Analyzing ratings of Google Play Store applications

Ivan Radonov
5967988
https://github.com/ikar1234/DataLiteracyProject
ivan.radonov@student.uni-tuebingen.de

Abstract

This study uses a dataset containing both free and paid apps from Google Play Store and looks at how users ratings differ between both categories. Further, the most popular paid app categories are explored and the relationship between application rating and the price, size, and last updated date is assessed.

1 Introduction

Google Play Store is the official app store for Android devices. Apps can be uploaded from certified users (developers) and can be downloaded from any user, provided that they meet the age restriction, their device fulfills all technical requirements, and they pay for it before use, in case the app is paid. App creators can decide whether their apps are free or paid. Any app can still have paid features inside it which are independent of its price.

A rating of an app provides a crowdsourced indicator of its quality [5], hence it can be regarded as a key feature to optimize. To this end, Android developers need to make some important decisions while developing an app regarding its other features, for example, whether the app should be paid and whether it should meet some size constraints. In this study, an app dataset from the Google Play Store was used to answer some of these questions and help Android developers make better decisions.

There exist similar studies, which focus on specific app categories, such as nutritional or finance apps [3], [4]. Other studies focus on sentiment analysis on the user reviews [1]. No studies were found which emphasize how the app price reflects user satisfaction.

2 Methods

2.1 Origin and structure of app data

The data was downloaded from Kaggle:

https://www.kaggle.com/lava18/google-play-store-apps

It contains a snapshot of 13 characteristics from 10841 apps available in the Google Play Store as of August 2018. Notable features include:

- the app name, category, rating, number of installations and reviews, type (paid or free), price in US dollars, size, and last updated date.

There are 34 app categories in total. The rating shows the average rating given by users of the app in the range from 1 (lowest) to 5 (highest). The app size is either given in KB or MB, or it varies with device.

2.2 Data preprocessing

Prior to conducting the analysis, the dataset was preprocessed. Missing values for some apps were imputed based on internet research. The columns containing the rating, number of reviews, number of installs, and the price were converted to numerical values. The app size was converted to the same unit (MB). To obtain the number of days since an app was last updated, the difference in days between the last updated date for the app and the last updated date for any app was calculated. Then, only apps updated in the last 1000 days were taken to remove outliers. In order to mitigate bias, apps with fewer than 20 reviews were removed from the analysis. Apps with missing ratings were also removed.

2.3 Data analysis

The dataset was split into free and paid apps. A two-sample Kolmogorov-Smirnov (KS) test was used to determine whether both samples are drawn from the same distributions. The KS test is a nonparametric test used for one-dimensional probability distributions [2]. It can be either one-sample, where a sample is compared to a reference distribution, or two-sample, where two samples are compared. The null hypothesis assumes both samples come from the same distribution. The test statistic is the maximum absolute difference of both cumulative distributions. An implementation of the KS test is provided in `scipy`, a Python library for scientific computing [6].

All visualizations were done using Python's `seaborn` library, which is used to create statistical graphics [7]. Linear regression fits for different features were done using the built-in functionalities of the library.

The full data processing and analysis pipeline can be found in the `RatingAnalysis.ipynb` notebook.

3 Results

3.1 Application price vs. rating

The rating distributions for both app types as well as the corresponding mean values are displayed in Figure 1.

