

Executive Summary

This project delivers a comprehensive and insight-driven Exploratory Data Analysis (EDA) of the Google Play Store dataset. The goal is to understand the dynamics of the app marketplace—uncovering what drives installs, ratings, user engagement, and category performance. The analysis provides both business insights and technical depth, making it valuable for product managers, data analysts, and app developers.

Key Highlights

- **Market Dominance:** Categories like *Games*, *Tools*, and *Productivity* lead in sheer volume and install metrics.
- **Free vs Paid Apps:** Free apps overwhelmingly dominate installs, while paid apps often have slightly higher ratings.
- **Rating Drivers:** App quality, UI experience, and frequent updates strongly influence ratings.
- **User Behavior:** High installs correlate with small app sizes, brand recognition, and utility-driven features.
- **Data Cleanup Achievements:** Successfully converted complex fields like *Installs*, *Reviews*, *Size*, and *Price* into clean numerical values.

Summary of Methodology

- Performed **extensive data cleaning**, removing duplicates and handling inconsistent formatting.
- Transformed textual numeric columns using regex and mapping functions.
- Conducted **statistical analysis** to understand distributions and outliers.
- Built visually rich charts—bar charts, pair plots, heatmaps, and category-based comparisons.
- Derived **actionable insights** for app developers and marketers.

Business Implications

- Frequent updates and lightweight app design improve retention and install rates.
- Free apps dominate onboarding; paid apps must rely on premium features and niche audiences.
- High-performing categories provide an excellent landscape for new app launches.