Fantasy, Farms, and Freemium What Game Data Mining Teaches Us About Retention, Conversion, and Virality

Jim Whitehead



Software Introspection Laboratory University of California, Santa Cruz

Why study games?

Facebook games have discovered powerful techniques for quickly gaining large number of players.

Launch of CityVille



- On December 2, 2010, Zynga launched CityVille
 - A social network based city simulation game, similar to SimCity
- In its first 24 hours, over 290,000 people played the game
 - Organic growth, mostly from players sharing status updates and inviting their friends
- After 8 days, there were 6 million people playing the game every day
- Currently around 19 million players every day, with 88.9 million players in the last month
- Among the most successful software launches ever

Software is becoming volitional

Increasingly, software use is an enjoyable leisure activity, not some tool people have to use.

Examples of Volitional Software

- Games are volitional
 - The quintessential example of leisure-time software
- Most phone and tablet apps are volitional
 - While some apps are serious tools, many others are there for fun
- Many web sites are volitional
 - Facebook, YouTube, Flickr, blogs, news sites,



Games are networked

- Games are increasingly networked, and played over the Internet
- Game players generally do not have strong concerns about privacy of their game play
 - This may change after recent PSN security problems
- Game companies are starting to persistently record gameplay telemetry for most games
- Creates an opportunity to learn how people play games at a fine grain level









Overview

- This talk explores three aspects of games and data mining
- Mining gameplay data to be more efficient at making game software
 - Project Gotham Racing 4
- Understanding how to structure games to acquire new users quickly
 - CityVille
- Understanding how game design decisions affect player retention
 - Madden NFL 2011







Project Gotham Racing 4

- Car and motorcycle racing videogame
- Single and multiplayer races
 - Multiplayer quick races
 - Arcade mode
 - Time attack challenge
 - Racing against ghosts
 - Ranked matches
- Career mode
 - Player earns money by competing in races
 - Unlocking of cars and races over time



PGR4 Box Art Bizarre Creations (2007)

Vehicles and Routes in Project Gotham Racing 4

I34 different vehicle types

- Organized into 7 classes A-G
- A: high performance, difficult to master
- G: lower performance, easier to drive



Race tracks

- 9 in-game locations
 - Tokyo, New York, London, Las Vegas, Nürburgring, Shanghai, St. Petersburg, Quebec City, Macau, Michelin Test Track
- 121 routes spread over these locations



PGR4 Street Race



Business of PGR4

 Actual costs and revenues from PGR4 are not publically available



- But...
 - In July, 2007, Bizarre Creations Business Director Brian Woodhouse

"...admitted the studio has already run up huge costs creating Project Gotham Racing 4," and has, "spent a fortune building this game already."

Would it have been possible to develop PGR4 for less money, and still have players be very satisfied?

Analysis of PGR4 Data

- Over the summer of 2010 four people analyzed PGR4 data
 - Ken Hullett (UCSC), Nachi Nagappan (Microsoft Research), Eric Schuh (Microsoft Game Studios), John Hopson (Bungie Studios)
 - See their NIER paper at ICSE 2011, "Data Analytics for Game Development"
- Start of Race dataset
 - Contains 3.1 million entries, once for each time a players starts a race
- Information recorded
 - Type of event
 - Route selected
 - Vehicle selected
 - Number of vehicles in race
 - Player's career rating
 - Number of previous events completed by player
 - Total kudos earned by player



PGR4 Findings: Game Modes

| Game Mode | Races | % of total |
|--------------------------------|-----------|------------|
| Offline career | 1,479,586 | 47.63% |
| Arcade | 566,705 | 18.24% |
| Network Playtime | 584,201 | 18.81% |
| Network Online Career | 193,091 | 6.22% |
| Single Player Playtime | 185,415 | 5.97% |
| Time Attack | 43,942 | 1.41% |
| World Challenge Mode | 36,581 | 1.18% |
| Network Tournament Qualify | 13,847 | 0.45% |
| Network Tournament Elimination | 2,713 | 0.09% |

Four game modes are used by less than 1.5% of the player population Two are used by less than 0.5%.

