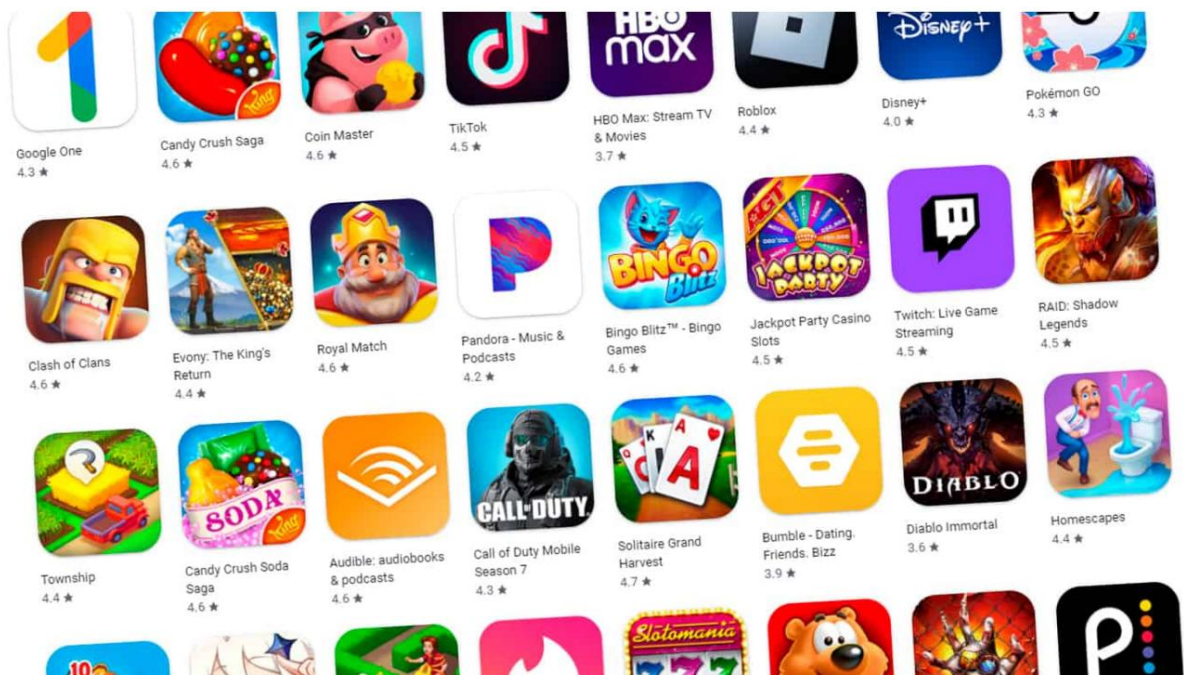


## Hồ sơ ứng dụng có lợi nhuận cho App Store và Thị trường Google Play 📱

Mục tiêu của chúng tôi trong dự án này là tìm hồ sơ ứng dụng dành cho thiết bị di động mang lại lợi nhuận cho thị trường App Store và Google Play. Chúng tôi đang làm việc với tư cách là nhà phân tích dữ liệu cho một công ty xây dựng ứng dụng dành cho thiết bị di động Android và iOS và công việc của chúng tôi là hỗ trợ nhóm các nhà phát triển đưa ra quyết định dựa trên dữ liệu đối với loại ứng dụng mà họ xây dựng.

Tại công ty của chúng tôi, chúng tôi chỉ xây dựng các ứng dụng miễn phí để tải xuống và cài đặt, đồng thời nguồn doanh thu chính của chúng tôi bao gồm quảng cáo trong ứng dụng. Điều này có nghĩa là doanh thu của chúng tôi cho bất kỳ ứng dụng cụ thể nào chủ yếu bị ảnh hưởng bởi số lượng người dùng sử dụng ứng dụng của chúng tôi. Mục tiêu của chúng tôi cho dự án này là phân tích dữ liệu để giúp các nhà phát triển của chúng tôi hiểu loại ứng dụng nào có khả năng thu hút nhiều người dùng hơn.



Hãy bắt đầu bằng cách mở hai bộ dữ liệu

Trong [6]: từ trình đọc nhập csv

```
open_app_store=open('AppleStore.csv')
read_app=reader(open_app_store) ios
=list(read_app)
ios_header= ios[0]
ios_data = ios[1:]

open_google_play=open('googleplaystore.csv')
read_gg=reader(open_google_play)
android =list(read_gg)
android_header = android[0]
android_data = android[1:]
```

Tạo một explorer\_data giúp in một vài hàng của bộ dữ liệu:

Trong [7]:

```
def explorer_data(dataset, start, end, rows_and_columns=True):
    data_slice=dataset[start:end] cho
    hàng trong data_slice:
        print(row)
        print('\n')
    if rows_and_columns:
        print('Number of rows:',len(dataset))
        print('Số cột:',len(dataset[0]))
```

Trong [8]: explorer\_data(ios\_data, 0, 2, True)

```
['284882215', 'Facebook', '389879808', 'USD', '0.0', '2974676', '212', '3.5', '3.5',
'95.0', '4+', 'Mạng xã hội', '37', '1', '29', '1']
```

```
['389801252', 'Instagram', '113954816', 'USD', '0.0', '2161558', '1289', '4.
5', '4.0', '10.23', '12+', 'Ảnh & Video', '37', '0', '29', '1']
```

Số hàng: 7197

Số cột: 16

Chúng ta có thể thấy rằng của hàng iOS có 7197 ứng dụng và 16 cột. Các cột thú vị nhất có thể là: track\_name, price, rating\_count\_tot, user\_rating, cont\_rating, prime\_genre, size\_bytes,

Trong [9]: `explorer_data(android_data,0,2,True)`

```
['Trình chỉnh sửa ảnh & Máy ảnh kẹo & Lư ới & Sủ lư u niệu', 'ART_AND_DESIGN', '4.1', '159',
'19M', '10.000+', 'Miễn phí', '0', 'Mọi ngư ời', 'Nghệ thuật & Thiết kế ', 'Ngày 7 tháng 1 năm
2018', '1.0.0', '4.0.3 trở lên']
```

```
['Sách tô màu rêu', 'ART_AND_DESIGN', '3.9', '967', '14M', '500,000+', 'F ree', '0', 'Mọi ngư ời',
'Nghệ thuật & Thiết kế;Chơ i giả vờ', '15/01/2018', '2.0. 0', '4.0.3 trở lên']
```

Số hàng: 10841

Số cột: 13

Cửa hàng Google có 10841 ứng dụng và 13 hàng. Các cột tư ơng ối giúp chúng tôi điều tra bức tranh là: Ứng dụng, Danh mục, Xếp hạng, Bài đánh giá, Kích thư ớc, Lư ợt cài đặt, Loại, Giá, Xếp hạng nội dung, Thẻ loại, Android Ver.

## Xóa dữ liệu sai

Tập dữ liệu Google Play có một phần thảo luận chuyên dụng và chúng ta có thể thấy rằng một trong các cuộc thảo luận nêu ra lỗi ối với hàng 10472. Hãy in hàng này và so sánh nó với tiêu đề và một hàng khác xem có đúng không.

Trong [10]: `print(android_data[10472])`  
`print('\n')`  
`print(android_header)`  
`print('\n')`  
`print(android_data[0])`

```
['Khung ảnh màn hình cảm ứng WI-Fi Life Made', '1.9', '19', '3.0M', '1.000+', 'F ree', '0', 'Mọi
ngư ời', '', 'Ngày 11 tháng 2, 2018', '1.0.19', '4.0 trở lên']
```

```
['Ứng dụng', 'Danh mục', 'Xếp hạng', 'Bài đánh giá', 'Kích thư ớc', 'Số lư ợt cài đặt', 'Loại', 'Giá',
'Xếp hạng nội dung', 'Thẻ loại', 'Cập nhật lần cuối', 'Phiên bản hiện tại', 'Phiên bản Android']
```

```
['Trình chỉnh sửa ảnh & Máy ảnh kẹo & Lư ới & Sủ lư u niệu', 'ART_AND_DESIGN', '4.1', '159',
'19M', '10.000+', 'Miễn phí', '0', 'Mọi ngư ời', 'Nghệ thuật & Thiết kế ', 'Ngày 7 tháng 1 năm
2018', '1.0.0', '4.0.3 trở lên']
```

Hàng 10472 tư ơng ứng với ứng dụng Life Made WI-Fi Touchscreen Photo Frame và chúng ta có thể thấy rằng xếp hạng là 19. Điều này rõ ràng bị tắt vì xếp hạng tối đa cho ứng dụng Google Play là 5 (như đã đề cập trong phần thảo luận, điều này vẫn đề là do thiếu giá trị trong cột 'Danh mục'). Do đó, chúng tôi sẽ xóa hàng này.

```
Trong [11]: print(len(android_data)) del
           android_data[10472]          # không chạy cái này nhiều hơn một lần
           print(len(android_data))
           print(android_data[10472])

10841
10840

['osmino Wi-Fi: WiFi miễn phí', 'CÔNG CỤ', '4.2', '134203', '4.1M', '10.000.000+', 'Miễn phí',
 '0', 'Mọi ngư ời', 'Công cụ' , 'ngày 7 tháng 8 năm 2018', '6.06.14', '4.4 trở lên']
```

