Data analysing for Android and iOS mobile apps.

The aim of that project is to analyze data of apps which are free to download and install,that could help developers understand what type of apps are likely to attract more users.

The Google Play data set

```
In [7]: from csv import reader

opened_file = open('googleplaystore.csv')
read_file = reader(opened_file)
android = list(read_file)
android_header = android[0]
android = android[1:]
```

The App Store data set

```
In [8]: opened_file = open('AppleStore.csv')
read_file = reader(opened_file)
ios = list(read_file)
ios_header = ios[0]
ios = ios[1:]
```

```
In [9]: def explore_data(dataset, start, end, rows_and_columns=False):
            dataset_slice = dataset[start:end]
            for row in dataset_slice:
                print(row)
                print('\n') # adds a new (empty) line between rows
```

```python
    if rows_and_columns:
        print('Number of rows:', len(dataset))
        print('Number of columns:', len(dataset[0]))

print(android_header)
print('\n')
explore_data(android, 0, 3, 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']

['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: 10841
Number of columns: 13

In [10]:
```python
print(android_header, sep= "   ")

print(ios_header)
print("\n")
explore_data(android,0,3,True)
print("\n")
explore_data(ios,0,3,True)
```

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


['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']


['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: 10841
Number of columns: 13


['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']


['529479190', 'Clash of Clans', '116476928', 'USD', '0.0', '2130805',
'579', '4.5', '4.5', '9.24.12', '9+', 'Games', '38', '5', '18', '1']
```

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

Delecting wrong Data from Android' dataset. the outlines an error for row 10472. Let's print this row and compare it against the header and another row that is correct.#

```
In [11]: print(android[10472])   # incorrect row
         print('\n')
         print(android_header)   # header
         print('\n')
         print(android[0])        # correct 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']
```

```
['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']
```

Deleting unwanted data.

```
In [17]: print(len(android))
         del android[10472]   # don't run this more than once
         print(len(android))
```

```
10841
10840
```

Explore Duplicate Entries.

```python
In [18]: for app in android:
             name = app[0]
             if name == 'Instagram':
                 print(app)
```

```
['Instagram', 'SOCIAL', '4.5', '66577313', 'Varies with device', '1,00
0,000,000+', 'Free', '0', 'Teen', 'Social', 'July 31, 2018', 'Varies wi
th device', 'Varies with device']
['Instagram', 'SOCIAL', '4.5', '66577446', 'Varies with device', '1,00
0,000,000+', 'Free', '0', 'Teen', 'Social', 'July 31, 2018', 'Varies wi
th device', 'Varies with device']
['Instagram', 'SOCIAL', '4.5', '66577313', 'Varies with device', '1,00
0,000,000+', 'Free', '0', 'Teen', 'Social', 'July 31, 2018', 'Varies wi
th device', 'Varies with device']
['Instagram', 'SOCIAL', '4.5', '66509917', 'Varies with device', '1,00
0,000,000+', 'Free', '0', 'Teen', 'Social', 'July 31, 2018', 'Varies wi
th device', 'Varies with device']
```

```python
In [19]: Duplicate_apps = []
         Unique_apps = []

         for app in android:
             name = app[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[:15])
         print('\n')
         print("Number of Unique_apps is :",len(Unique_apps))
```

Number of Duplicate apps: 1181

Examples of Duplicate apps: ['Quick PDF Scanner + OCR FREE', 'Box', 'Google My Business', 'ZOOM Cloud Meetings', 'join.me - Simple Meetings', 'Box', 'Zenefits', 'Google Ads', 'Google My Business', 'Slack', 'FreshBooks Classic', 'Insightly CRM', 'QuickBooks Accounting: Invoicing & Expenses', 'HipChat - Chat Built for Teams', 'Xero Accounting Software']

Number of Unique_apps is : 9659

Select the hights reviews of the duplicates data inside the App.

In [20]:
```python
Max_reviews = {}

for app in android:
    name = app[0]
    n_reviews = float(app[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

print(len(Max_reviews))
print(len(android))
```

9659
10840

Remove deplicate data of the highist reviews of

App.

```
In [23]: clean_android = []
         already_added = []

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

             if (Max_reviews[name] == n_reviews) and (name not in already_added
         ):
                 clean_android.append(app)
                 already_added.append(name)


         print(len(clean_android))
```

9659

filter out non-English apps of each data sets in order to extract only English words witnin max three emojis.

```
In [24]: def in_English(string):
             non_ascii = 0

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

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

```
In [25]: android_English = []
         ios_English = []

         for app in clean_android:
             name = app[0]
             if in_English(name):
                 android_English.append(app)

         for app in ios:
             name = app[1]
             if in_English(name):
                 ios_English.append(app)


         print(len(ios_English))
         print(len(android_English))
```

```
6183
9614
```

```
In [26]: explore_data(ios_English, 0,3,True)
         print('\n')
         explore_data(android_English, 0,3,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']


