DATA ANALYTICS REPORT

EXPLORATORY DATA ANALYSIS OF GOOGLE PLAY STORE APPS

Market Trends, App Performance & User Behavior

App Store

TOOLS USED

Python, Pandas, Matplotlib, Seaborn & Plotly

PRESENTED BY

Tejas Jadhav

GITHUB

TABLE OF CONTENTS

PROBLEM STATEMENT & OBJECTIVE

Problem Statement:

The Google Play Store hosts millions of apps across different categories. It’s challenging to understand trends, user preferences, and which types of apps are most successful.

Objective:

To explore the dataset of Google Play Store apps and uncover patterns, insights, and trends that can help app developers and businesses make better decisions.

EDA WORKFLOW

For this analysis, a structured workflow was followed, involving data collection, understanding, cleaning, exploration, and summarization of insights to allow a clear understanding of the dataset and its trends.

01 Data Collection

- Gathered the Google Play Store dataset from Kaggle

02 Data Understanding & Anomaly Detection

- Looked at data distributions
- Found missing values, outliers, and unusual patterns

03 Data Cleaning & Treatment

- Fixed missing or incorrect values
- Standardized formats for consistency

04 Exploratory Analysis

- Univariate Analysis: e.g., Ratings, Installs, Prices
- Bivariate Analysis: Rating vs Reviews
- Multivariate Analysis: Reviews vs Rating vs Installs

05 Insights & Reporting

- Summarized patterns and trends
- Highlighted key findings for developers and businesses

KEY QUESTIONS EXPLORED

- Which App Store categories are most saturated and where do underrepresented categories offer opportunity?
- How are free and paid apps distributed across the App Store?
- How are paid apps priced, and how common are premium price points?
- Which content ratings dominate the App Store ecosystem?
- How are app sizes distributed, and do most apps remain lightweight?
- How do user ratings distribute across apps?
- How does review volume relate to rating stability and perceived quality?

DATA OVERVIEW

The dataset provides key information about iOS App Store apps, including ratings, review counts, pricing, categories, and app type. Additional attributes such as content rating, supported devices, app size, language availability and version details describe app characteristics, without any time-based update information.

Data Source: Kaggle

Dataset Size

9660
Records

13
Features

App Diversity

8196
Apps covered

33
App Categories

115
App Genres

DATA OVERVIEW

Below is a detailed description of the feature set:

Dataset Features	Type	Feature Description
App	String	Name of the application as listed on the Google Play Store
Category	Categorical	Primary category under which the app is classified
Rating	Numerical (Continuous)	Average user rating of the app on a scale of 1 to 5
Reviews	Numerical (Discrete)	Total number of user reviews submitted for the app
Size	Numerical (Continuous)	Storage size of the app as reported on the Play Store
Installs	Numerical (Discrete)	Approximate number of times the app has been installed by users
Type	Categorical	Indicates whether the app is Free or Paid
Price	Numerical (Continuous)	Price of the app (0 indicates a free app)
Content Rating	Categorical	Target age group suitability (e.g., Everyone, Teen, Mature)
Genres	Categorical	Specific genre or sub-genres associated with the app
Last Updated	Date/time	Date when the app was most recently updated
Current Ver	Categorical	Current version of the application
Android Ver	Categorical	Minimum Android OS version required to run the app

DATA QUALITY CHALLENGES & ANOMALIES

Few inconsistencies were found in the dataset, which could have affected the analysis if left unaddressed.

DATA ANOMALIES

- **Columns requiring cleaning or type conversion**:
  - **size_bytes -** needed conversion from bytes to MB/GB for readability
- **Columns with notable anomalies or invalid entries**:
  - track_name - had 2 duplicate names (7197 IDs vs 7195 unique app names)

DATA CLEANING & TREATMENT

Inconsistent were addressed, and key features were cleaned and standardized for analysis.