PGR4 Findings: Event types

 There are 29 total event types, each being a specific kind of challenge within a mode

| Event Type | Races | % of Total |
|-----------------------------------|---------|------------|
| Street Race | 795,334 | 25.60% |
| Network Street Race | 543,491 | 17.50% |
| Elimination | 216,042 | 6.95% |
| Hotlap | 195,949 | 6.31% |
| ••• | | |
| Testtrack Time | 7,484 | 0.24% |
| Networked Cat and Mouse Free Roam | 3,989 | 0.13% |
| Cat and Mouse | 53 | 0.00% |

I2 of the 29 event types were used in less than 1% of races

PGR4 Findings: Routes

- Within PGR4, there are 9 in-game locations, but many of these locations have multiple routes
 - For example, different configurations of city streets within the location of Quebec
- Findings:
 - 47 of the routes (39%) were each used in less than 0.5% of races
 - I9 of the routes (16%) were each used in less than 0.25% of races
 - The 47 routes which individually used in less than 0.5% of races account as a group for 13% of overall usage





PGR4 Findings: Cars

- Out of 134 unique vehicles, 50 were used in less than 0.25% of races
 - I6 were used in less than 0.1%
- Each vehicle represents a significant investment
 - 3d modeling and texturing
 - Play testing and performance tweaking
- Could reduce number of vehicles by more than 20% and still have box say "game contains more than 100 vehicles"





Long tail of content in PGR4

- Across many types of content (game modes, event types, routes, cars) in PGR4, the same trend:
 - Some content used quite a bit
 - A long tail of content that is used infrequently

- Clear implication:
 - A successor to PGR4 could save substantial development cost by eliminating little used content and play modes
 - Effort spent on performing data mining of player data would have clear and large return on investment
 - ~\$50-100k in analysis yields an estimated \$0.5m-\$2m in potential savings
- Interesting to think about
 - Instead of a pre-packaged game on a disk, what if the game were online...
 - ... and could be tweaked based on this research to increase gameplay of little used content?

Understanding how to structure games to **acquire new users** quickly



Goal of CityVille

- Core goal: build up your city
 - Not well motivated: assumption is if you're playing the game, you find this intrinsically satisfying
- Attract people
 - Build houses
 - Costs money
- Businesses make coins
 - Require supplies
- Farms make supplies



Game is comprised of multiple interlocking gameplay systems

Energy System

- Many in-game actions cost energy to perform
 - Harvesting crops
 - Collecting rent from houses
 - Collecting profits from businesses
 - Building new structures
 - Collecting from community buildings
- Energy is earned
 - Over time
 - With gifts from friends
 - Occasional payout in collections
 - Reward for visiting neighbors (friends)
 - Reward for playing multiple days in a row
- Neighbors can help by performing energyrequiring actions on your behalf
 - This is "free" for friends when they visit your city





Business System

- Businesses provide the primary source of coins
- Use energy to collect coin profits from businesses
 - Businesses produce a profit after a certain amount of time has elapsed (customers have visited)
 - With more people in a city, businesses produce faster
- Businesses must be supplied with goods to reset their ability to produce coin profits
- Goods come from farms, factories, ships, or trains





Land System

- Many items in game consume real estate
 - Homes, businesses, farms, community buildings all have a footprint
- Players begin with a fixed amount of land that is quickly used up
- To expand, players must buy an expansion
 - Requires:
 - Specific population level
 - Building permit (must obtain this as a gift from a friend)
 - Coins
 - Or, pay with cash (real money)



Population System

Build housing

- Once built, people move in
- More move in periodically over time
- Buildings available at higher levels are higher density, more people for same land footprint
- Max. population is determined by the number of community buildings
 - Each community building increases population ceiling by a different



