Android App Metrics Analysis (2010-2018 Data)

by Dr. Mouhamad Ibrahim & Eng. Rashad Al Aych

Introduction

Imagine a world where every download is a step towards success, and every app developer has the key to understanding user preferences. This is the domain we explore, where the Google Play Store becomes the canvas for our analysis, revealing patterns and understanding the factors that send an app to the peak of popularity.

This project utilizes a Kaggle dataset of over ten thousand Google Play apps. The primary objective is to uncover correlations between app downloads and diverse parameters within this dynamic and competitive ecosystem.

This analysis was preceded by a comprehensive data cleaning process that resulted in a 75.65% improvement in data quality. This foundational enhancement enabled accurate trend identification and reliable insights into the factors driving app success.

In the Android-dominated era from 2010 to 2018 (Figure 1), the Google Play Store was the epicenter of the mobile app universe. By 2018, it firmly secured a commanding 74% share of the global smartphone market, presenting developers with the challenge of understanding user preferences in a highly competitive market. Our exploration goes beyond statistics; it's a quest to decode the language of downloads. We aim to unravel the reasons behind app success or failure, providing insights to guide developers in strategies for increased downloads.

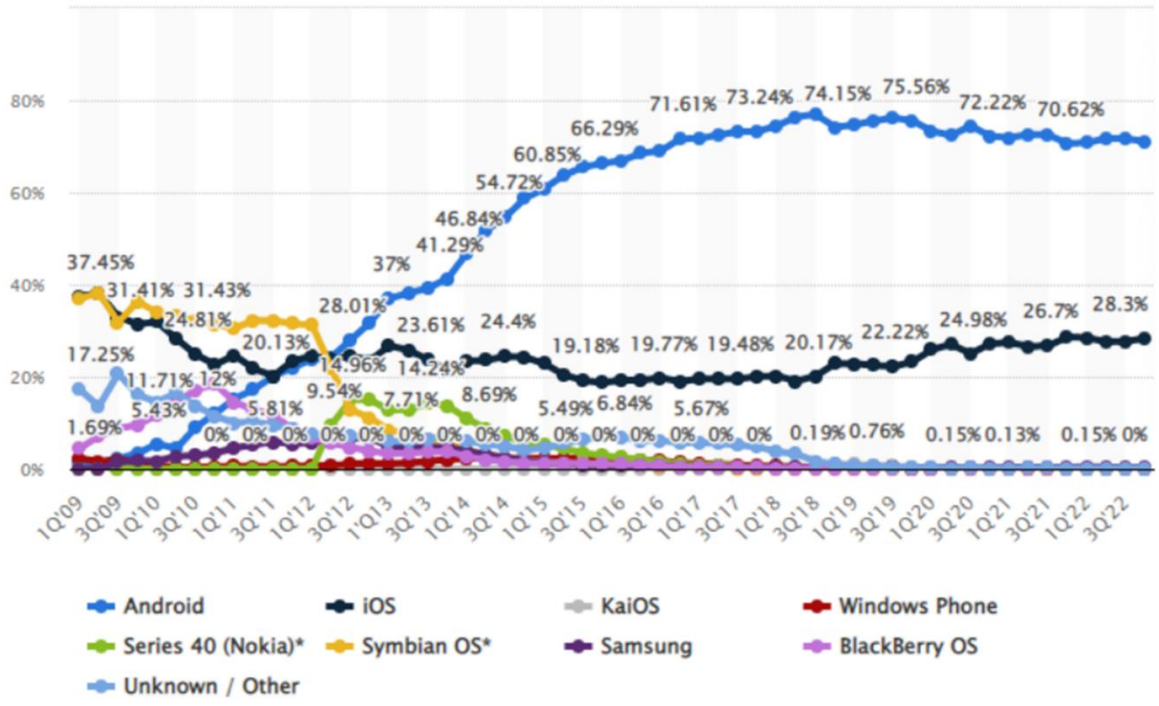

Figure 1: Growth rate of android applications compared with other applications from 2009 to 2022[1]

Problem Statement

While previous research primarily delved into rating, our approach takes a distinct path. We refrained from rating analysis due to the consistently high average ratings. Instead, our focus was on understanding how downloads can be increased by studying correlations with parameters such as genre, target audience, and reviews. A novel aspect of our study is the

[1]Growth rate of android applications compared with other applications from 2009 to 2022.

