Designing Freemium and Free Trials For Digital Subscriptions

Ian N. Weaver[1], Vineet Kumar[2], and Sreelata Jonnalagedda[3]

[1,2]Yale School of Management

[3]Indian Institute of Management, Bangalore

Date: Spring 2023

Abstract

In digital product settings, freemium (with a permanently free plan) and free trial are among the most commonly used and effective strategies used by both start-up companies and established firms. The design of the free plan and trial is of primary importance, but there are few studies of its causal impact on metrics that businesses value. We conduct a field experiment in collaboration with a company that provides video-based learning for performing arts space to identify the impact of free content availability on both engagement and monetization outcomes. We find that the generosity of the free plan drives engagement higher, whereas conversion rates are inversely related. However, the monetization impact of the generosity demonstrates an inverted-U shaped relationship. We document the underlying mechanism for user behavior that drives these results. Finally, we undertake a content analysis, adapting recent advances in using raw unstructured video data to characterize rich content features so we can understand how content characteristics explain differential treatment effects.

1 Notes, 1st Aug 2023 - SJ

Storyline:

1. Generosity leads to long-term relationship (people who get more free minutes convert to 6 month plan)

2. Can we say anything about acquisition and retention? Acq we can because we observe all visitors. Retention may be tricky because we only have a month's worth of data.
3. Freemium more impactful than free-trial (aligns with Hema's paper)

Directions for exploration

1. Plotting engagement over time for each cell
2. Capture ideas of depth vs. variety (perhaps repeat views vs unique views)
3. What explains which plan people sign up for?
4. Does clumpy engagement predict conversion?
5. Engagement + Monetization Copula, Bayesian model.
6. Can we model live videos as an exogenous shock and evaluate effects on engagement/conversion?
7. repeat views/timing between views as DVs – what are the IVs?
8. Some IVs to consider: device + location of the user, weekend + weekday sign=up, other cust. vars
9. What is the value of personalized plan (similar to Hema paper)?

2 Things To Do

1. Gini coefficient – Preference for variety.
2. Customer Journey
3. Time to conversion
4. Live versus Library content.
5. Do they shift the kind of content consumed during and after free trial?
6. One aspect to add in the analysis is that not just conversion but monetization into different plans is important. Another aspect is that retention is also important, and we see higher retention rates when X plans are used. All this builds into greater CLV for the intermediate 7-minute plan compared to the other plans.
7. Can you do heterogeneous effects in terms of behavioral variables? what they have done etc.? See Ascarza Netzer paper.
8. Can we model plan choice? Cannot use the basic discrete choice model

3 Introduction

We investigate the tradeoffs in designing product plans for digital platforms, specifically in the freemium business model, and examine how plan design impact consumer choices and firm outcomes. In markets with digital products and services, we commonly see firms offering both a free product and a paid premium product, which is termed a freemium business model. For example, Spotify offers a free plan as well as a premium plan with additional features for a monthly fee. Similarly, companies with digital products across a wide range of industries like the New York Times, LinkedIn, Zoom, Dropbox and Google Docs also use this model. Offering a *perpetually available free plan (freemium)* makes financial sense when the cost structure includes relatively high fixed costs for product development and perhaps more importantly, low variable costs associated with usage. Other businesses do not offer a perpetually free plan, but rather offer a *time-limited free trial* of a paid product. Although such a strategy may on the surface appear similar, the underlying rationale is often quite different. Free trials are often used when consumers face uncertainty in terms of the value proposition, since the products are often experience goods (Nelson, 1970). Allowing users to try a sample (free trial) in such cases allows them to evaluate the product and determine fit and value before they purchase the product.

Freemium and free trial are the most commonly used and significant business models for digital products and services. In the app store, more than 90% of gaming apps use a freemium, and on Google Play, it was estimated that 98% of app revenue was driven by freemium apps (Kosner, 2015; Savov, 2014). Firms in digital product markets typically care about both engagement and conversion as metrics of short-run and long-run monetization. Companies often get valued and acquired for eyeballs, not just current revenue, e.g. Instagram was acquired by Facebook for over $1 billion in 2012 even though it had no revenues. Engagement has also been shown to be important for attracting venture funding, monetization important for short-term survival (Valliere and Peterson, 2004; Wiltbank, 2005). More broadly, the academic literature has examined the question of what makes engagement valuable from the perspective of both the firm and the investor (Holm and Günzel-Jensen, 2017; John et al., 2017). However, it is important to understand the drivers of monetization metrics like revenue and lifetime value separately as well (Datta et al., 2015).

We examine some of the primary questions related to the design of the perpetual free plan and time-limited free-trial in digital product markets. Firms in such markets typically care about both user engagement as well as monetization, and aim to design their plans to improve along these measures. However, there are often tradeoffs companies face here in designing the free plan.

Increasing the value of the free offering might be beneficial from an engagement viewpoint, but might cannibalize the purchase of the premium product, resulting in lower monetization effectiveness. Thus, we might expect a direct tradeoff between engagement and monetization. A free trial might seem less likely to result in such cannibalization, but its relative effectiveness compared to a perpetually free product is not well understood. Moreover, firms often distinguish between short-run and long-run outcomes, with varying emphasis depending on the goals of the company. The impact of these design choices across timeframes is therefore also important to understand in such cases.

With this framing in mind, we undertake a careful experimental study of the following questions related to design of the freemium and free-trial strategies over both the short and long run. We examine the impact of freemium plan design, or how the amount of (perpetual) free content impacts engagement and monetization outcomes. One goal is compare the relative effectiveness of freemium and free trial strategies. We also examine whether there is evidence of complementarities across freemium and free trial strategies, in the sense of whether they might be more effective when used in conjunction with each other.

Our research focus is on understanding the outcomes corresponding to different design choices for the free plan in a freemium business model. The outcomes we are interested in include short and long run measures of engagement, as well as monetization metrics of conversion, revenue and customer lifetime value. We vary the value proposition available to the free consumer in several ways, and characterize the individual-level and aggregate outcomes. First, we are interested in examining how varying the amount of content offered in the free plan impacts the above outcomes. Second, we evaluate whether a time limited (24 hours) free trial allowing access to all content impacts engagement and monetization. An important question arises whether the (permanent) free plan and the (temporary) free-trial serve complementary roles in achieving engagement and monetization, or whether they serve as substitute strategies from the firm's perspective. We also investigate the impact of permanently offering a full-length sample. Finally, we conduct a content-level analysis using the recently developed generic machine learning inference (Chernozhukov et al., 2020) to examine whether and how different types of content obtain *differentially higher engagement* under the experimental conditions.

Field Experiment Setting: We conduct a field experiment in collaboration with an online platform (name withheld) for dance enthusiasts with video lessons that enables learners to take dance lessons online. The platform offered 180+ classes across 30+ different dance forms from beginner to advanced level. The content on the platform comprised of live classes or library (recorded) classes. Live classes are typically lower quality and offered for free at a pre-specified time on specific days, and are aimed at a wide audience. The library classes included both higher quality studio classes and recorded versions of live classes (live-recorded). The studio classes were categorized into either moves, choreographies, or courses based on content. The firm used a freemium business model where all live classes and a part of all the library content were free for all users. To access library classes beyond the minutes allowed in the free plan, the consumers had to subscribe to one of the following three premium plans at different price levels: 7 days (¤199), 30 days (¤499) and 180 days (¤2499).[1] The studio classes available to premium users involved high quality production with significant involvement by the company, including production resources and monetary support. Users were typically acquired through digital and social media, specifically Facebook/Instagram and Google/Youtube, which served as the primary marketing channels.

There are a few studies that use observational data and modeling to get to similar questions, but a field experiment would offer an ideal approach to isolate the causal impact of different plans. The experiment also allows us to investigate alternative plan designs that go well beyond the current design. We design a field experiment using classic Randomized Controlled Trial (RCT) principles to vary the design of the free plan along 4x3 aspects, and run it over the course of 15 weeks. Randomization is done at the user-level when a user first signs up to access the product, and the user is permanently allocated to a condition. There are many ways to design free plans. The firm operates in a performing arts space, and perhaps unlike passive video consumption experiences like Instagram or Youtube, the value comes from learning and practicing dance. Thus, we might expect users to derive value from repeated viewing of the same content. We design free plans that feature different amounts of permanently free video content (number of minutes) available for each studio video content. We also include a limited-time (24 hour) trial that allows the user to access all content, so they can obtain a complete and comprehensive experience of the entire content available to them if they upgraded to the premium plan. In addition, we have a condition where one complete video class was provided to the free users so that they could experience a complete lesson.

The experimental conditions (cells) were created to apply from the time an individual user landed on the site. Each user upon sign up (registration) was randomized into one of the experimental cells, and stayed in that condition permanently. Thus, the user did not even know of the existence of other experimental conditions except their own, which alleviates consumer confusion. Any user who upgraded to a premium plan would have unrestricted and complete access to all content and

[1]Currency units have been disguised to avoid identification.

not be impacted by the design and constraints of the free plan.

We characterize business outcomes at two distinct and unique levels of analyses: user-level and content-level. The user-level analysis informs us how the individual user journey is different across the different experimental conditions (free plan + free trial designs). In contrast, the content-level analysis looks at engagement for each content video across the conditions. In the *user-level analysis*, our focus in on metrics of engagement and monetization, aggregated across the users, corresponding to each experimental condition (cell). We evaluate 3 different metrics of engagement, including total view time, repeat views and variety of content consumed. We find that as the amount of permanently free content is increased, engagement across each of the metrics uniformly increases. However, we find that the proportion of consumers upgrading to a premium plan decreases as the free plan becomes more generous with permanent content. This is suggestive of the idea that users may be getting sufficient value from a more generous free plan, which leads to a drop in the conversion rate to premium plans. However, this is not the complete story from a monetization perspective. The company offers premium plans of varying duration, and we find that when the free plan is *somewhat more generous*, users are more likely to upgrade to a premium plan with a longer duration, which brings in greater and more predictable revenue to the company. The revenue seems to have an inverted-U shaped relationship with free plan generosity, suggesting that an intermediate amount works best from a monetization perspective. When we account for retention likelihood in computing customer lifetime value for long-run monetization, we find results similar to, and stronger than revenue, in suggesting free plans of intermediate generosity work best. In examining the engagement and conversion results for a time-limited (24 hours) free trial, we find that although the free trial drives greater engagement during the 24 hour interval where the user can access all premium content, there is no long run impact that follows this greater engagement, whether in terms of engagement or monetization metrics.

In examining whether freemium and free trial work better together, i.e. as complements, we find that the relative improvement from a free trial is actually lower when a permanently free (freemium) plan is offered. Thus, we find no evidence for complementarity between these two commonly used business strategies. Overall, we find that whereas free trial might drive short run engagement, free plan design with freemium is more impactful than free trial in driving long run outcomes in terms of both engagement and monetization.

Next, we conduct a *content-level analysis*, which would be helpful in helping a firm *choose a more appropriate plan based on the kind of content present in its library*. For the content-level analysis, there were several differences between the characteristics of videos associated with largest increase in watch time as users were given access to more content (i.e. a more generous free plan). Most notably, additional viewing went disproportionately to videos where they instructor was more positive both in their speech and in facial emotions. Users also preferred videos that were more demanding and challenging (i.e. longer *courses* rather than the shorter *moves* videos), and a more collaborative linguistic teaching style. Additionally, the location of instructor relative to the viewer was also a factor in driving differential engagement.

Although there are several related studies on freemium and free trial, field experiments that causally examine how plan design leads to different business outcomes across both short and long run are relatively few. We examine this in a subscription setting rather than individual products available for purchase. Our field experimental design tests both a 24-hour (time-limited) free trial and freemium (part of each content is available for free) model. We also study the potential complementarities between the two strategies. From a user-level perspective, we examine how value on a number of metrics is impacted by plan design:- engagement (depth, variety and repeat viewing), conversion and separately monetization. We find that conversion and monetization can yield different results as plan duration varies.

