Large-scale Joint Modeling of Player Attributes In Freemium Mobile Games

Gourab Mukherjee

University of Southern California

21st October 2019

Informs Annual Meeting 2019, Seattle

Joint work with: Trambak Banerjee (USC), Shantanu Dutta (USC) & Pulak Ghosh (IIM Bangalore)

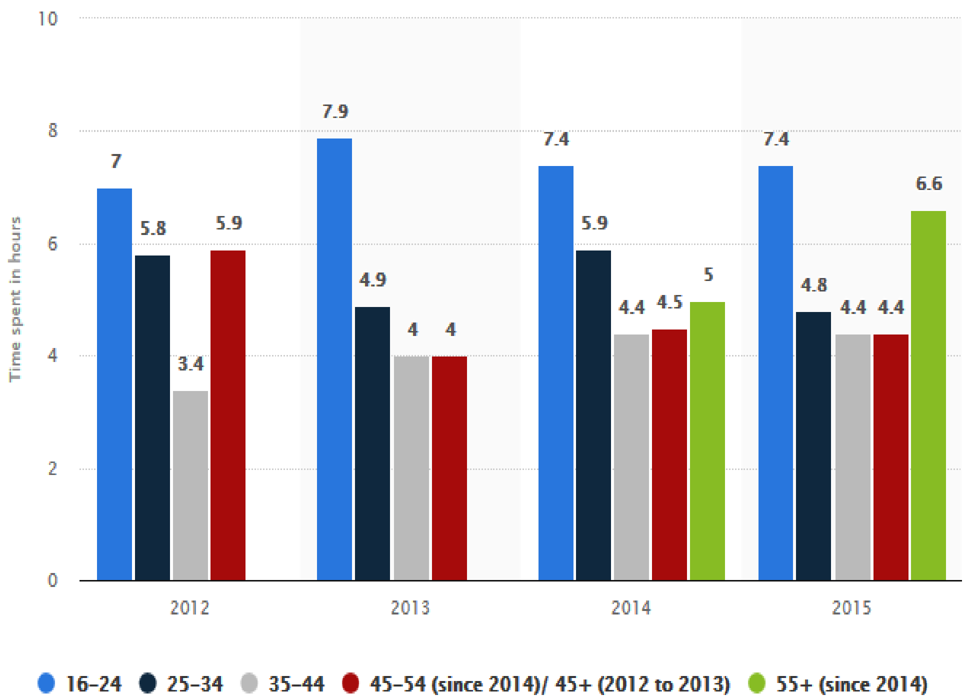
Mobile games - an integral part of modern life

Scientists in Antarctica are downloading mobile games

billion clicked apps categorized in 16 groups: 56% mobile games

Mobile games - an integral part of modern life

How many hours in a typical week would you say you play games?*



DOWNLOAD SETTINGS SHARE

PNG + PDF + XLS + PPT +

DESCRIPTION SOURCE MORE INFORMATION

This statistic illustrates the average weekly time spent playing games in the United Kingdom from 2012 to 2015, broken down by age group (inclusive weekdays and weekends). In 2012, the average volume of gaming among individuals aged over 45 years was 5.9 hours per week. This number fell to 4 hours in 2013. Statista provides a curated dossier of statistics covering the video game market in the United Kingdom.

Expanding Market

GLOBAL GAMES MARKET EXPANDING AT 6.2%

$128.5 billion by 2020.

In 2017, mobile games will generate

$46.1Bn

or **42%** of the global market.

Monetization Challenges

POCKET GAMER.biz | Features (Articles & Opinion) | News (Hot topics) | Events (Networking) | MGU / (Data, Deals &

determined by the application release date. While this can be changed by the developer before release, it usually isn't. Note that due to approval delay, the last few weeks may not be properly represented and are not displayed.

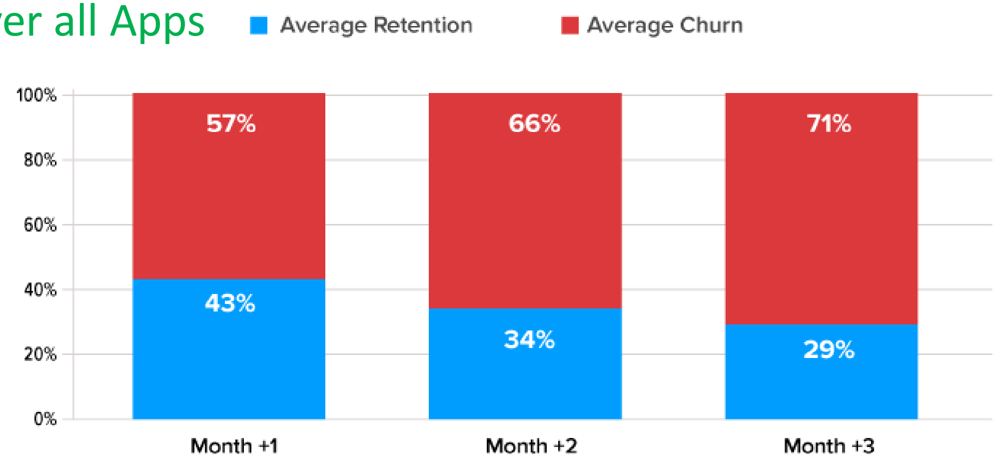
App Submissions By Month

Month	# Apps	# Games	# Total	Avg/Day
2018-03	8,716	1,930	10,646	343
2018-02	7,759	1,868	9,627	311
2018-01	11,267	3,026	14,293	461
2017-12	10,495	2,981	13,476	435
2017-11	10,234	2,633	12,867	415
2017-10	10,244	2,829	13,073	422
2017-09	9,272	2,612	11,884	383

