Empirical Study and Business Model Analysis of Successful Freemium Strategies in Digital Products.

ARCHIVES
MASSACHUSETTS INSTITUTE OF TECHNOLOLGY
JUN 24 2015
LIBRARIES

By

Pronitha Mysore Shankarananda

B.E. Electrical & Electronics Engineering
National Institute of Technology, Karnataka 2006
MBA, Faculty of Management Studies, 2009

SUBMITTED TO THE MIT SLOAN SCHOOL OF MANAGEMENT IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN MANAGEMENT STUDIES
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 2015

©2015 Pronitha Mysore Shankarananda. All rights reserved.

The author hereby grants to MIT permission to reproduce and to distribute publicly paper and electronic copies of this thesis document in whole or in part in any medium now known or hereafter created.

Signature of Author: Signature redacted
MIT Sloan School of Management
May 8, 2015

Certified by: Signature redacted
Sinan K. Aral
David Austin Professor of Management
MIT Sloan School of Management

Accepted by: Signature redacted
Michael A. Cusumano
SMR Distinguished Professor of Management
Program Director, M.S. in Management Studies Program
MIT Sloan School of Management

Empirical Study and Business Model Analysis of Successful Freemium Strategies in Digital Products.

By

Pronitha M Shankarananda

Submitted to MIT Sloan School of Management on May 8, 2015 in partial fulfillment of the requirements for the Degree of Master of Science in Management Studies.

Abstract

"Freemium" has become a popular business model for digital and internet startups to acquire customers quickly and monetize with limited marketing resources. In this model, basic functionalities and features of the product are offered free of cost and advanced features are monetized. Typically firms rely on the size of the free user network since the conversion rate in these models is very low (2%-7%). This thesis analyzes successful business models and uses predictive analytics on a dataset from a freemium product to determine critical success factors.

Empirical study is conducted on 79,033 total users and 4,217 premium subscribers of Last.fm, a Musica as a service (MaaS) product which employs the freemium pricing model, to predict the probability of a user being a subscriber. Classification and regression trees (CART) with k-fold cross validation is used to model on the training data and is validated on the test dataset to understand the influence of demographic, engagement, retention and social factors on subscribers. Five successful premium companies LinkedIn, Zynga, Evernote, Spotify and DropBox are studied to understand the company background, value proposition and freemium model. The analyses was based on secondary case studies and information.

The CART model designed yielded an accuracy of 74.84%, sensitivity of 67.6%, false negative rate (FNR) of 32.4% and Area under the curve (AUC) of 0.7474937. On fine-tuning the penalty matrix, FNR is reduced to 22% and sensitivity increased to 78% with an accuracy of 65.8%. The empirical study showed that age (demographic factor), playcount (engagement factor) and registration unixtime (retention) were the most significant variables in predicting outcome. Study of successful firms showed that freemium is not a one-size-fits-all strategy for internet startups. The strategy needs to be crafted specifically for each firm. Many factors go into formulating this plan including a strong product catalogue, focused customer acquisition engine, referrals, social and community features, continuous testing, data driven approach and commitment to continuous innovation.

Thesis Supervisor: Professor Sinan K. Aral
Title: David Austin Professor of Management

To my advisor Prof. Sinan Aral for his invaluable guidance.

To my parents Shankarananda and Sreemathi and my little sister Bhargavi for their support, encouragement and unwavering faith in me.

Finally to my husband Nikhil for his extraordinary patience and endless love.

Table of Contents

Abstract **2**
Table of Contents **4**
List of Figures **5**
List of Tables **6**
1. Introduction **7**
1.1 Birth of Free 7
1.2 Internet Businesses 8
1.3 Inherently Free and Perpetually Networked 8
1.5 Structure of Thesis 8
2. Literature Review **9**
2.1 Types of Freemium 9
2.2 Freemium Sales Cycle 10
2.3 Freemium as a Multi-sided Market 11
3. Freemium Business Model **14**
3.1 Elements of Freemium Business Model 14
3.2 Freemium Lifecycle 18
3.3 Metrics in Freemium Model 18
4. Research Design **24**
5. Business Model Analysis **25**
5.1 LinkedIn 25
5.2 Zynga 26
5.3 Evernote 27
5.4 Spotify 29
5.5 DropBox 30
5.6 Summary 31
6. Empirical Study **34**
6.1 Research Plan 34
6.2 Last.fm 35
6.3 Data Collection and Preparation 36
6.4 Data Description 36
6.4 Predictive Analytics 42
6.5 Interpretation of Model Results 46
7. Key Findings and Implications **48**
8. Limitations **53**

8.1 Business Model Analysis 53
8.2 Empirical Study 53
9. References 54
10. Appendix 59
10.1 Business Model Canvas 59
10.2 CART Model with penalty matrix altered 61
10.3 R Scripts 62

List of Figures

Figure 1 Time differentiation in freemium (Pujol, 2010) 9
Figure 2 Feature differentiated Freemium (A=free, B=paid) (Pujol, 2010) 10
Figure 3 Traditional Sales Cycle 10
Figure 4 Freemium Sales Cycle 11
Figure 5 Transactions in Freemium (Pujol, 2010) 12
Figure 6 Elements of Freemium Business Model 14
Figure 7 Product Design - WTP vs Perceived Value 15
Figure 8 Perceived value of product - TLF 15
Figure 9 Perceived value of product - FLF 16
Figure 10 App Monetization Models 17
Figure 11 Freemium life cycle of upgrades (Kumar 2014) 18
Figure 12 Continuous Monetization Curve 20
Figure 13 Revenue based user segments 21
Figure 14 LinkedIn's network effects 26
Figure 15 Evernote - Users region wise 2014 28
Figure 16 Estimated revenues and users - Dropbox, 2013 (Viswanathan, 2013) 31
Figure 17 Conversion from free to paying user 34
Figure 18 User distribution by country 38
Figure 19 Age Distribution - All Users and Subscribers 38
Figure 20 Playcount Distribution - All Users and Subscribers 39
Figure 21 Playlists - Density Plot - All Users and Subscribers 39
Figure 22 Registration Unixtime - All Users and Subscribers 40
Figure 23 Friendcount Distribution 40
Figure 24 Eventcount Distribution - All Users and Subscribers 41
Figure 25 Shoutcount Distribution - All Users and Subscribers 41

Figure 26 CART Model with Cross validation 44
Figure 27 Receiver Operator Characteristic curve for the CART – CV Model (Test data) 45
Figure 28 CART Tree - Significant Variables 47
Figure 29 Key findings of the study 51
Figure 30 Business Model Canvas - LinkedIn 59
Figure 31 Business Model Canvas - Zynga 59
Figure 32 Business Model Canvas - Evernote 60
Figure 33 Business Model Canvas - Spotify 60
Figure 34 Business Model Canvas - Dropbox 61
Figure 35 CART Model with Penalty Matrix altered 61

List of Tables

Table 1 Summary features and statistics of freemium models 32
Table 2 Summary of key findings - firm wise 33
Table 3 Dataset Variable Description 36
Table 4 Categories of Data variables 37
Table 5 Summary Statistics of variables - All users and Subscribers 37
Table 6 Classification Matrix 42

1. Introduction

1.1 Birth of Free

The software and internet market ecosystem have undergone drastic changes in the past decade. Performance of products increased, user interfaces became friendlier, and prices decreased which led to software becoming ubiquitous. Internet and smartphone penetration increased at an astounding rate which led to the birth of new online distribution models: app stores and online software consumption models like Software-as-a-Service (SaaS) and online feedback models (Dellarocas, 2003). Along with this dramatic change in the industry, a new software business model called freemium (Anderson, Free: the future of a radical price, 2009) materialized.

Freemium, a portmanteau of the words "free" and "premium", is a business model in which a service or a product is offered free of charge, but a premium is charged for advanced features, functionality, or related products and services (Hayes, 2008) (Chang & Learmonth, 2009). This business model is widely used today in digital-based products like games, music streaming, software etc., where a free version of product is available for download and use, but to avail the advanced features of the product, the user needs to upgrade to the premium version or make in-app purchases. Freemium models are spreading quickly in the software industry, especially among Web start-ups (Miller, 2009).

About 65.2 million US consumers will pay to download an app on their smartphones in 2015. However, this represents only 35.8% of total smartphone users in the US (Protalinski, 2015). The preference for free, ad-supported apps and freemium apps is on the rise. Despite growth in mobile and tablet purchases there is a decrease in paid app downloads. The industry invariably is moving towards the freemium model to create and market its products.

The freemium model is popular in B2C products, successful ones being Zynga Farmville, LinkedIn, Candy Crush Saga, Dropbox, Spotify, Pandora, Last.fm, Flickr, Evernote to name a few. Even media companies like the New York Times and mobile payment companies like Paypal utilize the freemium model. The success of this model has been further validated by B2B internet products like Box, Splunk and Yammer.

'Free' attracts customers. Freemium provides a strong customer acquisition strategy for businesses. Even internet startups with low capital can choose this model instead of investing capital in traditional advertising, sales and marketing channels. If this model is tied with social networking and referral program, the effectiveness of customer acquisition is usually amplified as a free product is easy to recommend to a friend.

Freemium products typically have over 90% free users and only a very small number of paid users. It is only this small subscriber base that contributes to the firm's turnover. Hence, the challenge requires balancing two tasks: growing consumer base by offering a free service and maintaining premium services to incentivize upgrades in order to stay profitable. (Needleman & Loten, 2012)

1.2 Internet Businesses

Marketers have always used free products to entice customers to try and experience the product. Earlier, the business model was to convert most of the free product users into paid users. In today's Web 2.0 world, this has changed. The freemium business model relies on a large scale user base and a small percentage of its user base as subscribed or paid users.

The inherent characteristics of internet based businesses are aiding the popularity of freemium models. Firstly, the low or negligible marginal cost of production and distribution of digital products makes it commercially possible for businesses to give out free versions of the product to the customers. Secondly, software products are built in a modular fashion which lets the business control, group and lock features. Thirdly, digital products are *experience* goods, so by sampling the product users learn its features, attributes and functionalities before paying for it. Freemium models let the users try most of the feature-set of the product.

Freemium model reaches a large user base rather quickly and also act as an efficient marketing channel because of network effects (Pulkkanen & Seppänen, 2012). Being 'free' creates a strong virality hook.

1.3 Inherently Free and Perpetually Networked

In today's inherently free and perpetually networked internet culture, many businesses have seen value in marrying the freemium model with social networking. In freemium-based Online Social Networks (OSNs), users are very engaged due to social connections, activities and interactions. Strong network effects are in play and product referrals show high conversions. The user sees high value in the product due to his network which directly effects the willingness to pay (Wang, Chin, & Wang, 2011).

The network effect in freemium based OSNs increases the engagement with the product. In online social networks a set of friends is about 100 times more powerful in influencing a user to join a group than the same number of strangers (Hui & Buchegger, 2009). In enterprise social media like Yammer, recent manager and coworker activities influence employees to start or increase participation (Brzozowski, Sandhol, & Hogg, 2009).

1.5 Structure of Thesis

Section 2 reviews the literature on freemium models and section 3 discusses the freemium business model, its components and metrics. Section 4 formulates the research plan for the thesis. Section 5 analyses five successful freemium business models: LinkedIn, Zynga, Evernote, Spotify and Dropbox to understand crucial factors that determine success of freemium business model. Section 6 describes the empirical study conducted on a dataset of 79,033 users and attributes from a freemium music streaming site Last.fm to study the factors effecting the conversion of a free user to a premium user. Key findings of business model analysis and empirical study are discussed in section 7 with their implication on freemium business strategy. Section 8 concludes the research, reviews limitations and offers suggestions for future research.

2. Literature Review

In the ever changing landscape of internet businesses, stiff competition and rapid product life cycle, new business models and ways to organize operations are tried and tested often. Freemium is one of them and though little literature exists in this model's analysis, there are signs of increased literature being written now as this model becomes a novel way to create and capture value. The freemium concept dates back to the 1980s, but the word freemium was first used in 2006 (Wilson, 2006).

The crux of the business model is conversion rate which is the percentage of paid users of the product to the total users (free and paid) of the product. Depending on the business (B2C or B2B, niche or mass), product functionalities and cost structure, the conversion rate for breakeven of the firm varies. The conversion rate of 2-5% is common for most of the business (Shaw T. , 2010). There are references to 5% percent rule in these models (Seufert, 2014). Thus freemium is typically a feasible option when the use base is large. However, bigger question is monetization of users than scale of users.

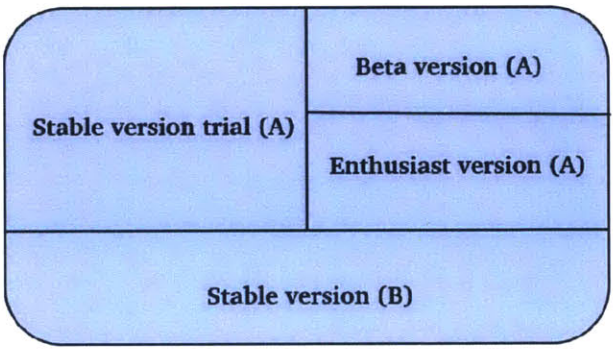
2.1 Types of Freemium

There are four differentiation strategies: quantity, feature, distribution or time (Pujol, 2010).

Quantity Limited: Quantity differentiation has been around the longest and has been achieved through product samples to accelerate or start sales of the product. A sample is the zero priced, quantity limited version of the complete, paid version of the product.

Time Limited: With the dawn of internet and digitized products and software came the time limited freemium model (TLF). The TLF model allows customers to access the complete version of the product for a set time period (1 week trial, 30 day trial). Once the trial period expires, the customer cannot access any features of the product and is required to purchase the product to use its features. Microsoft Office Suite is a good example of TLF model. It comes with a 60-day free trial.

Figure 1 Time differentiation in freemium (Pujol, 2010)

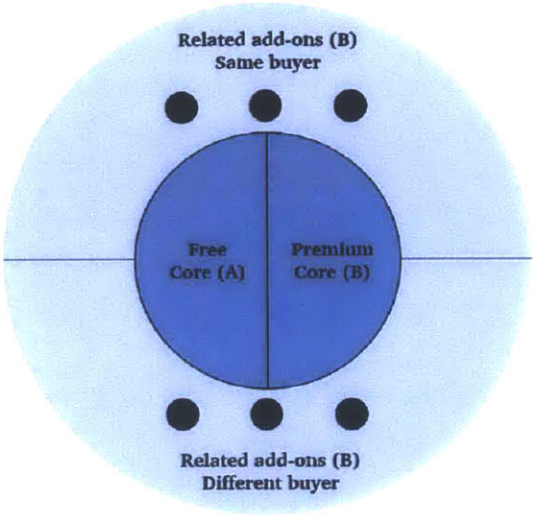


The TLF model led to launch of beta versions, enthusiast versions of a product (especially software) to a small set of customers like software developers which contains features that have not been time tested. The value comes from the quantity of time that it has been used, which reflects the reliability testing by the users. Customers pay for maturation of time (Pujol, 2010).

Feature limited: In feature limited freemium (FLF) models, customers use the fundamental core components for free (free core) and pay to use the advanced functionalities (premium core). Add on services or features can be released that promotes in-app purchases. These add-on features can be targeted at the existing user base or completely new users. Skype provides free voice chats and communication between PCs and text messaging but charges voice calls to landlines or mobile phone

numbers. This is a difficult model to design as it needs tradeoffs between expanding customer base and driving revenue stream.

Figure 2 Feature differentiated Freemium (A=free, B=paid) (Pujol, 2010)

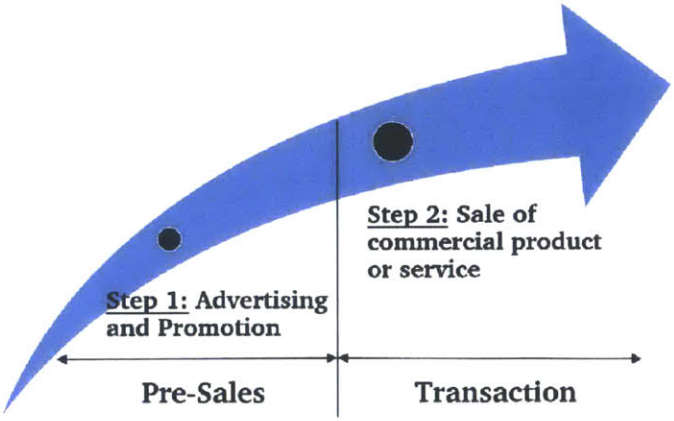


Distribution limited: In this model, the differentiation occurs in the mode of distribution. This is mostly seen in packaged softwares. For example, software discs come along with user licenses which state the use cases for the product, if it can be re-distributed or not. The software itself may be free to use but not to be re-distributed or may require a license fee for usage in commercial purposes.

Most commonly used freemium models in internet businesses are feature-limited freemiums (FLF) and time-limited freemiums (TLF) (Anderson, Free: the future of a radical price, 2009). Hybrids between FLF and TLF models also exist. For example LinkedIn lets its free users access the Job Seeker Premium for the first time free for one month. User then has access to premium features of higher number inmails, premium insights etc. However, at the end of the month, if the user chooses not to upgrade, he/she still has accesses to the basic version.

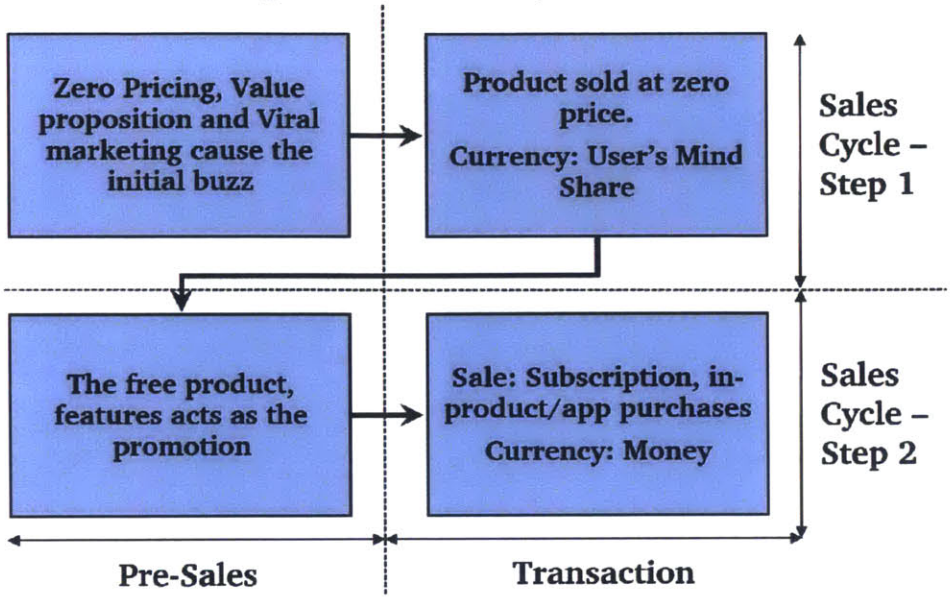
2.2 Freemium Sales Cycle

Figure 3 Traditional Sales Cycle



At first glance, freemium looks like an odd transaction where user gets the product for free with no requirement to purchase and the business doesn't receive payment (immediately). This would be a very cursory way of looking at this transaction. The transaction needs to be looked at in a longer time horizon and the currency of exchange need not be money. In fact, the transaction or sales cycle, can be split into two parts. In the first part, when the user begins the transaction by downloading or using the free product, he/she pays by an intangible currency: Mind share (Pujol, 2010). Mind share is awareness and recall of the brand in user's mind. In second part of the transaction, brand equity developed in user's mind acts as a promotional material or advert to promote purchase of the paid version of product.

