'Project': **Profitable App Profiles for the App Store and Google Play Markets**

About: This project aims to analyze data about mobile apps available on Google Play and the App Store.

Goal: To understand what type of apps are likely to attract more users on Google Play and the App Store.

OPENING DATASET AND PRINTING HEADER Dataset of Google Play Android apps and iOS apps dataset from the App Store is downloaded from public forum Kaggle and imported here. Links to datasets:

Link iOS apps dataset

Link Google Play Android apps dataset

In [1]:
```
## Dataset AppleStore.csv

opened_file = open(r'C:\Users\anuja\Downloads\AppleStore.csv', encoding
='utf8')
from csv import reader
read_file = reader(opened_file)
ios = list(read_file)
ios_header = ios[0]
ios = ios[1:]

print(ios_header)
```

```
['id', 'track_name', 'size_bytes', 'currency', 'price', 'rating_count_t
ot', 'rating_count_ver', 'user_rating', 'user_rating_ver', 'ver', 'cont
_rating', 'prime_genre', 'sup_devices.num', 'ipadSc_urls.num', 'lang.nu
m', 'vpp_lic']
```

In [2]:
```
##Dataset googleplaystore.csv
```

```
opened_file = open(r'C:\Users\anuja\Downloads\googleplaystore.csv', enc
oding='utf8')
from csv import reader
read_file = reader(opened_file)
android = list(read_file)
android_header = android[0]
android = android[1:]

print(android_header)
```

```
['App', 'Category', 'Rating', 'Reviews', 'Size', 'Installs', 'Type', 'P
rice', 'Content Rating', 'Genres', 'Last Updated', 'Current Ver', 'Andr
oid Ver']
```

DEFINING A NEW FUNCTION WITH THE AIM TO EXPLORE DATASETS

In [6]:
```
def explore_data(dataset,start,end,rows_and_columns='True'):
    dataset_slice = dataset[start:end]
    for row in dataset_slice:
        print(row)
        print('\n')
    if rows_and_columns:
        print('Number of rows:', len(dataset))
        print('Number of columns:', len(dataset[0]))

##EXPLORING DATASETS Finding header, rows, columns, no of rows, no of c
olumns in both datasets
## Dataset googleplaystore.csv

print(android_header)
print('\n')
explore_data(android, 0, 2, 'True')
print('\n')

## Dataset AppleStore.csv

print(ios_header)
```

```
print('\n')
explore_data(ios, 0, 2, 'True')
```

```
['App', 'Category', 'Rating', 'Reviews', 'Size', 'Installs', 'Type', 'P
rice', 'Content Rating', 'Genres', 'Last Updated', 'Current Ver', 'Andr
oid Ver']


['Photo Editor & Candy Camera & Grid & ScrapBook', 'ART_AND_DESIGN',
'4.1', '159', '19M', '10,000+', 'Free', '0', 'Everyone', 'Art & Desig
n', 'January 7, 2018', '1.0.0', '4.0.3 and up']


['Coloring book moana', 'ART_AND_DESIGN', '3.9', '967', '14M', '500,000
+', 'Free', '0', 'Everyone', 'Art & Design;Pretend Play', 'January 15,
2018', '2.0.0', '4.0.3 and up']


Number of rows: 10841
Number of columns: 13


['id', 'track_name', 'size_bytes', 'currency', 'price', 'rating_count_t
ot', 'rating_count_ver', 'user_rating', 'user_rating_ver', 'ver', 'cont
_rating', 'prime_genre', 'sup_devices.num', 'ipadSc_urls.num', 'lang.nu
m', 'vpp_lic']


['284882215', 'Facebook', '389879808', 'USD', '0.0', '2974676', '212',
'3.5', '3.5', '95.0', '4+', 'Social Networking', '37', '1', '29', '1']


['389801252', 'Instagram', '113954816', 'USD', '0.0', '2161558', '128
9', '4.5', '4.0', '10.23', '12+', 'Photo & Video', '37', '0', '29',
'1']


Number of rows: 7197
Number of columns: 16
```

So Google Play data set has 10841 apps and 13 columns. At a quick glance, the columns that might be useful for the purpose of our analysis are: 'App', 'Category', 'Reviews', 'Installs', 'Type', 'Price', and 'Genres'.

So ios data set has 7197 and 16 columns. We identified following columns that could help us with our analysis : 'track_name', 'currency', 'price', 'rating_count_tot', 'rating_count_ver', and 'prime_genre'.

DATA CLEANING

- Detect inaccurate data, and correct or remove it.
- Detect duplicate data, and remove the duplicates

In [8]:
```
## Detecting inaccurate data

for row in android:
    android_headerlength = len(android_header)
    rowlength = len(row)
    if rowlength != android_headerlength:
        print(row)

        print(android.index(row))


## Removing inaccurate data

print(len(android))
del android[10472]
print(len(android))

##Similarly checking accuracy of data in iOS apps dataset
for row in ios:
    ios_headerlength = len(ios_header)
    rowlength = len(row)
    if rowlength != ios_headerlength:
```

```
        print(row)
```

```
['Life Made WI-Fi Touchscreen Photo Frame', '1.9', '19', '3.0M', '1,000
+', 'Free', '0', 'Everyone', '', 'February 11, 2018', '1.0.19', '4.0 an
d up']
10472
10841
10840
```

The row 10472 corresponds to the app 'Life Made WI-Fi Touchscreen Photo Frame'

We can see that the rating is 19. This is clearly off because the maximum rating for a Google Play app is 5. As a consequence, we'll delete this row.