Further, taking a content-level perspective, using state-of-the-art tools for heterogeneous treatment effect (HTE) using machine learning, we examine how different types of content receive differential engagement (consumption) across plan designs. We characterize video content, which in raw form is very high-dimensional unstructured data into a number of interpretable characteristics / features, and evaluate how these impact both engagement overall and variation of engagement with free plan designs. The content-level perspective is novel and can inform several platforms, which have content of various types, similar to our setting. Firms would be able to better evaluate *which plan works best based on characterization of their unique content using raw unstructured data*. For instance, Hulu uses a freemium strategy and offers both movies and Tv shows in equal proportion, whereas Netflix, which uses a free trial (but no perpetual free product) is more skewed towards movies and has a larger library.

Our empirical setting involves a free plan and a set of premium plans. Our approach is causal since we have exogenous and random allocation of users to different designs of the *free plans*. We also do not have any differences in promotion or messaging across the experimental conditions that might confound any results. First, the results we obtain are limited by the number of possible plan designs we are able to test. They are arguably applicable to other plans that lie within the range we test but operational feasibility (and sample size considerations) would limit the number of possible conditions. Second, we are able to follow users for the entire 15 week duration of the experiment, but not beyond it. Depending on when a user signed up, we viewed their behavior between 1 and 15 weeks. The data shows that much of the new user behavior occurs within the first week, however, there may be longer run effects that we are not capturing in our study. Third, our free trial is conducted at the time of registration. Some companies, especially those who use only free trials, operate similarly, but others especially firms using freemium allow users to choose a free trial at a later time. Our study design does not address this interesting situation.

4 Literature

Freemium and Free Trial: Pricing is a first-order issue for most firms, and free or zero-price products are an especially important aspect of pricing in digital product markets. Consumers in experimental settings are shown to treat zero prices as distinct from any non-zero price, implying demand discontinuity effects at price zero (Ariely et al., 2018; Shampanier et al., 2007). Other studies demonstrate conditions under which a a permanently free product (freemium) can be a rational strategy for firms (Shi et al., 2019), and that product-line extensions in a freemium setting can overcome the zero-price effect (Gu et al., 2018). In addition to the effect of 'free' product on demand, our work also identifies the impact of the *amount of free content* on demand in a field experimental setting. In this context, prior research examined how sampling strategy under quality uncertainty (Halbheer et al., 2014) and peak and low demand periods (Lambrecht and Misra, 2017). Ascarza et al. (2020) also examine how much to give free, in large scale field experiment, in a mobile game context. In addition to studying the amount of free content, we also examine the duration and quality of free content. All the above-mentioned papers examine the impact of quantity of free content on the trade-off between advertisement and subscription revenues.

There is a significant distinction between the freemium and free trial strategies. A free trial, by definition, is time-limited whereas Freemium involves a *perpetually free version* that is always available to all users. In a free trial model, after the trial period expires, users have the choice between purchasing the product or not, whereas in freemium, they always have the option to use the free version (Kumar, 2014). The underlying purpose for using these strategies is also quite often different. With free trial, it primarily helps users who have a lot of uncertainty about the functionality, quality or experience with a product to try and resolve that uncertainty within a limited time. With freemium, the free version provides a value proposition that serves to continually engage the user, and provides perpetual opportunities for the user to upgrade to the premium version and obtain greater value.

There are a few closely related papers that also use a field experimental setting in studying different empirical settings and complementary questions. First, a study focusing on experimentally varying the length of the free trial in a software as a service setting (Yoganarasimhan et al., 2022) finds that shorter trials lead to higher conversion rates, and that longer trials rarely do better, even when choosing personalized lengths. We view this study as complementary to ours since they do not examine a freemium business model (with perpetually free products), and we explore freemium and a free trial, but we do not vary the trial length like they do. Second, a study focusing on book content also examines a free sampling strategy (Li et al., 2019). Specifically, they examine the impact of sample quality and product prices on premium conversions. There are several connections between our studies, but several distinctions too. First, their focus is more on the premium product design, whereas ours is on the free product design. They look at freemium design for a product market, where each product (book) is purchased separately, whereas we examine choices in a subscription setting in which a number of videos are included and users watch many of them. This also affects the level of experimentation, which is at the product versus platform level in our case. Our work combines both free trial and freemium design in our experimental design. An important distinction is that our study examines both short run and long run engagement, revenue and CLV metrics, more relevant to subscription settings, whereas their focus is on revenue. Finally, none of the related studies do a content level analysis characterizing and examining different features of the content and the corresponding impact.

In terms of what business outcomes are important, clearly monetization like revenue and lifetime value are important. In addition, customer engagement is also a highly valued metric in digital product settings. Although customer engagement has significant bearing on long-term survival and profitability of a platform, there is limited experimental research on this front. A related study examines whether engagement (likes) leads to other metrics closer to monetization in social media (John et al., 2017). Two studies by Bapna et al. (2016) and Barnes and Kirshner (2021) are the closest to our work that examine customer engagement in a freemium setting. Both these studies identify the impact of premium users on free user engagement and conversions. While the emphasis in the above studies is social engagement in digital communities as the primary variable of interest, our work focuses on user level engagement as indicators of learning and longevity on a platform targeting skill development.

4.1 Audiovisual Analytics

The proliferation and growing impact of unstructured multimedia content on social media (Li and Xie, 2020) is drawing the attention of marketing academics and practitioners alike. Image, video, and audio analytics are evolving into a promising platform to understand consumer behavior online and improve company's decision making. There is substantial research on methods for visual/audio data analysis; Barnes and Kirshner (2021) for example present a tool to efficiently extract 109 video-based variables, that may be of interest for business research. Balkan and Kholod (2015) make a case for intelligent video analytics as a way to enhance decision-making and reducing cognitive biases inherent in quantitative primary data. Chakraborty et al. (2022) combine AI prediction with human judgments using a Bayesian approach, to compare predictive performance of the hybrid AI-human model to a pure human panel benchmark, and find that AI can significantly replace human effort in screening but may have a limited role in selection. Zhang et al. (2021) use verified and unverified images from Airbnb property data to quantify 12 human-interpretable image attributes that have a bearing on consumer perception, consideration, and choice. Hu and Ma (2021) process full pitch videos of start-ups using machine learning algorithms, to quantify persuasion in visual, vocal, and verbal dimensions. Cao et al. (2022) study how image-text congruence affects consumer preference using a deep learning model with data from a reading platform.

Focusing on marketing outcomes, a study examining difference between storytelling and selling video ads shows the former lead to higher performance (Coker et al., 2017). Others have studied the impact of visual aesthetics on consumer engagement and video popularity in online learning (Zhou et al., 2021).

Using social media posts in two different categories, Barnes and Kirshner (2021) use deep learning to identify characteristics of host-faces for a Airbnb that impact market outcomes. Cheng and Zhang (2022) extract a rich set of features of user-generated YouTube videos created by influencers in the beauty and style category, and demonstrate the difference in impact of sponsored versus self-created content on consumer behavior. Similarly, Boughanmi and Ansari (2021) develop a novel machine learning framework that combines multimedia data to predict the success of musical albums and playlists, where as Yu et al. (2022) combine video analytics, machine learning, and econometric analysis, on click-stream data from a leading e-commerce platform to understand the effectiveness of outstream video ads, a fast growing category of advertisements in the digital world. In this context, our study relates content video characteristics to effectiveness of different plans.

4.2 Heterogeneous Treatment Effects

Heterogeneous treatment effects are very useful in many studies as the average treatment effect may hide vast differences between how individuals are affected. In general, especially without a pre-analysis plan, it is insufficient to differentiate between noise and true effects by separately conditioning on many variables. One approach typically used in marketing is to classify individuals into discrete segments (Kamakura and Russell, 1989; Sriram et al., 2010). However, a downside of this approach is that it just an approximation of heterogeneity and may be inconsistent with the notion that everyone is different. An alternative Bayesian approach that has a continuous form that fit the data better in various studies was provided by Allenby and Rossi (1998). However, rather than segment individuals, in our case it would be of more interest to see how the treatment effects differ from the numerous variables we have available. In recent years, machine learning has proved useful in understanding the intricacies of heterogeneity for large sets of variables. Famously, Wager and Athey (2018) introduced *causal forests*, which constructs asymptotic confidence intervals for the true treatment effect that are centered at the random forest estimates (Breiman, 2001), allowing for an understanding of heterogeneity based on splits in the trees. This idea of combining machine learning and causal inference for heterogeneity was extended upon with *generic machine learning inference* (Chernozhukov et al., 2020), which is used in this paper. Importantly, it allows for any machine learning method to be used (such as elastic net, neural networks, ensembles, etc.) and claims to work better with higher dimensional settings.

5 Field Experiment Setting

5.1 Company and Market Background

Our experiment was conducted in collaboration with a digital platform that offered online dance lessons by a number of instructors across various dance forms. At the time of the experiment, the company had a base of over 100,000 users from over 100 countries, and over 200 instructors across 40 different dance forms, ranging from beginner to advanced levels. The platform had 390+ videos included both library or live streamed classes with synchronous audience participation. The studio classes were were typically high quality productions recorded in a dance studio and involved significant resource investments from the company. In contrast, live classes were generally low quality videos, as they are directly streamed from instructors' dance space.

The company's business model relied on a freemium design, where all live classes and some portion of the recorded content were free for all users. Other content on the platform was only available for premium subscribers. The freemium pricing plan had a three-tier structure where users paid ¤199 / ¤499 / ¤2499 in order to access premium content for 7 days / 30 days / 180 days respectively. The platform relied on digital and social media marketing to reach their target segment of dance enthusiasts across the world, and used predominantly two channels namely Facebook/Instagram and Google/Youtube. While Instagram helped in brand building and engagement, the YouTube channel served as the platform's primary source for new users. At the time of the platform's YouTube videos had an aggregate viewership of millions of users, and their Instagram handle had tens of thousands of followers.

6 Data

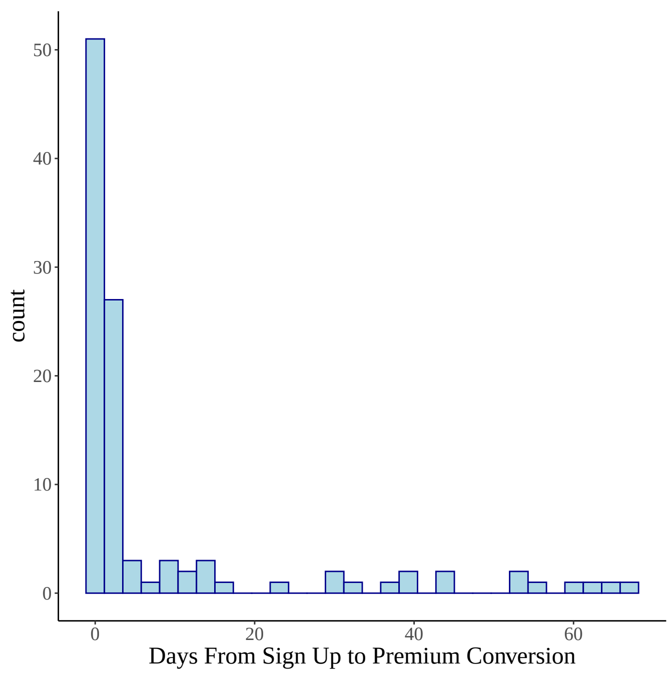
The company uses an analytics backend to capture the journey of the user as they navigate the various elements of the site. The data captured include engagement metrics like the amount of time spent watching each video, replays of segments of the video, and user directed navigation across videos. We are therefore able to create several engagement metrics based on each user's individual journey. The user level of metrics also aggregated at the level of content videos to obtain engagement metrics for each content video. These metrics are then used in our content level analysis to examine how are the class characteristics, which include information about the class as well as the raw video data, interact with different experimental conditions (design of the free plan). Thus, we collect a number of structured data variables, including those which capture the business outcomes of interest, and a variety of data from unstructured video content. We collect data from N = 4960 users over the course of 15 weeks over the time period of December 2021 - March 2022.

