

How can we increase revenue from Catch the Pink Flamingo?

METAN KONE

Problem Statement

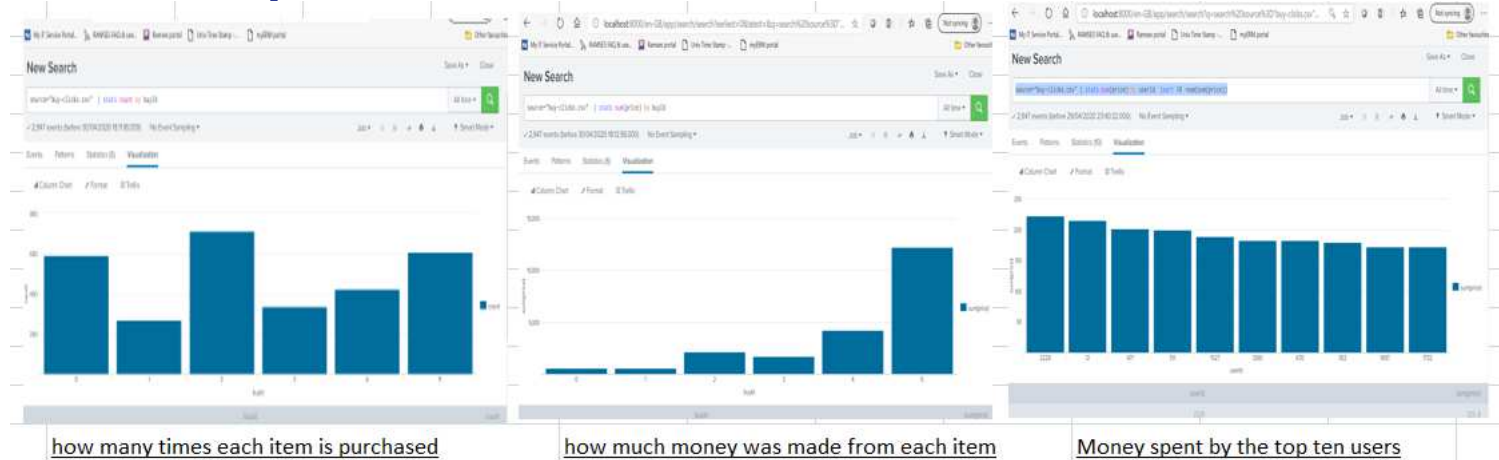
How can we use the following data sets to understand options for increasing

buy-clicks.csv: A line is added to this file when a player makes an in-app purchase in the Flamingo app.
game-clicks.csv: A line is added to this file each time a user performs a click in the game.
user-session.csv: Each line in this file describes a user session, which denotes when a user starts and stops playing the game.
ad-clicks.csv: A line is added to this file when a player clicks on an advertisement in the Flamingo app.
users.csv: This file contains a line for each user playing the game.
team.csv: This file contains a line for each team terminated in the game.
team-assignments.csv: A line is added to this file each time a user joins a team. A user can be in at most a single team at a time.
level-events.csv: A line is added to this file each time a team starts or finishes a level in the game

The datasets presented to us come from logs resulting from the **Catch the pink flamingos** game application. These logs are records of the players actions, and capture the players habits and behaviour, even on which platform (android, iphone) they use to play. These data also capture the events related to all the purchases done on the application (which players category do lot of purchases, ...).

So by exploring and analyzing these big data with machine learning technics, we can identify patterns based on connections between nodes, and then do recommendations in line with business strategies in order to increase the revenue of company.

Data Exploration Overview



Using **Splunk**, We computed aggregated key indicators to identify for example the amount spent buying items or how many times each item is purchased. We used histograms visualization for a better overview of these data.

Studying user id, platform, and hit-ratio percentage for the top three buying users, we identified those top 3 buying users all use Iphone.

What have we learned from classification?