Similarly we checked accuracy of data in iOS apps dataset and found iOS apps dataset has no wrong data.

ios apps dataset (checking for duplicate entries)

In [11]:
```
## ios apps dataset (checking for duplicate entries)

for row in ios:
    name = row[1]
    if name == 'SCRABBLE Premium':
        print(row)



## Google Play Android apps dataset

for row in android:
    name = row[0]
    if name == 'Facebook':
        print(row)
```

```
['284815117', 'SCRABBLE Premium', '227547136', 'USD', '7.99', '105776',
'166', '3.5', '2.5', '5.19.0', '4+', 'Games', '37', '0', '6', '1']
```

```
['Facebook', 'SOCIAL', '4.1', '78158306', 'Varies with device', '1,000,
000,000+', 'Free', '0', 'Teen', 'Social', 'August 3, 2018', 'Varies wit
h device', 'Varies with device']
['Facebook', 'SOCIAL', '4.1', '78128208', 'Varies with device', '1,000,
000,000+', 'Free', '0', 'Teen', 'Social', 'August 3, 2018', 'Varies wit
h device', 'Varies with device']
```

ios dataset: Only one entry per app. So no duplicate entries. google play dataset: There are multiple entries per app in the google play store dataset. Let's find out no of duplicate entries in code cell below:

In [13]:
```
duplicate_apps = []
unique_apps = []

for row in android:
    name = row[0]
    if name in unique_apps:
        duplicate_apps.append(name)
    else:
        unique_apps.append(name)

print('Number of duplicate apps:', len(duplicate_apps))
print('\n')
print('Examples of duplicate apps:', duplicate_apps[:10])

print('Number of unique apps:', len(unique_apps))
```

```
Number of duplicate apps: 1181


Examples of duplicate apps: ['Quick PDF Scanner + OCR FREE', 'Box', 'Go
ogle My Business', 'ZOOM Cloud Meetings', 'join.me - Simple Meetings',
'Box', 'Zenefits', 'Google Ads', 'Google My Business', 'Slack']
Number of unique apps: 9659
```

Now in order to clean the dataset, We will keep only one entry per app (which has highest number of reviews) among all duplicate entries.

To do that, we will:

*Create a dictionary where each key is a unique app name, and the value is the highest number of reviews of that app

*Use the dictionary to create a new data set, which will have only one entry per app

*To clean data and to remove duplicate appa data, We must first create a dictionary and then separate the data of highest review in order to clean it for further analysis.

In [14]:
```
max_reviews = {}

for row in android:
    name = row[0]
    n_reviews = float(row[3])
    if name in max_reviews and max_reviews[name] < n_reviews:
        max_reviews[name] = n_reviews
    elif name not in max_reviews:
        max_reviews[name] = n_reviews
```

In [15]:
```
print(len(android))
print(len(max_reviews))
print(len(android)-len(duplicate_apps))
```

```
10840
9659
9659
```

Now we need to clean it further by keeping only only one entry per app (which has highest number of reviews) among all duplicate entries

For the duplicate cases, we'll only keep the entries with the highest number of reviews. In the code cell below:

We start by initializing two empty lists, android_clean and already_added. We loop through the android data set, and for every iteration: We isolate the name of the app and the number of

reviews. We add the current row (app) to the android_clean list, and the app name (name) to the already_added list with a condition.

In [16]:
```
android_clean = []
already_added = []

for row in android:
    name = row[0]
    n_reviews = float(row[3])

    if (max_reviews[name] == n_reviews) and (name not in already_added
):
        android_clean.append(row)
        already_added.append(name)
```

In [18]:
```
print(len(android_clean))
print(len(already_added))
```

```
9659
9659
```

Exploring the new data set, and confirm that the number of rows is 9,659.

In [20]:
```
explore_data(android_clean, 0, 2, True)
```

```
['Photo Editor & Candy Camera & Grid & ScrapBook', 'ART_AND_DESIGN',
'4.1', '159', '19M', '10,000+', 'Free', '0', 'Everyone', 'Art & Desig
n', 'January 7, 2018', '1.0.0', '4.0.3 and up']


['U Launcher Lite – FREE Live Cool Themes, Hide Apps', 'ART_AND_DESIG
N', '4.7', '87510', '8.7M', '5,000,000+', 'Free', '0', 'Everyone', 'Art
& Design', 'August 1, 2018', '1.2.4', '4.0.3 and up']


Number of rows: 9659
Number of columns: 13
```

Removing Non English Apps Data

Here in below code cell, we are creating a common function to remove Non English Apps data from both datasets and minimise the loss of data

In [21]:
```
def english_app(string):
    non_ascii = 0

    for character in string:
        if ord(character) > 127:
            non_ascii += 1

    if non_ascii > 3:
        return False
    else:
        return True
```

Lets explore some apps using above function

In [22]:
```
print(english_app('Instagram'))
print(english_app('爱奇艺PPS -《欢乐颂2》电视剧热播'))
print(english_app('Docs To Go™ Free Office Suite'))
print(english_app('Instachat 😜'))
```

```
True
False
True
True
```

Let's separate out English Apps Data in both datasets.