6.1 Structured Data: User Behavior

During the field experiment, we obtain and use multiple aspects of user behavior as outcome variables of interest. These variables include engagement metrics, conversion, and revenue. We also obtain a derivative measure of customer lifetime value at a user level and then aggregate it to the experimental condition. *Engagement* is multi-dimensional and not just captured by a single metric. This construct thus includes variables characterizing which the user's journey through the site as they watch different video classes.Thus, for each user, the *total engagement* is the number of minutes a user watched of each video. Engagement components then include *variety* - the number of unique videos a user watched, *depth* - the average duration of time each video is watched by the user, and *repeat viewing* measures the number of different sessions.[2] Variety measures how many videos a user watches which is indicative of how much they engaged with the entire breadth portfolio of videos. If a user only comes to the site for a couple videos, they might be less likely to re-subscribe once they have completed viewing. Depth measures how long a user watches each video. Repeat viewing is important because premium subscriptions are more likely for users who make learning dance part of their routine. Mastery of a dance move and choreography requires involvement and practice, so users who are spending more time and repeatedly viewing a specific dance video are likely to be trying to perform the dance themselves and thus gaining utility from the product. This aspect distinguishes our empirical setting from a passive viewing medium like Netflix.

The summary statistics show that half the users who convert do so within their first day on the site (Figure 1). Another 25% convert the second day, with the other 25% of converts happening throughout the length of the experiment. The amount of engagement (number of free minutes watched) went up as the number of free minutes increased (Figure 2) and when the user was given a 24 hour free trial (Figure 3). Revenue had an inverted-U shape (Figure 4).

Figure 1: Days to Conversion



Last, demographic variables collected by the firm were sparse, but they did include the IP location of the user as well as their referring domain. Users had the option to not allow tracking

[2] A new session begins when a user has been inactive on the site for at least one hour.

Figure 2: Engagement by Number of Free Minutes

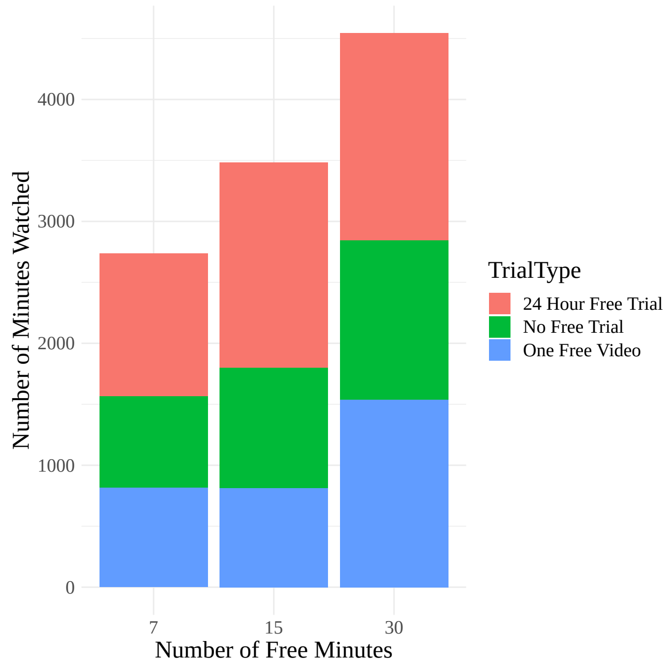
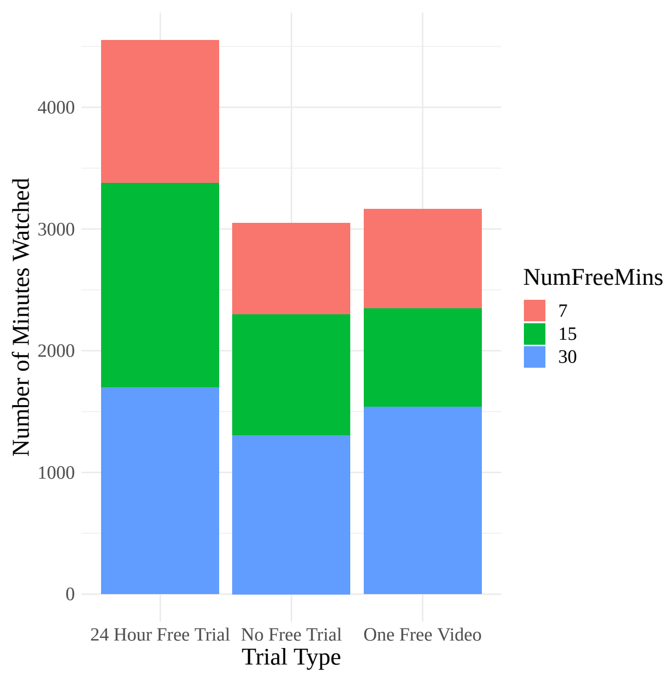
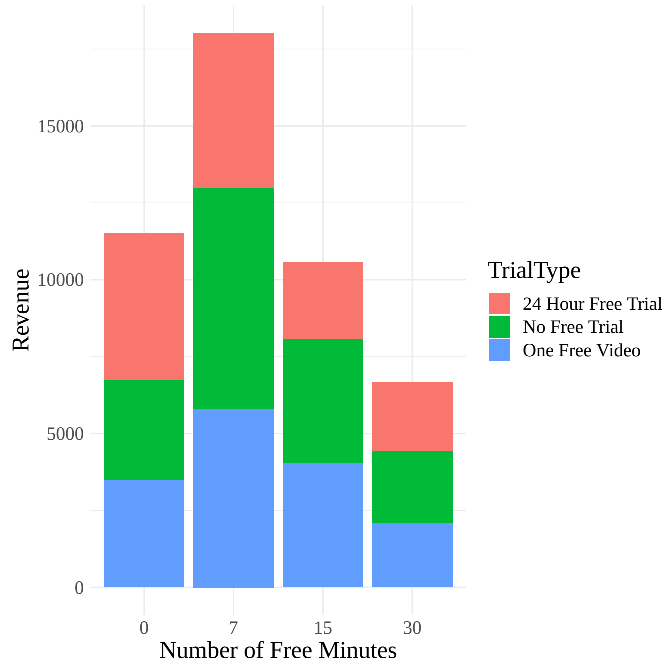


Figure 3: Engagement by Free Trial Type



by adjusting their cookies, but only 4% of users opted to not be tracked. Of the users that could be tracked, almost all came domestically (95.4%). The most common referring domain was Google which was the source of 46.0% of users, YouTube with 25.4% and direct landings accounting for 20.6% of users. Only 8.0% of tracked users came from another source.

Figure 4: Revenue by Free Minutes



The engagement and conversion for users in the experiment are shown in Table 1. The most striking difference is that users who could not be tracked were highly correlated with upgrading to premium. One possible explanation is that users who spend more time on the site are more likely to alter their cookies settings. A second finding, is that international users have a higher conversion rate compared to domestic users.

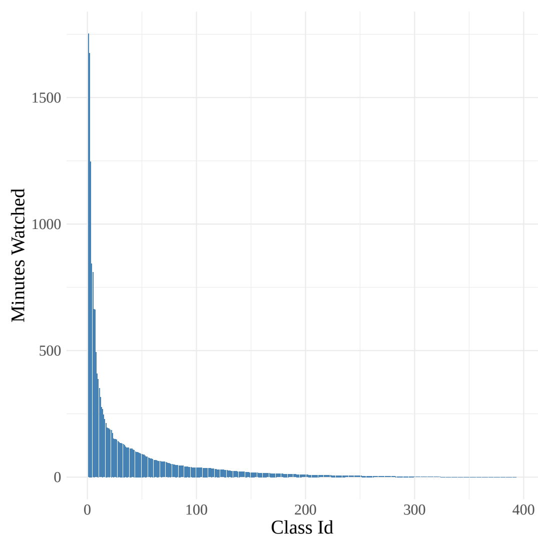
Table 1: Engagement and Conversion by Referring Domain and Country

	Minutes Watched Per User (Free Users)	Minutes Watched Per User (Premium Users)	Conversion Rate
Referring Domain			
Google	2.8	72.2	2.2%
YouTube	3.3	22.4	0.8%
Direct	3.1	68.6	1.2%
Other	0.9	39.8	1.4%
Unknown	4.9	64.2	16.0%
Country			
Domestic	2.8	68.0	1.4%
International	4.6	28.6	4.3%
Unknown	4.9	64.2	16.0%

6.2 Unstructured Data: Content

Since the content that makes the site valuable to users is dance instructional videos, we would like to characterize and identify features encoded in these videos. In the course of the experiment, there were 391 dance videos on the site covering 42 different dance styles taught by 173 different instructors, providing a high degree of variation. There was also significant heterogeneity in the popularity of videos (Figure 5) with 70 videos (17.9%) receiving no views and the top 10 videos receiving 43.4% of all minutes watched. Since we have access to the video content from the platform, we aim to characterize each video along a number of dimensions that might be important in terms of treatment effects, i.e. what types of videos receive differentially higher engagement when the free plan is more generous.

Figure 5: Minutes Watched per Video by Experiment Users



Following recent literature in using video data (Chakraborty et al., 2022; Cheng and Zhang, 2022; Hu and Ma, 2021; Zhou et al., 2021), we decompose the information embedded in the videos into three components: text transcript, audio, and visuals. Interpretable metrics are then extracted from each component and calculated for the duration of the video. We augment this information with basic video properties provided by the firm.

Basic Video Properties. Each video has been classified into one of four different types by the company as described in Table 2 and is either pre-recorded in a dance studio or streamed live by the instructor during a live session and later uploaded to the site.[3] Videos feature an introduction

[3]To avoid confusion, when referring to a particular type of video we will italicize and capitalize the word (eg: *Choreographies*) to distinguish it from the literal use of the word.

(a brief explanation of the class and other logistics - involves no dancing) and the main content (all other parts of the video where instruction is being provided on the core content of the class). We obtain the length and proportion of the video that is dedicated to the introduction. Additionally, each video has a breakdown with corresponding timestamps provided by the site allowing the videos to be classified into sections. Other properties include the instructor, dance style, and class level (e.g. advanced), which are structured data.

Table 2: Overview of Video Types

Type	Description	Number of Videos	Min Length (min)	Max Length (min)	Mean Length (min)
Moves	Choreography for one dance move	47	2	7	4
Choreographies	Choreography for one song	47	23	130	58
Foundations	Basics of a particular dance style	283	34	106	68
Courses	In depth series of videos teaching one dance style	14	72	277	174

Text Transcript. We next examine the speech of the instructor in the dance video, specifically the word content. For each video we extracted the speech from the audio data using the speech-to-text conversion API provided by Google Cloud.[4] The ML algorithm provides a transcript that includes the a list of the words said, their corresponding time stamps (onset, offset, and duration) as well as punctuation (such as a period or a question mark). To improve the accuracy of the results, we also input some additional context related words to the Google speech-to-text algorithm. This customization boosts the chance of the algorithm successfully recognizing some domain specific such as "choreography" as well as some popular dance types including "waacking", "locking", "popping", and "krump".