VERY CROWDED MARKET

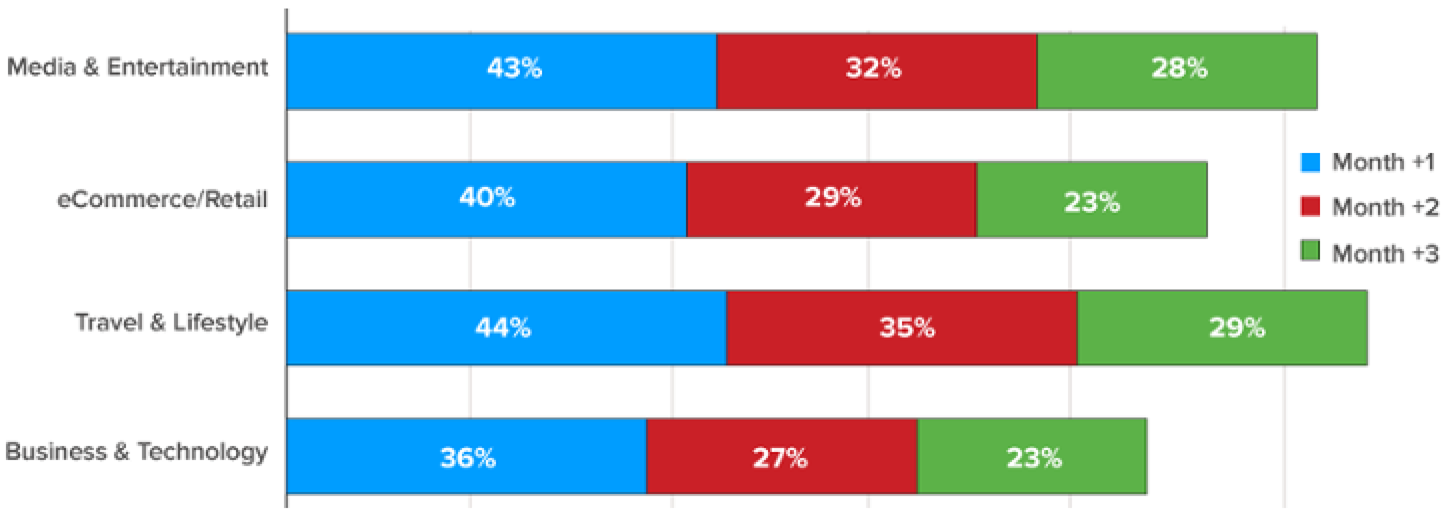
Monetization Challenges

Average over all Apps

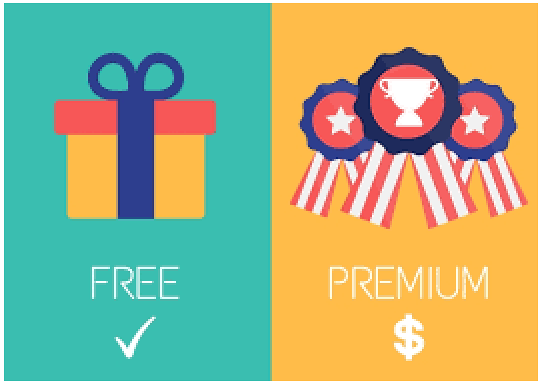


HIGH
CHURN (DROP-OUT)
RATES

Industry Specific



Monetization Challenges



Mobile App Store Revenue Share Worldwide, by Business Model and Platform, Jan 2013 & Nov 2013
% of total

	Jan 2013		Nov 2013	
	Google Play	**Apple App Store**	**Google Play**	**Apple App Store**
Free apps with in-app purchases	89%	77%	98%	92%
Paid apps	6%	11%	1%	4%
Paid apps with in-app purchases	5%	12%	1%	4%

Note: excludes ad revenues
Source: Distimo, "2013 Year in Review," Dec 17, 2013

167322 www.eMarketer.com

People Spend Way More

On Purchases In Free Apps

Than They Do

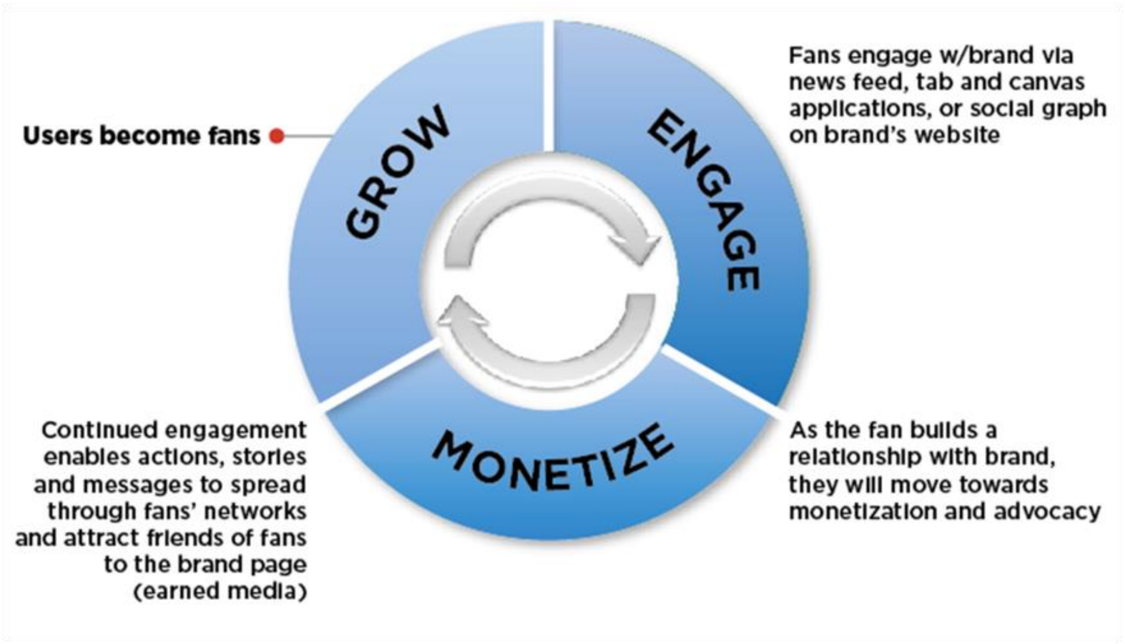
Downloading Paid Apps

Monetization Strategies

1) Monetization through Premium Purchase: rare unless it's a popular game.

2) Monetization through ads:
- Banner ads are omnipresent and easy to place in games but do not earn much.
- Video ads make the highest gain but lowest click through rate.

3) Monetization through social media: Provide in-game incentive to do social media errands

