# Overview

This report summarizes a quick analysis of the Google Play Store dataset you uploaded

(googleplaystore (1).csv). The file contains app-level metadata such as App name, Category, Rating, Reviews, Size, Installs, Type (Free/Paid), Price and other fields. I cleaned numeric fields (Installs, Size, Rating, Reviews, Price) where possible and produced visualizations that reveal how installs, ratings and sizes vary across apps and categories.

# Key takeaway (summary first)

 The Play Store in this dataset is dominated by free apps; a small number of apps account for most installs.

 Ratings cluster at the high end, but rating distributions differ notably by category.

 Total installs vary widely between categories; a few categories capture the bulk of installs.

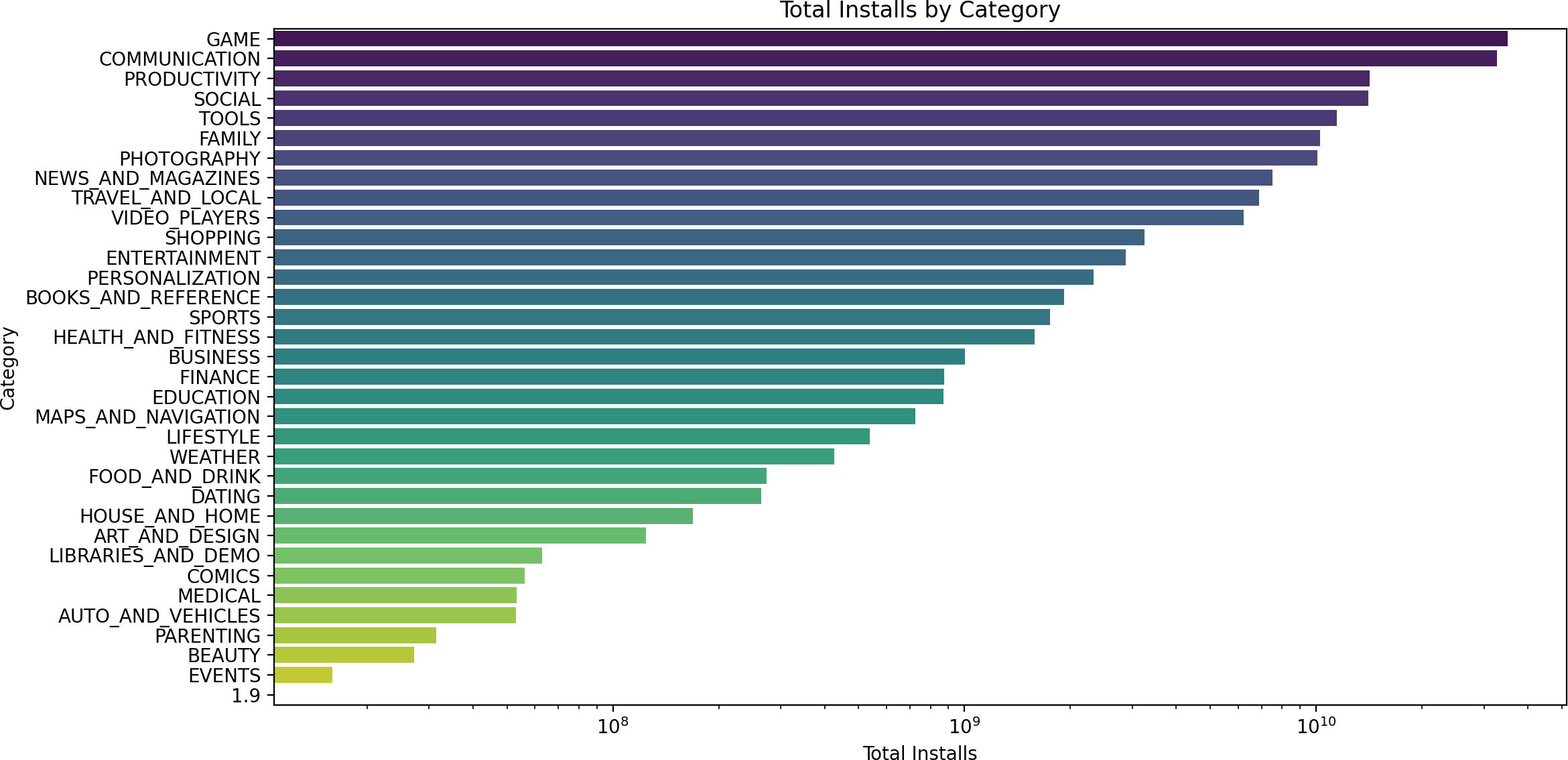
 App size has only a weak and noisy relationship with installs — many small apps still get very large install counts. Below I alternate insights with the charts and tables that