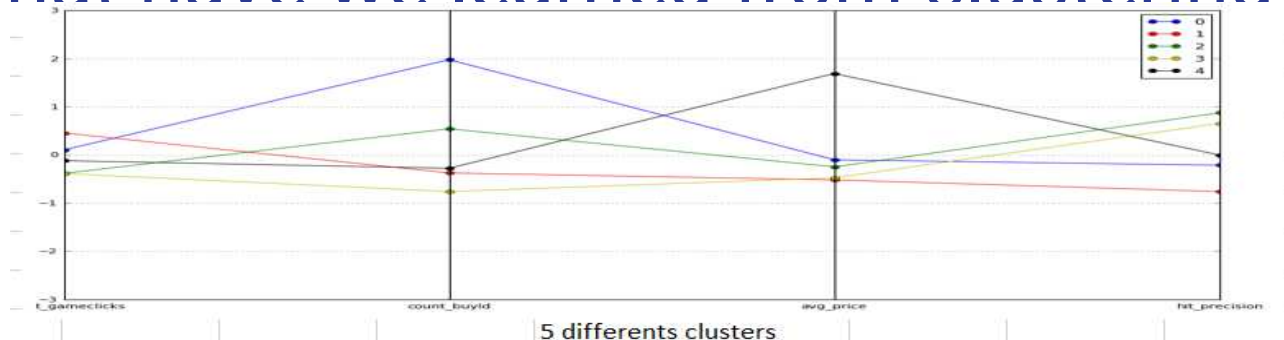


In the classification analysis, a new categorical attribute was created to enable analysis of players as broken into 2 categories (HighRollers and PennyPinchers).

HighRollers do more purchases while PennyPinchers spend very little money on the application.

Using Knime, we predicted which users falls into HighRollers category, or in PennyPinchers category. **We identified that players on iphone are high HighRollers while users on Linux are high PennyPinchers.**

What have we learned from clustering?



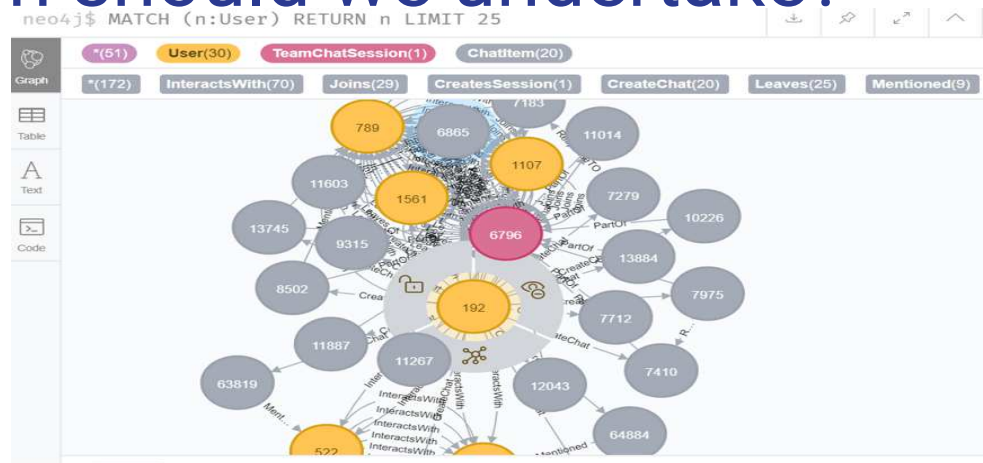
Using Apache Spark, we have done classification based on the 4 following features identified : **count_gameclicks**, **count_buyId**, **avg_price**, **hit_precision**.

Using elbow method, we saw optimum number of clusters was 5 :

Cluster 0 has the highest number count_buyId, but we can notice the revenue is coming most from cluster 4. Cluster 1 members are most playing . We can therefore notice they have the less precision. Also , Cluster 1 members spend a lot of time playing but they are NOT giving revenue.

Cluster 3 members who plays in a moderate way , have more hit precision.

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



Graph analytics shows there are many interactions between the players and their teams. Neo4J allowed to identify communities formed from density of conversations inside the graph. In these communities, some users are more chattier and engage longer conversations.

Chattier users and initiators of longer conversations can be more valuable, because of their potential to spread information to other communities members. Eglence, Inc. can target those members to increase its revenue by showing them more expensive items to such users, and showing them more adverts.

Recommendation

From classification analysis, we identify that more advert should be presented to iphone users as they are purchasing more on the application.

From clustering analysis , it was seen that users who play a lot are also the users who spend less and click less on ads.

If we increase ads to users who play a lot, it will promote these users to spend more and therefore increase the revenue.

From graph analytics findings, chattier users and initiators of longer conversations can be more valuable, because of their potential to spread information to other communities members. Eglence, Inc. can target those members to increase its revenue by showing them more expensive items to such users, and showing them more adverts.