Figure 4 Freemium Sales Cycle

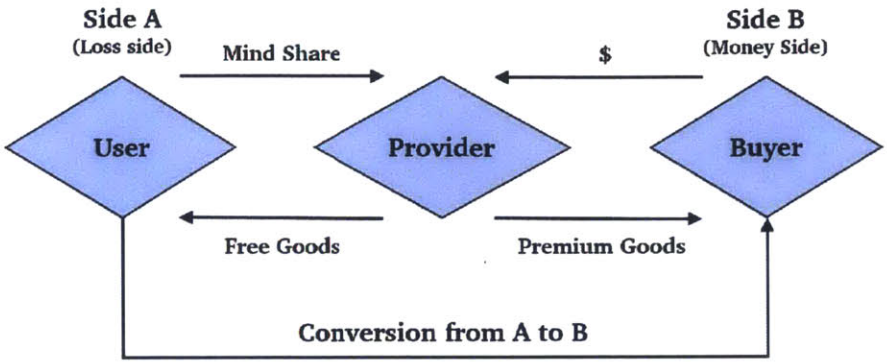


2.3 Freemium as a Multi-sided Market

Economics of multi-sided platforms (MSP) has emerged over the past decade and has quickly become a popular research domain in economics, strategy and technology. MSP can be defined as an organization that creates value primarily by enabling direct interactions between two (or more) distinct types of affiliated customers (Hagiu & Wright, 2011). Decision of each of the affiliated customer affects the outcome of the other set, typically through an externality (network effect). There are four kinds of MSPs (Evans & Schmalensee, 2007), Exchanges (ebay), advertiser supported media (google, newspapers), transaction systems (credit-card companies) and software platforms (adobe pdf reader and writer, smartphone OS and app stores).

At the outset, a freemium model does not look like a typical MSP, however a deeper understanding of the customers shows that the system behaves like an MSP. Early adopters and free users of the product, (represented by 'A' – Loss side) use and test the product, create a brand, enable viral diffusion of the product through referrals. There is a distinct set of customers (B – Money side) who use the paid, advanced functionalities of the product. The two sides have distinct behavior and needs and show interdependence. Free users benefit from the presence of the paid users for the sustenance of the business or product itself. The paid users value the fact that the early adopters and free users have tested the product, built the brand equity and gives them the confidence to pay and subscribe to the product. This interdependence mirrors the indirect network effects seen on platforms (Pujol, 2010).

Figure 5 Transactions in Freemium (Pujol, 2010)



There is growing literature on two sided markets and platform strategies which aim to accelerate the adoption of premium features by subsidizing the basic functionalities. In doing so and compared with charging both sides, the firms may profit more overall due to cross-side positive network effects (Niculescu & Wu, 2011). The literature tries to understand pricing mechanisms and identify the money side (Rochet & Tirole, 2003) (Rochet & Tirole, 2006) (Parker & Van Alstyne, 2005).

In FLF models, the free version is essentially a lower quality version, as it is limited by features or content or performance. There is literature understanding versioning of software products (Raghunathan, 2000), (Bhargava & Choudhary, When is versioning optimal for information goods, 2008) (Bhargava & Choudhary, Information goods and vertical differentiation, 2001), (Wei & Nault, 2005) , (Cheng & Tang, 2010).

Effects of freemium strategy in software market have been studied in the literature from early 2000s (Haruvy & Prasad, 2001) (Faugère & Giri Kumar, 2007) (Cheng & Tang, 2010). The freemium model is employed to stimulate product diffusion process (Jiang & Sarkar, 2009), brand building (Kapil & Robert, 2004) and positive network effects (Gallaugher & Wang, 2002). Increasing product visibility is a potent way of increasing demand in competitive market (Anderson, The Long Tail, 2004).

Freemium models let users experience the product before paying for the same. Product sampling and free demonstration in digital products help the user determine the value of the product (Chellappa & Shivendu, 2005). In a TLF model, the user experiences all the featured for a limited time. Hence, the user learns the true value of the product. In FLF models, based on the use of free functionalities, the user adjusts the value of the premium features and this can be lower than the true value of the product. Thus the business model chosen effects user's perception of the product value which in turn effects the revenue.

Analysis has been made on reduced uncertainty vs demand cannibalization and free trial duration in more recent literature (Cheng & Tang, 2010) (Cheng & Liu, 2010). They find that if network externalities are present, businesses should offer free trials over feature limited options. Optimal stopping time for free trial and profit comparison between time-limited and feature limited versions are also discussed.

Freemium models have many potential revenue streams: in-app purchases, subscription and advertising. Empirical analysis of freemium strategy for mobile apps on google play market has shown that offering consumers trial version of a mobile app for free is positively associated with higher sales rank and revenue of its paid version (Liu, Au, & Choi, 2012). With the deluge of apps in the market, offering a free version would increase visibility of lower ranking or newly launched app and increase the revenue of its paid version. When the free version of an app is offered, importance of review rating reduces, which may be attributed to the fact that consumers have the ability to test the app before purchase, significantly reducing uncertainty associated with the app, and hence dependence on user reviews. (Liu, Au, & Choi, 2012).

One stream of literature studies freemium as a possible revenue model and determines business strategies in different situations (Semenzin, Meulendijks, Seele, Wagner, & Brinkkemper, 2012). Comparison of revenue models of traditional industries with technology industries have been made and conclusion that freemium is an encouraging revenue model has been drawn (Teece, 2010). Difference between freemium revenue models for startup and established companies have been analyzed and results show that former drives most of their revenue from paid customers (Hung, 2010).

A study of 132 freemium music as a service (MaaS) customers was conducted to determine the influence of various features of the product on their willingness to pay (Dörr, Benlian, Vetter, & Hess, 2010). They conclude that sound quality and contract period significantly influence users' willingness to pay. In popular freemium content driven sites, the effect of the community and social network has been analyzed and connection between customer willingness to pay and community identified (Oestreicher-Singer & Zalmanson, 2013). One study used Dual Mediation Hypothesis to investigate whether free versions work as advertisements for premium versions within freemium models (Wagner, Benlian, & Hess, 2013). It shows that choosing a high premium fit, or in other words small functional difference would increase premium cognitions, which may lead to free users upgrading. If users are able to evaluate more facets of premium version, the persuasive effect is stronger. So, offering a trial period of premium version to the free users increases the probability of the free user upgrading (Wagner, Benlian, & Hess, 2013).

There is a growing literature in the field of freemium products and strategy. This thesis is an attempt to understand the model from successful case study perspective as well as though empirical study and add on to the literature.

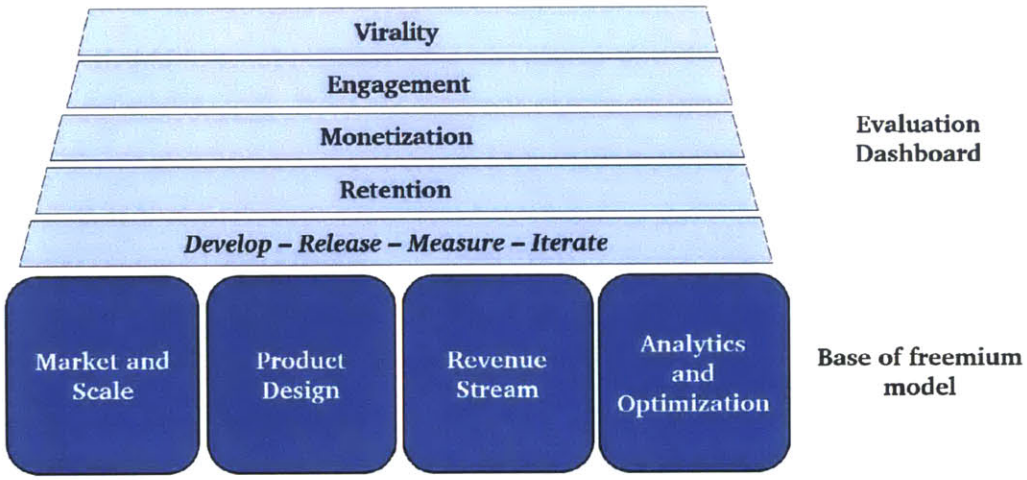
3. Freemium Business Model

Freemium model has become a dominant means to generate revenue on mobile devices and internet businesses. This section aims to understand the business model, its components, revenue generation, user engagement and growth.

3.1 Elements of Freemium Business Model

The design of a freemium business model hinges on four key elements that form the base of the structure: understanding of the market and its scale, product design, monetization and analytics and feature optimization. This structure and analysis is based on work by (Seufert, 2014) in this area.

Figure 6 Elements of Freemium Business Model



3.1.1 Scale

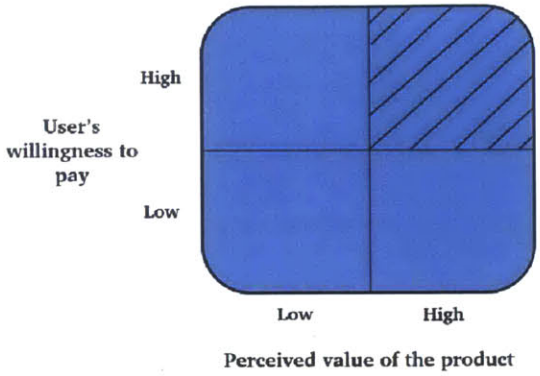
Review of literature on freemium economics have clarified that in typical B2C freemium products, paying customers form a very small portion of the total user base (2-5%). Hence the need for scale of the customer base is crucial for the success of the business. This does not mean that the scale of the user base is a compulsory condition for success, but that it creates conditions to generate more paid users. There are cases of freemium models operating in B2B businesses and niche businesses which do not have a high customer base. They succeed owing to higher conversion rate of a moderately sized, dedicated and highly engaged user base. However, in practice the proportion of users paying for premium features is very low and the business should prepare its revenue strategy and potential to scale its customer base accordingly. Thus the 5% rule is a design decision that businesses take to ensure the model accommodates realities of business. Hence, larger the user base, higher number of users pay for features and higher the revenue.

3.1.2 Product Design

There is burgeoning literature on lean methodology and lean startup which hinges on Build-Measure-Learn loop at its core. This iterative and high speed method of product development centers on Minimum Viable Product (MVP) (Ries, 2011). MVP is the "version of a new product which allows a

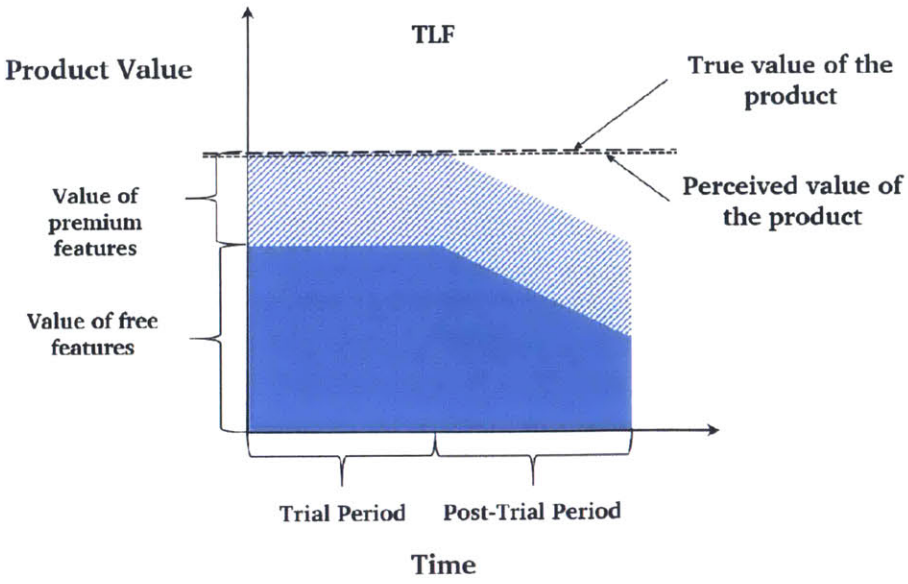
team to collect the maximum amount of validated learning about customers with the least effort" (Ries, 2011). This concept lets an agile organization release a basic or 'lite' version of the product, learn from the wisdom of the crowd and iteratively fine tune it. The cycle of Build-Measure-Learn helps the team achieve incremental benefits in product features. Develop-Release-Measure-Iterate cycle is another version of the lean methodology for freemium product design (Seufert, 2014). Freemium product design does not have a cookie cutter solution. The design can have multiple questions and hypotheses that needs to be tested in the market before creating a successful product. Hence the MVP concept and Develop-Release-Measure-Iterate cycle is of utmost importance to freemium businesses

Figure 7 Product Design - WTP vs Perceived Value



First critical question that needs to be answered during freemium product design is: *'How much value should the free product provide to consumers relative to the premium product?'* A better free product would encourage more to join its service, but also would cannibalize sales of the premium version, by reducing the likelihood of upgrading to the premium version (Lee, Gupta, & Kumar, 2013). Hence the design of features and product should be such that the customer sees value in the feature set, but also sees that the experience will be enhanced with premium feature set. When the perceived value of the premium product is high and it is matched by willingness to pay, users pay for the advanced functionalities. If the free feature set is very rich or poor, it effects the user's perception of the premium product. A rich feature set would make the user question if there is a need for premium version at all. Whereas a poor feature set doesn't give an optimal experience of the product and effects the perception of the advanced features. Striking this delicate balance can increase conversion rates.

Figure 8 Perceived value of product - TLF

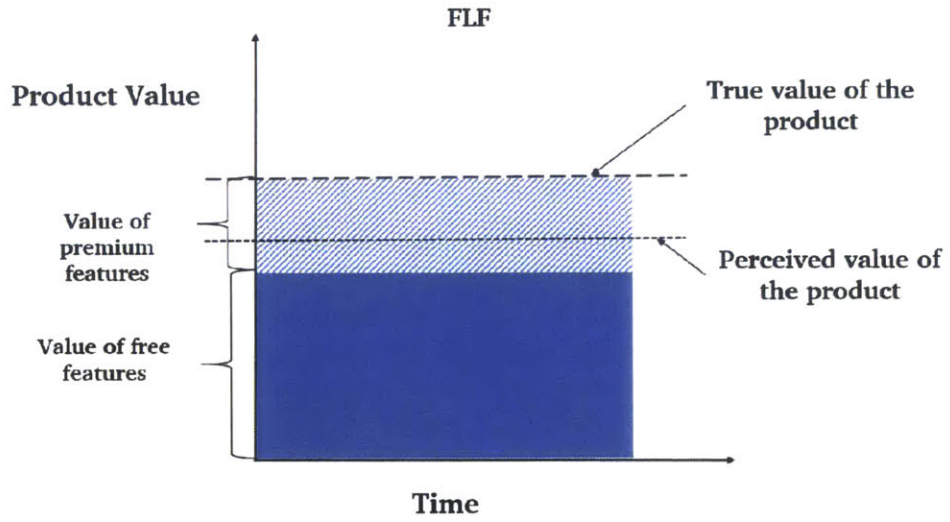


The next question that needs an answer is *'Do I need TLF or FLF or Hybrid freemium business model?'* To answer this, we first need to understand that digital goods are *experience* products. The TLF model lets the user experience all the feature sets available and hence the user's perceived value of the product is close to true value. However, if the

user does not purchase the product at the end of the free trial period, passage of time can reduce the perceived value of the product.

In FLF model, users experience basic feature set and hence the user's perception of the product value may not reach the true value at all. However, since the user always has an access to the free version, there is no decay in the perceived value of the product. Today, businesses employ a hybrid of TLF and FLF to rope in the benefits of both the models.

Figure 9 Perceived value of product - FLF



Finally, the monetization question comes into play: *'How do I monetize? Subscription? In-app purchases? Ads?'* Monetization plays an integral part in the product and feature design. Since this is a critical element in the model design, it is discussed in detail in section 3.1.3.

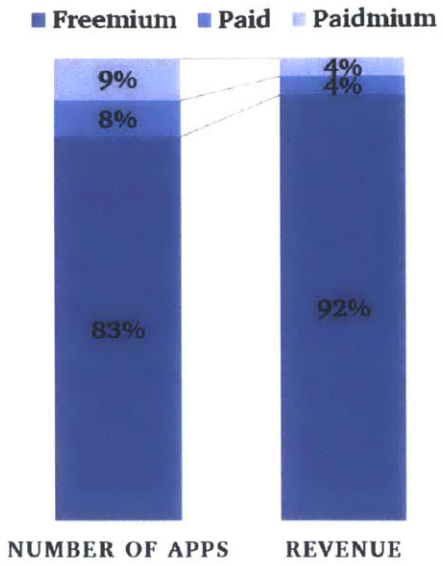
3.1.3 Monetization

Monetization is a tricky element to design in freemium businesses as the model relies on revenue streams from a relatively small proportion of the total user base. This is the central element around which the product is designed and features are optimized. Freemium and paid access are the two commonly used monetization models currently in mobile apps. 'Paidmium' is a relatively new monetization model where the user pays to gain access to the app and during the course of app usage, can make in-app-purchases (IAP). Freemium model is by far the most popular business model in mobile apps. However not all these apps have a successful monetization strategy. The difference between freemium and paid access revenue models is that in paid access, the user pays to gain access to the feature set of the product and may or may not fulfill user's needs and expectations. However in the freemium model, the user already has access to the basic feature set and is engaged with the product enough to pay and access the advanced feature set. This tends to enhance the user experience.

The paid access model has a fundamental barrier due to its direct cost, even if the cost is very small, which can deter the user from using the product at all. It has also been demonstrated that there truly exists 'penny gap': for users even one cent price has totally different psychological response compared

to free (Anderson, Free: Why $0.00 Is the Future of Business, 2008). In a paid access model, the user will pay only if his perceived value of the product is higher than the product cost and his disposable income is higher than product cost (Seufert, 2014). In case of freemium model the second condition of disposable income still holds good. However the first condition changes a little as the perceived value of the product tends to be high as they are already exposed to product features.

Figure 10 App Monetization Models



Appstore composition by business model
Worldwide, December 2013
Source: IDC and App Annie

Freemium model can be monetized in many ways. Advertisements in the free version of the product is a popular one. Pandora uses this model and when the user upgrades to the premium version, ads are removed and few advanced feature sets are provided. However more often than not, digital products have ads that can be annoying and intrusive and the business models depend on the 'annoyance' factor to upgrade (Learmonth, 2009). This is a wrong kind of inducement and drives the customers away to premium version. Ads if done right and targeted right, can add huge value to the product and can increase click through rates and ad revenue. But the current model of annoying ads does more harm than good to the product perception and ad revenue takes a hit. If the business model chooses ads as a monetization route, they need to be relevant and targeted rather than annoying and intrusive.

Another popular means of monetization in freemium is subscription. Subscription models can provide a steady stream of income than others, however the number of users willing to pay can reduce with continuous payment cycle. This can also induce customers to subscribe only for a short period of time when they need the advanced features and turn it off when they do not need.

In-app purchases are popular in gaming products where the user pays to buy virtual goods like upgrades or speed-ups, purchase additional functionality like levels or content, more time to use the app, pay to remove ads, or a combination of them all (Koetsier, 2014). Since in-app purchases are not as steady as the subscription model, it works when the user is very engaged with the product.

3.1.4 Analytics and Optimization

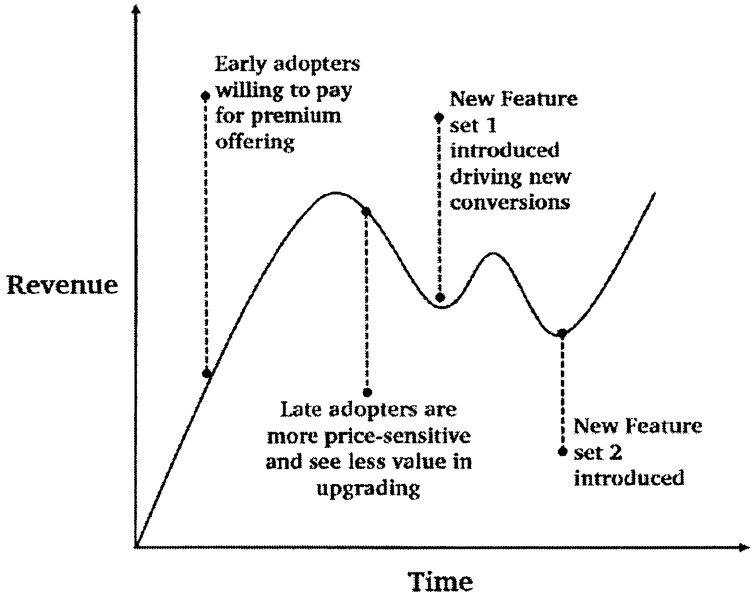
Data is the driver of freemium model. The design-measure-learn-iterate model relies on a systematic, quantitative understanding of user behavior. Data collected at every stage of product design, iteration and usage should be analyzed through an array of tools and procedures to optimize the product and its features. Target metrics for features and products set the benchmark for optimization. Data collection and analytics make up only a small part of the process. Picking up relevant data, making sense of it, comparing against key metrics, converting into insights and fine-tuning the features based on the insights forms the core of data analytics and optimization. This process of converting usage and

behavior data into product design features in an iterative manner can provide incremental benefits for the product and in turn increase user engagement and conversion rate.

