Data Science for Business
Lecture #6
Freemium Case Introduction

Prof. Alan L. Montgomery

The University of Hong Kong & Carnegie Mellon University, Tepper School of Business

email: alanmontgomery@cmu.edu

1

Case Objectives

Illustrate how to use data-based methods to target customers

Develop a promotional campaign and other insights based upon data mining methods to encourage migrating customers from free to paid

Experience the use of classification methods like logistic regression and decision trees

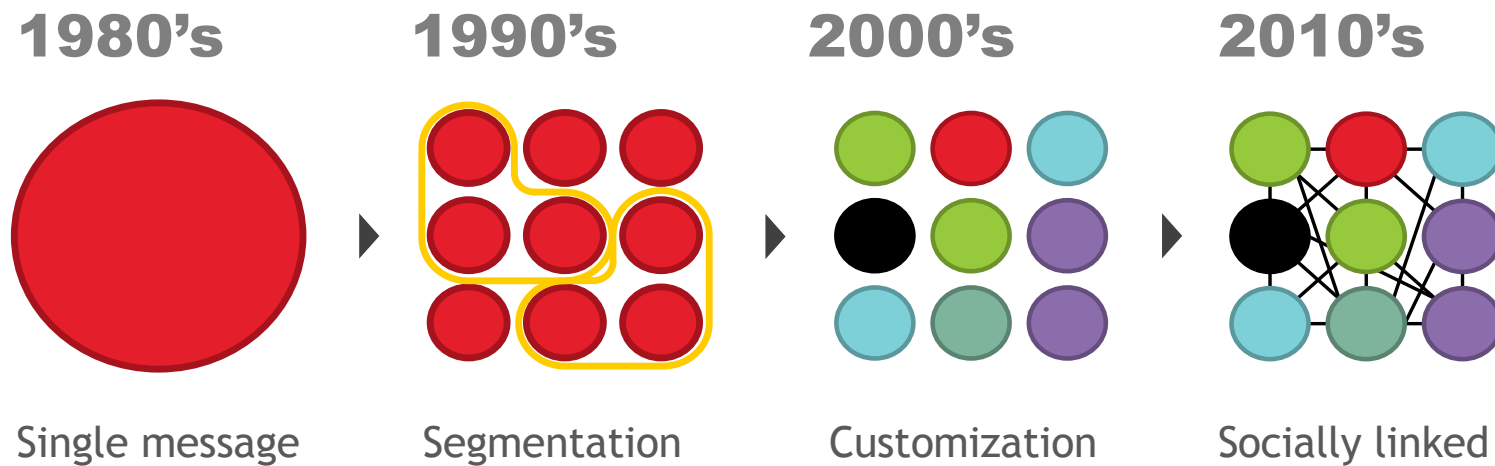Practice the communication of these results to marketing managers

Freemium

"High Note" is a fictitious name for data that comes from a real company. Similar to spotify, Last.fm, or Pandora.

Four Decades
of Consumer Involvement

1980's	1990's	2000's	2010's

Single message   Segmentation   Customization   Socially linked

4

Online Business Freemium Models

How do we get people from free to fee?

Oestreicher-Singer & Zalmanson (2013)

Marketing Objectives

How to monetize freemium?

What's new and different about freemium business models?

Hard to Monetize

Premium users are rare.

> < 3% on last.fm
>
> pandora - 1.6%
>
> dropbox - 4%
>
> evernote - 2%,
>
> linkedIn - 1%
>
> **25% on spotify**

But, it's where the money is: Premium users are

24_x more profitable

on last.fm

What do we know?

Free accounts generate advertising revenue.

Companies that offer a free version of the basic site or features do so to build an audience that will attract advertisers and advertising revenue. Offering a free version of the service is a common strategy to attract users and build a community.

Premium subscribers are more profitable than free users.

Users who pay a premium to subscribe get access to additional and enhanced features, which often include the ability to "turn off" advertising. These users tend to be more profitable. For example, the company on which this case is based showed premium subscribers to be 24 times more profitable compared to free (ad-supported) users.

Premium subscribers are rare.

