EXPLORATORY DATA ANALYSIS OF GOOGLE PLAY STORE APPS

Market Trends, App Performance & User Behavior

Google Play

TOOLS USED

Python, Pandas, Matplotlib, Seaborn & Plotly

PRESENTED BY

Tejas Jadhav

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
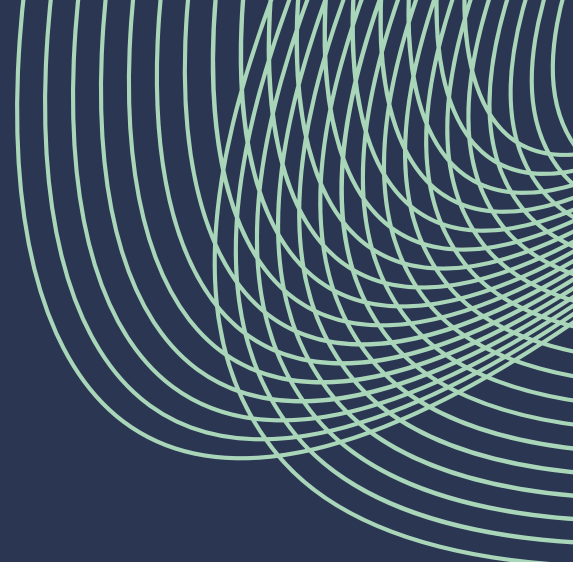- Standardized formats for consistency

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 categories dominate the Play Store, and where is competition highest?

- What is the distribution of Free vs Paid apps across categories and ratings?

- Who are the target audiences of apps (content rating distribution)?

- How do app sizes vary, and what is the optimal size range?

- How do app ratings behave across categories, review volumes, and time?

- What monetization models work best by category and audience?

DATA OVERVIEW

The dataset provides key information about Google Play Store apps, including ratings, reviews, installs, price, category, and type. Additional details like content rating, genres, app size, last update, and Android version capture app characteristics and trends.

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

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

DATA ANOMALIES

- **Missing values** are mainly present in the Rating column, with very few missing entries in Current Ver, Android Ver, Content Rating, and Type.
- **Columns requiring cleaning or type conversion**: Reviews, Size, Installs, Type, Price, Last Updated, Android Ver
- **Columns with notable anomalies or invalid entries**:
  - Category (e.g., 1.9)
  - Rating (e.g., 19.)
  - Type (e.g., 0)
  - Installs (e.g., Free)
  - Price (e.g., Everyone)
  - Size and Android Ver (e.g., Varies with device)

DATA CLEANING & TREATMENT

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

DATA CLEANING SUMMARY

- Missing values, mainly in **Rating** and a few version-related fields, were handled using median imputation.
- Features such as **Reviews, Size, Installs, Price, Type, Last Updated, and Android Ver** were cleaned and standardized.
- Inconsistent and invalid entries (e.g., incorrect categories, unrealistic ratings, mixed text values) were identified and corrected.
- A new feature, **Minimum Android Version Required**, was created from Android version data.
- **Current Ver** was removed as it is a technical detail with limited business relevance.
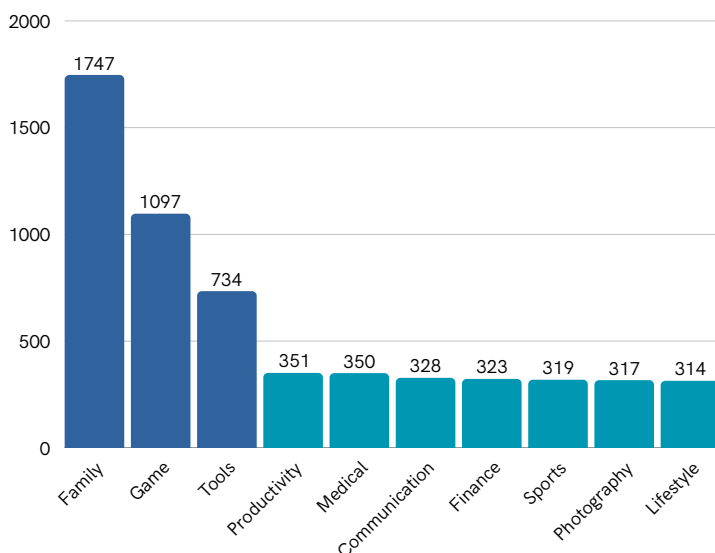
INSIGHTS