SOURCE: Koala et al., Software Vulnerabilities' Detection by Analyzing Application Execution Traces, IJACSA, Vol. 14, No. 6, 2023

emphasis on genre rather than the traditional category variable, recognizing its greater accuracy. Furthermore, we diverged from revenue analysis as the majority of apps are free.

Notably, none of the previous research conducted a comprehensive investigation into strategies for increasing app downloads. This became the focal point of our study, seeking to fill a gap in existing literature and provide actionable insights for developers navigating the competitive landscape of the Google Play Store. In response to the question of what constitutes successful downloads, we introduced a Deviation Metric. This metric quantifies the difference between an app's average downloads within specific variables, such as genre and target audience, and the overall average downloads within those variables. Our aim is to identify which variables have the most significant impact on the Deviation Metric.

Methodology

Dataset Introduction

We work with dataset scraped from the Google Play Store during the period 2010-2018 *(Table 1)*.

Key Variables

Variable	Description
App	Application name
Category	Category of the app. Family, game...etc.
Rating	How users rate the app out of 5, with 1 being the lowest rating
Reviews	Number of users reviews each app has received
Size	Memory size needed to install the application
Installs	Number of times each app has been downloaded by users
Type	Whether the app is free or a paid app
Price	Price of the app
Content Rating	Specifying the intended audience for the app. Can be for teens, mature audience, or everyone
Genres	Sub-category, e.g., Educational under Family category, Arcade under Game category
Last Updated	Release date of the most recent update for the app

Table 1: Unlocking App Insights: Kaggle Dataset Overview

Experimental setup

The experimental setup involves two parts, each contributing to a comprehensive analysis of the Google Play Store app data.

First part: Data Preprocessing using R

The initial data preprocessing step involves using R, to handle missing values, clean outliers, and ensure data consistency. Before delving into the analysis, we start by loading the dataset into data frame.

Addressing Duplicate Entries & Duplicate Apps Name

To ensure data integrity, we remove duplicate entries. Organizing the data frame by reviews prioritizes entries by popularity, facilitating the removal of duplicate app names. Subsequently, we systematically eliminate duplicate app entries, retaining only the one with the highest number of reviews.

Adjusting Category Value for Outliers

An examination of the category column reveals an outlier, specifically the entry "Life Made WI-Fi Touchscreen Photo Frame." The category value is adjusted to align with Google Play Store information, rectifying any discrepancies in the dataset.

Handling Missing Data

Our approach to missing values involves a thorough examination, ensuring data completeness. Missing values in the 'Rating' column are handled strategically by replacing with mean ratings grouped by genre.

Data Type Standardization

Columns underwent data type standardization for consistency and meaningful analyses. "Reviews" and "Installs" were converted to numeric, "Type" to a factor for clarity, "Price" to numeric, "Content Rating" to a factor, "Size" to numeric with missing values addressed by replacing with mean size grouped by genre.

Export

We exported the cleaned data frame CSV file for further analysis and operations.

Second part: Visualizing Data and their Implications using Power BI

The second part of experimental setup involves leveraging Power BI for visualizing the processed data and deriving actionable insights. This step is crucial for understanding the implications of the data and communicating findings effectively.

Data Import

In this phase, we imported the cleaned dataset into Power BI. Ensuring data integrity, we validated and confirmed proper data types during the import process, laying the foundation for accurate analyses.

Creating New Variables

New variables like the Deviation Metric were introduced.

Deviation Metric = variable avg downloads – overall avg downloads

This metric quantifies the difference between an app's average downloads within specific variables, such as genre and target audience, and the overall average downloads within those variables. Our aim is to identify which variables have the most significant impact on the Deviation Metric. This enhances visualization, providing a more insightful perspective.

Visual Exploration

Utilizing Power BI's powerful visualization tools, we explored data patterns, and trends through diverse charts, graphs, and dashboards.

Results and Findings

Exploring the App Ecosystem

In our analysis, we identified a diverse landscape with 33 categories spread across 48 genres, encompassing a total of 9,639 unique apps in the dataset.

The average rating across all apps is approximately 4.2 (Figure 2), showcasing a consistently high level, and considering the maximum rating is 5, we refrained from further investigation into the rating aspect.

