



Analysis of Google Play Store Apps and User Reviews

Android, the leading mobile operating system with a staggering 85% market share, owes much of its success to the Google Play Store, the largest hub for mobile applications. This project focuses on analyzing two critical datasets: Play Store app metadata and user reviews. By exploring app performance metrics, user sentiment, and prevailing market trends, the analysis aims to deliver valuable insights for enhancing app development, refining marketing strategies, and improving user satisfaction. This approach empowers Android app developers and marketers to make data-driven decisions, fostering a robust and thriving app ecosystem.

By Chandan Aruk

Objective of the Project

- **Goal:**
 - To analyze the Google Play Store apps and user reviews to extract actionable insights for app development and marketing.
- **Key Areas of Focus:**
 - Data cleaning and preprocessing
 - Sentiment analysis and its impact on ratings
 - Key app categories and their performance





Datasets Overview

- **Google Play Store Dataset:**
 - Contains app details like name, category, rating, installs, size, etc.
 - **Rows:** 10841 | **Columns:** 13
- **User Reviews Dataset:**
 - Contains reviews, sentiments, polarity, and subjectivity.
 - **Rows:** 64295 | **Columns:** 5



Data Cleaning and Wrangling

- **Play Store Dataset:**
 - Handled duplicate rows and Dropping Invalid rows
 - Data Type Conversion
 - Handled missing values for **Rating**, **Size**, and **Installs**, etc.
- **User Reviews Dataset:**
 - Filled missing values in **Translated_Review** and **Sentiment** using mode.
 - Filled **Sentiment_Polarity** and **Sentiment_Subjectivity** with mean.
 - Dropped rows with all missing values in app groups.



Big Data Analytics



Understanding the Dataset

This project leverages two datasets to analyze app performance and user feedback:

Google Play Store Dataset:

- This dataset provides comprehensive information about apps, including categories, ratings, reviews, size, price, content ratings, and the number of installs.
- Key columns include:
 - **Rating:** Indicates the average user rating (scale of 1 to 5).
 - **Installs:** Reflects the popularity of apps by showing download counts.
 - **Size:** Provides insights into app file size, impacting user preferences.
 - **Content Rating:** Categorizes apps based on their target audience (e.g., Everyone, Teen).
- Purpose: This dataset helps identify trends, high-performing app categories, and factors influencing app success.

User Reviews Dataset:

- This dataset contains detailed user feedback, including translated reviews, sentiment labels, and sentiment scores.
- Key columns include:
 - **Translated Review:** The textual feedback provided by users.
 - **Sentiment:** Categorized as Positive, Neutral, or Negative.
 - **Sentiment Polarity:** A numeric value representing the sentiment's intensity (positive or negative).
 - **Sentiment Subjectivity:** Indicates whether feedback is opinion-based or factual.
- Purpose: This dataset helps analyze user satisfaction and sentiment trends.

Importance of Understanding the Dataset

The Play Store dataset allows us to examine app performance through metrics like ratings and installs, while the User Reviews dataset provides insights into user opinions and satisfaction. Together, they offer a holistic view of app success, enabling data-driven decisions for app development and marketing.



Understanding App Performance

1. Metrics to Assess Performance:

- **Ratings:** Reflect user satisfaction and app quality.
- **Reviews:** Indicate user engagement and provide qualitative feedback.
- **Installs:** Measure app popularity and reach.
- **Sentiments:** Analyze user opinions (positive, neutral, negative) from reviews.

2. Key Findings from the Play Store Dataset:

- **Top Performing Categories:**
 - **Social, Communication, and Games** apps dominate in installs and reviews, showing high user engagement.
 - Categories like **Medical** and **Events** receive fewer installs and reviews, indicating a niche audience.
- **Content Rating Impact:**
 - Apps rated "Everyone" consistently received higher ratings and installs, making them the most successful among all age groups.
 - "Teen" and "Everyone 10+" also showed strong performance but slightly lower than "Everyone" apps.

