**Play-store-data-analysis**

[](https://user-images.githubusercontent.com/95841292/182010746-e342365c-4930-4f1c-afe6-728979504318.png)

Analysis of play store app data and its sentiment analysis.

**Introduction**

In this project, i investigate the different variables of Apps on Google Play Store that affect the application and the top relevant user reviews. i attempt to use our analysis to answer the following questions:

**What makes an app popular?**

What are some interesting trends that we can observe about user behaviour & sentiment on app usage? Can we predict app popularity if given a set of features about the app? Mobile app market has grown ~20% in the last 4 years. Android apps market comprises 90% of its market share. Our analysis has the potential to be scaled and applied to identify and solve the following problems for the much wider mobile app industry:

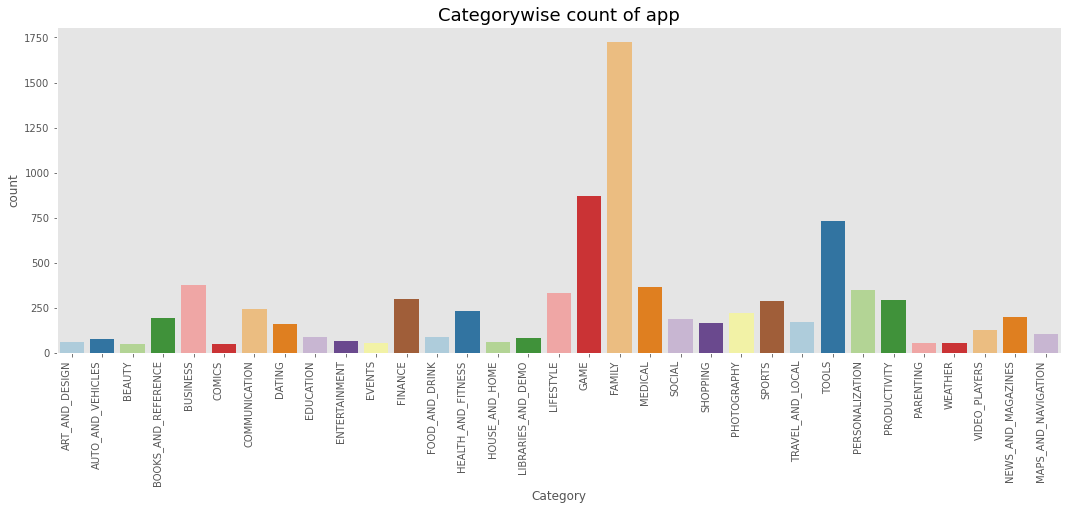
Prediction of app popularity to gauge revenue generated & optimize investment strategy for app development Identifying untapped segments in the app market Identifying fake/junk apps that spam the play store

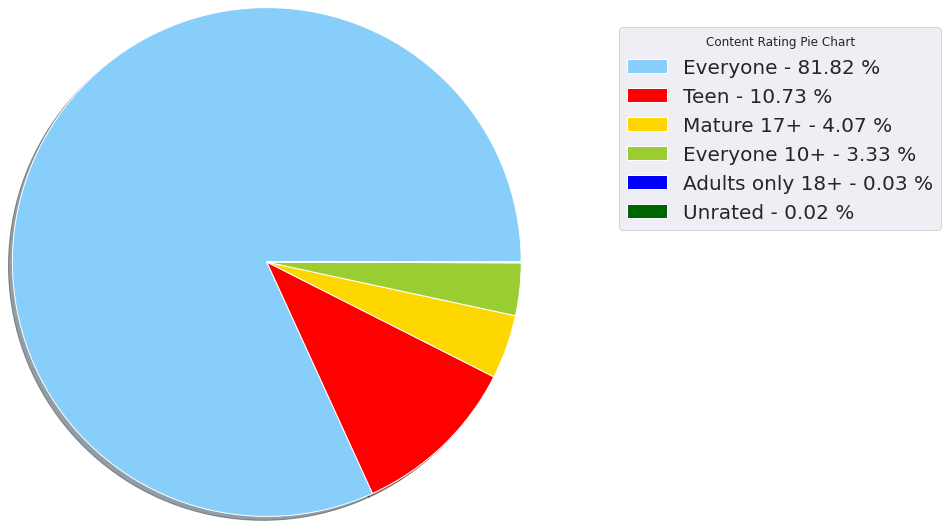
**Data Cleaning**

We started the analysis by creating a data dictionary to understand the structure of the dataset and what each feature represents. Post that, we handled the missing values in some of the columns by either dropping the rows or imputing them with the mode values, depending on the percentage of nulls in each feature. We transformed and some of the columns like Installs, Size and Price to numeric type for ease of analysis.

**Exploratory Data Analysis**

After establishing a good sense of each feature, we proceeded with plotting a pairwise plot between all the quantitative variables to look for any evident patterns or relationships between the features.

[](https://user-images.githubusercontent.com/95841292/203572427-6d19cacd-7ac5-415d-876c-7535159d6d57.png)

[](https://user-images.githubusercontent.com/95841292/203572755-fd50c371-198e-45e9-9f19-164c91b567c9.png)