A notable observation is that a significant majority, around 92%, of the apps are offered for free, while only 8% are paid (Figure 2). As a result, we opted not to delve into price analysis, given the dominance of free apps.

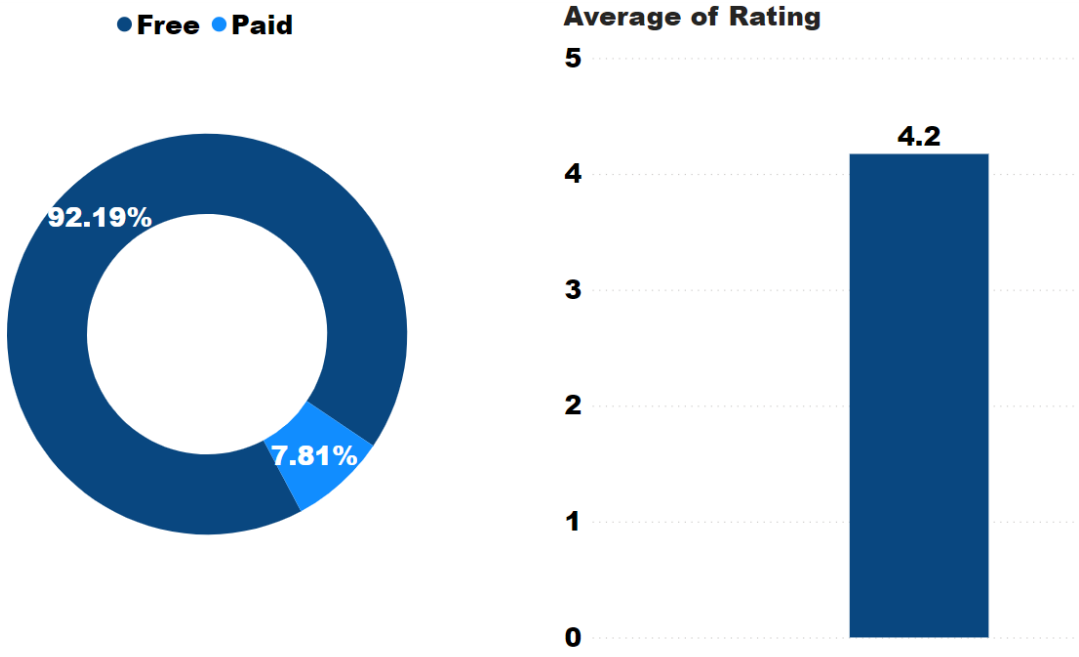

Figure 2: Rating and Pricing Trends in Mobile Apps

Download Distribution, %

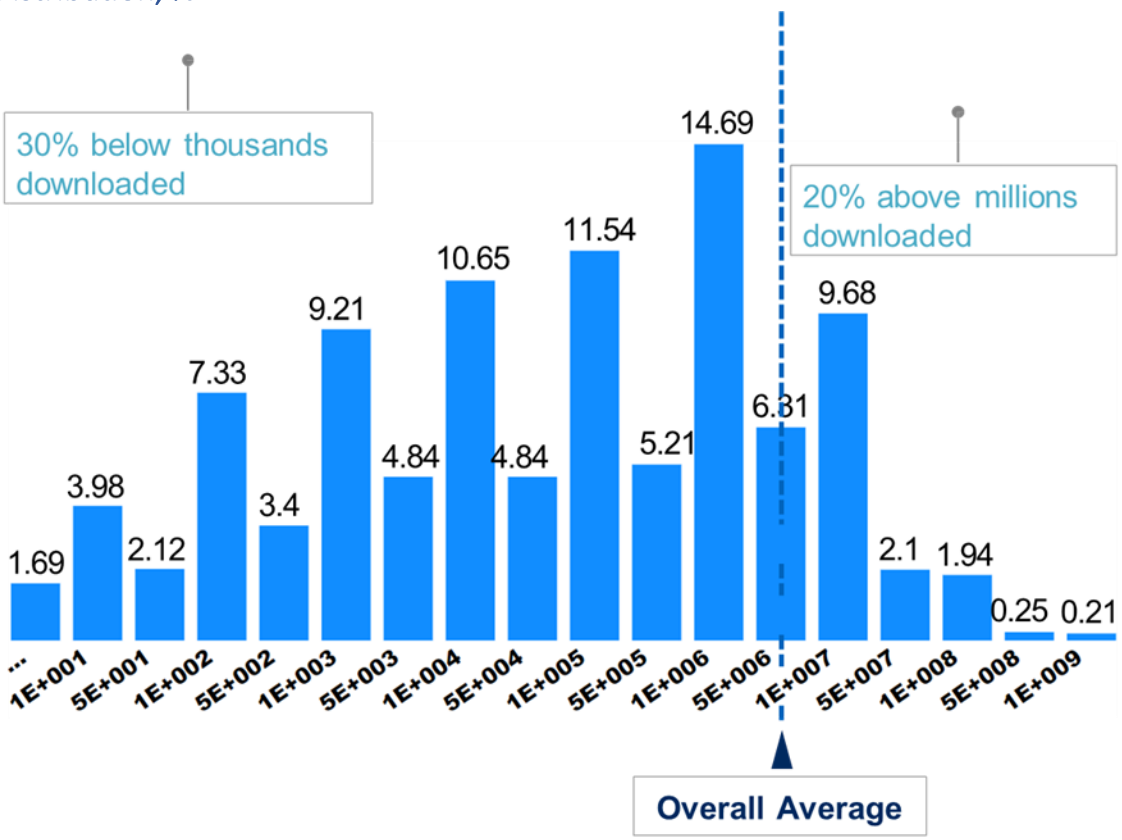

Figure 3: App Download Diversity: A Comprehensive Overview

Examining the data reveals a diverse landscape of app downloads (Figure 3). The majority of apps, approximately 80%, garner fewer than a million downloads, while a noteworthy 20% exceed this threshold. Conversely, 30% of apps face challenges reaching a thousand downloads. The average number of downloads per app stands at approximately 7.78 million.

Deviation Metric by Genre, 48 genres

Our examination of genre deviations sheds light on the standout performers in the app landscape (Figure 4). Notably, communication, video players, social, and arcade games emerge as the genre giants, each making a significant impact. The communication genre, in particular, exhibits a remarkable 450% deviation from the average, underscoring the paramount importance of connectivity in users' preferences. Following closely is the arcade games genre, boasting a 250% deviation, indicating its efficacy in providing stress-relief experiences, as suggested by Dr. Warburton, a gaming expert from Macquarie University[1]. However, not all genres experience the same level of success; educational and medical apps encounter a decline, signaling the necessity for innovative strategies to capture and maintain user interest.