In [23]:
```
##ios apps dataset
ios_english = []

for row in ios:
    string = row[1]
```

```
    if english_app(string):
        ios_english.append(row)

print(explore_data(ios_english, 0, 2, True))
```

```
['284882215', 'Facebook', '389879808', 'USD', '0.0', '2974676', '212',
'3.5', '3.5', '95.0', '4+', 'Social Networking', '37', '1', '29', '1']


['389801252', 'Instagram', '113954816', 'USD', '0.0', '2161558', '128
9', '4.5', '4.0', '10.23', '12+', 'Photo & Video', '37', '0', '29',
'1']


Number of rows: 6183
Number of columns: 16
None
```

In [24]:
```
##google play store dataset
android_english = []

for row in android_clean:
    string = row[0]
    if english_app(string):
        android_english.append(row)

print(explore_data(android_english, 0, 2, True))
```

```
['Photo Editor & Candy Camera & Grid & ScrapBook', 'ART_AND_DESIGN',
'4.1', '159', '19M', '10,000+', 'Free', '0', 'Everyone', 'Art & Desig
n', 'January 7, 2018', '1.0.0', '4.0.3 and up']


['U Launcher Lite – FREE Live Cool Themes, Hide Apps', 'ART_AND_DESIG
N', '4.7', '87510', '8.7M', '5,000,000+', 'Free', '0', 'Everyone', 'Art
& Design', 'August 1, 2018', '1.2.4', '4.0.3 and up']


Number of rows: 9614
```

```
Number of columns: 13
None
```

Finally we have 9614 Android apps and 6183 iOS apps.

Isolating the Free Apps

In [25]:
```
android_final = []
ios_final = []

for row in android_english:
    price = row[7]
    if price == '0':
        android_final.append(row)

for row in ios_english:
    price = row[4]
    if price == '0.0':
        ios_final.append(row)

print(len(android_final))
print(len(ios_final))
```

```
8864
3222
```

Now after data cleaning we have 8864 Android apps and 3222 iOS apps for analysis.

Most Common Apps by Genre

In [26]:
```
def freq_table(dataset, index):
    table = {}
    total = 0

    for row in dataset:
        total += 1
```

```
        value = row[index]
        if value in table:
            table[value] += 1
        else:
            table[value] = 1

    table_percentages = {}
    for key in table:
        percentage = (table[key] / total) * 100
        table_percentages[key] = percentage

    return table_percentages


def display_table(dataset, index):
    table = freq_table(dataset, index)
    table_display = []
    for key in table:
        key_val_as_tuple = (table[key], key)
        table_display.append(key_val_as_tuple)

    table_sorted = sorted(table_display, reverse = True)
    for entry in table_sorted:
        print(entry[1], ':', entry[0])
```

In [27]:
```
display_table(ios_final, -5)
```

```
Games : 58.16263190564867
Entertainment : 7.883302296710118
Photo & Video : 4.9658597144630665
Education : 3.662321539416512
Social Networking : 3.2898820608317814
Shopping : 2.60707635009311
Utilities : 2.5139664804469275
Sports : 2.1415270018621975
Music : 2.0484171322160147
Health & Fitness : 2.0173805090006205
Productivity : 1.7380509000620732
Lifestyle : 1.5828677839851024
```

```
News : 1.3345747982619491
Travel : 1.2414649286157666
Finance : 1.1173184357541899
Weather : 0.8690254500310366
Food & Drink : 0.8069522036002483
Reference : 0.5586592178770949
Business : 0.5276225946617008
Book : 0.4345127250155183
Navigation : 0.186219739292365
Medical : 0.186219739292365
Catalogs : 0.12414649286157665
```

We can see more than half is for Games apps. Let's examine Genres and Category columns of the Google Play dataset

In [28]:
```
# Category

display_table(android_final, 1)
```

```
FAMILY : 18.907942238267147
GAME : 9.724729241877256
TOOLS : 8.461191335740072
BUSINESS : 4.591606498194946
LIFESTYLE : 3.9034296028880866
PRODUCTIVITY : 3.892148014440433
FINANCE : 3.7003610108303246
MEDICAL : 3.531137184115524
SPORTS : 3.395758122743682
PERSONALIZATION : 3.3167870036101084
COMMUNICATION : 3.2378158844765346
HEALTH_AND_FITNESS : 3.0798736462093865
PHOTOGRAPHY : 2.944494584837545
NEWS_AND_MAGAZINES : 2.7978339350180503
SOCIAL : 2.6624548736462095
TRAVEL_AND_LOCAL : 2.33528880866426
SHOPPING : 2.2450361010830324
BOOKS_AND_REFERENCE : 2.1435018050541514
DATING : 1.861462093862816
```

```
VIDEO_PLAYERS : 1.7937725631768955
MAPS_AND_NAVIGATION : 1.3989169675090252
FOOD_AND_DRINK : 1.2409747292418771
EDUCATION : 1.1620036101083033
ENTERTAINMENT : 0.9589350180505415
LIBRARIES_AND_DEMO : 0.9363718411552346
AUTO_AND_VEHICLES : 0.9250902527075812
HOUSE_AND_HOME : 0.8235559566787004
WEATHER : 0.8009927797833934
EVENTS : 0.7107400722021661
PARENTING : 0.6543321299638989
ART_AND_DESIGN : 0.6430505415162455
COMICS : 0.6204873646209386
BEAUTY : 0.5979241877256317
```

Family category (which accounts for almost 19% of the apps), opn checking play store..we found it as games for kids.