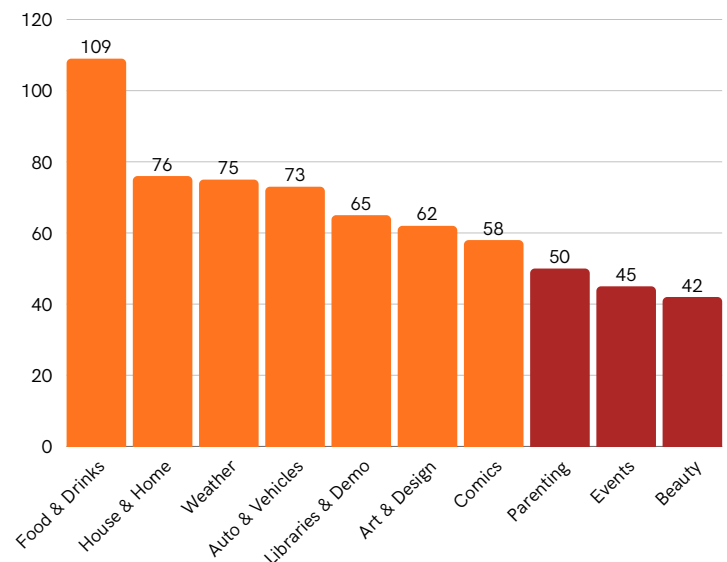# Family, Game and Entertainment categories dominates the Play Store

3,578

Total Apps in Top 3 Categories

Number of Apps per Category (Top 10)

Number of Apps per Category (Bottom 10)

Key observations

- Out of 33 categories, a small subset holds the bulk of apps, with Family, Games, and Tools topping the list.
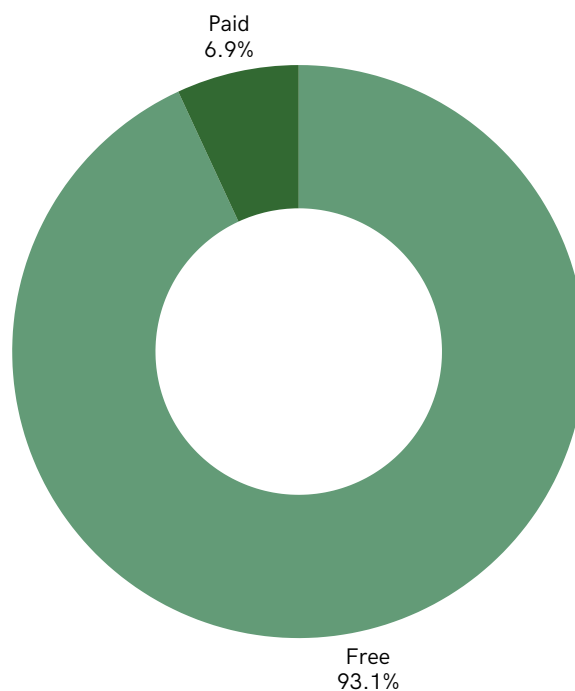
Business Insights

- Highly populated categories mean stiff competition.

- Less crowded categories, such as Auto & Vehicles, Events, and Beauty etc offer new apps a better chance to stand out.