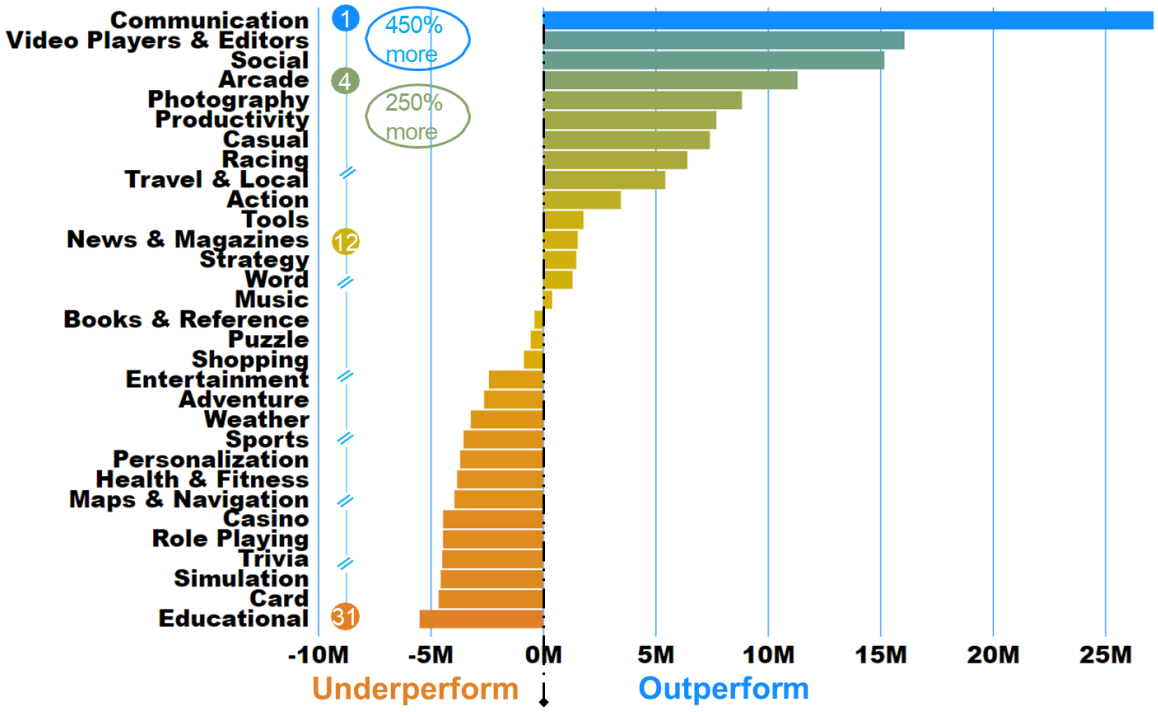

Figure 4: Genre Deviations: Unveiling Dominant Trends in App Landscape

Deviation Metric by Audience target

The data highlights a surprising trend in Android app downloads, with a distinct focus on the younger demographic (Figure 5). Specifically, pre-teens and teens emerge as the dominant audience, attracting a remarkable 200% more downloads compared to the average across all audience segments. This unforeseen revelation challenges preconceptions, as adults appear to be less dominant in this dynamic landscape.

[1]Dr. Warburton, Macquarie Univ., gaming expert

SOURCE: https://highlandsec.com.au/game-changer-the-6-surprising-benefits-of-amusement-arcade-games

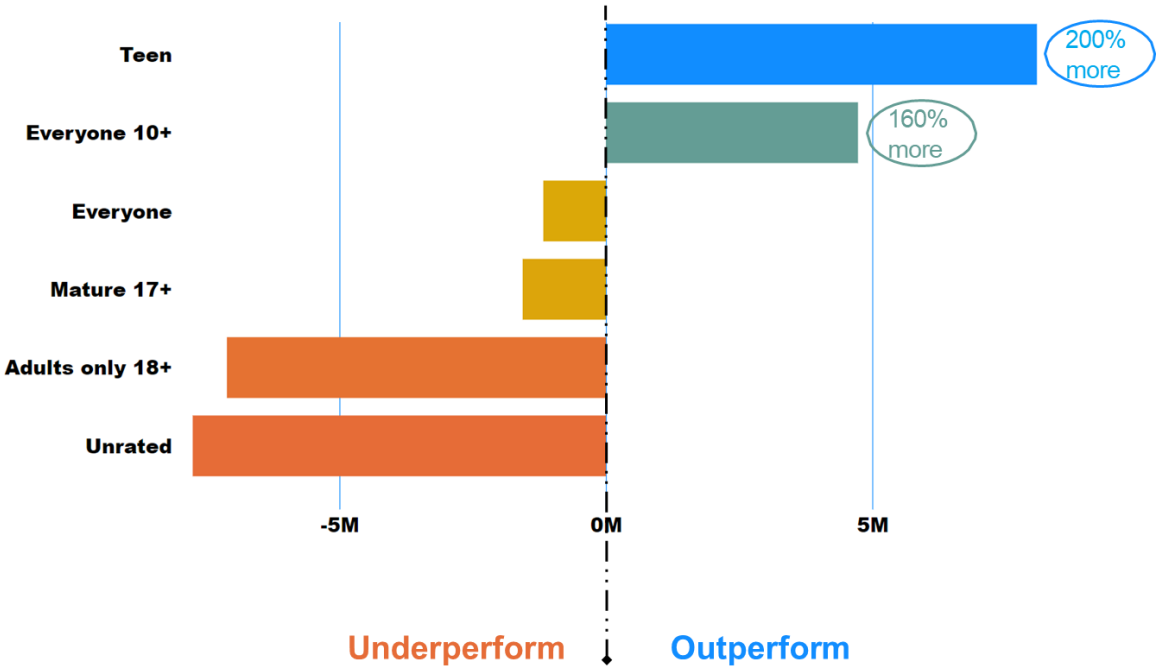

Figure 5: Youthful Surge: Pre-Teens and Teens Lead Android Downloads

Deviation Metric by Size

Our analysis unveils an unexpected trend in user behavior (Figure 6). Users willingly download larger apps that deliver substantial value. The 70MB arcade game, surpassing the average size by 300%, stands out as a chart-topping phenomenon. This insight suggests that file size alone may not deter users when compelling and engaging experiences are at stake.

