The Economics of App Success: How Revenue Streams Influence Downloads and User Ratings

Group Assignment - Strategy and Business Models

Group 5

28/11/2024

Jheronimus
Academy
of Data Science

Contents

1 Introduction

With the every-growing popularity of cellphones (Richter 2023), the popularity of mobile applications is also steadily increasing. In 2024, mobile applications are estimated to generate over $900 billion in revenue ("Mobile App Revenue Worldwide by Segment (2019-2027)" 2023). Generally, mobile applications ('*apps*' from here on) tend to be categorized in three different categories (Roma and Ragaglia 2016). Paid apps are the most transparent; they revenue is based on an up-front purchase by the user. Free apps, on the other hand, require no purchase by the user at any stage. According to Roma and Ragaglia (2016), these apps make their revenue from deals with third-parties, either through advertisement or other purposes such as market information.

Finally, freemium apps are, as the name suggests, a middle-ground between free and premium. Users get access to a basic version of the application first and can unlock more features through an in-app payment (Kumar 2014). Of these three revenue models, freemium is the most commonly used and the most (Salehudin and Alpert 2021) and leads to more downloads as well as revenue (Liu, Au, and Choi 2014).

1.1 Academic Background

Most research uses these three established categories—paid, freemium, and free—when discussing revenue models for apps. However, by limiting the discussion to these three terms, nuances within these categories might be missed.

In a review paper from 2023 (Djaruma et al. 2023), different levels of monetization are suggested based on previous literature. These levels provide a clear framework for the revenue models of mobile apps, and are briefly summarized in Table 1.

Strategy	Description
Level 5: Premium	Pay to use the application. This either happens up-front, or after a trial period.
Level 4: Semi-premium	Use a limited number of features for free. Unlock the app with all features through an in-app purchase.
Level 3: In-app advertisement and in-app purchases	Free application with ads, encouraging users to remove ads or to make in-app purchases.
Level 2: Sample and premium	Two different versions of the same app. One is a version with limit features and/or ads. The other version is a premium version.
Level 1: In-app advertisement	Only one version of the app, with only ads and no in-app purchases.
Level 0: Free	The app has no monetization. However, money can still be made through selling user information.

Table 1: Six levels of monetization for apps

The monetization levels as shown in Table 1 are not meant to illustrate the revenue a firm will earn. It is meant to show how much customers may engage in paying for features or services at different levels.

Level 0 is an app that is completely free to use. This requires no payment nor ads, but may still make revenue through selling user data. This risks both ethical pitfalls, as well as scrutiny from customers (Djaruma et al. 2023).

Level 1 is an app that requires no user payment, and only makes revenue through ads. These may also make revenue by selling user information. Large social media apps such as Facebook and Instagram fall in this category (Djaruma et al. 2023).

Level 2 allows the user to purchase either a free or premium version. The free app has limit features; if the user wants to experience the app in full, they will have to purchase the premium version. However, this does run the risk of 'cannibalization', wherein users never upgrade to the premium version. Therefore, it is of the utmost importance to find the right amount of features to include in the free version (Djaruma et al. 2023).

Level 3 combines advertisement and purchases. Many pay-to-win games use this strategy (Nieborg 2016): users can sit through ads and wait a long time for certain rewards, or they can pay for immediate rewards.

Level 4 is similar to level 2: a user has a free trial version, and can upgrade to a premium version. This is often seen in subscription-style apps (Chen 2023).

Level 5 is a totally premium app, where a user has to pay up-front to access as features. This level requires a high-quality app, as users will not pay for an app that is not polished.

1.2 Societal Background

Currently, most apps utilize the freemium revenue model (Salehudin and Alpert 2021). However, as discussed in Djaruma et al. (2023), there are many revenue models between completely premium and completely free. A more fine-grained classification of app revenue models beyond the traditional "paid-freemium-free" framework holds significant societal and business implications.

For society, such distinctions enhance transparency. Some monetization models, such as free or ad-filled apps, may rely on selling user information as a source of revenue (Bamberger et al. 2020). Therefore, clearer distinctions regarding the revenue model will empower consumers to make informed choices. It may also enable policymakers to identify and regulate exploitative practices, such as manipulative microtransactions or intrusive ad models, ensuring all applications align with ethical and legal standards (Mileros and Forchheimer 2024).

For businesses, this paper should unlock more insight into the effectiveness of different revenue streams. This will allow developers to tailor monetization strategies to specific audiences. Furthermore, both consumers and regulatory bodies are growing more concerned with the privacy concerns of apps, especially ones that rely on market information (Mileros and Forchheimer 2024). A granular understanding helps businesses adapt, aligning profitability with sustainability and ethical considerations.

1.3 Research Gap

In short, apps play an increasingly important role in our techno-centric society. To improve the user experience and increase profits, consideration of revenue models is key. Despite the great depth of research on this topic, literature tends to be focussed on the three big categories of paid, freemium, and free. This lack of nuance prevents us from understanding the fine-grained details that may help improve future apps.

The levels of monetization proposed by Djaruma et al. (2023) offer an opportunity to capture this nuance. However, no empirical study has yet applied their framework, as their paper was published only last year. By using this framework to examine how different revenue streams impact app popularity, this study aims to provide valuable insights into consumer preferences. Therefore, this paper seeks to answer the question: *How are the six different revenue models proposed by Djaruma et al. (2023) correlated with app success?*

2 Theory and Hypotheses

In this section, prior research into the topic of revenue streams and its correlation to success in apps will be discussed. As mentioned in the Introduction section, this paper will apply the six levels of revenue as proposed by Djaruma et al. (2023) to app data. The following section will contain a holistic overview of the existing research, as well as hypotheses that arise from this theoretical framework.

2.1 Literature Review

According to Djaruma et al. (2023), there are six distinct revenue strategies for apps. However, it is important to point out they were not the inventors of these revenue strategies; these arose from the literature they analyzed. In the following section, we shall look at the previous research into these revenue strategies.

Firstly, biggest portion of apps are free to download. In fact, this contributes over 95% of all apps in both the Google Play Store as well as the IOS App Store (Statista 2023). But, this does not mean the apps do not generate revenue; they might implement advertisements, in-app purchases or sample and premium versions to still make a profit.

Tower (2023) has predicted global spending on in-app advertisements will reach over 233 billion U.S. dollars in 2026. These funds are an important source of income for mobile app developers (Gao et al. 2022; Maddodi and Upadhyaya 2023).

Aside from advertisements, there are two different revenue strategies to consider: in-app purchases, and a trial and premium version of the same app. Both of these are considered “freemium” in recent literature. Kumar (2014) describes freemium as apps for which users get basic features for free and can update it using a payment. This can be done both through in-app payments, as well as purchasing the premium version of an app.

Another revenue strategy used in apps are in-app payments. This strategy is often used in games. More specifically, games that fall in the “pay-to-win” category (Nieborg 2016).

These games often require the user to wait or watch advertisements, which they can circumvent by paying. Some games also allow users to buy special features, such as the look of their avatar or special powers. Nonetheless, while in-app purchases are a proven revenue strategy, it does come with risks. Notably, in-app payments are also prone to security risks, as it requires more complex interactions and involves more participants than traditional payment (Yang et al. 2019). Therefore, developers need to make sure their payment methods are secure before launching an app with this revenue strategy.