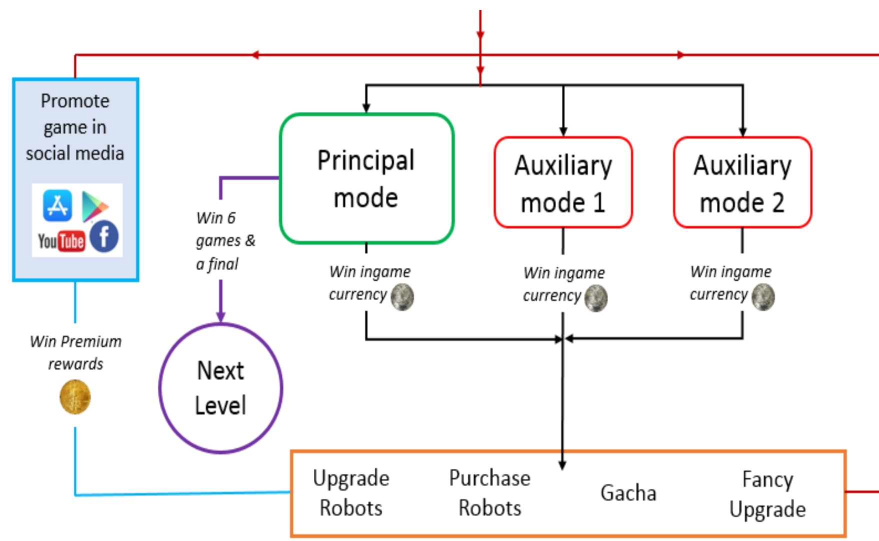


Our Game: Robot Boxing

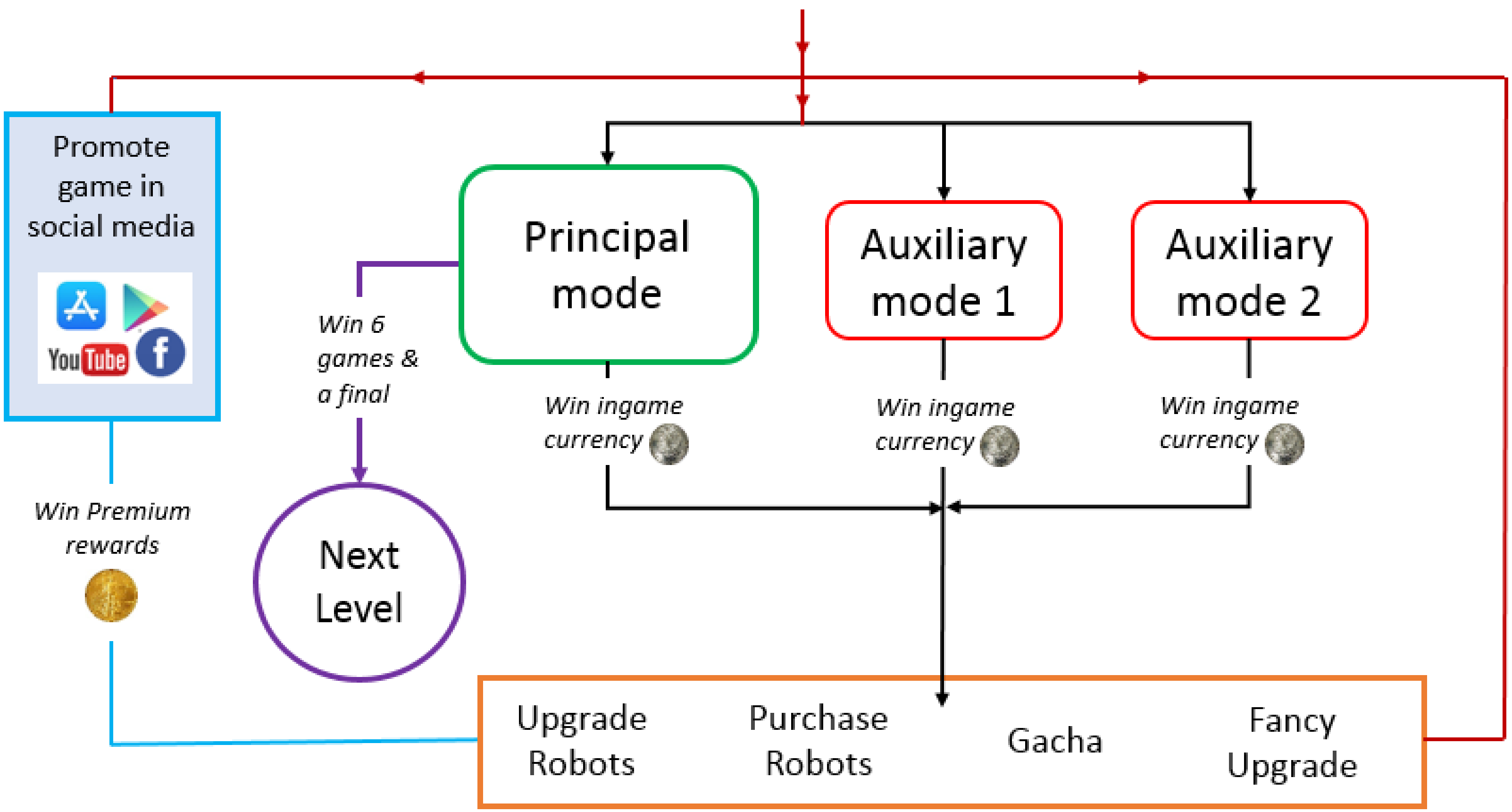
Users use robot avatars to fight other robots till one is destroyed

3 fighting modes:

1. Championship mode (Principal Game Level Progression mode)
2. Free Sparing mode (Auxiliary 1)
3. Time attacking mode (Auxiliary 2)



Our Game: Robot Boxing



Our Data-Set

Daily activities of 55K players over 60 days

- Longitudinal data over 60 time points
- Time to event: players drop out of the game after some time

We would like to model the following for an arbitrary player in gaming portals:

Time Spent (ACTIVITY)

Total time spent playing the game on day i

Revenue generated

Cash Spent for premium purchases +
Substantial Game promotion by Social Media Engagement

Dropouts:

If a player does not login for 30 consecutive days we say he has dropped out

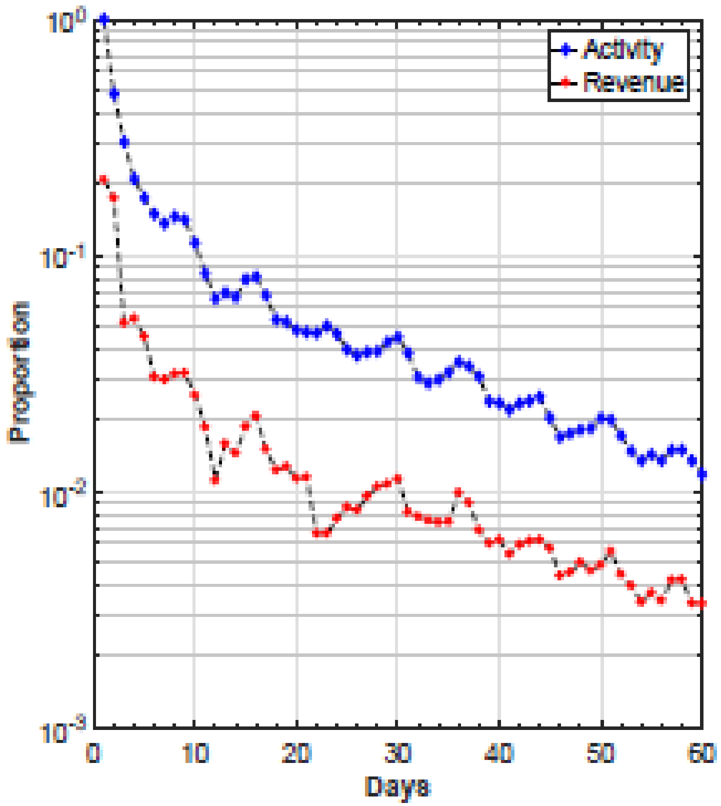
Our Data-Set: Responses

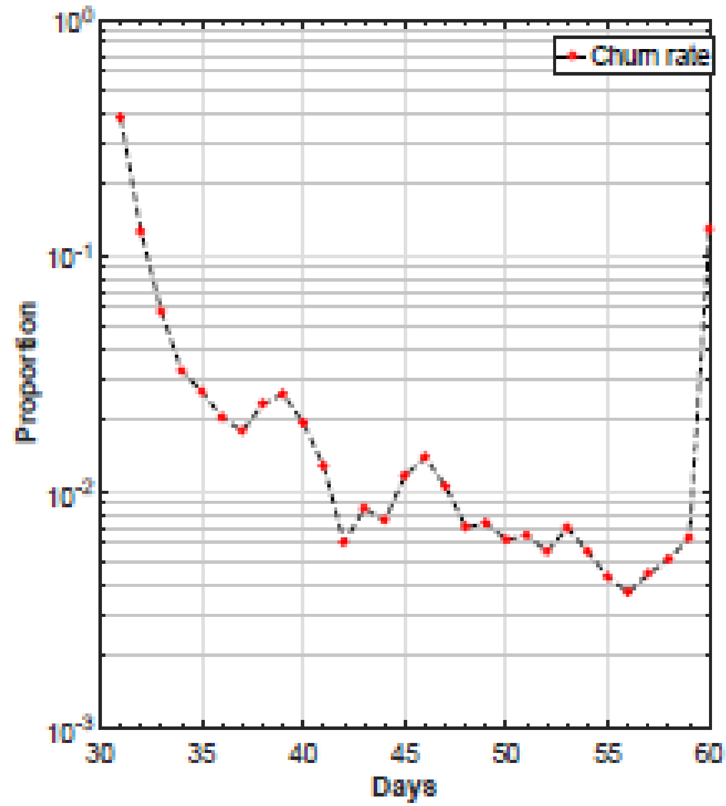
Time Spent (ACTIVITY)

Revenue generated

Dropouts

- **Freemium behavior implies activity very rarely translates into Revenue**
- **Drop-out rates are very high.**