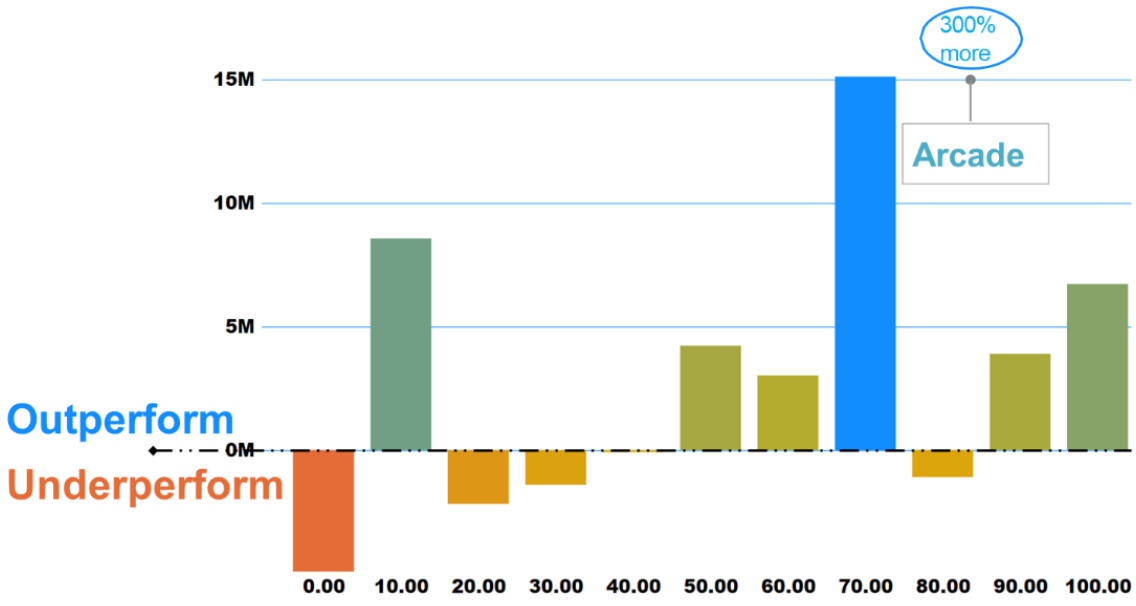

Figure 6: Decoding Downloads: The 70MB Arcade Game Trend

Deviation Metric by Release month

Our data unveils a notable spike during the summer months (Figure 7), with a 310% surge in downloads. This prompts us to delve into the factors that elevate July and August to prime time for app exploration. As schools take a break, app stores experience a significant uptick in user engagement and downloads.

