

Play store App Review Analysis

Kalpita Malviya,
Data science trainees,
AlmaBetter, Bhopal.

Abstract:

They are very useful for us to characterize

The Play store is the greatest source of Android apps. The most leading and safest apps are published here.

Our features-based analysis on the database can help to find the key factors which are responsible for the app engagement, popularity and success.

Keywords: *Exploratory Data Analysis (EDA), Reviews, Rating, Distribution.*

1. Problem Statement

Data provided by Play store, which is operated and developed by Google. Play store has a million number of similar or different kinds of apps which are tagged by different Genres. These apps are available for both free and paid. They are also providing different features for the users to choose and download the best apps from it.

- App name: This is the column which contains the name of the app.
- Category: The Category is a column by which the apps got

an app from different points of view and provide to the people across available options. They have been in operation since the year 2008. During this period, they have optimized a lot and are more secure from any android app resources.

The main objective is to build an analysis, which could help the ordinary people to find the best apps based on the previous data. This would in turn help them in matching the right cabs with the right customers quickly and efficiently.
Reviews: The distance for the trip requested by the customer

- Size: Size is a factor which is not the same for all the devices, and it fully depends on the developers.
- Installs: Installs gives us the data of total installed users count.
- Type: Type defines whether the app is free or paid.
- Price: Here price is mentioned for the paid apps, for free apps 0 will be given.

separated based on the application and purpose.

- Rating: Rating plays a huge role in finding the correct apps. It was manually given by the users.

- Genres: Genres are like tags, an app can be coming under more than one Genres, based on the usage.
- Android Version: This feature gives us the supported device versions of android.

2. Introduction

Play store is an Android Market serves as the official app store for certified devices running on the AndroidOperating system. Developed and Operated by Google, launched on 6th March, 2012.

Approximately 3.48 million apps are in the Play store. Play store apps have their own features such as Ratings, Reviews, Size and more. From the problem statement given, we should analyze the given database and should come up with the key factors that increased the number of users, long term usage etc., the objective of this project is to deliver insights to understand customer demands better and thus help developers to popularize the product.

3. Exploring the database

We have provided with two databases

- Shape of this database is (10841, 13).

- Out of this thirteen columns we have numeric

User reviews database

- Shape of this database is (64295, 5).
- Here there are only two numeric values found,
- Sentiment Subjectivity, Sentiment Polarity.

4. Features selection

The columns are also known as features, one or more different features are grouped together for different analyses to form a data frame.

Highest_Rating_App: It contains the data of highest rating apps based on the categories.

Number_of_App: Here, we have shown the number of applications based on different categories.

More_downloads_App: Here, We have shown the highest number of download apps.

Family_lowest_rating: This data shows the lowest rating application of family category.

Family_highest_rating: This data shows the highest rating application of family category.

Merged_df: We have merged play store data and user review data.

5. Top 10 Highest rating Apps in google play store in terms of categories

The analyze is done between the top 10 categories which was having more number of downloads

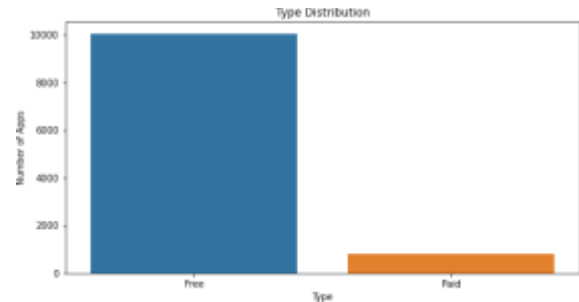
- As we saw here, the Play store having number of applications in the categories like, Games, Family, Business etc.
- The developers are mostly focusing on these categories because of the people's daily basis requirements.
- Categories like Games, Business, and Family are having comparatively more amount of apps count.

6. Free apps

- When comparing the both plots, people are showing more interest on free apps like art & design. And The number of free apps are more than 10000

7. Paid apps

- When it's coming, commercial people are preferring apps like Business, communication and Personalization and the number of paid apps are less than free apps.



8. 10 apps from the 'FAMILY' category are having the lowest rating and highest rating.

1. Lowest Rating

- In this data frame we have analyzed the lowest rating apps in family category, i.e. BG TV app, EB Mobile.