## Xóa các mục nhập trùng lặp

Phần một

Nếu khám phá bộ dữ liệu Google Play đủ lâu, chúng tôi sẽ thấy rằng một số ứng dụng có nhiều mục nhập. Chẳng hạn, ứng dụng Instagram có bốn mục:

```
Trong [12]: cho ứng dụng trong android_data:
           name=app[0]
           if name == 'Instagram':
               print(app)
```

```
['Instagram', 'XÃ HỘI', '4.5', '66577313', 'Khác nhau tùy theo thiết bị', '1.000.000,0 00+', 'Miễn
 phí', '0', 'Thanh thiếu niên', 'Xã hội', '31 tháng 7 , 2018', 'Khác nhau tùy theo thiết bị',
 'Khác nhau tùy theo thiết bị']
['Instagram', 'XÃ HỘI', '4.5', '66577446', 'Khác nhau tùy theo thiết bị', '1.000.000,0 00+', 'Miễn
 phí', '0', 'Thanh thiếu niên', 'Xã hội', '31 tháng 7 , 2018', 'Khác nhau tùy theo thiết bị',
 'Khác nhau tùy theo thiết bị']
['Instagram', 'XÃ HỘI', '4.5', '66577313', 'Khác nhau tùy theo thiết bị', '1.000.000,0 00+', 'Miễn
 phí', '0', 'Thanh thiếu niên', 'Xã hội', '31 tháng 7 , 2018', 'Khác nhau tùy theo thiết bị',
 'Khác nhau tùy theo thiết bị']
['Instagram', 'XÃ HỘI', '4.5', '66509917', 'Khác nhau tùy theo thiết bị', '1.000.000,0 00+', 'Miễn
 phí', '0', 'Thanh thiếu niên', 'Xã hội', '31 tháng 7 , 2018', 'Khác nhau tùy theo thiết bị',
 'Khác nhau tùy theo thiết bị']
```

Tổng cộng, có 1.181 trư ờng hợp một ứng dụng xuất hiện nhiều lần:

```
Trong [13]: Duplicate_apps=[]
            unique_apps=[] cho
            hàng trong android_data:
                name=row[0] nếu
                tên trong unique_apps:
                    Duplicate_apps.append(name) other:

                    unique_apps.append(tên)

            print('Số lượng ứng dụng trùng lặp là', len(duplicate_apps)) print('\n')
            print('Ví dụ
            về các ứng dụng trùng lặp:', Duplicate_apps[:10])
```

Số lượng ứng dụng trùng lặp là 1181

Ví dụ về các ứng dụng trùng lặp: ['Quick PDF Scanner + OCR FREE', 'Box', 'Google My Business', 'ZOOM Cloud Meets', 'join.me - Simple Meets', 'Box', 'Zene Fit', 'Quảng cáo Google', 'Google Doanh nghiệp của tôi', 'Slack']

Chúng tôi không muốn đếm một số ứng dụng nhiều lần khi phân tích dữ liệu, vì vậy chúng tôi cần xóa các mục nhập trùng lặp và chỉ giữ lại một mục nhập cho mỗi ứng dụng. Một điều chúng tôi có thể làm là loại bỏ các hàng trùng lặp một cách ngẫu nhiên, nhưng chúng tôi có thể tìm ra cách tốt hơn.

Nếu bạn kiểm tra các hàng mà chúng tôi đã in hai ô ở trên cho ứng dụng Instagram, sự khác biệt chính xảy ra ở vị trí thứ tư của mỗi hàng, tương ứng với số lượng đánh giá. Các con số khác nhau cho thấy dữ liệu được thu thập tại các thời điểm khác nhau. Chúng ta có thể sử dụng điều này để xây dựng một tiêu chí để giữ hàng. Chúng tôi sẽ không xóa các hàng một cách ngẫu nhiên mà thay vào đó, chúng tôi sẽ giữ các hàng có số lượng bài đánh giá cao nhất vì số bài đánh giá càng cao thì xếp hạng càng đáng tin cậy.

Để làm được điều đó, chúng tôi sẽ:

- Tạo một từ điển trong đó mỗi khóa là một tên ứng dụng duy nhất và giá trị là số lượng bài đánh giá cao nhất về ứng dụng đó
- Sử dụng từ điển để tạo tập dữ liệu mới, tập dữ liệu này sẽ chỉ có một mục nhập cho mỗi ứng dụng (và chúng tôi chỉ chọn những ứng dụng có số lượng bài đánh giá cao nhất)

Hãy bắt đầu bằng cách xây dựng từ điển.

```
Trong [14]: reviews_max={}
cho hàng trong android_data:
    name=row[0]
    n_reviews=(row[3]) nếu
    tên trong reviews_max và reviews_max[name]<n_reviews:
        reviews_max[name]=n_reviews
    tên elif không có trong reviews_max:
        reviews_max[name]=n_reviews

in(len(reviews_max))
```

9659

```
Trong [15]: print('Độ dài dự kiến:', len(android_data) - 1181) print('Độ dài
thực tế:', len(reviews_max))
```

Độ dài dự kiến: 9659

Chiều dài thực tế: 9659

Bây giờ, hãy sử dụng từ điển reviews\_max để loại bỏ các bản sao. Đối với các trường hợp trùng lặp, chúng tôi sẽ chỉ giữ lại các mục có số lượng đánh giá cao nhất. Trong ô mã bên dưới:

- Chúng tôi bắt đầu bằng cách khởi tạo hai danh sách trống, android\_clean và already\_added.
- Chúng tôi lặp qua tập dữ liệu Android và cho mỗi lần lặp lại:
  - Chúng tôi cô lập tên của ứng dụng và số lượng bài đánh giá.
  - Chúng tôi thêm hàng (ứng dụng) hiện tại vào danh sách android\_clean và tên ứng dụng (tên) vào danh sách đã
    - được thêm vào nếu: Số lượng bài đánh giá của ứng dụng hiện tại khớp với số bài đánh giá của ứng dụng đó như được mô tả trong từ điển
    - reviews\_max; và Tên của ứng dụng chưa có trong danh sách đã được thêm vào. Chúng tôi cần thêm điều kiện bổ sung này để tính đến những trường hợp trong đó số lượng bài đánh giá cao nhất của một ứng dụng trùng lặp giống nhau cho nhiều mục (ví dụ: ứng dụng Box có ba mục và số bài đánh giá là như nhau) . Nếu chúng tôi chỉ kiểm tra reviews\_max[name] == n\_reviews, chúng tôi vẫn sẽ nhận được các mục nhập trùng lặp cho một số ứng dụng.

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

cho ứng dụng trong android_data:
    name = app[0]
    n_reviews = (app[3])

    if (reviews_max[name] == n_reviews) và (tên chưa có trong already_added):
        android_clean.append(app)
        already_added.append(name) # đảm bảo rằng đây là bên trong khối if
```

Bây giờ, hãy nhanh chóng khám phá tập dữ liệu mới và xác nhận rằng số hàng là 9.659.

Trong [17]: `explorer_data(android_clean, 0, 3, True)`

```
['Trình chỉnh sửa ảnh & Máy ảnh kẹo & Lư ới & Sổ lư u niệu', 'ART_AND_DESIGN', '4.1', '159',
'19M', '10.000+', 'Miễn phí', '0', 'Mọi ngư ời', 'Nghệ thuật & Thiết kế ', 'Ngày 7 tháng 1 năm
2018', '1.0.0', '4.0.3 trở lên']
```

```
['U Launcher Lite - Chủ đề thú vị trực tiếp MIỄN PHÍ , Ấn ứng dụng', 'ART_AND_DESIGN', '4.
7', '87510', '8,7 triệu', '5.000.000+', 'Miễn phí', '0', 'Mọi ngư ời', 'Nghệ thuật & Thiết kế',
'Ngày 1 tháng 8 năm 2018', '1.2.4', '4.0 .3 trở lên']
```

```
['Phác thảo - Vẽ & Sơ n', 'ART_AND_DESIGN', '4.5', '215644', '25M', '50,000,0 00+', 'Free', '0',
'Teen', 'Art & Design' , 'Ngày 8 tháng 6 năm 2018', 'Khác nhau tùy theo thiết bị', '4.2 trở lên']
```

Số hàng: 9659

Số cột: 13

Chúng tôi có 9659 hàng, đúng như mong đợi.

## Xóa ứng dụng không phải tiếng Anh

### Phần một

Nếu khám phá đủ bộ dữ liệu, bạn sẽ thấy rằng có những ứng dụng không hướng đến đối tượng nói tiếng Anh. Dưới đây là một số ví dụ:

Trong [18]: 

