Assessment and Prediction of Engagement, Interaction and Revenue of a Mobile Social Game

Lucas Wu & Sophia He

Department of Statistics & Actuarial Science, Simon Fraser University

Introduction

The dataset we analyzed is on various demographic and event features for a mobile game. The goal of this analysis is to use distinctive or combination of these variables to predict three critical measures of this game: revenue, engagement and whether players return to play the game or not.

Demographic Features:

- Platform of Installation: installed on iPhone or iPad
- Date of Installation on Second Platform: whether they install the game on both platforms
- Facebook Connection: whether players link their Facebook account to the game
- Country: country where players are registered
- Gender: males, females or not shown (NA)

Event Features:

- Basic Events:
  - Tutorial Completion: players need to finish tutorial to proceed
  - First Game Played: indication of whether players started their first try of the game
  - First Win + Gift Exchange
  - First Bonus: achieved by accumulating energy to aid in winning the game
  - First Collection: achieved by obtaining various collections of artifacts
- 4 Types of Game: four variations of the game for players to choose
- 3 Types of Prize: three types of prize that can be earned when winning
- 7 Stages of Game: there are seven stages in total which players can pass each stage in different orders

Objectives and Methodology

Our Goals:

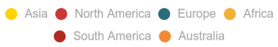
- Visualize the data set in a meaningful and interesting way
- Uncover hidden patterns from the data set by reconstructing variables and features
- Apply various statistical models to predict three metrics (engagement, revenue and return player) based on predictors and features identified

Interesting Findings

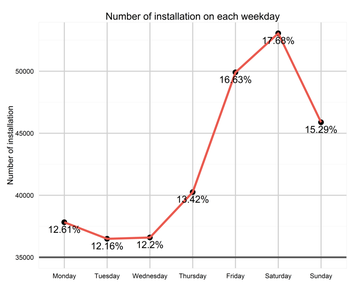
Demographic Features

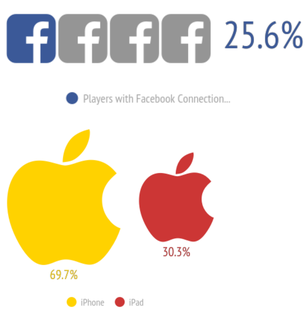
Among 300,000 users, the percentages of players on each continent are:

- **North America 59.6%**
- **Europe 20.0%**
- **Asia 9.4%**
- South America 4.7%
- Australia 3.7%
- Africa 0.2%



Among all players with gender recorded, **76.7%** of them are **females** and **23.3%** are **males**.



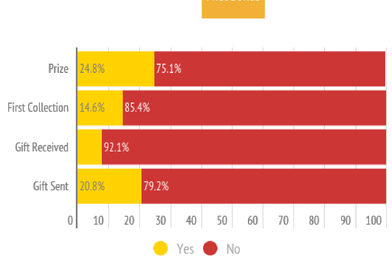


The game offers option of connecting Facebook account (**25.6%** of all users) with the game account to stimulate interactions between players.

Among 300,000 users, **69.7%** installed the game **on iPhone** compared to **30.3% on iPad**. **1.46%** of all users installed it **on both iPhone and iPad**.

Event Features

Among all players, **93.3% completed tutorial**. Among them, **92.8%** started their **first game**. Among them, **68.5%** achieved their **first win**. Among them **95.6%** obtained their **first bonus** within observation period.



Among all players, **24.8%** attained **at least 1 type of prize** during observation period; **14.6%** completed their **first collection**. Among players with Facebook connection, **20.8% sent out gifts** and **7.29% received gifts** from friends.

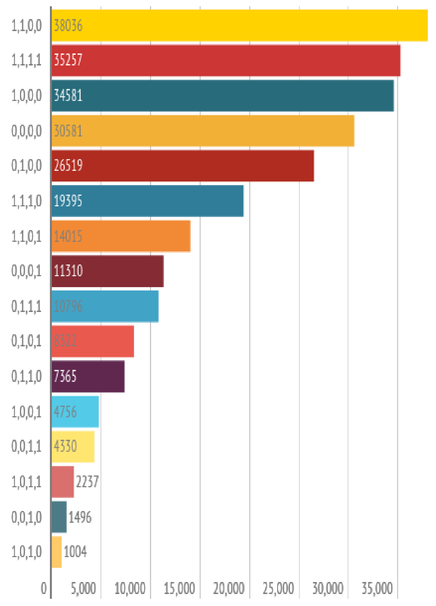
Game Type Features: Which Game Type is More Popular?