DATA CLEANING SUMMARY

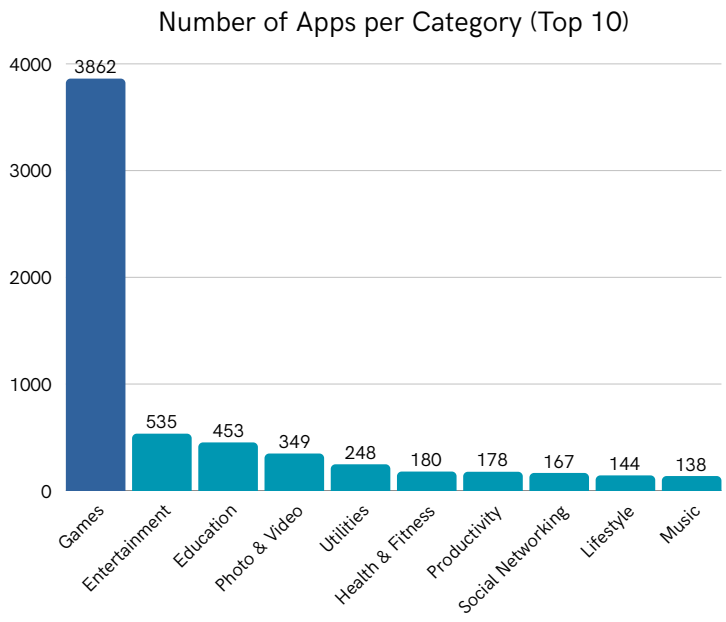
- No missing values were detected in the dataset, therefore, no missing value imputation was required.
- **size_bytes** column was converted from bytes to MB for readability
- In the **track_name** feature, there were two duplicates (7197 & 7195), which were removed
- Few of columns (**Unnamed: 0, id, size_bytes, vpp_lic**) were removed from the data frame as they didn't contribute much for the business insights

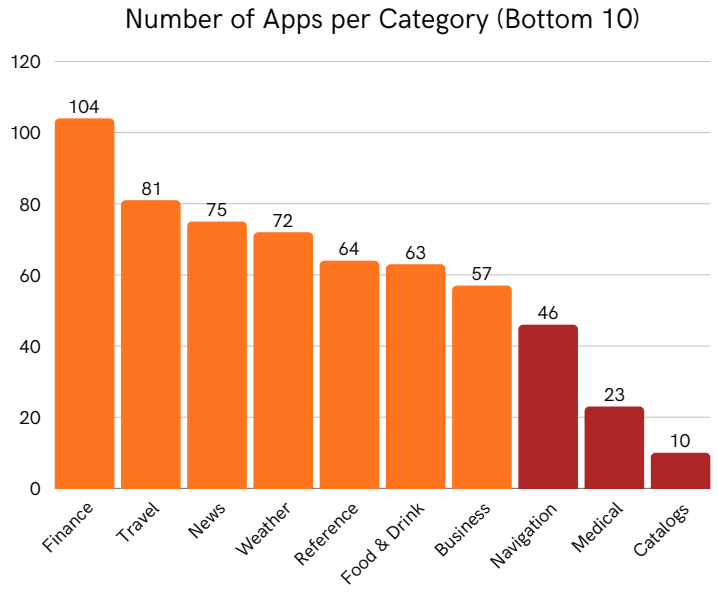
INSIGHTS

App Store supply is heavily concentrated in Games, leaving many functional categories wide open

3,578

Total Apps in Top 3 Categories





Key observations

- Games dominate the App Store by a wide margin, with a steep drop in app counts across all other categories.
- Several utility-focused categories such as Medical, Navigation, and Finance have very limited representation.