```
print(ios_data[814][1])
print(ios_data[6732][1])
print('\n')
print(android_clean[4412][0])
```

Filterra - Trình chỉnh sửa ảnh, Hiệu ứng cho ảnh /

Wowkwis aq Ka'qaquj

Vì vậy, chúng ta cần phải gỡ bỏ các ứng dụng này.

Một cách để giải quyết vấn đề này là xóa từng ứng dụng có tên chứa ký hiệu không được sử dụng phổ biến trong văn bản tiếng Anh – Văn bản tiếng Anh thường bao gồm các chữ cái trong bảng chữ cái tiếng Anh, các số gồm các chữ số từ 0 đến 9, dấu chấm câu (., ! , ?, ; , v.v.) và các ký hiệu khác (+, \*, / , v.v.). Tất cả các ký tự dành riêng cho văn bản tiếng Anh này đều được mã hóa bằng tiêu chuẩn ASCII.

Mỗi ký tự ASCII có một số từ ứng từ 0 đến 127 được liên kết với nó và chúng ta có thể tận dụng lợi thế đó để xây dựng một hàm kiểm tra tên ứng dụng và cho chúng ta biết liệu ứng dụng đó có chứa các ký tự không phải ASCII hay không.

Chúng tôi đã xây dựng chức năng này bên dưới và chúng tôi sử dụng chức năng `ord()` tích hợp để tìm ra số mã hóa tương ứng của mỗi ký tự.

```
Trong [19]: def is_english(string): cho
            ký tự trong chuỗi:
            nếu ord(ký tự) > 127:
                trả về Sai
            trả về Đúng
print(is_english('strg'))
print(is_english('Instagram'))
```

SAI

ĐÚNG VẬY

Chức năng này có vẻ hoạt động tốt, nhưng một số tên ứng dụng tiếng Anh sử dụng biểu tượng cảm xúc hoặc các ký hiệu khác (™, – (em dash), - (en dash), v.v.) nằm ngoài phạm vi ASCII. Do đó, chúng tôi sẽ xóa các ứng dụng hữu ích nếu chúng tôi sử dụng chức năng này ở dạng hiện tại.

```
Trong [20]: print(is_english(' Bộ Office miễn phí Docs To Go™'))
print(is_english('Instachat '))
print(ord('™'))
print(ord(' '))
```

SAI

SAI

8482

128540

## Phần hai

Chúng tôi thay đổi chức năng để các ứng dụng có tối đa 3 ký tự nằm ngoài phạm vi ASCII vẫn có thể được liệt kê là ứng dụng nói tiếng Anh



```
Trong [21]: def is_english(string):
            non_ascii=0

            cho ký tự trong chuỗi:
                nếu ord(ký tự)>127:
                    non_ascii+=1

            nếu non_ascii>3:
                trả về Sai
            khác:
                trả về Đúng

print(is_english('Instachat '))
print(is_english(' Bộ Office miễn phí Docs To Go™'))
print(is_english('PPS - 2 '))
```

ĐÚNG VẬY

ĐÚNG VẬY

SAI

Chức năng này vẫn chưa hoàn hảo và rất ít ứng dụng không phải tiếng Anh có thể vượt qua bộ lọc của chúng tôi, nhưng điều này có vẻ đủ tốt vào thời điểm này trong phân tích của chúng tôi – chúng tôi không nên dành quá nhiều thời gian cho việc tối ưu hóa vào thời điểm này.

Bên dư ới, chúng tôi sử dụng hàm `is_english()` để lọc ra các ứng dụng không phải tiếng Anh cho cả hai bộ dữ liệu:

```
Trong [22]: ios_english=[]

cho hàng trong ios_data:
    string=row[1]
    if is_english(string)==True:
        ios_english.append(row)

explorer_data(ios_english,0,3, True)

['284882215', 'Facebook', '389879808', 'USD', '0.0', '2974676', '212', '3.5', '3.5', '95.0',
'4+', 'Mạng xã hội', '37', '1', '29', '1']

['389801252', 'Instagram', '113954816', 'USD', '0.0', '2161558', '1289', '4.
5', '4.0', '10.23', '12+', 'Ảnh & Video', '37', '0', '29', '1']

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

Số hàng: 6183

Số cột: 16

Trong [23]:

```
android_english=[]

cho hàng trong android_clean:
    string=row[0] if
    is_english(string)==True:
        android_english.append(row)
explorer_data(android_english,0,3,True)
```

```
['Trình chỉnh sửa ảnh & Máy ảnh kẹo & Lư ới & Sủ lư u niệ', 'ART_AND_DESIGN', '4.1', '159',
'19M', '10.000+', 'Miễn phí', '0', 'Mọi ngư ời', 'Nghệ thuật & Thiết kế ', 'Ngày 7 tháng 1 năm
2018', '1.0.0', '4.0.3 trở lên']
```

```
['U Launcher Lite - Chủ đề thú vị trực tiếp MIỄN PHÍ , Ẩn ứng dụng', 'ART_AND_DESIGN', '4.
7', '87510', '8,7 triệu', '5.000.000+', 'Miễn phí', '0', 'Mọi ngư ời', 'Nghệ thuật & Thiết kế',
'Ngày 1 tháng 8 năm 2018', '1.2.4', '4.0 .3 trở lên']
```

```
['Phác thảo - Vẽ & Sơ n', 'ART_AND_DESIGN', '4.5', '215644', '25M', '50,000,0 00+', 'Free', '0',
'Teen', 'Art & Design' , 'Ngày 8 tháng 6 năm 2018', 'Khác nhau tùy theo thiết bị', '4.2 trở lên']
```

Số hàng: 9614

Số cột: 13

Trong [24]:

```
in(ios_header)

['id', 'track_name', 'size_bytes', 'currency', 'price', 'rating_count_tot', 'rating_count_ver',
'user_rating', 'user_rating_ver', 'ver', 'cont_rating', 'prime_genre', ' sup_devices.num',
'ipadSc_urls.num', 'lang.num', 'vpp_lic']
```

Chúng ta có thể thấy rằng chúng ta còn lại 9614 ứng dụng Android và 6183 ứng dụng iOS.

## Cô lập các ứng dụng miễn phí

Như chúng tôi đã đề cập trong phần giới thiệu, chúng tôi chỉ xây dựng các ứng dụng miễn phí để tải xuống và cài đặt, đồng thời nguồn doanh thu chính của chúng tôi bao gồm quảng cáo trong ứng dụng. Bộ dữ liệu của chúng tôi chứa cả ứng dụng miễn phí và không miễn phí và chúng tôi sẽ chỉ cần tách riêng các ứng dụng miễn phí để phân tích. Dưới đây, chúng tôi cô lập các ứng dụng miễn phí cho cả hai bộ dữ liệu của mình.

Trong [31]:

```
free_ios=[]
cho ứng dụng bằng ios_english:
    giá=ứng dụng[4]
    nếu giá == '0,0':
        free_ios.append(ứng dụng)
in(len(free_ios))
```

3222

```
Trong [32]: free_android=[] cho ứng  
dụng bằng android_english: price = (apps[7])  
    if price=='0':  
        free_android.append(apps)  
  
in(len(free_android))
```

8862

Chúng tôi còn lại 8862 ứng dụng Android và 3222 ứng dụng iOS, đủ cho phân tích của chúng tôi.

## Các ứng dụng phổ biến nhất theo thể loại

### Phần một

Như chúng tôi đã đề cập trong phần giới thiệu, mục tiêu của chúng tôi là xác định loại ứng dụng có khả năng thu hút nhiều người dùng hơn vì doanh thu của chúng tôi bị ảnh hưởng nhiều bởi số lượng người sử dụng ứng dụng của chúng tôi. ứng dụng.

Để giảm thiểu rủi ro và chi phí chung, chiến lược xác thực ý tưởng ứng dụng của chúng tôi bao gồm ba bước:

1. Tạo một phiên bản Android tối thiểu của ứng dụng và thêm nó vào Google Play.
2. Nếu ứng dụng nhận được phản hồi tốt từ người dùng, chúng tôi sẽ phát triển thêm.
3. Nếu ứng dụng có lãi sau sáu tháng, chúng tôi cũng xây dựng phiên bản iOS của ứng dụng và thêm nó vào đến Cửa hàng ứng dụng.

Vì mục tiêu cuối cùng của chúng tôi là thêm ứng dụng trên cả App Store và Google Play, nên chúng tôi cần tìm hồ sơ ứng dụng thành công trên cả hai thị trường. Ví dụ: một hồ sơ có thể hoạt động tốt cho cả hai thị trường có thể là một ứng dụng năng suất sử dụng trò chơi hóa.

Hãy bắt đầu phân tích bằng cách tìm hiểu các thể loại phổ biến nhất cho từng thị trường. Đối với điều này, chúng ta sẽ tạo một bảng tần suất cho cột prime\_genre của tập dữ liệu App Store và các cột Thể loại và Danh mục của tập dữ liệu Google Play.