['529479190', 'Clash of Clans', '116476928', 'USD', '0.0', '2130805',
'579', '4.5', '4.5', '9.24.12', '9+', 'Games', '38', '5', '18', '1']


Number of rows: 6183
Number of columns: 16
```

```
['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']


['Sketch - Draw & Paint', 'ART_AND_DESIGN', '4.5', '215644', '25M', '5
0,000,000+', 'Free', '0', 'Teen', 'Art & Design', 'June 8, 2018', 'Vari
es with device', '4.2 and up']


Number of rows: 9614
Number of columns: 13
```

Isolating the free apps for analysising of both data sets as mention in Introduction.

In [27]:
```python
android_free = []
ios_free = []

for app in android_English:
    price = app[7]
    if price == '0':
        android_free.append(app)

for app in ios_English:
    price = app[4]
    if price == '0.0':
        ios_free.append(app)

print(len(android_free))
```

```
print(len(ios_free))

explore_data(ios_free,-5,True)
```

8864
3222

Finding app profile that is successful on both markets in order to minimize risks and overhead, our validation strategy for an app idea is comprised of three steps:# # 1. Build a minimal Android version of the app, and add it to Google Play.# # 2. If the app has a good response from users, we then develop it further.# # 3. If the app is profitable after six months, we also build an iOS version of the app and add it to the App Store.# # Analaysing could be done by build frequency tables for a few columns in choosen data set which is the most common genres for each market. Therefor we select Android' App for that analysisng research.

```python
In [33]: def frq_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 = frq_table(dataset, index)
             table_display = []
             for key in table:
```

```
            key_val_as_tuple = (table[key], key)# convert into tuple to sor
ted.
            table_display.append(key_val_as_tuple)
            table_sorted = sorted(table_display, reverse = True)
        for entry in table_sorted:
            print(entry[1], ':', entry[0])
```

We start by examining the frequency table for the prime_genre column, genre, and Categroy of the App Store data set.

In [34]: `display_table(ios_free,11) #prime genre`

```
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 : 0.1241464928615 7665
```

In [35]: `display_table(android,1) # Category`

```
FAMILY : 18.19188191881919
GAME : 10.55350553505535
```

```
TOOLS : 7.776752767527675
MEDICAL : 4.271217712177122
BUSINESS : 4.243542435424354
PRODUCTIVITY : 3.911439114391144
PERSONALIZATION : 3.616236162361624
COMMUNICATION : 3.5701107011070112
SPORTS : 3.5424354243542435
LIFESTYLE : 3.523985239852398
FINANCE : 3.3763837638376386
HEALTH_AND_FITNESS : 3.1457564575645756
PHOTOGRAPHY : 3.0904059040590406
SOCIAL : 2.7214022140221403
NEWS_AND_MAGAZINES : 2.61070110701107
SHOPPING : 2.3985239852398523
TRAVEL_AND_LOCAL : 2.3800738007380073
DATING : 2.158671586715867
BOOKS_AND_REFERENCE : 2.1309963099630997
VIDEO_PLAYERS : 1.6143911439114391
EDUCATION : 1.4391143911439115
ENTERTAINMENT : 1.3745387453874538
MAPS_AND_NAVIGATION : 1.2638376383763839
FOOD_AND_DRINK : 1.1715867158671587
HOUSE_AND_HOME : 0.8118081180811807
LIBRARIES_AND_DEMO : 0.7841328413284132
AUTO_AND_VEHICLES : 0.7841328413284132
WEATHER : 0.7564575645756457
ART_AND_DESIGN : 0.5996309963099631
EVENTS : 0.5904059040590406
PARENTING : 0.5535055350553505
COMICS : 0.5535055350553505
BEAUTY : 0.488929889298893
```

In [36]: `display_table(android_free, -4)# Genres`

```
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
```

**Conclusion the analysing of both data_sets
show that App Store is dominated by apps
designed for fun, meanwhile Google Play shows**

a more balanced landscape of both practical and fun apps

Planning to find out which genres in ios have the most popular (have the most users) by calculating the average number of installs for each app genre but the information is missing on intalls so that get user ratings (rating_count_tot)as a proxy to accounting the average rating.

In [60]:
```python
genres_ios = frq_table(ios_free, -5)

print(genres_ios)
print("\n")
s=explore_data(ios_free, 0,4,True)