3. Key Findings from the User Reviews Dataset:

- **Sentiment Analysis:**
 - The majority of reviews are positive, indicating user satisfaction across most apps.
 - Negative reviews are concentrated in apps with low ratings or technical issues.
- **Polarity and Subjectivity:**
 - Apps with high ratings often exhibit high polarity (positive sentiment) and moderate subjectivity, showing a balance between opinionated and factual feedback.

4. Combined Insights:

- Apps with high installs and ratings often have positive sentiment, reinforcing their success.
- Categories with strong engagement can further improve by addressing user feedback to maintain positive sentiment.

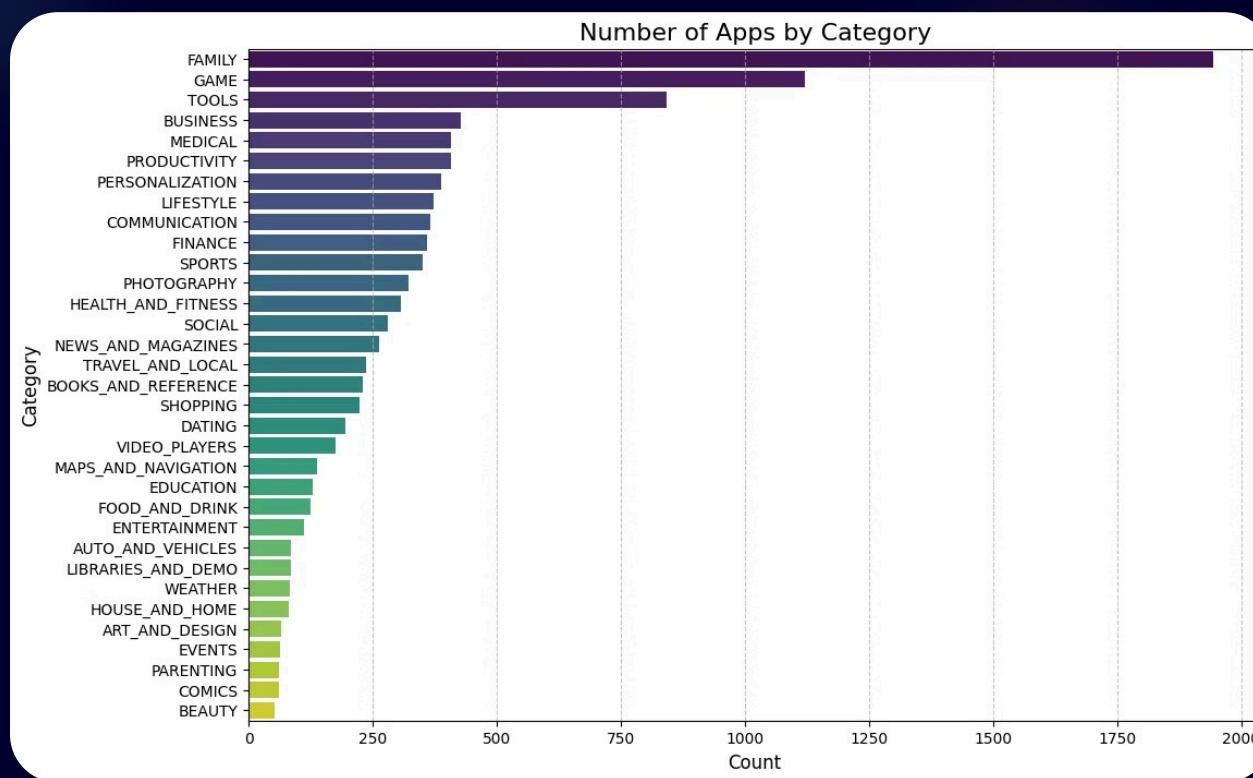
5. Business Implications:

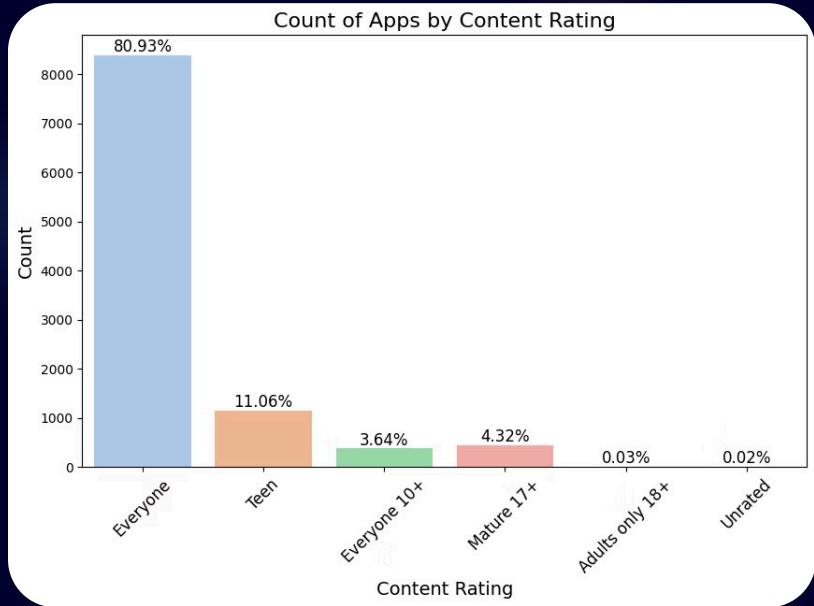
- **For Developers:** Focus on top categories (e.g., Social, Communication) and maintain high user engagement through updates and features.
- **For Marketers:** Highlight positive reviews and ratings to attract new users.
- **For Improvement:** Address critical feedback from negative reviews to improve app quality and user satisfaction.

Key Insights

Highest Number of Apps by Category

- The Family category has the highest number of apps, followed by Game and Tools.
- Categories like Comics, Beauty, and Parenting have the least number of apps, indicating niche markets.