2. Highest Rating

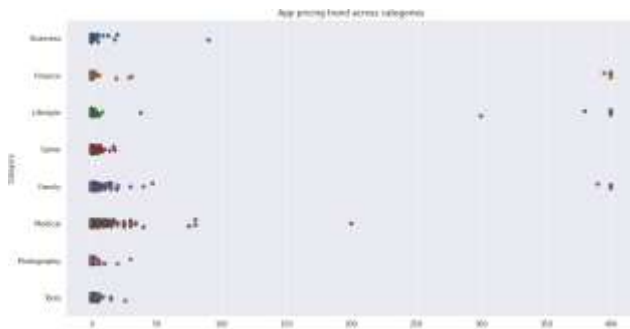
- In this data frame we have analyzed the highest rating apps in family category, i.e. Pyaar Ek Dhoka, Ek Bander Ne Kholi Dukan.

9. Top 10 apps which has more downloads:

- In this data we have shown the highest number of downloads in top10 applications.
- In this data we have analyzed the maximum number of downloads of app that is subway surfers.

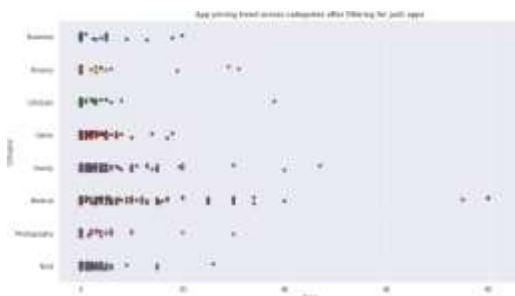
10. Relation between app category and app price

- In this data we have analyzed that which category has the highest price.
- Also we have shown the lowest category.



11. Filter out "junk" apps

- It looks like a bunch of the really expensive apps are "junk" apps. That is, apps that don't really have a purpose. Some app developer may create an app called I Am Rich Premium or most expensive app (H) just for a joke or to test their app development skills. Some developers even do this with malicious intent and try to make money by hoping people accidentally click purchase on their app in the store.



12. Sentiment analysis of user reviews:

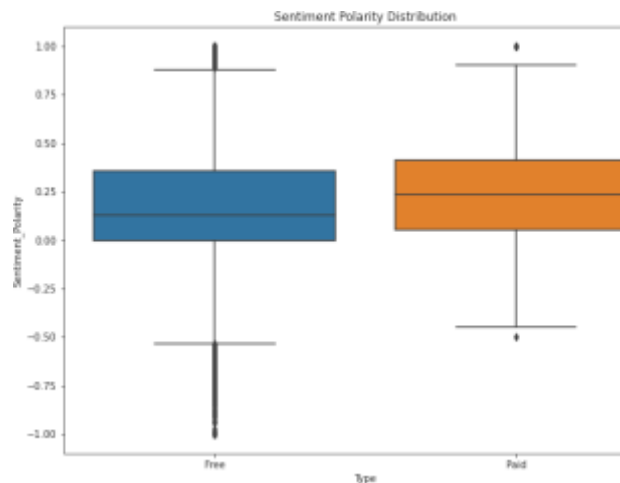
- User reviews for apps can be analyzed to identify if the mood is positive, negative or neutral about that app. For example, positive words in an app review might include words such as 'amazing', 'friendly', 'good', 'great', and 'love'. Negative words might be words like 'malware', 'hate', 'problem', 'refund', and 'incompetent'.
- By plotting sentiment polarity scores of user reviews for paid and free apps, we observe that free apps receive a lot of harsh comments, as indicated by the outliers on the negative y-axis. Reviews for paid apps appear never to be extremely negative. This may indicate something about app quality, i.e., paid apps being of higher quality than free apps on average. The median polarity score for paid apps is a little higher than free apps, thereby syncing with our previous observation.
- In this notebook, we analyzed over ten thousand apps from the Google Play

Store. We can use our findings to inform our decisions should we ever wish to create an app ourselves.

- g) Filter out "junk" apps
- h) Sentiment analysis of user reviews

References

1. [GeeksforGeeks](#)
2. [matplotlib.org](#)
3. [w3School](#)



Conclusion

After analyzing the dataset, we have got answers to some of the serious & interesting facts which any of the android users would love to know.

- a) Top 10 Highest rating Apps in google play store in terms of categories
- b) Number of Application in terms of Category
- c) Top 10 apps which has more downloads
- d) Which 10 apps from the 'FAMILY' category are having the lowest rating and highest rating.
- e) Free and Paid Apps
- f) Relation between app category and app price