Most apps are free to download

8,719

Apps are free to download

Paid
6.9%

Free
93.1%

Key observations

- Majority of apps on the Play Store are free (93.1%), with paid apps representing a much smaller share (6.9%).
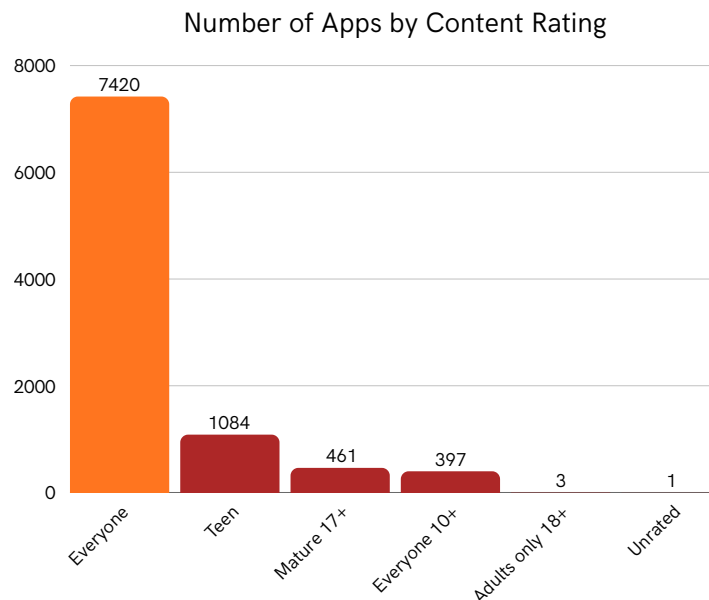
Business Insights

- The dominance of free apps suggests that ad-based or freemium models are more common than upfront paid pricing.

Apps primarily target general audiences

79.2%

Apps targeted general audience

Number of Apps by Content Rating

Key observations

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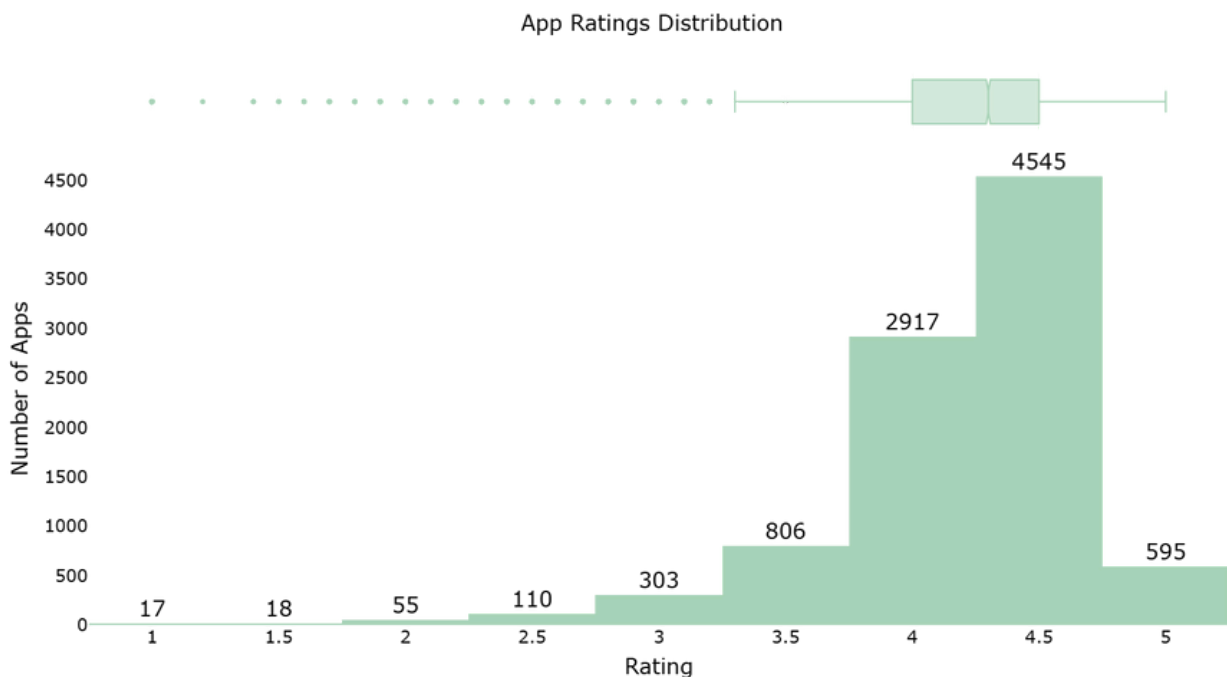
Business Insights

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

Most apps have high user ratings

4.3

Typical App Rating (Median)