Leveling Up

Experience

- In-game actions release blue stars (experience points)
- Level up at different XP counts
- Levels unlock building types
 - Better buildings at higher levels
- Reputation
 - Actions you do to help neighbors while visiting their cities generates reputation points
 - Level up at different reputation point counts
 - Isn't as well integrated into gameplay as XP, relatively few effects
 - A way of tracking social currency





City Cash System

- Players can spend real money to buy coins or energy
- City cash
 - Earn one city cash dollar for every level you increase (slow)
 - Can purchase with real money (fast, relatively cheap)
 - Or, take advantage of offers
- City cash uses
 - Exclusive items: some items can only be bought with city cash
 - Can hurry construction of community buildings
 - Can take a week or more to complete community buildings without
 - Allows your city to grow faster







Freemium model: can play for free, buy paying real money brings many advantages

So, why was this game so successful?

- So far, what has been described is a pretty straight-up city simulation game
 - Most of the game systems are pretty conventional, though they are certainly executed well
- If CityVille isn't that innovative of a game, why did it grow so quickly?
 - Some "easy" answers
 - Game launched with translations to multiple foreign languages
 - Zynga has huge base of existing players of other games, can cross-sell to them

User acquisition mechanics

User acquisition mechanics

- Required help
- Voluntary help
- Gifting
- Neighbor-only actions
- Broadcast to wall actions



- All provide motivation to invite friends into the game
 - Game is very challenging to play (or costs a fair amount of money) without having friends playing as well
- All provide ways to interact with friends via the game
 - A way of building out-of-game social currency via in-game help and gift systems

Required help

- Completing a community building requires things only available from friends (or City Cash)
 - People to staff positions within the community building
 - Items that can only be acquired as gifts from friends

STAFF YOUR CITY HALL

Community Buildings need to be staffed with employees. Hire your friends to work in your City Hall!



UC SANTA CRUZ

Voluntary Help

- Visiting cities of neighbors
- Performing actions in the city to help friends
- Requires neighbors
 - Which requires you to invite your friends into the game
- Business upgrades can be helped along by asking for help from friends
 - This isn't required, but speeds things up
- It feels good to help friends!





UC SANTA CRUZ

Gifting

- Can give a gift to any friend once a day
 - Is a nice way to say, I'm playing, and you're playing too
 - Gift giving UI gives you hints about friends you might invite to be neighbors
- Can request a gift from a friend once a day
- There is no cost for gifting
- It feels good to give and receive gifts!



Neighbor only actions

Franchise system

- Allows you to place a business in a neighbor's city
- Once a day, can collect a bonus from this business
- If you have a franchise of a friend in your city, it pays out very well
- Must keep inviting new neighbors to unlock the ability to have a franchise built in your city



Broadcast-to-wall actions

Train

- Send off train, returns after some period of time
- Can optionally broadcast a message to your wall asking people to have the train stop in their city
- If they do, the payout from the train increases substantially

Quests

 Some quests require items that can only be acquired by wall posts





Key Player Acquisition Metrics

- Virality (k-factor)
 - How "viral" is a given player?
 - A measure of how many people a given player invites into the game
- Player death
 - When a player stops playing the game
 - Not in-game death: this is impossible in CityVille
 - Important figure: average time to player death
- Conversion factor
 - Percentage of players who convert from free to paying players
 - Typically well under 10%, often under 5%
- DAU, MAU, DAU/MAU
 - Daily active users, monthly active users
 - The ratio indicates the daily active % of a user base









Source: www.appdata.com

Understanding how game design decisions affect **player retention**



Competition for leisure time attention

- People today have an enormous range of entertainment options
- New games, TV shows, movies, festivals, books, magazines, concerts, parties, family events and sporting events are released or occur every day
- How do you keep an audience focused on just one of these, over an extended period of time?