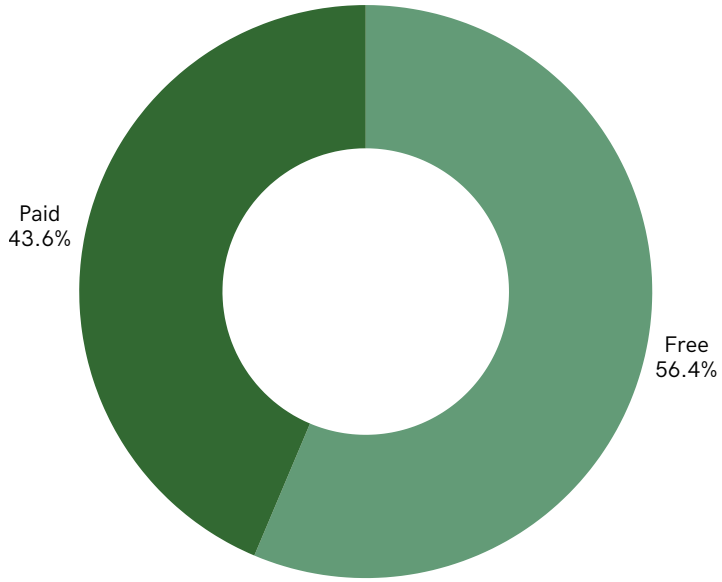
Business Insights

- Saturation in Games makes differentiation costly, requiring strong branding or heavy user acquisition spend.
- Underrepresented categories offer clearer entry points and better visibility for problem-solving or niche apps.

Free apps hold a slight majority on the App Store, but paid apps remain prevalent

8,719

Apps are free to download



Key observations

- Free apps outnumber paid apps, but the difference between the two is relatively small.
- Paid apps still account for a significant share of total App Store listings.

Business Insights

- Unlike other app ecosystems, the App Store supports paid distribution more strongly.
- Developers can still justify upfront pricing when the app delivers clear, differentiated value.

Paid apps cluster at very low prices, with only a few extreme high-price outliers

8,719

Apps are free to download

Key observations

- The vast majority of paid apps are priced at the lower end, with prices heavily concentrated near the minimum.
- A small number of apps appear at very high price points, creating a long right-skewed tail.

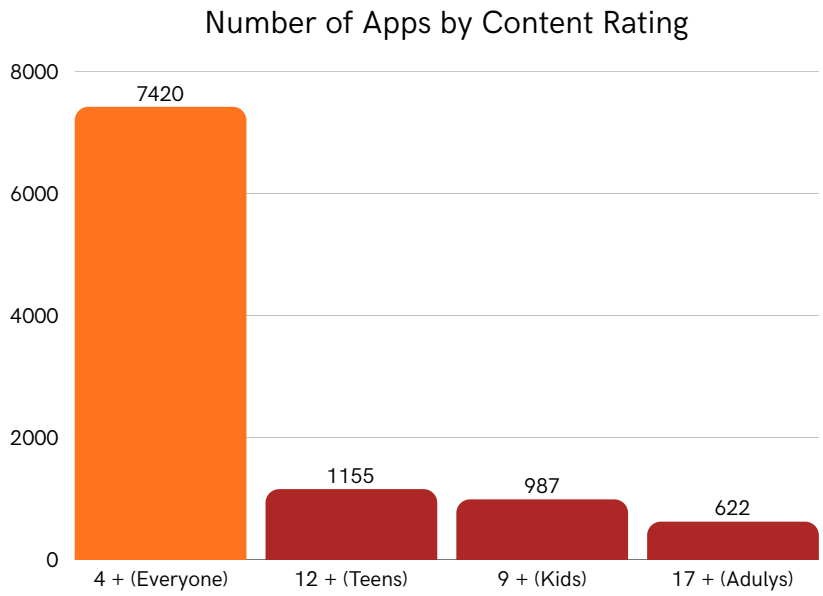
Business Insights

- Low pricing is the dominant strategy for paid apps, likely to reduce purchase friction.
- High-priced apps operate as niche or professional offerings and are not representative of the broader market.

Apps primarily target general audiences

79.2%

Apps targeted general audience



Key observations

- Apps rated for general audiences (Everyone) significantly outnumbered apps targeting teens & mature audiences.