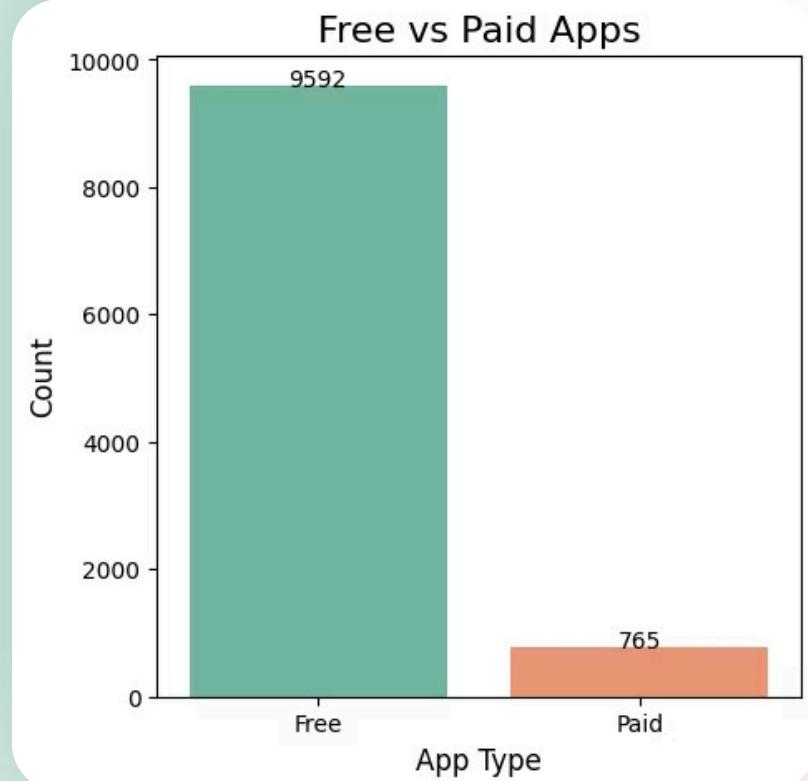


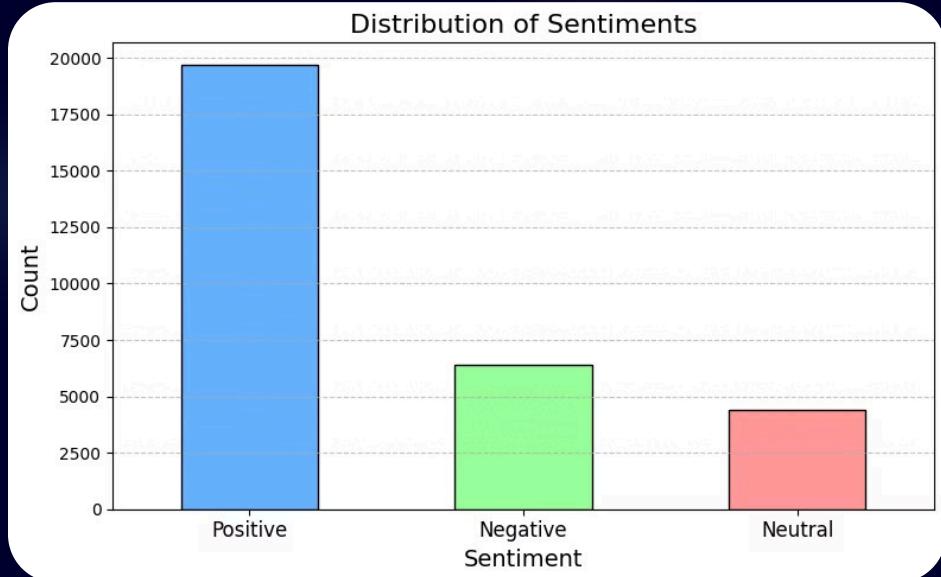
Highest Count of Apps by Content Rating

1. **Majority Apps for Everyone:** A significant majority (80.93%) of apps are rated as "Everyone," indicating that most developers aim to target a broad and general audience.
2. **Minor Categories:**
 - Content ratings like "Teen" (11.06%) and "Mature 17+" (4.32%) have relatively smaller shares, indicating niche markets.
 - "Adults only 18+" and "Unrated" categories are almost negligible.

Free vs Paid Apps

- Free apps dominate the market with 9,592 compared to 765 paid apps.
- Users prefer free apps, making them a better option for mass adoption.
- Businesses should focus on free or freemium models to maximize reach.
- Paid apps can succeed in niche markets by offering unique value.





Sentiment Distribution

- **Positive Sentiments Dominate:** The majority of sentiments are positive, with a count significantly higher than negative or neutral sentiments. This indicates an overall favorable reception or perception.
- **Negative vs Neutral:** Negative sentiments slightly outnumber neutral ones, suggesting a tendency to express dissatisfaction over remaining indifferent.
- **Engagement Patterns:** The prevalence of positive sentiment indicates high user satisfaction or effective strategies leading to positive feedback.

▼ Price Analysis:

- **Most apps are free** (median price: 0), but prices can go as high as **400**, reflecting a wide range of app pricing strategies.

▼ App Size Details:

- The **average app size** is about **22.06 MB**, ranging from **0.008 MB** to **100 MB**.
- Filtering out very small or excessively large apps (e.g., below 1 MB or above 100 MB) can help refine the dataset for meaningful insights.

▼ Last Updated Trends:

- On average, apps were last updated around **November 2017**, with most updates occurring between **2010 and 2018**.

▼ Genre Diversity:

- With over **120 genres**, the platform offers an impressive variety of apps, catering to diverse user needs and interests.

▼ Review Observations:

- Reviews are **predominantly positive**, although there's noticeable variation in user opinions.

