How can we increase revenue from Catch the Pink Flamingo?

Trung Ung

Jun 11, 2018

I’m TrungUng, I’m a developer, come from VietNam. I will be talking about my data analysis from Catch Pink Flamingo game and give the answer for question what have we learned from data exploration, classification, clustering and graph analysis?

Problem Statement

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

- *Ad-clicks*
- *Buy-clicks*
- *Game-clicks*
- *Users*
- *User-session*
- *Team*
- *Team-assigments*
- *Level-events*
- *Chat data*

- “***flamingo-data***”, it contains 8 CSV files containing simulated game data and log data for the Catch the Pink Flamingo Simulated Game Data. This data will be used for data exploration in **Splunk**.

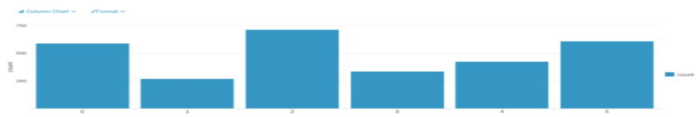
- “***combined-data***” contains a single CSV file. It is intended to be used for identifying big spenders with **KNIME**.

- “***chat-data***” contains 6 CSV files representing simulated chat data related to the Catch the Pink Flamingo game to be used in Graph Analytics with **neo4j**.

Data Exploration Overview

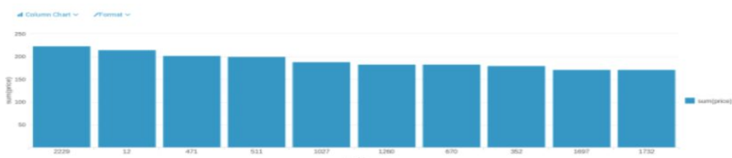
Aggregation

A histogram showing how many times each item is purchased



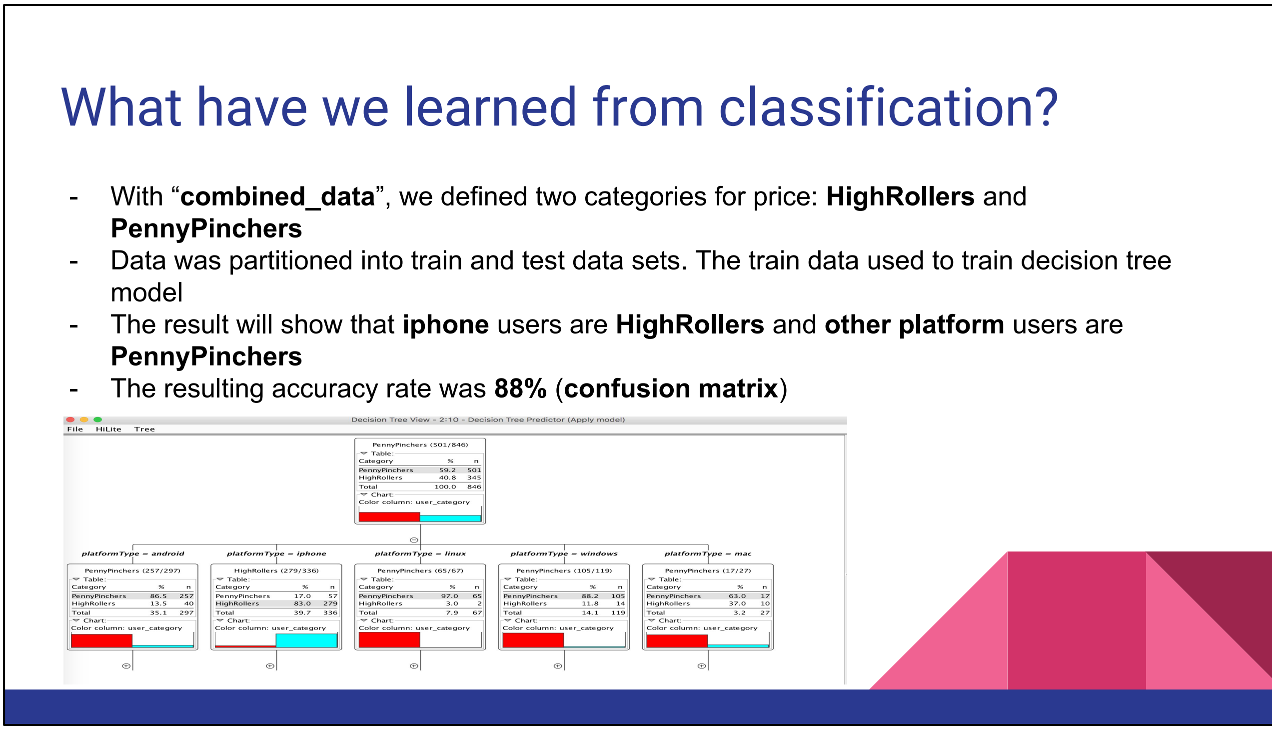
Filtering

A histogram showing total amount of money spent by the top ten users (ranked by how much money they spent).



The first Data exploration, Importing datasets of *”**flamingo-data**”* into **Splunk**, using aggregation functions to make histograms as follows.

- The first histogram is made by the data in ***buy-clicks***, which shows how many times each item is purchased. According to this graph, we can see that, **Item 2** is the most purchased and **Item 1** is the least purchased
- The second one is made by the same dataset. According to this graph, we get different amount of money that made from each item.



- Classification analysis is using a decision tree model in **KNIME** to identify big spenders in the game. The trained model was then applied to the test dataset
- In this graph, red stands for the PennyPinchers, blue stands for the HighRollers.
- According to the resulting decision tree, it obviously shows that the predicted **user_category** is different in various platforms, the users on the platform android, linux, windows and mac are almost **PennyPincher**, however, most users which on the platform iphone are **HighRoller**.
- Follow the confusion matrix, our model has a **88.5%** accuracy rate.

What have we learned from clustering?

Cluster	Total ad clicking	Total game clicking	Revenue
1	25.12	362.50	35.36
2	32.05	2393.95	41.20
3	36.47	953.82	46.16

Cluster 1 *is different from the others in that the users' ad-clicks, game-clicks and cost are all less than others* > Called "low level spending user".