3.2 Freemium Lifecycle

Freemium model does not typically generate a steady revenue stream. There is a rise and fall in revenue stream owing to the rise and fall in the subscriber base (Kumar, 2014). Product managers should be cognizant of this nature of revenue and accordingly introduce upgrades, updates and new features to trigger subscriptions by new users and more in-app purchases by existing customers.

Figure 11 Freemium life cycle of upgrades (Kumar 2014)



3.3 Metrics in Freemium Model

"What gets measured gets managed." – Peter Drucker.

The Minimum Viable Product (MVP) model is at the heart of freemium design and develop-release-measure-iterate is the engine driving the design. Product features are not introduced based on product manager's intuition but on methodical, systematic and quantitative analysis of data collected. Nirvana lies in using this data-driven iterative approach to find a sweet spot in conversion rate where the business is profitable. This section discusses metrics that can be used to achieve this goal.

3.3.1 Retention

Retention is a retroactive, time-based measure of product use (Seufert, 2014). In simple terms, customer retention rate is number of customers retained with respect to the total number of customers acquired at the start of the time period. New customer signups and acquisitions are not considered in this analysis. Retention metrics for a product are typically seen as day 1, day 3, day 7, day 28 values which essentially gives the percentage of users returning back to the product 1, 3, 7 and 28 days after

using the product for the first time. The intent of retention metrics is to understand length of product use before the user leaves or churns out of the product. Retention metrics are analyzed in cohorts of users and the resulting cohort analysis gives a rich set of information about frequency and length of product use. Actionable cohort analysis uses cohort date, identifiers and lifecycle stages or dates and gives cohort wise retention data that gives the ability to drill down to the users of each cohort to understand their behavior better.

Metrics

Daily new users (DNU) and Daily Active Users (DAU) are frequently used retention metrics. New user is any user who downloads or registers interacts with the product for the first time. Activity for measuring the DAU is a little ambiguous and depends on the definition of key activity by the business. DAU is the number of users interacting with the product on a given day and DNU are the new users acquired by the product. DNU and DAU can be both plotted and viewed on a line or bar graph. Retention chart tracking day 1, day 3, day 7 and day 30 coupled with full cohort analysis acts as a retention dashboard for the product team.

3.3.2 Monetization

Revenue optimization is critical to the success of freemium model and monetization metrics have to be tracked for the same. Once the team has clarity on the retention statistics of the product, emphasis needs to be shifted to monetization metrics. Metrics discussed here are not be looked at in sequential manner, but in a parallel fashion. However, user retention is pre-requisite for the functioning of the product and hence before delving into all metrics, user retention should be strengthened.

Conversion Rate

Conversion rate measures the number of paid users relative to the total user base of the product. It is usually shown as a percentage of users. This is a critical metric that the team must follow to ensure product profitability. The conversion rate in freemium products is very low, in the range of 2%-7% in most products. These users are referred to as 'converted users'. This metric is calculated over life time of the product and can also be calculated on a daily basis, cohort wise to see the variations over time which could mirror the changes in feature set. This metric can provide insights into the features and the way users respond to it.

Revenue Metrics – ARPU and ARPPU

Metrics that communicate money being spent by users are Average Revenue Per User (ARPU) and Average Revenue Per Paying User (ARPPU). ARPU captures a user's worth over the lifecycle. Lifetime ARPU provides insight into behavior of large user groups. However, considering the small percentage of paid users in freemium, 'average' metrics are not be the best indicator of distribution.

Daily ARPU is a better indicator as it can show the effect of feature changes or additions and its impact on the number of paid users. When tracked over time, daily ARPU becomes a valuable indicator. Lifetime ARPPU is a better indicator than the ARPU metric as this is evaluated over the base of total paying users. Daily ARPPU is a very contextual metric in freemium products as changes in features can

have drastic changes in the values owing to the relatively smaller denominator compared to daily ARPU. Together, these revenue metrics indicate the monetary health of the product.

However, average monetization metrics, do not provide keen insights into the user behavior and product behavior, especially in freemium models. Hence we need to look into few more monetization concepts – Lifetime Customer Value and Continuous Monetization curve to appreciate the complexity of the freemium monetary model.

LTV – Lifetime Customer Value

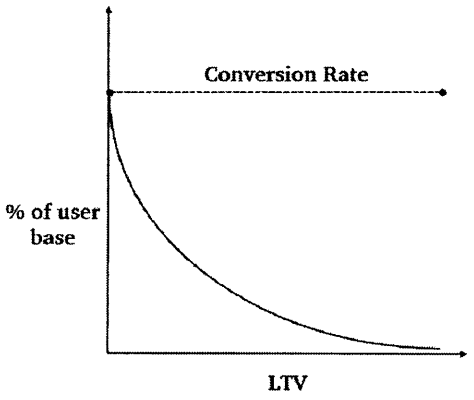
LTV is the total revenue that a user brings in over the course of a lifetime with the product. The LTV concept and equation was popularized in a 2006 article which defines LTV as the present value of all future cash flows attributable to a user for a given product (Gupta, et al., 2006). Simply put, LTV of paying users must be greater than the cost of acquisition and cost of servicing of all customers for the freemium model to be profitable.

LTV can be a complex metric to calculate in a freemium product as there can be multiple revenue streams – subscription and in app purchases. However, calculating LTV value as accurate as possible (with justifiable assumptions) helps increase paid user acquisition with a positive ROI and prioritize feature development to maximize product revenues (Seufert, 2014). LTV metric gives one metric that can help decide marketing spends in customer acquisition, choose feature sets that drive revenue as well as allocate resources between various products or product features.

The Continuous Monetization Curve

The concept behind the Continuous Monetization curve in freemium products is that the product catalogue is so complete that, at any given point in their usage time with a product, users are presented with a diverse and relevant set of potential purchasable items from which to choose (Seufert, 2014). A small catalogue essentially means that the user has fewer options to choose from and a large catalogue translates to higher combinations of monetization choices which are dynamic, personalized purchasable options. This increases the probability of a user making a purchase. The product breadth and appeal and the user's need are the bases used to draw the *continuous monetization curve.* This shows the product catalogue depth and possible monetization values. The curve involves plotting the percentage of user base and possible lifetime customer value.

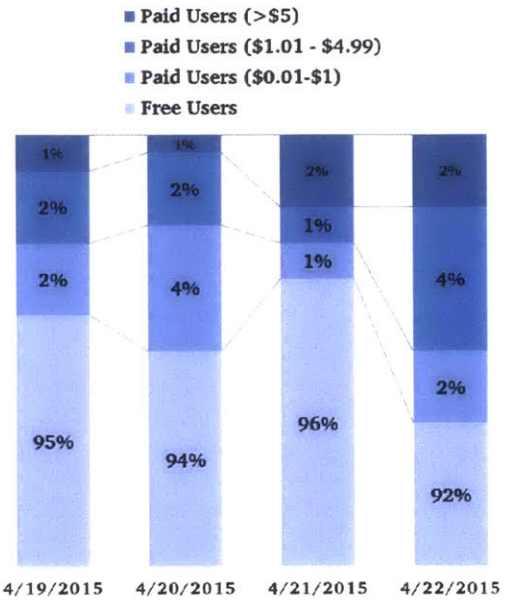
Figure 12 Continuous Monetization Curve



Revenue based user segments

The Continuous Monetization Curve is useful in design decision making. However, it is difficult to interpret the graph and hence does not work well as a reporting tool. A better way to use the same data and depict in a more interpretable format is by creating revenue base user segments and tracking it over time. Using a stacked bar charts to depict segments gives an intuitive way to view growth or deterioration of various user segments over the lifetime. Revenue generating segments can be divided into multiple segments to capture granular details of their behavior and this can be helpful if the monetization curve has a long tail. If the revenue graphs show lots of variation and volatility, it is a sign that the continuous monetization curve has not reached an optimal level of continuity and that the product catalogue needs some work.

Figure 13 Revenue based user segments



3.3.3 Engagement

Every digital product design team aims to make their product as *sticky* as possible to the user. A sticky product means higher engagement and this can translate to stronger revenue streams for business. Hence we need to capture engagement metrics which depict the user's behavior and interaction with the product. These metrics help the product team to craft an environment and experience that keeps the users engaged.

The ultimate goal is to have the user interacting with the product every day or even multiple times every day. This would mean that the product has a strong value proposition for the user and he derives genuine satisfaction from the features. This translates to more avenues and opportunities to monetize. Engagement metrics gives a pictures of how the users interact and also helps to curate those interactions.

The Onboarding Funnel

First session that a user spends with the product is very critical. This determines the user's lifetime with the product. Hence a great deal of thought needs to go into designing and optimizing the onboarding experience. In any freemium product—especially when massive scale is achieved through large-scale organic user base growth and aggressive paid acquisition—a percentage of new users will churn out during their first sessions (Seufert, 2014).

The onboarding funnel is a graph tracking key milestones in the user's onboarding journey (Issa, 2013). These are critical formative steps in the initial stage and can have high attrition rates if milestones have

high friction. The onboarding process familiarizes the user with the product and feature set and induces the user to sign up or register.

Firstly there should be a clear onboarding process and key milestones for the process. These milestones are the goals that need to be tracked in the funnel and conversion at each stage of the funnel calculated. This can be plotted on a graph which has onboarding milestones on the x-axis and users on the y-axis. Once the funnel is setup, problematic areas can be identified and optimized.

Session Metrics

Session metrics capture the length and frequency of a user's sessions. These metrics aim to understand user's interactions and engagement with features. Higher session length and higher frequency shows higher user engagement. Average session length can be calculated daily using total length of all sessions and number of sessions. But, like all average metrics for freemium, this is not very indicative as session lengths can differ greatly. Using median session length solves this problem to an extent. Another possible way is to calculate average session metrics for user segment based on revenues generated. Session metrics can also be analyzed cohort wise to see variations over time as the feature sets change. Session metrics should be evaluated along with retention and monetization metrics to derive insights and used to optimize features.

Net Promoter Score

Net Promoter Score is a simple, intuitive measure of customer loyalty. This is based on a principle that every product's users can be divided into one of the three categories: Promoters, Passives and Detractors. This category can be figured out just by asking one question: *How likely is it that you would recommend this product to a friend or colleague?* (Reichheld, 2003) This can be measured on a scale of 1 to 10 or -100 to +100. The question attempts to understand user's loyalty levels. This can be seen as a leading indicator of referrals or future in-app purchases or premium upgrade.

Promoters are extremely satisfied with the product and would love to refer it to their friends. This can help acquire new users, whereas a detractor or unsatisfied customer would spread harsh product experience. By taking a difference of these two groups, the business can derive a quantitative number for user loyalty.

NPS adds another qualitative metric to the dashboard which can be used as a signal of satisfaction and future referral. Low NPS or falling NPS should immediately shift attention to user's engagement and interaction with the product. It serves as a capable signal of how well the long tail of the monetization curve is developing (Seufert, 2014).

3.3.4 Virality

Virality, in freemium model refers to the tendency of the product to circulate rapidly and widely from one user to another, without any active marketing effort from the business. Five important aspects of virality are: propagation, network, speed, reach and self-sustainability (Feder, 2014). A passionate user can be a great tool in efficiently marketing the product through his networks and word-of-mouth

referrals. Hence virality is a big lever that the freemium product team can to rely on to reduce marketing costs.

Viral Hooks

The crucial aspect of embedding virality in the product is creation of *virality hooks.* A common way of implementing this hook is by integrating social networks in the product, where the user's activities and statuses on the product are updated directly on social networks. These hooks must be designed such that they are effective and traceable.

A superficial virality hook can be the integration of social network on to the product. Here the update on the network feed happens automatically when the user interacts with the freemium product. These hooks are difficult to audit or trace but takes little effort for implementation. Thus, tracking the effect of the viral loop in terms of new user registrations due to the hook is very difficult.

A deep virality hook on the other hand is the networking layer embedded in the product design itself. This is more engaging to the user and the auditability and effectiveness of the viral loop is high. Here referrals or posts are not as dis-engaged as in a superficial virality hook. The number of recipients of viral invites are fewer here, but the invites are personal and the probability of conversion is higher. Deep hooks are typically engineered into the product, like a feature set of the product and hence would require more effort and time for implementation. Though the production and implementation time of this hook is high, it produces actionable metrics and data that can be used to increase the virality of the product. In both kinds of hook implementations, viral invitations can be tracked and this will be an added arsenal in the metric toolkit.

4. Research Design

Based on the theoretical background and literature reviewed, there is an understanding of building blocks of freemium model. Next few sections try to dig deep into the model by analyzing successful freemium companies and by conducting empirical study of a freemium music company using CART model.

Business Model Analysis

Five successful freemium companies are chosen for analysis and the selected companies are small or medium sized and primarily internet based B2C products. The companies selected are LinkedIn, Zynga, Evernote, Spotify and Dropbox. The companies are analyzed by understanding their background, the value proposition and the Freemium business model. Business model canvas are drawn to discern the success factors.

Empirical Study

Second section of the research uses large dataset of user information (free and premium) from Last.fm and five sets of hypotheses on likelihood of conversion into a subscriber are written. This is a binary predictive classification problem and CART model will be developed using training dataset and validated on test dataset. The model will predict the likelihood of conversion of a free user to premium user and will assess the significance levels of demographic, engagement, retention and social features.

5. Business Model Analysis

This section of the thesis examines few successful freemium-centric companies and attempts to determine their key success factors. Each firm will be analyzed separately as each business operates in different market, addresses different consumer and has a distinct value offering. The business model canvas (Osterwalder, 2004) tool is used to analyze the firms and are included in the appendix. Finally, the aim is to highlight the factors that prevail across the case studies and determine freemium model success.

5.1 LinkedIn

5.1.1 Company Background

Reid Hoffman and four co-founders launched LinkedIn in May 2003 in Mountain View, California as a website for career management and professional networking activities (LinkedIn files for IPO, 2011). Within a month of its founding, over 4500 members joined LinkedIn (DeBaise, 2009). The company's revenue stream was through advertisements in 2004 and job postings and premium subscriptions in 2005. By early 2007, LinkedIn had nearly 10 million subscribers.

LinkedIn received significant funding in its formative years. Sequoia Capital, Greylock partners, Bessemer Venture partners, European Founders Fund, Bain Capital and Goldman Sachs were some of the prominent venture capitalists who invested in the growing business. By June 2007, LinkedIn was valued at $1 billion (Buckman, 2008). Since its early days, LinkedIn was global in nature and launched the site in six languages and currently is available in sixteen languages. In January 2011, LinkedIn filed for an Initial Public Offering (IPO). The company acquires small companies in the space to strengthen and broaden its offerings. LinkedIn currently has more than 300 million members and has offices around the world.

5.1.2 Value Proposition

Jeff Weiner, the CEO of LinkedIn describes LinkedIn's member value proposition as below:

"LinkedIn is focused on three primary value propositions: Professional identity, Network and Knowledge. Professional Identity: We want to enable people to represent their experience, skills, and ambitions via LinkedIn profile. Network: Connect all of the world's professionals and Knowledge: Be the definitive publishing platform (Yoffie & Kind, 2013)."

5.1.3 Business Model

LinkedIn is a multi-sided platform operating on a freemium model. LinkedIn earns revenue from three streams: talent solutions, marketing solution and premium subscriptions (Brochet & Weber, 2012). Members, businesses and advertisers form the three sides of the platform.

Talent Solutions: LinkedIn provides recruiters with access to profiles and contact details of members who are looking for jobs. These services are sold on a subscription basis. Recruiters can also post job listings on the website for $195 per posting or $1250 per 10 postings (LinkedIn files for IPO, 2011).

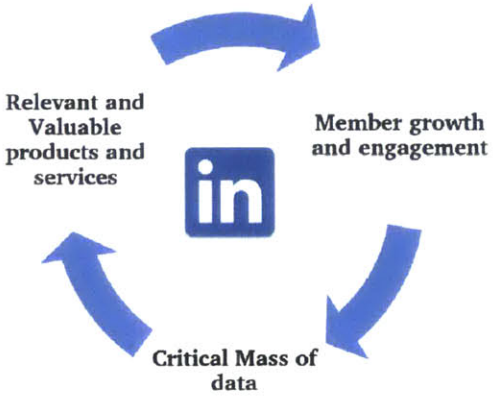
Marketing Solutions: LinkedIn provides advertising space on its webpage and targeted ads to members.

Premium Subscriptions: LinkedIn provides free basic service to all its members and advanced feature sets for a subscription fee. These features include enhanced search tools and results, ability to communicate with recruiters and improved customer support. These subscription options are provided to businesses too. Individuals and businesses are provided three different levels of premium subscriptions (Brochet & Weber, 2012).

Talent solutions provide the biggest stream of revenue and represented 55% of the revenue in 2013, followed by marketing solutions at 25% (LinkedIn Quarterly Earnings, 2013).

Deep Nishar, SVP of products and user experience for LinkedIn notes that "We provide the majority of our solutions to members at no cost. We believe this approach provides the best way to continue to build a critical mass of members, resulting in beneficial network effects that promote greater utilization of our solutions, higher levels of engagement and increased value for our members"

Figure 14 LinkedIn's network effects



5.2 Zynga

5.2.1 Company Background

Originally incorporated in January 2007 as Presidio Media, Zynga initially created games that were adaptation of popular board games. In September 2007 Zynga introduced an adaptation of a popular poker game: Texas Hold'em and released it on Facebook platform. Zynga offered innovative monetization opportunities by selling its virtual good: a set of Poker chips, immediately making the company profitable. Zynga built on its portfolio of games which used virtual currencies and in-game purchases. Zynga offered the broadest array of games to the social network and mobile platforms with almost 40 games on Facebook alone.

Mark Pincus, the founder secured funding from Union Square Ventures and several other sources starting early 2008. Zynga began trading on NASDAQ in December 2011 and generates US$873.266 million (Zynga Inc - Form 10K, 2013) and currently boasts of 240 million Monthly Active Users (MAUs) and 20 million Monthly Active Users (Zynga-AppData).

5.2.2 Value Proposition

The value proposition of Zynga is to make playing an integral part of the way people experience their relationship with other people on the Internet, and become the dominant player in that market.

5.2.3 Business Model

By 2010, Zynga was a highly successful freemium based gaming company with estimated revenues of nearly $50 million per month and its flagship 'Farmville' sporting more than 80 million monthly active users (Zynga Revenue, 2010) (Piskorski & Chen, 2013). Zynga's secret sauce for success was combining an engaging game with social network and in-game monetization. The key element of Farmville's economics is the virtual currency purchases in-game rather than purchasing the items directly. Zynga used this method to showcase the value of premium content by allowing users to collect limited amounts of in-game currency during gameplay without spending real cash. A satisfied user experience here promoted actual payments repeatedly. (Shaw T.)

Social Gaming

Farmville featured a number of characteristics that encouraged social interactions between players and online friends. Inviting facebook friends to the game created social bond which kept the players engaged and increased frequency and length of logins. Games involved chatting and were designed to take advantage of the 'social graph' or a map of people and how they are related (Chang & Mendelson, 2010). To quote Pincus, "Gaming is fundamentally a social experience, not a single player experience and not a technology experience. We are bringing gaming back to its roots." (Chang & Mendelson, 2010)

Virtual goods and currencies

Zynga encouraged playing and earning of virtual currency by completion of jobs in the game and then redeem currency for virtual goods. Inviting friends to the game rewarded users with virtual currencies. Users could also buy additional functionalities using credit cards or paypal.

Viral features

Farmville had an embedded virality hook which asked users to post or broadcast their activities to all their friends on facebook and these posts contributed to the game's popularity. Supported by these viral features, the game became an instant hit, amassing 10 million downloads within days of being released. Within a month Farmville became the eighth most popular game on facebook (Piskorski & Chen, 2013).

Monetization

Right from the company's inception, Pincus had set three underlying goals for Zynga: reach, retention and revenue. Monetization relied on three streams: affiliate offers, virtual good sales and advertising (Piskorski & Chen, 2013). Zynga offered virtual currencies to a player for signing up for an affiliate offer from another company. Zynga was paid for every such signup by the affiliate. Virtual goods provided 90% of the revenue in 2010. 0.25% to 5% of players bought virtual goods depending on the game (Piskorski & Chen, 2013). Advertising was a small revenue stream.