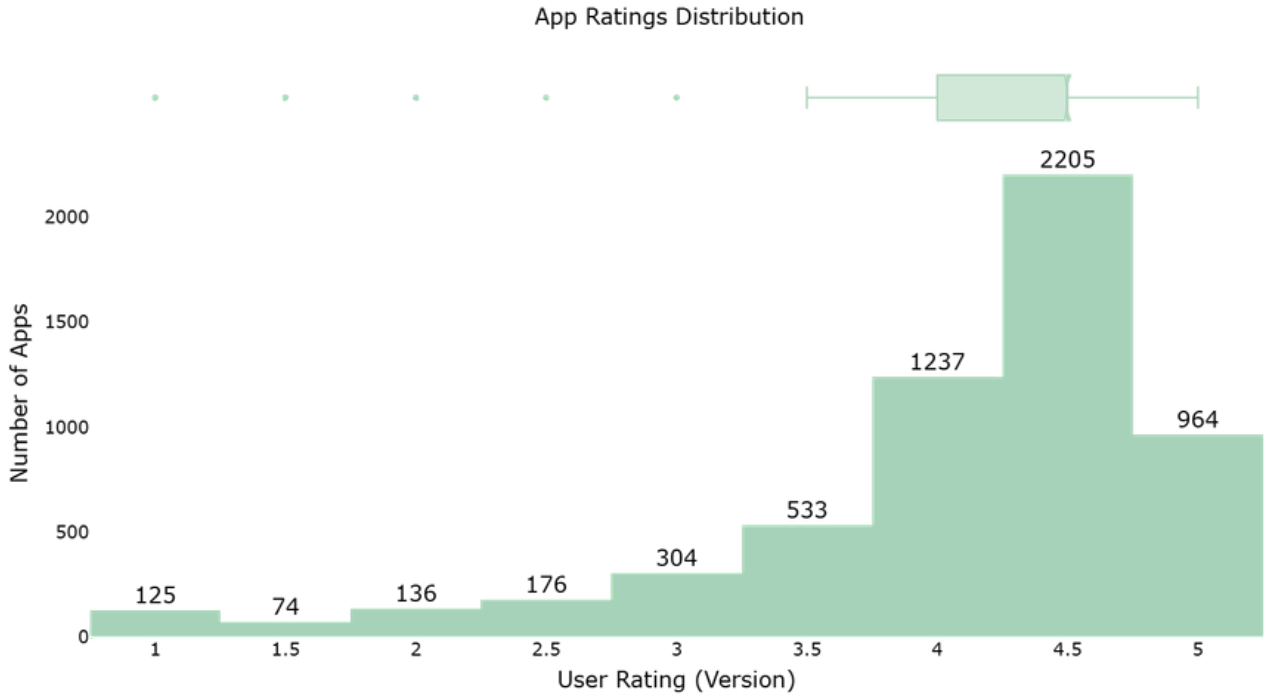
Business Insights

- Developers may favour "Everyone"-rated apps to reach a broader audience, increase downloads, and generate higher ad revenue due to fewer content restrictions.

Most App Store apps maintain high user ratings, clustering near the top end of the scale

4.3

Typical App Rating (Median)



Key observations

- App ratings are heavily concentrated between 3.5 and 5 stars, with a clear peak around 4–4.5.
- Very few apps fall below 3 stars, indicating a strong left-skewed distribution.