In [29]:
```
##Genres

display_table(android_final, -4)
```

```
Tools : 8.449909747292418
Entertainment : 6.069494584837545
Education : 5.347472924187725
Business : 4.591606498194946
Productivity : 3.892148014440433
Lifestyle : 3.892148014440433
Finance : 3.7003610108303246
Medical : 3.531137184115524
Sports : 3.463447653429603
Personalization : 3.3167870036101084
Communication : 3.2378158844765346
Action : 3.1024368231046933
Health & Fitness : 3.0798736462093865
Photography : 2.944494584837545
News & Magazines : 2.7978339350180503
Social : 2.6624548736462095
```

```
Travel & Local : 2.3240072202166067
Shopping : 2.2450361010830324
Books & Reference : 2.1435018050541514
Simulation : 2.0419675090252705
Dating : 1.861462093862816
Arcade : 1.8501805054151623
Video Players & Editors : 1.7712093862815883
Casual : 1.7599277978339352
Maps & Navigation : 1.3989169675090252
Food & Drink : 1.2409747292418771
Puzzle : 1.128158844765343
Racing : 0.9927797833935018
Role Playing : 0.9363718411552346
Libraries & Demo : 0.9363718411552346
Auto & Vehicles : 0.9250902527075812
Strategy : 0.9138086642599278
House & Home : 0.8235559566787004
Weather : 0.8009927797833934
Events : 0.7107400722021661
Adventure : 0.6768953068592057
Comics : 0.6092057761732852
Beauty : 0.5979241877256317
Art & Design : 0.5979241877256317
Parenting : 0.4963898916967509
Card : 0.45126353790613716
Casino : 0.42870036101083037
Trivia : 0.41741877256317694
Educational;Education : 0.39485559566787
Board : 0.3835740072202166
Educational : 0.3722924187725632
Education;Education : 0.33844765342960287
Word : 0.2594765342960289
Casual;Pretend Play : 0.236913357400722
Music : 0.2030685920577617
Racing;Action & Adventure : 0.16922382671480143
Puzzle;Brain Games : 0.16922382671480143
Entertainment;Music & Video : 0.16922382671480143
Casual;Brain Games : 0.13537906137184114
Casual;Action & Adventure : 0.13537906137184114
```

```
Arcade;Action & Adventure : 0.12409747292418773
Action;Action & Adventure : 0.10153429602888085
Educational;Pretend Play : 0.09025270758122744
Simulation;Action & Adventure : 0.078971119133574
Parenting;Education : 0.078971119133574
Entertainment;Brain Games : 0.078971119133574
Board;Brain Games : 0.078971119133574
Parenting;Music & Video : 0.06768953068592057
Educational;Brain Games : 0.06768953068592057
Casual;Creativity : 0.06768953068592057
Art & Design;Creativity : 0.06768953068592057
Education;Pretend Play : 0.056407942238267145
Role Playing;Pretend Play : 0.04512635379061372
Education;Creativity : 0.04512635379061372
Role Playing;Action & Adventure : 0.033844765342960284
Puzzle;Action & Adventure : 0.033844765342960284
Entertainment;Creativity : 0.033844765342960284
Entertainment;Action & Adventure : 0.033844765342960284
Educational;Creativity : 0.033844765342960284
Educational;Action & Adventure : 0.033844765342960284
Education;Music & Video : 0.033844765342960284
Education;Brain Games : 0.033844765342960284
Education;Action & Adventure : 0.033844765342960284
Adventure;Action & Adventure : 0.033844765342960284
Video Players & Editors;Music & Video : 0.02256317689530686
Sports;Action & Adventure : 0.02256317689530686
Simulation;Pretend Play : 0.02256317689530686
Puzzle;Creativity : 0.02256317689530686
Music;Music & Video : 0.02256317689530686
Entertainment;Pretend Play : 0.02256317689530686
Casual;Education : 0.02256317689530686
Board;Action & Adventure : 0.02256317689530686
Video Players & Editors;Creativity : 0.01128158844765343
Trivia;Education : 0.01128158844765343
Travel & Local;Action & Adventure : 0.01128158844765343
Tools;Education : 0.01128158844765343
Strategy;Education : 0.01128158844765343
Strategy;Creativity : 0.01128158844765343
Strategy;Action & Adventure : 0.01128158844765343
```

```
Simulation;Education : 0.01128158844765343
Role Playing;Brain Games : 0.01128158844765343
Racing;Pretend Play : 0.01128158844765343
Puzzle;Education : 0.01128158844765343
Parenting;Brain Games : 0.01128158844765343
Music & Audio;Music & Video : 0.01128158844765343
Lifestyle;Pretend Play : 0.01128158844765343
Lifestyle;Education : 0.01128158844765343
Health & Fitness;Education : 0.01128158844765343
Health & Fitness;Action & Adventure : 0.01128158844765343
Entertainment;Education : 0.01128158844765343
Communication;Creativity : 0.01128158844765343
Comics;Creativity : 0.01128158844765343
Casual;Music & Video : 0.01128158844765343
Card;Action & Adventure : 0.01128158844765343
Books & Reference;Education : 0.01128158844765343
Art & Design;Pretend Play : 0.01128158844765343
Art & Design;Action & Adventure : 0.01128158844765343
Arcade;Pretend Play : 0.01128158844765343
Adventure;Education : 0.01128158844765343
```

We're only looking for the bigger picture at the moment, so we'll only work with the Category column moving forward. Up to this point, we found that the App Store is dominated by apps designed for fun, while Google Play shows a more balanced landscape of both practical and for-fun apps. Now we'd like to get an idea about the kind of apps that have most users.

