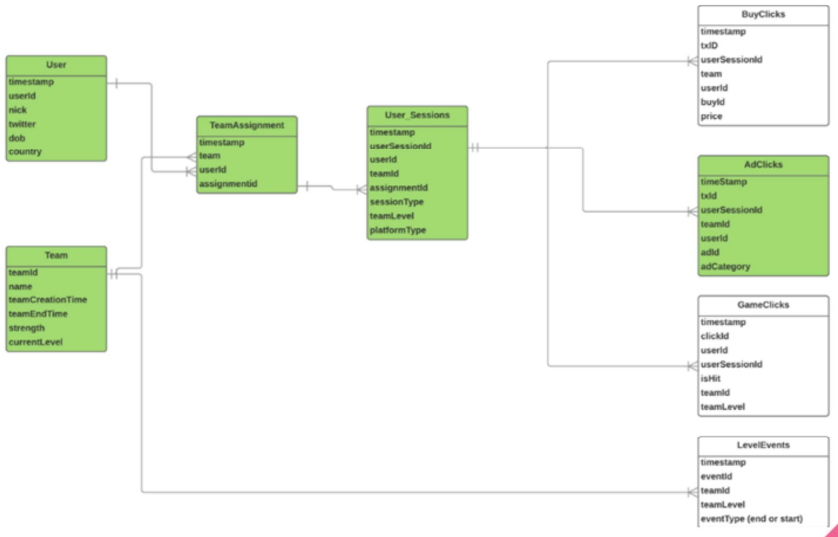
How can we increase revenue from Catch the Pink Flamingo?

By Abha Saxena

Problem Statement

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



This slide shows the relational schema of the datasets which would be considered to increase revenue.
The User dataset contains the details of each user i.e., userid, nickname, date of birth, country,twitter handle and timestamp (which is an indicator of his/her activity). Similarly, the Team dataset consists of the details of each team. The Team Assignment table includes the timestamp when a user is assigned to a team. The User_Sessions dataset contains all the information about sessions of users in the game, including the timestamp, userSessionId, sessionType, teamLevel, platformType, among other ids. This dataset helps in finding out the most active players and grouping them according to the platform they are playing on. The next two datasets are BuyClicks and AdClicks. These are the indicators of the game purchases and learning which kind of advertisements appeal more to different groups of players. The GameClicks dataset contains click information of each user with timestamp and sessionId. Lastly, the LevelEvents dataset contains information about teams' level events in game with timestamps.

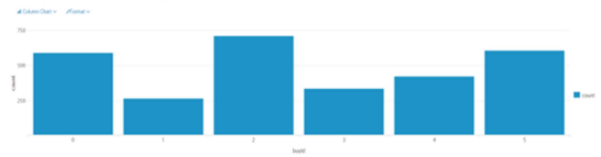
Data Exploration Overview

Using Splunk to explore data, some tables and histograms were generated using aggregation and filtering queries to observe datasets closely. For example –

Aggregation

Amount spent buying items	21407.0
Number of unique items available to be purchased	6

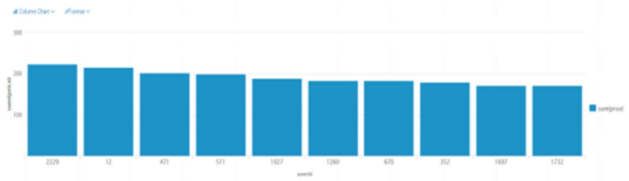
A histogram showing how many times each item is purchased:



A histogram showing how much money was made from each item:

Filtering

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



The following table shows the user id, platform, and hit-ratio percentage for the top three buying users:

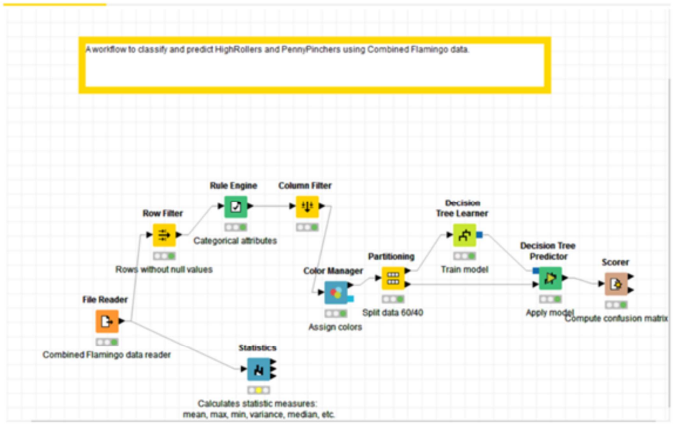
Rank	User Id	Platform	Hit-Ratio (%)
1	2229	iphone	11.597
2	12	iphone	13.0682
3	471	iphone	14.5038

Data Exploration is important so as to understand the data we are working on better, observe trends and make useful conclusions. The aggregation queries group data based upon some common attributes and produce results in form of aggregate functions like sum, average, max, min and so on. Likewise, filtering queries enable us to observe a subset of the entire dataset. Using Splunk, aggregation is used to find out how much money is made from each item in in app purchases and filtering is used to generate statistics of top 10 buying players. It is observed that top 3 buyers use iPhone and this knowledge is useful to make better offers for iPhone users to generate more income.