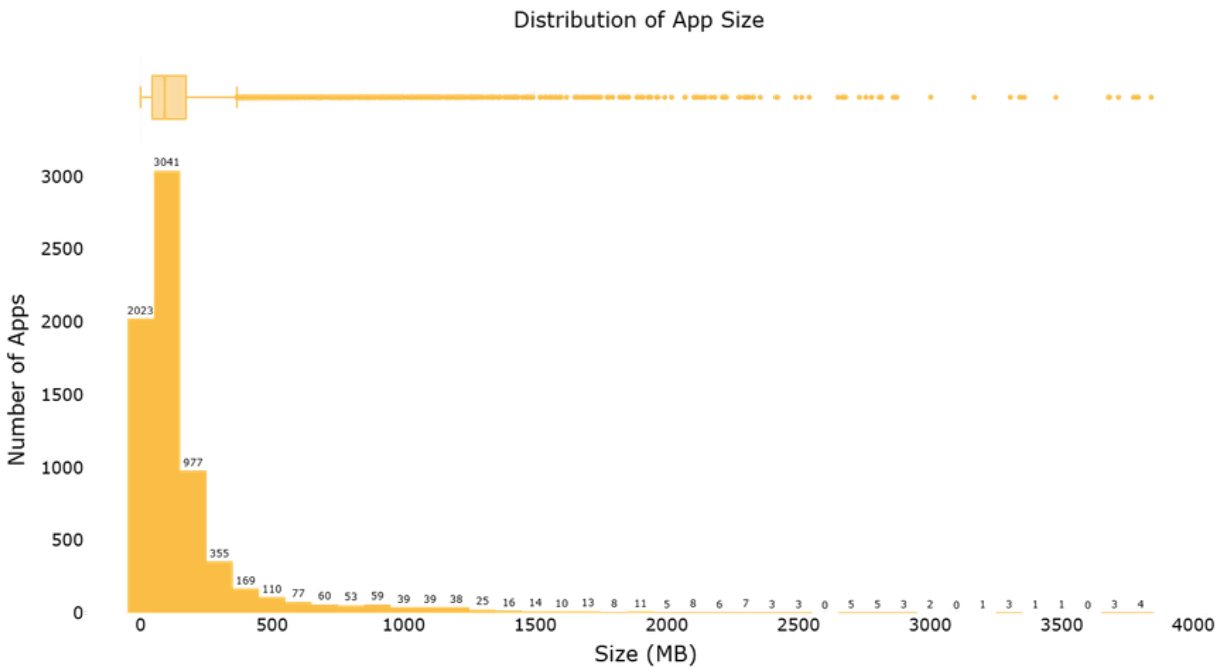
Business Insights

- High rating concentration suggests strong quality benchmarks and elevated user expectations on the App Store.
- Apps with lower ratings are likely to face discoverability and retention challenges in a quality-driven ecosystem.

App sizes are heavily skewed toward smaller builds, with a long tail of large apps

4.3

Typical App Rating (Median)



Key observations

- Most apps cluster at the lower end of the size range, with counts dropping sharply as size increases.
- A small number of very large apps create a long right-skewed distribution.

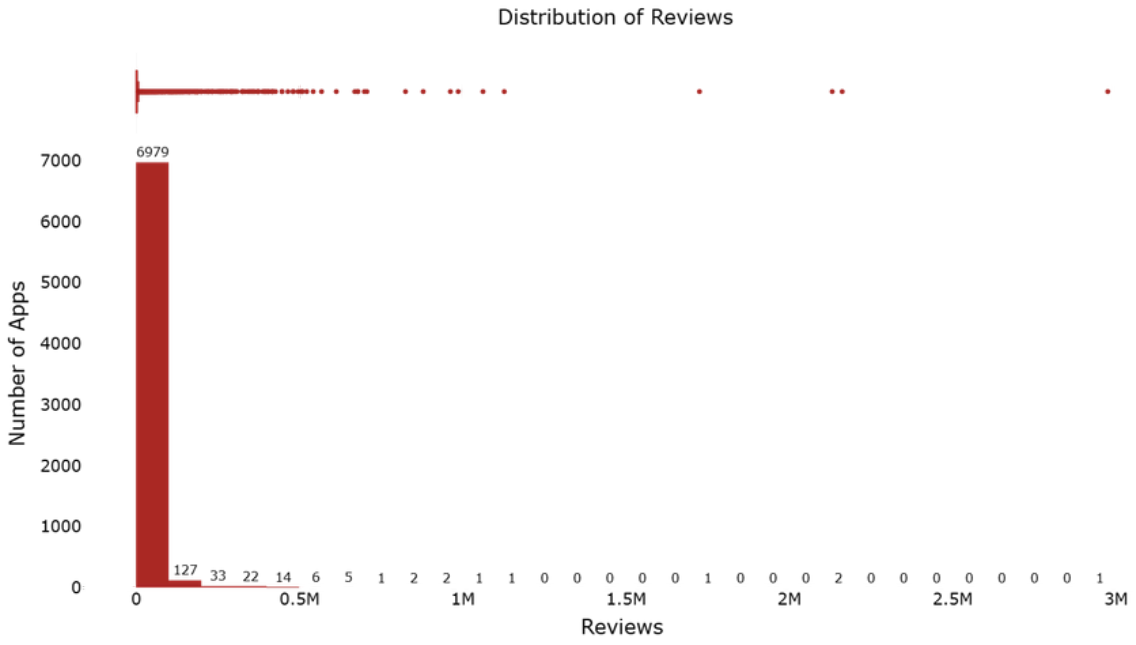
Business Insights

- Developers prioritize keeping apps lightweight to reduce download friction and storage concerns.
- Larger app sizes are likely tied to feature-rich or media-heavy use cases rather than the norm.