In-app purchases are not the only solution for offering freemium services. A different method is offering two different versions: a free one, and a paid one. The free version often has limit features combined with ads, meaning you make a profit by advertisements (Appel et al. 2020). This is combined with the revenue of the premium version.

To answer the question "*How are the six different revenue models as proposed by Djaruma et al. (2023) correlate to the success of an app?*", we must first define what constitutes to success. In this paper, success will be defined by a couple of factors: popularity, rating, and estimated revenue.

2.1.1 Popularity

The popularity of an app can be measured by the number of downloads. It is important to note the popularity of an app is complex, and is not solely dependent on the chosen revenue model. Other features, such as whether an app is featured on charts, whether it has frequent updates, and word-of-mouth awareness, will also impact the popularity of an app (Aydin Gokgoz, Ataman, and Bruggen 2021). However, despite these other variables, to versions of the same app will still have drastically different performances with different revenue streams (Liu, Au, and Choi 2014).

Building on this, we hypothesize that revenue models associated with fewer barriers to entry will drive higher downloads overall, but the impact of these models may vary by app type and market context.

H1a: Apps that allow the user to have free access to all features (level 0 and 1) will have the highest amount of downloads overall.

As Djaruma et al. (2023) has shown, the most highly ranked apps are level 1. These apps are the big social media platforms such as Instagram and Facebook. Therefore, we expect this to be reflected within our data and the following to be true.

H1b: The apps with the most downloads will be level 1.

Freemium apps can be implemented with either in-app purchases or different versions. For the latter, we expect more downloads to be generated by the free app, than its premium counterpart. As demonstrated by Liu, Au, and Choi (2012), users tend to download the trial version before committing to a premium version. Thus, the following is likely to be true.

H1c: For apps that utilize a sample and a premium version of the same app (level 2), the free versions of an app will have more downloads than their paid-for counterpart.

Finally, many games use a pay-to-win mechanism (Nieborg 2016). Therefore, we expect these in-app payment constructions to be used for most highly downloaded games.

H1d: The most downloaded apps in the gaming category will likely fall under level 3 and 4.

2.1.2 Rating

The downloads of an app are not everything. An app can be downloaded often, but may not be highly rated. Ratings provide valuable insights into user satisfaction, which often reflects the perceived quality and value of an app. Therefore, we hypothesize revenue models that prioritize short-term gains may achieve high download counts but could negatively affect ratings if users feel misled or dissatisfied.

The main draw of a freemium model is to attract users, and have them update to a paid version (Kumar 2014). However, as Kumar (2014) points out, this can be a double-edged sword. Too few features, and it may not be attractive to users. Too many features, and the users will not update. This leads to the following hypothesis.

H2a: Apps that require the user to pay to unlock features (level 2, 3, and 4) will tend to have lower ratings than the version that requires payment upfront (level 5).

The majority of the apps are free-to-use (Djaruma et al. 2023). However, this means there might be more difference in quality between these apps, as the barrier to downloading is lower for free apps than paid or freemium apps (Mileros and Forchheimer 2024). Therefore, we postulate the following.

H2b: Apps that allow the user to have free access to all features (level 0 and 1) will have more variance in their rating, as quality can vary for free-to-access apps.

In the same vein as H2a, users have more realistic expectations of paid apps compared to apps that require you to unlock features (Kumar 2014). Therefore, more users downloading premium apps will be satisfied with their purchase, leading to less variance. Thus, the following should follow from our data.

H2c: Fully premium apps (level 5) and premium version of level 2 will have less variance in their ratings, while all other levels will have more.

While popularity and ratings are distinct measures of success, they are inherently connected. Revenue models that increase downloads can also shape user expectations, which in turn influence ratings. For example, level 1 apps may dominate in terms of downloads (H1a) but face fluctuating ratings (H2c) due to quality variation. Similarly, level 2 apps may achieve high download counts for their free versions (H1c), but the gap between free and paid features could drive dissatisfaction (H2a).

2.1.3 Revenue Estimation

It is important to point out downloads and ratings likely do not directly correlate to the actual revenue of an app. The revenue of apps "premium" apps that require an upfront payment, the revenue is relatively simple to track and compare. However, for apps that rely on advertisement, in-app purchases and/or selling market information, this is harder to track.

For apps that solely on advertisement, time retention can be a good measure of revenue (Ross 2018). However, this only works if the app solely relies on ads. An example of this given by Djaruma et al. (2023) is TikTok: this app relies not only on advertisement, but also on users purchasing products through its shop. Therefore, using solely the time retention would not accurately capture the revenue of an app with both revenue streams. Furthermore, the selling of user data is usually not publicized, meaning it is not possible to know the revenue from this.

Unfortunately, our data only contains the price of “premium” app versions. The data does not include any details regarding in-app purchases nor time-retention. Because of this lack of sufficient data, solely downloads and ratings will be taken into account as indicators of success.

3 Methods and Data

In this section, we will discuss the dataset and methods used to test the hypotheses outlined in the previous section. The focus lies on providing a comprehensive description of the dataset, including its structure and the variables it contains, followed by an explanation of the variable selection process. Additionally, we outline the statistical methods applied and discuss how assumptions, such as missing values and potential biases, were addressed to ensure the robustness of our analysis.

3.1 Dataset Description

The dataset used in this research comprises 1,016,666 instances and 27 variables, offering a comprehensive overview of mobile applications across various revenue models. Each instance corresponds to a single app, capturing details about its characteristics, user engagement, and monetization strategies. Key variables include the app’s unique identifier (my_app_id), the number of downloads (num_downloads), the average rating (rating_app), and the number of ratings received (nb_rating). These variables provide critical insights into app performance metrics.

Additional variables include pricing details (price_gplay), information about in-app purchases (in_app), and whether the app contains advertisements (has_ads). Other attributes, such as content ratings (content_rating_app) and metadata about the app developer (developer_name), further enhance the dataset’s richness by adding contextual information.

Several variables, such as whats_new (completely null) and in_app_product (89.57% null), were excluded from the analysis due to their high proportions of missing data. Conversely, variables with minor missingness (e.g., date_published, privacy_policy) were retained after appropriate preprocessing. This curated dataset is the foundation for examining monetization strategies and their relationship to app performance metrics like downloads and user ratings.

3.2 Variable Selection

The analysis focuses on a subset of 13 variables selected from the dataset, as shown in Table 2 below. These variables were chosen for their relevance to our research questions, capturing information about app characteristics, user engagement, and monetization strategies.

Variable Name	**Description**
`my_app_id`	Unique identifier for each app.
`num_downloads`	Number of downloads for the app. Key indicator of app success.
`rating_app`	Average user rating. Measures user satisfaction.
`nb_rating`	Number of ratings received. Reflects user engagement.
`price_gplay`	Price of the app. Used to differentiate free vs premium apps.
`in_app`	Boolean indicating whether the app has in-app purchases.
`has_ads`	Boolean indicating whether the app includes advertisements.
`content_rating_app`	Content rating of the app (e.g., Everyone, Teen).
`categ_app`	App category (e.g., Productivity, Games). Groups apps by functionality.
`developer_name`	Name of the app developer. Provides context on developer reputation.
`developer_info`	Additional metadata about the developer (e.g., location, website).