From text transcript, as with any unstructured data, there are a number of potential variables that can be created. We obtain three metrics of basic speech pattern that have commonly been used in the literature: the speech rate (number of words said per second), average sentence length (number of words per sentence), and the percentage of complex words used (words longer than 6 letters). Speech rate contains the rate at which information is being relayed to the viewer, and past research (Guo et al., 2014) has found that faster speech tends to correlate with higher user engagement. Average sentence length and percentage of complex words pertain to linguistic complexity (Pogacar et al., 2018) as ease of understanding has been associated with simpler words and sentence structure (Laham et al., 2012). Other text features include various types of linguistic

[4]See https://cloud.google.com/speech-to-text for access.

styles, which we obtain by extending the idea of LIWC dictionaries (Pennebaker et al., 2001; Tausczik and Pennebaker, 2010) with words specific to our context. In the learning context, an linguistic style is how *individualistic* and *collaborative* an instructor speaks. *Individualistic* and *collaborative* speech styles are operationalized the same way as in Chakraborty et al. (2022) by counting the rate the instructor uses first person singular pronouns ("I, "me", "my" "mine", "myself") and first person collective pronouns ("we", "us", "our", "ourselves"). Other linguistic styles which could be of importance in teaching are *politeness* (Hobjilă, 2012) and *Asking Questions* (Jarvis, 1985). *Politeness* is measured as the rate the instructor uses polite words ("thanks", "thank you", "please"), while *Asking Questions* is measured as the rate of sentences spoken by the instructor that were a question. Lastly, we measured the text sentiment using the NRC Emotion lexicon (Mohammad and Turney, 2010, 2013). This lexicon allows for text to be broken into eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive).[5] Each segment of text is partitioned among these eight emotions such that the sum adds to 1; the same is true of two sentiments.

Audio. The audio for each video was processed using pyAudioAnalysis (Giannakopoulos, 2015) which provides the mean and variance of 34 selected features in one second intervals. These includes features of interest such as the zero crossing rate (ZCR), energy and spectral centroid of the audio signal. ZCR is known to be important in characterizing percussion instruments and variability more broadly (Hwang et al., 2021).[6] Energy represents loudness or amount of a signal there is at a given time (Gerhard, 2003). The spectral centroid is a measure of the average frequency of the signal (Gerhard, 2003) and represents the impression of "brightness" of the sound (Schubert and Wolfe, 2006). We compute and use the mean of each of these three metrics were calculated across each video.

Visuals. For dance instruction, the video content is of critical importance. While the audio plays a supportive role in learning, in understanding what the dance move involves and how to perform it incuding the movement of body parts, the visuals are the primary source of learning and instruction.

To characterize videos, typically images are sampled (we obtain 2 samples per second). We then used two different approaches for characterizing image content. First, we estimate emotion

[5]The analysis was conducted using the *syuzhet* package in R (Jockers, 2015).

[6]The zero crossing rate is the rate at which the audio signal amplitude changes from the positive spectrum to negative (Subramanian et al., 2004). In theory, for monophonic audio this would correspond with the fundamental frequency (pitch) as the ZCR will cross over twice per cycle (Gerhard, 2003). However, for spectrally rich samples it is possible for the ZCR to cross more than twice per cycle (Roads, 1996).

Table 3: Variable Descriptions

Type	Feature	Description
Basic Video Properties	Class Level	Difficulty of the class as decided by the firm (either 'beginner', 'intermediate', 'advanced' or 'open')
	Style ID	Id for the dance style of the class
	Instructor ID	ID for the lead instructor of the class
	Class Type	One of the four types of classes: 'Choreographies', 'Foundations', 'Moves', and 'Courses'
	Pre-Recorded	Binary indicator for whether the class was pre-recorded or a livestream later uploaded
	Intro Length	Number of seconds the instructor spends on the introduction
	Class Length	Total length of the class in seconds
Text Features	Words Per Sentence	Measures how long each sentence is
	Complex Words	Percentage of words that are longer than 6 letters
	Speech Rate	The number of words spoken per second
	Individualistic	Rate which words "I", "me", "my", "mine", and "myself" are used
	Collaborative	Rate which words "we", "us" "our", and "ourselves" are used
	Politeness	Rate which words "thank you", "thanks," and "please" are used
	Asking Questions	Proportion of sentences that are a question
	Text Emotion	Using the NRC emotion lexicon, the emotion of the text is partitioned into 8 parts (anger, joy, anticipation, trust, sadness, disgust, fear, and surprise) such that the sum is 1
	Text Sentiment	Using NRC emotion lexicon, text is partitioned into negative and positive such that the sum is 1
Audio Features	Energy	The spoken energy of the instructor
	Spectral Centroid	Measures the brightness of the voice of the instructor
	Zero Crossing-Rate	Measures the number of times the voice signal goes from positive to negative (voice modulation)
Visual Features	Facial Emotion	Using the Microsoft Face API, the emotion of a face is partitioned into 8 parts (anger, contempt, happiness, neutral, sadness, disgust, fear, surprise) such that the sum is 1
	Closeness	A measure of how much of the frame the head of an instructor is occupying
	Head Movement	Distance traveled by the nose from frame t to frame $t+1$ scaled by the closeness

and physical characteristics of the instructor using the Microsoft Face API.[7] Second, the motion of the instructor is measured using OpenPose (Cao et al., 2017) which extracts the pixel locations of various body parts.

The emotion expressed through the instructor's facial expression can have a large impact on student engagement. Prior research has shown that the primary way emotions are communicated is

[7]See https://azure.microsoft.com/en-us/services/cognitive-services/face.

through facial expressions (Ekman and Oster, 1979) and that viewers are highly responsive to such facial changes (Dimberg et al., 2000). However, it is unclear how these emotions affect learning. In classroom settings, previous studies have found that anger can disrupt student behavior (Lewis, 2001), though others have shown that it can lead to better performance (Goldenberg, 1989). It is unclear, however, whether and how these results translate to optional self-motivated learning in a digital setting.

We applied the Microsoft Face API to one frame per second for each video. The emotion recognition model partitions the emotion of each detected face into eight categories: anger, contempt, disgust, fear, happiness, neutral, sadness and surprise. For each face, every emotion is a given a score between 0 and 1, corresponding to the proportion of the emotion that belongs to each category; overall they must sum to 1. We then averaged over the entire video to obtain an overall score for the each emotion for the instructor. Only the facial expressions for the lead instructor were used with those detected for the backup dancers (if any) being discarded.

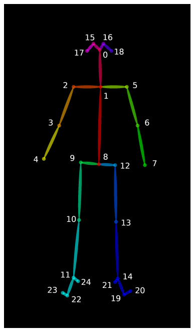
The level of motion has been found to be an important video characteristic across multiple studies (Chakraborty et al., 2022; Cheng and Zhang, 2022; Yang et al., 2021; Zhou et al., 2021). In a product like dance instruction, the importance is even more so. There are several ways of capturing motion. One approach is capture the entire motion by measuring the rate at which pixels change between frames. Another approach is to only track the motion of the focal person by using OpenPose Cao et al. (2017) which calculates the pixel locations of various body parts (nose, ears, neck, shoulders, hips, elbows, wrists, knees, and ankles).

An example from the OpenPose documentation[8], is shown in Figure 6. These pixel locations can then be used to calculate the motion of each of the body parts in the video. OpenPose also has the ability to distinguish between people, allowing us to only track the main instructor (making comparison between videos with and without backup dancers easier).

Following Chakraborty et al. (2022), head movement speed is defined as the pixel distance traveled by the nose between frame t and frame $t+1$. Frames were taken at one half second intervals, since we observed that significant movement can happen if a lower frequency is used. Similar to the calculation of the emotion metrics, only the motion of the lead instructor was used with motion metrics from backup dancers (if any) being excluded. Results were scaled to a 1920×1080 pixel resolution to allow for comparisons between videos of different qualities.

[8]See `https://cmu-perceptual-computing-lab.github.io/openpose/web/html/doc/md_doc_02_output.html`

Figure 6: Body parts detected by OpenPose



Head movement speed encompasses two different aspects.[9] Instructors standing closer to the camera will have their heads take up more of the frame and as a result it requires a smaller movement for their nose to move x pixels than someone standing further away from the camera. To parse out these two effects, we created a measure of closeness to the camera and then scaled the head movement by the closeness.

To create a measure for closeness, we used the average of the distance between the base of the neck and the left and right ears of the focal person. Using Figure 6 as reference, this would be the average of the distance between points 1 and 18 and points 1 and 17. Closeness is then defined as the ratio of this distance divided by the length of the x-axis on the video frame. For example, a value of 0.1 would mean the distance between the base of the neck and ears is 1/10 of that of the x-axis of the video frame. Having created this metric, head movement can be scaled to measure how much the focal person's nose is moving between frames in relation to their own body, rather than just pixel distance.[10]

Summary Statistics. The summary statistics for all the video features can be found in Table 4. For each video, the average of each metric is calculated across the length of the video. We use the first 7 minutes of each video[11] for comparison, as this amount is visible to all users except those in the 0 minute condition. Viewers are also most likely to start at the beginning, meaning the first few minutes is crucial to their decision to continue the video. Five videos were of low quality so that metrics could not be obtained for them, limiting our sample to 386 videos.

[9] This was not an issue in Chakraborty et al. (2022) as all sales videos had the camera and participants in the same place.

[10] There are limitations to this parsing of the two effects. First, bodies come in all shapes and sizes so the distance between the base of the neck and the ears may differ between people even if they have the same proximity to the camera. Second, head rotations can cause certain features to appear closer together as a video is viewed in 2-d. While we believe that taking the average of the distance between the two ears and the neck is fairly robust to head rotations, it remains to be tested.

[11] Or the entire video for the few videos that were less than 7 minutes

Table 4: Descriptive Statistics of Video Features (First 7 Minutes)

Type	Statistic	Mean	St. Dev.	Min	Max
Text Features	Mean Sentence Length	8.688	2.477	2.500	20.880
	Speaking Rate	1.517	0.794	0.081	3.564
	% of Big Words	0.089	0.035	0.011	0.196
	Individualistic	0.032	0.021	0.000	0.130
	Collaborative	0.016	0.011	0.000	0.054
	Politeness	0.001	0.003	0.000	0.025
	Question Rate	0.054	0.047	0.000	0.474
	Text Anger	0.057	0.062	0.000	0.484
	Text Anticipation	0.250	0.114	0.000	1.000
	Text Disgust	0.027	0.030	0.000	0.250
	Text Fear	0.070	0.057	0.000	0.464
	Text Joy	0.199	0.076	0.000	0.429
	Text Sadness	0.081	0.055	0.000	0.333
	Text Surprise	0.081	0.048	0.000	0.241
	Text Trust	0.236	0.086	0.000	0.500
	Text Negative	0.238	0.148	0.000	1.000
	Text Positive	0.762	0.148	0.000	1.000
Visual Features	Visual Anger	0.004	0.008	0.000	0.069
	Visual Contempt	0.005	0.007	0.000	0.051
	Visual Disgust	0.001	0.002	0.000	0.020
	Visual Fear	0.001	0.002	0.000	0.024
	Visual Happiness	0.204	0.183	0.000	0.894
	Visual Neutral	0.727	0.194	0.097	0.998
	Visual Sadness	0.009	0.017	0.000	0.134
	Visual Surprise	0.050	0.058	0.000	0.397
	Head Movement	0.320	0.162	0.021	0.952
	Closeness	0.087	0.043	0.026	0.310
Audio Features	ZCR	0.023	0.023	0.001	0.132
	Energy	3.039	0.067	2.831	3.202
	Spectral Centroid	0.150	0.010	0.128	0.185
Basic Video Properties	Choreographies	0.720	0.449	0	1
	Courses	0.036	0.187	0	1
	Foundations	0.122	0.327	0	1
	Moves	0.122	0.327	0	1
	Intro as Proportion of Video	0.249	0.264	0.010	1.000

7 User Behavior Analysis

Given the field experimental setting where we exogenously locate incoming users randomly across conditions, reuse a regression based approach analyze user level outcomes like engagement metrics, conversion rate, revenue and customer lifetime value. Our analysis also distinguishes between short run and long run outcomes because we would like to understand the temporal impact of the experimental interventions.