App Ratings Distribution

Key observations

- The rating distribution is left-skewed, with most apps clustered between 3.5 and 5 stars. Few apps have low ratings.
- The median rating is 4.3, slightly higher than the mean of 4.1, reflecting the left skew.
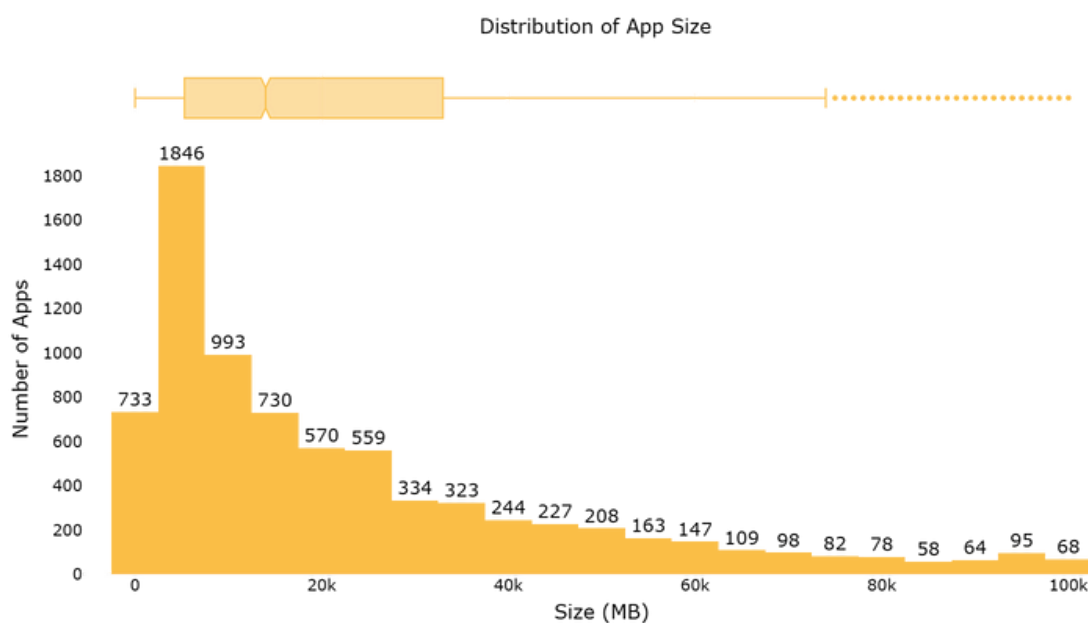
Business Insights

- High average ratings reflect strong competition and high user expectations, while lower-rated apps may struggle to attract or retain users.

Most apps fall within a moderate size range

4.3

Typical App Rating (Median)

Distribution of App Size

Key observations

- Most apps are moderately sized, with few very small or very large ones.
- The median app size is 14 MB, and the mean is 23 MB, indicating that a small number of very large apps pull the average upward.
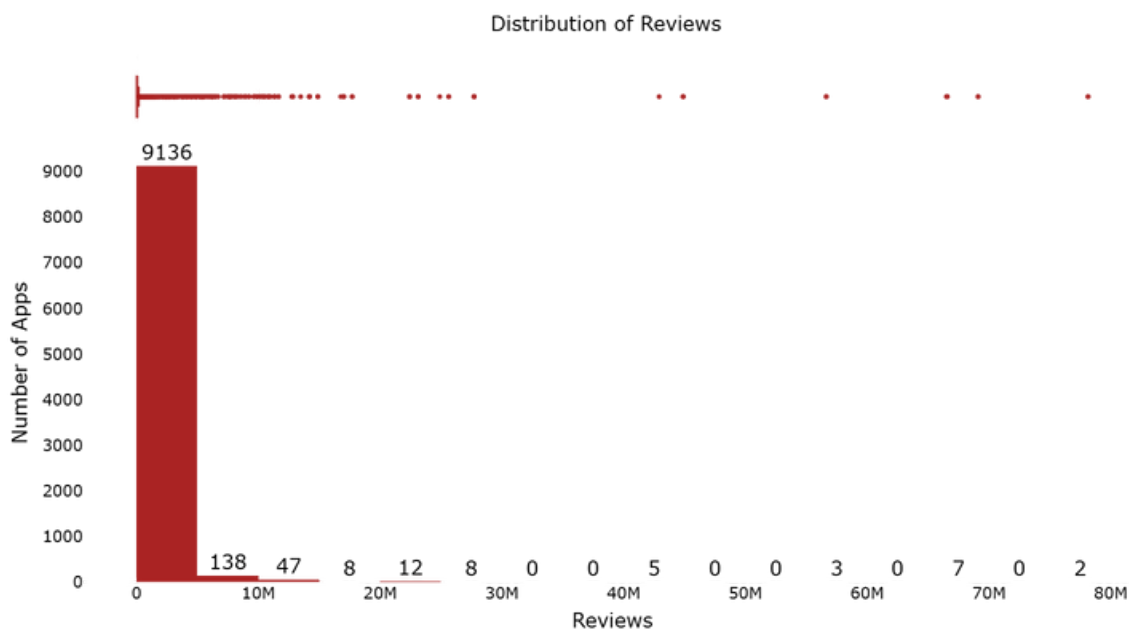
Business Insights