### Phần hai

Chúng ta sẽ xây dựng hai hàm mà chúng ta có thể sử dụng để phân tích các bảng tần suất:

- Một chức năng để tạo bảng tần suất hiển thị tỷ lệ phần trăm
- Một chức năng khác mà chúng ta có thể sử dụng để hiển thị tỷ lệ phần trăm theo thứ tự giảm dần

```

Trong [33]: def freq_table(dataset, index): table={}
            total=0

            cho hàng trong tập dữ liệu:
                tổng+=1
                giá trị=hàng[chỉ mục]
                nếu giá trị trong bảng:
                    bảng[giá trị] += 1 khác:

                    bảng [giá trị] = 1

            table_percentage = {} cho
            khóa trong bảng:
                percent=(table[key]/total)*100
                table_percentage[key]=percentage

            trả về table_percentage

```

```

Trong [44]: freq_table(free_ios, 11)

```

```

Ra[44]: {'Mạng xã hội': 3.2898820608317814, 'Ảnh & Video':
4.9658597144630665, 'Trò chơi': 58.16263190564867,
'Âm nhạc': 2.04841713222160147, 'Tham
khảo': 0.558659 2178770949, 'Sức khỏe
& Thể hình': 2.0173805090006205, 'Thời
tiết' : 0.8690254500310366, 'Tiện ích':
2.5139664804469275, 'Du lịch':
1.2414649286157666, 'Mua sắm':
2.60707635009311, 'Tin tức':
1.3345747982619491, 'N hàng không':
0.186219739292365, 'Phong cách sống':
1.5828677839851024, 'Giải trí':
7.883302296710118, 'Đồ ăn thức uống' :
0.8069522036002483, 'Thể thao': 2.1415270018621975,
'Sách': 0.4345127250155183, 'Tài chính':
1.1173184357541899, 'Giáo dục':
3.662321539416512, ' Năng suất':
1.7380509000620732, 'Kinh doanh':
0.5276225946617008, 'Danh mục':
0.12414649286157665, 'Y tế': 0.186219739292365 }

```

Chúng tôi tiếp tục thêm chức năng display\_table.

```
Trong [39]: 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) cho mục nhập
            trong table_sorted: print(entry[1],
                                      '...', entry[0])
```

Một phần ba

Chúng tôi bắt đầu bằng cách kiểm tra bảng tần suất cho cột prime\_genre của dữ liệu App Store bộ.

```
Trong [49]: in(ios_header)
            in(android_header)
```

```
['id', 'track_name', 'size_bytes', 'currency', 'price', 'rating_count_tot', 'rating_count_ver',
 'user_rating', 'user_rating_ver', 'ver', 'cont_rating', 'prime_genre', 'sup_devices.num',
 'ipadSc_urls.num', 'lang.num', 'vpp_lic']
['Ứng dụng', 'Danh mục', 'Xếp hạng', 'Đánh giá', 'Kích thước', 'Số lượt cài đặt', 'Loại', 'Giá',
 'Xếp hạng nội dung', 'Thể loại', 'Cập nhật lần cuối', 'Phiên bản hiện tại', 'Phiên bản Android']
```

```
Trong [51]: display_table(free_ios, -5) #prime_genre
```

```
Trò chơi : 58.16263190564867
Giải trí : 7.883302296710118 Ảnh & Video :
4.9658597144630665
Giáo dục : 3.662321539416512 Mạng xã
hội : 3.2898820608317814 Mua sắm : 2.60707635009311
Tiện ích : 2.5139664804469275
```

```
Thể thao : 2.1415270018621975 Âm
nhạc : 2.0484171322160147 Sức khỏe
& Thể hình : 2.0173805090006205 Năng suất :
1.7380509000620732 Lối sống :
1.5828677839851024 Tin tức : 1.334574798
2619491 Du lịch :
1.2414649286157666 Tài chính :
1.1173184357541899
Thời tiết : 0.8690254500310366 Đồ ăn & Đồ
uống : 0.8069522036002483 Tham khảo :
0.5586592178770949 Kinh doanh : 0.5276225946617008
```

```
B k 0 4345127250155183
```

Có thể thấy trong số các ứng dụng tiếng Anh miễn phí trên iOS store, hơn một nửa (58,16%) là game. Tiếp theo là các ứng dụng giải trí chiếm gần 8%, ứng dụng ảnh và video chiếm gần 5%. Chỉ 3,66% ứng dụng được thiết kế cho giáo dục, tiếp theo là ứng dụng mạng xã hội chiếm 3,29% ứng dụng trong bộ dữ liệu của chúng tôi.

Ấn tượng đầu tiên là App Store (ít nhất là một phần chứa ứng dụng tiếng Anh miễn phí) bị chi phối bởi các ứng dụng được thiết kế để giải trí như trò chơi, giải trí, ảnh & video, trong khi các ứng dụng có mục đích thực tế như giáo dục, mua sắm, năng suất, sức khỏe và thể dục lại nhiều hơn hiếm.

Tuy nhiên, số lượng ứng dụng thú vị nhiều nhất không có nghĩa là loại này có số lượng người dùng nhiều nhất. Nhu cầu có thể không giống như lời đề nghị.

Hãy tiếp tục bằng cách kiểm tra cột Thể loại và Danh mục của tập dữ liệu Google Play (hai cột dư thừa như có liên quan với nhau)

Trong [47]: display\_table(free\_android, 1) #Category

```
GIA ĐÌNH : 18.934777702550214
TRÒ CHƠI : 9.693071541412774 CÔNG CỤ :
8.451816745655607
KINH DOANH : 4.5926427443015125
DẠNG SỐNG : 3.9043105393816293
NĂNG SUẤT : 3.8930264048747465 TÀI CHÍNH :
3.7011961182577298
Y TẾ : 3.5206499661475967
THỂ THAO : 3.39652448657188
CÁ NHÂN HÓA : 3.3175355450236967 TRUYỀN THÔNG :
3.238546603475513
HEALTH_AND_FITNESS : 3.080568720379147 ẢNH : 2.945159106296547

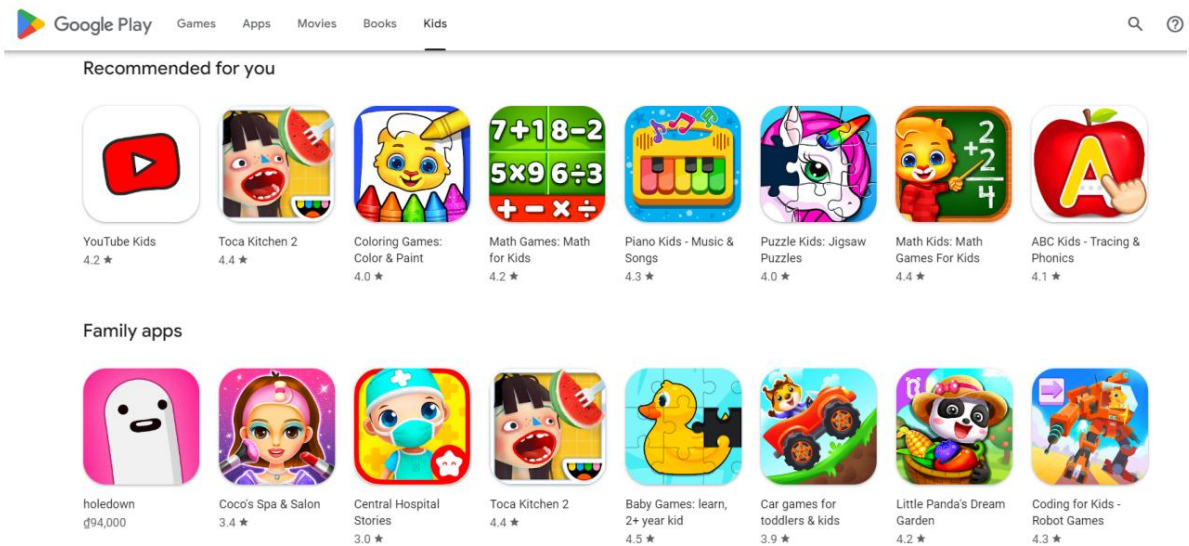
NEWS_AND_MAGAZINES : 2.798465357707064 XÃ HỘI : 2.663055743624464

TRAVEL_AND_LOCAL : 2.335815842924848 MUA SẮM :
2.2455427668697814
BOOKS_AND_REFERENCE : 2.143985556307831 NGÀY : 1.8618821936357481

NGƯỜI CHƠI VIDEO 1 7941773865944481
```

Nhìn vào kết quả, chúng ta có thể thấy rằng ứng dụng Gia đình là loại phổ biến nhất trên cửa hàng Google Play với hơn 19%, tiếp theo là Trò chơi với gần 10% và Công cụ với gần 9%.