To systematically explore monetization strategies, we classified the apps into six distinct levels based on their monetization models. These levels reflect varying approaches to generating revenue, ranging from completely free apps to fully premium paid apps.

- **Level 0**: Apps with no monetization, offering free services without ads or in-app purchases.
- **Level 1**: Free apps monetized solely through ads.
- **Level 2**: Freemium model, employing both free sample apps with limited functionality (and potentially ads) and paid premium apps with full features.
- **Level 3**: Apps combining ads and in-app purchases, monetizing through both strategies.
- **Level 4**: Freemium apps monetized entirely through in-app purchases, removing ads for a seamless user experience.
- **Level 5**: Fully premium paid apps, with no ads or in-app purchases, delivering a premium experience.

This classification is grounded in theoretical frameworks, such as the monetization levels proposed by Djaruma et al. (2023), and allows for a nuanced analysis of how different revenue models impact app success metrics like user ratings and downloads. Our systematic categorization facilitates a deeper understanding of the relationship between monetization strategies and app performance.

3.3 Statistical Methods

To test our hypotheses, we employed a combination of descriptive statistics, text processing, and machine learning techniques, ensuring a rigorous approach to analyzing the dataset. Descriptive statistics were used to explore distributions and trends in key metrics such as num_downloads, rating_app, and price_gplay. Applications were categorized into six monetization levels using binary indicators (is_free, in_app, and has_ads), while preprocessing of price data facilitated the distinction between free and paid applications. These steps established a structured foundation for investigating relationships between monetization strategies and app performance.

For the paired application analysis within Level 2, we applied Term Frequency-Inverse Document Frequency (TF-IDF) vectorization combined with cosine similarity to app names, allowing us to measure textual similarity (Widianto et al. 2023). Logical pairings were refined through prefix matching and the identification of indicative terms like "Free" or "Pro." This process was further validated by ensuring paired apps shared the same developer, using metadata from developer_name and developer_info. While effective, this approach acknowledged potential limitations, such as ambiguity in naming conventions, which were flagged as edge cases for transparency.

To address potential biases in app ratings caused by low review counts, we calculated a Bayesian average, adjusting raw ratings by combining the global average with individual app ratings weighted by review volume (Moyeed and Clarke 2005). This provided a more balanced measure of user satisfaction, reducing the influence of apps with disproportionately few reviews. Scatter plots and box plots were employed to visualize the relationships between monetization levels and user engagement metrics, offering valuable preliminary insights into patterns within the data. These visualizations, alongside the Bayesian adjustments, laid the groundwork for robust and interpretable results, which are elaborated in the results section.

3.3.1 Handeling of Assumptions

We addressed missing values by removing rows with critical nulls, such as those in num_downloads, to maintain data integrity. Text-based variables like content_rating_app were standardized to ensure consistency. For price_gplay, currency symbols were removed to facilitate the classification of applications into free or paid categories.

Outliers in metrics like num_downloads were retained if they represented industry-leading applications, as their exclusion could skew the analysis. The use of Bayesian averages mitigated bias in rating_app due to low review counts, providing a more accurate reflection of user satisfaction. Covariance checks were conducted to ensure the absence of multicollinearity among numerical variables, thereby enhancing the reliability of correlation and regression analyses.

Some applications exhibited rare combinations of is_free, in_app, and has_ads that did not fit within the predefined monetization levels. These applications were excluded from the analysis but documented as a limitation. Edge cases in level 2 application pairing were flagged for potential mismatches due to naming ambiguities, ensuring transparency in the classification process.

These methodologies facilitated a systematic and accurate exploration of monetization models and their impact on application performance.

4 Results

In this section, we will analyze the data through tables and visualizations. These plots largely explore the data around the hypotheses and research question discussed in the previous sections. The aim of this section is to test the hypotheses.

The results aim to provide an answer to the question "*How are the six different revenue models as proposed by Djaruma et al. (2023) correlate to the success of an app?*". To do so, the six different revenue models defined by Djaruma et al. (2023). Short descriptions of these levels can be found in Table 1.

In order to analyze the Google Play Store data, it was first divided into these different levels. More than 75% of all the apps belong to level 0 and 1 (see Figure 1). The smallest portion of apps belongs to level 2, which included both version of the same app.

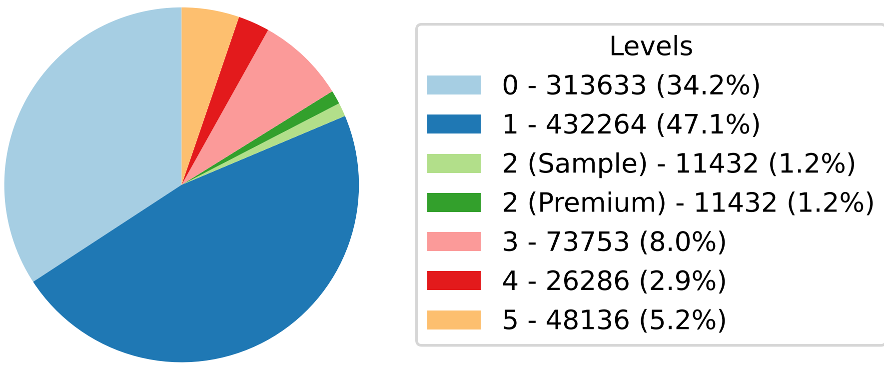


Figure 1: Distribution of the amount of apps across the revenue levels

To find the success of an app, metrics like popularity, the rating and the revenue estimation of an application can be used. However, as discussed in Section 2.1, the revenue estimation will be disregarded due to a lack of data on this variable. The number of downloads gives us a measure of popularity, and will be used to answer hypotheses H1a through H1d. The average rating given by users will allow for evaluation of hypotheses H2a through H2d.

4.1 Downloads

In this section, we will venture to answer hypotheses H1a through H1d. This will be done in several subsections. Firstly, in Section 4.1.1 the distribution of the downloads will be analyzed, in order to answer hypotheses H1a and H1b. In Section 4.1.2, the differences in downloads between the sample and premium versions of level 2 will be explored to answer hypotheses H1c. Lastly, in Section 4.1.3 the downloads of the gaming apps will be further examined in order to answer hypothesis H1d.

4.1.1 Distribution of app downloads

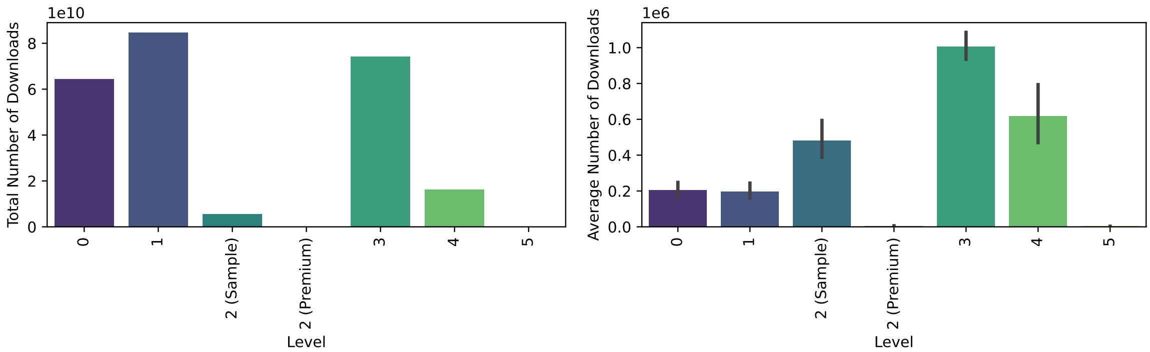
This section will venture to answer the following hypotheses.