5.3 Evernote

5.3.1 Company Background

Evernote is a suite of software and services that allows users to capture, organize, and find information across multiple platforms (Evernote - Crunchbase, 2015) by means of note-taking and archiving. A *note* can be a piece of formatted text, a full webpage or webpage excerpt, a photograph, a voice memo, or a handwritten "ink" note (Evernote - Wiki, 2015). Evernote is an independent, privately held company and founded in 2007 by Stepan Pachikov. In early 2008, Evernote introduced invitation-only private beta version. The strength and features of beta product led to twelve times the target beta users to sign up. The team iterated and fine-tuned the product in a closed beta phase and in June 2008 released an open beta version. Since its founding, it has received $290 million in 12 rounds from 15 investors (Evernote - Crunchbase, 2015).

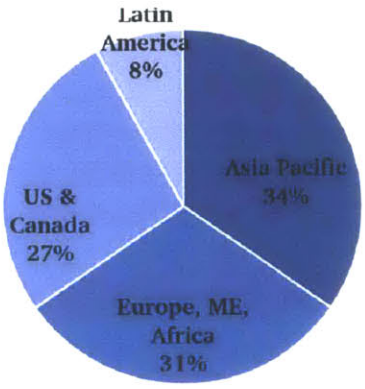
5.3.2 Value Proposition

The product is positioned as a tool to help the user *remember everything*. This is achieved by *capturing anything; accessing it anywhere* be it phone, tablet or computer; finding *things fast* by searching keyword, tag and images and *presenting*. Evernote is ideal for jotting notes, clipping web pages, taking camera phone snapshots, creating to-do lists, and recording audio notes. Once in Evernote, notes are analyzed and made searchable, even text within images.

5.3.3. Business Model

Evernote is a freemium note-taking productivity software. It provides free version and two versions of premium functionalities – Evernote *premium* with enhanced search features and offline access to notes on mobile at $5 per month per user and *Evernote Business* with collaboration features for $10 per month per user.

Figure 15 Evernote - Users region wise 2014



Clear, simple and intuitive product features, cloud infrastructure and unequaled image search technology has helped the user base cross 100 million by May 2014 (Libin, Evernote Blog, 2014). Evernote launched few more productivity related softwares and currently the user base is distributed very well across the globe. Internationalization is a key element of their strategy as the product is fully localized in 14 different languages, android app in 17 languages and a dozen more language customizations in progress. The company didn't spend on any marketing, SEM or advertising (Libin, Freemium for Consumer Internet Business, 2010).

High Retention

Evernote has high conversion rate in customers who have lasted longer and see value in the product. 0.5 percent of people who sign up in a given month go premium, but 2 percent of people who signed up a year ago upgrade to premium features (Gannes, 2010). The range of conversion rate is 0.5% - 6.0% depending on the user cohort (Westaway, 2013).

Virality and Network Effects

Evernote has none of the social networking aspects of Zynga, Twitter or Facebook which led to their viral spread. As the company CEO, Phil Libin, likes to say, "Evernote is anti-social" and that it was created "for you, not your friends" (Little, 2011). However, the product has referral feature and a highly satisfied refers it to a friend. Over 90% of the growth comes from word-of-mouth-referrals. This is not a product whose success is based on network effects, but essentially unique value proposition and excellent execution of the same. The product is the company's message and trump card.

5.4 Spotify

5.3.1 Company Background

Spotify is a commercial music streaming service that provides digital rights management-restricted content from record labels. Daniel Ek and Martin Lorentzon launched Spotify in 2008 in Sweden and several other European markets after announcing deals with all major labels and several independent companies that allowed Spotify to stream their catalog (Spotify-Blog, 2008).

Spotify launched it through the controlled beta version and it was free for all users. The company periodically sent "invites" to their existing users that they could use to invite their friends. Spotify lifted these limitations in 2009 in UK and proceeded to do the same in most of the markets in the following months.

Spotify was launched in the United States when it already had 10 million users throughout Europe. Spotify grew quickly and by March 2013, it claimed 1 million paid users in the US and 6 million worldwide. Internationalization is a key part of their strategy and by 2014, they had presence in 56 countries spanning the Americas, Europe, Asia and Oceania.

5.3.2 Value Proposition

"Spotify, it's easy to find the right music for every moment – on your phone, your computer, your tablet and more."

5.3.3. Business Model

Freemium

Spotify offers two service versions in the United States: free and premium. Both these versions offer on-demand unlimited listening of all the songs in the Spotify catalogue. Free service had their playback interrupted every five songs by a short audio advertisement. Premium users do not have ads, have access to high quality audio and enjoy other features.

When it introduced the offering in the United States, it offered six months of free premium use to all US customers. After the trial expiration, users could enroll in a paid subscription or start using free, ad supported version of the service. Spotify has a conversion rate of 25% (6 million paid users in 24 million total users) which is one of the highest in the industry.

Social and Viral features

Spotify has several social features where a user can follow their friends, follow artists, bands, integrate their account with twitter and facebook and automatically broadcast their Spotify listening activity on social network. Users can save and share their playlists and non-users could click and listen to these playlists by registering on Spotify.

5.5 DropBox

5.3.1 Company Background

"It's hard to imagine Tom Cruise in Minority Report sending himself files via Gmail or lugging around a USB thumbdrive." – Drew Houston, CEO, Dropbox.

Drew Houston got the idea for Dropbox while waiting for a bus at Boston in December 2006 when he had forgotten his USB flash drive back home. Soon Houston recruited Arash Ferdowsi, a student at MIT and the pair spent the next four months coding a prototype (Eisenmann, Pao, & Barley, 2014). The company was incorporated in June 2007.

Dropbox was a late entrant to the fiercely competitive online backup and cloud storage space. By mid-2007 the tech blog Mashable published a list of more than 80 online backup and storage services (Aune, 2007). However, the founder's vision was to build a single version of Dropbox that would be targeted to both consumers and business users and to use freemium pricing model. Bulk of the players in 2007 were in the enterprise storage solution space. A private beta program was rolled out to a limited group of users and these early adopters provided insights into user behavior and feature changes. In September 2008, Dropbox publicly launched its Windows and Mac clients and added a Linux version. Dropbox has received $1.1 billion in investments in 6 rounds from 24 investors. The company has 300 million users, 70% of which are international (Lynley, 2014).

5.3.3 Value Proposition

Dropbox for consumers is known for its ease-of-use and reliability. Dropbox for Business enables users to share the right files with the right people; put documents, photos, and other digital products into a central location and work on or share the items from any connected device (Crunchbase - Dropbox , 2014).

5.3.3. Business Model

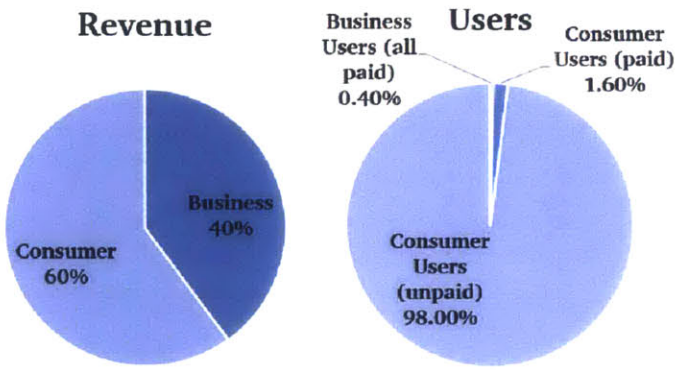
Freemium

Dropbox uses a freemium business model, where users are offered a free account with a set storage size and paid subscriptions for accounts with more capacity. All basic users are offered an initial 2 GB of free online storage space (Wikipedia - Dropbox, 2015). An estimated 1.6 to 4.0% of its users provided revenue to the company (Westaway, 2013). 2GB solution is offered for free with paid tiers of 50GB for $9.99 per month and 100GB for $19.99 per month.

Enterprise solutions

Dropbox initially decided to avoid red tape and concentrate only on the individual consumer market rather than corporate clients. Houston's goal was to get individuals to use and like Dropbox so much that they, in turn, got their employers to sign up as well. Only after five years of individual customer acquisition, Dropbox began to target corporate customers (Teixeira & Watkins, 2014). Dropbox for business became another company success with roughly 40% of annual speculated $400 million revenue came from corporates (Vishwanathan, 2013).

Figure 16 Estimated revenues and users - Dropbox, 2013 (Viswanathan, 2013)



Marketing

Early on, Dropbox attempted to acquire new customers through paid search advertising. However, it cost Dropbox more than $300 to acquire one paying customer. The team experimented with display ads and affiliate programs. Even these programs yielded unacceptably high customer acquisition costs. The team decided to give these marketing methods and instead look at other ways to grow the user base organically.

Referrals

Word-of-mouth-referrals and viral marketing efforts rather than paid advertising led to the acquisition of the vast majority of users. The two sided incentive structure in the referral program, viral features like password-protected shared folders encouraged users to refer Dropbox (Eisenmann, Pao, & Barley, 2014) and these generated strong results. By April 2010, 35% of Dropbox's recent signups, 35% came from the referral program and 20% from shared folders and other viral features. In 2010 four million existing users produced 2.8 million direct referral invites, a referral rate of 70%.

Analytics

Houston invested in analytics and optimizing customer acquisition efforts. The team believes strongly in tracking the user metrics cohort wise. They relied on the Startup metrics for Pirates: AARRR (Acquisition, Activation, Retention, Referral and Revenue) developed by investor Dave McClure (Eisenmann, Pao, & Barley, 2014). The team heavily used A/B testing to fine-tune page layout, content and feature set. The company launched "Votebox" on its site, which allowed users to vote and comment on new features to be added.

5.6 Summary

Key statistics and features of the five case studies discussed are tabulated in table 1.

Table 1 Summary features and statistics of freemium models

Company	Product	No. of paid tiers	Free tier freemium	1st paid tier features	Conversion Rate	Revenue Source
LinkedIn	Networking tool	3	Basic Profile. Limited mail	More access to user info. More communication services	0.8% [1]	23% ads, 77% talent solutions and premium member
Zynga	Online game	Multiple	Basic features of game.	In app purchases, virtual currency to access levels	2.1%[2]	10% ads, 90% in app purchases
Evernote	Note-taking tool	1	60MB/per month	1GB/month, Communication services, use on mobile devices	1.0 – 6.0%[3]	100% subscribers
Spotify	Online radio	3	Unlimited streaming with ads. On computer	No ads. Mobile streaming. Offline play.	20%	17% ads, 23% subscribers[4]
Dropbox	Cloud storage	4	2 GB storage	50 GB storage	1.6 – 4.0%	100% subscribers

Analysis of the five firms shows common factors that determined their success. These factors are listed in table 2 and the five companies' strength in each one of them is denoted by the color in the table. Darker the color, higher is the firm's competency in the factor. The table gives insights into specific strategies adopted by firms to chart their way to success.

[1] Forbes Online, 'Where To Find French Designers And German Engineers. Hint: It's Not On LinkedIn', Posted on October 13, 2013. Accessed at http://www.forbes.com/sites/realspin/2013/10/13/where-to-find-french-designers-and-german-engineers-hint-its-not-on-linkedin/

[2] Gamesbrief, 'Zynga's conversion rate' by Lovell, Nicholas, posted on September 23, 2014. Accessed at http://www.gamesbrief.com/2014/09/zyngas-conversion-rate/

3 Phil Libin, Keynote address – Founder Showcase, 2011. Accessed at: https://vimeo.com/11932184

4 Daniel Ek, 'spotify: No web freemium business like us.', The Wall Street Journal, posted on July 31, 2013. Accessed at: http://blogs.wsj.com/digits/2013/07/31/spotify-no-web-freemium-business-like-us/

Table 2 Summary of key findings - firm wise

Company	MVP Launch & Interate	Organic Acquisition	Focus on retention	Multi platform presence	Evangelism	Social features	Data driven method	Broad product catalogue	Compeling migration path	Multi sided monetization	Continuous Innovation
LinkedIn											
Zynga											
Evernote											
Spotify											
Dropbox											

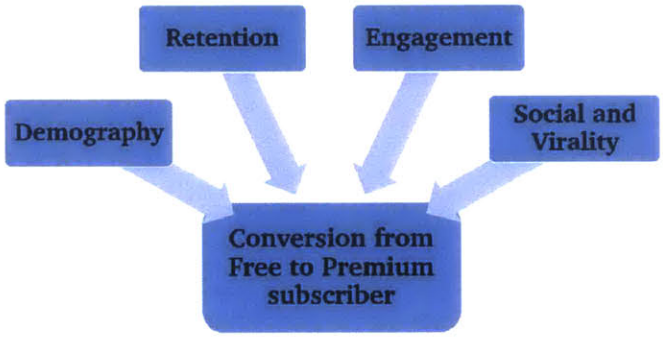
6. Empirical Study

In this section, predictive analytics on Last.fm user dataset is performed to analyze significant variables that drive conversion from free users to premium users.

6.1 Research Plan

Factors that convert a free user to a premium user can be divided into four categories: demographic factors, factors indicating engagement with the product, factors indicating retention and factors indicating virality and social features of the product.

Figure 17 Conversion from free to paying user



Research and data analysis begins with five hypotheses. The first hypothesis deals with demographic variables like the age, gender of subscribers. Considering that users would all belong to a specific target segment, demographic factors will be associated with low propensity to subscribe. Accordingly hypothesis 1 is as below:

> ***Hypothesis 1:*** *Demographic factors will not have a strong association with the likelihood of a free user subscribing to premium services.*

Second hypothesis addresses the retention factors of the product. A user who is invested in the product for a long time tends to have higher perceived value of the product and in turn higher willingness to pay. Hence, hypothesis 2 is as below:

> ***Hypothesis 2:*** *Retention is positively associated with the likelihood of a free user subscribing to premium services.*

Third and fourth hypotheses deals with engagement and social factors of the product. A highly engaging product with social and viral features leads to users engaging with the product for longer durations. The stickiness factor arising due to these features will be associated with the propensity to subscribe.

Also, the social and community factors effects the conversion to premium user. The hypotheses 3 and 4 are as below:

Hypothesis 3: *Engagement is positively associated with the likelihood of a free user subscribing to premium services.*

Hypothesis 4: *Social features (Virality) is positively associated with the likelihood of a free user subscribing to premium services.*

The next set of hypotheses compares the different factors or categories of variables.

Hypothesis 5a: *Retention factors will have a stronger association with the decision to upgrade to a premium subscriber than demographic factors.*

Hypothesis 5b: *Engagement factors will have a stronger association with the decision to upgrade to a premium subscriber than retention and demographic factors.*

Hypothesis 5c: *Social factors will have a stronger association with the decision to upgrade to a premium subscriber than engagement, retention and demographic factors.*

6.2 Last.fm

Data for this research was collected from Last.fm, an online music radio site that has social community aspect built into it. Last.fm is a music website, founded in the United Kingdom in 2002. Using a music recommender system called "*Audioscrobbler*", Last.fm builds a detailed profile of each user's musical taste by recording details of the tracks the user listens to, either from Internet radio stations, or personal computer or tables or mobile devices. This information is transferred ("scrobbled") to Last.fm's database either via the music player itself (Rdio, Spotify, Clementine, Amarok, MusicBee) or through a plugin installed into the user's music player (Wikipedia - Last.fm, 2015). Last.fm was purchased by CBS for $280 million in 2007.

6.1.1 Features

Last.fm has a host of features in its product that tries to keep the user engaged in it. For this data analysis, we use few of the total features to determine significant variables driving premium subscription.

User Accounts: Free user account gives access to all the main features of the product, whereas the premium users (€3 per month) get access to post on Last.fm forums, send and receive private messages and use the music player.

Profile: A Last.fm user can develop their music profile by listening to music on personal device with audioscrobbler plugin or on lastfm player. The user can update profile picture, use the "shoutbox" to update public messages or shouts. Profile page shows their friends, musical neighbors, favorite tags, groups and events.

Events: This functionality lets the user create and edit music events in any area. Users can set their status as *attending* the event. Reviews and photographs of the event can be added to the page and the event page can be linked to the performing artist page too.

6.3 Data Collection and Preparation

6.3.1 Last.fm API

Last.fm API allows anyone to build their own programs using Last.fm data, whether they're on the web, the desktop or mobile devices. The data was collected strictly adhering to the guidelines of API usage.

Data collection involved two separate steps. Firstly, a web crawler was specially programmed to collect list of userids by using group.getmembers() and user.getFriends() methods and recursively using these userids to find friends in their network. This method yielded 79,033 userids. Finally, a web crawler was programmed that used 79,033 userids to get specific user profile information. Following methods were used to procure the user information: user.getInfo(), user.getPastEvents(), user.getEvents() and user.getShouts().

The final dataset consists of 79,033 observations of all users with 13 variables. Of these users, 4217 are subscribers or premium users. Subscribers make up 5.33% of the total users which is close to typical conversion rates on digital freemium products.

6.4 Data Description

6.4.1 Dataframe

Quick details of the variables in the dataframe are given below:

Table 3 Dataset Variable Description

Variable	Type	Description
$ id	Factor	Unique identifier associated with the user
$ country	Factor	Two character ISO2 codes for the country of the user
$ age	Integer	Age of the user
$ gender	Factor	Factor variable with three possible values: F/M/N (Not disclosed)
$ subscriber	Factor	Factor variable that takes the value of 0 if the user is a free user and 1 if the user is a premium subscriber
$ playcount	Integer	Number of songs the user has played since joining lastfm
$ playlists	Integer	Number of playlists the user has created
$ isimageincluded	Factor	Factor variable that takes the value 'True' if the user has included the image in his/her profile and 'False' is not
$ unixtime	Integer	The Unixtime value of user registration date and time
$ friendscount	Integer	Number of friends the user has on Spotify

$ eventscount	Integer	Number of events that the user plans to attend
$ pasteventscount	Integer	Number of events that the user attended in the past
$ shoutscount	Integer	Number of shouts made by the user on the lastfm account

6.4.2 Variable Categories

Data variables mentioned above can be classified and placed into one or more categories: Demographics, Engagement, Retention and Social/Virality.