We also conduct placebo tests to evaluate the impact of the experimental condition where we might expect no effect. For instance, since the experimental conditions vary the design of the free plan, we might expect that the live content that is constantly available across all of the conditions would have the same engagement levels across all the conditions. Similarly, we might expect that after users upgrade to the premium plan they are not impacted by the restrictions imposed under the free plan.

The randomized procedure for assigning users, allows the empirical analysis to be straightforward. For user i assigned to testing cell k, the variable of interest y_i can be modeled as:

$$y_i = \alpha + \beta_1 NumFreeMins_i + \beta_2 FreeTrialType_i + \beta_3 X_i + \epsilon_i \tag{1}$$

In this specification, the outcome variable y_i can refer to different metrics being evaluated (engagement, conversions, revenue or customer lifetime value). The explanatory variables include $NumFreeMins_i$ and $FreeTrialType_i$, the two dimensions for the experimental condition the user was placed, and X_i are some demographic variables corresponding to user i including *Country* and *ReferringDomain*.[12] More precisely, the variables are described as follows:

The first set of regressions, explores how engagement changes as the free plan was changed (Table 6). Both the number of free minutes and the free trial type were set as categorical variables. Recall that we have a 4×3 experimental design, but our analysis here focuses on specifying *permanent free plan minutes* separately from the presence or absence of *free trial*. Therefore, we define both a row (out of 4) and a column (out of 3) to be the control in this analysis. Our controls corresponded to the "no free trial" condition and the "free minutes equal to 7" condition. We find that engagement went up among free users as the amount of permanent free content available increases. The increase was present not just for the total engagement metric, across all operationalizations of the engagement variable, from variety to depth and repeat viewing behavior. This is also true when we compare the 24-hour free trial relative to the no free trial. The key takeaway is that engagement is positively related to the generosity of the free plan, and also holds for the free trial. In terms of the magnitude of the effect size, we find that they were quite large (for example, a change from 7 minutes to 30 minutes led to a 79.3% increase in minutes watched among free users).

[12]The data provided by the company does not include demographic variables that are commonly used in such analysis.

Table 5: Description of Variables

Variable	Description
y_i	Dependent variable that either an engagement metric or a monetization metric
Engagement Metrics:	$TotalEngagement_i$: total number of minutes watched by user i $Variety_i$: total number of unique videos watched by user i $Depth_i$: average number of minutes watched per video by user i $RepeatViewing_i$: number of unique viewing sessions
Monetization Metrics:	$Conversion_i$: binary variable whether user i ever converted to premium during the experiment $Revenue_i$: amount of money made from user i during the experiment CLV_i: projected life time customer value of user i given their behavior during the experiment
$NumFreeMins_i$	Categorical variable for the number of free minutes available at the state of each video for user i
$FreeTrialType_i$	Categorical variable indicating the free trial type for user i (NoFreeTrial, 24HourFreeTrial, OneFreeVideo)
$Country_i$	a categorical variable that is either Domestic, International, or Unknown
$SourceDomain_i$	Categorical variable denoting the source for the user (either Direct, Google, YouTube, or other)

The second set of regressions in Table 7 examines the impact of free plan design on monetization metrics. Specifically, we examine how conversion, revenue and CLV changes as the free plan becomes more generous. We find that the conversion rate decreases as the number of free minutes increased. However, interestingly, the 24 hour free trial did not show a significantly different conversion rate compared to the no free trial condition. While this may be surprising, recent research on free trials by Yoganarasimhan et al. (2022) has also found that the effectiveness of free trials can be weak, especially as the timeframe becomes more generous.

We then turn focus to revenue, where we find that total revenue had an inverted U shaped relationship with generosity of the free plan. Initially, when the free plan increases from 0 free minutes to 7 minutes across all videos, we find that revenue increases. However, when the free plan is increased from 7 to 15 minutes, we find that revenue decreases. In terms of magnitude, the revenue decrease from 7 to 15 minutes was roughly equivalent to the revenue increase from 0 to 7 minutes. Overall, the effects were large with an over 50% decrease in revenue from going from the 7 minute condition to either 0 or 15.

Table 6: Engagement - Free Users

	Dependent variable:			
	TotalMins	UniqueVids	AvgMinsPerVideo	UniqueWatchSessions
	(1)	(2)	(3)	(4)
FreeTrial24Hours	1.575***	0.126***	0.916***	0.161***
	(0.326)	(0.029)	(0.215)	(0.037)
OneFreeVideo	0.086	0.019	0.031	0.010
	(0.327)	(0.029)	(0.215)	(0.037)
FreeMins_0	-1.339***	-0.256***	-1.042***	-0.290***
	(0.378)	(0.034)	(0.249)	(0.043)
FreeMins_15	0.576	0.047	0.215	0.077
	(0.377)	(0.034)	(0.248)	(0.043)
FreeMins_30	1.436***	0.081*	0.720**	0.137**
	(0.377)	(0.034)	(0.248)	(0.043)
CountryInternational	1.001	0.077	1.008*	0.130
	(0.659)	(0.059)	(0.434)	(0.076)
CountryUnknown	3.232***	0.444***	1.664**	0.501***
	(0.841)	(0.076)	(0.553)	(0.096)
DomainGoogle	-0.280	0.029	0.085	0.011
	(0.361)	(0.032)	(0.237)	(0.041)
DomainOther	-1.776**	-0.213***	-0.979**	-0.270***
	(0.554)	(0.050)	(0.365)	(0.064)
DomainYoutube	0.335	0.058	0.607*	0.062
	(0.401)	(0.036)	(0.264)	(0.046)
Constant	1.809***	0.299***	1.237***	0.355***
	(0.424)	(0.038)	(0.279)	(0.049)
Observations	4,853	4,853	4,853	4,853
R^2	0.023	0.038	0.022	0.036
Adjusted R^2	0.021	0.036	0.020	0.034
Note:				*p<0.05; **p<0.01; ***p<0.001

It is interesting to understand why the revenue shows a different pattern from conversion rate,

Table 7: Conversion, Revenue and CLV

	Dependent variable:		
	Conversion	Revenue	CLV
	probit	*OLS*	*OLS*
	(1)	(2)	(3)
FreeTrial24Hours	-0.107	-1.412	-4.336
	(0.106)	(2.772)	(4.215)
OneFreeVideo	-0.068	-0.916	-1.627
	(0.104)	(2.772)	(4.215)
FreeMins_0	-0.039	-6.190	-10.994*
	(0.114)	(3.201)	(4.868)
FreeMins_15	-0.109	-6.381*	-9.907*
	(0.118)	(3.201)	(4.867)
FreeMins_30	-0.254*	-8.375**	-12.435*
	(0.129)	(3.203)	(4.870)
CountryInternational	0.437**	8.318	20.855*
	(0.168)	(5.544)	(8.430)
CountryUnknown	1.300***	78.082***	119.548***
	(0.112)	(5.556)	(8.449)
Constant	-2.071***	11.679***	18.683***
	(0.100)	(2.791)	(4.244)
Observations	4,960	4,960	4,960
R^2		0.040	0.041
Adjusted R^2		0.039	0.040
Log Likelihood	-452.484		
Akaike Inf. Crit.	920.968		
Note:	*p<0.05; **p<0.01; ***p<0.001		

when we might expect both to move in the same direction. Recall that we have multiple premium plans, which vary in terms of length. Now, the reason for the disparity between revenue and conversion rate arises due to premium plan length. Specifically, users in the 0 minute condition are much more likely to purchase the shorter weekly plan. Their purchase behavior was likely due to an inability to experience much of the product, and in a sense they purchased a shorter premium

Table 8: Engagement (First 24 Hours) - Free Users

	Dependent variable:			
	TotalMins	UniqueVids	AvgMinsPerVideo	UniqueWatchSessions
	(1)	(2)	(3)	(4)
FreeTrial24Hours	1.543***	0.118***	0.904***	0.158***
	(0.272)	(0.024)	(0.196)	(0.028)
OneFreeVideo	-0.039	0.012	-0.032	0.006
	(0.272)	(0.024)	(0.196)	(0.028)
FreeMins_0	-1.076***	-0.207***	-0.898***	-0.220***
	(0.314)	(0.027)	(0.227)	(0.032)
FreeMins_15	0.254	0.032	0.148	0.033
	(0.314)	(0.027)	(0.227)	(0.032)
FreeMins_30	0.672*	0.040	0.430	0.066*
	(0.314)	(0.027)	(0.226)	(0.032)
CountryInternational	0.782	0.051	0.944*	0.095
	(0.548)	(0.048)	(0.396)	(0.056)
CountryUnknown	3.307***	0.443***	1.727***	0.494***
	(0.699)	(0.061)	(0.505)	(0.071)
DomainGoogle	0.069	0.040	0.156	0.034
	(0.300)	(0.026)	(0.217)	(0.030)
DomainOther	-1.295**	-0.176***	-0.856*	-0.208***
	(0.461)	(0.040)	(0.333)	(0.047)
DomainYoutube	0.625	0.071*	0.690**	0.074*
	(0.334)	(0.029)	(0.241)	(0.034)
Constant	1.318***	0.241***	1.048***	0.266***
	(0.353)	(0.031)	(0.255)	(0.036)
Observations	4,853	4,853	4,853	4,853
R^2	0.022	0.041	0.021	0.041
Adjusted R^2	0.020	0.039	0.019	0.039
Note:				$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

plan effectively as a "paid trial." Consistent with this explanation, many of these premium users

Table 9: Engagement (Post 24 Hours) - Free Users

	Dependent variable:			
	TotalMins	UniqueVids	AvgMinsPerVideo	UniqueWatchSessions
	(1)	(2)	(3)	(4)
FreeTrial24Hours	0.032	0.009	0.019	0.003
	(0.156)	(0.016)	(0.096)	(0.020)
OneFreeVideo	0.125	0.013	0.028	0.003
	(0.156)	(0.016)	(0.096)	(0.021)
FreeMins_0	-0.263	-0.062***	-0.194	-0.070**
	(0.180)	(0.018)	(0.111)	(0.024)
FreeMins_15	0.322	0.026	0.204	0.044
	(0.180)	(0.018)	(0.110)	(0.024)
FreeMins_30	0.764***	0.053**	0.422***	0.071**
	(0.180)	(0.018)	(0.110)	(0.024)
CountryInternational	0.219	0.049	0.185	0.036
	(0.315)	(0.031)	(0.193)	(0.041)
CountryUnknown	-0.075	0.017	-0.016	0.007
	(0.401)	(0.040)	(0.246)	(0.053)
DomainGoogle	-0.349*	-0.009	-0.174	-0.024
	(0.172)	(0.017)	(0.106)	(0.023)
DomainOther	-0.480	-0.049	-0.300	-0.062
	(0.265)	(0.026)	(0.162)	(0.035)
DomainYoutube	-0.290	-0.009	-0.132	-0.012
	(0.192)	(0.019)	(0.117)	(0.025)
Constant	0.491*	0.065**	0.321**	0.089***
	(0.203)	(0.020)	(0.124)	(0.027)
Observations	4,853	4,853	4,853	4,853
R^2	0.009	0.011	0.008	0.010
Adjusted R^2	0.007	0.009	0.006	0.008
Note:				*p<0.05; **p<0.01; ***p<0.001

did not re-subscribe after a making them very low revenue conversions. Meanwhile, weekly plan purchases were uncommon among users who had a positive amount of free minutes. These results were even stronger when we considered customer lifetime value or CLV instead of revenues as a metric of interest. [13] Overall, users who spent more to begin with were more likely to spend in the future and the increase in CLV from 7 minutes compared to 0 minutes is even larger than the increase in observed revenue.

