# Play Store App Review Analysis

**(pradyumna vikram singh)**

**Data science trainees,**

**AlmaBetter, Bangalore**

**Abstract:**

Mobile Applications are an integral part of our daily lives. As they are easy to create and are lucrative, it becomes essential to do comprehensive analysis of play store app data and understand the demand of users. Through our exploratory data analysis, we tried our level best to bring out some important insights from the data to devise strategies for growth and retention. **Keywords: Play store apps, Exploratory Data Analysis, Price, Installs, Reviews, Rating, Size, Sentiment polarity.**

**1. Problem Statement**

The marketplace is crowded with all types of mobile apps as it has enormous potential to drive app-making businesses to success. The key performance indicators for app engagement and its success are:

• **Rating**: Overall user rating of the application. It lies between 1 to 5.

• **Reviews:** Total number of users reviews each app has received.

**• Size**: The memory size needed to install the application.

• **Installs:** Number of installs- An install takes place when a user has downloaded

an app and successfully opens it for the first time.

**• Price**: The price of the app. Thus, this play store app review analysis is performed to know, the categories, the genres

which is most appealing to users. And also, how does the current react to certain parameters such as size of apps and their price

**2. Introduction**

User ratings are the key components to conversion rate. The ratings of the app directly influence the visitors in their decisions to download an app, thus acting as an important component of conversion rate. When taking a decision about downloading an app, 90% of users take into account its rating

Similarly, a comprehensive mobile app review exactly reflects what an app offers. App reviews are a major decision-making factor for users and app owners alike

Apart, from this, installs are important to mobile marketing for a number of reasons. Firstly, installs help advertisers identify sources that successfully deliver traffic to their app. Tracking installs into an app helps advertisers identify the time when campaigns produce results. Installs can also be used to identify the overall health of an application.

Also, users often avoid downloading apps that seem too large, especially in an era where each application has come up with its ‘lite’ version. So, this can also be one of the major factors of its usage.

Thus, we can see that actionable insights can be drawn for developers to work on and capture the Android market

The flow of analyzing the dataset will be as follows:

➢ **Problem Comprehension**: It involves loading the dataset and do a logical analysis of variables

**Data Exploration:** Exploring the data using head, tail, info and describe.

**Data Cleaning:** Looking for and handling NaN/Null/Missing values.

➢ **Extracting Statistics**: Finding the key statistical values to fill up the missing values and making important insights.

➢ **Exploratory analysis**: Getting some conclusion from the given dataset, finding correlation and trends that fulfills the purpose of analysis.

**➢ Data Visualization**: Visualizing the analysis with various graphs such as Histograms, Line chart, Pi chart, Boxplot etc.

➢ **Conclusio**

**3. Play store user reviews Dataset**

Our dataset consists of play store application and user reviews

• **Play Store Data.csv**: contains all the details of the applications on Google Play. There are 13 features that describe a given app.

**User Reviews.csv:** contains reviews for each app, most helpful first. The text in each review has been preprocessed and attributed with three new features: Sentiment (Positive, Negative or Neutral), Sentiment Polarity and Sentiment Subjectivity.

4**. Data Cleaning**

It is a crucial step which includes:

• Dropping Duplicates

• Finding NaN, Nulls and missing values

• Validating data to a standard pattern

The three features that we will be working with most frequently henceforth are Installs, Size, and Price. A careful glance of the dataset reveals that some of these columns mandate data cleaning. Specifically, the presence of special characters (, $ +) and letters (M k) in the Installs, Size, and Price columns are making their conversion to a numerical data type difficult

We also looked for nulls in columnssentiment, sentiment polarity and sentiment subjectivity in second dataset based on user reviews.

So, once the data of both datasets was cleaned, we started with exploring to reach to some conclusions

**5. Exploratory Data Visualization**

Exploratory data visualization is performed when we do not have a clue about what information lies within our dataset

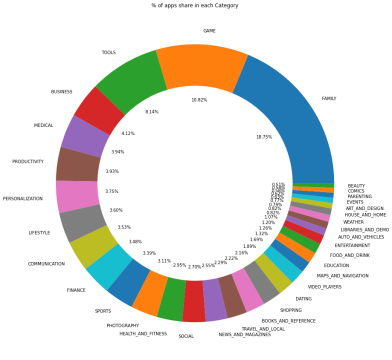
**5.1 Exploring 'Categories' of Apps**

With more than 1 billion active users in 190 countries around the world, Google Play continues to be an important distribution platform to build a global audience. For businesses to get their apps in front of users, it's important to make them more quickly and easily discoverable on Google Play. To improve the overall search experience, Google has introduced the concept of grouping apps into categories.