So với cửa hàng iOS, bối cảnh có vẻ khác. Tuy nhiên, nếu tìm hiểu sâu hơn, chúng ta thấy hầu hết các ứng dụng trong danh mục Gia đình trên Google Play đều là trò chơi dành cho trẻ em.



[. \(https://play.google.com/store/apps/category/FAMILY2hl=vi\)](https://play.google.com/store/apps/category/FAMILY2hl=vi) .

Mặc dù vậy, các ứng dụng thực tế dường như có đại diện tốt hơn trên Google Play so với Ứng dụng Cửa hàng. Hình ảnh này cũng được xác nhận bởi bảng tần suất mà chúng ta thấy cho cột Thể loại:

```
Trong [54]: display_table(free_android, -4) #Genres
```



Công cụ : 8.440532611148726 Giải trí : 6.070864364703228

Giáo dục : 5,348679756262695 Kinh doanh : 4,5926427443015125 Năng suất : 3,8930264048747465

Lối sống : 3,8930264048747465

Tài chính : 3,7011961182577298 Y tế : 3,520649966147 5967 Thể thao : 3.4642292936131795

Cá nhân hóa : 3.3175355450236967

Truyền thông : 3.238546603475513 Hành động : 3.1031369893929135

Sức khỏe & Thể hình : 3.080568720379147 Nhiếp ảnh : 2.945159106296547

Tin tức & Tạp chí : 2.798465357707064 Xã hội : 2.66305574362 4464

Du lịch & Địa phương : 2.324531708417964 Mua sắm : 2.2455427668697814

Sách & Tài liệu tham khảo : 2.143985556307831

Mô phỏng : 2.0424283457458814

Hẹn hò : 1.8618821936357481

Trò chơi điện tử : 1.850598 0591288649

Trình phát và chỉnh sửa video: 1.7716091175806816

Thông thư ờng: 1.7490408485669149

Bản đồ & Điều hướng: 1.399232678853532

Đồ ăn & Đồ uống: 1.2412547957571656

Xếp hình : 1.128413450688332 Đua xe : 0.9930038366057323

Nhập vai : 0.9365831640713158

Thư viện & Trình diễn : 0.9365831640713158

Ô tô & Xe cộ : 0.9252990295644324

Chiến lược : 0.9140148950575491

Nhà & Trang chủ : 0.8237418190024826

Thời tiết : 0.8011735499887158

Sự kiện : 0,7109004739336493

Phiêu lưu : 0,6770480704129994

Truyện tranh : 0,6093432633716994

Làm đẹp : 0,598059128864816

Nghệ thuật & Thiết kế : 0,598059128864816

Nuôi dạy con cái : 0,4965019183 0286616

Thẻ : 0.4400812457684496

Sòng bạc : 0.4287971112615662

Thông tin bên lề : 0,41751297675468296

Giáo dục;Giáo dục : 0,3949447077409162

Giáo dục : 0,3723764387271496

Hội đồng quản trị : 0,3723764387271496

Giáo dục;Giáo dục : 0,3385240352064997

Word : 0,2595350936583164

Bình thư ờng;Chơi giả vờ : 0,23696682464454977

Âm nhạc : 0,2031144211238998

Puzzle;Trò chơi trí tuệ : 0.18054615211013314

Đua xe;Hành động & Phiêu lưu : 0.16926201760324985

Giải trí;Âm nhạc & Video : 0.16926201760324985

Thông thư ờng;Trò chơi trí não : 0.13540961408259986

Thông thư ờng;Hành động & Phiêu lưu : 0.1354096 1408259986

Arcade;Hành động & Phiêu lưu : 0.12412547957571654

Hành động;Hành động & Phiêu lưu : 0.1015572105619499

Giáo dục;Chơi i giả vờ: 0,09027307605506657 Trò chơi trí tuệ: 0,09027307605506657 Mô phỏng;Hành động & Phiêu lưu: 0,07898894154818326 Nuôi dạy con cái;Giáo dục: 0,07898894154818326 Giải trí;Trò chơi trí tuệ: 0,078988 94154818326 Nuôi dạy con cái;Âm nhạc & Video : 0,06770480704129993 Giáo dục;Trò chơi trí tuệ: 0,06770480704129993 Thông thư ờng;Sáng tạo: 0,06770480704129993 Nghệ thuật & Thiết kế;Sáng tạo : 0,06770480704129993 Giáo dục;Chơi i đóng vai : 0,056420672534416606 Đóng vai;Chơi i đóng vai : 0,045136538027533285 Giáo dục;Sáng tạo : 0,045136538027533285 Đóng vai;Hành động & Phiêu lưu : 0,033 852403520649964 Giải i;Hành động & Phiêu lưu : 0,033852403520649964 Giải trí;Sáng tạo : 0,033852403520649964 Giải trí;Hành động & Phiêu lưu : 0,033852403520649964 Giáo dục;Sáng tạo : 0,033852403520649964 Giáo dục;Hành động & Phiêu lưu : 0,033852403520649964 Giáo dục;Âm nhạc & Video : 0,033852403520649964 Giáo dục;Trò chơi trí tuệ : 0,033852403520649964 Giáo dục;Hành động & Phiêu lưu : 0,0338 52403520649964 Phiêu lưu;Hành động & Phiêu lưu : 0,033852403520649964 Trình chỉnh sửa & Trình phát video;Âm nhạc & Video : 0,022568269013766643 Thể thao ;Hành động & Phiêu lưu : 0,022568269013766643 Mô phỏng;Chơi i giả vờ : 0,022568269013766643 Xếp hình;Sáng tạo : 0,022568269013766643 Âm nhạc;Âm nhạc & Video : 0,022568269013766643 Giải trí;Chơi i giả vờ : 0,022 568269013766643 Bình thư ờng;Giáo dục: 0,022568269013766643 Trò chơi i cờ;Hành động & Phiêu lưu: 0,022568269013766643 Trình phát & biên tập video;Sáng tạo: 0,011284134506883321 Thông tin bên lề;Giáo dục : 0,011284134506883321 Du lịch & Địa phư ơ ng;Hành động & Phiêu lưu : 0,011284134506883321 Công cụ;Giáo dục : 0,011284134506883321 Chiến lư ợc;Giáo dục : 0,011284134506883321 Chiến lư ợc;Sáng tạo : 0,01 1284134506883321 Chiến lư ợc;Hành động & Phiêu lưu : 0,011284134506883321 Mô phỏng;Giáo dục : 0,011284134506883321 Nhập vai;Trò chơi i trí não : 0,011284134506883321 Đua xe; Chơi i đóng vai : 0,011284134506883321 Xếp hình;Giáo dục : 0,011284134506883321 Nuôi dạy con cái;Trò chơi i trí tuệ : 0,011284134506883321 Âm nhạc & Âm thanh;Nhạc & Video : 0,011284134506883321 Lối sống;Chơi i đóng vai : 0,01 1284134506883321 Lối sống;Giáo dục : 0,011284134506883321 Sức khỏe & Thể hình;Giáo dục : 0,011284134506883321 Sức khỏe & Thể hình;Hành động & Phiêu lưu : 0,011284134506883321 Giải trí;Giáo dục : 0,011284134506883321 Truyền thông;Sáng tạo : 0,011284134506883321 Truyện tranh;Sáng tạo : 0,011284134506883321 Thông thư ờng;Âm nhạc & Video : 0,0112841345068 83321 Thẻ;Trò chơi i trí tuệ: 0,011284134506883321 Thẻ;Hành động & Phiêu lưu: 0,011284134506883321 Sách & Tài liệu tham khảo;Giáo dục: 0,011284134506883321 Nghệ thuật & Thiết kế;Chơi i đóng vai: 0.011284134506883321 Nghệ thuật & Thiết kế;Hành động & Phiêu lưu : 0.011284134506883321

Arcade;Chơi giả vờ : 0,011284134506883321 Phiêu lưu;Giáo dục :  
0,011284134506883321

Sự khác biệt giữa cột Thể loại và Thể loại không rõ ràng, nhưng có một điều chúng ta có thể nhận thấy là cột Thể loại chi tiết hơn nhiều (nó có nhiều danh mục hơn). Chúng tôi chỉ đang tìm kiếm bức tranh lớn hơn vào lúc này, vì vậy chúng tôi sẽ chỉ làm việc với cột Danh mục trong tương lai.

Cho đến thời điểm này, chúng tôi nhận thấy rằng App Store bị chi phối bởi các ứng dụng được thiết kế để giải trí, trong khi Google Play thể hiện bối cảnh cân bằng hơn về cả ứng dụng thiết thực và giải trí. Bây giờ chúng tôi muốn có ý tưởng về loại ứng dụng có hầu hết người dùng.

## Ứng dụng phổ biến nhất theo thể loại trên App Store

Một cách để tìm hiểu thể loại nào phổ biến nhất (có nhiều người dùng nhất) là tính số lượt cài đặt trung bình cho mỗi thể loại ứng dụng. Đối với tập dữ liệu Google Play, chúng tôi có thể tìm thấy thông tin này trong cột Lượt cài đặt, nhưng đối với tập dữ liệu App Store thì thông tin này bị thiếu.