print(s)
```

{'Social Networking': 3.2898820608317814, 'Photo & Video': 4.9658597144630665, 'Games': 58.16263190564867, 'Music': 2.0484171322160147, 'Reference': 0.5586592178770949, 'Health & Fitness': 2.0173805090006205, 'Weather': 0.8690254500310366, 'Utilities': 2.5139664804469275, 'Travel': 1.2414649286157666, 'Shopping': 2.60707635009311, 'News': 1.3345747982619491, 'Navigation': 0.186219739292365, 'Lifestyle': 1.5828677839851024, 'Entertainment': 7.883302296710118, 'Food & Drink': 0.8069522036002483, 'Sports': 2.1415270018621975, 'Book': 0.4345127250155183, 'Finance': 1.1173184357541899, 'Education': 3.662321539416512, 'Productivity': 1.7380509000620732, 'Business': 0.5276225946617008, 'Catalogs': 0.12414649286157665, 'Medical': 0.186219739292365}

['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', '1289', '4.5', '4.0', '10.23', '12+', 'Photo & Video', '37', '0', '29', '1']

['529479190', 'Clash of Clans', '116476928', 'USD', '0.0', '2130805', '579', '4.5', '4.5', '9.24.12', '9+', 'Games', '38', '5', '18', '1']

```
['420009108', 'Temple Run', '65921024', 'USD', '0.0', '1724546', '384
2', '4.5', '4.0', '1.6.2', '9+', 'Games', '40', '5', '1', '1']


Number of rows: 3222
Number of columns: 16
None
```

In [62]:
```python
genres_ios = frq_table(ios_free, -5)

for genre in genres_ios:
    total = 0
    len_genre = 0
    for app in ios_free:
        genre_app = app[-5]
        if genre_app == genre:
            n_ratings = float(app[5]) # column of (rating_count_tot).
            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
```

On average, 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 [65]:  for app in ios_free:
              if app[-5]== "Navigation":
                  print(app[1],':',app[5])# print name and number of ratings
```

```
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
```

```
In [66]:  for app in ios_free:
              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
```

Planning to find out which genres have the most popular (have the most users) by calculating the average number of number of installs for the Google Play market.

In [108]: `display_table(android_free,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
```

```python
In [137]:  categories_android = frq_table(android_free, 1)

           for category in categories_android:
               total = 0
               len_category = 0
               for app in android_free:
                   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 : 253542.22222222222
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
```

The results show above that the average of communication apps have the most installs by 38,456,119. as accounting for many different apps that have over one billion installs which including (WhatsApp, Facebook Messenger, Skype, Google Chrome, Gmail, and Hangouts)

In [124]:
```python
for app in android_free:
    if app[1]=="COMMUNICATION" and (app[5]== '1000,000+' or
                                    app[5]== '100,000+' or
                                    app[5]== '500,000'):
        print (app[0],':', app[5])
```

```
Antillean Gold Telegram (original version) : 100,000+
K-@ Mail - Email App : 100,000+
M star Dialer : 100,000+
T-Mobile DIGITS : 100,000+
PlacarTv Futebol Ao Vivo : 100,000+
WiFi Access Point (hotspot) : 100,000+
Baby Monitor AV : 100,000+
AZ Browser. Private & Download : 100,000+
BT MeetMe with Dolby Voice : 100,000+
CB On Mobile : 100,000+
Channel 19 : 100,000+
QRZ Assistant : 100,000+
Morse Code Reader : 100,000+
Goodbox - Mega App : 100,000+
Everbridge : 100,000+
Portable Wi-Fi hotspot Free : 100,000+
```

If we removed all the communication apps that have over 100 million installs, the average would be reduced roughly ten times:

In [136]:
```python
app_under_100m =[]
for app in android_free:

    n_installs =app[5]
    n_installs =  n_installs.replace('+','')
    n_installs =  n_installs.replace(',','')

    if app[1]=="COMMUNICATION"  and  float(n_installs) < 100000000:
        app_under_100m.append(float(n_installs))
avr= sum(app_under_100m)/ len (app_under_100m)
print(avr)
```

3603485.3884615386

In [ ]:
```
Display the maximum rate THAN (100,000) of installs in categroy of VIDE
O PLAYER.
```

In [148]:
```python
sorted_table=[]
for app in android_free:
    if app[1]== 'VIDEO_PLAYERS' and (app[5]>="100,000") :
        sorte= [app[5],app[0]]
        sorted_table.append(sorte)
        sor=sorted(sorted_table, reverse=True)
print(sor)


# Presenting the number of installs sorted as ascending.
```

[['500,000,000+', 'MX Player'], ['500,000+', 'Video Player All Format f
or Android'], ['500,000+', 'GoPlus Cam'], ['500,000+', 'DR TV'], ['500,
000+', 'CJ VLC HD Remote (+ Stream)'], ['500+', 'YourTube Video Views B
G'], ['500+', 'Video Wallpaper Show'], ['500+', 'DG Screen Recorder'],
['500+', 'CX Monthly Tech News'], ['500+', 'CJ Camcorder'], ['500+', 'A
-Z Screen Recorder -'], ['50,000,000+', 'Vote for'], ['50,000,000+', 'V