▼ Translated Reviews:

- The translated reviews further emphasize the **positive sentiment**, as users frequently use words like "**good**," "**love**," and "**great**" to describe their experiences.
- A recurring theme in user feedback revolves around **ease of use**, **functionality**, and **regular updates**. These factors seem to play a key role in maintaining user satisfaction and loyalty.
- Apps that deliver engaging content, enjoyable gameplay, and seamless usability tend to receive the most praise, suggesting that these are critical areas for developers to focus on.



▼ Sentiment Subjectivity:

- Reviews exhibit a **moderate level of subjectivity**, which reflects a healthy mix of opinions and factual observations.
- While some users focus on sharing personal experiences, others highlight functional aspects of the app, making the reviews a combination of opinion-based and objective feedback.

▼ Sentiment Polarity:

- The majority of reviews fall between **neutral to slightly positive**, suggesting a generally favorable perception of the apps on the platform.
- There is minimal evidence of **extreme sentiment values**, indicating that users rarely provide overly negative or highly emotional feedback. This trend implies a balanced user experience where most users are neither completely dissatisfied nor overwhelmingly impressed.



Solution to Business Objective

Targeting Broad Reach

The FAMILY, GAME, and ENTERTAINMENT categories attract diverse users, increasing potential downloads and engagement. Businesses can focus on these categories for greater visibility and growth.

Niche Market Strategies

For premium pricing, FINANCE and LIFESTYLE offer opportunities to cater to specific needs and desires. Unique features and high-value offerings attract users willing to pay more for specialized solutions.

From the analysis, the **FAMILY** category offers significant potential as it appeals to users of all age groups, making it highly inclusive. Similarly, the **GAME** category holds substantial growth opportunities due to its broad engagement and consistent popularity among Android users. These categories provide businesses with opportunities to create apps that attract diverse audiences and drive high user interaction.

Categories like **FAMILY**, **SOCIAL**, **COMMUNICATION**, and **ENTERTAINMENT** currently dominate in terms of installs, ratings, and reviews. This reflects user preferences for apps that provide utility, social connectivity, and entertainment. However, these popular categories do not appear among the top five most expensive apps, which are primarily from **FINANCE** and **LIFESTYLE**. This suggests that premium-priced apps tend to succeed in specialized niches that cater to specific needs, such as financial tools or luxury lifestyle solutions.

In conclusion, businesses aiming for broader market reach should focus on developing apps in the FAMILY, GAME, and ENTERTAINMENT categories, which align with user trends and high engagement. Meanwhile, those targeting higher revenues through premium pricing can explore opportunities in niche categories like FINANCE and LIFESTYLE by offering unique, high-value features that appeal to specialized audiences.

Conclusion



Trending Apps

Most of the trending apps (in terms of users' installs) are from the categories like GAME, COMMUNICATION, and TOOL even though the amount of available apps from these categories are twice as much lesser than the category FAMILY.



App Quality

The trending of these apps are most probably due to their nature of being able to entertain or assist the user. Besides, it also shows a good trend where we can see that developers from these categories are focusing on the quality instead of the quantity of the apps.



High Ratings

Most of the apps having good ratings of above 4.0 are mostly confirmed to have high amount of reviews and user installs.



Social and Game Apps

There are some spikes in term of size and price but it shouldn't reflect that apps with high rating are mostly big in size and pricy as by looking at the graphs they are most probably are due to some minority. Furthermore, most of the apps that are having high amount of reviews are from the categories of SOCIAL and GAME like Instagram, Clash of Clans etc.

About me



Hello! I'm Chandan Aruk

I'm passionate about exploring the world of Data Science, Machine Learning, and Artificial Intelligence.



What I'm Learning

Currently pursuing:

Advanced Certification in Full Stack Data Science and AI



My Interests

- Data Science
- Data Analysis
- Machine Learning & Deep Learning
- Artificial Intelligence (AI)



Connect with Me

- LinkedIn: www.linkedin.com/in/chandan-aruk-0865052b2
- Email: aruk5638@gmail.com