Như một giải pháp thay thế, chúng tôi sẽ lấy tổng số xếp hạng của người dùng làm proxy mà chúng tôi có thể tìm thấy trong ứng dụng `rating_count_tot`.

Dưới đây, chúng tôi tính toán số lượng xếp hạng trung bình của người dùng cho mỗi thể loại ứng dụng trên App Store:

Đầu tiên, chúng tôi cô lập các ứng dụng theo thể loại. Sau đó cộng xếp hạng của người dùng cho các ứng dụng thuộc thể loại đó. Chia tổng cho số lượng ứng dụng thuộc thể loại đó.

```
Trong [56]: thể_loại_ios = freq_table(free_ios, -5)
```

```
Trong [57]: cho thể loại trong thể loại_ios:
            tổng = 0

            len_genre=0
            cho ứng dụng trong
                free_ios: thể loại_app =
                ứng dụng[-5] if thể loại_app
                    == thể loại: n_rating =
                    float(app[5]) tổng
                    cộng += n_rating
            len_genre += 1 average_ratings = tổng số /
            len_genre print(thể loại, ': ', trung bình_xếp hạng)
```

Mạng xã hội : 71548.34905660378 Ảnh & Video :  
28441.54375 Trò chơi : 22788.6696905016

Âm nhạc : 57326.530303030304 Tham khảo :  
74942.11111111111 Sức khỏe & Thể hình :  
23298.015384615384 Thời tiết : 52279.892857142855

Tiền ích : 18684.456790123455 Du lịch :  
28243.8 Mua sắm :  
26919.690476190477 Tin tức :  
21248.023255813954

Điều hướng : 86090.33333333333 Lối  
sống : 16485.764705882353 Giải trí :  
14029.830708661417 Đồ ăn thức uống :  
33333.92307692308

Thể thao : 23008.898550724636 Sách :  
39758.5 Tài chính :  
31467.944444444445 Giáo dục :  
7003.983050847458

Năng suất : 21028.410714285714 Kinh doanh :  
7491.117647058823 Danh mục : 4004.0 Y  
tế : 612.0

Trung bình, các ứng dụng điều hướng có số lượng đánh giá cao nhất của người dùng, nhưng con số này bị ảnh hưởng nhiều bởi Waze và Google Maps, những ứng dụng có gần nửa triệu đánh giá của người dùng cùng nhau:

```
Trong [58]: cho ứng dụng trong
            free_ios: if app[-5] == 'Navigation':
                print(app[1], ': ', app[5]) # in tên và số xếp hạng
```

Waze - Điều hướng GPS, Bản đồ & Giao thông thời gian thực : 345046  
Google Maps - Điều hướng & Chuyển tuyến : 154911  
Geocaching® : 12811  
CoPilot GPS - Điều hướng ô tô & Bản đồ ngoại tuyến : 3582  
ImmobilienScout24: Tìm kiếm Bất động sản ở Đức : 187  
Tìm kiếm tuyến đường sắt : 5

Mô hình tư duy tự cũng áp dụng cho các ứng dụng mạng xã hội, trong đó số lượng trung bình bị ảnh hưởng nhiều bởi một số gã khổng lồ như Facebook, Pinterest, Skype, v.v. số trung bình.

Mục đích của chúng tôi là tìm các thể loại phổ biến, nhưng các ứng dụng điều hướng, mạng xã hội hoặc âm nhạc có vẻ phổ biến hơn thực tế. Số lượng xếp hạng trung bình dường như bị sai lệch bởi rất ít ứng dụng có hàng trăm nghìn xếp hạng của người dùng, trong khi các ứng dụng khác có thể phải vật lộn để vượt qua ngưỡng 10.000. Chúng tôi có thể có được hình ảnh tốt hơn bằng cách xóa các ứng dụng cực kỳ phổ biến này cho từng thể loại và sau đó tính lại mức trung bình, nhưng chúng tôi sẽ để lại mức độ chi tiết này sau.

Các ứng dụng tham khảo có 74.942 xếp hạng trung bình của người dùng, nhưng thực tế là Kinh thánh và Dictionary.com đã làm sai lệch xếp hạng trung bình:

```
Trong [59]: cho ứng dụng trong free_ios:
           nếu ứng dụng[-5] == 'Tham khảo':
               in(ứng dụng[1], ':', ứng dụng[5])
```

Kinh thánh : 985920  
Dictionary.com Dictionary & Thesaurus : 200047 Dictionary.com  
Dictionary & Thesaurus for iPad : 54175 Google Translate : 26786 Muslim Pro:  
Ramadan 2017 Thời gian cầu nguyện,  
Azan, Kinh Qur'an : 18418 New Furniture Mods - Pocket Wiki & Game Tools cho Minecraft  
PC Phiên bản : 1758  
...  
Từ điển Merriam-Webster : 16849  
Bầu trời đêm : 12122  
Bản đồ thành phố cho Minecraft PE - Bản đồ tốt nhất cho Minecraft Pocket Edition (MCP  
E): 8535  
LUCKY BLOCK MOD ™ cho Minecraft PC Edition - The Best Pocket Wiki & Mods Inst Công cụ hỗ trợ: 4693  
  
GUNS MODS cho Minecraft PC Edition - Mods Tools : 1497  
Hướng dẫn về Pokémon GO - Tin tức và cheat Pokemon GO : 826  
WWDC : 762 Bản  
đồ kinh dị cho Minecraft PE - Tải xuống Bản đồ đáng sợ nhất cho Minecraft Pocket Edition (MCPE) Miễn phí : 718 VPN  
Express : 14 Kinh thực tế ảo Real Bike  
Traffic Rider : 8  
goo : 0 Jishokun-Từ điển tiếng Anh-Nhật Bản & Trình biên dịch : 0

Tuy nhiên, niche này dường như cho thấy một số tiềm năng. Một điều chúng tôi có thể làm là lấy một cuốn sách phổ biến khác và biến nó thành một ứng dụng nơi chúng tôi có thể thêm các tính năng khác ngoài phiên bản thô của cuốn sách. Điều này có thể bao gồm các trích dẫn hàng ngày từ cuốn sách, phiên bản âm thanh của cuốn sách, câu đố về cuốn sách, v.v. Ngoài ra, chúng tôi cũng có thể nhúng từ điển vào ứng dụng để người dùng không cần phải thoát khỏi ứng dụng của chúng tôi để tra cứu lên các từ trong một ứng dụng bên ngoài.

Ý tưởng này có vẻ phù hợp với thực tế là App Store bị chi phối bởi các ứng dụng giải trí. Điều này cho thấy thị trường có thể hòa trộn với các ứng dụng giải trí, điều đó có nghĩa là một ứng dụng thực tế có thể có nhiều cơ hội nổi bật hơn trong số lượng lớn ứng dụng trên App Store.

Các thể loại khác có vẻ phổ biến bao gồm thời tiết, sách, đồ ăn thức uống hoặc tài chính. Thể loại sách dường như trùng lặp một chút với ý tưởng ứng dụng mà chúng tôi đã mô tả ở trên, nhưng các thể loại khác dường như không quá thú vị đối với chúng tôi:

- Ứng dụng thời tiết – mọi người thường không dành quá nhiều thời gian trong ứng dụng và cơ hội kiếm lợi nhuận từ các phần bổ sung trong ứng dụng là thấp. Ngoài ra, việc nhận dữ liệu thời tiết trực tiếp đáng tin cậy có thể yêu cầu chúng tôi kết nối ứng dụng của mình với các API không miễn phí.
- Đồ ăn và thức uống – ví dụ ở đây bao gồm Starbucks, Dunkin' Donuts, McDonald's, v.v. Vì vậy, việc tạo ra một ứng dụng đồ ăn và thức uống phổ biến yêu cầu dịch vụ nấu ăn và giao hàng thực tế, điều này nằm ngoài phạm vi của công ty chúng tôi.
- Ứng dụng tài chính – những ứng dụng này liên quan đến ngân hàng, thanh toán hóa đơn, chuyển tiền, v.v. Xây dựng ứng dụng tài chính yêu cầu kiến thức về lĩnh vực và chúng tôi không muốn thuê chuyên gia tài chính chỉ để xây dựng ứng dụng

## Ứng dụng phổ biến nhất theo thể loại trên Google Play

Đối với thị trường Google Play, chúng tôi thực sự có dữ liệu về số lượt cài đặt, vì vậy chúng tôi có thể có được bức tranh rõ ràng hơn về mức độ phổ biến của thể loại này. Tuy nhiên, số lượt cài đặt dường như không đủ chính xác – chúng ta có thể thấy rằng hầu hết các giá trị đều là kết thúc mở (100+, 1.000+, 5.000+, v.v.):