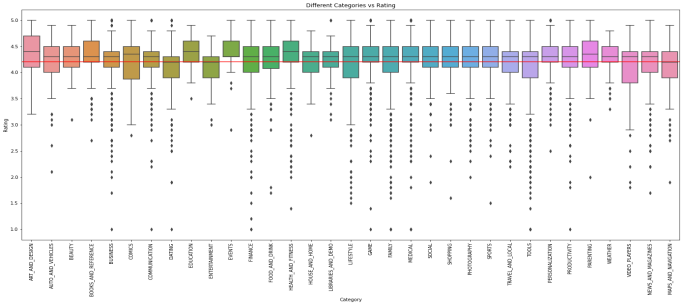
According to this univariate analysis, we observed that there were a greater number of apps related to categories

• Family- 18.75%

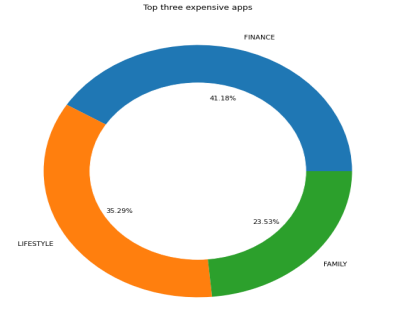
• Game- 10.82%

• Tools-8.14%

But, out of these categories apps related to Games and medical has the highest market prevalence according to their ratings, though the no of apps is more related to family and tools as well



Now, let us look at the top three most expensive apps as there are many factors to consider when selecting the right pricing strategy for your mobile app. It is important to consider the willingness of your customer to pay for your app. A wrong price could break the deal before the download even happens. Potential customers could be turned off by what they perceive to be a shocking cost, or they might delete an app they’ve

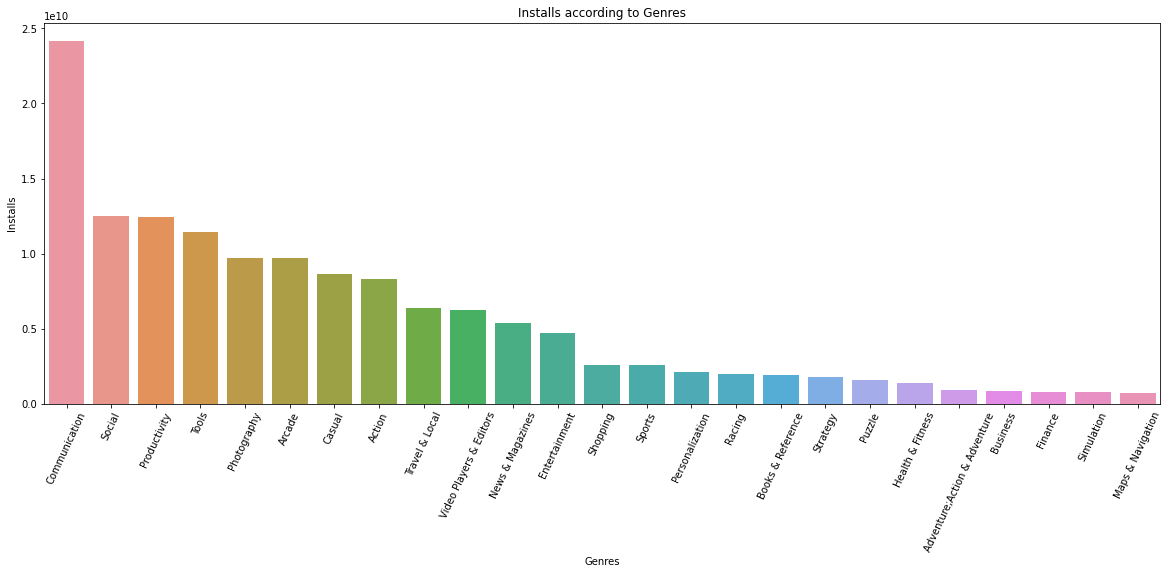
downloaded after receiving too many ads or simply not getting their money's worth. 

From above it is good to see that apps related to finance, lifestyle and family are the most expensive apps, while daily usage apps for mass use are free for the users.