support them.

# Top categories drive most installs

 Insight: A small set of categories account for the majority of installs; totals span several orders of magnitude.

 Supporting visual: Total installs by category (log scale)

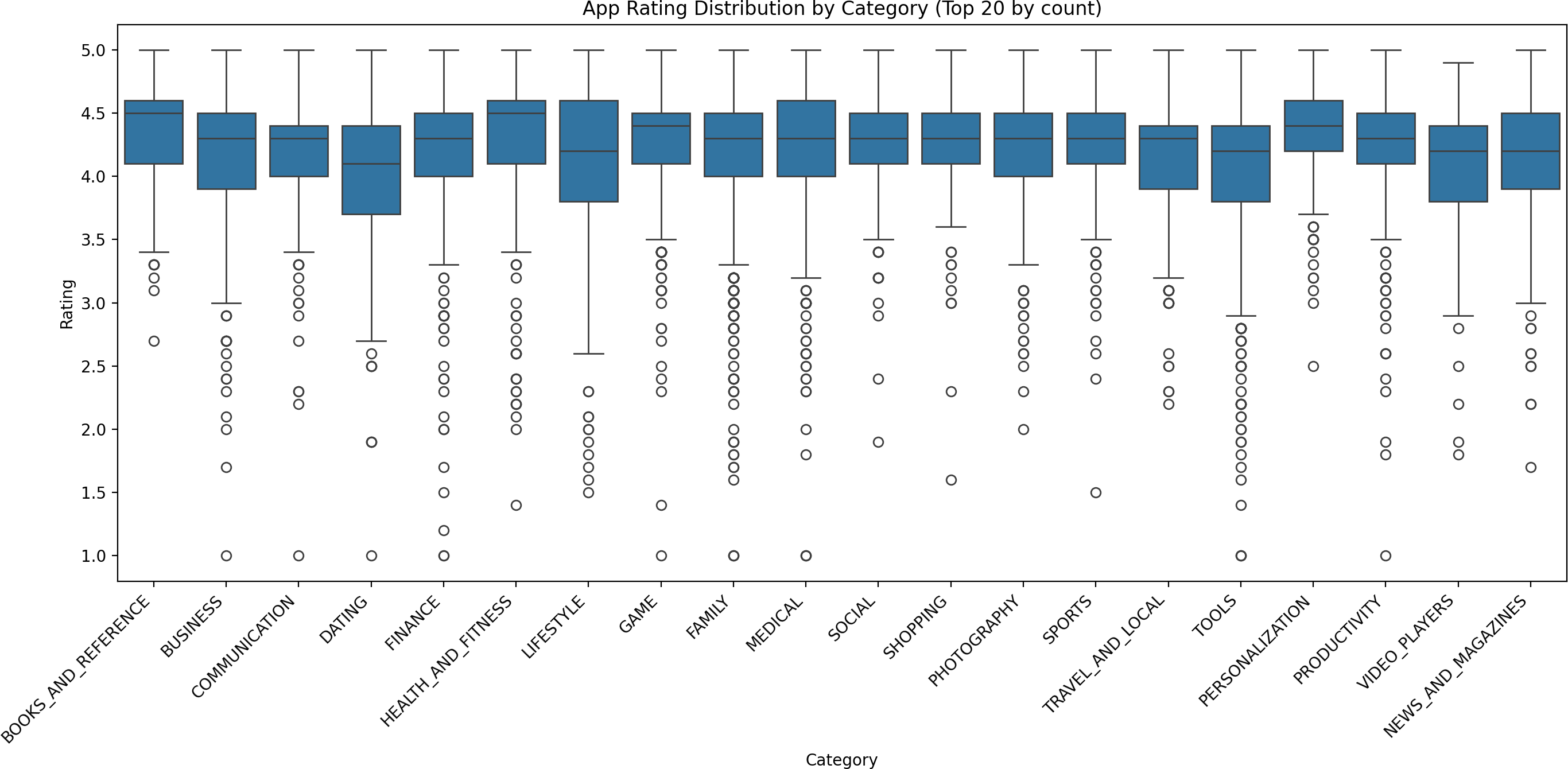