- Developers appear to balance functionality with storage constraints when designing apps, aiming to provide rich features without creating excessively large files.

Popularity Is Concentrated Among Few Apps

5,930

Typical Number of Reviews (Median)

Distribution of Reviews

Key observations

- Most apps receive few reviews, while a small number get extremely high counts.
- The median number of app reviews is 5,930, and the mean is 514,050, indicating that a small number of apps with extremely high review counts pull the average upward.

Business Insights

- User engagement is concentrated among a few popular apps, making visibility a key challenge for new apps.

Ratings Stabilize as Review Volume Increases

Rating vs Reviews (Normal Scale)
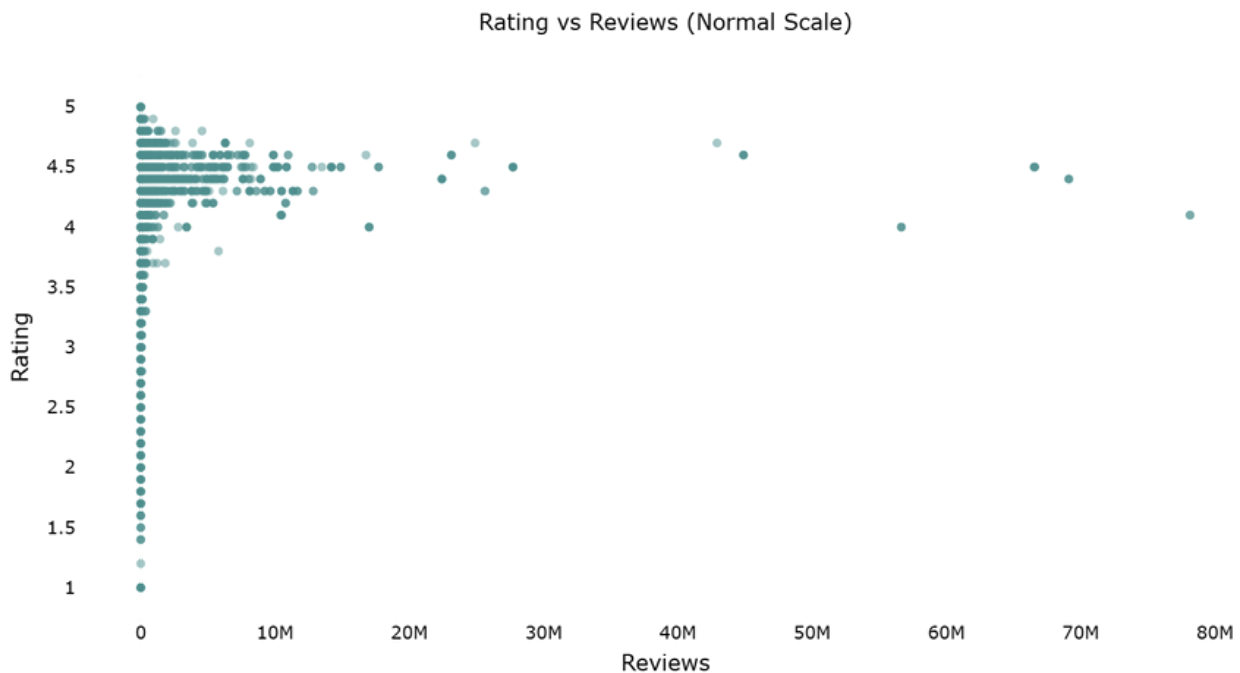
Key observations

- Most apps have few reviews with widely varying ratings, while high-review apps are rare and clustered above 4.