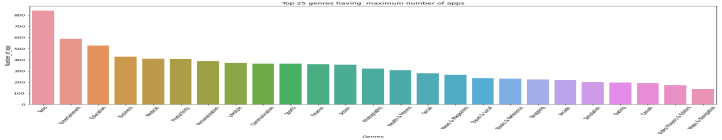
**5.2 Diving into Genres**

Among the 119 genres we found in our dataset, we found that Tools, Entertainment and education share maximum number of Apps. With the help of bar plot, we plotted the top 25 genres having maximum number of apps.

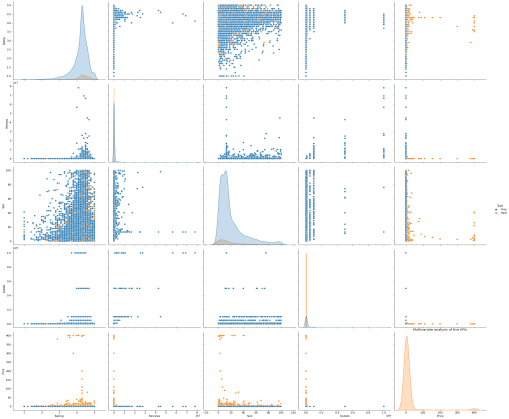
But, the number of installations of apps are more for Communication, Social and productivity.

****

But, the number of installations of apps are more for Communication, Social and productivity.

****

• users have paid more for lite apps • Users have paid highly for good rated apps and it does not depend on number of reviews.

****

And for further detailed analysis we plotted heatmap to find the correlation between each of these five KPIs.

Through the above heatmap we found that

Users have installed more apps of the ones having high reviews i.e installs and reviews are highly corelated.

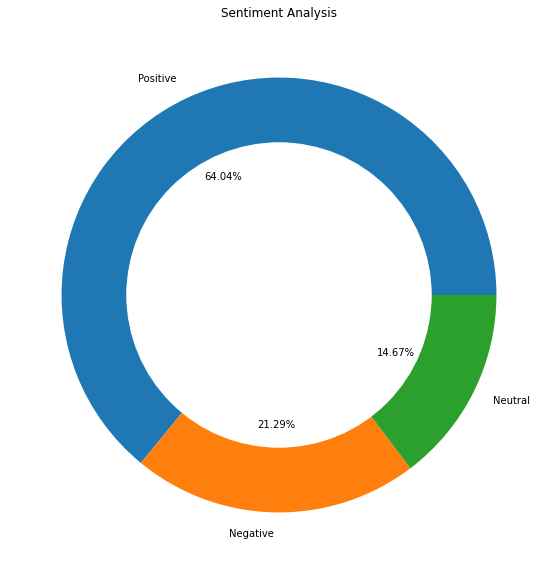
• Installs of app doesn't depends on Size i.e. there is no correlation between them

**5.4 Sentiment Analysis**

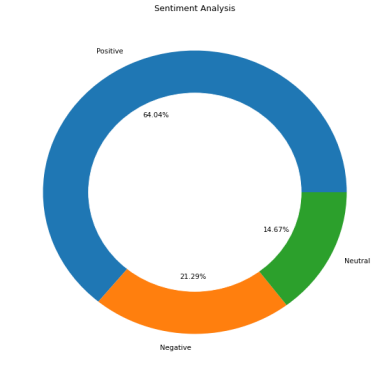
Mining user review data to determine how people feel about any product, brand, or service can be done using a technique called sentiment analysis. User reviews for apps can be analyzed to identify if the mood is positive, negative or neutral about that app. For example, positive words in an app review might include words such as 'amazing', 'friendly', 'good', 'great', and 'love'. Negative words might be words like 'malware', 'hate', 'problem', 'refund', and 'incompetent'.

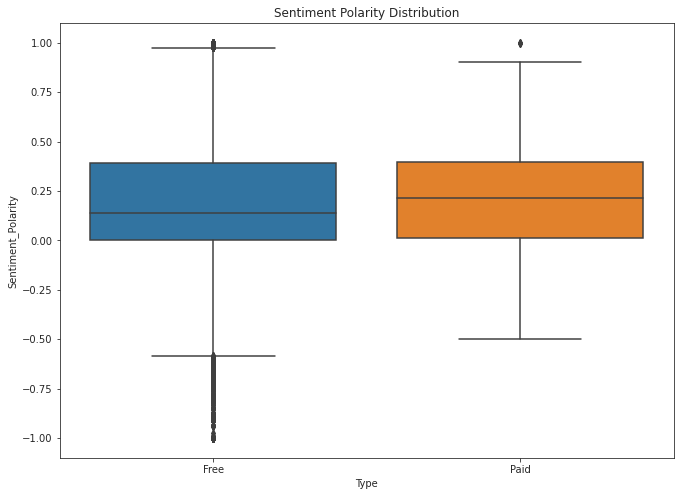
During the analysis we found that generally users give

