App Store and Google Play Store App Profitability

Download: <mark>*Jupyter Notebook file, ReadMe*</mark>

Abstract: 10,000 Play Store apps and 7,200 App Store apps datasets from Kaggle that include 13 features from the Play Store and 10 features from the App Store that describe the given app.

Data Set Characteristics:	Multivariate	Number of Instances:	17200	Area:	Mobile Apps
Attribute Characteristics:	N/A	Number of Attributes:	23	Missing Values:	1

Source:
https://www.kaggle.com/lava18/google-play-store-apps
https://www.kaggle.com/ramamet4/app-store-apple-data-set-10k-apps?select=AppleStore.csv

Data Set Information:
Details of the applications on Google Play. There are 13 features that describe a given app. License information is this work is licensed under the Creative Commons Attribution 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by/3.0/.

For App Store, the ever-changing mobile landscape is a challenging space to navigate. . The percentage of mobile over desktop is only increasing. Android holds about 53.2% of the smartphone market, while iOS is 43%. To get more people to download your app, you need to make sure they can easily find your app. Mobile app analytics is a great way to understand the existing strategy to drive growth and retention of future users. With millions of apps around nowadays, the following data set has become very key to getting top trending apps in the iOS app store. This data set contains more than 7000 Apple iOS mobile application details. The data was extracted from the iTunes Search API at the Apple Inc website. R and linux web scraping tools were used for this study. Data collection date (from API); July 2017. License information is http://www.gnu.org/licenses/old-licenses/gpl-2.0.en.html.

Data variables include:

App Store DataSet

Label	Data Type	Description
'id'	Numeric	App ID
'track_name'	Numeric	Application name

'size_bytes'	Numeric	Memory size (in Bytes)
'currency'	String	Currency of app
'price'	Numeric	Cost of app
'rating_count_tot'	Numeric	Rating count total
'rating_count_ver'	Numeric	Rating count by version
'user_rating'	Numeric	User rating
'user_rating_ver'	Numeric	User rating by version
'ver'	Numeric	Version
'cont_rating'	String	Content rating
'prime_genre'	String	Genre
'sup_devices.num'	Numeric	Number of supported devices
'ipadSc_urls.num'	Numeric	No description given by source
'lang.num'	Numeric	No description given by source
'vpp_lic'	Numeric	No description given by source

Google Play Store DataSet

Label	Data Type	Description
App	String	Application name
Category	String	Category the app belongs to
Rating	Numeric	Overall user rating of the app (as when scraped)
Reviews	Numeric	Number of user reviews for the app (as when scraped)
Size	String	Size of the app (as when

		scraped)
Installs	String	Number of user downloads/installs for the app (as when scraped)
Type	String	Paid or Free
Price	String	Price of the app (as when scraped)
Content	String	Age group the app is targeted at - Children / Mature 21+ / Adult
Genres	String	An app can belong to multiple genres (apart from its main category). For eg, a musical family game will belong to Music, Game, Family genres.
Last Updated	Date	Date when the app was last updated on Play Store (as when scraped)
Current Version	String	Current version of the app available on Play Store (as when scraped)
Android Version	String	Min required Android version (as when scraped)

Data Summaries Calculated:
- Mean
- Frequency

Units for Data Summaries and Variables:
- User Ratings
- Number of Installs

Data Transformations:
- Data was downloaded from https://www.kaggle.com/lava18/google-play-store-apps, https://www.kaggle.com/ramamet4/app-store-apple-data-set-10k-apps?select=AppleStore.csv
- Files "applestore.csv", "googleplaystore.csv" were imported into Jupyter Notebook.
- The Google Play Store apps data set has an error in one of its rows: a "CATEGORY" value is missing. Removed the row with the error.
- Duplicate entries were removed using the criterion of "entries with number of reviews lower than entry with highest number of reviews".
- Non-English language apps were removed and a new data set was created.
- Removed all paid iOS and Android apps.
- Identified app profiles that were successful on both iOS and Android platforms which included
    - iOS and Android data sets with most common genres in each market visualized in frequency tables.
    - iOS and Android apps by genre by average number of installs.