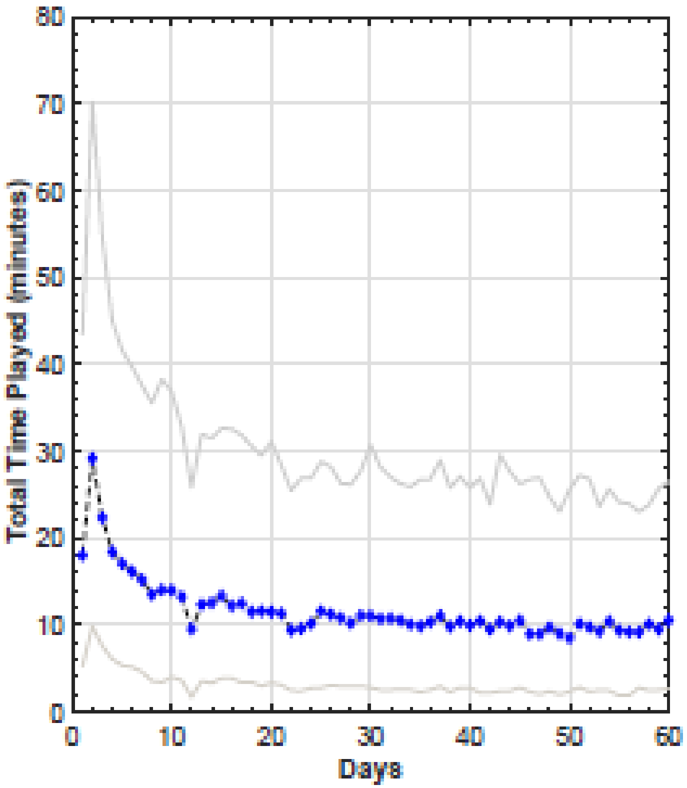
Our Data-Set: Responses

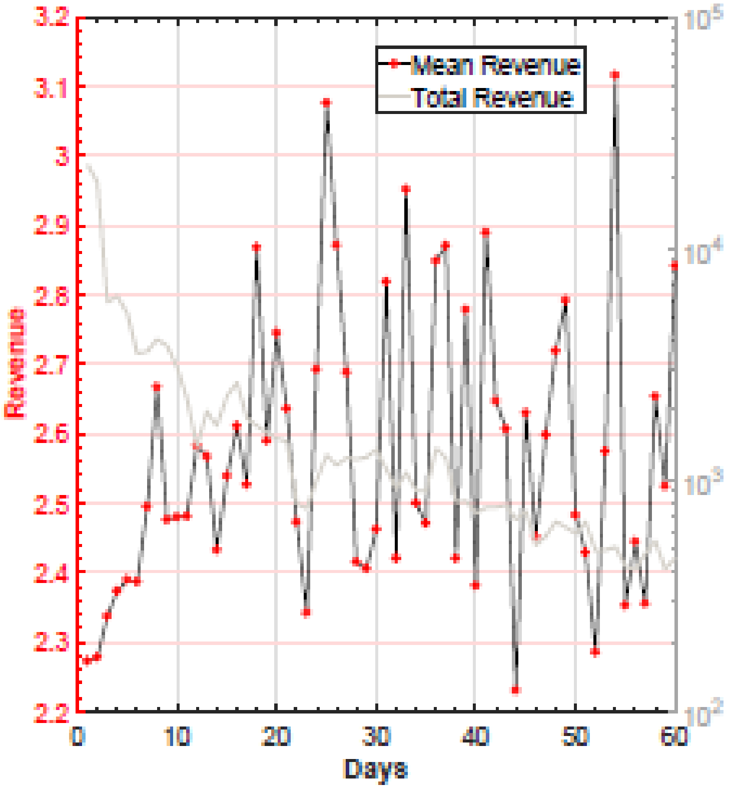
Time Spent (ACTIVITY)

Revenue generated

Dropouts

- **Freemium behavior implies activity very rarely translates into Revenue**
- **Drop-out rates are very high.**





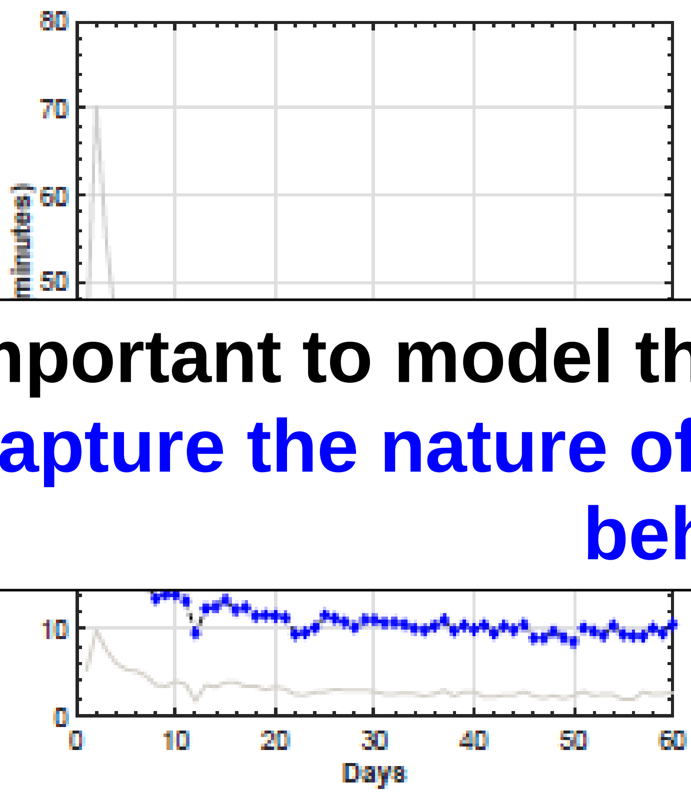
Our Data-Set: Responses

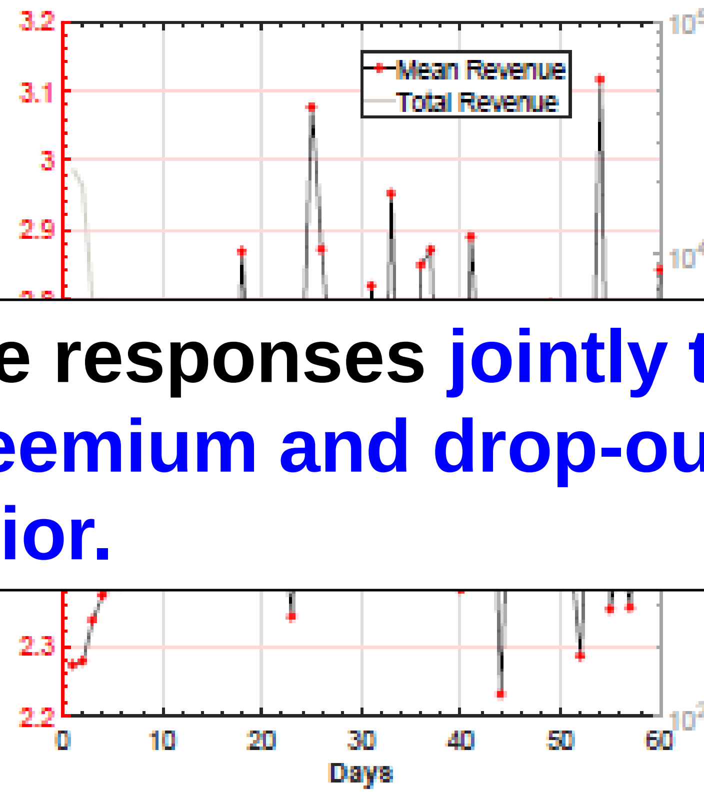
Time Spent (ACTIVITY)

Revenue generated

Dropouts

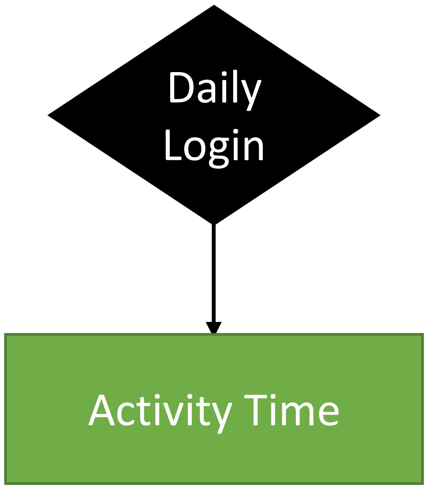
- **Freemium behavior implies activity very rarely translates into Revenue**
- **Drop-out rates are very high.**





Important to model these responses jointly to capture the nature of freemium and drop-out behavior.