Trong [62]: `display_table(free_android, 5) # cột Lượt cài đặt`

```
1.000.000+ : 15.741367637102236 100.000+ :
11.554953735048521 10.000.000+ :
10.516813360415256 10.000+ : 10.200857594222523
1.000+ : 8.395396073121193 100+ :
6.917174452719477

5.000.000+ : 6.838185511171294 500.000+ :
5.574362446400361 50.000+ :
4.773188896411646 5.000+ :
4.513653802753328 10+ : 3.543
2182351613632
500+ : 3.2498307379823967

50.000.000+ : 2.2906793048973144 100.000.000+ :
2.1214172872940646 50+ : 1.9183028661701647

5+ : 0,7898894154818324 1+ :
0,5077860528097494

500.000.000+ : 0,2708192281651997
1.000.000.000+ : 0,22568269013766643 0+ :
0,045136538027533285 0 :
0,011284134506883321
```

Một vấn đề với dữ liệu này là nó không chính xác. Chẳng hạn, chúng tôi không biết liệu một ứng dụng có hơn 100.000 lượt cài đặt có 100.000 lượt cài đặt, 200.000 hay 350.000. Tuy nhiên, chúng tôi không cần dữ liệu quá chính xác cho mục đích của mình – chúng tôi chỉ muốn biết thể loại ứng dụng nào thu hút nhiều người dùng nhất và chúng tôi không cần độ chính xác hoàn hảo đối với số lượng người dùng.

Chúng tôi sẽ giữ nguyên các con số, điều đó có nghĩa là chúng tôi sẽ xem xét rằng một ứng dụng có hơn 100.000 lượt cài đặt sẽ có 100.000 lượt cài đặt và một ứng dụng có hơn 1.000.000 lượt cài đặt sẽ có 1.000.000 lượt cài đặt, v.v.

Tuy nhiên, để thực hiện tính toán, chúng tôi sẽ cần chuyển đổi từng số lượt cài đặt thành float – điều này có nghĩa là chúng tôi cần xóa dấu phẩy và ký tự dấu cộng, nếu không thì chuyển đổi

```
Trong [63]: loại_android = freq_table(free_android, 1)

cho danh mục trong danh mục_android: tổng = 0 len_category
    = 0 cho ứng
    dụng trong free_android:
    danh mục_app = ứng dụng[1] if danh
        mục_app == danh mục: n_installs =
        ứng dụng[5] n_installs = n_installs.replace(',', ' ',
            ' ') n_installs =
            n_installs .replace('+', ' ') tổng cộng += float(n_installs)
            len_category += 1

    avg_n_installs = tổng số / len_category print(category, ': ',
    avg_n_installs)
```

ART\_AND\_DESIGN : 1986335.0877192982 AUTO\_AND\_VEHICLES :  
647317.8170731707 ĐẸP : 513151.88679245283

BOOKS\_AND\_REFERENCE : 8767811.894736841 KINH DOANH :  
1712290.1474201474  
TRUYỆN TRANH : 817657.2727272727 TRUYỀN THÔNG :  
38456119.167247385  
HẸN HÒ : 854028.8303030303  
GIÁO DỤC : 1820673.076923077  
GIẢI TRÍ : 11640705.88235294 SỰ KIỆN : 253542.2222222222

TÀI CHÍNH : 1387692.475609756  
FOOD\_AND\_DRINK : 1924897.7363636363 HEALTH\_AND\_FITNESS :  
4188821.9853479853 HOUSE\_AND\_HOME : 1331540.5616438356  
LIBRARIES\_AND\_DEMO : 638503.734939759 LIFESTYLE : 143781  
6.2687861272

TRÒ CHƠI : 15560965.599534342 GIA ĐÌNH :  
3694276.334922527  
Y TẾ : 120616.48717948717  
XÃ HỘI : 23253652.127118643  
MUA SẮM : 7036877.311557789 CHỤP ẢNH :  
17805627.643678162  
THỂ THAO : 3638640.1428571427  
TRAVEL\_AND\_LOCAL : 13984077.710144928 CÔNG CỤ :  
10682301.033377837 CÁ NHÂN HÓA :  
5201482.6122448975  
NĂNG SUẤT : 16787331.344927534  
CHA MẸ : 542603.6206896552  
THỜI TIẾT : 5074486.197183099 VIDEO\_PLAYERS :  
24727872.452830188 NEWS\_AND\_MAGAZINES : 9549178.467741935  
MAPS\_AND\_NAVIGATION : 4056941.7741935486

Trung bình, các ứng dụng giao tiếp có nhiều lượt cài đặt nhất: 38.456.119. Con số này bị sai lệch nhiều bởi một số ứng dụng có hơn một tỷ lượt cài đặt (WhatsApp, Facebook Messenger, Skype, Google Chrome, Gmail và Hangouts) và một số ứng dụng khác có hơn 100 và 500 triệu

cài đặt:

```
Trong [65]: cho ứng dụng trong free_android:
            nếu ứng dụng[1] == 'TRUYỀN THÔNG' và (ứng dụng[5] == '1.000.000.000+' hoặc ứng
                                                    dụng[5] == '500.000.000+' hoặc ứng
                                                    dụng[5] == '100.000.000+'):

            in(ứng dụng[0], ':', ứng dụng[5])
```

WhatsApp Messenger : 1.000.000.000+ cuộc gọi  
và tin nhắn miễn phí imo beta: 100.000.000+  
Tin nhắn Android : 100.000.000+  
Google Duo - Cuộc gọi video chất lượng cao : 500.000.000+  
Messenger - Trò chuyện bằng văn bản và video miễn phí: Hơn 1.000.000.000  
Cuộc gọi và trò chuyện video miễn phí imo: 500.000.000+  
Skype - cuộc gọi IM & video miễn phí: 1.000.000.000+  
Ai : 100.000.000+  
GO SMS Pro - Messenger, Chủ đề miễn phí, Biểu tượng cảm xúc: 100.000.000+  
LINE: Gọi và nhắn tin miễn phí : 500.000.000+  
Google Chrome: Nhanh & An toàn : 1.000.000.000+  
Trình duyệt Firefox nhanh & riêng tư : 100.000.000+  
UC Browser - Tải xuống nhanh Riêng tư & Bảo mật : 500.000.000+  
Gmail : 1.000.000.000+  
Hangout : 1.000.000.000+  
Messenger Lite: Tin nhắn & Cuộc gọi Miễn phí : 100.000.000+  
Kik : 100.000.000+  
KakaoTalk: Gọi & Nhắn tin miễn phí : 100.000.000+  
Opera Mini - trình duyệt web nhanh : 100.000.000+  
Trình duyệt Opera: Nhanh và An toàn : 100.000.000+  
Điện tin : 100.000.000+  
Truecaller: ID người gọi, chặn thư rác SMS & Trình quay số: 100.000.000+  
UC Browser Mini -Tiny Nhanh Riêng tư & An toàn : 100.000.000+  
Viber Messenger : 500.000.000+  
WeChat : 100.000.000+  
Yahoo Mail - Luôn ngăn nắp : 100.000.000+  
BBM - Gọi và nhắn tin miễn phí : 100.000.000+

Nếu chúng tôi xóa tất cả các ứng dụng giao tiếp có hơn 100 triệu lượt cài đặt, mức trung bình sẽ giảm khoảng 10 lần:

```
Trong [66]: under_100_m = []

            đối với ứng dụng trong
            free_android: n_installs
            = app[5] n_installs = n_installs.replace(',', '')
            n_installs = n_installs.replace('+', '') if (app[1]
            == 'COMMUNICATION') và (float(n_install) < 100000000):
            under_100_m.append(float(n_installs))

            tổng(dưới_100_m) / len(dưới_100_m)
```

Ra[66]: 3603485.3884615386

Chúng tôi thấy mô hình tương tự đối với danh mục trình phát video, là danh mục về nhì với 24.727.872 lượt cài đặt. Thị trường bị chi phối bởi các ứng dụng như Youtube, Google Play Phim & TV hoặc MX Player.

Mô hình này được lặp lại cho các ứng dụng xã hội (nơi chúng ta có những gã khổng lồ như Facebook,



Instagram, Google+, v.v.), ứng dụng chụp ảnh (Google Photos và các trình chỉnh sửa ảnh phổ biến khác) hoặc ứng dụng năng suất (Microsoft Word, Dropbox, Lịch Google, Evernote, v.v.).

Một lần nữa, mối quan tâm chính là các thể loại ứng dụng này có vẻ phổ biến hơn n thực tế. Hơn nữa, những thị trường ngách này dường như bị chi phối bởi một vài gã khổng lồ khó cạnh tranh.

Thể loại trò chơi này có vẻ khá phổ biến, nhưng từ đây chúng tôi phát hiện ra rằng phần này của thị trường có vẻ hơi bão hòa, vì vậy chúng tôi muốn đưa ra một đề xuất ứng dụng khác nếu có thể.