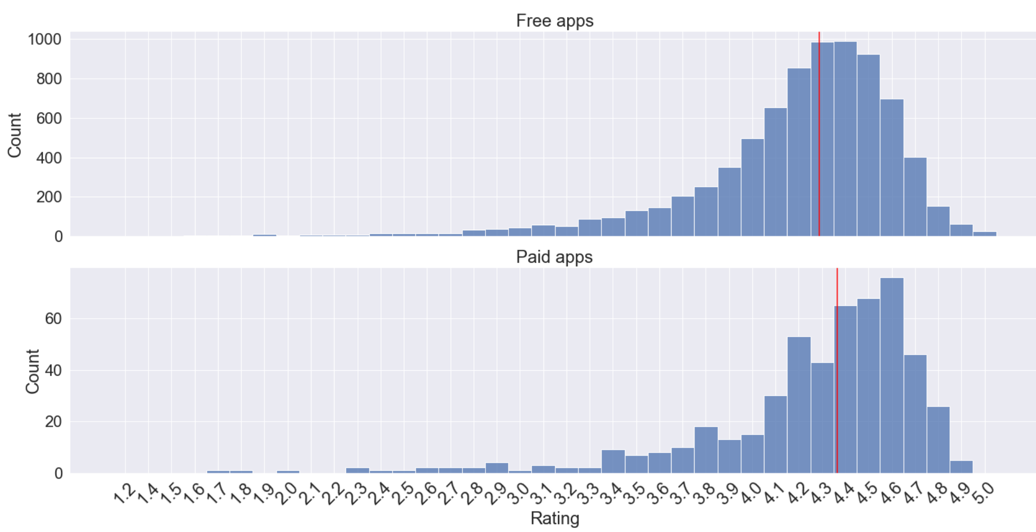


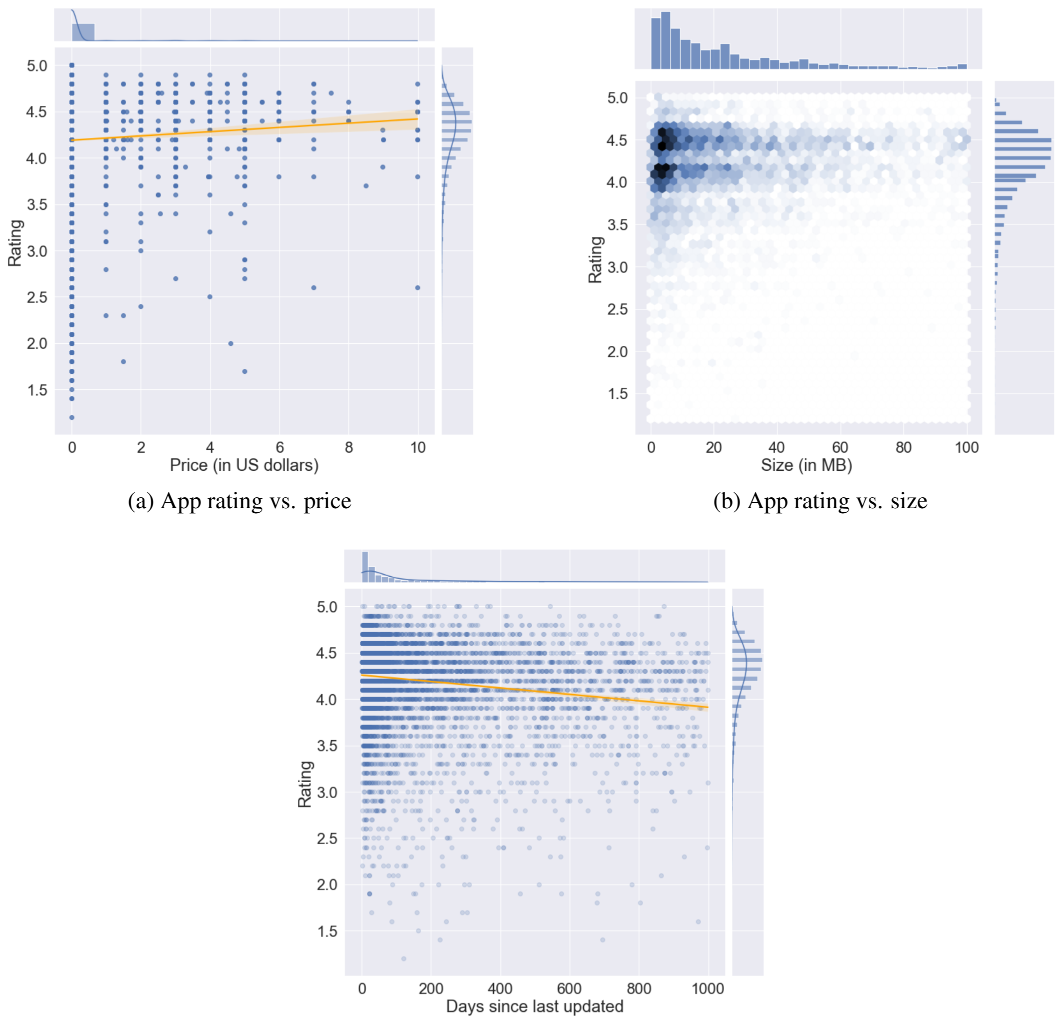
Figure 1: Rating distribution of free (top picture) and paid (bottom picture) apps, with corresponding rating means across apps given as vertical red lines. Apps with missing rating values or fewer than 20 reviews are not shown. Each bin corresponds to a discrete rating value.

As can be seen from this figure, the rating of the vast majority of the apps lies in the range from 3.4 to 5.