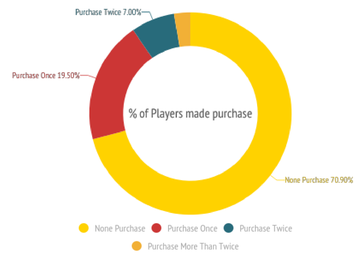
Game Type Combination Index:

4 different types of game that players can play. "Played" is coded by "1" and "not played" is coded by "0" . The ith position of the index corresponds to the ith type game played. For Example, "1,1,0,0" means players who played Type 1 and Type 2 games but not Type 3 and Type 4.

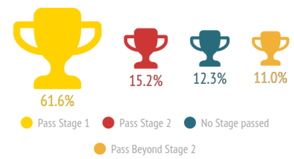
Observations:

- Largest number of players participated in game combination of **Type 1 and Type 2 Game (38036)** followed by participation of **all types of game (35257)** and participation in **solely Type 1 Game (34581)**
- **Type 3 Game alone** attracts the least amount of players **(1496)** compared to **Type 1 Game alone (34581), Type 2 Game alone (26519)** and **Type 4 Game alone (11310)**





Among all players, **70.9%** (212580) **didn't make any in-app purchase** followed by **19.5%** (58552) and **7.0%** (21064) of them made **1 and 2 purchase**. **Less than 3%** (7804) of them made **more than 2 purchases.**



Among all players, **more than half passed Stage 1**; **15.2% passed Stage 2** and **11%** of them **reached beyond Stage 2**.

Model and Prediction

Model:

Y = Demographics Features + Player Activities + Player Interactions

- **Demographics Features:** Platform (iPhone or iPad) + Date Of Installation On 2nd Platform + Facebook Connection
- **Player Activities:** Completion Of Tutorial + First Game Played Indicator + Game Type Combination Index + Total Number Of Game Type Played + First Win Indicator + First Bonus Indicator + In-app Purchase Indicator+ Total Number Of Purchase + Total Number Of Stage Passed
- **Player Interactions:** Gift Sent Indicator + Gift Received Indicator + Uken Gift Received + Collection Indicator +Prize Received Indicator + Total Number Of Prize

Prediction:

R^2 on Test Data Set Using Different Methods

	Linear Regression	Regression Tree	Random Forest	Boosting
Revenue	**13.0%**	**25.2%**	**17.0%**	**29.8%**
Engagement	**63.4%**	**65.6%**	**67.5%**	**71.7%**

Prediction Accuracy Rate on Test Data Set Using Different Methods

	Logistic Regression	Classification Tree	Random Forest	Boosting
Return Player	**92.57%**	**92.61%**	**92.64%**	**91.16%**

Top 3 Important Independent Variables

	No. 1 Predictor	No. 2 Predictor	No. 3 Predictor
Revenue	Total Number of Purchase	Total Number of Stage Passed	Game Type Combination Index
Engagement	First Collection	Total Number of Stage Passed	Total Number of Prize Won
Return Player	Total Number of Stage Passed	First Collection	Total Number of Prize Won

Conclusion

Our approach is to transform most variables into binary and combination index variables to predict revenue, engagement and return player. After determining important independent variables, we applied different statistical techniques which assume linear or non-linear relationship between the Y and X_i. By comparing R^2 and prediction accuracy rate, we chose the best techniques which produce the highest results. It turns out that Boosting is the best method to predict Revenue and Engagement, while Random Forest is the best method to predict whether a player returns or not.

Acknowledgements: We would like to acknowledge support from: Uken Games, the Department of Statistics and Actuarial Science and Charith Karunarathna.