

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

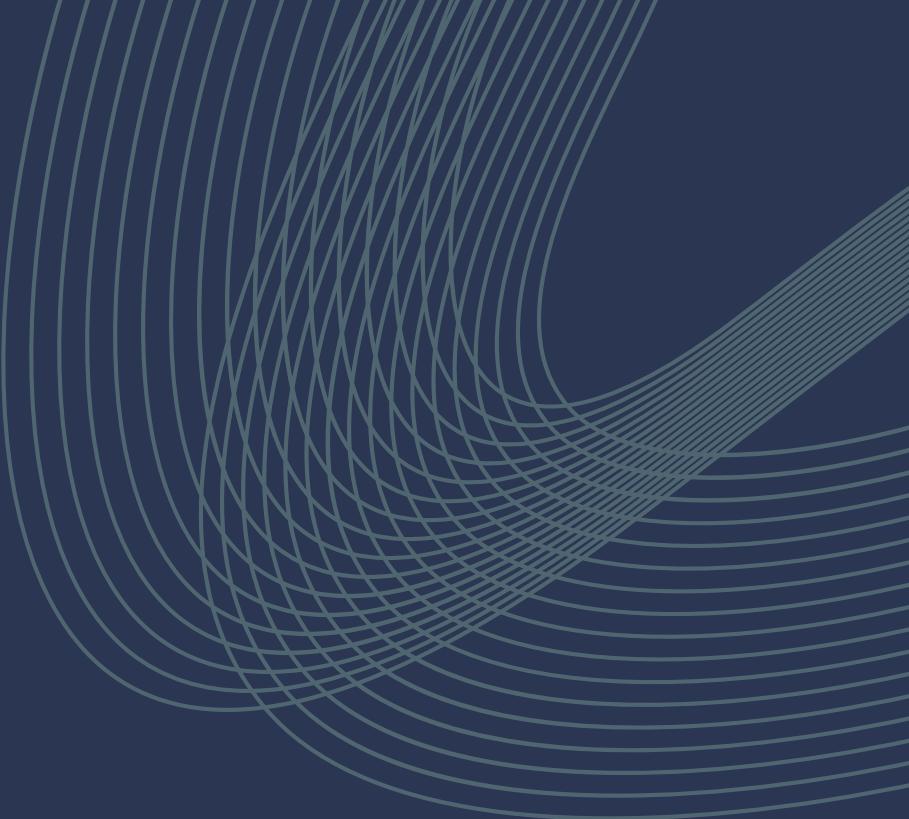


TABLE OF CONTENTS

Problem Statement & Objective	3
<hr/>	
EDA Workflow	4
<hr/>	
Key Questions Explored	6
<hr/>	
Data Overview	
<hr/>	
Insights	7
<hr/>	
Business/Developer Takeaways	8
<hr/>	
Future Work	11
<hr/>	
Conclusion	12

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.



KEY QUESTIONS EXPLORED

- Which app categories are most popular ?
- Do free apps get more installs than paid apps ?
- How does price affect ratings ?
- What factors influence high ratings ?
- Are more reviews linked to better ratings ?



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.



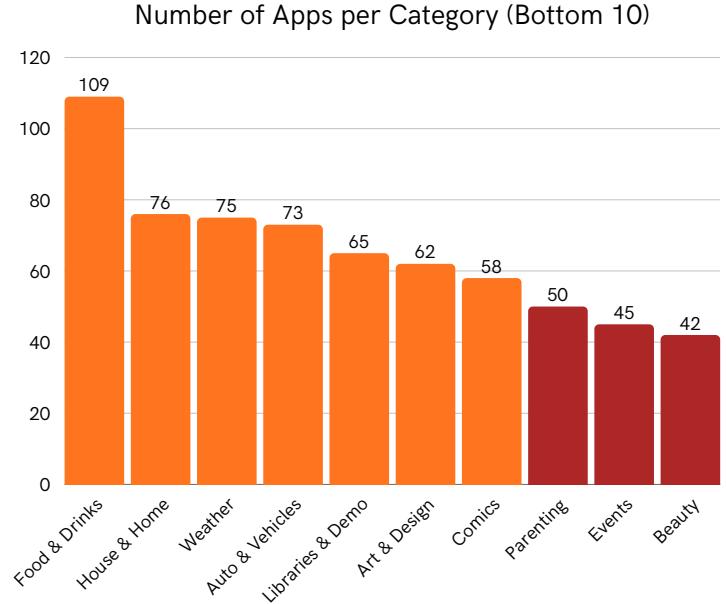
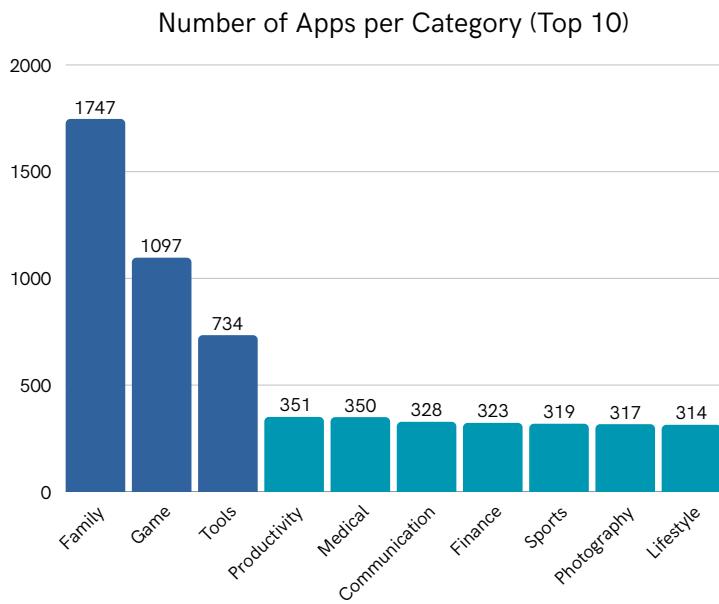
INSIGHTS



Family, Game and Entertainment categories dominates the Play Store

3,578

Total Apps in Top 3 Categories



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



Key observations

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

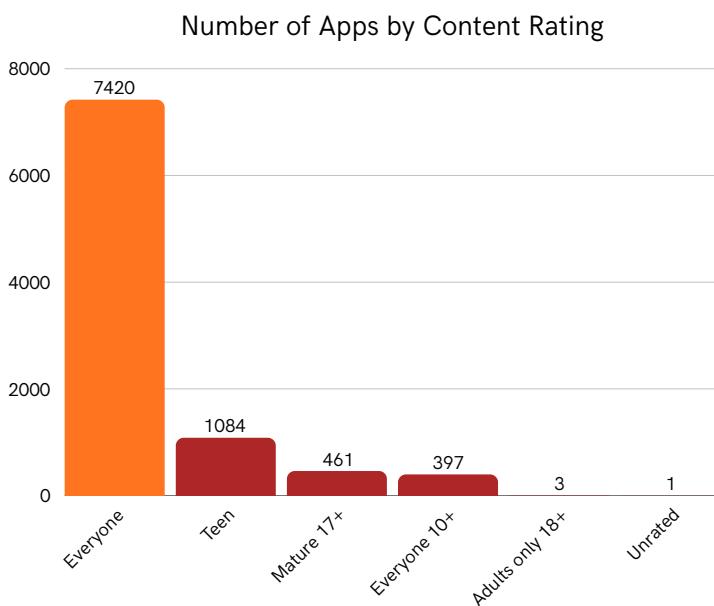
Business Insights

- 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



Key observations

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

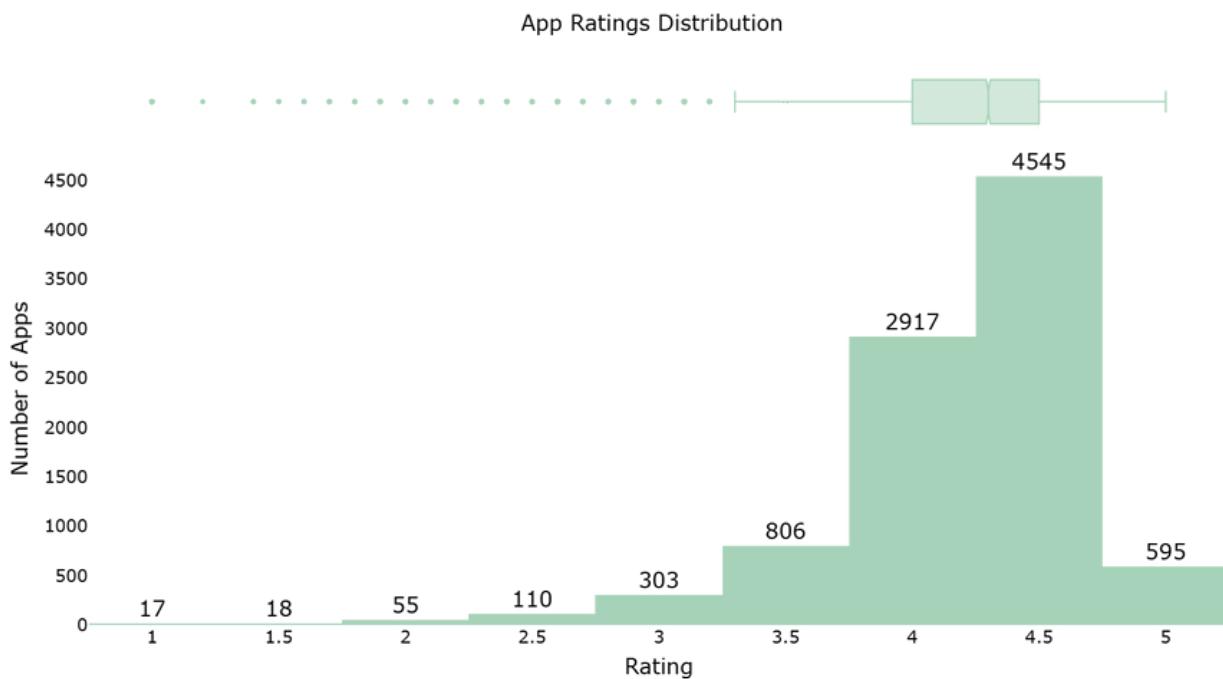
Business Insights

- 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 have high user ratings

4.3

Typical App Rating (Median)



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.

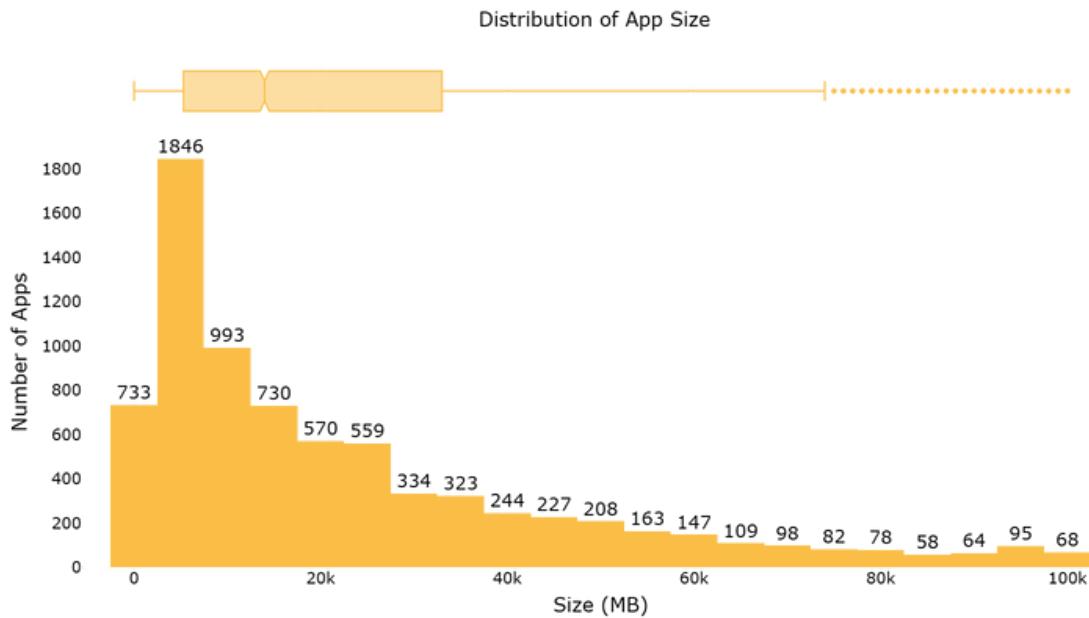
Business Insights

- 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)



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.

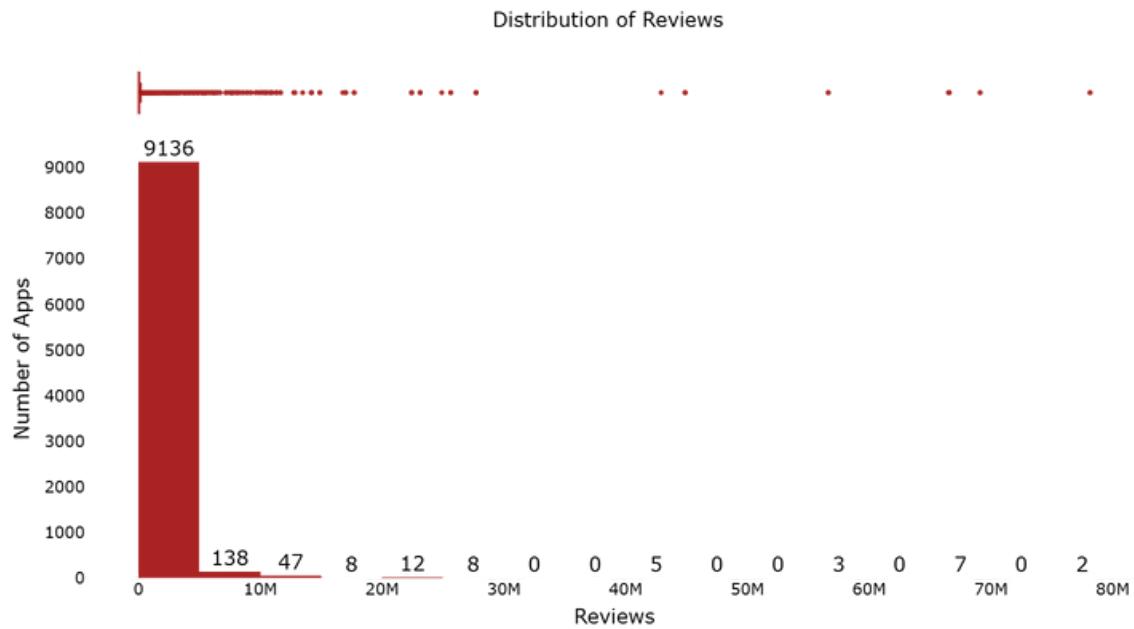
Business Insights

- 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)



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.