Table 4 Categories of Data variables

Variable	Demographics	Engagement	Retention	Social/Virality
$ age	✓			
$ gender	✓			
$ playcount		✓		
$ playlists		✓		
$ isimageincluded		✓		
$ unixtime			✓	
$ friendscount		✓		✓
$ eventscount		✓		✓
$ pasteventscount		✓		✓
$ shoutscount		✓		✓

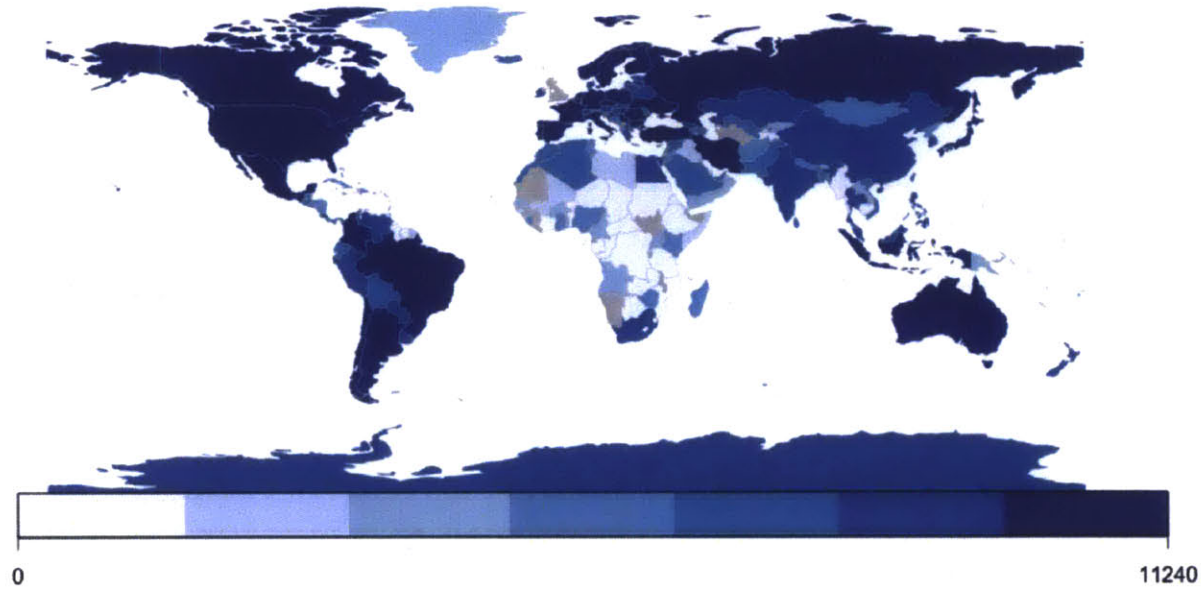
6.4.3 Summary Statistics

Table 5 Summary Statistics of variables - All users and Subscribers

Variable	All user Mean	All user Median	Subscriber Mean	Subscriber Median
$ age	27.34	25	35.38	32
$ playcount	65,100	28,400	88,950	59,440
$ playlists	1.159	0	7.724	1
$ unixtime	1.252e+09	1.246e+09	1.210e+09	1.200e+09
$ friendscount	128.5	40	126.6	52
$ eventscount	0.4122	0	1.206	0
$ pasteventscount	16.81	2	44.16	9
$ shoutscount	166.4	20	382.9	41

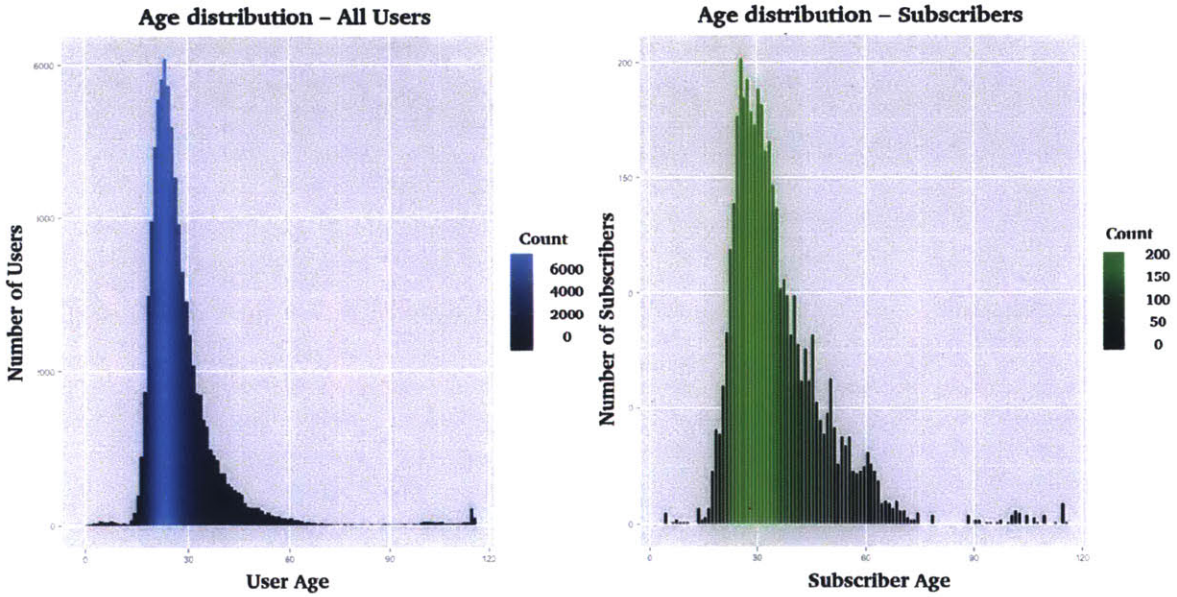
Country: The dataset has been randomly sampled and the users are distributed all over the world. The total country of origin of users in the dataset is 197 countries. The density of distribution of users country-wise is shown in the below map.

Figure 18 User distribution by country



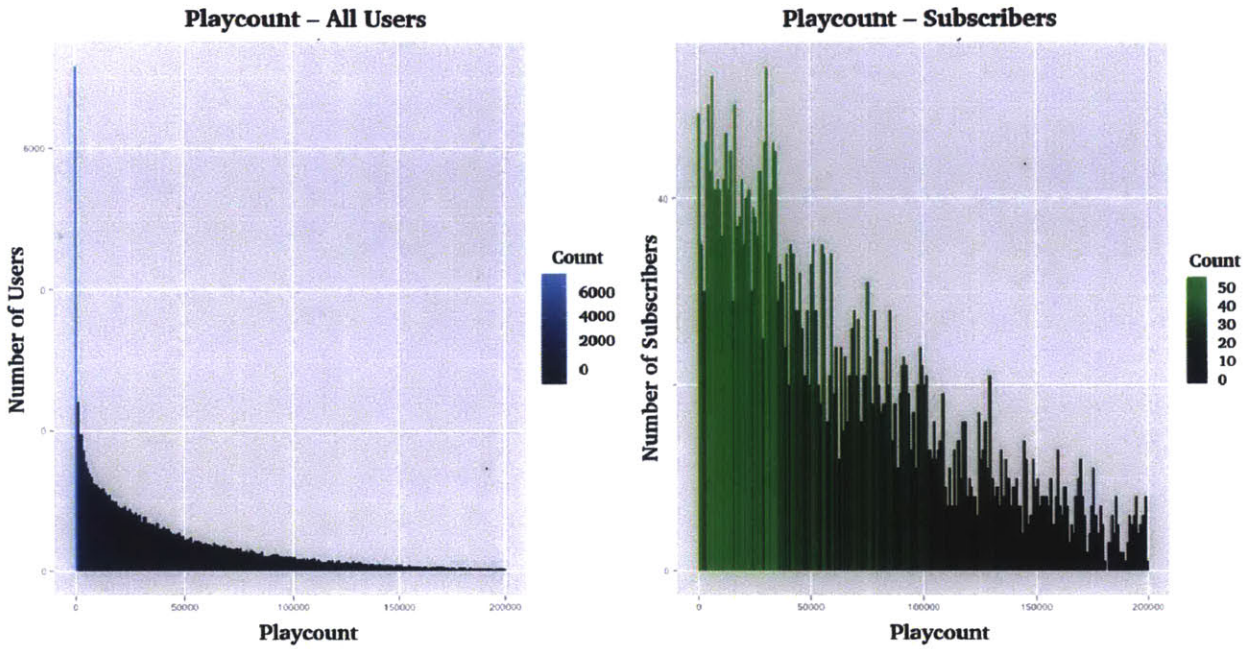
Age: The mean and median age of all users of Last.fm is lower than the mean and median age of the subscribers. The age distribution of the users (all users - blue and subscribers - green) in figure 19 clearly shows the median and mean age of the subscribers shifting to a higher value. The normal distribution depicts the target segment age range of the business.

Figure 19 Age Distribution - All Users and Subscribers



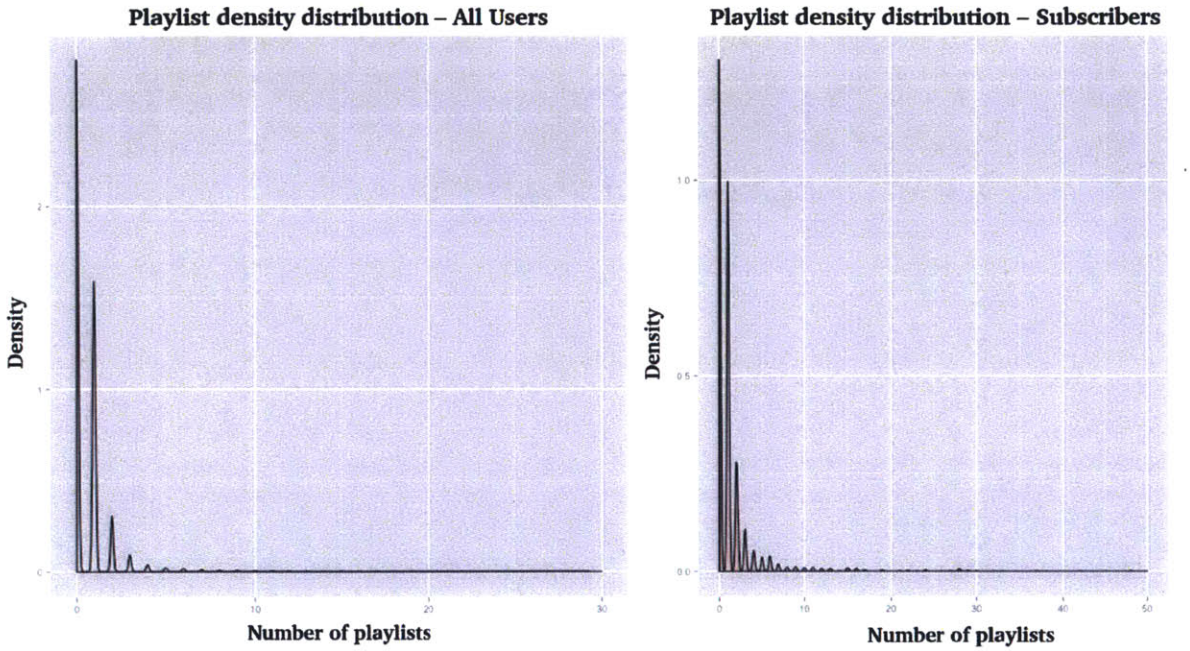
Playcount: Playcount variable refers to the total number of tracks played by the user of Last.fm. The mean value subscribers is 88,950 versus 65,100 songs for all the users. Subscribers definitely show a tendency to listen to more songs.

Figure 20 Playcount Distribution - All Users and Subscribers



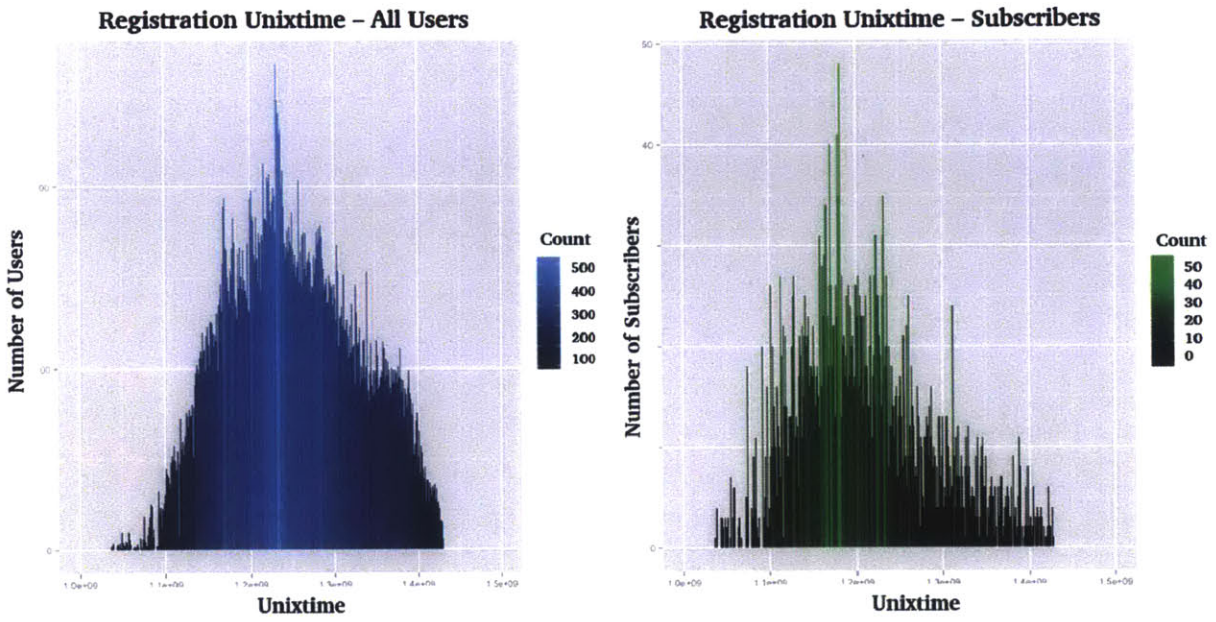
Playlists: Subscribers tend to create more playlists on the site (Mean value of 7.724) when compared to all users (Mean value of 1.159).

Figure 21 Playlists - Density Plot - All Users and Subscribers



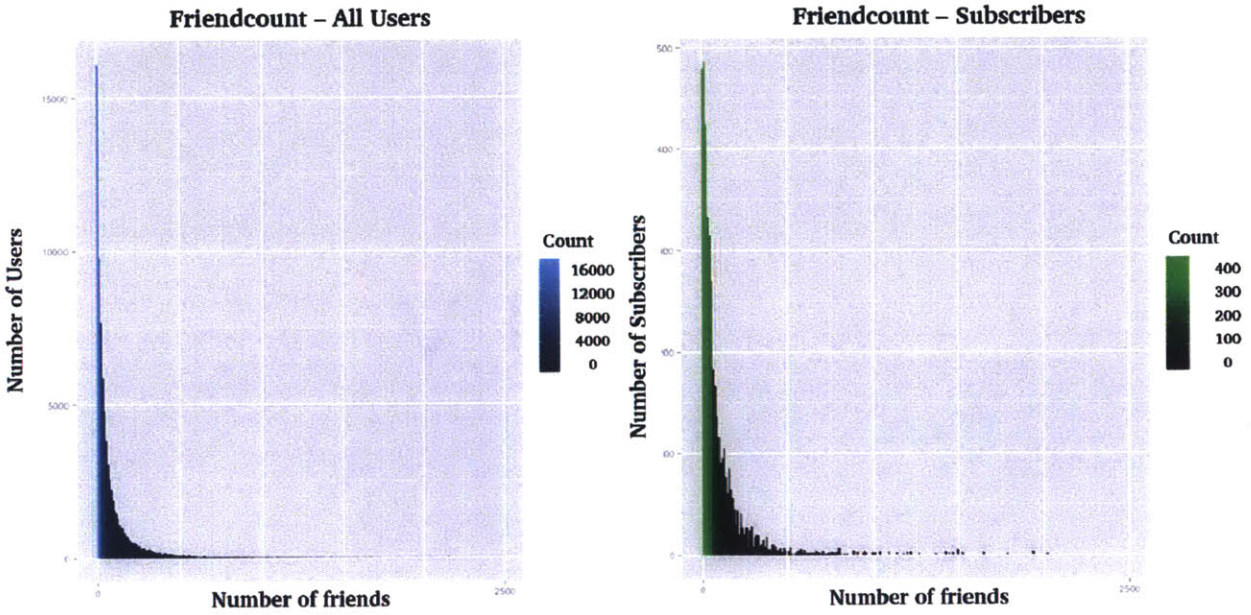
Registered Time: The registration unixtime of a user is the time in seconds since 1st January, 1970 UTC till the instant of user registration on Last.fm website. Small number indicates, early registration. Subscriber mean value (1.210e+09) is lower than all user mean value (1.252e+09).

Figure 22 Registration Unixtime - All Users and Subscribers



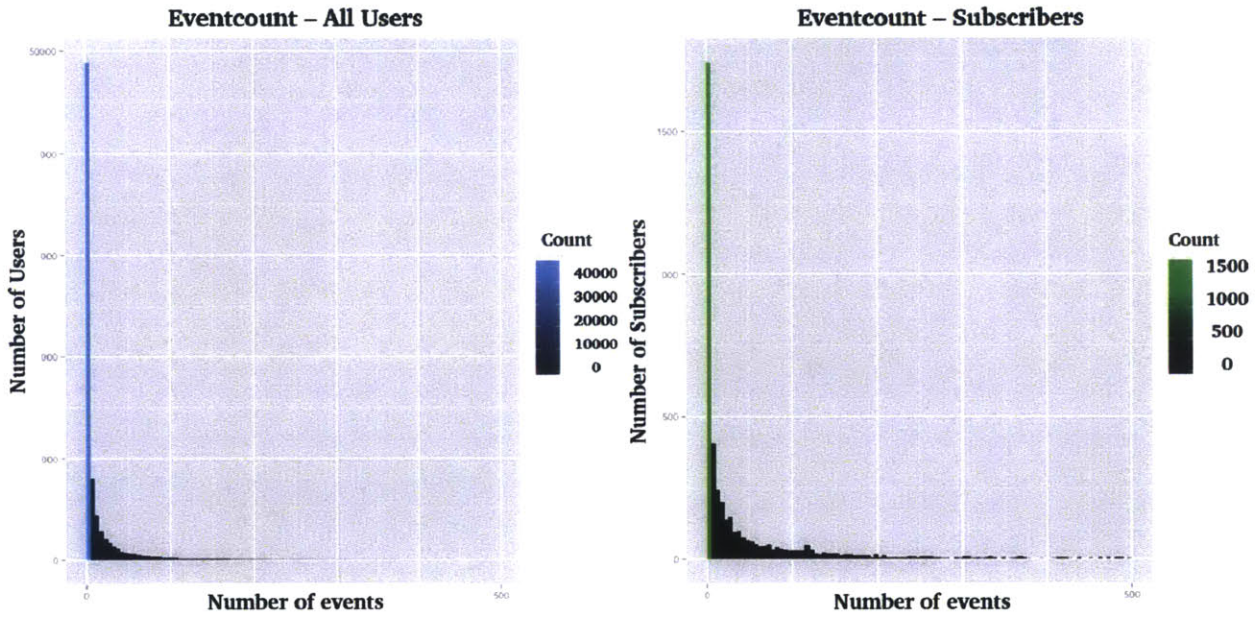
Friends: Last.fm allows users to integrate their facebook account with their profile or manually search for friends on Last.fm and add to their profile. The mean value of friend count for subscribers is 126.6 which is lower than mean for all users – 128.5. On looking at the data for all users, it is clear that few users have friendcounts in the range of 16,000 which could be increasing the mean friend count of All Users. This high count could only be possible by linking their facebook accounts with profile, assuming facebook allows 5,000 plus friends on a user's profile.

Figure 23 Friendcount Distribution



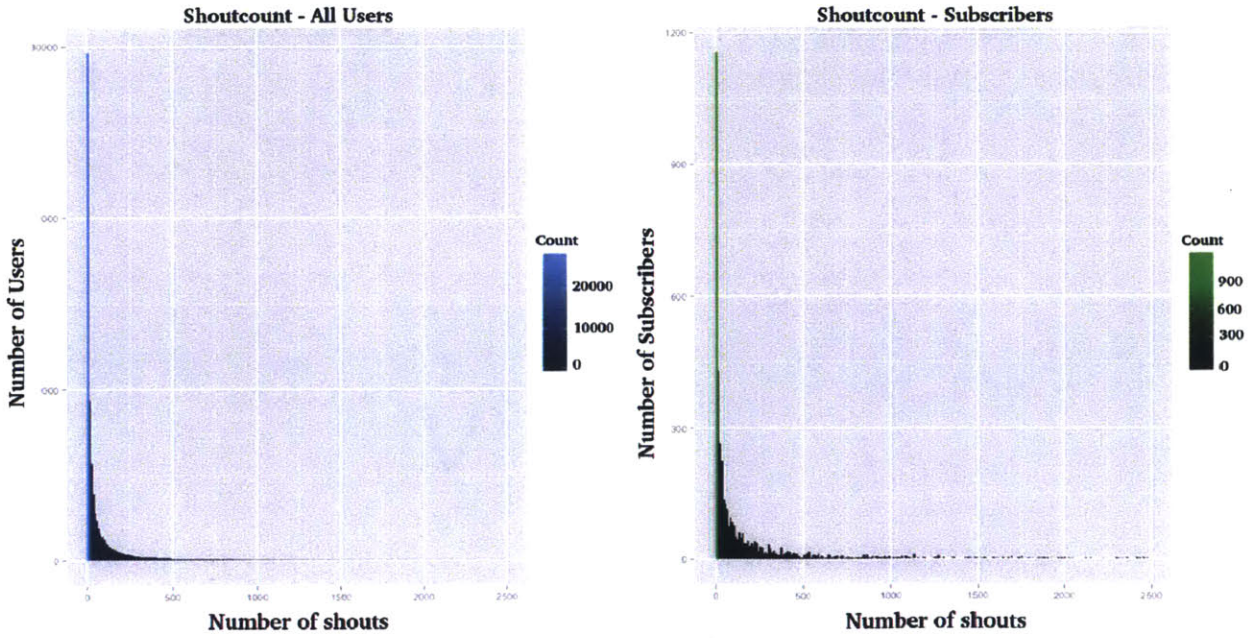
Events: Last.fm lets users create, join and edit event pages. The API gives details of past events that user had joined and new events that the user is planning to attend. Total event count is calculated by adding past event counts and event count. On looking at summary statistics and the histogram distributions of events, it is seen that subscribers have higher mean values in both past events (44.16) and events (1.206) versus all users (past events: 16.81 and events: 0.4122).