Joint Modeling Framework



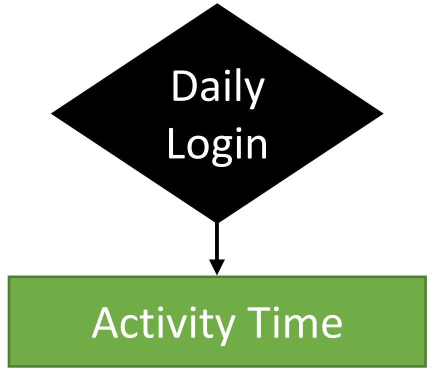
Daily Login: binary variable

If positive, then it lead to positive Activity time

Activity Time is a continuous random variable but it has Zero-value if the player has not logged in.

Model Zero-inflated Activity Time by a mixture distribution:

Joint Modeling Framework



Daily Login: binary variable

If positive, then it lead to positive Activity time

Activity Time is a continuous random variable but it has Zero-value if the player has not logged in.

Model Zero-inflated Activity Time by a mixture distribution:

- Atom of probability **login.prob** at 0
- Continuous distribution at non-zero values which are modeled by **log-normal**
- We use generalized linear mixed models (glmm) with **fixed &** random effects

$\boldsymbol{\beta}$: $\boldsymbol{fixed}$; b: random

fixed effect: does not have individual player effect

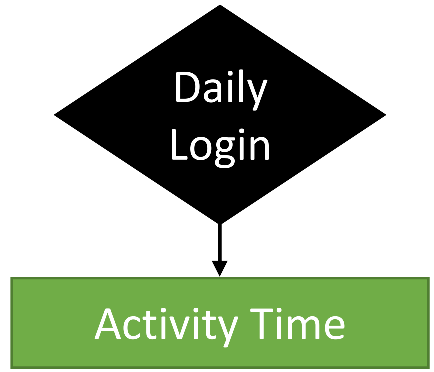
Random effect: model individual player characteristics; but assumed to have a higher level contiguity among players

For player k:

login.prob = logit $(X_F\boldsymbol{\beta}_1 + X_R b_{1,k})$

Activity Time = login.prob δ_0 **+ (1 - login.prob) lognormal**$(X_F\boldsymbol{\beta}_2 + X_R b_{2,k})$

Joint Modeling Framework



Daily Login: binary variable

If positive, then it lead to positive Activity time

Activity Time is a continuous random variable but it has Zero-value if the player has not logged in.

Model Zero-inflated Activity Time by a mixture distribution:

- Atom of probability **login.prob** at 0
- Continuous distribution at non-zero values which are modeled by **log-normal**
- We use generalized linear mixed models (glmm) with **fixed &** random effects

$\boldsymbol{\beta}$: $\boldsymbol{fixed}$; $\boldsymbol{b}$: random

fixed effect: does not have individual player effect

Random effect: model individual player characteristics; but assumed to have a higher level contiguity among players

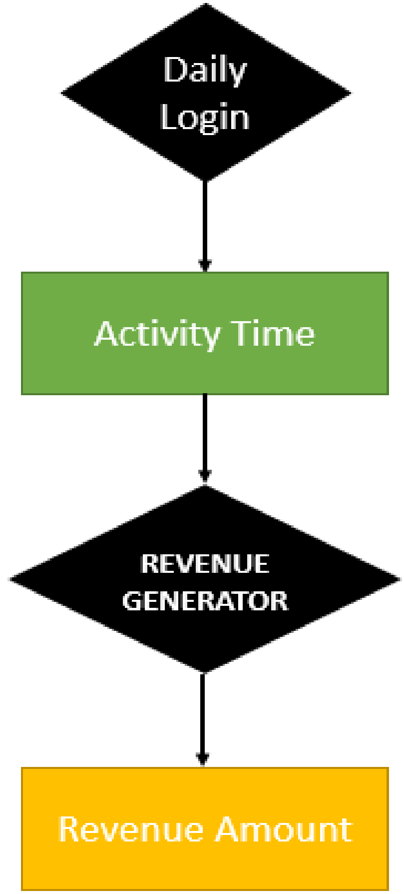
For player k:

{($b_{1,k}$, $b_{2,k}$): k=1,2,….. } i.i.d. N(0, Σ)

$$\text{login.prob} = \text{logit}\,(X_F\beta_1 + X_R b_{1,k})$$

$$\text{Activity Time} = \text{login.prob}\ \delta_0 + (1 - \text{login.prob})\ \text{lognormal}(X_F\beta_2 + X_R b_{2,k})$$

A Joint Model



Daily Login: binary variable

If positive, then it lead to positive Activity time

Positive Activity time seldom Leads to positive revenue

So, revenue is not only zero-inflated but **extremely zero inflated.**

Revenue.probability = Probability of generating positive revenue given positive activity time.

CURES ZERO INFLATION

revenue.probability = logit ($X_F\beta_3$ +$X_R b_{3,k}$)

Unconditional probability of positive revenue (r.prob)

= revenue.probability*login.prob

Revenue = (1-r.prob) δ_0 + r.prob * lognormal($X_F\beta_4$ +$X_R b_{4,k}$)

For player k:

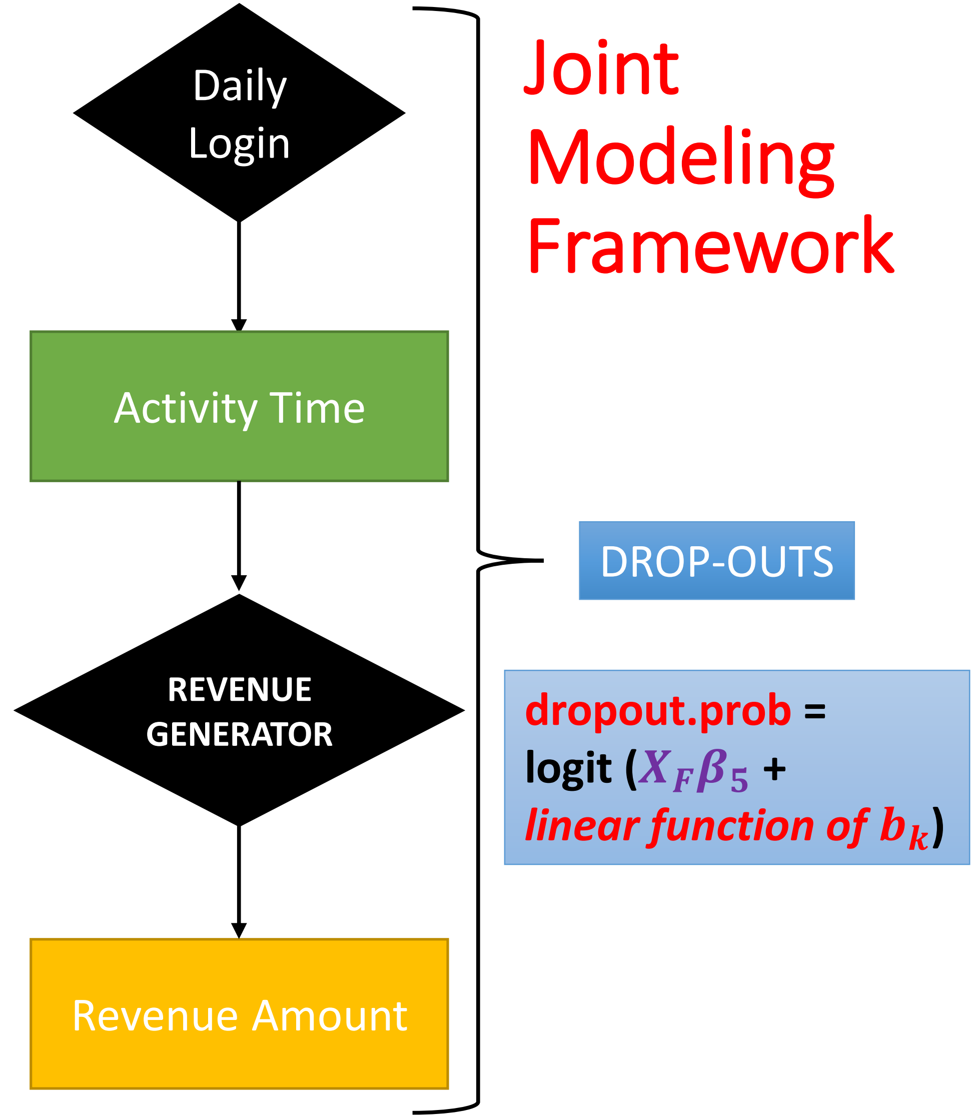
login.prob = logit ($X_F\beta_1$ + $X_R b_{1,k}$)