H1a: Apps that allow the user to have free access to all features (level 0 and 1) will have the highest amount of downloads overall.

H1b: The apps with the most downloads will be level 1.

In order to find whether these hypotheses are true for this dataset, the number of downloads per revenue model were visualized (Figure 2). Figure 2a displays the total amount of downloads, while the Figure 2b displays the average amount of downloads.

The total number of downloads (Figure 2a) and the average number of downloads (Figure 2b) vary wildly for levels 0 and 1. This is largely a result of the fact that 75% of the total apps fall under these levels, as can be seen in Figure 1. Thus, by looking solely at the average amount of downloads per level, it does not seem that level 0 and 1 are the most downloaded apps.



(a) Total Number of Downloads by Level

(b) Average Number of Downloads by Level

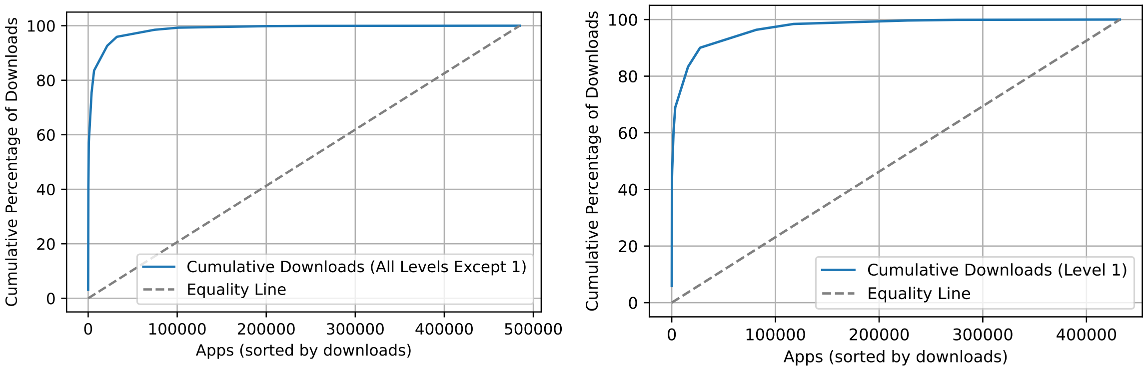
Figure 2: Comparison of Total and Average Downloads by Level

As the majority of apps fall under level 0 and 1, these do have the highest overall downloads, as shown in Figure 2a. Table 3 shows just how big the difference between the total downloads is. The apps that are completely free to use (level 0 and 1) have over 40 billion more downloads than the apps that are paid to some extend (levels 2 through 5).

	category	num_downloads
0	Free all access	149140785038
1	Freemium and Premium	96093512270

Table 3: Difference Total Number of Downloads

Figure 3 provides insights into distribution inequality. The distribution of the app downloads are highly concentrated, meaning a small proportion of the apps make up for the vast majority of the downloads. Figure 3a shows relatively the same inequality as Figure 3b. This indicates a small amount of apps dominate the numbers. However, it is important to note this might not be exclusive to level 0 and 1; the other levels might exhibit similar patterns.



(a) All levels except 1

(b) Level 1 Only

Figure 3: Distribution of App Downloads (Top-Heavy Analysis)

The download range of apps in Table 4 illustrates that level 1 dominates in terms of apps with over 1 billion downloads. Furthermore, level 0 also has a few apps with over 1 billion downloads. Level 0 has the most app downloads between 500 million and 1 billion with 18.

	0-1M	1M-100M	100M-500M	500M-1B	1B+
0	312119	1469	17	18	1
1	429133	3105	7	7	3
2 (Premium)	11431	0	0	0	0
2 (Sample)	11140	291	1	0	0
3	69699	4040	10	3	0
4	25738	538	8	2	0
5	48124	0	0	0	0

Table 4: Number of Downloads by Level

Taking a closer look at the apps provided by Djaruma et al. (2023) in Table 5. We do see that Facebook, Instagram, Spotify, Snapchat and Amazon Shopping fall under level 1.

	App Name	Downloads (in B)	Category	Level
0	WhatsApp	1	Communication	0
1	Facebook Lite	1	Social	1
2	Instagram	1	Social	1
3	Facebook Messenger	1	Communication	4

	App Name	Downloads (in B)	Category	Level
4	Spotify	0.5	Music & Audio	1
5	Snapchat	0.5	Social	1
6	Netflix	0.5	Entertainment	4
7	Amazon Shopping	0.1	Shopping	1

Table 5: Top 8 Apps by Downloads (in Billions)

To summarize these findings, free apps are downloaded more than paid apps (Table 3). However, these apps are not downloaded more on average, as can be seen in Figure 2b.

This means that H1a is partially true. Apps with level 0 and 1 do have the most amount of downloads, but this does not hold true for the average amount of downloads. This can be contributed to the fact that the majority of the dataset is either level 0 (34.2%) or level 1 (47.1%) (Figure 1). Therefore, the support for H1a remains inconclusive, and the hypothesis must be discarded.

The most downloaded apps do fall under level 1, as can be seen in Figure 2a. Further analysis shows that there is high inequality, resulting in a small proportion of the apps accounting to a majority of the downloads (Figure 3). Furthermore, Table 4 demonstrates that apps in level 1 are considered among the top downloaded apps. Looking at the apps provided by Djaruma et al. (2023), most social network platforms indeed fall under level 1 (see Table 5). Given these observations, it can be concluded that there is enough evidence to support hypothesis H1b.

4.1.2 Sample and Premium Apps

This section will test the following hypothesis.

H1c: For apps that utilize a sample and a premium version of the same app (level 2), the free versions of an app will have more downloads than their paid-for counterpart.

Levels 2 (premium) and 5 have the least amount of total and average downloads, as shown in Figure 2. Furthermore, we also see a high disparity in average downloads between level 2 sample and premium, as shown in Figure 2b.

As shown in Figure 4 , just 1.1% of the downloads are attributed to premium apps within all apps in level 2. This illustrates the incredible discrepancy between these download rates.