Figure 24 Eventcount Distribution - All Users and Subscribers



Shouts: Last.fm lets users share personal messages on their personal profiles called 'shouts'. Mean value shouts-count for subscribers is 382.9, more than twice the mean value for all users – 166.4.

Figure 25 Shoutcount Distribution - All Users and Subscribers



6.4 Predictive Analytics

To understand the interplay of engagement, retention, social features and demographics and predict the probability of premium subscriber, regression tree (CART) is used. K-Fold cross validation technique is used for parameter selection.

6.4.1 CART Model with k-fold Cross Validation

Regression tree method is used as a predictive method to predict the independent variable: **subscriber** using dependent variables: (demographic) **age, gender,** (retention) **unixtime,** (engagement) **playcount, playlists, isimageincluded** (social) **friendscount, eventscount, pasteventscount and shoutscount.** The independent variable has two possible outcomes or classes: 0 (free user) and 1(premium subscriber). This is a binary classification problem and can be solved using logistic regression or CART or Random Forest methods. However, due to the high interpretability of results in CART models, this method is used and fine-tuned with k-fold cross validation. R programming is used to perform the analysis. The script file is included in the appendix for reference.

CART

CART (Classification and regression tree) is a flexible, data-driven, tree-based method for determining an outcome using splits, or logical rules, on the independent variables. The interpretability and ability to capture non-linearities in the data makes it a good modelling technique to be used in this case.

Train and Test

The dataset is split randomly based on the dependent variable – subscriber into training set and test set. The training dataset consists of 70% of observations and the test set contains the balance. The training dataset is used to develop the CART model. Once the model is created, performance is assessed on test set: true out-of-sample set.

Classification Matrix and Model Accuracy

The predicted outcomes can be tabulated into classification or confusion matrix to assess the accuracy, sensitivity, specificity and false negative rate of the model. Since the predicted outcome in this model is either 0 or 1, the classification matrix is a two by two matrix.

Table 6 Classification Matrix

	Predicted Class = 0	**Predicted Class = 1**
Actual Class = 0	True Negatives (TN)	False Positives (FP)
Actual Class = 1	False Negatives (FN)	True Positives (TP)

Using this matrix, following critical metrics are calculated:

$$Accuracy = (TN + TP) / (TN + TP + FN + FP)$$

$$Sensitivity = TP / (TP + FN)$$

$$Specificity = TN / (TN + FP)$$

$$False\ Negative\ Rate = FN / (FN + TP) = 1 - Sensitivity$$

Higher the model accuracy, better is the predictivity of the model. Ideal accuracy will be 100% and similarly is the model perfectly predicts the outcome, there would be no false positives or false negatives which would yield 100% sensitivity and specificity metric.

However, a perfect model is real world data is highly unlikely to predict perfectly and would have FN or FP or both. Depending on the prediction problem at hand, the requirement will be to reduce FN or FP. This can be done though weights and penalty matrix.

Weights and Penalty Matrix

The CART model needs to be fine-tuned by attaching weights and penalty matrix. In this prediction problem, a 100% accurate model classifies the subscribers and non-subscribers perfectly. However, if it is not perfect, there will be instances of FN and FP. False negative means that a user who is actually a subscriber gets classified as a non-subscriber. This a costly mistake for the model as it is of utmost importance to correctly recognize all the subscribers. By associating penalties with FN, the model can be fine-tuned which can also result in higher false positives and lower accuracy. For the model below, a penalty matrix of $\begin{matrix} 0 & 8 \\ 10 & 0 \end{matrix}$ is used.

CP (Complexity Parameter) and K-Fold Cross validation

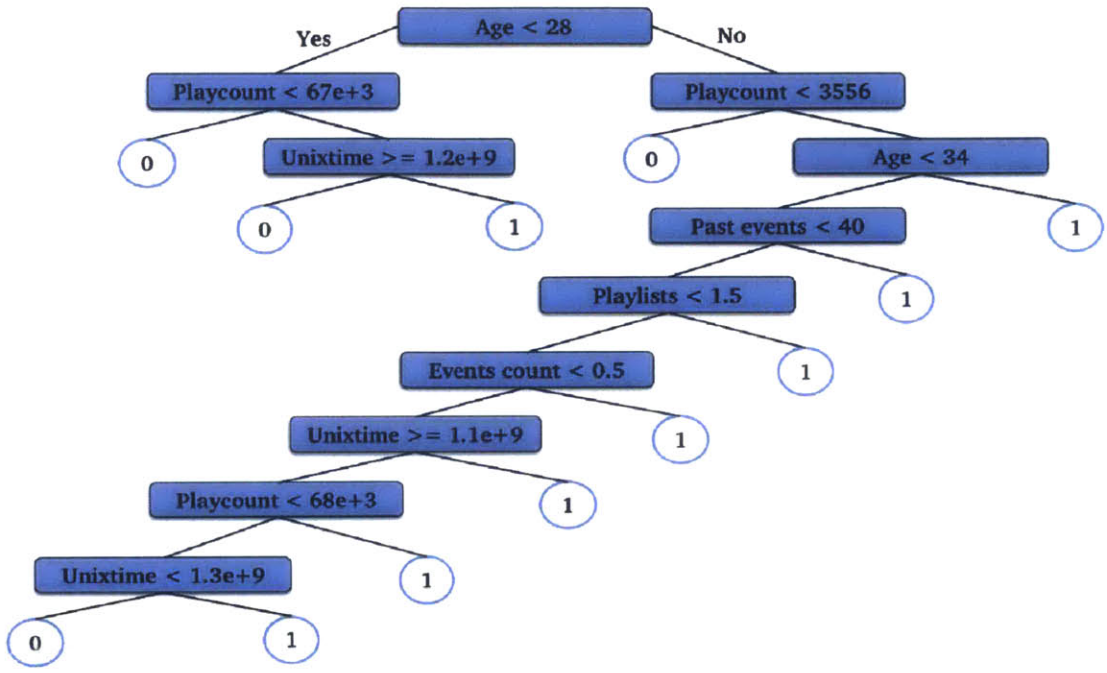
The model can be built such that each bucket or leaf of the resulting CART model is very 'pure'. High purity of a bucket means that higher proportion of observations belong to the same class. However, this could lead to over-fitting to the training set. *CP* (complexity parameter) is used to avoid this problem. To train the model better, k-fold cross validation method is used on training set.

6.4.2 Results

The CART Model

Cross validated CART model yields the following regression tree (Figure 26). The splits represent the independent variable used to make the decision. Higher the split, higher is the significance of the independent variable associated with the split. If the answer to a plot decision is *Yes*, the decision moves to the left side, else to the right side.

Figure 26 CART Model with Cross validation



Classification Matrix

Running the model on the test dataset provides the following results:

	Predicted Class = 0	Predicted Class = 1
Actual Class = 0	16890 (TN)	5555 (FP)
Actual Class = 1	410 (FN)	855 (TP)

Using this matrix, following metrics are calculated:

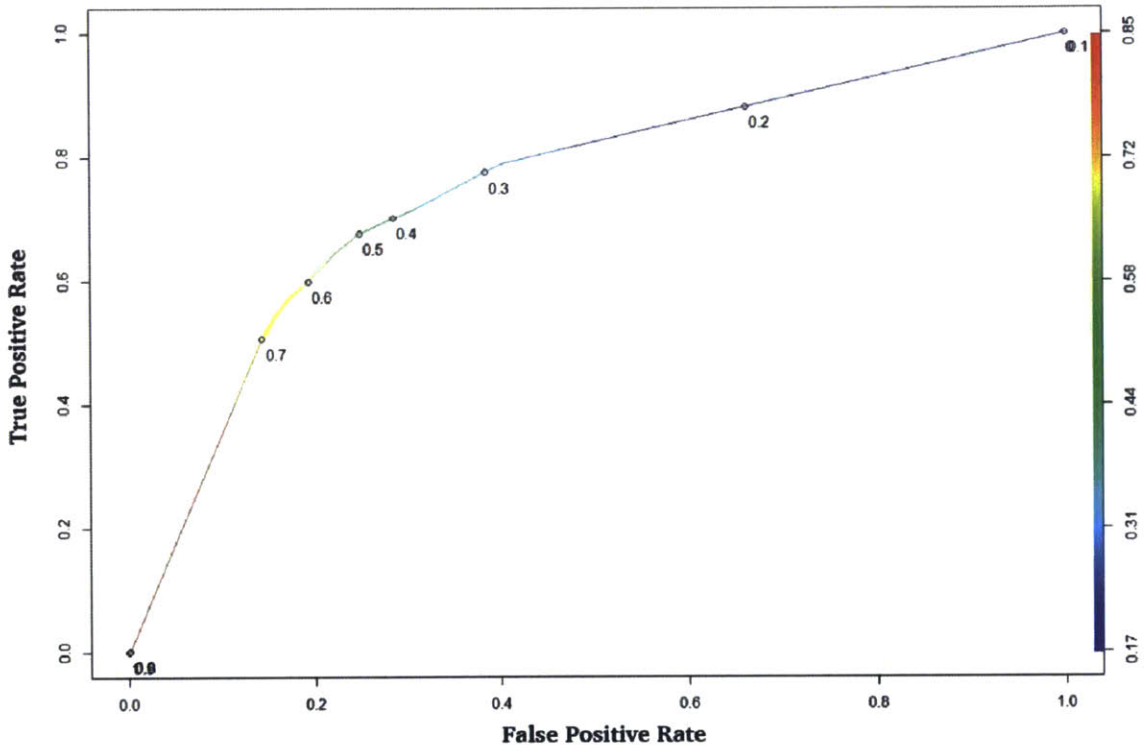
$$Accuracy = \underline{74.84\%}$$
$$Sensitivity = TP / (TP + FN) = \underline{67.6\%}$$
$$Specificity = TN / (TN + FP) = \underline{75.25\%}$$
$$FNR = \underline{32.4\%}$$

The classification matrix shows that there are 410 false negatives and 5555 false positives and the model has an accuracy of 74.84%. In this case, a freemium business would want to accurately predict all subscribers so that targeted and personalized marketing efforts can be aimed at users who show characteristics of turning into a subscriber. At the same time precious dollars should not be wasted on false positives i.e. users who will not upgrade to become a subscriber, but the model erroneously classified them as subscribers. But missing a potential subscriber (false negative) is a costlier error for a business. The business should aim at increasing the 'sensitivity' of the model and decreasing the false negative rate of the model. Hence the model can be further tuned with penalty values to reduce false negatives.

ROC Curve and AUC

ROC (Receiver Operator Characteristic) curve plots the false positive rate of a model on the x-axis, and the true positive rate (the percentage of class 1 observations that the model classified correctly) of the model on the y-axis for different threshold possibilities. The ROC curve for the test data is shown below. The line shows how the false positive rate and true positive rate vary with different threshold values for a particular model. Ideally, we want the ROC curve to be as close to the top left corner of the plot as possible, since it gives a high true positive rate and a low false positive rate.

Figure 27 Receiver Operator Characteristic curve for the CART – CV Model (Test data)



Using the ROC curve, another critical metric – AUC (Area under the curve) is calculated. AUC is a number between 0 and 1. Higher the AUC, higher is area under ROC curve, and the better is a model in distinguishing between the two different classes. AUC gives the percentage of time that we expect the model to predict the two observations correctly. Random guessing implies an AUC of 0.5. A model with AUC greater that 0.5 is doing something smarter than guessing. Closer the AUC is to 1, better is the model.

AUC of the model = <u>0.7474937</u>
(Calculated on test dataset)

6.4.3 Tuning with Penalty Matrix

The model is reconstructed with a penalty matrix of $\begin{matrix} 0 & 6 \\ 10 & 0 \end{matrix}$. This matrix assigns a higher penalty for FN than FP. This new model has lower accuracy of **65.7%,** but higher sensitivity of **78%** and lower false

negative rate of 22%. Regression tree for the same is included in the appendix (Figure 35). Based on the priorities for prediction: higher accuracy or lower false negative rate, one of the two models can be chosen.

6.5 Interpretation of Model Results

The CART model developed to predict free user or premium user outcome has fair to good predictive performance. The model results are now interpreted and tested against the hypotheses.

Demographics

First hypothesis stated that *'Demographic factors will not have a strong association with the likelihood of subscribing to premium services.'* The CART tree clearly shows that the primary split in the model is because of a demographic factor – Age, which means that age is the most significant variable determining the outcome. This disproves the first hypothesis. Association of age to prediction could be because higher age translates to higher disposable income and hence willingness to pay. By starting with the right age segment of current free users, targeted marketing can help convert them to premium users.

Retention

Second hypothesis states that *'Retention is positively associated with the likelihood of subscribing to premium services.'* The CART tree places unixtime, which a measure of user retention in this case, high in the list of significant variables. This proves the second hypothesis.

Engagement

Third hypothesis states that *'Engagement is positively associated with the likelihood of subscribing to premium services.'* This is proved by the 'playcount' variable which features in one of the top splits of the CART model. Higher the playcount, greater is the probability of outcome as a subscriber.

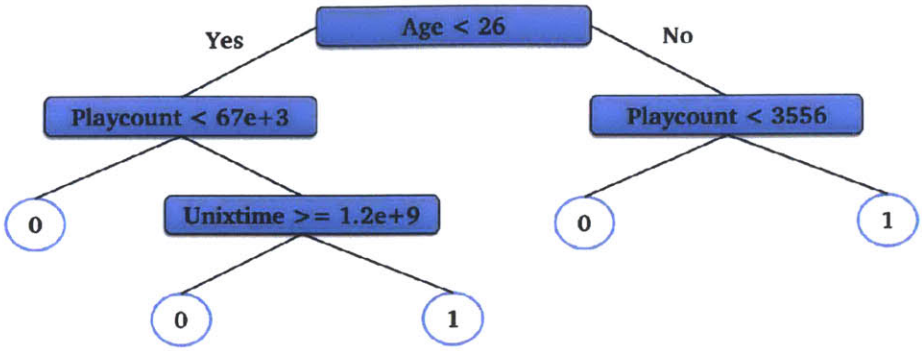
Social and Virality

Fourth hypothesis states that *'Social features (Virality) is* positively *associated with the likelihood of subscribing to premium services.'* This is true as eventscount and pasteventscount varibles are significant in predicting the outcome.

Comparison of Variable Categories

When the model is redesigned to lower complexity parameter in order to avoid over fitting, it yields a regression tree with only top three significant variables as below.

Figure 28 CART Tree - Significant Variables



These variables are age (demographic), playcount (engagement) and unixtime (retention). Other variables, including the social variables are not as significant as the three variables appearing in CART tree: figure 25. Based on this model demographic variable is more significant than engagement variable which in turn is more significant than retention variable. This disproves 5a, 5b and 5c hypotheses.

Conclusion

Predictive analytics on Last.fm user dataset reveals interesting information on significant independent variables driving user conversions. These can be referred to when designing marketing activities and fine-tuning the feature sets to increase conversion rates.

7. Key Findings and Implications

The purpose of this study is to use business model analysis and empirical study of freemium companies to identify critical factors that determine success. Freemium business models look deceptively simple but they have their inherent challenges. This section consolidates and distills the study into key takeaways.

One-size-does-NOT-fit-all

The Freemium model is not a *one-size-fits-all* strategy for digital startups. Strategy that works for one company can fail miserably for another freemium product or may not produce the staggering results it did for the first company. The nature of product or service or industry can change the levers that drive success and hence every company needs to find its own secret sauce for success.

Case in point: The analysis of Spotify showed that social features and referrals helped in the success of the business model. However, in the same MaaS industry, Last.fm's user data analysis showed that social features were of lower significance when compared to retention, demographic and engagement features.

Developing Organic Acquisition Engine

Acquiring new free users is as important as converting free users to premium users. Companies should invest in finding novel ways to acquire customers organically. Referrals are a popular model for organic growth strategy. Once this acquisition strategy falls in place, word-of-mouth effects start rolling leading to high user registrations. When a product has a high number of users, network effects come into being and switching costs increase.

Case in point: Dropbox initially began its marketing efforts with search advertising which was the obvious strategy for internet startups. However, the company soon realized that it had to create demand and not harvest demand and switched their strategy to developing and excelling in customer acquisition. The trick was in devising a great referral strategy by incentivizing both the sender and receiver of the referral. In 2010, Dropbox's 4 million users produced 2.8 million direct referral invites, a referral rate of 70%.

MVP

The principles of lean startup and minimum viable product is highly applicable to digital startups today. Quick MVP release to market can help the founders test their product in a closed set of beta users, receive invaluable feedback, tune feature set, release and iterate. Even though a bare bone version of the product is initially launched, the lean startup method will quickly help build the feature set. Keeping the product in a perpetual beta/testing phase will help the product team edit, modify and upgrade its offering based on customer needs.

Case in point: Zynga, Evernote and Dropbox launched as very simple, bare minimum products and went through multiple iterations to reach the stage they are in now.

Great Long-term Retention Rate

Retention is one of the primary metrics for evaluating a freemium product. If a user stays with the product long, the value of the product, stickiness and switching costs increase over time. Having a sound strategy to retain users and increase the retention rate increases the likelihood of higher conversions.

Case in point: Evernote's CEO Phil Libin believes that the key to their success is a great long term retention rate. Evernote is not a social product, but a personal product. Over time the value of this product grows very high and becomes an indispensable part of the person. In these kinds of situations, the user sticks to the product over longer periods and sees the value in converting into a paying user.

Multi-platform Presence

Today users access services on multiple platforms, from tablet to laptop to phone, each one with a different operating system. The likelihood of a person to convert into a premium user increases if the user accesses the product on multi platforms.

Case in point: After launch, product team at Evernote quickly realized that if they were to be true to their value proposition of 'external brain to remember everything' they had to ensure that users could have access to their product no matter what device they used. The app was launched for all operating systems which made users engaged with the service irrespective of the device.

Engineering a Freemium Product Catalogue

A critical question that needs to be answered during product design is if the offering is a-la-carte mode or tiered levels. Accordingly the freemium product catalogue should be engineered.

Having a diverse catalogue or a-la-carte offering lets the product interact with the user at multiple touch points, giving opportunities for monetization. A large, refined product offering lets a user extract value from the product in smaller and more relevant transactions and meantime gives the business more monetization opportunities. This broad catalogue gives the user a sense of personalization by fulfilling the needs and leverages the freemium economic model to the fullest. This is essentially the crux of the continuous monetization curve and by building a broad assortment of economic transaction opportunities for the users to experience advanced features, both the user and company benefits.

Case in point: The case studies discussed do not have a broad product catalogue except for Zynga which comes closest to offering multiple opportunities to pay and use in-app features.

Evangelism

Free users are of great value to a freemium company. First, they can become premium users and generate revenue. Second, their presence can increase the value of the product for others by the virtue of network effect and third, they can invite new users through referrals. Studies (Kumar, 2014) have shown that a free user can be worth 15% to 25% as much as a premium subscriber. Hence, businesses should offer appropriate referral bonus to incentivize customers.

Shared Product Use and Network Effects

Social features can play a huge role in keeping the customers engaged with the product or acquiring new customers though viral feature or word of mouth. This social or shared use of product can influence the likelihood of users upgrading. Firms can look at incentivizing users to use the social functionalities leading to positive externalities on other users. Firms can create platforms in which users are encouraged to go beyond passive content consumption.

Case in point: Zynga successfully integrated social features in its game. The social nature of Farmville helped in the viral growth of the game's usage. Social features however are not necessary conditions for success. Dropbox and Evernote do not have any social or community features built in them, but are highly successful freemium products.

Evolution with Data

Data driven, iterative approach in product and feature design of can increase the likelihood of designing a solution where the user is very satisfied and engaged.

Case in point: Zynga's constant evolution was one of the reasons for its success. Every click that a player made on the game was captured and clicking patterns analyzed carefully to understand user behavior, engagement and features that were successful with customers and ones that were not. On the basis of these observations, Zynga tweaked the feature sets and released upgrades twice a week. This approach by Zynga made the product catalogue strong, games complex and appealing to novice as well as advanced players.