Most Popular Apps by Genre on the App Store

More the no of installs, more popular is the app. We have Installs column in Google Play dataset but not in ios dataset. So we'll take the total number of user ratings as a proxy, which we can find in the rating_count_tot app. Below, we calculate the average number of user ratings per app genre on the App Store:

In [30]:
```
genres_ios = freq_table(ios_final, -5)

for genre in genres_ios:
```

```
    total = 0
    len_genre = 0
    for app in ios_final:
        genre_app = app[-5]
        if genre_app == genre:
            n_ratings = float(app[5])
            total += n_ratings
            len_genre += 1
    avg_n_ratings = total / len_genre
    print(genre, ':', avg_n_ratings)
```

```
Social Networking : 71548.34905660378
Photo & Video : 28441.54375
Games : 22788.6696905016
Music : 57326.530303030304
Reference : 74942.11111111111
Health & Fitness : 23298.015384615384
Weather : 52279.892857142855
Utilities : 18684.456790123455
Travel : 28243.8
Shopping : 26919.690476190477
News : 21248.023255813954
Navigation : 86090.33333333333
Lifestyle : 16485.764705882353
Entertainment : 14029.830708661417
Food & Drink : 33333.92307692308
Sports : 23008.898550724636
Book : 39758.5
Finance : 31467.944444444445
Education : 7003.983050847458
Productivity : 21028.410714285714
Business : 7491.117647058823
Catalogs : 4004.0
Medical : 612.0
```

Navigation apps have the highest number of user reviews, but this figure is heavily influenced by Waze and Google Maps, which have close to half a million user reviews together

In [31]:
```
## print name and number of ratings

for app in ios_final:
    if app[-5] == 'Navigation':
        print(app[1], ':', app[5])
```

```
Waze - GPS Navigation, Maps & Real-time Traffic : 345046
Google Maps - Navigation & Transit : 154911
Geocaching® : 12811
CoPilot GPS – Car Navigation & Offline Maps : 3582
ImmobilienScout24: Real Estate Search in Germany : 187
Railway Route Search : 5
```

The same pattern applies to social networking apps, music apps. Reference apps have 74,942 user ratings on average, but it's actually the Bible and Dictionary.com.

We could get a better picture by removing these extremely popular apps for each genre

In [32]:
```
for app in ios_final:
    if app[-5] == 'Reference':
        print(app[1], ':', app[5])
```

```
Bible : 985920
Dictionary.com Dictionary & Thesaurus : 200047
Dictionary.com Dictionary & Thesaurus for iPad : 54175
Google Translate : 26786
Muslim Pro: Ramadan 2017 Prayer Times, Azan, Quran : 18418
New Furniture Mods - Pocket Wiki & Game Tools for Minecraft PC Edition
: 17588
Merriam-Webster Dictionary : 16849
Night Sky : 12122
City Maps for Minecraft PE - The Best Maps for Minecraft Pocket Edition
(MCPE) : 8535
LUCKY BLOCK MOD ™ for Minecraft PC Edition - The Best Pocket Wiki & Mod
s Installer Tools : 4693
GUNS MODS for Minecraft PC Edition - Mods Tools : 1497
```

```
Guides for Pokémon GO - Pokemon GO News and Cheats : 826
WWDC : 762
Horror Maps for Minecraft PE - Download The Scariest Maps for Minecraft
Pocket Edition (MCPE) Free : 718
VPN Express : 14
Real Bike Traffic Rider Virtual Reality Glasses : 8
教えて!goo : 0
Jishokun-Japanese English Dictionary & Translator : 0
```

Other genres that seem popular include weather, book, food and drink, or finance. The book genre seem to overlap a bit with the app idea we described above, but the other genres don't seem too interesting to us:

*Weather apps — people generally don't spend too much time in-app, and the chances of making profit from in-app adds are low.

*Food and drink, Finance apps —So making company app is outside the scope of our company.

Now let's analyze the Google Play market a bit.

Most Popular Apps by Genre on Google Play

In [33]:
```
##Installs

display_table(android_final, 5)
```

```
1,000,000+ : 15.726534296028879
100,000+ : 11.552346570397113
10,000,000+ : 10.548285198555957
10,000+ : 10.198555956678701
1,000+ : 8.393501805054152
100+ : 6.915613718411552
5,000,000+ : 6.825361010830325
500,000+ : 5.561823104693141
50,000+ : 4.7721119133574
5,000+ : 4.512635379061372
10+ : 3.5424187725631766
```

```
500+ : 3.2490974729241873
50,000,000+ : 2.3014440433213
100,000,000+ : 2.1322202166064983
50+ : 1.917870036101083
5+ : 0.78971119133574
1+ : 0.5076714801444043
500,000,000+ : 0.2707581227436823
1,000,000,000+ : 0.22563176895306858
0+ : 0.04512635379061372
0 : 0.01128158844765343
```

In [34]:

```
##converting each install number to float

categories_android = freq_table(android_final, 1)

for category in categories_android:
    total = 0
    len_category = 0
    for app in android_final:
        category_app = app[1]
        if category_app == category:
            n_installs = app[5]
            n_installs = n_installs.replace(',', '')
            n_installs = n_installs.replace('+', '')
            total += float(n_installs)
            len_category += 1
    avg_n_installs = total / len_category
    print(category, ':', avg_n_installs)
```

```
ART_AND_DESIGN : 1986335.0877192982
AUTO_AND_VEHICLES : 647317.8170731707
BEAUTY : 513151.88679245283
BOOKS_AND_REFERENCE : 8767811.894736841
BUSINESS : 1712290.1474201474
COMICS : 817657.2727272727
COMMUNICATION : 38456119.167247385
DATING : 854028.8303030303
EDUCATION : 1833495.145631068
ENTERTAINMENT : 11640705.88235294
```

```
EVENTS : 253542.2222222222
FINANCE : 1387692.475609756
FOOD_AND_DRINK : 1924897.7363636363
HEALTH_AND_FITNESS : 4188821.9853479853
HOUSE_AND_HOME : 1331540.5616438356
LIBRARIES_AND_DEMO : 638503.734939759
LIFESTYLE : 1437816.2687861272
GAME : 15588015.603248259
FAMILY : 3695641.8198090694
MEDICAL : 120550.61980830671
SOCIAL : 23253652.127118643
SHOPPING : 7036877.311557789
PHOTOGRAPHY : 17840110.40229885
SPORTS : 3638640.1428571427
TRAVEL_AND_LOCAL : 13984077.710144928
TOOLS : 10801391.298666667
PERSONALIZATION : 5201482.6122448975
PRODUCTIVITY : 16787331.344927534
PARENTING : 542603.6206896552
WEATHER : 5074486.197183099
VIDEO_PLAYERS : 24727872.452830188
NEWS_AND_MAGAZINES : 9549178.467741935
MAPS_AND_NAVIGATION : 4056941.7741935486
```

communication apps have the most installs: 38,456,119.