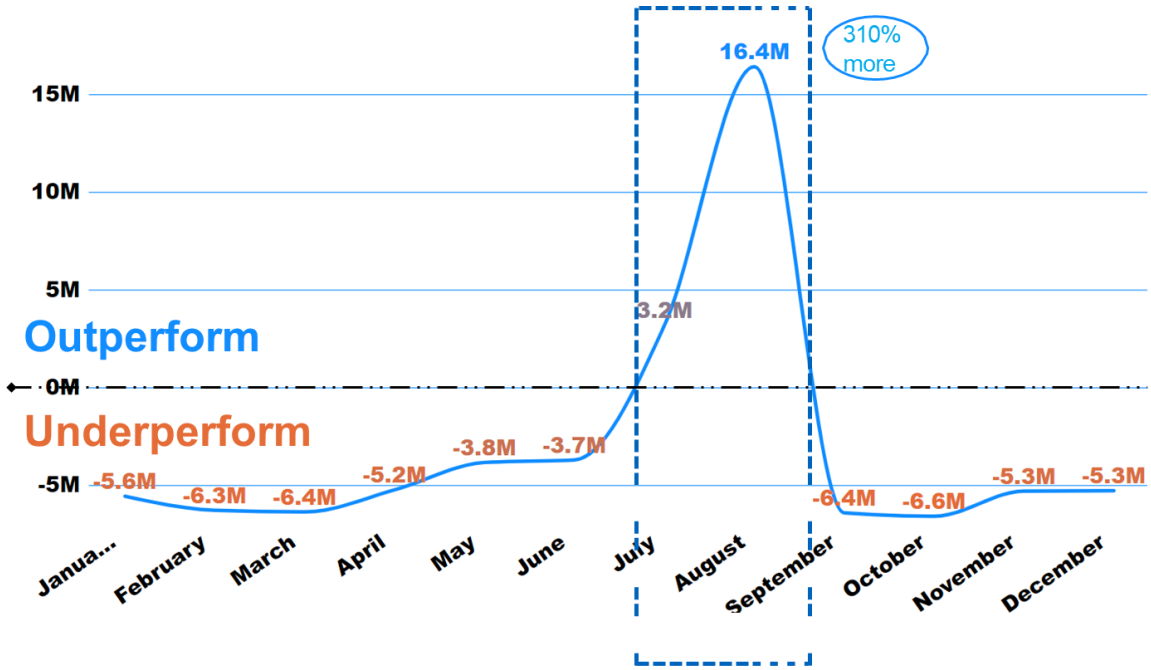

Figure 7: Summer Surge: Exploring the 310% Increase in App Downloads

Downloads vs # Reviews, log scale

Our analysis underscores the pivotal role of reviews in enhancing app visibility (Figure 8), revealing a strong correlation coefficient of 0.63 between downloads and reviews (Figure 9). This highlights the substantial impact of reviews on influencing download numbers.