Balancing Free and Premium Value Proposition

The purpose of free features is to attract users to the product. If the product is not drawing customers, it essentially means that the free feature set is not persuasive enough. If the product is attracting lots of customers but the conversion rate is low, it means the free offerings are too rich and it is cannibalizing on the revenue streams. Users should have a clarity on the value proposition of the premium offering. It the users do not grasp this, the conversion rates will be low. Hence, separation of free and premium versions must be clear and based on user needs.

Case in point: LinkedIn and Spotify. Spotify has very engaging free functionalities. At the same time, the value proposition of the premium features are compelling to a user to upgrade. With 24 million users and 6 million subscribers, Spotify has very high conversion rates for music streaming products. At the same time LinkedIn does not have a very strong communication of its premium features and hence the advantages of upgrading are very vague. LinkedIn has potential of monetizing further from member subscriptions with better, clear cut communication of premium features.

Compelling Migration Path

Once the firm has clear free and premium value propositions, it needs to create a compelling event (call to action) to ensure migration to paid. The free version has to be good enough to be sticky and engaging. At the same time it should leave the user wanting for more and the migration path and transition for such user should be easy, intuitive and logical.

Case in point: Spotify has a conversion rate of about 20% in a highly competitive MaaS industry and this is heavily owed to the clear distinction of free and premium features and a seamless, persuasive migration path.

Multi-sided Money Potential

If the business model involves two- or multi sided networks, there is potential for monetizing more than one side of the platform. Typically in a two sided network, one is the monetized side (money side) and the other side (loss side) is subsidized to create network effect and stickiness in the platform. However, it is possible to monetize more than one side of the platform.

Case in point: LinkedIn is a multi-sided platform. The members, recruiters and advertisers form the three pillars of their business and all the three sides are monetized.

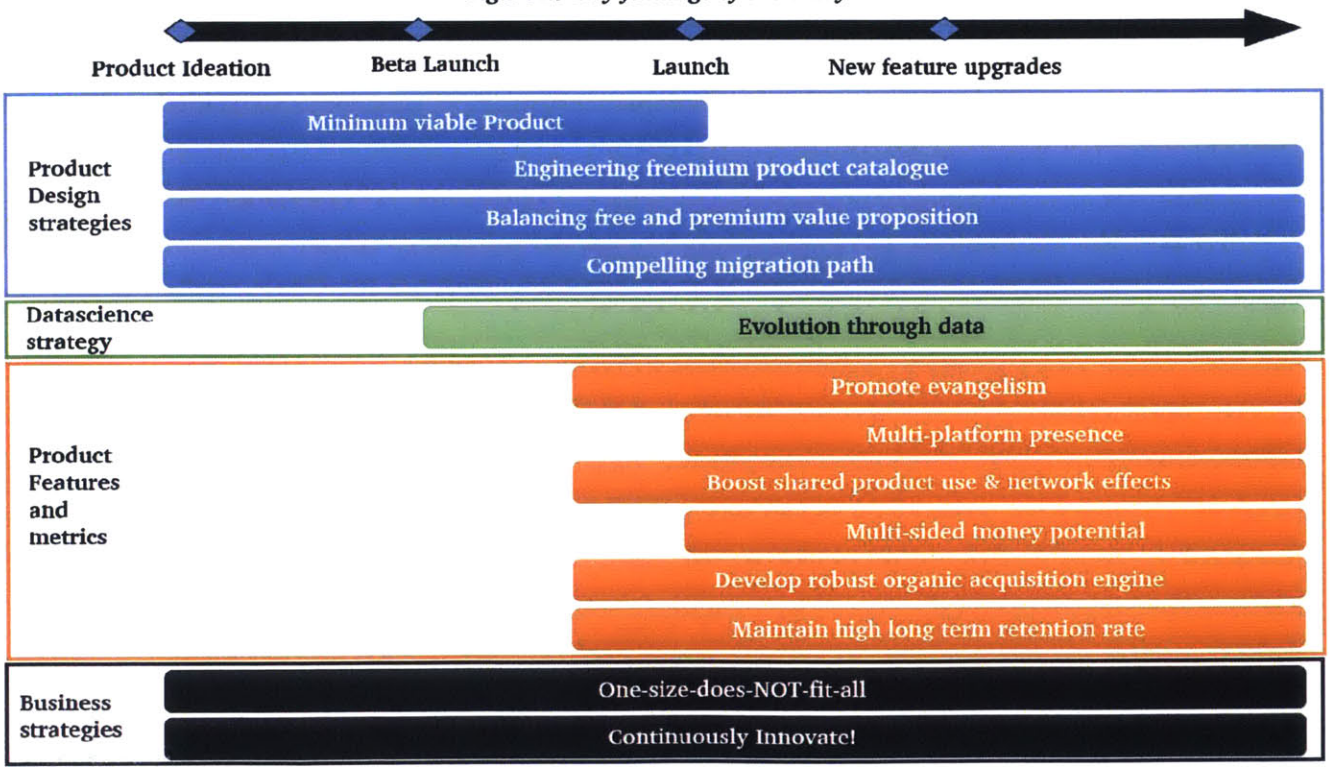
Cycle of Innovation

The freemium business model should not be looked at only as a customer acquisition and premium conversion strategy. Survival and success in the model requires continuous innovation in the product and features. The freemium lifecycle discussed in the section 3.2 mirrors the need to continuously upgrade. Freemium is not just monetization model, but also a commitment to innovation.

Case in point: Dropbox was launched in 2008 with bare minimum features. Over time it kept reinventing and upgrading its features and is a poster child of a successful cloud storage solution in the aggressive storage industry.

Key takeaways discussed above are presented below in one, easy to understand framework.

Figure 29 Key findings of the study



Conclusion

The marginal costs of production and distribution of digital products continue to drop, hence increasing number of firms will resort to freemium model. Size of the subscriber network is an important factor for success as the conversion rates are extremely low. Various tactics can be used alongside to increase engagement, retention, virality and revenue. Freemium model is indeed attractive and the key findings of the study hopes to boost the odds of succeeding.

8. Limitations

8.1 Business Model Analysis

This study analyzed five successful business models where the data was sourced from secondary research on the internet and not primary research. Further in-depth analysis can be done by conducting primary research to get specific details of the critical variables. Also, five firms aren't sufficient to generalize and deduce critical success factors. The study can be done on a wider range of companies and industries. Including failed freemium models will lend another dimension and depth to the analysis.

8.2 Empirical Study

The empirical study was conducted on data collected from Last.fm's site. The analysis had following limitations:

Quantity: The model was built on 79,033 users. However, to build a better model the study needs to be conducted on a larger dataset. Higher the number of data points, better is the learning in the model. Since the converted users in a freemium product is low, the model needs much larger dataset to predict better.

Features included: Last.fm api allows access to a rich set of user attributes like tags, groups. The model can use all the features to predict outcome. The current model uses a small subset of these features.

Better model: The model can be tuned further to increase sensitivity, accuracy and reduce the false negatives. Currently basic method of introducing penalty/loss matrix is used to improve sensitivity. Cluster-then-predict method can be used to understand the behavior of clusters of users and predictive model can be developed for each cluster.

9. References

Anderson, C. (2004). The Long Tail. *Wired.*

Anderson, C. (2008, February 25). *Free: Why $0.00 Is the Future of Business.* Retrieved from Wired: http://archive.wired.com/techbiz/it/magazine/16-03/ff_free?currentPage=all

Anderson, C. (2009). *Free: the future of a radical price.* Hyperion Books.

Aune, S. (2007, July 27). *Online Storage.* Retrieved from Mashable: http://mashable.com/2007/07/28/online-storage/

Belic, D. (2011, 12 22). *Freemium apps to continue flourishing in 2012.* Retrieved from Intomobile: http://www.intomobile.com/2011/12/22/freemium-apps-continue-flourishing-2012/

Bhargava, H., & Choudhary, V. (2001). Information goods and vertical differentiation. *Journal of Management Information System*, 89-106.

Bhargava, H., & Choudhary, V. (2008). When is versioning optimal for information goods. *Management Science 54(5)*, 1029-1035.

Brochet, F., & Weber, J. (2012). *LinkedIn Corporation.*

Brzozowski, M., Sandhol, T., & Hogg, T. (2009). Effects of Feedback and Peer Pressure on Contributions to Enterprise Social Media. 61-70.

Buckman, R. (2008, June 18). LinkedIn gets $53 Million infusion. *The Wall Street Jounal.*

Chang, R., & Learmonth, M. (2009). Are your ads being used to drive consumers to ad-free services? *Advertising Age*, 44-46.

Chang, V., & Mendelson, H. (2010). *Social Games.*

Chellappa, R., & Shivendu, S. (2005). Managing piracy: Pricing and sampling strategies for digital experience goods in vertically segmented markets. . *Inform. Systems Res.* , 400–417.

Cheng, H. K., & Liu, Y. (2010). Optimal software free trial strategy: The impact of network externalities and consumer uncertainty. *Inform. Systems Res.*

Cheng, H., & Tang, Q. (2010). Free Trial or No Free Trial: Optimal Software Product Design with Network Effects. *European Journal of Operational Research*, 437-447.

Crunchbase - Dropbox . (2014). Retrieved from Crunchbase: https://www.crunchbase.com/organization/dropbox

DeBaise, C. (2009, October 21). *Launching LinkedIn from a Living Room.* Retrieved from Wall Street Journal Online: http://www.wsj.com/articles/SB10001424052748704224004574487750190885402

Dellarocas, C. (2003). The digitization of word of mouth: Promise and challenges of online feedback. *Management Science*, 1407-1424.

Dörr, J., Benlian, A., Vetter, J., & Hess, T. (2010). Pricing of content services - an empirical investigation of music as a service. *Proceedings of the sixteenth Americas Conference on Information Systems.* Lima, PEru.

Dörr, J., Wagner, T. M., Benlian, A., & Hess, T. (2013). Music as a service: an alternative to music piracy? An empirical investigation of the intention to use music streaming services. *Business & Information Systems Engineering, 5(6)*, 383-396.

Eisenmann, T. R., Pao, M., & Barley, L. (2014). Dropbox: "It just works".

Evans, D. S., & Schmalensee, R. (2007). The Industrial Organization of Markets with Two-Sided Platforms.

Evernote - Crunchbase. (2015). Retrieved from Crunchbase: https://www.crunchbase.com/organization/evernote#sthash.FZCUcTNM.dpufA

Evernote - Wiki. (2015). Retrieved from Wikipedia: http://en.wikipedia.org/wiki/Evernote#Accounts_and_pricing

Faugère, C., & Giri Kumar, T. (2007). Designing Free Software Samples: A Game Theoretic Approach. *Information Technology & Management*, 263-278.

Feder, Y. (2014, September 18). *Social Network 301: What is Virality?* Retrieved from LinkedIn Pulse: https://www.linkedin.com/pulse/20140918141439-7859692-social-network-301-what-is-virality

Gallaugher, J., & Wang, Y. M. (2002). Understanding Network Effects in Software Markets: Evidence from Web Server Pricing. *MIS Quarterly*, 303-327.

Gannes, L. (2010, March 26). *Case Studies in Freemium: Pandora, Dropbox, Evernote, Automattic and MailChimp.* Retrieved from Gigaom: https://gigaom.com/2010/03/26/case-studies-in-freemium-pandora-dropbox-evernote-automattic-and-mailchimp/

Gayo, L. M., & Labra, J. (2009). Doing Business by Selling Free Services. *Web 2.0*, 1-14.

Ghose, A., & Sundararajan, A. (2005). Software versioning and quality degradation? An exploratory study of the evidence.

Gupta, S., Hanssens, D., Hardie, B., Kahn, W., Kumar, V., & Lin, N. (2006). Modeling customer lifetime value. *Journal of Service Research.*

Hagiu, A., & Wright, J. (2011). Multi-Sided Platforms.

Haruvy, E., & Prasad, A. (2001). Optimal Freeware Quality in the Presence of Network Externalities: An Evolutionary Game Theoretical Approach. *Journal of Evolutionary Economics*, 231-248.

Hayes, T. (2008). *Jump Point: How Network culture is revolutionizing business.* McGraw Hill.

Houston, D. (2013, October 12). *Dropbox - Startup lessons learnt.* Retrieved from Slideshare: http://www.slideshare.net/gueste94e4c/dropbox-startup-lessons-learned-3836587

Hui, P., & Buchegger, S. (2009). Groupthink and Peer Pressure: Social Influence in Online Social Network Groups. 53-59.

Hung, J. (2010). Economic essentials of online publishing with associated trends and patterns. . *Publishing Research Quarterly, 26(2)*, 79-95.

Issa, N. (2013, April 10). *How To Build And Optimize An Onboarding Funnel.* Retrieved from Woopra: https://www.woopra.com/blog/2013/04/10/how-to-build-and-optimize-an-onboarding-funnel/

Jiang, Z., & Sarkar, S. (2009). Speed Matters: The Role of Free Software Offer in Software Diffusion. *Journal of Management Information Systems*, 207-239.

Jones, R., & Mendelson, H. (2010). Information goods vs. industrial goods: Cost structure and competition. *Management Science.*

Kapil, B., & Robert, S. (2004). The Effects of Free Sample Promotions on Incremental Brand Sales. *Marketing Science*, 345.

Koetsier, J. (2014, March 27). *Mobile app monetization: Freemium is king, but in-app ads are growing fast.* Retrieved from Venture Beat: http://venturebeat.com/2014/03/27/mobile-app-monetization-freemium-is-king-but-in-app-ads-are-growing-fast/

Kumar, V. (2014). Making "Freemium" Work. *Harvard Business Review*, 27-29.

Learmonth, M. (2009, September 21). *Are Your Ads Being Used to Drive Consumers to AdFree Services?* Retrieved from Advertising Age: http://adage.com/article/digital/freemium-ads-drive-consumers-pay/139148/

Lee, C., Gupta, S., & Kumar, V. (2013). Designing Freemium: A model of consumer usage, upgrade and referral dynamics.

Libin, P. (2010). Freemium for Consumer Internet Business. Freemium Summit 2010.

Libin, P. (2014, May 13). *Evernote Blog.* Retrieved from Evernote: https://blog.evernote.com/blog/2014/05/13/evernote-reaches-100-million-users/

LinkedIn files for IPO. (2011, February 4). Retrieved from NextUp Research: www.nextupresearch.com

LinkedIn Quarterly Earnings. (2013). Retrieved from LinkedIn: http://investors.linkedin.com/results.cfm

Little, G. (2011, July 13). *Why Evernote Is Winning With The Soft Stuff.* Retrieved from Forbes: http://www.forbes.com/sites/nicoleperlroth/2011/07/13/why-evernote-is-winning-with-the-soft-stuff/

Liu, C. Z., Au, Y. A., & Choi, H. S. (2012). An empirical study of the freemium strategy for mobile apps: Evidence from the google play market. *Paper presented at the Proceedings of the thirty third International Conference on Information Systems.* Orlando, FL, USA.

Lynley, M. (2014, November 18). *As Online Storage Prices Race To Zero, Dropbox Turns To Business.* Retrieved from Buzzfeed: http://www.buzzfeed.com/mattlynley/the-future-of-business-according-to-dropbox#.fgDgGgOG0

Lyons, K., Messinger, P., Niu, R., & Stroulia, E. (2012). A tale of two pricing systems for services. *Information Systems and e-Business*, 19-42.

Madrigal, D., & McClain, B. (2010, December 6). *Getting the most out of freemium model.* Retrieved from UX Matters: http://uxmatters.com/mt/archives/2010/12/getting-the-most-out-of-the-freemium-model.php

Miller, C. C. (2009, May 25). *Ad revenu on the Web? No sure bet.* Retrieved from New york Times: New York Times: http://www.nytimes.com/2009/05/25/technology/start-ups/25startup.html?r=1

Needleman, S., & Loten, A. (2012, August 22). *When Freemium Fails.* Retrieved from The Wall Street Journal: http://www.wsj.com/articles/SB10000872396390443713704577603782317318996

Niculescu, M. F., & Wu, D. J. (2011). When Should Software Firms Commercialize New Products via Freemium Business Models?

Oestreicher-Singer, G., & Zalmanson, L. (2013). Content or community? A digital business strategy for content providers in the social age. *MIS Quarterly, 37(2)*, 591-616.

Osterwalder, A. (2004). The Business Model Ontology - A Proposition In A Design Science Approach.

Parker, G. G., & Van Alstyne, M. W. (2005). Two-sided network effects: A theory of information product design. *Management Science*, 1494 - 1504.

Piskorski, M. J., & Chen, D. (2013). *Zynga (A).*

Protalinski, E. (2015, February 5). *eMarketer: Only a third of U.S. mobile users will pay for apps in 2015.* Retrieved from Venture Beat: http://venturebeat.com/2015/02/05/emarketer-only-a-third-of-us-mobile-users-will-pay-for-apps-in-2015/

Pujol, N. (2010). Freemium: attributes of an emerging business model.

Pulkkanen, A., & Seppänen, M. (2012). Freemium Business Models in Technology Product Markets.

Raghunathan, S. (2000). Software editions: An application of segmentation theory to the packaged. *Management Information System*, 87-114.

Reichheld, F. F. (2003). One number you need to grow. *Harvard Business Review.*

Ries, E. (2011). *The Lean Startup.* Crown Business.

Rochet, J. C., & Tirole, J. (2006). Two-sided markets: a progress report. *Journal of European Economic Association*, 645-667.

Rochet, J., & Tirole, J. (2003). Platform competition in two-sided markets. *Journal of European Economic association*, 990 - 1029.

Semenzin, D., Meulendijks, E., Seele, W., Wagner, C., & Brinkkemper, S. (2012). Differentiation in freemium: Where does the line lie? *Paper presented at the Proceedings of the third International Conference on Software Business,* . Cambridge, MA, USA.

Seufert, E. B. (2014). *Freemium Economics: Leveraging Analytics and User Segmentation to Drive Revenue.* Morgan Kaufmann.

Shaw, T. (2010). Understanding successful utilisation of the 'freemium' model in digital media and games.

Shaw, T. (n.d.). *Understanding successful utilisation of the 'freemium' model in digital media and games.*

Spotify-Blog. (2008). *We've only just begun.* Retrieved from Spotify - Blog: https://news.spotify.com/us/2008/10/07/weve-only-just-begun/

Teece, D. J. (2010). Business models, business strategy and innovation. *Long Range Planning, 43(2–3)*, 172–194.

Teixeira, T., & Watkins, E. A. (2014). Freemium pricing at Dropbox.

Vishwanathan, B. (2013, February 7). *Why Is Dropbox Worth More Than $4 Billion?* Retrieved from Forbes: http://www.forbes.com/sites/quora/2013/02/07/why-is-dropbox-worth-more-than-4-billion/

Wagner, T. M., Benlian, A., & Hess, T. (2013). The Advertising Effect of Free–Do Free Basic Versions Promote Premium Versions within the Freemium Business Model of Music Services? *Proceedings of the forty sixth Hawaii International Conference on System Sciences,*, (pp. 1-10). Maui, HI, USA.

Wagner, T. M., Benlian, A., & Hess, T. (2014). Converting freemium customers from free to premium—the role of the perceived premium fit in the case of music as a service.

Wang, H., Chin, A., & Wang, H. (2011). Social Influence on Being a Pay User in Freemium-based Social Networks.

Wei, X., & Nault, B. (2005). Product differentiation and market segmentation of information goods.

Westaway, K. (2013, Mar 1). *A Billion Play, A Million Pay.* Retrieved from WSJ Blogs: http://blogs.wsj.com/accelerators/2013/03/01/a-billion-play-a-million-pay/

Wikipedia - Dropbox. (2015). Retrieved from Wikipedia: http://en.wikipedia.org/wiki/Dropbox_%28service%29

Wikipedia - Last.fm. (2015). Retrieved from Wikipedia: http://en.wikipedia.org/wiki/Last.fm

Wilson, F. (2006, March 23). *The Freemium Business Model.* Retrieved from avc.com: http://avc.com/2006/03/the_freemium_bu/

Yoffie, D., & Kind, L. (2013). *LinkedIn Corporation, 2012.*

Zynga Inc - Form 10K. (2013). Retrieved from SEC.GOV.