Another important issue is the difference between short run and long run outcomes, specifically whether engagement varies between them. Engagement during the first 24 hours for free users in shown in 8. There is a major lift for users in the 24 hour free trial, which has a larger effect than even the most generous free plan, i.e. 30 free minutes making it seem like the 24-hour plan would be better in driving engagement. However, after the first 24 hours, we find that there is *no intertemporal spillover* or state-dependence from being in the 24 hour free trial. After the 24 hour free trial, free users watched the same amount regardless of whether they had been in the free trial condition or no free trial condition. In contrast, offering a higher the number of free minutes permanently (freemium) has a persistent increase in engagement over the long run. This finding has significant implications for companies that want to use the plan design to achieve different short run or long run goals.

8 Content Analysis

We would like to investigate whether there are videos with specific characteristics that perform especially well under some of the experimental conditions. Companies in similar (or the same) market often adopt significantly different strategies. In this case, we explore whether the content that a firm owns in its library impacts the effectiveness of the plan, in terms of engagement. Consider, for example, it may be that Netflix has more movies than TV shows, whereas Hulu has more shows. It is possible that they therefore might find different plan designs optimal based on the content in their library. The content analysis helps answer in a principled and systematic way the question of how specific content types perform differently under different plans.

We note that the previous experimental analysis captures *average treatment effects* across users. In the content analysis, we switch focus from the user as the unit of experimentation to the *content video as the unit of experimentation*. Thus, looking at videos under control and treatment (or different plans), we examine whether they receive differential engagement.

[13] CLV is a projection of their future spending based on their behavior during the experiment, and requires us to make a further prediction on future purchase rates.

While free users watch more when given more access to content, it is possible this effect is not equal among video types. Such heterogeneity could be used to better determine the level of free content. To assess heterogeneity we follow Chernozhukov et al. (2020), which provides a strategy for determining heterogeneity among treatment effects. The only difference here is that instead of getting heterogeneous treatment effects across users, we will look at the heterogeneous treatment effects across videos. Specifically, in our context, we ask which types of videos had a higher increase in the amount of minutes watched when there was a change to the free plan? Generally, the problem with finding heterogeneous treatment effects *a posteriori* without a clear ex-ante hypothesis or pre-analysis plan is that there are many ways to split high dimensional data making distinguishing true effects and noise problematic. Crucially, Chernozhukov et al. (2020) provides a solution to this issue. The main idea is to create predictions of individual treatment effects and then use these predictions to make inferences on the variables of interest. Crucially, a predictive set of covariates Z_j are needed, for which we use the features described in section 6.2.

Their approach works as follows. First the data is split into a training set and a test set using a 50/50 split. Using the training set, a machine learning method is used to estimate the expected number of minutes watched in the control condition using covariates Z_j. Using the notation of Bryan et al. (2021), we call this $B(Z_j)$. Similarly, the training data is used to estimate the expected number of minutes watched with covariates Z_j for videos assigned to the treatment. These estimates are used to predict the outcome for both control and treatment assignments using the covariates for each video in the test set. The difference gives the predicted video-level treatment effects, denoted $S(Z_j)$. The method is agnostic to the type of the machine learning method used as it is just a tool for prediction. This means that any ML method could be used including elastic net, neural net, random forest, gradient boosting, etc. ML methods can be compared using a goodness-of-fit measure λ as defined in Chernozhukov et al. (2020).

Whether there is evidence of heterogeneity of the treatment based on covariates can be determined using the following regression on the test data:

$$Y_j = \alpha\, X_j + \beta_1\, T_j + \beta_2\, T_j\, S(Z_j) + \epsilon_j \tag{2}$$

Here, X_j is a set of covariates that includes $B(Z_j)$ (the predicted control effects) and other controls (if desired), T_j is an indicator for whether the observation was in the treatment. The primary use of this specification is to test the null hypothesis of no heterogeneity which would occur if $\beta_2 = 0$. The average treatment effect (ATE) is estimated by β_1. This equation is called the BLP (best linear predictor).

If the null hypothesis can be rejected, then there is evidence of heterogeneous treatment effects, but we still do not know much about how or why. One analysis presented by Chernozhukov et al. (2020) known as "(sorted) group average treatment effect" (GATES) provides a way to understand how the treatment effects differ across well defined groups. More precisely, by sorting the sample by size of the predicted treatment effect, $S(Z_j)$, into K equal groups, we can obtain the difference between the $k\%$ most affected videos and the $k\%$ least affected videos. The GATES effects are estimated using the following regression:

$$Y_j = \alpha X_j + \sum_{k=1}^{K} \gamma_k \, T_j \, \mathbf{1}(S_j \in I_k) + \eta_j \tag{3}$$

where I_k is the set of videos in the kth group, and γ_k measures the sorted groups average treatment effect for each group. From these sorted groups, it can determined which covariates are associated with the heterogeneity using a process known as “classification analysis” (CLAN). This is done by comparing the average characteristics of the most and least affected groups of videos where significance is determined by two sample t-tests.[14]

We explore how plan design interacts with video-level features, Z_j, (see Section 6.2) by seeing how the total number of minutes watched per video Y_j changes. Using the video features, Z_j outlined in Section 6.2, we test how plan design interacts with these video features with regards to the total engagement, Y_j. This is done in two ways: first, by having increasing the number of free minutes from 0 (control) to 30 (treatment) and second, by giving the user a 24 hour free trial (treatment) rather than no free trial (control). We considered two ML methods to estimate the proxy predictors: elastic net and random forest.[15] The goodness-of-fit statistics show that elastic net had superior performance to random forest for our data in both situations tested (Table 10). The first goal is to run the BLP (eq. 2) and test the null hypothesis that $\beta_1 = 0$ (ATE differs from 0) and the null hypothesis that $\beta_2 = 0$ (that our predictions $S(Z_j)$ are able to capture heterogeneity). In both our analyses, we detect the heterogeneous treatment effect (HTE) is significant with p-values less than 0.001 (Table 11).

In order to do the CLAN, we first sort the videos by the *size of the predicted treatment effect*;

[14] While the steps above provide a method to estimate heterogeneity it could depend on the splitting of the training and test set. For this reason, we run the analysis 1000 times each with a different split randomly chosen split. The result is 1000 estimates of β_2 and γ_k along with associated confidence intervals, standard errors and p-values. For the parameter estimates, the *median* across the 1000 runs is reported along with the median of the confidence intervals. The p-values account both for the estimation uncertainty and the uncertainty induced by the data splitting.

[15] These methods are implemented in R using the *mlr3* package (Lang et al., 2019). For each split of the data, hyperparameters are tuned only using the training sample. Scaling of the outcomes and covariates is done within the *mlr3* package.

Table 10: Comparison of ML Methods: Free Minutes and Free Trial

	Free Minutes		Free Trial	
	Elastic Net	Random Forest	Elastic Net	Random Forest
Best BLP (λ)	43.75	15.20	8.00	2.07
Best GATES ($\bar{\lambda}$)	10.63	8.73	4.16	3.85

Notes: Medians over 1000 splits in half.

Table 11: BLP of Free Minutes and Free Trial

Free Minutes		Free Trial	
ATE (β_1)	HET (β_2)	ATE(β_1)	HET (β_2)
2.277	0.704	1.481	0.439
(1.136, 3.436)	(0.573, 0.828)	0.439	(0.251, 0.647)
[0.000]	[0.000]	[0.003]	[0.000]

Notes: Medians over 1000 splits in half. Median confidence interval (α = .05) is in parenthesis. P-values for the hypothesis that the parameter is equal to zero against the two-sided alternatives is in brackets.

the videos are then partitioned into quartiles. From the regression in eq. 3, the average treatment effect of each group k, γ_k, is estimated. Notably, we test whether there is a significant difference between the average treatment effect of the 25% of videos with the highest predicted treatment effect, γ_4, is statistically different than the average treatment effect of the 25% of videos with the lowest predicted treatment effect. The results are shown in Table 12.

The CLAN show the characteristics associated with the heterogeneity, but it does not attempt to quantify the importance of each. In general, this can be a very difficult problem as many characteristics are likely to be highly correlated (for example, happiness in speech and happiness in facial emotions). The CLAN results are presented in Tables 13 and 14, which show multiple results that are robust across both an increase in the number of free minutes from 0 to 30 and an increase from no free trial to a 24 hour free trial. Videos with the following characteristics are those associated with the largest increase in watch time as more free time was allotted to the users:

1. A longer sentence length and a faster speech rate
2. More collaborative language and less individualistic language
3. A lower rate of questions being asked
4. Speech emotions that had less sadness and more trust
5. Speech sentiment that was more positive
6. Videos that were of higher quality (pre-recorded rather than live-recorded)
7. Classes that were longer and more difficult (more courses and less moves)
8. Visual features that were non-neutral (especially happy visual features)

9. Instructors that stood further away from the camera

In general, as the amount of free content available increased, engagement was allocated disproportionately to videos where the instructor was more positive both in their speech and in facial emotions. Videos that were more demanding also received higher engagement (i.e. longer *courses* rather than the shorter *moves* videos). An interesting takeaway from the linguistic teaching style were that videos where the instructor used more collaborative language fared better (i.e. more plural and fewer individual first person pronouns).

Table 12: GATES of 25% Most and Least Affected Groups (Free Minutes and Free Trial)

	Free Minutes			**Free Trial**		
	25% Most (I_4)	25% Least (I_1)	Difference	25% Most (I_4)	25% Least (I_1)	Difference
γ_k	6.14	0.51	5.67	3.54	0.60	2.94
	(3.75, 8.56)	(-1.85, 2.94)	(2.23, 9.08)	(1.58, 5.52)	(-1.35, 2.57)	(0.23, 5.64)
	[0.000]	[0.676]	[0.001]	[0.000]	[0.554]	[0.034]

Notes: Medians over 1000 splits in half. Median confidence interval ($\alpha = .05$) is in parenthesis. P-values for the hypothesis that the parameter is equal to zero against the two-sided alternatives is in brackets.