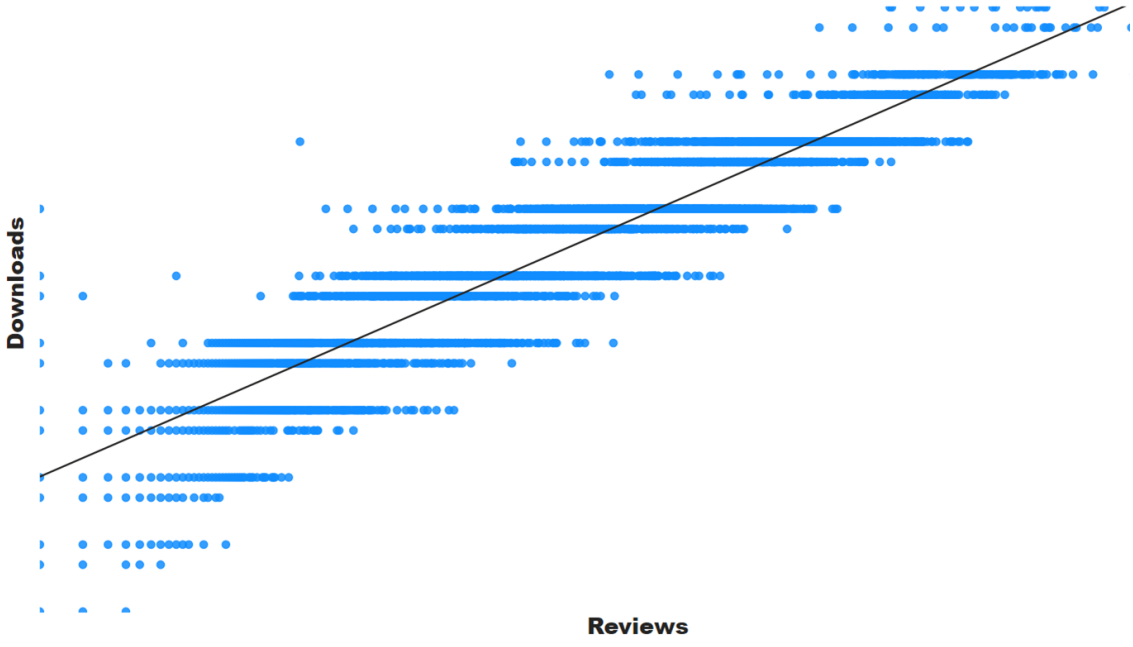

Figure 8: Reviews' Impact: Correlation with Downloads

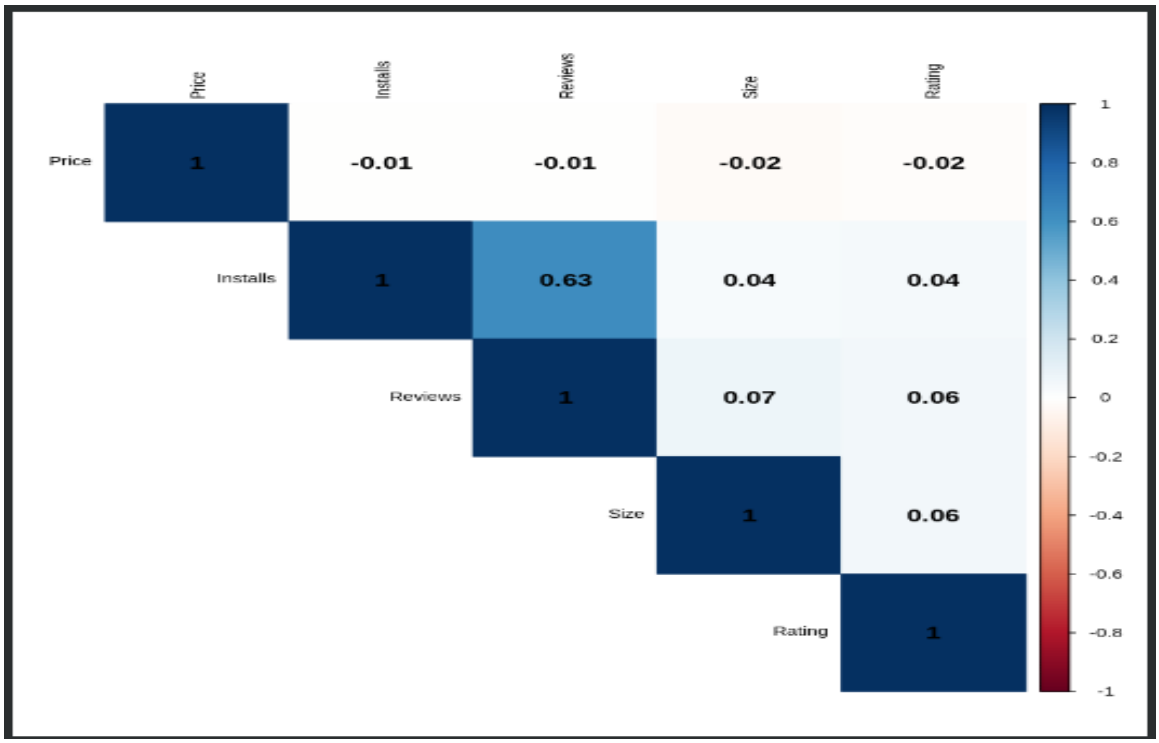

Figure 9: Correlation coefficient between downloads(installs) and other variables

Final Thoughts

Strategies for App Success – Key takeaways

To achieve fourfold growth and user appreciation, focus on developing value-centric apps like WhatsApp and Subway Surfers (Figure 10). Captivate the teen market, unlocking a 200% increase in engagement by tailoring designs to their needs rather than relying on stereotypes. Maximize the potential of summer by incorporating sunshine, special offers, and enjoyable features, resulting in a remarkable 310% surge in downloads. Encourage positive user reviews to propel your app to new heights, recognizing the significant impact of reviews as the rocket fuel for app success.

	Actions	Impact
Value-centric apps WhatsApp, Subway Surfers	▪ Innovate and stand out ▪ Polish for effortless use	↑ 400% grow your app with user love ↑ 250% skyrocket downloads with real value
Target audience Instagram, Clash of Clans	▪ Decode teen behavior, not just age ▪ Design for their needs, not stereotypes	↑ 200% turn teens into raving fans ↑ 160% watch downloads climb
Summer App Google Play Games	▪ Add a summer touch! ▪ Introduce in-app events and special offers	↑ 310% downloads soar like a sunlit swim!
Reviews Rocket Boost Facebook	▪ Use in-app prompts and rewards to encourage immediate reviews	↑ App Growth Explodes! ↑ Happy advocates bring you more to adore

Figure 10: Strategic Formulas for App Success: Growth, Engagement, and Reviews