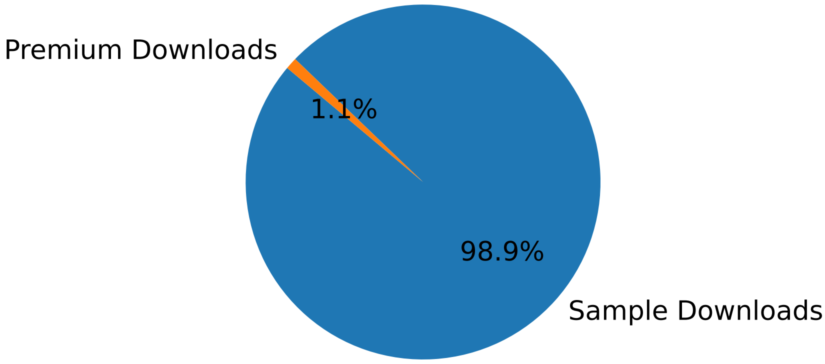


Figure 4: Proportion of Total Downloads: Sample vs Premium

High disparity is also evident in the average downloads, as shown in Table 6. Where the average download difference is close to 500.000. With an average download ratio of nearly 27%. Meaning that about one in fourth users that download the sample app, also download the premium app.

	Metric	Value
0	Average Sample Downloads	480358
1	Average Premium Downloads	5213.22
2	Average Download Difference	-475145
3	Average Download Ratio	0.265126

Table 6: Sample and Premium download metrics

In hypothesis H1c, we claim that the sample version should have more downloads than their paid-for counterpart. From Figure 2 and Figure 2b, it is apparent there is indeed an imbalance in the download rate between these classes. By examining the proportion of the sample vs premium downloads more closely (see Figure 4), this further strengthens this argument. Additionally, the high disparity across metrics in Table 6 is even more evidence to support H1c.

Therefore, hypothesis H1c is accepted.

4.1.3 Games in the Google Play Store

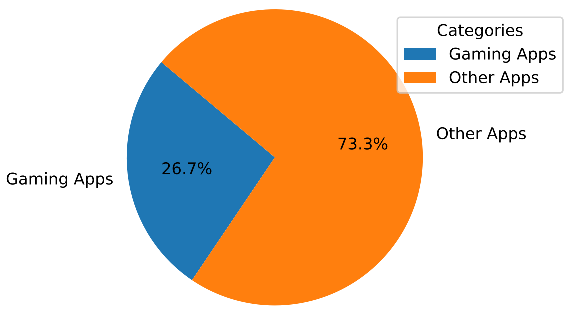
In this section, the following hypothesis shall be tested:

H1d: The most downloaded apps in the gaming category will likely fall under level 3 and 4.

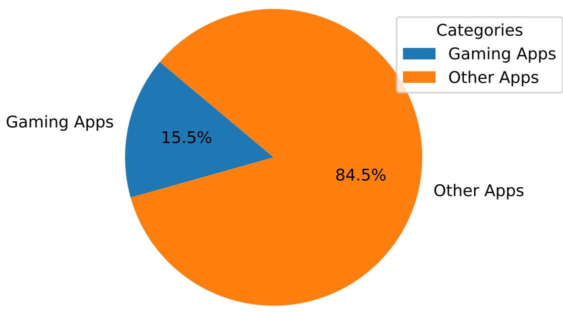
As is shown in Figure 2, there is no major difference between total and average number of downloads in level 3 and 4. Both the total and average number of downloads of level 3 and

4 are relatively high compared to other categories. Hypothesis 1d, which concerns these 2 revenue levels, claims that many popular games fall under these 2.

According to Figure 5b, 15.5% of the total apps are games. These games attribute to 26.7% of the total app downloads, as shown in Figure 5a. On average, the amount of downloads for games is 98.6% higher than the average downloads of the other apps (see Table 7).



(a) Total Downloads: Gaming vs Other Apps



(b) Total Amount: Gaming vs Other Apps

Figure 5: Proportion of Games compared to other apps

	Category	Avg Downloads	% Diff Gaming vs Other
0	Gaming	460828	98.6
1	Other	232041	-98.6
2	All	267450	-

Table 7: Average Downloads Games vs Other Apps

Up until now, these findings suggest that games are hugely popular, regardless of their revenue level. Figure 6 shows that particularly games in levels 3 and 4 accrue a lot of downloads, compared to other levels.

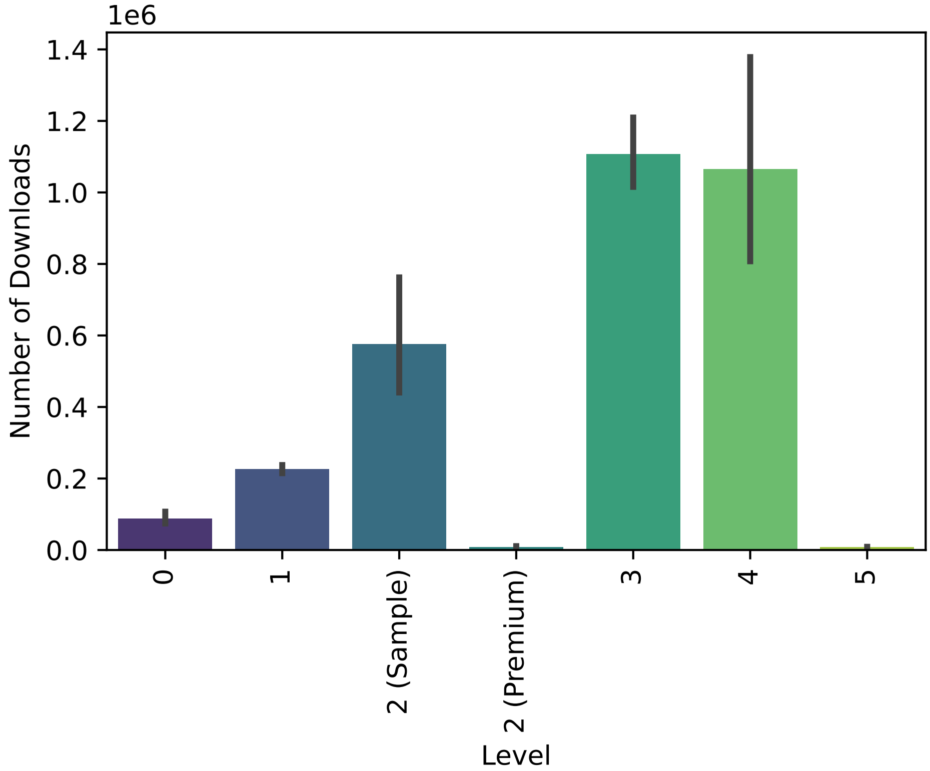


Figure 6: Average amount of Downloads for Gaming Apps by Level

All in all, games in the Google Play Store contribute a sizeable part of all downloads. As shown in Figure 5a, it attributes to 26.7% of the total downloads. Furthermore, the average downloads of games is 98.6% more than any other apps with an average close to 0.5 million downloads (see Table 7). In level 3 and 4, the average downloads jumps to over 1 million average downloads (see Figure 6). Given these observations, it can be concluded that there is enough evidence to support hypothesis H1d.

4.2 Ratings

In this section, we will venture to answer hypotheses H2a through H2c. To start, in Section 4.2.1, we will try to find fluctuations in ratings that might be by the quality in order to answer hypothesis H2a and H2b. Moreover, in Section 4.2.2, the variance in ratings will be analyzed to test hypothesis H2c.

4.2.1 Quality of Free and Paid Apps

This section of the variance of free and premium apps, will test the following hypothesis.

H2a: Apps that require the user to pay to unlock features (level 2, 3, and 4) will tend to have lower ratings than the version that requires payment upfront (level 5).