 What to notice: The use of a log scale highlights that while many categories have measurable installs, the largest categories (top bars) dominate the total volume.

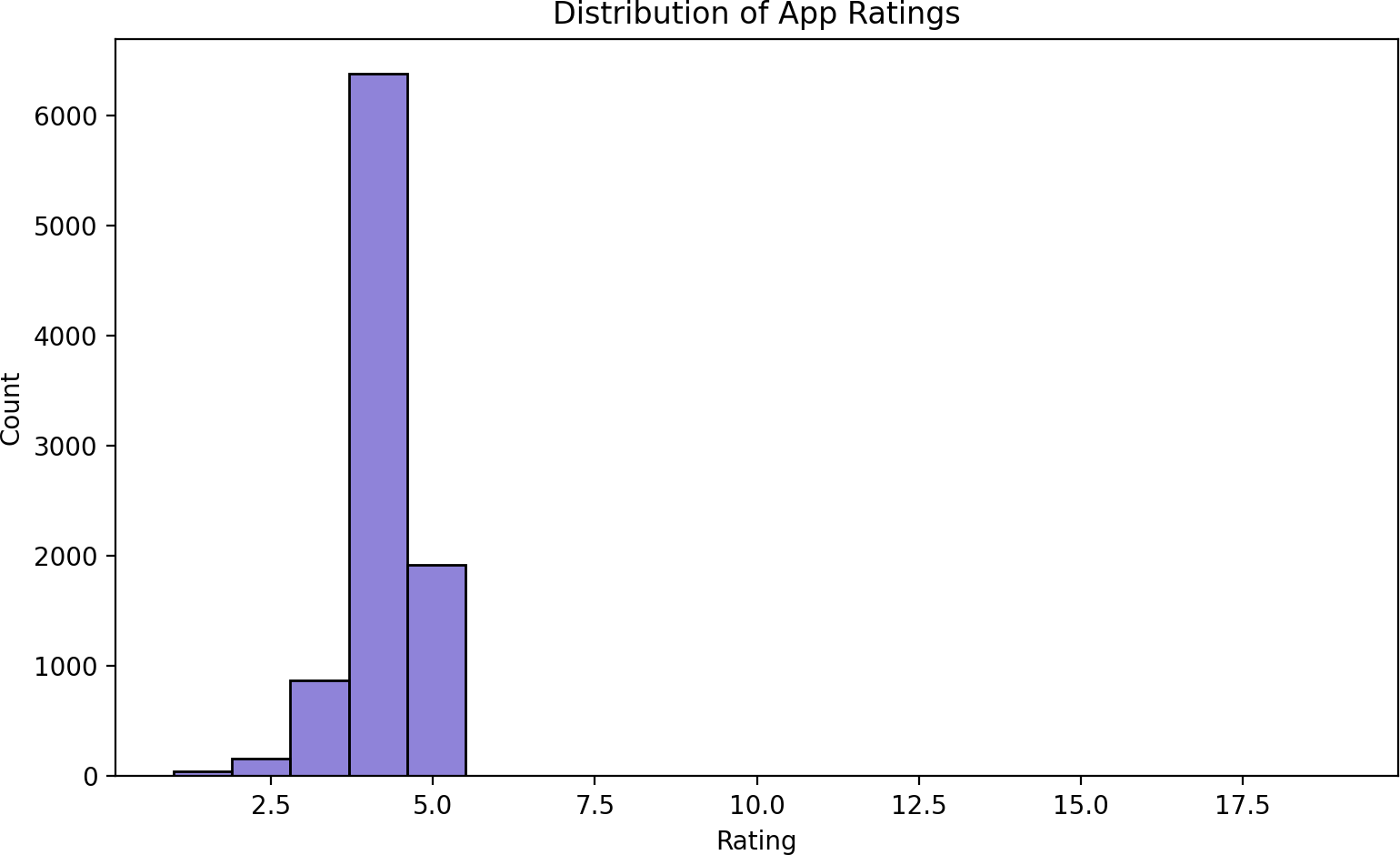
# Ratings are skewed high overall, but vary by category