In [35]:
```
for app in android_final:
    if app[1] == 'COMMUNICATION' and (app[5] == '1,000,000,000+'
                                      or app[5] == '500,000,000+'
                                      or app[5] == '100,000,000+'):
        print(app[0], ':', app[5])
```

```
WhatsApp Messenger : 1,000,000,000+
imo beta free calls and text : 100,000,000+
Android Messages : 100,000,000+
Google Duo - High Quality Video Calls : 500,000,000+
Messenger – Text and Video Chat for Free : 1,000,000,000+
imo free video calls and chat : 500,000,000+
```

```
Skype - free IM & video calls : 1,000,000,000+
Who : 100,000,000+
GO SMS Pro - Messenger, Free Themes, Emoji : 100,000,000+
LINE: Free Calls & Messages : 500,000,000+
Google Chrome: Fast & Secure : 1,000,000,000+
Firefox Browser fast & private : 100,000,000+
UC Browser - Fast Download Private & Secure : 500,000,000+
Gmail : 1,000,000,000+
Hangouts : 1,000,000,000+
Messenger Lite: Free Calls & Messages : 100,000,000+
Kik : 100,000,000+
KakaoTalk: Free Calls & Text : 100,000,000+
Opera Mini - fast web browser : 100,000,000+
Opera Browser: Fast and Secure : 100,000,000+
Telegram : 100,000,000+
Truecaller: Caller ID, SMS spam blocking & Dialer : 100,000,000+
UC Browser Mini -Tiny Fast Private & Secure : 100,000,000+
Viber Messenger : 500,000,000+
WeChat : 100,000,000+
Yahoo Mail – Stay Organized : 100,000,000+
BBM - Free Calls & Messages : 100,000,000+
```

In [36]:
```
##removing communication apps(eg whatsapp)

under_100_m = []

for app in android_final:
    n_installs = app[5]
    n_installs = n_installs.replace(',', '')
    n_installs = n_installs.replace('+', '')
    if (app[1] == 'COMMUNICATION') and (float(n_installs) < 100000000):
        under_100_m.append(float(n_installs))

sum(under_100_m) / len(under_100_m)
```

Out[36]:
```
3603485.3884615386
```

Video players category : 24,727,872 installs. Big Players domination (in categories-social apps,

photography apps or productivity apps) can be seen.

The game genre seems pretty popular, but previously we found out this part of the market seems a bit saturated, so we'd like to come up with a different app recommendation if possible.

The books and reference genre with an average number of installs of 8,767,811. It's interesting to explore this in more depth, since we found this genre has some potential to work well on the App Store, and our aim is to recommend an app genre that shows potential for being profitable on both the App Store and Google Play.

Let's take a look at some of the apps from this genre and their number of installs:

In [37]:
```
for app in android_final:
    if app[1] == 'BOOKS_AND_REFERENCE':
        print(app[0], ':', app[5])
```

```
E-Book Read - Read Book for free : 50,000+
Download free book with green book : 100,000+
Wikipedia : 10,000,000+
Cool Reader : 10,000,000+
Free Panda Radio Music : 100,000+
Book store : 1,000,000+
FBReader: Favorite Book Reader : 10,000,000+
English Grammar Complete Handbook : 500,000+
Free Books - Spirit Fanfiction and Stories : 1,000,000+
Google Play Books : 1,000,000,000+
AlReader -any text book reader : 5,000,000+
Offline English Dictionary : 100,000+
Offline: English to Tagalog Dictionary : 500,000+
FamilySearch Tree : 1,000,000+
Cloud of Books : 1,000,000+
Recipes of Prophetic Medicine for free : 500,000+
ReadEra – free ebook reader : 1,000,000+
Anonymous caller detection : 10,000+
Ebook Reader : 5,000,000+
Litnet - E-books : 100,000+
Read books online : 5,000,000+
English to Urdu Dictionary : 500,000+
```

eBoox: book reader fb2 epub zip : 1,000,000+
English Persian Dictionary : 500,000+
Flybook : 500,000+
All Maths Formulas : 1,000,000+
Ancestry : 5,000,000+
HTC Help : 10,000,000+
English translation from Bengali : 100,000+
Pdf Book Download - Read Pdf Book : 100,000+
Free Book Reader : 100,000+
eBoox new: Reader for fb2 epub zip books : 50,000+
Only 30 days in English, the guideline is guaranteed : 500,000+
Moon+ Reader : 10,000,000+
SH-02J Owner's Manual (Android 8.0) : 50,000+
English-Myanmar Dictionary : 1,000,000+
Golden Dictionary (EN-AR) : 1,000,000+
All Language Translator Free : 1,000,000+
Azpen eReader : 500,000+
URBANO V 02 instruction manual : 100,000+
Bible : 100,000,000+
C Programs and Reference : 50,000+
C Offline Tutorial : 1,000+
C Programs Handbook : 50,000+
Amazon Kindle : 100,000,000+
Aab e Hayat Full Novel : 100,000+
Aldiko Book Reader : 10,000,000+
Google I/O 2018 : 500,000+
R Language Reference Guide : 10,000+
Learn R Programming Full : 5,000+
R Programing Offline Tutorial : 1,000+
Guide for R Programming : 5+
Learn R Programming : 10+
R Quick Reference Big Data : 1,000+
V Made : 100,000+
Wattpad 📖 Free Books : 100,000,000+
Dictionary - WordWeb : 5,000,000+
Guide (for X-MEN) : 100,000+
AC Air condition Troubleshoot,Repair,Maintenance : 5,000+
AE Bulletins : 1,000+
Ae Allah na Dai (Rasa) : 10,000+