User engagement is concentrated in a small number of highly popular apps

5,930

Typical Number of Reviews (Median)



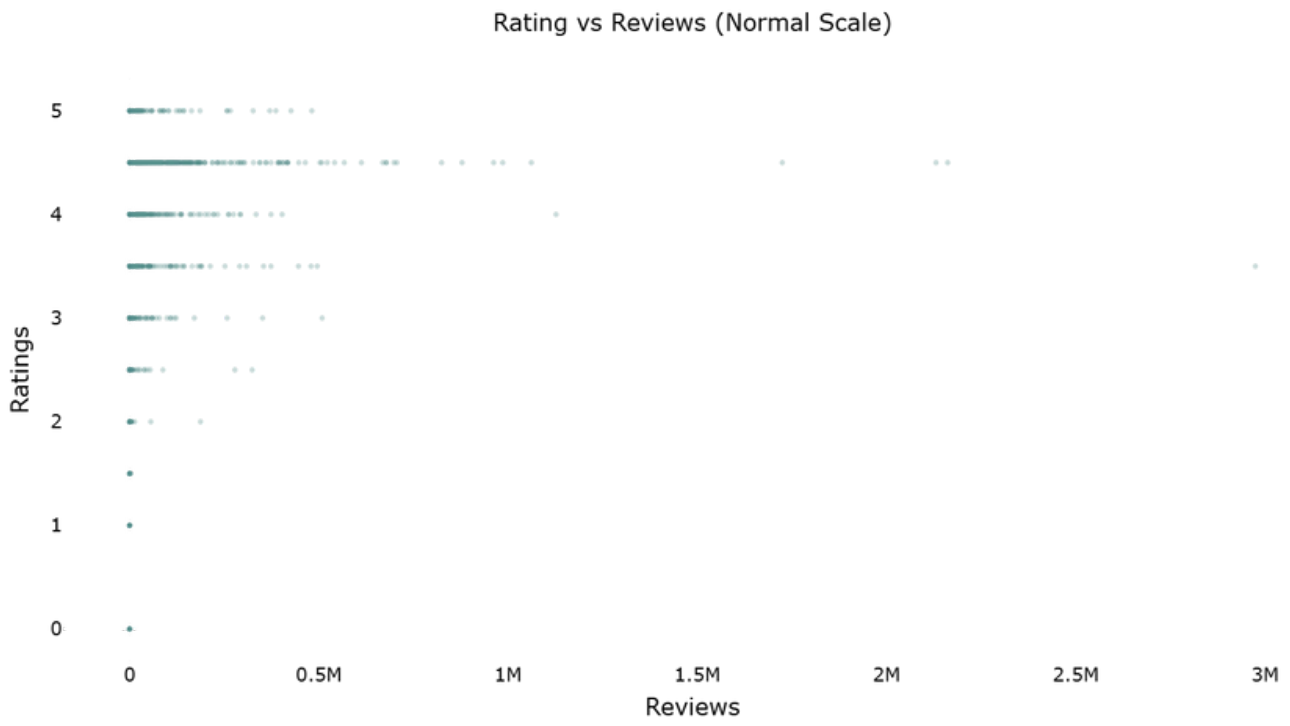
Key observations

- Most apps receive very few reviews, with counts heavily clustered near the lower end.
- A small number of apps accumulate extremely high review counts, creating a long right-skewed distribution.

Business Insights

- Visibility and user acquisition strongly favor already popular apps on the App Store.
- New apps must rely on strong differentiation or external marketing to break through low initial engagement.