This is the challenge of **retention**

Typical console game retention curve



- Key challenge: improving this curve
- Jim's suspicion: this curve may be typical of all volitional software

Madden Football

- An American Football simulation game
- Updated yearly with new players and functionality
- Networked and single player play
- Individual games as well as playing an entire season



UC SANTA CRUZ



Plays in Madden

- Each play (down) the player on each side selects a play
 - One team chooses an offensive play, the other one a defensive play
 - Plays can be modified on the fly using audibles just before a play is executed
- Executing a play correctly involves some eye-hand skill
 - E.g., deciding when to make a pass
 - Plays have differing success percentages
- Madden 2011 features a large number of plays
 - A feature called Gameflow helps the player deal with this by automatically selecting a play based on the current game situation





UC SANTA CRUZ

Madden Data Analysis

- In Fall 2010 an analysis was performed of Madden 2011 gameplay data
 - By Ben Weber (UC Santa Cruz) and Michael John (Electronic Arts), along with Michael Mateas (UCSC) and Arnav Jhala (UCSC)
 - "Modeling Player Retention in Madden NFL II", To appear: Innovative Applications of Artificial Intelligence (IAAI), August 2011
- Collected gameplay data for individual games from release of game on August 10 through November 1, 2010
- Data includes a summary of every play in the game
 - Starting conditions
 - Formations and playcalls executed by each team
 - A subset of the actions executed during the play,
 - The outcome of the play

Modeling the Player

- Players are modeled as a feature vector
- Mode preference features
 - A player's preference for different gameplay modes
 - Madden 2011 has 8 of these, variations of single and networked multiplayer
- Control usage features
 - A player's competency at using the controls
 - Pre-snap and intra-play commands
- Performance features
 - Ability of the player to make successful plays
 - Turnovers (changes in possession), average yards gained, average yards allowed, ratio of possession, and ratios of down conversions
- Playcalling features
 - A plalyer's playcalling preferences
 - Includes record of manual vs Gameflow choices



Drew Brees playing Madden in Times Square, NYC www.sfheat.com/?gclid=COCGw_202p8CFRMXawodVXLIHQ

Can # of games played be predicted?



The player model works well at predicting the number of games a player will play.

Graph of predicted vs actual games played, developing using additive regression (correlation = 0.9, RMSE = 24.4, Mean error = 12.6)

Features vs Retention Regression Analysis



- Individually varied weighting [0,1] of various features in regression model and noted effect
- Each line above is result of modifying weight of one feature, holding others constant

- By analyzing gameplay data, it is possible to see correlations between design choices and player retention
- These observations can directly drive design choices
- Very clear return on investment for performing this kind of data mining



Relevance to other types of software

User Acquisition Mechanics Outside of Games

- It may be possible to adapt the user acquisition mechanics from CityVille for use in non-game software
- Especially for web-based software, would permit replication of CityVille's rapid adoption curve
- How many current software industry segments would be seriously disrupted if a new entrant grew to millions of users in just a few weeks?
- For existing web applications, suggests a range of mechanics for increasing user engagement

- As more and more software use is volitional, user engagement rises in importance
- Retention engineering is concerned with how to design software so users have high engagement, and continue to use it
 - A new subfield that draws from human computer interaction, software data mining, game design
 - A shift in emphasis away from correctness and meeting requirements towards overall deeply understanding users, and increasing user satisfaction with the software experience

MSR: Mining End-User Experience Data

- The MSR community has traditionally focused its analysis on software artifacts
- When we look at people, it's either:
 - Analysis of what software engineers do, or,
 - The bugs submitted by end users
- We have, at times, struggled to establish clear return on investment for the analyses we perform
- Mining end-user experience data for PGR4 and Madden 2011 yielded:
 - Clear return on investment
 - Insights into software design that were deeply interesting and exciting to stakeholders
- Recommendation: The MSR community should start performing data analysis of the behavior of end-users