50000 Free eBooks & Free AudioBooks : 5,000,000+
Ag PhD Field Guide : 10,000+
Ag PhD Deficiencies : 10,000+
Ag PhD Planting Population Calculator : 1,000+
Ag PhD Soybean Diseases : 1,000+
Fertilizer Removal By Crop : 50,000+
A-J Media Vault : 50+
Al-Quran (Free) : 10,000,000+
Al Quran (Tafsir & by Word) : 500,000+
Al Quran Indonesia : 10,000,000+
Al'Quran Bahasa Indonesia : 10,000,000+
Al Quran Al karim : 1,000,000+
Al-Muhaffiz : 50,000+
Al Quran : EAlim - Translations & MP3 Offline : 5,000,000+
Al-Quran 30 Juz free copies : 500,000+
Koran Read &MP3 30 Juz Offline : 1,000,000+
Hafizi Quran 15 lines per page : 1,000,000+
Quran for Android : 10,000,000+
Surah Al-Waqiah : 100,000+
Hisnul Al Muslim - Hisn Invocations & Adhkaar : 100,000+
Satellite AR : 1,000,000+
Audiobooks from Audible : 100,000,000+
Kinot & Eichah for Tisha B'Av : 10,000+
AW Tozer Devotionals - Daily : 5,000+
Tozer Devotional -Series 1 : 1,000+
The Pursuit of God : 1,000+
AY Sing : 5,000+
Ay Hasnain k Nana Milad Naat : 10,000+
Ay Mohabbat Teri Khatir Novel : 10,000+
Arizona Statutes, ARS (AZ Law) : 1,000+
Oxford A-Z of English Usage : 1,000,000+
BD Fishpedia : 1,000+
BD All Sim Offer : 10,000+
Youboox - Livres, BD et magazines : 500,000+
B&H Kids AR : 10,000+
B y H Niños ES : 5,000+
Dictionary.com: Find Definitions for English Words : 10,000,000+
English Dictionary - Offline : 10,000,000+
Bible KJV : 5,000,000+

```
Borneo Bible, BM Bible : 10,000+
MOD Black for BM : 100+
BM Box : 1,000+
Anime Mod for BM : 100+
NOOK: Read eBooks & Magazines : 10,000,000+
NOOK Audiobooks : 500,000+
NOOK App for NOOK Devices : 500,000+
Browsery by Barnes & Noble : 5,000+
bp e-store : 1,000+
Brilliant Quotes: Life, Love, Family & Motivation : 1,000,000+
BR Ambedkar Biography & Quotes : 10,000+
BU Alsace : 100+
Catholic La Bu Zo Kam : 500+
Khrifa Hla Bu (Solfa) : 10+
Kristian Hla Bu : 10,000+
SA HLA BU : 1,000+
Learn SAP BW : 500+
Learn SAP BW on HANA : 500+
CA Laws 2018 (California Laws and Codes) : 5,000+
Bootable Methods(USB-CD-DVD) : 10,000+
cloudLibrary : 100,000+
SDA Collegiate Quarterly : 500+
Sabbath School : 100,000+
Cypress College Library : 100+
Stats Royale for Clash Royale : 1,000,000+
GATE 21 years CS Papers(2011-2018 Solved) : 50+
Learn CT Scan Of Head : 5,000+
Easy Cv maker 2018 : 10,000+
How to Write CV : 100,000+
CW Nuclear : 1,000+
CY Spray nozzle : 10+
BibleRead En Cy Zh Yue : 5+
CZ-Help : 5+
Modlitební knížka CZ : 500+
Guide for DB Xenoverse : 10,000+
Guide for DB Xenoverse 2 : 10,000+
Guide for IMS DB : 10+
DC HSEMA : 5,000+
DC Public Library : 1,000+
```

