Apple Store Genre Analysis

Kevin Wang, Aashika Balaji, Zhenzhen Liu, Minyue Fan

Introduction

The Mobile Apps Statistics Data

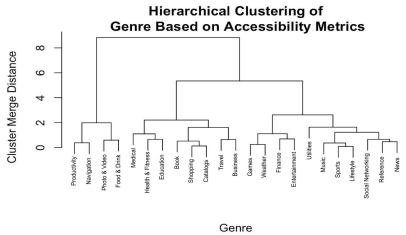
Our dataset contains 7200 mobile apps from Apple's app store. The variables depict the apps' price, user ratings, content ratings, number of languages supported, app descriptions, etc. The data were created on July 2017 through iTunes search API, and were made available on Kaggle.

Our Approach

Our analysis compares the accessibility and descriptions across different genres. We first explore the similarities among the 23 nuanced genres, and then shift our focus on 4 generalized genres.

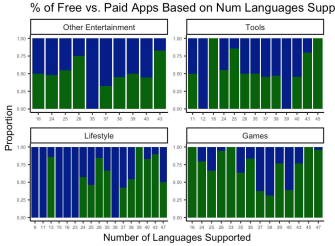
Accessibility of Apps

Clustering on Number of Supported Languages and Devices



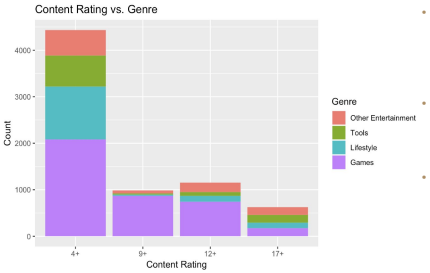
- This dendrogram uses nuanced the genres as the data points, and the distance was calculated based on number of languages supported and number of devices supported
- In general, genres belonging to the same generalized genres are clustered together, e.g. Medical, Health & Fitness, Education are all Lifestyle apps, etc.

% of Free/Paid Apps Based on Num Languages Supported



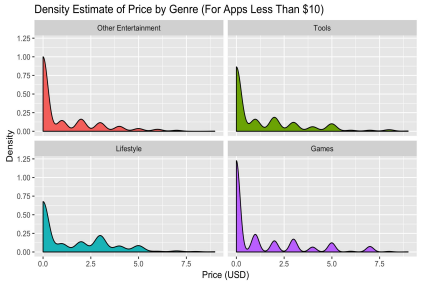
- Plot examines the number of supported languages across apps with respect to the pricing of the apps.
- The bars are broken up by the percentages of free vs paid and then faceted on the generalized genres.
- Some genres have more paid options than others (specifically Games and Tools), but are not significantly affected by number of languages supported.

Content Rating



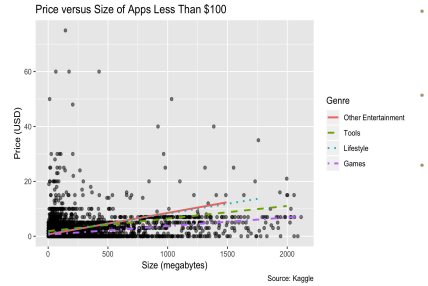
- The bar chart displays the conditional distribution of generalized genres by content ratings.
- The majority of apps are suitable for people with age 4+.
- For apps rated 4+, 9+, 12+, games have the most counts; for apps rated 17+, the 4 genres have roughly equal number of apps.

Price and Genre



- Only prices less than $10 were considered since there are comparatively few apps above that mark.
- Lifestyle is more uniformly and smoothly distributed than the other three genres.
- Games has the most pronounced peaks and is more clustered around 0 than the others.

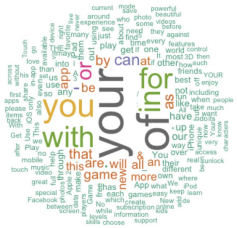
Price and App Size



- Only prices less than $100 were considered since there were a few extreme outliers above that mark.
- All genres seem to have a positive linear trend between size and price.
- The trend are similar around the different genres, with other entertainment and lifestyle showing stronger, and similar, trends.

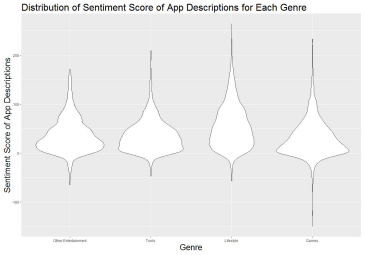
Text Analysis on The Descriptions of Apps

Word Cloud of App Description



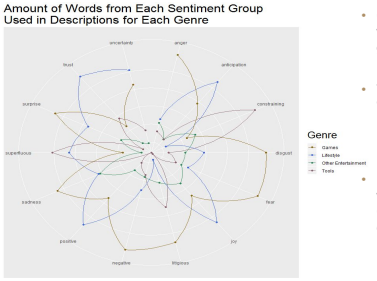
- The word cloud plots words that show up in the descriptions of apps. All particles and prepositions are not shown.
- Among the most common words, one type is to describe what is in an app: "games", "characters", "levels", "music", "photo", etc.
- Another type is positive adjectives: "powerful", "beautiful", "unique", "real", "easy", etc.

Sentiment Score



- The sentiment score for an app's description is calculated by the sum of scores (positive= +1, neutral=0, negative = -1) for all words.
- For all genres, most descriptions are positive, but some are negative, because some apps use "free for limited time", etc. to stimulate download.
- Descriptions of games have a wider range of sentiment scores, because some games include the games stories the description, and lead to more complex sentiments.

Sentiment Types Used in Descriptions



- Games involve various negative words because many games tell the game stories in their descriptions.
- Tools involve "constraining", and "superfluous" words, because they identify how they improve the inconveniences in people's lives.
- Lifestyle apps mostly use positive words like "joy" and "anticipation". Other entertainments are rather constrained in use of words.

ACKNOWLEDGEMENTS

Dataset: Perumal, Ramanathan, Mobile App Store (7200 Apps), https://www.kaggle.com/ramamet4/app-store-apple-data-set-10k-apps/home, 2018