 Insight: Overall app ratings cluster toward higher values, but the spread and outliers differ across categories — some categories show wider rating variability.

 Supporting visual: Rating distribution by category (box plots for top 20 categories by rated-app count)



Supporting visual: Overall rating histogram



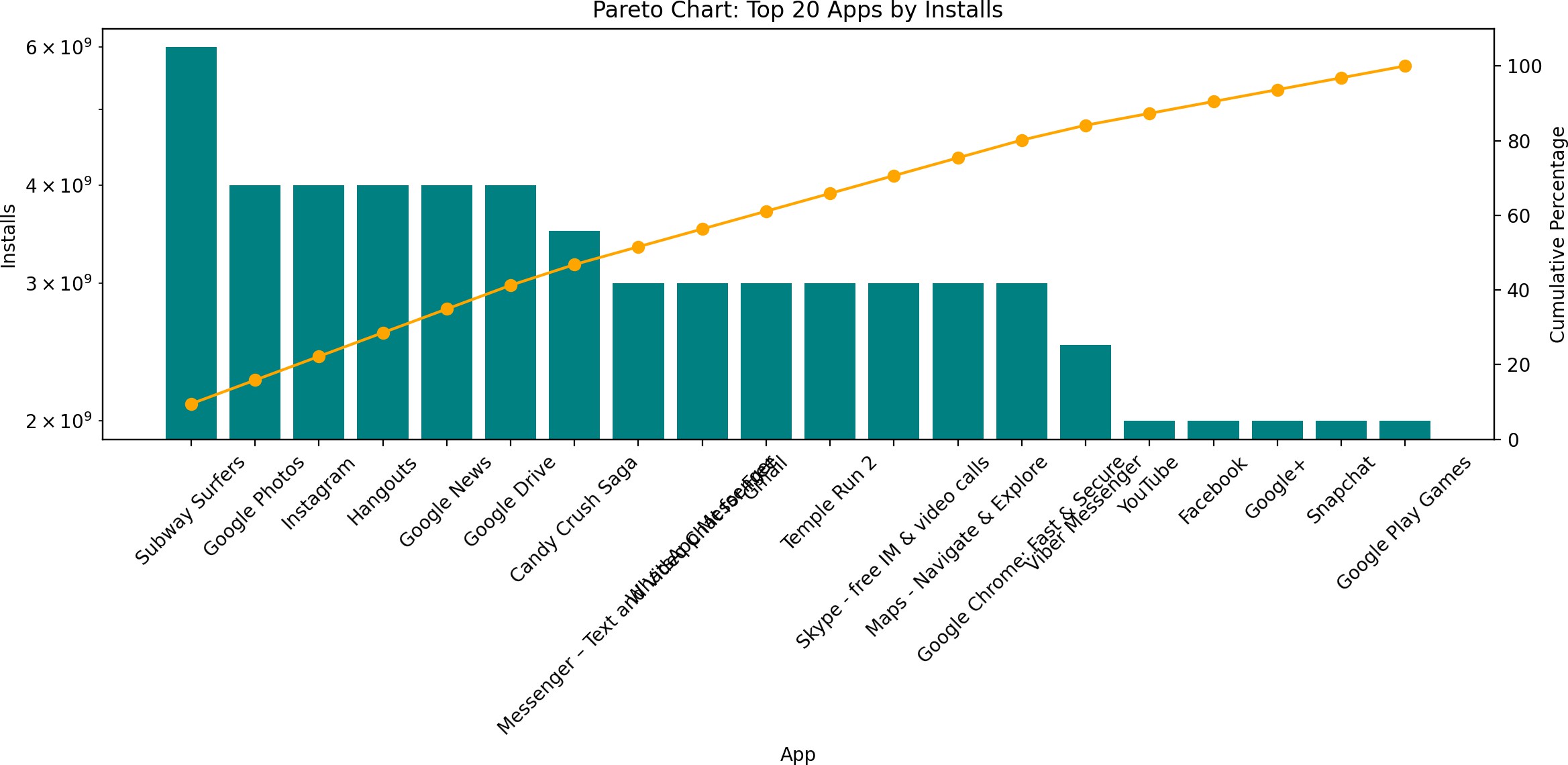
 What to notice: The histogram shows most ratings are clustered around the upper