Painting Lulu DC Super Friends : 1,000+
Dictionary : 10,000,000+
Fix Error Google Playstore : 1,000+
D. H. Lawrence Poems FREE : 1,000+
Bilingual Dictionary Audio App : 5,000+
DM Screen : 10,000+
wikiHow: how to do anything : 1,000,000+
Dr. Doug's Tips : 1,000+
Bible du Semeur-BDS (French) : 50,000+
La citadelle du musulman : 50,000+
DV 2019 Entry Guide : 10,000+
DV 2019 - EDV Photo & Form : 50,000+
DV 2018 Winners Guide : 1,000+
EB Annual Meetings : 1,000+
EC - AP & Telangana : 5,000+
TN Patta Citta & EC : 10,000+
AP Stamps and Registration : 10,000+
CompactiMa EC pH Calibration : 100+
EGW Writings 2 : 100,000+
EGW Writings : 1,000,000+
Bible with EGW Comments : 100,000+
My Little Pony AR Guide : 1,000,000+
SDA Sabbath School Quarterly : 500,000+
Duaa Ek Ibaadat : 5,000+
Spanish English Translator : 10,000,000+
Dictionary - Merriam-Webster : 10,000,000+
JW Library : 10,000,000+
Oxford Dictionary of English : Free : 10,000,000+
English Hindi Dictionary : 10,000,000+
English to Hindi Dictionary : 5,000,000+
EP Research Service : 1,000+
Hymnes et Louanges : 100,000+
EU Charter : 1,000+
EU Data Protection : 1,000+
EU IP Codes : 100+
EW PDF : 5+
BakaReader EX : 100,000+
EZ Quran : 50,000+
FA Part 1 & 2 Past Papers Solved Free – Offline : 5,000+

```
La Fe de Jesus : 1,000+
La Fe de Jesús : 500+
Le Fe de Jesus : 500+
Florida - Pocket Brainbook : 1,000+
Florida Statutes (FL Code) : 1,000+
English To Shona Dictionary : 10,000+
Greek Bible FP (Audio) : 1,000+
Golden Dictionary (FR-AR) : 500,000+
Fanfic-FR : 5,000+
Bulgarian French Dictionary Fr : 10,000+
Chemin (fr) : 1,000+
The SCP Foundation DB fr nn5n : 1,000+
```

The book and reference genre includes a variety of apps: software for processing and reading ebooks, various collections of libraries, dictionaries, tutorials on programming or languages, etc. It seems there's still a small number of extremely popular apps that skew the average:

In [38]:
```
for app in android_final:
    if app[1] == 'BOOKS_AND_REFERENCE' and (app[5] == '1,000,000,000+'
                                            or app[5] == '500,000,000+'
                                            or app[5] == '100,000,000+'
):
        print(app[0], ':', app[5])
```

```
Google Play Books : 1,000,000,000+
Bible : 100,000,000+
Amazon Kindle : 100,000,000+
Wattpad 📖 Free Books : 100,000,000+
Audiobooks from Audible : 100,000,000+
```

In [39]:
```
for app in android_final:
    if app[1] == 'BOOKS_AND_REFERENCE' and (app[5] == '1,000,000+'
                                            or app[5] == '5,000,000+'
                                            or app[5] == '10,000,000+'
                                            or app[5] == '50,000,000+'
):
        print(app[0], ':', app[5])
```

```
Wikipedia : 10,000,000+
```

Wikipedia : 10,000,000+

Cool Reader : 10,000,000+
Book store : 1,000,000+
FBReader: Favorite Book Reader : 10,000,000+
Free Books - Spirit Fanfiction and Stories : 1,000,000+
AlReader -any text book reader : 5,000,000+
FamilySearch Tree : 1,000,000+
Cloud of Books : 1,000,000+
ReadEra – free ebook reader : 1,000,000+
Ebook Reader : 5,000,000+
Read books online : 5,000,000+
eBoox: book reader fb2 epub zip : 1,000,000+
All Maths Formulas : 1,000,000+
Ancestry : 5,000,000+
HTC Help : 10,000,000+
Moon+ Reader : 10,000,000+
English-Myanmar Dictionary : 1,000,000+
Golden Dictionary (EN-AR) : 1,000,000+
All Language Translator Free : 1,000,000+
Aldiko Book Reader : 10,000,000+
Dictionary - WordWeb : 5,000,000+
50000 Free eBooks & Free AudioBooks : 5,000,000+
Al-Quran (Free) : 10,000,000+
Al Quran Indonesia : 10,000,000+
Al'Quran Bahasa Indonesia : 10,000,000+
Al Quran Al karim : 1,000,000+
Al Quran : EAlim - Translations & MP3 Offline : 5,000,000+
Koran Read &MP3 30 Juz Offline : 1,000,000+
Hafizi Quran 15 lines per page : 1,000,000+
Quran for Android : 10,000,000+
Satellite AR : 1,000,000+
Oxford A-Z of English Usage : 1,000,000+
Dictionary.com: Find Definitions for English Words : 10,000,000+
English Dictionary - Offline : 10,000,000+
Bible KJV : 5,000,000+
NOOK: Read eBooks & Magazines : 10,000,000+
Brilliant Quotes: Life, Love, Family & Motivation : 1,000,000+
Stats Royale for Clash Royale : 1,000,000+
Dictionary : 10,000,000+
wikiHow: how to do anything : 1,000,000+

```
EGW Writings : 1,000,000+
My Little Pony AR Guide : 1,000,000+
Spanish English Translator : 10,000,000+
Dictionary - Merriam-Webster : 10,000,000+
JW Library : 10,000,000+
Oxford Dictionary of English : Free : 10,000,000+
English Hindi Dictionary : 10,000,000+
English to Hindi Dictionary : 5,000,000+
```

It looks like there are many very popular apps, so this market shows more competition.

We also notice there are quite a few apps built around the book Quran, which suggests that building an app around a popular book can be profitable. It seems that taking a popular book (perhaps a more recent book) and turning it into an app could be profitable for both the Google Play and the App Store markets.

Moreover,there is no doubt that one book app integrating some special features besides the raw version (eg. daily quotes from the book, an audio version of the book, quizzes on the book, a forum where people can discuss the book, etc.) or other feature,may be profitable app profile.

Conclusion

In this project,we analyzed data about Google Play the App Store apps and try to find which genre that can be with profitable potential for both markets.

In the end we conclude that a BOOK app with new features like forum where people can discuss the book, daily quotes from the book, an audio version of the book, quizzes on the book, etc.could be profitable for both the Google Play and the App Store markets.Although the market in this category are full of libraries, we may enrich the app with new features.