Table 13: CLAN for Text Features

	Free Minutes			**Free Trial**		
	25% Most (δ_4)	25% Least (δ_1)	Difference ($\delta_4 - \delta_1$)	25% Most (δ_4)	25% Least (δ_1)	Difference ($\delta_4 - \delta_1$)
Sentence Length	9.378	8.246	1.128***	9.109	8.284	0.816***
	(9.002, 9.975)	(8.008, 8.488)		(8.784, 9.431)	(8.082, 8.489)	
Speaking Rate	1.686	1.497	0.183**	1.753	1.480	0.273***
	(1.593, 1.780)	(1.407, 1.588)		(1.672, 1.834)	(1.401, 1.558)	
% of Big Words	0.097	0.082	0.015***	0.091	0.087	0.004
	(0.093, 0.101)	(0.078, 0.086)		(0.087, 0.095)	(0.083, 0.090)	
Individualistic	0.025	0.036	-0.011***	0.027	0.033	-0.006***
	(0.023, 0.027)	(0.033, 0.038)		(0.025, 0.029)	(0.031, 0.035)	
Collaborative	0.020	0.011	0.009***	0.017	0.013	0.004***
	(0.018, 0.021)	(0.010, 0.012)		(0.016, 0.018)	(0.012, 0.014)	
Politeness	.0011	.0015	-.0004	.0009	.0014	-.0005**
	(.0008, .0014)	(.0012, .0019)		(.0007, .0011)	(.0011, .0018)	
Question Rate	0.044	0.072	-0.028***	0.045	0.066	-0.021***
	(0.040, 0.048)	(0.066, 0.079)		(0.041, 0.49)	(0.061, 0.072)	
Text Anger	0.051	0.069	-0.018***	0.059	0.060	-0.001
	(0.045, 0.056)	(0.060, 0.077)		(0.053, 0.065)	(0.054, 0.067)	
Text Anticipation	0.247	0.246	-0.000	0.239	0.258	-0.020*
	(0.236, 0.258)	(0.232, 0.261)		(0.230, 0.248)	(0.246, 0.272)	
Text Disgust	0.030	0.023	0.007**	0.036	0.020	0.016***
	(0.026, 0.034)	(0.020, 0.026)		(0.032, 0.039)	(0.017, 0.022)	
Text Fear	0.074	0.067	0.006	0.062	0.079	-0.017***
	(0.066, 0.081)	(0.060, 0.073)		(0.056, 0.067)	(0.072, 0.085)	
Text Joy	0.0196	0.0196	0.001	0.203	0.186	0.018***
	(0.188, 0.204)	(0.0187, 0.025)		(0.196, 0.211)	(0.179, 0.194)	
Text Sadness	0.070	0.091	-0.021***	0.070	0.090	-0.021***
	(0.065, 0.076)	(0.084, 0.098)		(0.065, 0.074)	(0.085, 0.096)	
Text Surprise	0.076	0.083	-0.007	0.069	0.088	-0.019***
	(0.071, 0.081)	(0.078, 0.089)		(0.065, 0.074)	(0.083, 0.094)	
Text Trust	0.253	0.222	0.031***	0.259	0.216	0.043***
	(0.243, 0.262)	(0.211, 0.232)		(0.251, 0.267)	(0.207, 0.224)	
Text Negative	0.207	0.281	-0.074***	0.217	0.280	-0.064***
	(0.194, 0.221)	(0.260, 0.302)		(0.204, 0.230)	(0.263, 0.297)	
Text Positive	0.793	0.719	0.074***	0.783	0.720	0.064***
	(0.779, 0.806)	(0.698, 0.740)		(0.770, 0.796)	(0.703, 0.737)	

Notes: Medians over 1000 splits in half. Median confidence interval ($\alpha = .05$) is in parenthesis. *p<0.05; **p<0.01; ***p<0.001

Table 14: CLAN for Basic Video Properties, Audio Features and Visual Features

	Free Minutes			Free Trial		
	25% Most (δ_4)	25% Least (δ_1)	Difference ($\delta_4 - \delta_1$)	25% Most (δ_4)	25% Least (δ_1)	Difference ($\delta_4 - \delta_1$)
Basic Video Properties						
Pre-Recorded	.511	.231	.284***	.592	.221	.370***
	(.453, .568)	(.182, .280)		(.543, .641)	(.179, .262)	
Intro Proportion	.280	.251	.028	.265	.229	.035
	(.246, .315)	(.219, .283)		(.236, .293)	(.202, .255)	
Courses	.143	.000	.143***	.143	.000	.143***
	(.107, .177)	(.000, .000)		(.107, .177)	(.000, .000)	
Foundations	.158	.057	.097***	.091	.084	.005
	(.116, .148)	(.029, .082)		(.062, .119)	(.056, .111)	
Moves	.045	.209	-.167***	.049	.204	-.153***
	(.021)	(.161, .254)		(.028, .071)	(.163, .243)	
Audio Features						
ZCR	.023	.019	.003	.025	.019	.005***
	(.020, .025)	(.017, .022)		(.022, .027)	(.017, .021)	
Energy	3.039	3.036	0.003	3.048	3.035	.013**
	(3.031, 3.047)	(3.028, 3.044)		(3.041, 3.055)	(3.029, 3.042)	
Spectral Centroid	.1522	.1472	.005***	.1506	.1492	.0014
	(.1510, .1535)	(.1463, .1482)		(.1496, .1515)	(.1483, .1502)	
Visual Features						
Visual Anger	.0041	.0052	-.0010	.0046	.0040	.0006
	(.0033, .0050)	(.0040, .0064)		(.0039, .0054)	(.0031, .0049)	
Visual Contempt	.0066	.0038	.0027***	.0045	.0064	-.0020**
	(.0055, .0077)	(.0032, .0044)		(.0038, .0051)	(.0055, .0074)	
Visual Disgust	.0009	.0009	.0000	.0009	.0009	.0000
	(.0007, .0012)	(.0006, .0012)		(.0006, .0011)	(.0006, .0012)	
Visual Fear	.0015	.0006	.0001***	.0013	.0008	.0005**
	(.0011, .0019)	(.0005, .0008)		(.0010, .0016)	(.0006, .0010)	
Visual Happiness	.252	.168	.084***	.234	.184	.052***
	(.230, .274)	(.149, .187)		(.215, .254)	(.167, .201)	
Visual Neutral	.669	.0762	-.094***	.680	.753	-.073***
	(.645, .692)	(.742, .782)		(.659, .699)	(.734, .771)	
Visual Sadness	.0114	.0082	.0032*	.009	.011	-.002
	(.0090, .0138)	(.0067, .097)		(.007, .011)	(.009, .013)	
Visual Surprise	.053	.049	.004	.064	.039	.025***
	(.047, .059)	(.043, .0557)		(.057, .071)	(.035, .043)	
Head Movement	.307	.318	-.010	.317	.305	.012
	(.288, .327)	(.299, .336)		(.300, .334)	(.289, .321)	
Closeness	.081	.095	-.015***	.078	.090	-.011***
	(.076, .085)	(.090, .101)		(.074, .083)	(.085, .094)	

Notes: Medians over 1000 splits in half. Median confidence interval ($\alpha = .05$) is in parenthesis. *p<0.05; **p<0.01; ***p<0.001

9 Discussion and Conclusion

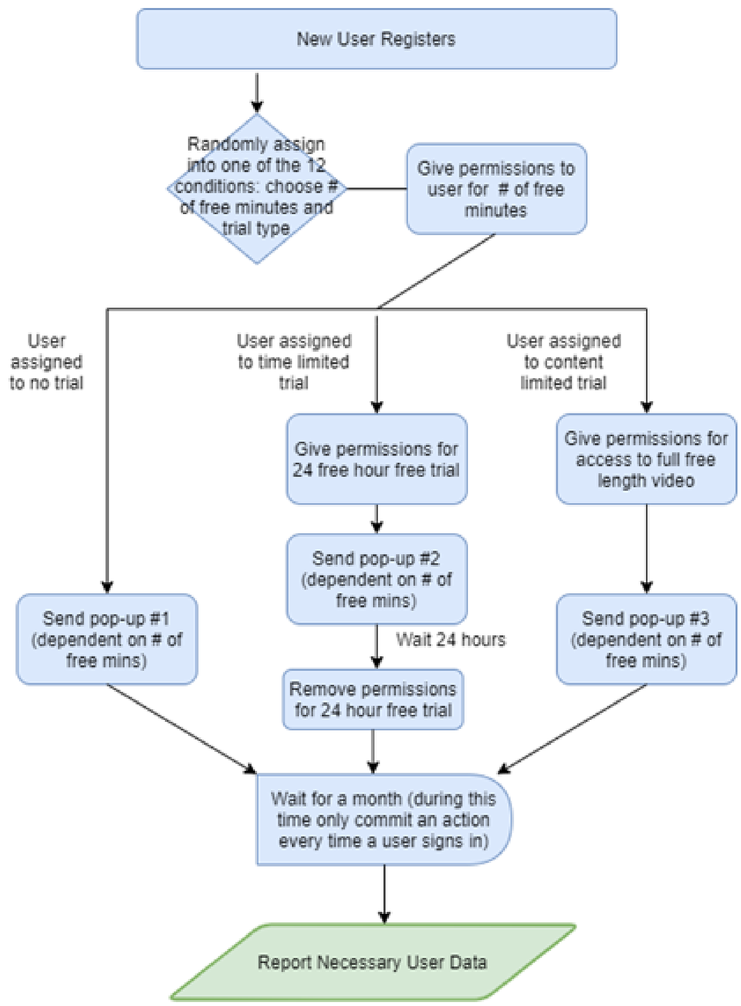
In digital business settings, the freemium business model permanently available free option as well as are time limited trial are among the most common monetisation strategies used. However, despite its importance, there are very few studies that identify the causal impact changes in the design of the plans offered. We have examined but the long run and short run consequences of different designs of the free plan and free trial in a freemium business model using a carefully designed field experiment with a performing arts learning platform. We find that a short term free trial drives engagement in the short run but does not have long run impact either on engagement or monetization. However, the amount of content *permanently available* in the free plan does impact engagement come on conversion rates, the type of plan chosen, and both short run and long run monetization. We also evaluate the heterogeneous impact of different content types using text, audio and visual characteristics from , and find that many content characteristics are associated with higher treatment effects.

There are a number of questions would be interesting to explore and further research related to this topic. First is the question of the user journey how consumers make choices to watch specific content videos how do you navigate from one content video to another in their learning journey. Second, while we have examined a 24 hour free plan, it would be interesting to see the engagement impact extending the trial length to examine whether the results in the performing arts learning context differ from those examined in prior research. Third, whereas the current experimental design imposes the same kinds of restrictions across all premium content, it would be Beautiful to examine whether search content restrictions could be customized, either at the content level or at the user level. Finally, it would be interesting to incorporate price variations in the experimental design to understand the value generated by different elements of the premium plan. This Issue is especially relevant sense most subscription businesses vary prices infrequently, if at all, making a challenging to identify the willingness to pay distribution (Chou and Kumar, 2023).

A Experiment Implementation Flowchart

We detail the implementation of the field experiment in Figure 7. First, when a user registers, they are *placed permanently* into one of the 12 conditions corresponding to the 4×3 experimental design. Users who are not registered cannot access the site or even see the details about the free plan, so registration is a necessary first step. Thus, we have (almost) the same number of users allocated into each of the conditions.

Figure 7: Experiment Design



B Placebo Regressions

In addition to the main engagement and monetization metrics, we conduct placebo tests here. In the main analysis, the idea is that when we change the design of the free plan, the engagement would potentially change. Here, the focus is on identifying metrics that we would expect **should not depend** on the design of the free plan. Recall that Live content was always included in the free plan across all plan designs and for all users. Thus, we might expect that user engagement with live viewing would not depend on the design of the free plan. In Table 15, we detail the results of this regression, and find that live viewing indeed does not depend on plan design. Similarly, we might expect that when users upgrade to the premium plan, their engagement would not be impacted by the free plan design since they are not subject to its constraints. Again, we find this to be the case, as shown in Table 16.