H2b: Apps that allow the user to have free access to all features (level 0 and 1) will have more variance in their rating, as quality can vary for free-to-access apps.

Hypothesis 2a suggests that the apps where users need to pay to unlock features tend to have a lower rating than apps that require payment upfront.

As shown in Figure 7, premium apps have a higher rating on average compared to free apps with users needing to pay to unlock features. The free apps show a greater spread in ratings.

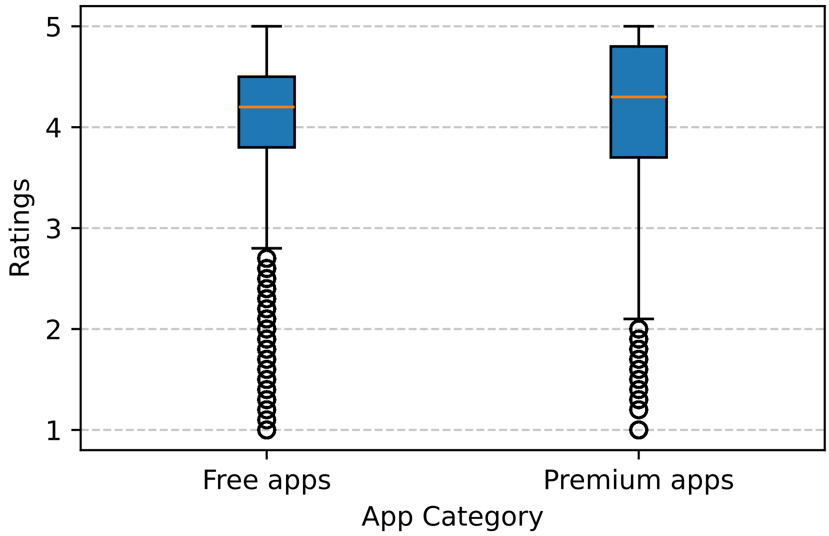


Figure 7: Boxplot of Ratings for Free and Premium Apps

The apps that require upfront payment receive a higher rating on average then free apps, shown in Figure 7. This can indicate that users perceive as being of higher quality. The greater spread of ratings for free apps, may indicate that the app quality varies more. This provides sufficient evidence in support of hypothesis H2a.

Level 0 and 1 have the highest overall downloads, as shown in Table 3 by quite a margin (40 billion downloads (0.4e11)) compared to the other apps. These apps under level 0 and 1 have free access to all features. Hypothesis 2b suggests that the variance in rating between "free access to all feature apps" and "free access to not all features or paid apps" is a sign of quality disparity.

Table 8 shows the variance and standard deviation between "free access to all feature apps" and "free access to not all features or paid apps". From which we can conclude that there is not a big statistical difference. Roughly, 0.02 and 0.01 difference in variance and standard deviation rating respectively.

	category	variance in rating	standard deviation in rating
0	Free all access	0.531282	0.728891
1	Freemium and Premium	0.511318	0.715065

Table 8: Variance (and std) difference. Between free access to all feature apps and free access to not all features or paid apps

Figure 8 visualizes variance with outliers. It shows that both categories have a similar median rating close to 4.5, with “Free/Paid not all feature apps” being slightly lower. Both categories seem to have a lot of outliers under 3, with ‘free all feature apps’ having more outliers in 2 and 1.

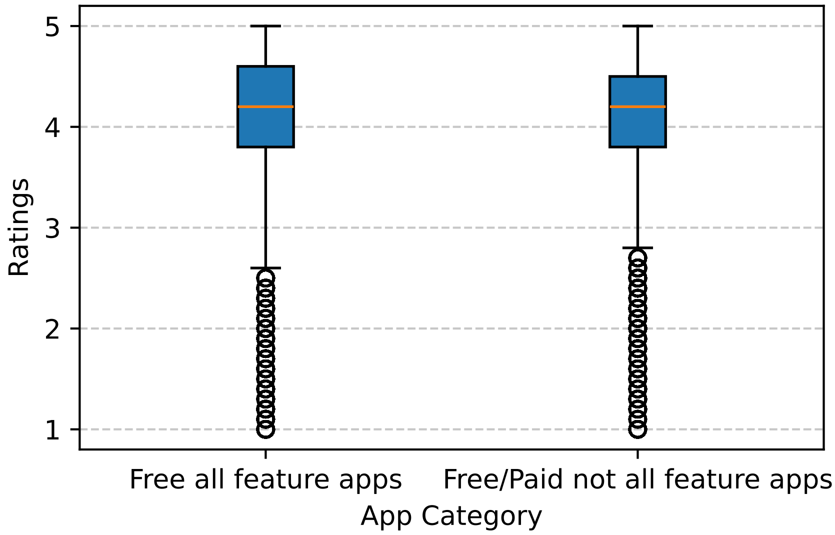


Figure 8: Boxplot of Ratings by App Category

The variance and standard deviation is minimal between the two categories, as shown in Table 8 . Figure 8 illustrates that both categories receive high ratings, with significant number of outliers. These outliers can be interpreted as variability in user satisfaction. The slighty higher median in ‘Free all feature apps’ may imply that users prefer this over ‘Free/Paid not all feature apps’. However, this is marginal and would probably require statistical validation. Thus, the evidence to suggest level 0 and 1 have more variance is inconclusive, leading to rejection of hypothesis H2b.

4.2.2 Variance in Ratings

This section will attempt to answer the following hypothesis regarding the variance in ratings

H2c: Fully premium apps (level 5) and premium version of level 2 will have less variance in their ratings, while all other levels will have more.

The variance across levels shows some major differences, as illustrated in Figure 9. However, when adjusted for total reviews using a Bayesian average rating, the differences in variance diminishes. This adjustment smooths out fluctuations, offering a more consistent representation of the ratings. For level 5, we see that the variance is one of the lowest when adjusted by Bayesian average. This difference may arise because users who are satisfied with the app tend to rate the app highly, while users who are not satisfied, tend to rate it much lower, having paid for the app.

