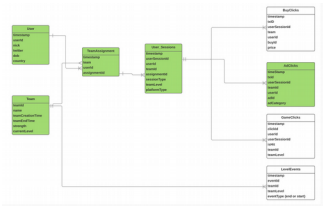
How can we increase revenue From Catch the Pink Flamingo?

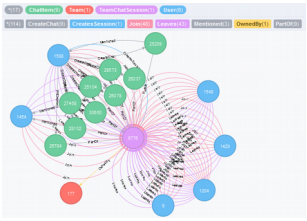
Markus Jung

Problem Statement

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

- Log files recording the activities of people playing Catch the Pink Flamingo
- Chatting activity of active users





We collect different data from our users of “Catch the pink flamingo”.

The first data we collect is the activity of people playing the game. We do this by generating and collecting Logfiles form our Webpage.

The second source we use is the chatting activity of our users in twitter.

With the come up of Big Data we can now use different data science techniques to interpret these data, par example Data Exploration, Classification, Clustering and Graph Analysis. The different kinds and sources of data are very important for us to be able to identify new revenue opportunities. Because the more we know about our users the more we can use this information for our business.

Data Exploration Overview

With Data Exploration we can get insights that can help us understand our players better.

- What is the Distribution of Operating Systems Used by Users?
- What are the two most Commonly-Clicked Ads?
- Hit-ratio percentage for the top three buying users?

With Data Exploration of the Log files we can get insights that can help make us more profitable.

We loaded the log files into Splunk, a leading platform for Operational Intelligence.

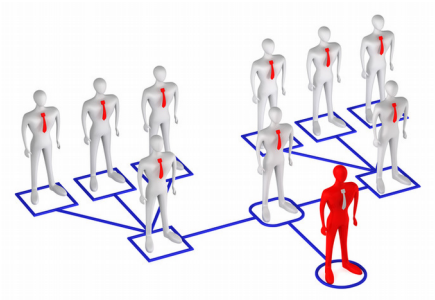
We asked different questions and got answers from the data exploration.

We saw e.g. that most of our users uses iPhone as platform type. And that the most commonly clicked ads are computers and games.

Also we can answer questions about the hit-ratio of our top buying users. It is over 10 % for the top three users.

What have we learned from classification?

What makes a user a HighRoller?



- About 40 % who purchase are HighRoller
- About 83 % with iPhone are HighRoller
- Only 3 % of the users with Linux are Highroller
- Only the platform type is relevant for prediction

With KNIME, a open solution for data-driven innovation helping to discover the potential hidden in data, we first classified users as buyers of big-ticket items (“HighRollers”) vs. buyers of inexpensive items (“PennyPinchers”).

For Predicting which user is likely to purchase big-ticket items while playing Catch the Pink Flamingo we build a decision tree. With a accuracy of over 88 % we can predict only with the platform type if a user is a HighRoller or not.

Other results are:

About 40 % of the users who pruchase are HighRoller.

About 83 % of the users with iPhone as type of platform are HighRoller.

Only 3 % of the users with linux as type of platform are Highroller.

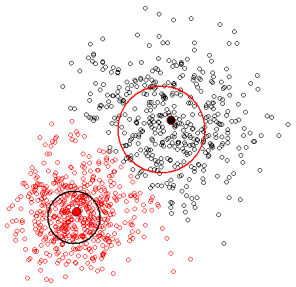
What have we learned from clustering?

We can cluster our users into groups based on characteristics to find any ‘significant’ differences.

With these techniques we can see correlations like

- Start Time of Session and purchasing behavior
- Click on ads and spending money

There is not only one way for clustering and selection of criteria to provide actionable information.



We used a K-Means clustering model in Spark Mllib, to cluster our users into groups based on characteristics such as their game playing behavior, purchase behavior, inclination to click on displayed ads, etc. MLlib is Apache Spark's scalable machine learning library.

The challenge is to find the kind of criteria that might provide actionable information.

A example we analyzed the start Time of a Session and the purchasing behavior. We learned that The center of the start time of a userSession is between 12 and 13 o'clock. So advertising can be priced with a model that determines the Start time when users spend most money.

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

With the analysis of chat graphs we can answer questions like

- Longest conversation chain
- How Active are Groups of Users

The next step is to combine these information of our users and groups with the recording of the activities of people playing Catch the Pink Flamingo.

For Graph Analytics of the Twitter Chat Data of our users we used Neo4j. Neo4j is a one of the popular Graph Databases.

With the analysis of chat graphs we can answer questions like which is the longest conversation chain or if the top 10 chattiest users belong to the top 10 chattiest teams? Although we can analyze how Active are Groups of Users.

The next step is to connect the twitter information of our users and groups with the recording of the activities of playing Catch the Pink Flamingo.

With this information connected we could e.g. give special offers to active user groups.

Recommendation

With these analysis we can better understand options for increasing revenue from game players. One concrete action could be:

- different offers for users with different platform type

Recommendations for next steps:

Close collaboration between departments and data scientists to

- Ask the right questions
- Find the the right conclusions

With this data science approach we learn a lot about our users and can better understand the options for increasing revenue for the company.

One action we can recommend are different offers for users with different platform type. This could be one way to stabilize the revenue from our iPhone Users and increase the revenue of users others platform types.

The next step to extract even more from our “big data” is a close collaboration from experts of the departments and our data scientists. Together we can ask the right questions to get the right answers for our business.