Activity Time = (1-login.prob) δ_0 +
login.prob*lognormal($X_F\beta_2$ + $X_R b_{2,k}$)

revenue.probability = logit ($X_F\beta_3$ + $X_R b_{3,k}$)

r.prob = revenue.probability*login.prob

Revenue = (1-r.prob) δ_0
+ r.prob * lognormal($X_F\beta_4$ + $X_R b_{4,k}$)



Random effects: $b_k = (b_{1,k}, b_{2,k}, b_{3,k}, b_{4,k})$

$\boxed{\{b_k : k=1,2,\ldots..\} \text{ i.i.d. } N(0, \Sigma)}$

STATISTICAL CHALLENGES

MODELING ISSUES

- Extreme Zero-inflation
- Variable Selection: hierarchy of fixed and random effects
- Structural Constraints

COMPUTATIONAL CHALLENGES

- Big data: We use Divide and Conquer
- Prediction from such models

Domain & Promotional Constraints

Domain Knowledge:

Weekend effect. In games of these kind, players are more active on weekends than on week days.

Promotion Effects:

6 different kinds of promotion mainly to increase activity was used. They have different doses.

Strategy	Description	No. of days	%
0	No strategy	20	33.33
1	More energy or rewards	8	13.33
2	Sale of boss robots	4	6.67
3	Discounts on powerful robots	8	13.33
4	Holiday sale	7	11.67
5	Promotion via emailing and messaging	5	8.33
6	Sale on all robots	8	13.33

Promotion 6 >= Promotion 2

Promotions effects are incorporated as categorical fixed effects

We can incorporate domain & promotion constraints via convexity constraints on the fixed effects.

Variable Selection

- We use penalized quasi-likelihood criterion for variable selection. Penalty chosen by cross validation.

- The penalty uses a ℓ_1 penalty on the fixed effect coefficients and the diagonal of the covariance of the random effects. They have different weights. The criterion is:

$$\max_{\theta,\, \Sigma \succ 0} \ \ell^{\mathsf{ql}}(\Theta) - n\lambda \sum_{s=1}^{5} \sum_{k=1}^{p} \left(c_{sk} |\beta_{sk}| + d_{sk}\, \Sigma_{kk}^{(s)}\, \mathbb{I}\{k \in \mathcal{I}_c\} \right)$$
$$\text{subject to } \mathfrak{f}^{(s)}(\beta^{(s)}) \le 0,\ s = 1, \cdots, 5 \ .$$

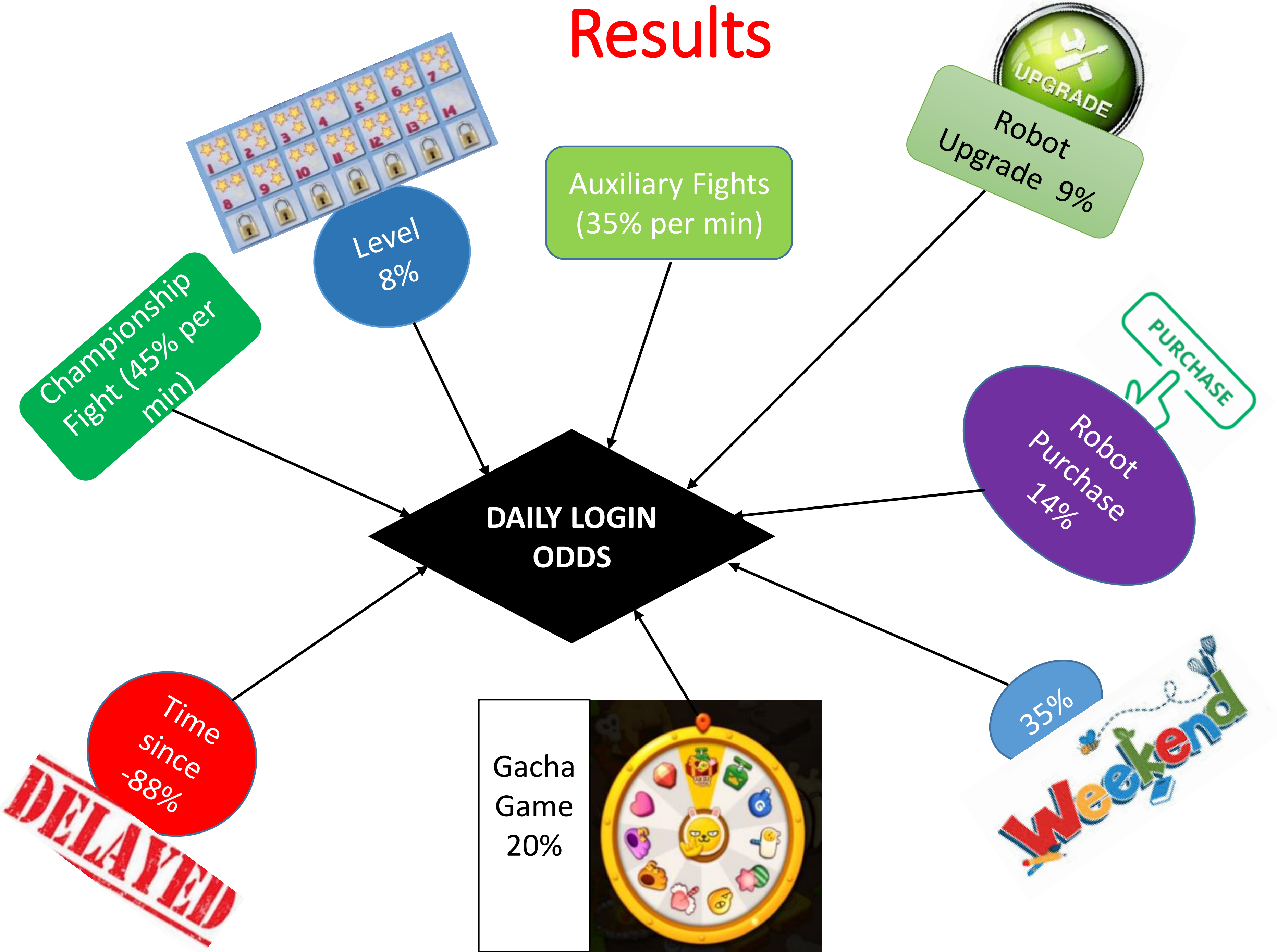
- The composite effects impose the following hierarchy between fixed and random effects: a random component can have a non-zero coefficient only if its corresponding fixed effect is non-zero.