Cluster 2 *is different from the others in that the ad-clicks is not the least, game-clicks is the most but their cost is not the most* > Called "neutral user".
Cluster 3 *is different from the others in that the users' ad-clicks, game-clicks and cost are all more than others* > Called "high level spending user".

- The results reflect that the ad clicking amount of "high level spending user" is **1.45 times** more than "low level spending user" and **1.14 times** more than "neutral user"; the game clicking amount of "high level spending user" is **2.63 times** more than "low level spending user" ; finally, the revenue from "high level spending user" is **1.31 times** higher than "low level spending user" and **1.12 times** higher than "neutral user"

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

- The CSV data files of *“chat-data”* are loaded into Neo4J. Find the top 10 Chattiest users and return the user’s Id and count the number of users.
- Analyzed the relationship between top 10 chattiest user and top 10 chattiest teams. Return the count of teams.
- Finally, report the top 3 most active users **based on cluster coefficients**

  > We can target these users with more **promotions/incentive**

- From Neo4j chat graph analysis, we can found the longest conversation chain. Have 5 unique users in this chain.
- Analysis the relationship between top 10 chattiest users and top 10 chattiest teams to see the user is in one of teams.
- Finally, I collected information of the 3 most active users based on cluster coefficient. This is useful for us to target specific users with more incentives or gather information about what kind of users love this game.

Recommendation

- Offer more products to iPhone users.
- Offer some promotions to PennyPinchers for attracting their consommation.
- Provide more products to “high level spending user”
  - Since they clicked less but buy more than others, we can provide more products to them for increasing the revenue
- Provide some fixed pay packages or promotion to users, especially to “low level spending user”
  - This action can stimulate consumption of users, and since the paying probability of “low level spending user” is low, the promotion can encourage them to purchase

Just give some recommendation for Eglence on this analysis game are:

- Offer more products to iPhone users: offering more products to them can increase our revenue.
- Offer some promotions to PennyPinchers for attracting their consommation.
- Provide more products to “high level spending user”: we can provide more products to them for increasing the revenue.
- Provide some fixed pay packages or promotion to users, especially to “low level spending user”: the promotion can encourage them to purchase.