	7.6.1	4.1 and up	
3248	Tools	July 31, 2018	v8.0.1.8.0629.1	4.2 and up	
3249	Tools	July 13, 2018	Varies with device	Varies with device	
3250	Tools	November 13, 2014	Varies with device	Varies with device	
3251	Tools	May 29, 2018	Varies with device	Varies with device	
3253	Tools	June 25, 2018	Varies with device	Varies with device	
3254	Tools	July 6, 2018	8.1.6	4.2 and up	

3255	Tools	July 30, 2018	4.5.28_ww	4.1 and up
3256	Tools	August 1, 2018	Varies with device	Varies with device
3257	Tools	July 18, 2018	1.0.204375696	5.0 and up
3259	Tools	August 1, 2018	1.7.4	7.0 and up
3260	Tools	August 3, 2018	v7.0.7.1.0625.1_06_0629	5.0 and up

	Genres_split
3233	Tools
3234	Tools
3235	Tools
3236	Tools
3237	Tools
3238	Tools
3239	Tools
3240	Tools
3241	Tools
3243	Tools
3244	Tools
3245	Tools
3246	Tools
3247	Tools
3248	Tools
3249	Tools
3250	Tools
3251	Tools
3253	Tools
3254	Tools
3255	Tools
3256	Tools
3257	Tools
3259	Tools
3260	Tools

We could also do some more work involving the rating. It might be the case the the majority of the Game apps are just terrible. But without some extra information, like retention or if the game has timers it would be a dice roll. The example I have in mind is someone who plays a game hits a timer, then switches to another game hits another timer, etc.

We could also search for apps by specific substrings, say apps in the 'Tools' category that screen out robocalls. Although it probably doesn't fit our funding model of having in app ads.

In []: