DATA ANALYTICS REPORT

JANUARY 2026

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

GITHUB

TABLE OF CONTENTS

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

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 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

9960
Apps covered

34
App Categories

120
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

FUTURE WORK

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.