Zynga Revenue. (2010, May 14). Retrieved from Engage Digital: http://www.engagedigital.com/blog/2010/05/14/zynga-revenue-estimated-at-50m-per-month/

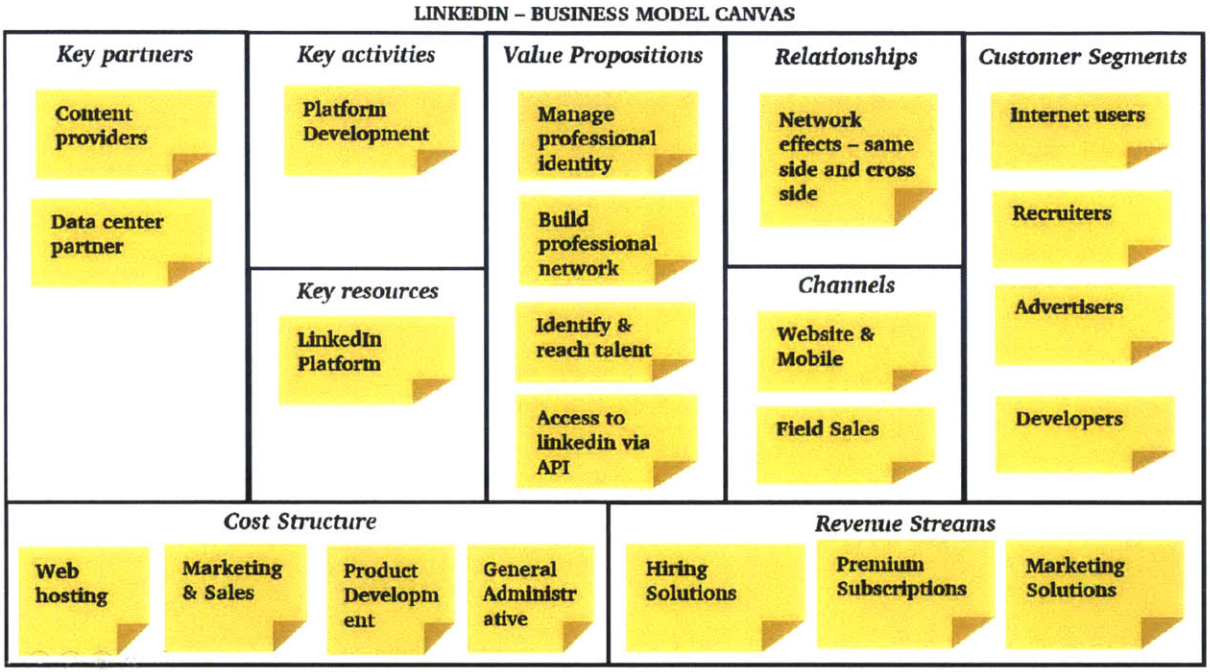
Zynga-AppData. (n.d.). Retrieved from AppData: http://www.appdata.com/devs/10-zynga

10. Appendix

10.1 Business Model Canvas

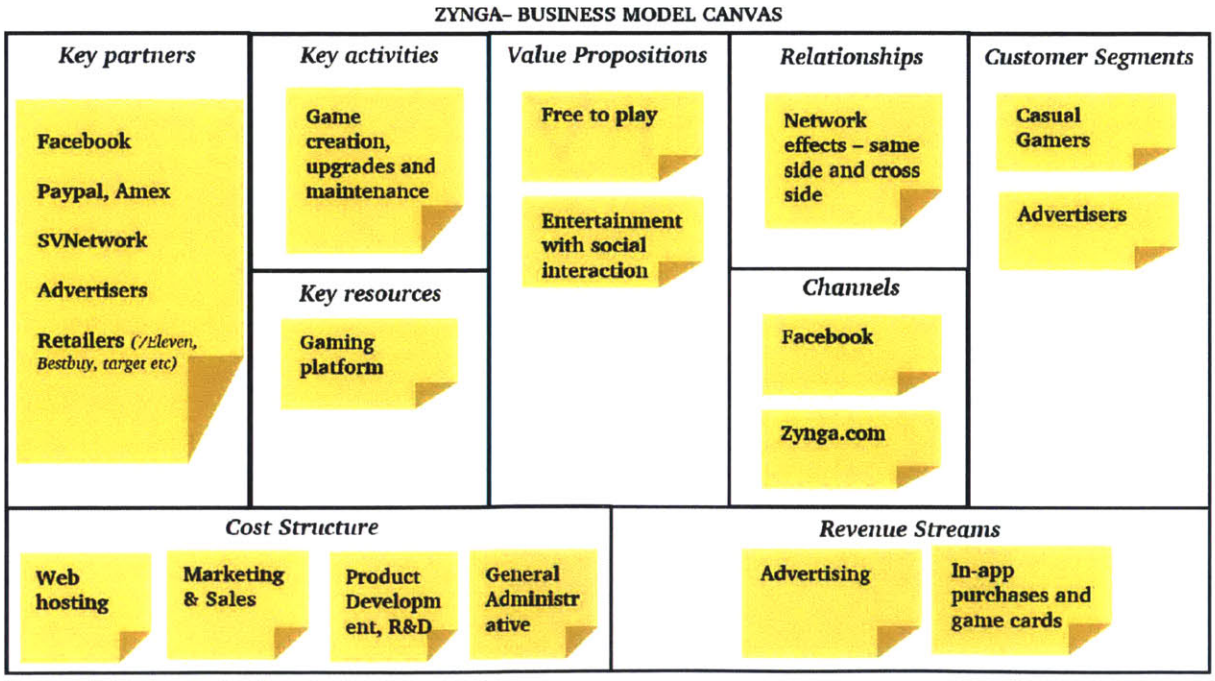
10.1.1 LinkedIn Business Model Canvas

Figure 30 Business Model Canvas - LinkedIn



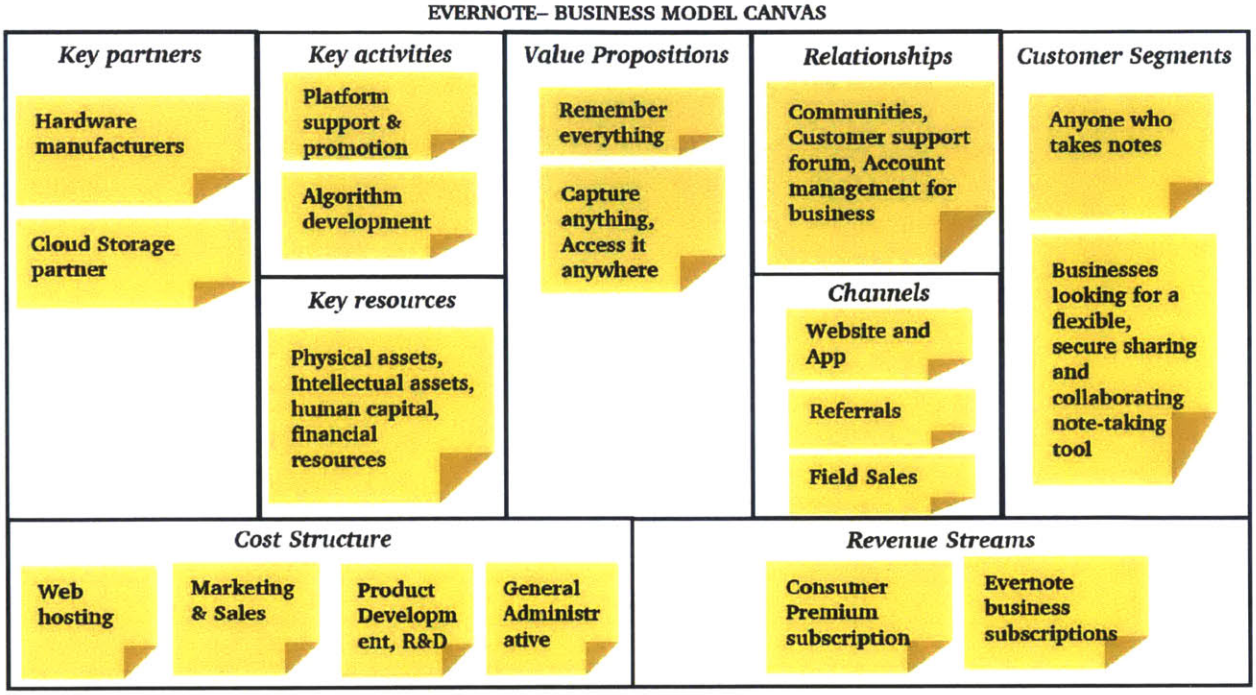
10.1.2 Zynga Business Model Canvas

Figure 31 Business Model Canvas - Zynga



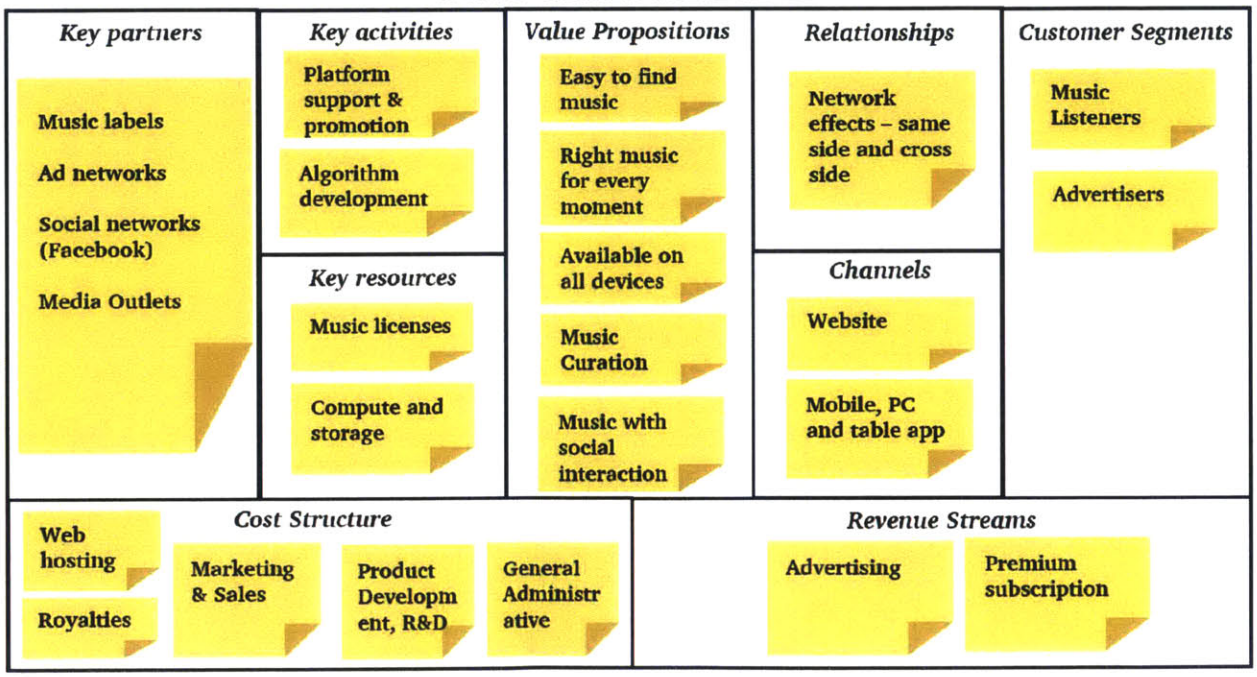
10.1.3 Evernote Business Model Canvas

Figure 32 Business Model Canvas - Evernote



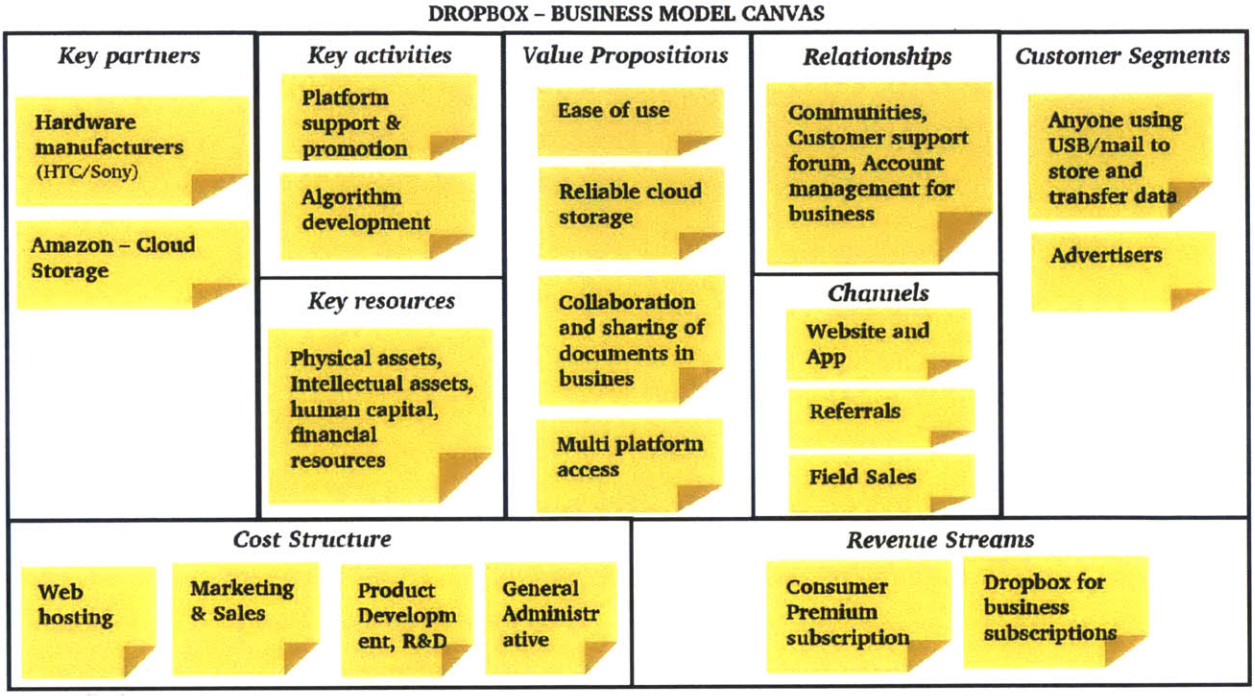
10.1.4 Spotify Business Model Canvas

Figure 33 Business Model Canvas - Spotify



10.1.5 DropBox Business Model Canvas

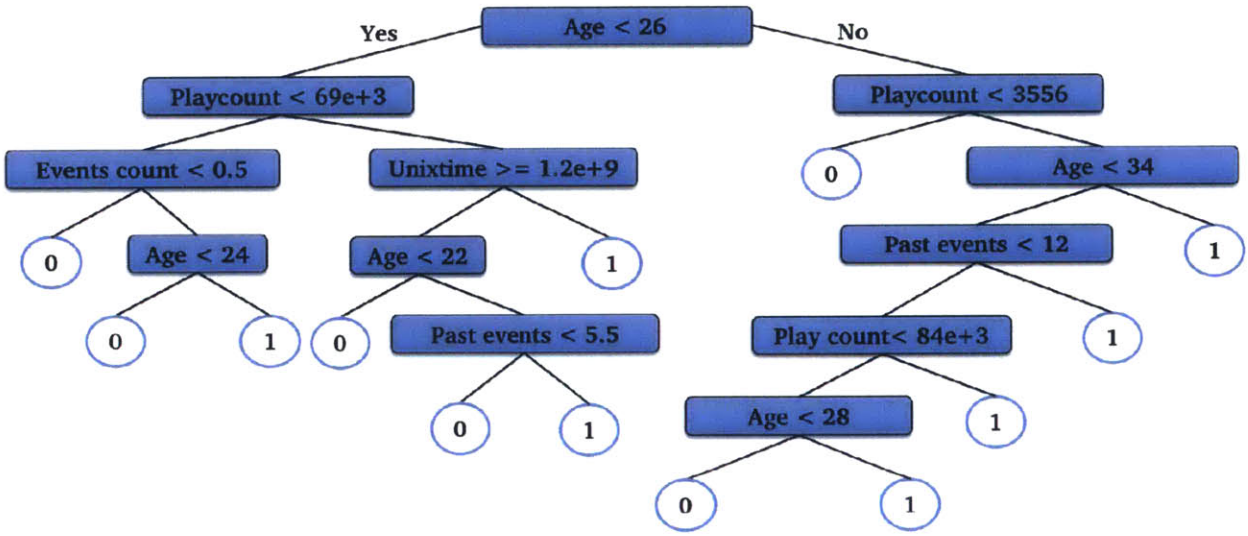
Figure 34 Business Model Canvas - Dropbox



10.2 CART Model with penalty matrix altered

The model is reconstructed with a penalty matrix of $\begin{matrix} 0 & 6 \\ 10 & 0 \end{matrix}$. Accuracy reduces to **65.7%.** But the model's false negative rate reduces to **22%.**

Figure 35 CART Model with Penalty Matrix altered



10.3 R Scripts

10.3.1 CART modelling code

The basic code for predictive CART modeling is included in this section.

```
#################### PREDITIVE MODELING ON LAST.FM ####################

# Reading Last.fm data
lastfm = read.csv("UserData.csv")
lastfm = na.omit(lastfm)
str(lastfm)
#Number of Subscribers
table(lastfm$subscriber)

# Splitting data into training set and test set
install.packages("caTools")
library(caTools)
set.seed(88)

# Randomly split the data
spl = sample.split(lastfm$subscriber, SplitRatio = 0.70)
lastfmtrain = subset(lastfm, spl == TRUE)
lastfmtest = subset(lastfm, spl == FALSE)

# Modelling with weights and penalty matrix
trainweights = 1 + 12 * as.numeric(lastfmtrain$subscriber)
#Penalty Matrix
PenaltyMatrix = matrix(c(0,8,10,0),byrow = TRUE, nrow = 2)
PenaltyMatrix

#################### CROSS VALIDATED CART MODEL ########################
install.packages("rpart")
library(rpart)
install.packages("rpart.plot")
library(rpart.plot)
install.packages("caret")
library(caret)
install.packages("e1071")
library(e1071)

# Defining number of folds
folds = trainControl(method = "cv", number = 10)
# Picking possible values for parameter cp
cpValues = expand.grid(.cp = seq(0.01,1,0.001))
# Performing cross validation
```

```
train(subscriber ~ age + gender + playcount + playlists +
isimageincluded + friendscount + unixtime + eventscount +
pasteventscount + shoutscount, data = lastfmtrain, method = "rpart",
trControl = folds, tuneGrid = cpValues)
# Creating a CART model with this value of cp
lastfmtreeCV = rpart(subscriber ~ age + gender + playcount + playlists
+ isimageincluded + friendscount + unixtime + eventscount +
pasteventscount + shoutscount, data = lastfmtrain, method="class", cp
= 0.0025, weights=trainweights, parms = list(split="information",
loss=PenaltyMatrix))
# CART Tree
prp(lastfmtreeCV)
summary(lastfmtreeCV)
# Making predictions on training set using this model
PredictCV = predict(lastfmtreeCV, type = "class")
# Confusion matrix
tabtrain = table(lastfmtrain$subscriber, PredictCV)
acctrain = sum(diag(tabtrain))/sum(tabtrain)

################ CHECKING MODEL ON TEST DATA ########################

PredictCARTCVTest = predict(lastfmtreeCV, newdata=lastfmtest, type =
"class")
tab = table(lastfmtest$subscriber, PredictCARTCVTest)
acc = sum(diag(tab))/sum(tab)
spec = tab[1,1]/(tab[1,1] + tab[1,2])
sens = tab[2,2]/(tab[2,2] + tab[2,1])
FNR = tab[2,1]/(tab[2,1] + tab[2,2])
FPR = tab[1,2]/(tab[1,2] + tab[1,1])

########################### ROC CURVES ################################
install.packages("ROCR")
library(ROCR)

# Plotting ROC
PredictCARTCVTest = predict(lastfmtreeCV, newdata = lastfmtest)
str(PredictCARTCVTest)
PredictCARTCVTest = PredictCARTCVTest[,2]
pred = prediction(PredictCARTCVTest, lastfmtest$subscriber)
perf = performance(pred, "tpr", "fpr")
plot(perf, colorize =TRUE, print.cutoffs.at=seq(0,1,by=0.1),
text.adj=c(-0.2,1.7))
# Calculating AUC
as.numeric(performance(pred, "auc")@y.values)
```