- For that purpose we use different weights on the penalty and reweight them in every iteration.

Results

	DAU $\widehat{\beta}^{(1)}$	Total Time $\widehat{\beta}^{(2)}$	Revenue $\widehat{\beta}^{(4)}$	Churn $\widehat{\beta}^{(5)}$
Intercept	-4.648*	0.932*	0.953*	-1.902
avg_session_length		0.269*		
p_fights	0.378*	0.169*		
a1_fights	0.303*	0.379*		
a2_fights	0.334*	0.216*		
level	0.084*	0.304*		
robot_played	–	–	–	–
gacha_sink		0.201*		0.129
gacha_premium_sink	–	–	–	–
pfight_source		0.144*		
a1fight_source	0.030	-0.239*		
a2fight_source	-0.240*	-0.192*	0.331*	
gacha_source	0.182*	-0.212*		
gacha_premium_source	–	–	–	–
robot_purchase_count	0.134*			
upgrade_count	0.093*	0.112*		
lucky_draw_wg	–	–	–	–
timesince	-2.065*	-0.641*		3.502
lucky_draw_og	-0.230*			
fancy_sink	–	–	–	–
upgrade_sink	0.037*			
robot_buy_sink	–	–	–	–
gain_gachaprem	–	–	–	–
gain_gachagrind	-0.127*	0.180*		
weekend	0.302*	0.358*		
promotion 1			-1.153	-0.894
promotion 2	0.178	0.134		-0.934
promotion 3	-0.129		-1.791	-3.500
promotion 4			-3.345	-0.673
promotion 5				0.828
promotion 6	0.290	0.249	-2.389	-1.509

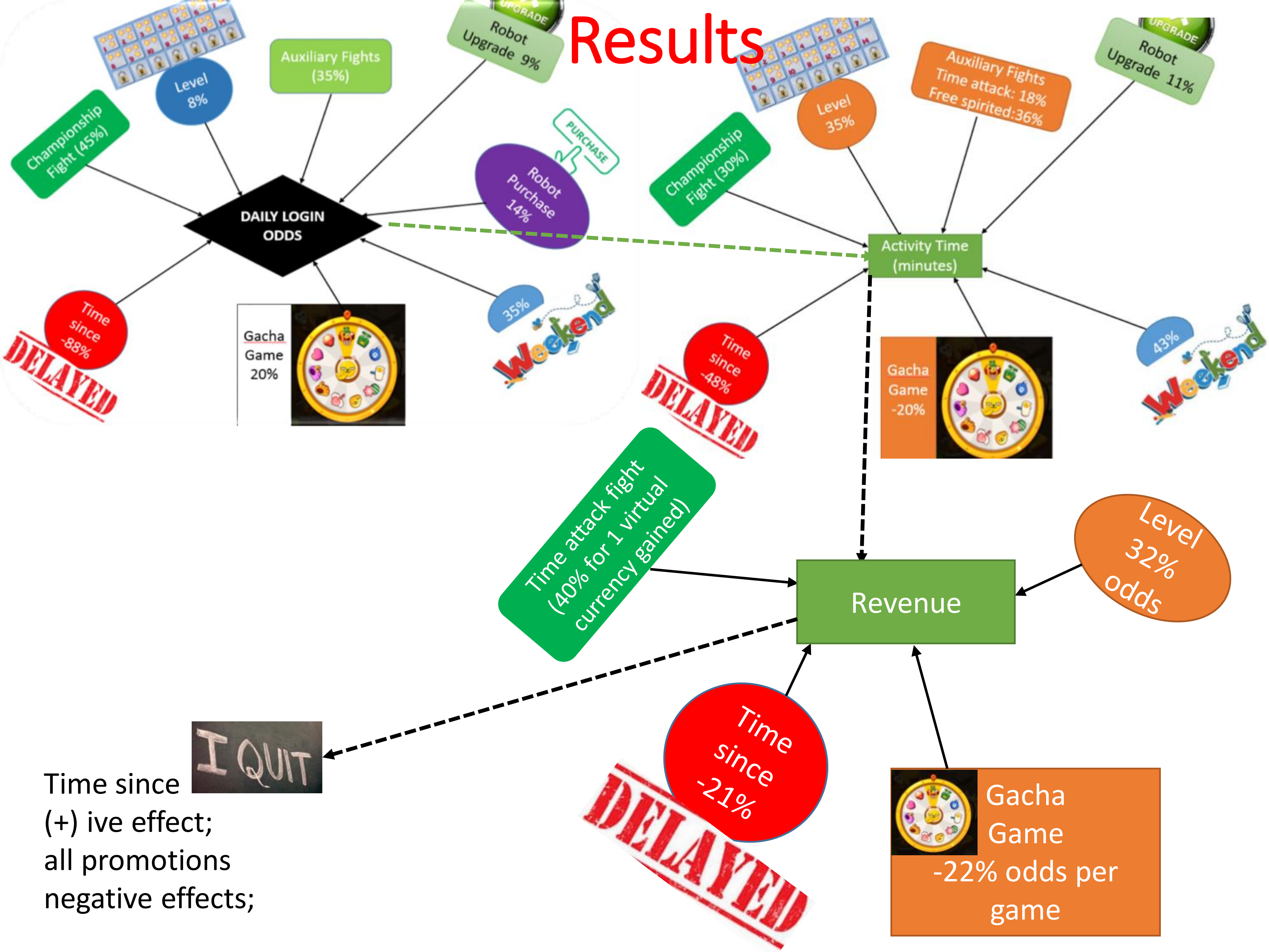
Results


UPGRADE
Robot Upgrade 9%
Auxiliary Fights (35% per min)
Level 8%
Championship Fight (45% per min)
PURCHASE
Robot Purchase 14%
DAILY LOGIN ODDS
Time since -88%
DELAYED
Gacha Game 20%
35%
Weekend