- Rating variability decreases as review count increases, indicating greater stability with more feedback.

Business Insights

- Ratings from low-review apps are noisy; review volume adds credibility and reliability to ratings.

- Encouraging early reviews helps stabilize perceived quality rather than directly boosting ratings.

# Upfront Paid Apps Are Uncommon; General Users Access Most Free Apps
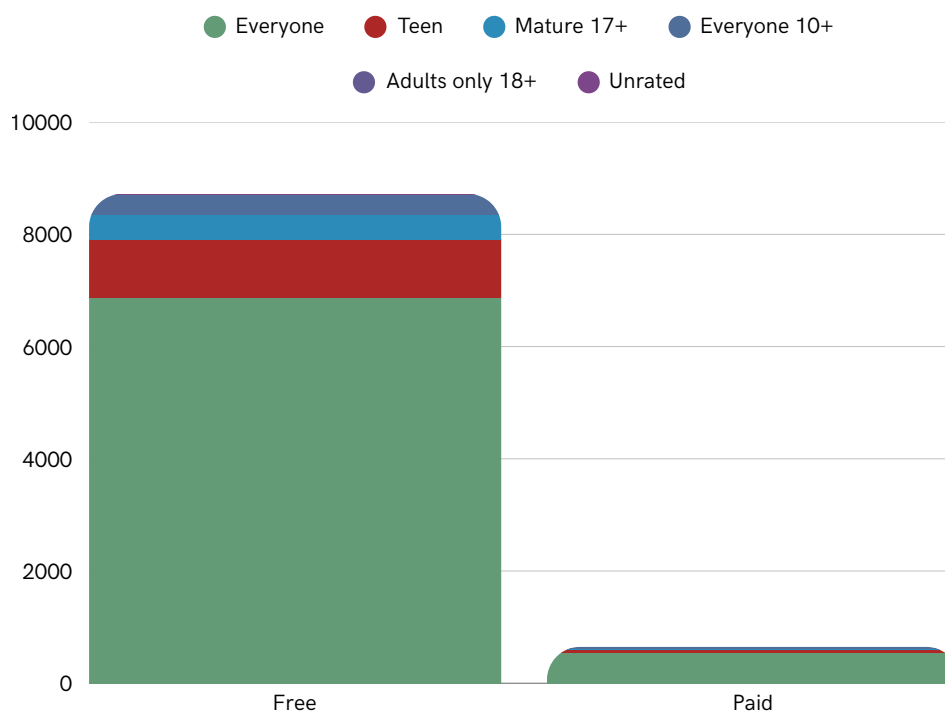
Key observations

- The Play Store ecosystem is heavily skewed toward free applications across all content ratings, with 'Everyone' dominating both free and paid types

Business Insights

- Paid apps constitute a small minority, indicating that upfront pricing is not the preferred monetization strategy regardless of content classification