Ratings Stabilize as Review Volume Increases



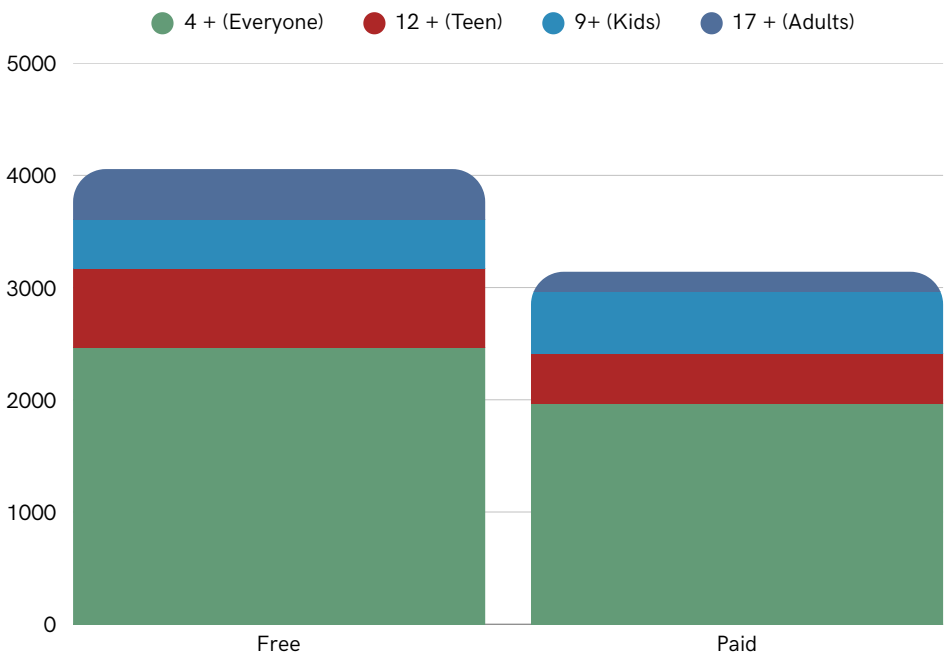
Key observations

- Ratings from low-review apps are less reliable due to high variability.
- Apps with large review counts are rare and are tightly clustered around higher ratings.

Business Insights

- Ratings from low-review apps are noisy; review volume adds credibility and reliability to ratings.
- Building early review volume improves rating stability and user trust more than chasing marginal rating gains.

Upfront Paid Apps Are Uncommon; General Users Access Most Free Apps



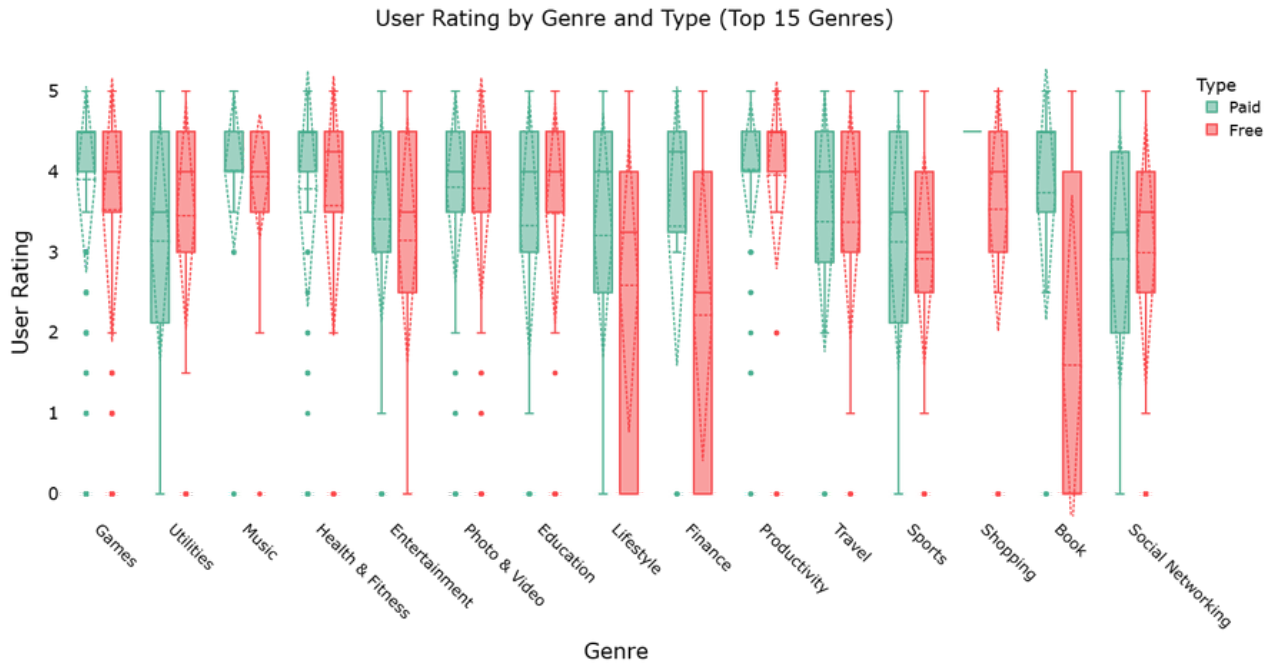
Key observations

- Apps rated 4+ form the largest share across both free and paid types.
- Free apps consistently outnumber paid apps across all content rating categories.