The Kolmogorov-Smirnov test returned a p-value of $1.22 \times e^{-8}$ and a test statistic of 0.138. The null hypothesis was thus rejected, and both samples were deemed different. This finding can be confirmed by observing Figure 1, although both sample means are almost equal.

3.2 Relationship between app rating and other features

As a further step, the study aimed to find whether there exists a relationship between the rating of an app and some of its other features. To this end, the app size, price, and last updated date were considered and compared to the app rating. Figure 2 visualizes the findings. Each subfigure shows a joint plot of the rating and the other feature, with marginal plots on the top and left.



(a) App rating vs. price

(b) App rating vs. size

(c) App rating vs. last updated date

Figure 2: The relationship between app rating and app price, size and last updated date. In each figure, the top/left plots show the marginal distributions of that feature. Figure (a) compares the app rating to the app price, given in US dollars. Figure (b) shows a density plot of the rating and app size, given in MB. Darker values correspond to higher density. Figure (c) compares the app rating to the last updated date. Orange lines in figures (a) and (b) show the best fitting line found by a standard linear regression model.

For all three above features, a standard linear regression model was used to determine a correlation. For the app price and last updated date, the best fitting line found was plotted in Figure 2 as an orange line. For the app size, no strong relationship could be found. For this reason, only a density plot is shown.

4 Discussion

Overall, it can be seen that an app's rating is correlated to its other features, and some causal relationships can be found.

The results from Section 3.1 imply that making an app paid can slightly improve its rating. Needless to say, this would decrease the user basis, so the price has to be chosen carefully.

The insights from Figure 2 can be summarized as follows: increasing an app's price has a beneficial influence on its rating. Apps that are updated more frequently receive substantially higher ratings. There was no substantial relationship found between the app size and its rating, although it can be seen that most apps are below 20 MB in size, which can serve as a guidance for the app scope.

Based on the results of this study it can be conjectured that, given enough (external) features, an app's rating can be almost perfectly predicted, even without knowing its content or actual quality.

These results allow for some suggestions to be given as to how to design Android applications: To achieve high ratings, an app should be frequently updated, of moderate size, and be paid, in case it provides some features missing in other similar apps.

Several limitations impeded achieving better results. First, since paid apps are affordable to fewer people, their ratings are biased by the user profile, but their quality is not necessarily better. Further, both rating and number of installs are imperfect measures of app quality or profitability, since they don't tell e.g., how long the app was installed or if the ratings were honest. Increasing the number of features would also allow for more findings. Further studies may consider e.g. sentiment analysis on the user reviews, for which there are also available datasets.

5 Conclusion

In this study, app data from the Google Play Store was analyzed and visualized. It was found that paid applications receive slightly higher ratings. The relationship between an app's rating and its price, size, and last updated date was assessed and discussed. Key findings were that app rating increases with the price and update frequency. Some limitations of the data, applications and extensions of the analysis were pointed out.

References

[1] Mir Riyanul Islam. "Numeric rating of Apps on Google Play Store by sentiment analysis on user reviews". In: *2014 International Conference on Electrical Engineering and Information Communication Technology*. 2014, pp. 1–4. DOI: `10.1109/ICEEICT.2014.6919058`.

[2] "Kolmogorov–Smirnov Test". In: *The Concise Encyclopedia of Statistics*. New York, NY: Springer New York, 2008, pp. 283–287. ISBN: 978-0-387-32833-1. DOI: `10.1007/978-0-387-32833-1_214`. URL: `https://doi.org/10.1007/978-0-387-32833-1_214`.

[3] Harleigh Schumer, Chioma Amadi, and Ashish Joshi. "Evaluating the dietary and nutritional apps in the google play store". In: *Healthcare informatics research* 24.1 (2018), pp. 38–45.

[4] Vincent F Taylor and Ivan Martinovic. "Short paper: A longitudinal study of financial apps in the google play store". In: *International Conference on Financial Cryptography and Data Security*. Springer. 2017, pp. 302–309.

[5] Rajesh Vasa et al. "A preliminary analysis of mobile app user reviews". In: *Proceedings of the 24th Australian computer-human interaction conference*. 2012, pp. 241–244.

[6] Pauli Virtanen et al. "SciPy 1.0: fundamental algorithms for scientific computing in Python". In: *Nature methods* 17.3 (2020), pp. 261–272.

[7] Michael L Waskom. "Seaborn: statistical data visualization". In: *Journal of Open Source Software* 6.60 (2021), p. 3021.