Ratings Stabilize as Review Volume Increases

Average Rating Over the Years
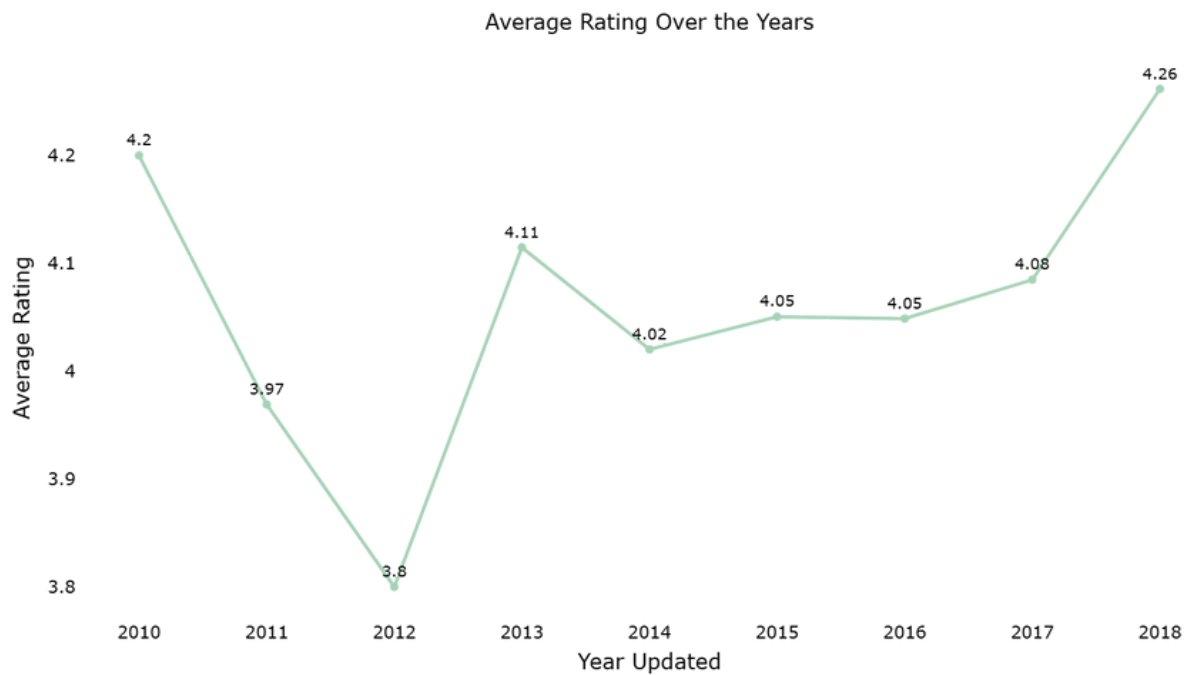
Key observations

- Average ratings stayed above 3.8 across all years, with a dip around 2011–2012 followed by recovery.
- From 2013 onwards, ratings show minimal fluctuation and a gradual upward trend, indicating stabilization.

Business Insights

- High and stable ratings suggest that the Play Store maintains consistent app quality and user satisfaction.
- The gradual upward trend signals platform maturity, indicating opportunities for user trust, app launches, and growth strategies.

Free Apps Dominate Most Categories, While Paid Apps Succeed Only in High-Trust or High-Value Segments

Rating by Category and Type (Top 15 Categories)
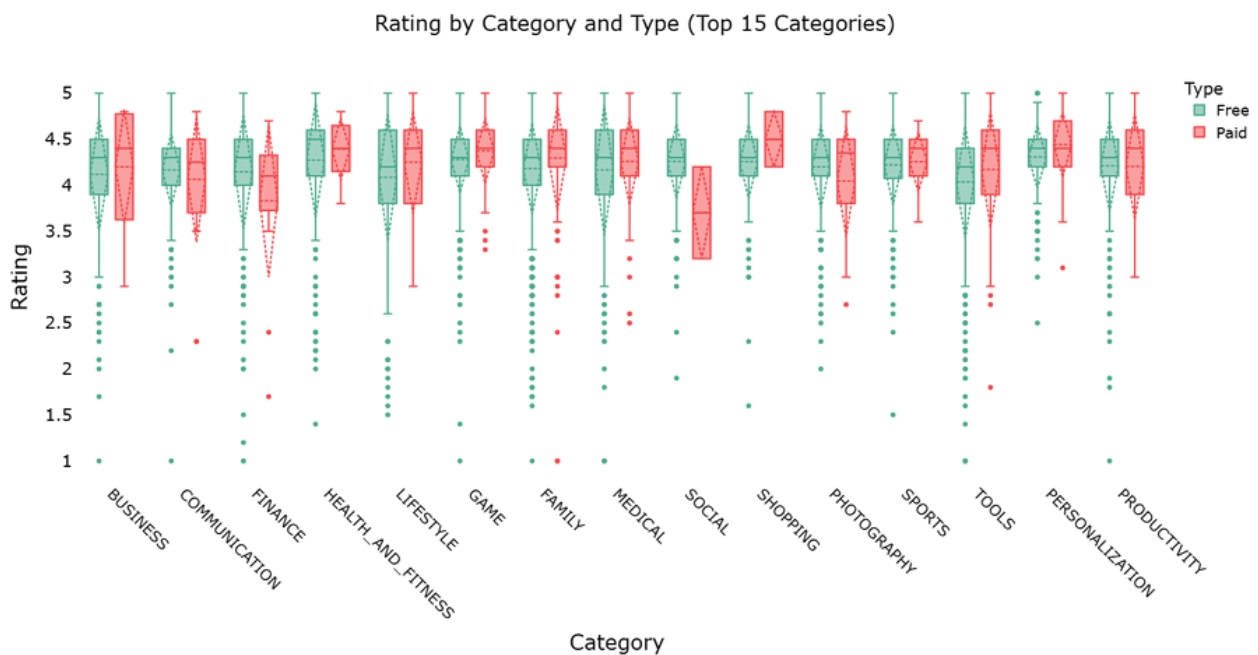
Key observations