• 64.04% - positive sentiments • 21.29% - Neutral sentiments • 14.67% - Negative sentiments

****

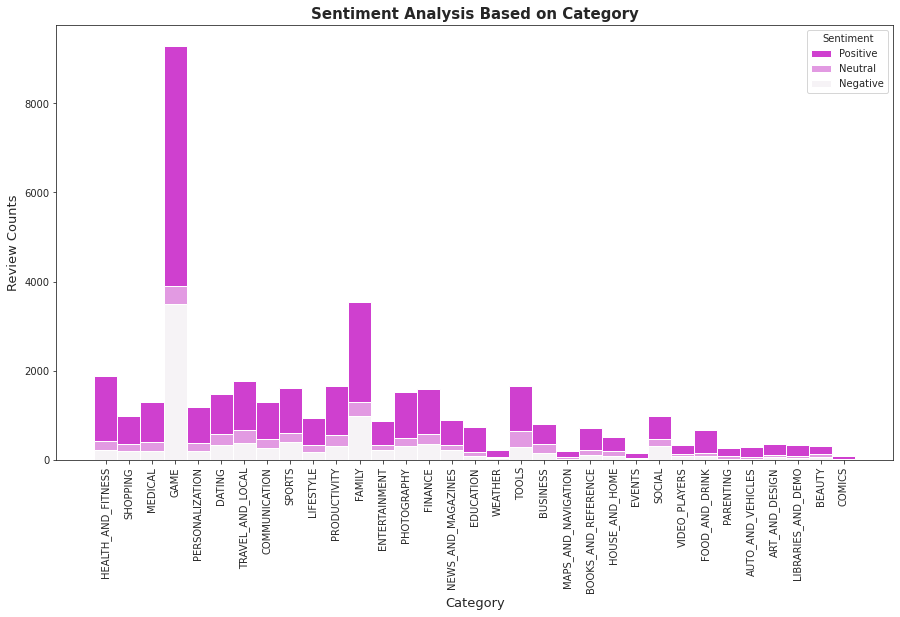
Now, to further dive deep into sentiment analysis and find that for which type of apps, users are generally harsh, we merged the two datasets i.e we joined category and type with user reviews dataset and plotted doughnut chart for analysis.plotted doughnut chart for

analysis.

By comparing above two plots, it was interesting to note that users gave more positive sentiments for paid apps. Also, looking below at the outliers of box plot analysis on sentiment polarity we found that users were harsh while reviewing free apps****

**5.5 Sentiment Analysis based on category and reviews**

The graph combining the mainstreams of both datasets clearly reveals that users gave maximum reviews for game category followed by family. Also, the positive and negative sentiments are highest for games category.

**** Thus, from these we can conclude that if the apps are good and user-friendly like for paid ones, users give positive sentiments. Thus, developers should try to provide good user experience through their apps and must belong to top 5 categories and genres to gain organic ratings and reviews

**6. Conclusion**

The marketplace is crowded with all types of mobile apps as it has enormous potential to drive app-making businesses to success. Actionable insights can be drawn for developers to work on and capture the Android market. Today, even businesses who never would’ve needed apps in the past are getting into the game to meet customer demand for an app that makes their lives easier or to stay ahead of the tech curve. Thus, through this exhaustive analysis of approx. 10k apps we discovered following key factors responsible for app engagement and success in recent study:

1. We found that among 33 unique app categories present in our dataset, apps related to Games and Medical have the highest market prevalence according to their ratings, though the number of apps are more for family and tools as well.
2. . But, the number of installs is higher for Communication, Social and productivity genres whereas Tools, Entertainment, Education related genres have maximum number of apps among all 119 genres.
3. t was interesting to note that o Users have paid more for lite and high rating apps. o Users have installed more apps of the ones having high reviews o Installs of app doesn't depends on Size.
4. . From the sentiment analysis we found users showed more positive sentiments for paid apps and were harsh while showing sentiments for free apps. Also, Neutral sentiments were less for paid ones i.e users were true to reviews once they have paid for it.
5. . Users gave maximum reviews for game category followed by family. Also, the positive and negative sentiments were highest for games category.