

Apple iOS Mobile Applications Data Exploration

Dataset

The data consists of information's regarding 7200 mobile applications that includes application names, ratings given by users, price, size, different genres of the applications, content rating groups, version numbers and other application related information's.

The dataset can be found in the Kaggle repository at

<https://www.kaggle.com/ramamet4/app-store-apple-data-set-10k-apps/home>

Summary of Findings

In my exploration on this dataset, my initial thought was that the ratings given by the users on the apps would be impacted by the Price and Size of the applications. But on analyzing it, I found that the two categoric variables on the dataset, the Content Rating Groups and the Prime Genre of the Apps have a strong relationship with the user ratings. Another observation, is that there are two categories of application types, one is free where the price is 0.0 and another is a Paid one. So, for detailed analysis, I created two datasets for each of these types. Coming back to price and sizes, applications that are prices below 10 USD have higher number of rating counts than those with higher prices. And for sizes, applications that are around 100 MB have higher rating counts and generally with high rating values. Regarding the number of languages supported by the apps, applications with less than 20 number of languages supported have a strong positive influence on the user ratings. So, below listed are my major findings:

For PAID APPS:

- For users in 4+ groups, shopping, book and references have a stronger relation with higher user ratings.
- Book, Education, Health & Fitness, Games, Photo & Video and News are stronger in 9+ groups.
- Books, Catalogs, Education, Food & Drink, Travel, Navigation, Games and Health & Fitness are strong.
- While for 17+, Books, Business, Lifestyle, Navigation, News, Photo & Video, Travel and Weather are pretty strong.
- I can say that users on average gives higher ratings to apps around 3.5 to 4.5 as compared to the lower ratings.

For FREE APPS:

- I can see Health & Fitness, Productivity, Reference forms a stronger relation with higher user ratings in 4+ groups.
- Books, Games, Lifestyle, Social Networking, Shopping and Utilities are the most popular in 9+ groups.
- Book, Games, Medical, Navigation, Photo & Video, Productivity, Shopping and Utilities are for 12+ groups.
- Lastly in 17+ groups, Book, Catalogs, Education, Food & Drink and Medical are the most popular ones.
- Users on average gives variant ratings across the apps of different content rating groups. 4+ groups give on average more ratings of 3.5, 4.5 and 5. 9+ groups on average gave more ratings of 3 and 4.5. 12+ with 2.5, 4 and 4.5 ratings and lastly 17+ groups give 2, 2.5, 3.5, 4.5 and 5 user ratings on average.

Lastly, the user ratings for the current version and the overall version of the apps have a strong positive relation. So, we can somewhat predict the ratings of the current version of an app if we know the overall version.

Key Insights for Presentation

For the presentation, initially I focused on the price and size distribution against the user ratings. I used Scatterplot to depict these relationships. Same types of plot are also used in determining the relationship between languages supported and user ratings, and then the relationship between ratings of the current version and the overall version since all of these variables are numeric variables.

Later on, once I started observing the influence of the categoric variables, I used a clustered bar chart to show the relationship between the two categoric variables. And then a heatmap is used to depict how the user ratings are impacted by Content Ratings and Prime Genres. This is also because, I wanted to use the annotations showing a kind of numeric value for the relations. Lastly, to find the ratio of how the total rating counts and the user ratings relate across the various content groups, I used a pointplot for the two types of applications, the free and the paid ones.