Table 15: Live Viewing - All Users

	Dependent variable:	
	WatchedLive	MinsLive
	probit	*OLS*
	(1)	(2)
FreeTrial24Hours	0.034	-0.089
	(0.083)	(0.434)
OneFreeVideo	0.063	0.022
	(0.082)	(0.434)
FreeMins_0	-0.048	-0.405
	(0.096)	(0.501)
FreeMins_15	0.051	-0.358
	(0.093)	(0.501)
FreeMins_30	-0.008	-0.355
	(0.096)	(0.501)
CountryInternational	-0.403	0.938
	(0.234)	(0.870)
CountryUnknown	-0.258	-2.241*
	(0.161)	(1.055)
DomainGoogle	-0.284***	-0.796
	(0.083)	(0.481)
DomainOther	0.254*	1.853*
	(0.108)	(0.741)
DomainYoutube	-0.730***	-1.885***
	(0.122)	(0.537)
Constant	-1.580***	2.321***
	(0.100)	(0.565)
Observations	4,960	4,960
R^2		0.007
Adjusted R^2		0.005
Log Likelihood	-776.164	
Akaike Inf. Crit.	1,574.328	
Note:	*p<0.05; **p<0.01; ***p<0.001	

Table 16: Engagement - Premium Users

	Dependent variable:			
	MinutesWatchedLib	UniqueVidsWatchLib	AverageMinsWatched	UniqueWatchSessLib
	(1)	(2)	(3)	(4)
FreeTrial24Hours	-34.191	-1.712	-0.799	-3.416
	(25.258)	(1.350)	(4.856)	(2.458)
OneFreeVideo	-30.639	-1.577	-1.300	-3.439
	(24.021)	(1.284)	(4.618)	(2.338)
FreeMins_0	57.794*	2.653	5.993	5.644*
	(25.555)	(1.366)	(4.913)	(2.487)
FreeMins_15	9.246	1.726	0.269	3.727
	(27.186)	(1.453)	(5.227)	(2.646)
FreeMins_30	51.564	0.828	6.065	3.290
	(31.216)	(1.668)	(6.002)	(3.038)
CountryInternational	-10.770	-2.225	2.191	-2.358
	(38.983)	(2.083)	(7.495)	(3.794)
CountryUnknown	-14.305	-2.273	-1.095	-3.742
	(50.215)	(2.684)	(9.655)	(4.887)
DomainGoogle	-12.889	-0.081	-2.660	0.246
	(37.544)	(2.006)	(7.218)	(3.654)
DomainOther	-20.650	2.126	-2.540	2.548
	(58.599)	(3.132)	(11.267)	(5.703)
DomainYoutube	1.225	1.658	-1.275	1.154
	(47.590)	(2.543)	(9.150)	(4.632)
Constant	87.767*	4.102*	18.364*	7.875*
	(37.492)	(2.004)	(7.208)	(3.649)
Observations	107	107	107	107
R^2	0.102	0.072	0.029	0.080
Adjusted R^2	0.008	-0.024	-0.072	-0.016
Note:				$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

References

Allenby, G. M. and Rossi, P. E. (1998). Marketing models of consumer heterogeneity. *Journal of econometrics*, 89(1-2):57–78.

Ariely, D., Gneezy, U., and Haruvy, E. (2018). Social norms and the price of zero. *Journal of Consumer Psychology*, 28(2):180–191.

Ascarza, E., Netzer, O., and Runge, J. (2020). The twofold effect of customer retention in freemium settings. *Columbia Business School Research Paper Forthcoming*.

Balkan, S. and Kholod, T. (2015). Video analytics in market research. *Information Systems Management*, 32(3):192–199.

Bapna, R., Ramaprasad, J., and Umyarov, A. (2016). Monetizing freemium communities: Does paying for premium increase social engagement? *Available at SSRN 2885681*.

Barnes, S. J. and Kirshner, S. N. (2021). Understanding the impact of host facial characteristics on airbnb pricing: Integrating facial image analytics into tourism research. *Tourism Management*, 83:104235.

Boughanmi, K. and Ansari, A. (2021). Dynamics of musical success: A machine learning approach for multimedia data fusion. *Journal of Marketing Research*, 58(6):1034–1057.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Bryan, G. T., Karlan, D., and Osman, A. (2021). Big loans to small businesses: Predicting winners and losers in an entrepreneurial lending experiment. Technical report, National Bureau of Economic Research.

Cao, J., Li, X., and Zhang, L. (2022). Is relevancy everything? a deep learning approach to understand the coupling of image and text. *A Deep Learning Approach to Understand the Coupling of Image and Text (June 27, 2022)*.

Cao, Z., Simon, T., Wei, S.-E., and Sheikh, Y. (2017). Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7291–7299.

Chakraborty, I., Chiong, K., Dover, H., and Sudhir, K. (2022). Ai and ai-human based salesforce hiring using interview videos. *Available at SSRN 4137872*.

Cheng, M. M. and Zhang, S. (2022). Reputation burning: Analyzing the impact of brand sponsorship on social influencers. *Shunyuan, Reputation Burning: Analyzing the Impact of Brand Sponsorship on Social Influencers*.

Chernozhukov, V., Demirer, M., Duflo, E., and Fernandez-Val, I. (2020). Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in india. Technical report, National Bureau of Economic Research.

Chou, C. and Kumar, V. (2023). Can willingness to pay be identified without price variation? what big data on usage tracking can (and cannot) tell us. *Marketing Science (Forthcoming)*.

Coker, K. K., Flight, R. L., and Baima, D. M. (2017). Skip it or view it: The role of video storytelling in social media marketing. *Marketing Management Journal*, 27(2).

Datta, H., Foubert, B., and Van Heerde, H. J. (2015). The challenge of retaining customers acquired with free trials. *Journal of Marketing Research*, 52(2):217–234.

Dimberg, U., Thunberg, M., and Elmehed, K. (2000). Unconscious facial reactions to emotional facial expressions. *Psychological science*, 11(1):86–89.

Ekman, P. and Oster, H. (1979). Facial expressions of emotion. *Annual review of psychology*.

Gerhard, D. (2003). *Audio signal classification: History and current techniques*. Citeseer.

Giannakopoulos, T. (2015). pyaudioanalysis: An open-source python library for audio signal analysis. *PloS one*, 10(12):e0144610.

Goldenberg, C. N. (1989). Making success a more common occurrence for children at risk for failure: Lessons from hispanic first graders learning to read. *Risk makers, risk takers, risk breakers: Reducing the risks for young literacy learners*, pages 48–78.

Gu, X., Kannan, P., and Ma, L. (2018). Selling the premium in freemium. *Journal of Marketing*, 82(6):10–27.

Guo, P. J., Kim, J., and Rubin, R. (2014). How video production affects student engagement: An empirical study of mooc videos. In *Proceedings of the first ACM conference on Learning@ scale conference*, pages 41–50.

Halbheer, D., Stahl, F., Koenigsberg, O., and Lehmann, D. R. (2014). Choosing a digital content strategy: How much should be free? *International Journal of Research in Marketing*, 31(2):192–206.

Hobjilă, A. (2012). Positive politeness and negative politeness in didactic communication–landmarks in teaching methodology. *Procedia-Social and Behavioral Sciences*, 63:213–222.

Holm, A. B. and Günzel-Jensen, F. (2017). Succeeding with freemium: strategies for implementation. *Journal of Business Strategy*.

Hu, A. and Ma, S. (2021). Persuading investors: A video-based study. Technical report, National Bureau of Economic Research.

Hwang, S., Liu, X., and Srinivasan, K. (2021). Voice analytics of online influencers—soft selling in branded videos.

Jarvis, P. (1985). Thinking critically in an information society: A sociological analysis. *Lifelong learning*, 8(6):11–14.

Jockers, M. L. (2015). *Syuzhet: Extract Sentiment and Plot Arcs from Text*.

John, L. K., Emrich, O., Gupta, S., and Norton, M. I. (2017). Does "liking" lead to loving? the impact of joining a brand's social network on marketing outcomes. *Journal of marketing research*, 54(1):144–155.

Kamakura, W. A. and Russell, G. J. (1989). A probabilistic choice model for market segmentation and elasticity structure. *Journal of marketing research*, 26(4):379–390.

Kosner, A. W. (2015). Apple app store revenue surge and the rise of freemium app pricing. *Forbes*.

Kumar, V. (2014). Making" freemium" work. *Harvard business review*, 92(5):27–29.

Laham, S. M., Koval, P., and Alter, A. L. (2012). The name-pronunciation effect: Why people like mr. smith more than mr. colquhoun. *Journal of Experimental Social Psychology*, 48(3):752–756.

Lambrecht, A. and Misra, K. (2017). Fee or free: When should firms charge for online content? *Management Science*, 63(4):1150–1165.

Lang, M., Binder, M., Richter, J., Schratz, P., Pfisterer, F., Coors, S., Au, Q., Casalicchio, G., Kotthoff, L., and Bischl, B. (2019). mlr3: A modern object-oriented machine learning framework in R. *Journal of Open Source Software*.

Lewis, R. (2001). Classroom discipline and student responsibility:: The students' view. *Teaching and teacher education*, 17(3):307–319.

Li, H., Jain, S., and Kannan, P. (2019). Optimal design of free samples for digital products and services. *Journal of Marketing Research*, 56(3):419–438.

Li, Y. and Xie, Y. (2020). Is a picture worth a thousand words? an empirical study of image content and social media engagement. *Journal of Marketing Research*, 57(1):1–19.

Mohammad, S. and Turney, P. (2010). Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. In *Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text*, pages 26–34.

Mohammad, S. M. and Turney, P. D. (2013). Crowdsourcing a word–emotion association lexicon. *Computational intelligence*, 29(3):436–465.

Nelson, P. (1970). Information and consumer behavior. *Journal of political economy*, 78(2):311–329.

Pennebaker, J. W., Francis, M. E., and Booth, R. J. (2001). Linguistic inquiry and word count: Liwc 2001. *Mahway: Lawrence Erlbaum Associates*, 71(2001):2001.

Pogacar, R., Shrum, L., and Lowrey, T. M. (2018). The effects of linguistic devices on consumer information processing and persuasion: A language complexity× processing mode framework. *Journal of Consumer Psychology*, 28(4).

Roads, C. (1996). *The computer music tutorial*. MIT press.

Savov, V. (2014). Freemium apps account for 98 percent of worldwide google play revenue. *The Verge*.

Schubert, E. and Wolfe, J. (2006). Does timbral brightness scale with frequency and spectral centroid? *Acta acustica united with acustica*, 92(5):820–825.

Shampanier, K., Mazar, N., and Ariely, D. (2007). Zero as a special price: The true value of free products. *Marketing science*, 26(6):742–757.

Shi, Z., Zhang, K., and Srinivasan, K. (2019). Freemium as an optimal strategy for market dominant firms. *Marketing Science*, 38(1):150–169.

Sriram, S., Chintagunta, P. K., and Agarwal, M. K. (2010). Investigating consumer purchase behavior in related technology product categories. *Marketing Science*, 29(2):291–314.

Subramanian, H., Rao, P., and Roy, S. (2004). Audio signal classification. *EE Dept, IIT Bombay*, pages 1–5.

Tausczik, Y. R. and Pennebaker, J. W. (2010). The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of language and social psychology*, 29(1):24–54.

Valliere, D. and Peterson, R. (2004). Inflating the bubble: examining dot-com investor behaviour. *Venture Capital*, 6(1):1–22.

Wager, S. and Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523):1228–1242.

Wiltbank, R. (2005). Investment practices and outcomesof informal venture investors. *Venture Capital*, 7(4):343–357.

Yang, J., Zhang, J., and Zhang, Y. (2021). First law of motion: Influencer video advertising on Tiktok. *Available at SSRN 3815124*.

Yoganarasimhan, H., Barzegary, E., and Pani, A. (2022). Design and evaluation of optimal free trials. *Management Science*.

Yu, Y., Wang, Y., Zhang, G., Zhang, Z., Wang, C., and Tan, Y. (2022). Outstream video advertisement effectiveness. *Available at SSRN 4098246*.

Zhang, S., Lee, D., Singh, P. V., and Srinivasan, K. (2021). What makes a good image? airbnb demand analytics leveraging interpretable image features. *Management Science*.

Zhou, M., Chen, G. H., Ferreira, P., and Smith, M. D. (2021). Consumer behavior in the online classroom: Using video analytics and machine learning to understand the consumption of video courseware. *Journal of Marketing Research*, 58(6):1079–1100.