Results

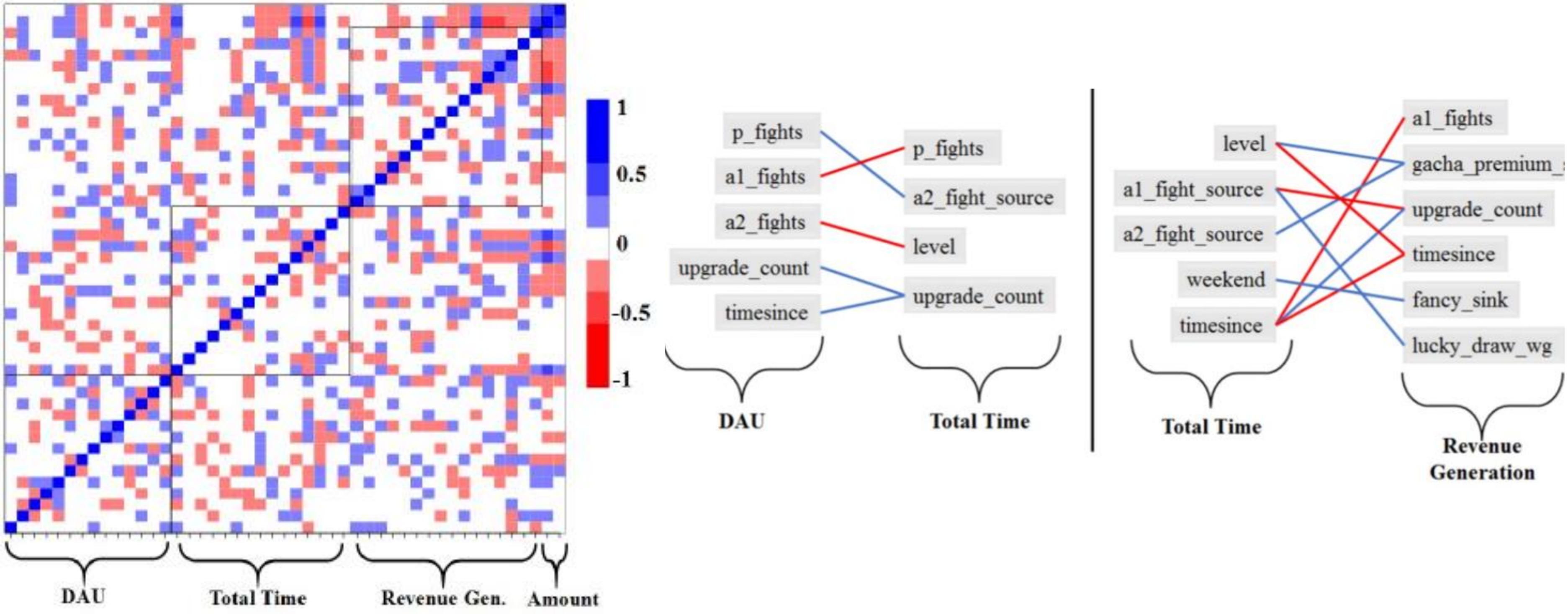
DAILY LOGIN ODDS
Championship Fight (45%)
Level 8%
Auxiliary Fights (35%)
Robot Upgrade 9%
UPGRADE
PURCHASE
Robot Purchase 14%
35%
Weekend
Gacha Game 20%
Time since -88%
DELAYED
Activity Time (minutes)
Championship Fight (30%)
Level 35%
Auxiliary Fights Time attack: 18% Free spirited:36%
Robot Upgrade 11%
UPGRADE
43%
Weekend
Gacha Game -20%
Time since -48%
DELAYED

Results



Time since
(+) ive effect;
all promotions
negative effects;

Results



Heatmap of the 47 dimensional correlation matrix estimate. On the horizontal axis are the selected composite effects of the four sub-models: DAU, Total Time, Revenue Generation & Amount.

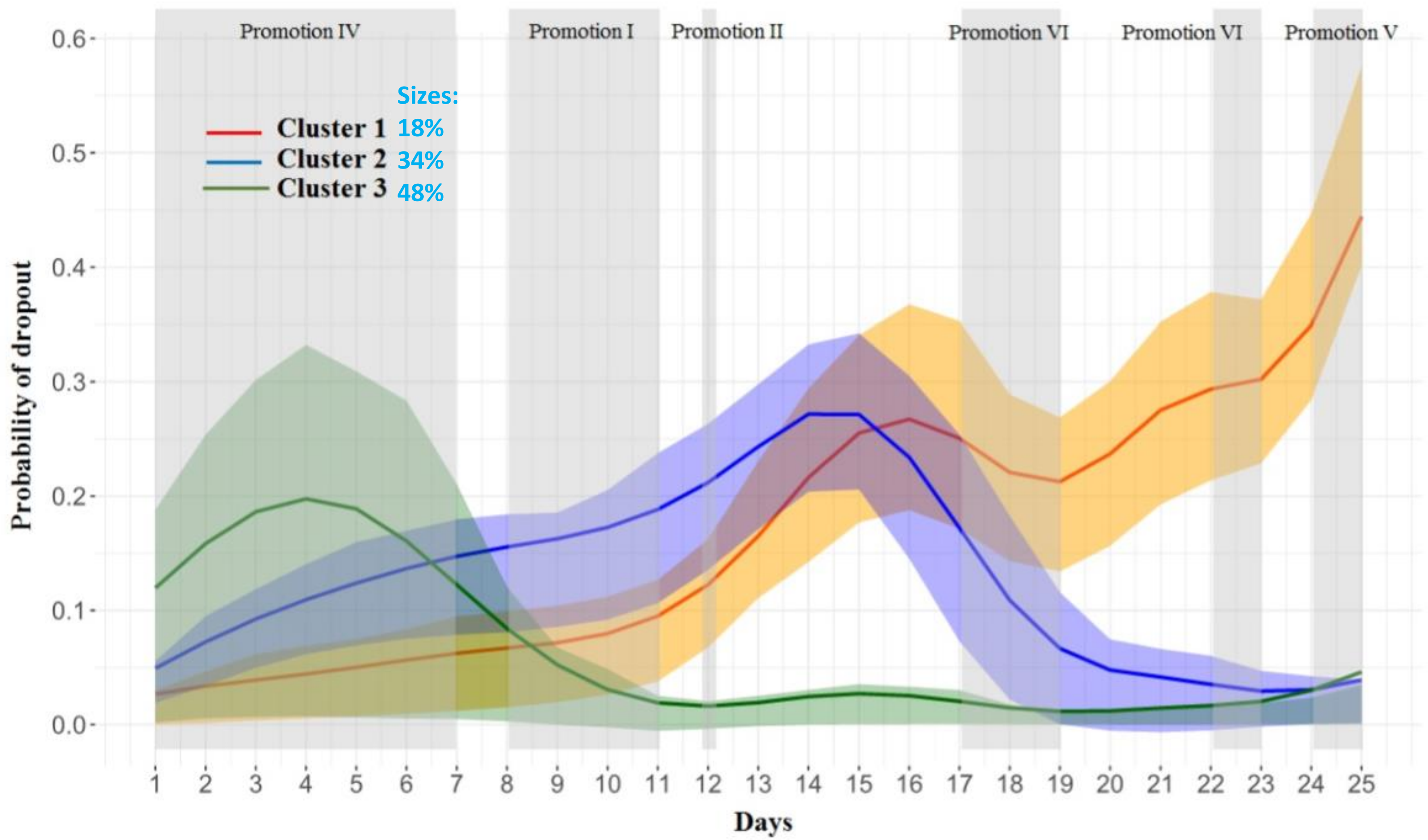
Significant correlations are found across models.

Better Predictive Performance Over Benchmarks

Table 2: Results of predictive performance of CEZIJ model and Benchmarks I, II. For activity and engagement indices, the false positive (FP) rate / the false negative (FN) rate averaged over the 30 time points are reported. For non-zero total time and engagement amounts, the ratio of prediction errors (2) of Benchmarks I and II to CEZIJ model averaged over the 30 time points are reported.

Sub-model	Benchmark I	Benchmark II	CEZIJ
Activity Index	1.19% / 6.33%	5.92% / 4.15%	5.86% / 4.12%
Positive Total Time Played	4.662	1.041	1
Engagement Index	0.05% / 1.89%	3.56% / 1.48%	3.54% / 1.47%
Positive Engagement Amount	1.217	1.067	1

Segmentation based on Joint Model



THANKS
☺

Reference

Banerjee T, Mukherjee G, Dutta S and Ghosh P. *A Large-scale Constrained Joint Modeling Approach For Predicting User Activity, Engagement And Churn With Application To Freemium Mobile Games.* Journal of American Statistical Association, 2019

MATLAB TOOLBOX: CEZIJ