How can we increase revenue from Catch the Pink Flamingo?

Tihomir Nikolov
Coursera and UCSandDiego

September 16, 2018

Dear management of Eglence Inc., after extensive analysis of the data from the game 'Catch the Pink Flamingo?', we found several interesing patterns, which could and should be utilized in order to improve the profitability of the game. We will present our analysis in the following slides, starting with that data that underpins it.

Problem Statement

How can we use the following data sets to understand options for increasing revenue from game players?

- ad-clicks.scv
- buy-clicks.csv
- game-clicks.csv
- users.csv
- teams.csv
- team-assignmnets.csv
- level-events.csv
- user-session.csv

Data format: coma separated values(csv)

- Several dimensions

1. User, team, session, level of the game
2. Frequency of ad clicking, information about purchases, how active are users during the game

- Perform analysis based on the different datasets and the relation between them
- Find insights *hidden* in the different aspects of the data

1. The management of the game is essentially a data science story, since we are observing the behaviour of the different players and would want to react adequately to it in order for our company Eglence to increase their revenues. We need to collect data regarding different aspects of the playing habits of the users.
2. We need to have different dimensions in the data, e.g. identifying each user, whether he/she belongs to a specific team and in which session they belong, on what level. That way we could identify important users, teams, connections with the level and session of the game. Also we could make observation about the spending behaviour, activity throughout the game, and propensity to clicking on ads. All this segmented based on teams, sessions or game levels.
3. We could increase the revenues for the owners of the game by finding hidden patterns in the data and exploit them in order to make the customers click more on ads and ultimately pay more/purchase more.

Data Exploration Overview

Two dimesions when exploring the data:

a Aggregation

- Total amount spent $21407, consiting of 6 items
- Most lucrative items are 5 and 4
- Most frequently bought items 2 and 5

b Filtering

- The total amount of money spent by top ten users ranges from $172 to $223
- Top three highest spending users all use Iphone platform
- The highest spending users have higher accuracy in playing the game

1. While exploring the data two dimensions are to be found: aggregation and filtering. We could see the total amount of money spent so far, as well as the number of items to be purchased. It could also be discerned that the most frequently bought items are not the most lucrative, which can be explained with the different prices. We could also see that the most lavish spenders purchased items for the amount of roughly $200.
2. From our basic exploratory analysis can be seen that the top 3 high rollers are all Iphone users, and also their accuracy in playing the game is somewhat higher than the average. From here we could expect our machine learning techniques to yield results that are roughly in unison with what we found with basic aggregation/filtering.

What have we learned from classification?

Purpose of doing the classification: find patterns in spending habits

a Procedure

- Select relevant features
- Create categorical variable, arbitrary boundary $5
- Accuracy of the model 88%

b Essential takeaways

- The main predictor of high spending is the platform type
- 83% of Iphone users are high spenders
- 37% of mac, and 14% of android users are also high spenders
- Target users based on their paltform

1. Our classification analysis uses an arbitrary constructed categorical variable in order to segment the users in different brunches, based on features considered relevant. It is also important to exclude features that are irrelevant such as identifiers. The model trained has a good accuracy of 88% and the main conclusion of it is that the Iphone users are really the high rollers among the the players of Eglence.
2. No other feature is considered important here, including the accuracy of the players. In addition we might pay attention on the small part of Mac and some android users.

What have we learned from clustering?

Goal: finding similar groups(clusters) of users based on numerical values

a Procedure

- Attribute slection: game level, activity, accuracy, revenue
- Selection of the number of clusters: 3
- Computing the cluster centers

b Major takeaways based on selected attributes and custers

1 Relatively good players, spending less than average

2 Users in this cluster are high spenders, very good players

3 Inexperienced payers, spending less tan average

Given our segmentation of users, Eglence could target those specific clusters of users

1. K-means clustering technique which is used here, uses numerical variables in order to segment groups of users which resemble each other. It is important to decide which features shuld be included when running the algorithm. The number of clusters is the other main consideration here. An arbitrary number of 3 clusters had been selected, even though the elbow method pointed to 4. Reason:hard to interpret.
2. The result is three clearly distinct clusters of users which can be conveniently charachterized with some basic economics. Some really good spenders, some users with potential, and some thrifty users.We could therefore craft some marketing strategies towards each.

From our chat graph analysis, what further exploration should we undertake?

The graph analysis scrutinizes the interaction between users, based on their chats.

- Finding the 10 most chattiest users
- Finding the 10 most chattiest teams
- Exploring the relations chattiest teams/users
- Studying the activity of groups of users through cluster coefficient
- Identifying long conversations and the users participating

Purpose of the graph analysis: targeting those influential users/teams

1. Our graph analysis explores entirely new dimension of our data, i.e. the interaction between the users through a message chat supplied by the game. Here we could gain extra insight into the behaviour of the players, find important users and/or teams. First obvious aspect is finding the most chattiest users and teams. That is important from management point of view, since those are some special users which have to be analysed. From our analysis only one most of the chattiest users belonged to chattiest teams, suggesting that individual users do not influence the teams as a whole.
2. When characterizing users we need to take into consideration the cluster coefficient which indicates whether an user interact with the entire graoup or only some part of it. Users with higher coefficient, i.e. communicating with everyone are more influential. Some extara analysys could be perform on some long chats and their participants.

Recommendation

- Data analysis could yield important insights

1. Pay attention on the platform users use, especially the Iphone users
2. Charge higher price for advertising to users belonging to cluster #2
3. Send precisely targeted ads to users belonging to cluster #1: most susceptible to prices
4. Promote the game among the users from cluster 3:they may become more like users from #2 and spend more
5. Send special advertising to chattiest users, take into consideration their cluster coefficients
6. Craft ads towards chattiest teams, they hold the most of the conversation and hence could be good communicators

1. All of our analysis shows that there are some important takeaways to be found, hidden in our data. First is that the platform type is really important in determining the spending habits of the players. Iphone users are clearly the high rollers and should be offered some appropriate ads and monitored closely. Apart from that we could segment our users into the 3 distinct clusters. For the cluster #2, higher charges to the advertisers are justified, since users can clearly afford to buy more, hence are more valuable. Cluster #3 represents the users with potential, they should be encourage to play more. Cluster #1 are the spendthrift users, which should be targeted with appropriate ads, based on the marketing strategy.
2. From the graph analysis the main recommendation is to find the most chattiest users and teams and to offer them crafted advertisment or encouragement to play the game. Regognizing that the teams are somewhat more important than the individual users, and use this when decisions on marketing strategy are taken. Prioritize the users based on their chattiness and also their cluster coefficient. That way really influential players could be found.
3. Once more here we should ephasize that the game is essentially a data science story. Since the mentioned recommendation are consitent with the data, and do not contradict each other, it follows that if the management of Eglence heeds our advice they would most certainly inclease their revenues.