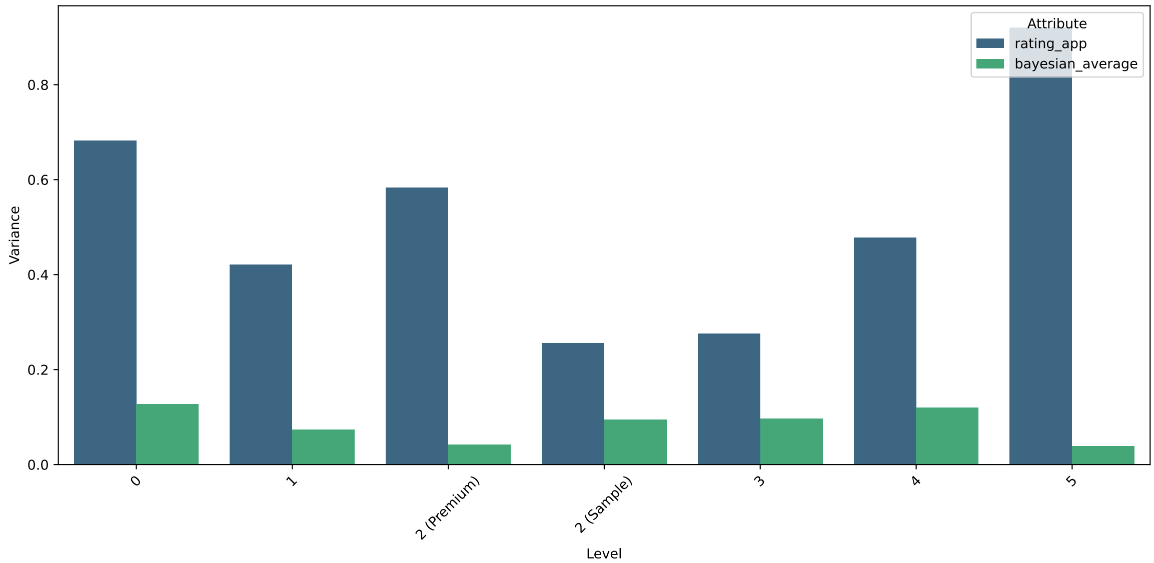


Figure 9: Variance of Different Attributes Grouped by Level

Figure 10 illustrates a significant difference in the variance of ratings. Premium apps tend to have a higher variance, with more outliers.

	Category	Variance in Ratings
0	Other apps	0.504639
1	Premium apps	0.858123

Figure 10: Variance in Ratings for Other vs Premium Apps

In conclusion, the variance across all levels show that level 5 has the highest variance for rating and the lowest variance for bayesian rating, as shown in Figure 9. The bayesian average seems more reliable regarding the overall app performance. The premium apps may have polarized feedback due to high user expectations. In short, this provides sufficient evidence to support hypothesis H2c.

4.3 Conclusion

The six revenue strategies proposed by Djaruma et al. (2023) show distinct relationships with app success, balancing accessibility and user satisfaction. Free models (levels 0 and 1) dominate in popularity, achieving ~40 billion downloads, but show variable ratings. Their accessibility attracts large audiences but often struggles with quality consistency. Freemium models (levels 2, 3, and 4) show more variability in ratings, reflecting differing user expectations, yet remain popular by allowing users to try features before committing. Premium apps (level 5) achieve the highest average ratings with minimal variance, reflecting their polished quality and clear value propositions, though they trade reach for exclusivity. Each model offers trade-offs: free models maximize reach, freemium models balance accessibility and satisfaction, and premium models excel in consistent quality, appealing to smaller, selective audiences.

4.4 Ethical Considerations

This research uses publicly available Google Play Store data without involving individual user data, ensuring privacy concerns are avoided (Tikkinen-Piri, Rohunen, and Markkula 2018). However, even public data can reveal sensitive business insights, such as metrics that expose proprietary strategies. To mitigate this, we avoid singling out specific apps. The findings could influence developers, marketers, and policymakers by revealing insights into profitable monetization methods like ads or in-app purchases. While these insights can help optimize revenue strategies, they also raise ethical concerns, as they might encourage exploitative practices that prioritize profits over user experience (Mileros and Forchheimer 2024). This underscores our dual responsibility to provide actionable insights while advocating for ethical applications of our conclusions.

5 Discussion

This study examined the relationship between six revenue strategies and app success, measured by downloads and ratings. The findings reveal clear trade-offs among these models, offering insights into app monetization strategies.

5.1 Reflection on the Findings

We found that free apps (Levels 0 and 1) dominate in terms of downloads. However, these apps also have higher variance in ratings, suggesting they struggle with consistent quality. While this aligns with existing literature suggesting that free apps attract a broad audience but often deliver inconsistent experiences (Mileros and Forchheimer 2024), it does raise the question of whether the amount of downloads should be the primary measure of success. Free apps may have more downloads, but without understanding the deeper dynamics of user engagement and satisfaction, these figures could be misleading. For example, apps with high ratings but fewer downloads could still represent significant success in specific niches.

The finding that freemium models (Levels 2 and 3) strike a balance between accessibility and profitability is promising, as it suggests a flexible approach that could cater to both casual users and those willing to pay for premium features. As noted in previous research (Kumar 2014), user retention and monetization are deeply interconnected in freemium models. This issue could explain why some freemium apps fail to reach their full potential despite high download numbers. Furthermore, we concluded the correlation between sample and premium ratings is likely due to the app quality both versions. But, this does on take into account the influence of other factors such as app category and external marketing efforts (Stocchi et al. 2022) on the ratings of the app.

The strong correlation found between premium apps (Level 5) and higher ratings aligns with the assumption that users expect a higher level of quality from paid apps. However, the trade-off here is the limited reach of premium apps, as shown by their lower download numbers. This finding suggests that while premium apps may offer superior experiences, their market appeal is often restricted by their pricing structure. It raises the issue of how to balance quality with accessibility in a market where competition is fierce, and user expectations are high.

This finding aligns with the broader literature on optimal distinctiveness, which explores how firms should manage the balance between conformity and differentiation in order to achieve superior performance (Zhao et al. 2018). On the one hand, products need to conform to category norms to build legitimacy, but on the other hand, they should differ enough to avoid intense competition. The study by (Angeren et al. 2022) further extends this line of inquiry by considering how the relationship between differentiation and performance varies based on the product's revenue model. For premium apps, moderate differentiation appears to be an optimal strategy, allowing for both quality perception and enough market fit, but with the trade-off of a smaller, more specific user base.

5.2 Limitations

Several limitations must be acknowledged. Downloads do not necessarily indicate revenue for freemium models (Djaruma et al. 2023). The time the user spends on an app and the purchases made within this app (Ross 2018) are better measures of the revenue for freemium applications.

Moreover, the data used for this project was collected before the launch of Tiktok, the social media platform with over 900 million users worldwide (Ceci 2024). As is mentioned in Section Section 2.1.3, Tiktok has a business model that relies more on purchases through their online shop than solely user retention, making it different from other social media platforms. The inclusion of TikTok might have lead to different results when analyzing the amount of downloads, particularly level 3.

Causality is another limitation. While correlations were observed between revenue models, downloads, and ratings, confounding factors may be at play. Variables such as marketing efforts, app category, and regional preferences could influence success. Consequently, the findings should not be interpreted as direct causal relationships.

5.3 Practical Implications for Businesses

For app developers, these findings offer valuable guidance. Selecting a revenue model should align with the app's target audience and objectives. Free models are ideal for maximizing reach, but consistent quality control is crucial. Freemium models provide flexibility, catering to diverse user preferences while driving profitability through in-app purchases. Premium models are suited for niche markets with high user expectations but require substantial upfront investment in quality.

Businesses must avoid over-reliance on these results, as they only reflect correlations. Nonetheless, revenue strategies should consider these trade-offs, balancing user satisfaction and profitability.

5.4 Future Research Directions

Future studies could address limitations by incorporating datasets from more recent app markets, including platforms like TikTok. Further exploration of hypotheses that were not supported, such as H2b, could provide deeper insights into rating variance. Expanding the analysis to include additional metrics, such as user retention and in-app purchases, would enhance the understanding of revenue models' impact on success.

Ultimately, app monetization is a dynamic field. Ongoing research is essential to keep pace with evolving user preferences and technological advancements. By building on these findings, future work can contribute to more nuanced strategies for app success.