range. The per-category boxplots indicate which categories have more consistent ratings (narrow boxes) versus those with wider variation or more outliers.

# A few apps contribute the lion’s share of installs (Pareto)

 Insight: The top 20 apps capture a large cumulative share of installs; the Pareto curve shows how quickly cumulative installs rise among top apps.

 Supporting visual: Pareto chart (top 20 apps by installs, with cumulative percentage)

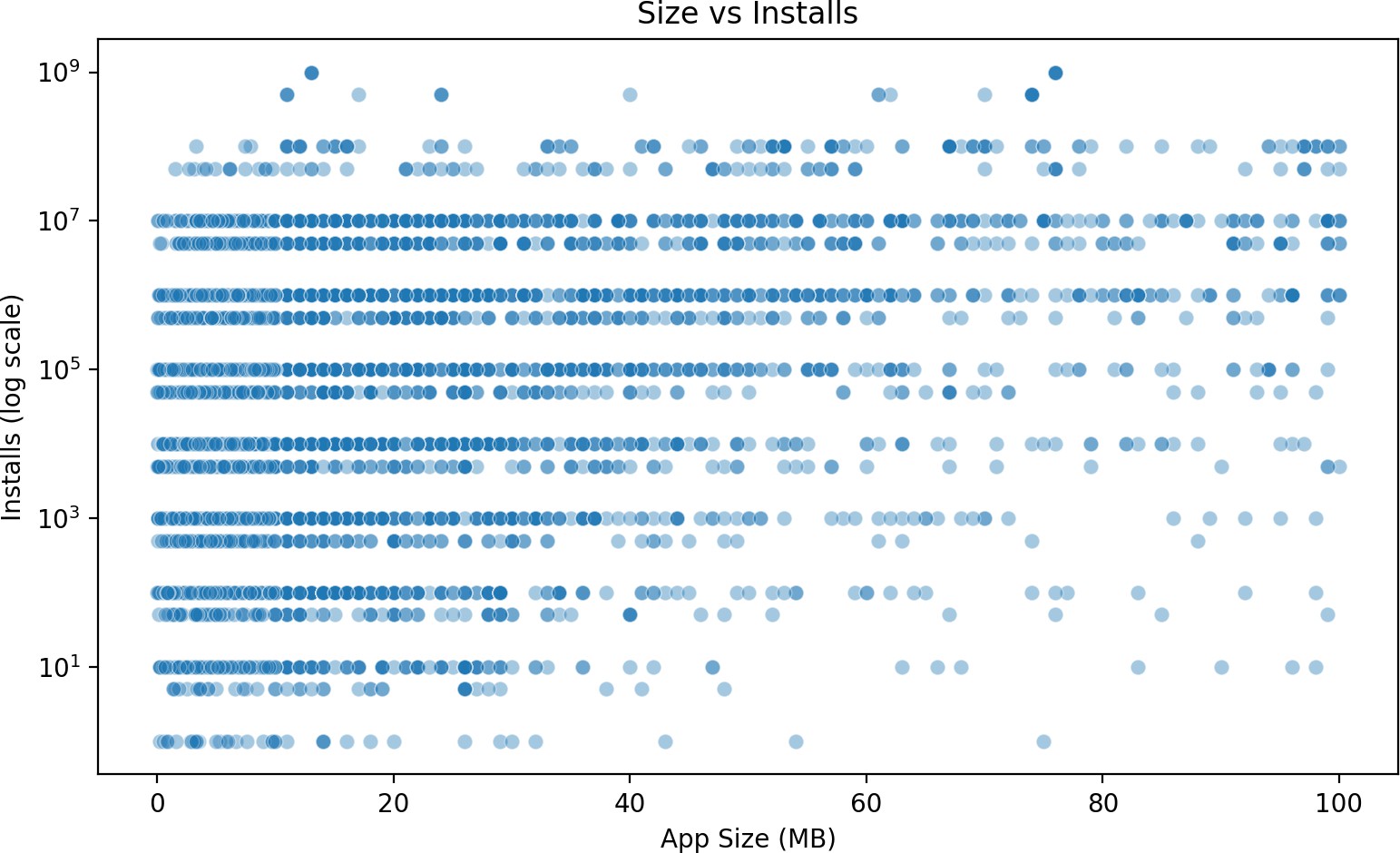


 What to notice: The bar heights (installs, log scale) and the cumulative line demonstrate that a handful of apps are responsible for a large fraction of total installs.

# App size and installs: weak/no clear linear relationship

 Insight: App size (MB) does not strongly predict install numbers. Many small apps have very large installs, and large apps can have either low or high installs.

 Supporting visual: Size vs Installs scatter plot (installs on log scale; sampled up to 5,000 points to reduce overplotting)



 What to notice: The cloud of points across sizes with installs spread on a log scale

indicates no simple monotonic relationship; product fit, marketing and category likely matter more.

# Example sample of raw rows (first 5 records)

 Insight: The dataset includes app names, categories, ratings, review counts, size strings, and install ranges. I converted Size to MB and Installs to numeric counts where possible.

 Supporting data (first five rows converted to table):

|  |  |  |
| --- | --- | --- |
| App | Category | Rating |
| Photo Editor & Candy Camera & Grid & ScrapBook | ART\_AND\_DESIGN | 4.1 |
| Coloring book moana | ART\_AND\_DESIGN | 3.9 |
| U Launcher Lite – FREE Live Cool Themes, Hide Apps | ART\_AND\_DESIGN | 4.7 |
| Sketch - Draw & Paint | ART\_AND\_DESIGN | 4.5 |
| Pixel Draw - Number Art Coloring Book | ART\_AND\_DESIGN | 4.3 |

# Final summary

 The dataset is heavily skewed to free apps; only a small fraction are paid.

 Installs and ratings are concentrated: a few categories and a few apps explain most installs, while ratings are generally high but variable by category.

 App size alone is not a reliable indicator of popularity (installs). These findings point

toward prioritizing category- and app-level analysis for any growth or product strategy: targeting top categories and benchmarking against the top apps will reveal the most

opportunity.