Business Insights

- The App Store strongly favors broad, family-safe content regardless of monetization model.
- Targeting general audiences increases reach, while paid apps must still align with mainstream content expectations.

Free apps compete on quality across most categories, narrowing the value gap for paid apps



Key observations

- Median ratings for free and paid apps are similar across most categories, with heavy overlap in distributions.
- Paid apps show limited rating advantages, appearing only in a few niche or trust-driven categories.

Business Insights

- Freemium or free-first models remain effective in categories where users do not associate price with quality.
- Upfront pricing works best in select niches where specialization, trust, or perceived value clearly justify payment.

BUSINESS / DEVELOPER TAKEAWAYS

- **Category Saturation vs Opportunity**: Highly saturated categories like Games make differentiation expensive, requiring strong branding or heavy user acquisition spend. Underrepresented categories offer clearer entry points & better visibility for problem-solving or niche apps.
- **Free Is the Default, Paid Is Still Viable on iOS**: While free apps slightly dominate, the App Store supports paid distribution more strongly than many ecosystems. Developers can justify upfront pricing when the app delivers clear, differentiated value.
- **Pricing Power Is Limited for Most Paid Apps**:
  - Free apps dominate most categories: A free-first or freemium approach works best for reach and adoption.
  - Paid apps thrive in high-trust, high-value segments: A paid- or premium-first strategy builds credibility and attracts willing buyers.
- **Monetization Strategy Should Be Category-Led:** Freemium or free-first models work best in categories where users do not associate price with quality. Paid or premium-first approaches are more effective in high-trust or high-value segments.

BUSINESS / DEVELOPER TAKEAWAYS

- **Broad, Family-Safe Content Maximizes Reach:** The App Store strongly favors "Everyone"-rated content across monetization models. Targeting general audiences increases reach, downloads, and monetization potential.
- **Quality Is a Baseline Requirement:** High rating concentration reflects strong quality benchmarks and elevated user expectations. Apps with lower ratings face significant discoverability and retention challenges.
- **Lightweight Apps Reduce Adoption Friction:** Most apps are skewed toward smaller sizes, reflecting sensitivity to download speed and storage constraints. Larger apps are exceptions, typically justified by feature-rich or media-heavy use cases.
- **Review Volume and Popularity Compound Visibility:** Ratings from low-review apps are noisy and less reliable, while engagement concentrates around already popular apps. Early review acquisition and strong differentiation are critical to break initial visibility barriers.

CHALLENGES & OPPORTUNITIES

Limitations

- The dataset represents a historical snapshot and may not reflect the current iOS App Store ecosystem.
- Geographic coverage may be limited as the dataset may not reflect global app store trends.
- Temporal relevance is limited: the dataset captures a snapshot in time, so trends from updates or recent releases may be missing.

Future Work / Opportunities:

- Perform Sentiment Analysis on user reviews to understand app perception.
- Incorporate time-based analysis using version history or update data to track changes in popularity and ratings over time.
- Explore deeper feature relationships such as price, app size, ratings, and review volume to uncover additional insights.
- Conduct cross-platform comparisons with Android apps to identify platform-specific monetization and quality dynamics.

CONCLUSION

This analysis of iOS App Store apps provided a comprehensive understanding of category dynamics, user ratings, and app characteristics. Key patterns and opportunities were identified, demonstrating how data-driven insights can inform strategic decisions in the mobile app market. Future analysis could include sentiment analysis of user reviews, time-series tracking of app performance, and cross-platform comparisons with the Apple App Store to uncover deeper insights.

This analysis was conducted using Python, Pandas, Matplotlib, Seaborn and Plotly.

Author:

Tejas Jadhav

GitHub: **@tejas-jadhav**

LinkedIn: **Tejas Jadhav**