```
igo Video'], ['50,000,000+', 'VMate'], ['50,000,000+', 'Samsung Video L
ibrary'], ['50,000,000+', 'Ringdroid'], ['50,000,000+', 'MiniMovie - Fr
ee Video and Slideshow Editor'], ['50,000,000+', 'LIKE — Magic Video Ma
ker & Community'], ['50,000,000+', 'KineMaster — Pro Video Editor'],
['50,000,000+', 'HD Video Downloader : 2018 Best video mate'], ['50,00
0,000+', 'DU Recorder — Screen Recorder, Video Editor, Live'], ['50,000
+', 'dv Prompter'], ['50,000+', 'bgtime.tv'], ['50,000+', 'ES-IPTV'],
['50,000+', 'BR Series'], ['50,000+', 'AX Video Player'], ['50,000+',
'ACTIVEON CX & CX GOLD'], ['5,000,000+', 'YouCut - Video Editor & Video
Maker, No Watermark'], ['5,000,000+', 'WiFi Baby Monitor - NannyCam'],
['5,000,000+', 'Video Editor'], ['5,000,000+', 'HTC Service—Video Playe
r'], ['5,000,000+', 'Droid Zap by Motorola'], ['5,000,000+', 'Adobe Pre
miere Clip'], ['5,000,000+', 'ActionDirector Video Editor - Edit Videos
Fast'], ['5,000+', 'W Box VMS HD'], ['5,000+', 'Videos downloader for F
acebook:fast fb video saver'], ['5,000+', 'Q-See Plus'], ['5,000+', 'M-
Sight Pro'], ['5,000+', 'HD Video Player - Video & MP3 Player | AV Play
er |'], ['5,000+', 'Furrion ES Control'], ['5,000+', 'EF Sidekick'],
['5,000+', 'DZ Popup Video Player'], ['5,000+', 'DV Lottery Photo'],
['5,000+', 'BR Video Player'], ['5,000+', 'BG video - floating video -
background video'], ['5,000+', 'Ay'], ['5,000+', 'A-B repeater'], ['5,0
00+', '4K VIDEO PLAYER ULTRA HD'], ['100,000,000+', 'VivaVideo - Video
Editor & Photo Movie'], ['100,000,000+', 'VideoShow-Video Editor, Video
Maker, Beauty Camera'], ['100,000,000+', 'VLC for Android'], ['100,000,
000+', 'Motorola Gallery'], ['100,000,000+', 'Motorola FM Radio'], ['10
0,000,000+', 'Dubsmash'], ['100,000+', 'amazer - Global Kpop Video Comm
unity'], ['100,000+', 'Video Downloader for FB : Video Download with Li
nk'], ['100,000+', "Sketch 'n' go"], ['100,000+', 'Nero AirBurn'], ['10
0,000+', 'GoAction'], ['100,000+', 'Free TV series'], ['100,000+', 'Ez
Screen Recorder (no ad)'], ['100,000+', 'EZ Web Video Cast | Chromecas
t'], ['100,000+', 'ES Audio Player ( Shortcut )'], ['100,000+', 'BGCN T
V'], ['100,000+', 'AW Screen Recorder No Root'], ['100,000+', 'AB Repea
t Player']]
```

We see the same pattern for the video players category, which is the runner-up with 24,727,872 installs. The market is dominated by apps like Youtube, Google Play Movies & TV, or MX Player.In other word there is no potential apps from this categroy to investment in AppStore. The books and reference genre seems closely popular as well, 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:

```python
In [149]:  for app in android_free:
               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 include a wide range of apps: software for processing and reading ebooks, different groups 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 [157]:
```python
for app in android_free:
    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+
```

The result above illustrates that, there are only a few very popular apps, so this market still shows potential. Let's try to get some app ideas based on the kind of apps that are somewhere in the middle in terms of popularity (between 1,000,000 and 100,000,000 downloads):

In [161]:
```python
for app in android_free:
    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+
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+
```

This niche seems to be covered by software for processing and reading ebooks, as well as various collections of libraries and dictionaries, so it's likely not a proper idea to build similar apps since there'll be some significant 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.

However, it seems that the market is already full of libraries, so perhaps needing to add some special features besides the raw version of the book. This might include daily quotes from the book, an audio version of the book, quizzes on the book, a forum where people can discuss the book, etc.

To sum up In this project, the data has been analyzed data regarding the App Store and Google Play mobile apps with the goal of recommending an app profile that can be profitable for both markets. We concluded 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. The markets are already full of libraries, so we need to add some special features besides the raw version of the book. This might include daily quotes from the book, an audio version of the book, quizzes on the book, a forum where people can discuss the book, etc.