Premium subscriptions are in the low single-digit percentages of total user populations for these popular freemium services.
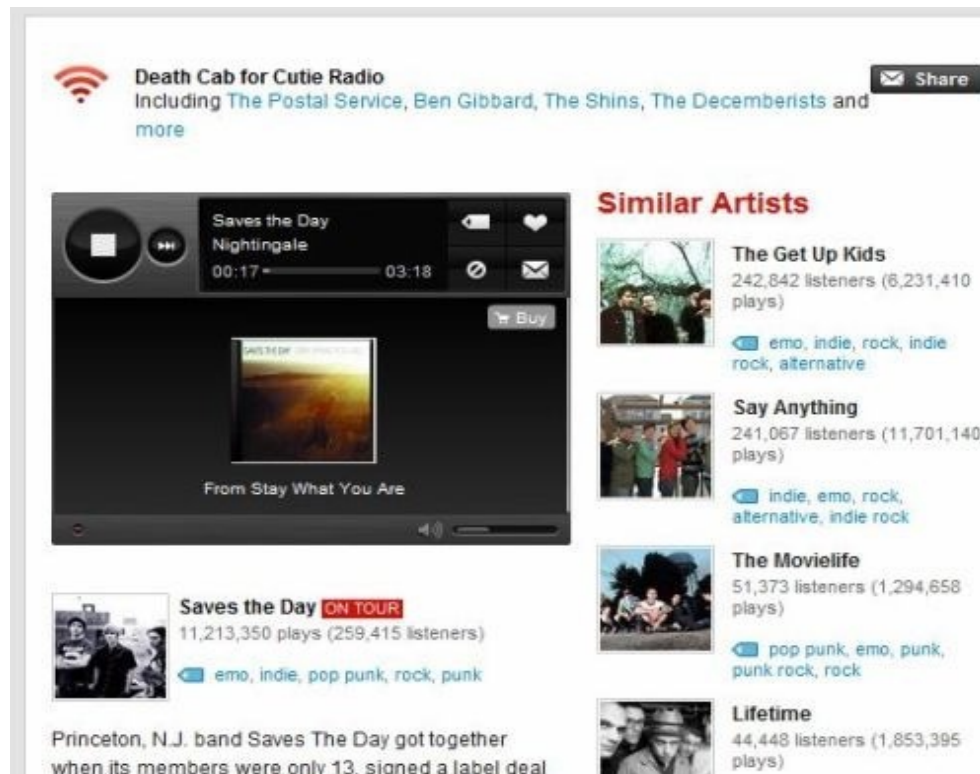
'Typical' High Note User

Last.fm media player is "blog"

shouts, upcoming events, …

Understanding our Data

Demographic characteristics
- ◦ Age, gender, country

Social network characteristics
- ◦ Number of friends a user has on the network

Engagement level data
- ◦ Activities performed when using the service,
- ◦ Number of songs the user listened to, playlists created, "shouts" sent to friends, etc.

Data Dictionary

Variable	Description	Values	Values	Values
net_user	highnote id	tinaj5920	tina machine	zzerbamtt
age	age at the beginning of the current period		22	20
male	if 1 then male else female	0	0	1
friend_cnt	number of friends	20	8	8
avg_friend_age	average age of the friends	30.29	22.57	20.13
avg_friend_male	what proportion of friends are male	0.74	0.43	0.75
friend_country_cnt	Number of different countries this user's friends are from	14	1	2
subscriber_friend_cnt	number of friends who are premium subscribers	1	0	0
songsListened	cumulative number of songs listened til the beginning of the current period	8414	9687	8856
lovedTracks	number of tracks loved at the start of CURRENT	348	194	56
posts	number of Q&A forum posts made at the start of CURRENT	0	0	0
playlists	number of playlists made till the current period	1	1	0
shouts	number of shouts received from other users till the current period	6	8	2
adopter	=1 if the user switched from being FREE to PREMIUM subscriber in CURRENT period (never been premium before)	0	0	1
tenure	how long has the user been on the site (in months)	59	59	30
good_country	= 1 if from US, UK or Germany, otherwise rest of the world	1	1	1

Generic strategy for data understanding

Look at your data
◦ Understand what each variable means
◦ Use your domain knowledge to consider potential relationships
◦ Recode and transform your data as needed

Exploratory analyses
◦ Compute summary statistics
◦ Separate summary statistics for free users and premium subscribers
◦ Graphical summaries

Model your data
◦ Use logistic regression and/or decision trees to gain insights
◦ Choose the "best" model using out-of-sample validation

Make recommendations
◦ Use insights from your model to suggest ways of encouraging upgrades
◦ Notice that you are not simply "predicting" but trying to "understand"

Questions to consider

What are the different types of data that emerge from modern online social communities?

Is there a systematic way to think about this?

What are the commensurate marketing strategies?

Can this big-data actually help?