6 References

Angeren, Joey van, Govert Vroom, Brian T. McCann, Ksenia Podoynitsyna, and Fred Langerak. 2022. "Optimal Distinctiveness Across Revenue Models: Performance Effects of Differentiation of Paid and Free Products in a Mobile App Market." *Strategic Management Journal* 43 (10): 2066–2100. https://doi.org/10.1002/smj.3394.

Appel, Gil, Barak Libai, Eitan Muller, and Ron Shachar. 2020. "On the Monetization of Mobile Apps." *International Journal of Research in Marketing* 37 (1): 93–107. https://doi.org/10.1016/j.ijresmar.2019.07.007.

Aydin Gokgoz, Zeynep, M. Berk Ataman, and Gerrit H. van Bruggen. 2021. "There's an App for That! Understanding the Drivers of Mobile Application Downloads." *Journal of Business Research* 123 (February): 423–37. https://doi.org/10.1016/j.jbusres.2020.10.006.

Bamberger, Kenneth A., Serge Egelman, Catherine Han, Amit Elazari Bar On, and Irwin Reyes. 2020. "Can You Pay for Privacy? Consumer Expectations and the Behavior of Free and Paid Apps." *Berkeley Technology Law Journal* 35: 327. https://heinonline.org/HOL/Page?handle=hein.journals/berktech35&id=339&div=&collection=.

Ceci, Laura. 2024. "TikTok: Number of Global Users 2020-2025." https://www.statista.com/statistics/1327116/number-of-global-tiktok-users/; Statista.

Chen, Li. 2023. "Analysis of Online Platforms' Free Trial Strategies for Digital Content Subscription." *Journal of Theoretical and Applied Electronic Commerce Research* 18 (4): 2107–24. https://doi.org/10.3390/jtaer18040106.

Djaruma, Heryan, Zaki Widyadhana Wirawan, Alexander Agung Santoso Gunawan, and Karen Etania Saputra. 2023. "2023 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT)." In, 577–83. https://doi.org/10.1109/COMNETSAT59769.2023.10420748.

Gao, Cuiyun, Jichuan Zeng, David Lo, Xin Xia, Irwin King, and Michael R. Lyu. 2022. "Understanding in-App Advertising Issues Based on Large Scale App Review Analysis." *Information and Software Technology* 142 (February): 106741. https://doi.org/10.1016/j.infsof.2021.106741.

Kumar, Vineet. 2014. "Making "Freemium" Work." *Harvard Business Review*, May. https://hbr.org/2014/05/making-freemium-work.

Liu, Charles Zhechao, Yoris A. Au, and Hoon Seok Choi. 2012. "An Empirical Study of the Freemium Strategy for Mobile Apps: Evidence from the Google Play Market." In *Proceedings of the 2012 International Conference on Information Systems (ICIS)*, 1–19. http://aisel.aisnet.org:80/cgi/viewcontent.cgi?article=1050&context=icis2012.

———. 2014. "Effects of Freemium Strategy in the Mobile App Market: An Empirical Study of Google Play." *Journal of Management Information Systems* 31 (3): 326–54. https://doi.org/10.1080/07421222.2014.995564.

Maddodi, Chetana Balakrishna, and Pallavi Upadhyaya. 2023. "In-App Advertising: A Systematic Literature Review and Implications for Future Research." *Spanish Journal of Marketing - ESIC* 28 (3): 334–55. https://doi.org/10.1108/SJME-05-2022-0120.

Mileros, Martin D., and Robert Forchheimer. 2024. "Free for You and Me? Exploring the Value Users Gain from Their Seemingly Free Apps." *Digital Policy, Regulation and Governance* ahead-of-print (ahead-of-print). https://doi.org/10.1108/DPRG-01-2024-0009.

"Mobile App Revenue Worldwide by Segment (2019-2027)." 2023. Statista. https://www.statista.com/forecasts/1262892/mobile-app-revenue-worldwide-by-segment.

Moyeed, R. A., and R. T. Clarke. 2005. "The Use of Bayesian Methods for Fitting Rating Curves, with Case Studies." *Advances in Water Resources* 28 (8): 807–18. https://doi.org/https://doi.org/10.1016/j.advwatres.2005.02.005.

Nieborg, David B. 2016. "Free-to-Play Games and App Advertising: The Rise of the Player Commodity." In. Routledge.

Richter, Felix. 2023. "Charted: There Are More Mobile Phones Than People in the World." World Economic Forum. https://www.weforum.org/stories/2023/04/charted-there-are-more-phones-than-people-in-the-world/.

Roma, Paolo, and Daniele Ragaglia. 2016. "Revenue Models, in-App Purchase, and the App Performance: Evidence from Apple's App Store and Google Play." *Electronic Commerce Research and Applications* 17 (May): 173–90. https://doi.org/10.1016/j.elerap.2016.04.007.

Ross, Nicholas. 2018. "Customer Retention in Freemium Applications." *Journal of Marketing Analytics* 6 (4): 127–37. https://doi.org/10.1057/s41270-018-0042-x.

Salehudin, Imam, and Frank Alpert. 2021. "No Such Thing As A Free App: A Taxonomy of Freemium Business Models and User Archetypes in the Mobile Games Market," December. https://papers.ssrn.com/abstract=4001100.

Statista. 2023. "Number of Applications for Mobile Phones." https://www.statista.com/statistics/263797/number-of-applications-for-mobile-phones/.

Stocchi, Lara, Naser Pourazad, Nina Michaelidou, Arry Tanusondjaja, and Paul Harrigan. 2022. "Marketing Research on Mobile Apps: Past, Present and Future." *Journal of the Academy of Marketing Science* 50 (2): 195–225. https://doi.org/10.1007/s11747-021-00815-w.

Tikkinen-Piri, Christina, Anna Rohunen, and Jouni Markkula. 2018. "EU General Data Protection Regulation: Changes and Implications for Personal Data Collecting Companies." *Computer Law & Security Review* 34 (1): 134–53. https://doi.org/10.1016/j.clsr.

2017.05.015.

Tower, Sensor. 2023. "App Market Forecast 2026." https://sensortower.com/blog/sensor-tower-app-market-forecast-2026.

Widianto, Adi, Eka Pebriyanto, Fitriyanti Fitriyanti, and Marna Marna. 2023. "Document Similarity Using Term Frequency-Inverse Document Frequency Representation and Cosine Similarity." *DINDA: Digital Information and Database* 4 (2): —. https://doi.org/10.20895/dinda.v4i2.1589.

Yang, Wenbo, Juanru Li, Yuanyuan Zhang, and Dawu Gu. 2019. "Security Analysis of Third-Party in-App Payment in Mobile Applications." *Journal of Information Security and Applications* 48 (October): 102358. https://doi.org/10.1016/j.jisa.2019.102358.

Zhao, Eric Yanfei, Masakazu Ishihara, P. Devereaux Jennings, and Michael Lounsbury. 2018. "Optimal Distinctiveness in the Console Video Game Industry: An Exemplar-Based Model of Proto-Category Evolution." *Organization Science* 29 (3): 492–510. https://doi.org/10.1287/orsc.2017.1194.