How can we increase revenue from Catch the Pink Flamingo?

SHUN-WEN CHANG

Hello, my name is Shunwen Chang. I’ll be talking about my data analysis from Catch the Pink Flamingo game and how we could possibly increase company revenue from this game based on my analysis.

Problem Statement

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

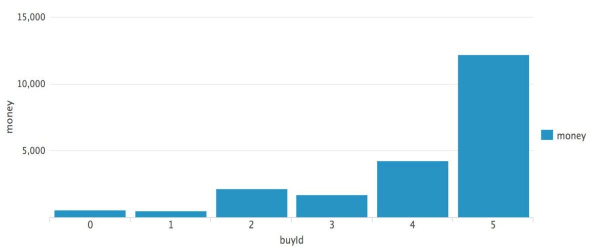
Data file	Description
ad-clicks.csv	Record when a player clicks an ad
buy-clicks.csv	Record when a player makes an in-app purchase
users.csv	List of users playing the game
team.csv	List of teams terminated in the game
team-assignments.csv	Record when an user joins a team, at most one at a time
level-events.csv	When a team starts or finishes a level
user-session.csv	Record when an user starts and stops in the game. Refreshes when a team goes to next level.
game-clicks.csv	Record when an user clicks on a flamingo in the game

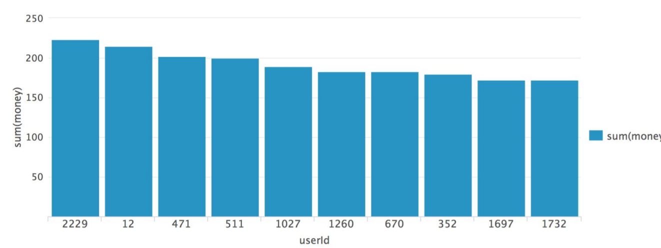
There are main 6 data sets showing in the table. We have ad-clicks, buy-clicks, user list, team list, team assignment, level event, user session, and game clicks. It is very important for Eglence to be able to identify new revenue opportunities with various kinds and sources of data rather than just one. Since in order to get a big picture of a data science story, we can't just look at one side of it. For example, we can't conclude what will make more money just based on ad-clicks data set. Even if you know specifically who clicked the ad, it won't tell you any information on how to target for the future selling. We need all the information to do the classification, clustering analysis and make our reasonable guess and further increase our revenue with our devised model.

Data Exploration Overview

Aggregation - left figure showing a histogram of how much money was made from each item.

Filtering - right figure showing a histogram of total amount of money spent by the top ten users (Ranked by how much they spent).





Now let's explore our data. I've done both aggregation and filtering on all 6 different data sets. Here are some key findings in these data. On the left figure is a histogram of how much money was made from each item. buyId represents the ID for each merchandise. You can see item 5 generates the most income. However, this is not entirely because item 5 has the most sell. It is partly due to the price of item 5 is the highest of all.

On the right side figure is a histogram of total amount of money spent by the top ten spending users. This chart tells us what is the most money one user has spent on buying items in this game, and based on the user ID we can also analyze what characterics these high spending users have in common. For example, the platforms they use and the hit accuracy they have in the game.

What have we learned from classification?

- A new categorical attribute was created based on categorization of avg-_price into 2 bins to enable analysis of players as broken HighRollers and PennyPinchers.
- Unrelated attributes (ave_price, user_id, session_id)were removed.
- Data was partitioned into train and test data sets. We used train data to train our decision tree model.
- Our result shows that iphone users are HighRollers and other platformType users are PennyPinchers.

I performed classification analysis using a decision tree model. First I created another categorical attribute that can break avg-price into 2 categories: HighRollers and PennyPinchers, where buyId>5 belongs to the HighRollers and buyId<=5 belongs to the PennyPinchers. The creation of this new categorical attribute was necessary because this is a classification problem, we can not use a continuous value field like avg--price, which range from 0-6. Then I discard attributes related to unique IDs. The data was then randomly partitioned into 60% and 40% for training part and test part in our decision tree model. This is important because when we do data analysis, we should test our model on a data set that was not used to train the model. The resulting accuracy was 88%. It shows that iPhone users are HighRollers and users with other Platform types are PennyPinchers.

What have we learned from clustering?

- We selected "team strength", "revenue", and "team current level" as 3 attributes for clustering analysis
- 3 clusters were created
- The resulting clusters location indicates that teams with less strength tend to spend more money on in-app purchases.

For the clustering analysis, I selected 3 attributes: team strength, revenue, and team current level. The dimension of my training data set is 44 by 4. And after training, there were 3 clusters generated centered at (0.37, 1, 702), (0.55, 1, 177), (0.54, 1, 418). The result shows that teams with lower strength tend to spend more on in-app purchases. From this finding, it is then suggested to have more ads or more expensive selling items at lower level games.

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

- We used Neo4j Graphics to find the longest conversation chain and its participants. This is important for the company to collect information on what topics attracts more conversations and set new business plans based on those.
- We analyzed the relationship between top 10 chattiest users and top 10 chattiest teams. There is only one such user belongs to such team. No direct relation between chattiest user and chattiest team can be seen here.
- We found the 3 most active users based on cluster coefficient. This is useful information since we can target these users with more promotions/incentive.

From my Neo4j chat graph analysis, I found the longest conversation chain to have a length of 10. Furthermore, there are 5 unique users in this chain of conversation. This kind of search is useful for Eglence because it can tell them what kind of subjects or topics users are enthusiastic about and they can therefore set business plan targeting on these subjects. I also analyzed the relationship between the top 10 chattiest users and the top 10 chattiest teams to see if any of the user is in one of the teams. There is only one chattiest user who also belongs to the chattiest team. It seems that there is no strong relationship between chattiest users and chattiest teams. This analysis is also important as to knowing if Eglence should target more on teams or individuals. Finally, I collected information of the 3 most active users based on cluster coefficient. This kind of analysis is very useful for Eglence to target specific users with more incentives or gather informations about what kind of users love this game.

Recommendation

Increase in-app merchandize for lower level games.

Increase price for ads in iphone platforms.

In conclusion, my recommendation for Eglence on this game are: 1. Increase in-app ads or increase the price for ads in lower-level games since according to my analysis, users who achieve higher levels tend to spend less on in-app purchases. 2. Increase in-app ads or increase the price for ads in iPhone platform games, because from my analysis, iphone users spend much more than all other users with other platform types.

THANK YOU !

Thank you for your attention. Please don’t hesitate to ask if you have any questions.