Thể loại sách và tài liệu tham khảo cũng khá phổ biến, với số lượt cài đặt trung bình là 8.767.811. Thật thú vị khi khám phá điều này sâu hơn, vì chúng tôi nhận thấy thể loại này có một số tiềm năng hoạt động tốt trên App Store và mục tiêu của chúng tôi là đề xuất một thể loại ứng dụng cho thấy tiềm năng sinh lời trên cả App Store và Google Play.

Chúng ta hãy xem một số ứng dụng thuộc thể loại này và số lượt cài đặt của chúng:

```
Trong [67]: cho ứng dụng trong free_android:
           nếu ứng dụng[1] == 'SÁCH_AND_REFERENCE': in(ứng
               dụng[0], ':', ứng dụng[5])
```

Đọc sách điện tử - Đọc sách miễn phí : 50.000+  
 Tải xuống sách miễn phí với sách xanh: 100.000+  
 Wikipedia : 10.000.000+  
 Đầu đọc thú vị : 10.000.000+  
 Đài phát thanh Panda miễn phí Âm nhạc: 100.000+  
 Cửa hàng sách : 1.000.000+  
 FBReader: Trình đọc sách yêu thích: 10.000.000+  
 English Grammar Complete Handbook : 500,000+  
 Sách miễn phí - Truyện thần linh và truyện: 1.000.000+  
 Google Play Sách : 1.000.000.000+  
 AlReader - bất kỳ trình đọc sách văn bản nào: 5.000.000+  
 Từ điển tiếng Anh ngoại tuyến: 100.000+  
 Ngoại tuyến: Từ điển tiếng Anh sang tiếng Tagalog: 500.000+  
 Cây tìm kiếm gia đình: 1.000.000+  
 Đám mây sách : 1.000.000+  
 Công thức Thuốc tiên tri miễn phí : 500.000+  
 ReadEra - trình đọc sách điện tử miễn phí: 1.000.000+  
 Phát hiện người gọi ẩn danh: 10.000+  
 Trình đọc sách điện tử: 5.000.000+  
 thấp sáng E bk 100 000+

Thể loại sách và tài liệu tham khảo bao gồm nhiều ứng dụng: phần mềm xử lý và đọc sách điện tử, nhiều bộ sưu tập thư viện, từ điển, hướng dẫn về lập trình hoặc ngôn ngữ, v.v. Có vẻ như vẫn còn một số ít ứng dụng cực kỳ phổ biến làm sai lệch mức trung bình:

```
Trong [68]: cho ứng dụng trong free_android:
            nếu ứng dụng[1] == 'BOOKS_AND_REFERENCE' và (ứng dụng[5] == '1.000.000.000+' hoặc ứng
                                                    dụng[5] == '500.000.000+' hoặc ứng
                                                    dụng[5] == '100.000.000+'):

                in(ứng dụng[0], ':', ứng dụng[5])
```

Google Play Sách : 1.000.000.000+

Kinh thánh : 100.000.000+

Amazon Kindle : 100.000.000+

Wattpad Sách miễn phí : 100.000.000+

Sách nói từ Audible: 100.000.000+

Tuy nhiên, có vẻ như chỉ có một số ứng dụng rất phổ biến nên thị trường này vẫn cho thấy tiềm năng. Hãy thử lấy một số ý tưởng về ứng dụng dựa trên loại ứng dụng nằm ở mức trung bình về mức độ phổ biến (từ 1.000.000 đến 100.000.000 lượt tải xuống):

```
Trong [69]: cho ứng dụng trong free_android:
            nếu ứng dụng[1] == 'BOOKS_AND_REFERENCE' và (ứng dụng[5] == '1.000.000+' hoặc ứng
                                                    dụng[5] == '5.000.000+' hoặc ứng
                                                    dụng[5] == '10.000.000+' hoặc ứng
                                                    dụng[5] == '50.000.000+'):

            in(ứng dụng[0], ':', ứng dụng[5])
```

Wikipedia : 10.000.000+ Cool  
 Reader : 10.000.000+ Cửa hàng  
 sách : 1.000.000+ FBReader:  
 Trình đọc sách yêu thích : 10.000.000+ Sách miễn phí -  
 Truyện thần linh và truyện : 1.000.000+ AlReader -trình đọc sách văn bản  
 bất kỳ : 5.000.000+ FamilySearch Tree : 1.000.000+ Cloud  
 of Books : 1.000.000+ ReadEra - trình  
 đọc sách điện tử miễn phí :  
 1.000.000+ Trình đọc sách điện tử : 5.000.000+ Đọc  
 sách trực tuyến : 5.000.000+  
 eBook: trình đọc sách fb2 epub zip :  
 1.000.000+ Tất cả các công thức toán học : 1.000.000+ Tổ  
 tiên : 5.000.000+ Trợ giúp HTC : 10  
 ,000,000+ Mặt trăng+ Ngủ ời  
 đọc : 10,000,000 + Từ điển  
 tiếng Anh-tôi Miễn phí): 10.000.000+  
 Al Qur'an Indonesia: 10.000.000+ Al'Quran Bahasa  
 Indonesia: 10.000.000+ Al Qur'an Al karim:  
 1.000.000+ Al Qur'an: EAlim - Bản dịch & MP3 Ngoại  
 tuyến: 5.000.000+ Đọc Koran & MP3 30 Juz  
 Ngoại tuyến: 1.000.000+ Hafi zi Kinh Qur'an  
 15 dòng trên mỗi trang: 1.000.000+ Kinh Qur'an cho Android:  
 10.000.000+ Vệ tinh AR: 1.000.000+  
 Oxford AZ của tiếng Anh Cách sử dụng:  
 1.000.000+ Dictionary.com: Tìm định nghĩa cho các  
 từ tiếng Anh: 10.000.000+ Từ điển tiếng  
 Anh - Ngoại tuyến: 10.000.000+ Kinh thánh KJV: 5.000.000+ N XEM : Đọc sách  
 điện tử & Tạp chí : Hơn 10.000.000 Trích dẫn hay: Cuộc  
 sống, Tình yêu, Gia đình & Động lực : Hơn 1.000.000 Số  
 liệu thống kê Royale cho Clash Royale :  
 1.000.000+ Từ điển : 10.000.000+  
 wikiHow: cách làm bất cứ điều gì : 1.000.000+ Bài  
 viết EGV : 1.000.000+ My Little Pony Hư ớng dẫn AR : 1.000.000+ Trình dịch tiếng Anh  
 Tây Ban Nha : 10.000.000+ Từ điển - Merriam-Webster :  
 10.000.000+ Thư viện JW :  
 10.000.000+ Từ điển tiếng Anh Oxford : Miễn phí :  
 10.000.000+ Từ điển tiếng Anh Hindi : 10.000.000+ Từ điển tiếng Anh sang tiếng  
 Hindi : 5.000.000+

Thị trường sách này dường như bị chi phối bởi phần mềm xử lý và đọc sách điện tử, cũng như các bộ sưu tập thư viện và từ điển khác nhau, vì vậy có lẽ không nên xây dựng các ứng dụng tương tự vì sẽ có một số đối thủ cạnh tranh đáng kể.

Chúng tôi cũng nhận thấy có khá nhiều ứng dụng được xây dựng dựa trên cuốn sách Kinh Qur'an, điều này cho thấy rằng việc xây dựng ứng dụng dựa trên cuốn sách nổi tiếng có thể mang lại lợi nhuận. Có vẻ như việc sử dụng một cuốn sách nổi tiếng (có lẽ là một cuốn sách mới hơn) và biến nó thành một ứng dụng có thể mang lại lợi nhuận cho cả thị trường Google Play và App Store.

Tuy nhiên, có vẻ như thị trường đã đầy thư viện, vì vậy chúng tôi cần thêm một số tính năng đặc biệt bên cạnh phiên bản thô của cuốn sách. Điều này có thể bao gồm các trích dẫn hàng ngày từ cuốn sách, phiên bản âm thanh của cuốn sách, câu đố về cuốn sách, diễn đàn nơi mọi người có thể thảo luận về cuốn sách, v.v.

## kết luận

Trong dự án này, chúng tôi đã phân tích dữ liệu về ứng dụng dành cho thiết bị di động trên App Store và Google Play với mục tiêu đề xuất cấu hình ứng dụng có thể mang lại lợi nhuận cho cả hai thị trường.

Chúng tôi kết luận rằng việc sử dụng một cuốn sách nổi tiếng (có lẽ là một cuốn sách mới hơn) và biến nó thành một ứng dụng có thể mang lại lợi nhuận cho cả thị trường Google Play và App Store. Các thị trường đã đầy thư viện, vì vậy chúng tôi cần thêm một số tính năng đặc biệt bên cạnh phiên bản thô của cuốn sách. Điều này có thể bao gồm các trích dẫn hàng ngày từ cuốn sách, phiên bản âm thanh của cuốn sách, câu đố về cuốn sách, diễn đàn nơi mọi người có thể thảo luận về cuốn sách, v.v.