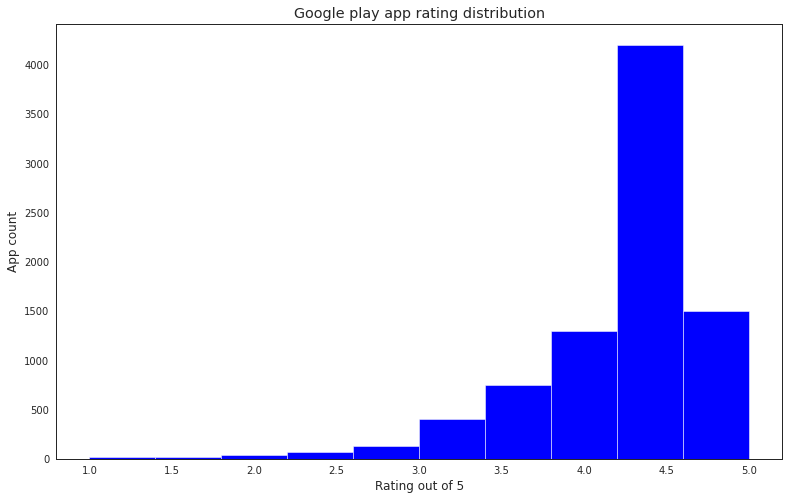
**Android Market Breakdown**

We broke down the apps by category and found that the Family and Game categories have the highest market prevalence. The Business, Tools and Medical apps are also catching up. We also checked how each category performed in terms of number of reviews, size and installation count. We found that the gaming category has the maximum number of reviews and install count but the family apps consume more space in the play store.

Findings and Conclusion Developers should aim to keep the apps optimally sized (between 2 MB to 40 MB) to increase the likelihood of it being popular. Paid apps designed to fulfil specific functions tend to be small sized and are more likely to meet user expectations. Majority of the apps should be free. For apps with no ads, apps can be priced under $10. Free apps tend to outperform paid apps both in volume, install counts and ratings. There is a positive correlation between installs and ratings The category (Game) is a potential unsaturated space for app developers. Apps in this category tend to have high installs. Highly installed apps such as Communication, Productivity & Photography are not as highly rated indicating dissatisfaction among customers and potential untapped market segments. For paid apps, users review harshly. Basic sentiment analysis revealed some common issues like loading time and inconvenience with apps may contribute to a negative rating. Usability was one of the main reasons apps were rating positive.

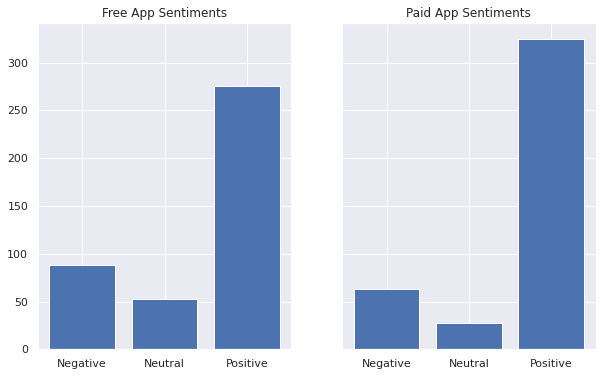
**Rating Distribution**

The rating distribution revealed that most apps perform reasonably well with an average rating of 4.17. We broke down the average rating by category to check if any category performs exceedingly good or bad. We conducted a One-way Anova Test and confirmed that the average ratings across categories is statistically different. The Health and Fitness and Books and Reference produce the best apps with 50% apps having a rating greater than 4.5. Interestingly, half of the Dating apps have a rating lower than the average.

[](https://user-images.githubusercontent.com/95841292/203572922-2f50e80e-7339-4d67-ba94-5ed372e5e475.png)

**Basic Sentiment Analysis – User Reviews**

We plotted the fraction of positive, negative and neutral reviews for each category and observed that the Health and Fitness apps perform the best with more than 85% positive reviews. On the other hand, Game and Social apps have a higher fraction of negative reviews. We compared the reviews between free and paid apps and found that people are harsher towards free apps whereas users are more tolerant when they are paying for it.

[](https://user-images.githubusercontent.com/95841292/203572522-bbe8b0e1-edd7-4d89-8858-d45cdd0083ea.png)

**Frequency of Words in Reviews**

We created a word cloud of commonly occurring words in positive and negative reviews and found that the words – “love”, “great” and “good” were the most commonly occurring words in the positive reviews. On the other hand, the negative words that were prevalent were “bad”, “hate” and “ads”. Our aim was to analyse the reviews and get a better idea of the common issues that people face with apps or the attributes that make an app popular. We extracted phrases from the reviews and observed that positive reviews had phrases like “user friendly”, “free version”, “works great” and “highly recommend”. The negative reviews contained phrases like “waste time”, “many ads”, “spend money” and “takes forever”. We can see that loading time and ads were one of the main concerns amongst users. On the other hand, usability is one of the reasons that users give positive reviews.

**Positive comments**

[](https://user-images.githubusercontent.com/95841292/203573242-27f4d604-609f-407a-9894-a72d412bff85.png)

**Findings and Conclusion**

1)Developers should aim to keep the apps optimally sized (between 2 MB to 40 MB) to increase the likelihood of it being popular.

2)Paid apps designed to fulfil specific functions tend to be small sized and are more likely to meet user expectations.

3)Majority of the apps should be free. For apps with no ads, apps can be priced under $10.

4)Free apps tend to outperform paid apps both in volume, install counts and ratings.

5)There is a positive correlation between installs and ratings

6)For paid apps, users review harshly.

7)Basic sentiment analysis revealed some common issues like loading time and inconvenience with apps may contribute to a negative rating. Usability was one of the main reasons apps were rating positive.

**📜 Credits**

Jasmeen kaur | Data enthusiast

Special thanks to AlmaBetter

Contact me for Data Science Project Collaborations

[image](https://www.linkedin.com/in/sameersatpute/) [image](https://github.com/sameersat96) [image](https://medium.com/@sameersatpute) [image](https://user-images.githubusercontent.com/95841292/202914940-5d5eba71-e45d-4e95-8dfe-65e45d255aec.png)

**📚 References**

1. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5756316/>
2. <https://towardsdatascience.com/sentiments-of-google-play-store-apps-reviews-13f9ee7246c6>