- Free apps dominate most categories (Education, Art & Design, Events, House & Home) with high ratings and low variance, while Paid apps rarely show significant rating advantages.

- Paid apps outperform Free apps in certain sensitive or niche categories (Finance, Health & Fitness) and in some high-value categories (Food & Drink, Beauty), but Free apps remain essential for reach and virality (Games, Communication).

Business Insights

- Adopt Free-first or Freemium models in categories where Free apps are strong, leveraging optional premium features to reduce risk while maintaining user adoption.

- Use Paid or Premium-first models in high-trust or high-value categories to ensure quality perception, capitalize on user willingness to pay, and maintain credibility, while balancing risk and reach with hybrid models (Free to Paid upgrades) where appropriate.

BUSINESS / DEVELOPER TAKEAWAYS
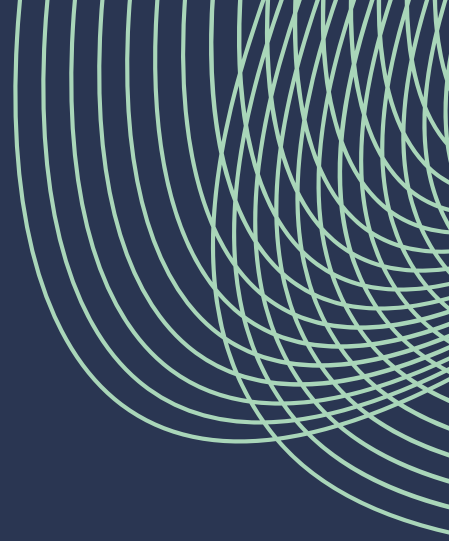
- **High-Competition Categories**: Family, Games, and Entertainment dominate. Consider exploring less crowded categories like Auto & Vehicles, Events, and Beauty to stand out.

- **Free Apps Lead the Market**: Most apps are free. Freemium or ad-based models help reduce risk while maximizing reach.

- **Free vs. Paid Strategy by Category:**
    - **Free apps dominate most categories**. A free-first or freemium approach works best.
    - **Paid apps thrive in high-trust, high-value segments**. A paid- or premium-first strategy builds credibility and attracts willing buyers.

- **Target Broad Audiences:** "Everyone"-rated apps dominate. Focusing on general users helps boost downloads and ad revenue.

BUSINESS / DEVELOPER TAKEAWAYS

- **Optimize App Size & UX:** With a median size of ~14 MB and a mean of ~23 MB, balancing rich features with storage efficiency is key to avoiding download friction.

- **Ratings Require Volume & Quality:** Apps with few reviews show unreliable ratings. Early feedback helps stabilize perceived quality, while a consistent UX builds user trust.

- **Platform Maturity & Opportunity:** Ratings have stabilized over time. Leverage platform improvements to maintain quality, while exploring less crowded categories to increase visibility.

CHALLENGES & OPPORTUNITIES

Limitations

- The dataset is outdated, which may not reflect the current app ecosystem.
- Most apps are in the free category, resulting in a bias toward free apps.
- 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.
- Conduct Time-Series Analysis using Last Updated and installs to track trends in app popularity over time.
- Compare insights with the Apple App Store to get a cross-platform perspective.
- Explore feature correlations (e.g., rating vs installs, price vs reviews) to uncover additional insights.
- Investigate impact of app size and Android version requirements on user adoption.

CONCLUSION

This analysis of Google Play Store apps provided a comprehensive understanding of market trends, user behaviour 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 popularity, 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