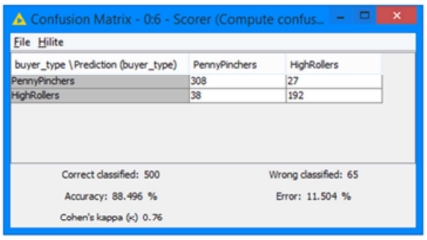
What have we learned from classification?

Using KNIME, users are classified as buyers of big-ticket items ("HighRollers") vs. buyers of inexpensive items ("PennyPinchers"). Big-ticket items are those with a price of more than $5.00, and inexpensive items are those that cost $5.00 or less.

The final KNIME workflow is shown below:



A screenshot of the confusion matrix can be seen below:



As seen in the screenshot above, the overall accuracy of the model is **88.496%**

- PennyPinchers/PennyPinchers: 308 actual Penny Pinchers were correctly predicted as Penny Pinchers.
- PennyPinchers/HighRollers: 27 PennyPinchers were incorrectly predicted as HighRollers.
- HighRollers /PennyPinchers: 38 HighRollers were incorrectly predicted as PennyPinchers.
- HighRollers / HighRollers: 192 actual HighRollers were correctly predicted as HighRollers

What makes a HighRoller vs. a PennyPincher?

The model predicts a whether a user is a HighRoller or a PennyPincher based on the in app purchases. The platform seems to strongly determine if a user is a High Roller. The majority of iPhone users are High Rollers whereas in other platforms, the majority of users are Penny Pinchers. Following this observation, the following steps can be taken to increase revenue.

First, personalized suggestions can be provided to users for items to purchase (Especially to iPhone users as they will be willing to spend if suggestions are even more related to their liking).

Second, cheap combo offers can be provided for users of other platforms i.e., except iPhone users to attract PennyPinchers.

What have we learned from clustering?

Using Spark Mllib, the following attributes were selected for clustering. It is observed that users buying products of lesser range also buy lesser number of items and those buying more number of items are those buying a wide range of items.

Attribute Selection

Attribute	Rationale for Selection
Number of Purchases	The number of purchases by each user shows the purchasing record of the user and thus will help in predicting future purchasing power.
game clicks per hour by each user	Decides how active the player is.
Range of purchases	For learning the range of items a user purchases, whether he is confined to similar ones or tries different ones.

Recommended Actions

Action Recommended	Rationale for the action
Customized ads for all users	The users having a lesser range of purchases should be recommended more similar items rather than of variety and those with huge range of purchases should be recommended a nice variety of items. This will increase purchases of both type of users.
Increase variety of products	As the range of purchasing items increases, the number of items purchased increases too. Therefore, increasing variety will increase range and in turn, increase purchases.
Consider team purchases	Consider the purchases made by a team as a whole and suggest combo packs of items according to their demands.

Clustering the datasets provided valuable information that users buying products of lesser range also buy lesser number of items and those buying more number of items are those buying a wide range of items. This knowledge can be exploited to increase sales by giving customized advertisements to users. For users buying lesser variety of items, similar products could be advertised and to those buying wide range of products, different variety of items could be advertised. This strategy would appeal both the groups. Further, variety of products should be increased as the users appealed by more variety are the ones who buy the maximum number of items. This will increase the sales too.

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

Exploration of chat data yielded the following key information:

- the longest conversation chain and its participants.
- the relationship between top 10 chattiest users and top 10 chattiest teams
- Most Active Users (based on Cluster Coefficients)

Further, we can explore groups of users frequently chatting with one another.

The interactions involved with chats occurring in the game are modelled using this Graph Data Model. The nodes are ChatItem (a chat text), Team (to which users belong), TeamChatSession (the session involving a user on the team) and User. Edges signify timestamps of a chat event. The actual chat text is not used in this model. Edges involving users include: CreatesSession, Joins, Leaves or CreateChat. TeamChatSession is OwnedBy a team. A ChatItem is part of a TeamChatSession and ResponseTo is an edge between two ChatItems. An additional edge created for analysis denotes InteractsWith between Users and does not have a timestamp. The observations made about the chattiest users and chattiest teams, the longest conversation chains and most active users, can be used to mark users' favourites and notify them when they are online. In-app purchases related to chats like stickers or themes may be included and advertised more to the chattiest users.

Recommendation

Eglence Inc. should pay attention to the products available for in-app purchases and their advertising.

Using the trends observed and predictive analysis of different datasets, more range of products should be available and customized advertisements should be made to different groups of users based upon their purchasing power, interests and activity.

The main objective of any business is to maximize the owner's profits. Hence using the analysis of these datasets, Eglence Inc. should improve its market based upon the observations made. More expensive items can be produced and advertised especially to highRollers, which are mainly iPhone users. Further, combo items should be launched to attract teams as a whole. Users buying a narrow range of products should be suggested similar ones and those buying a wide range should be suggested a variety of items. More active users should be appealed more. Thus, Eglence should use all possible trends observed to improve its market and in turn, generate more revenue.