How can we increase revenue

from

Catch the Pink Flamingo?

By Nchang Valery

Problem Statement

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

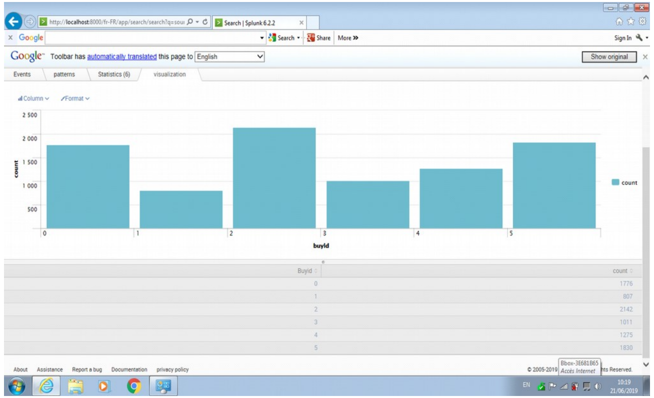
The catch the pink Flamingo big data capstone project is made up of three main datasets of which we can derive insights that could lead to an increase revenue from the Game players.They’re as follows ;

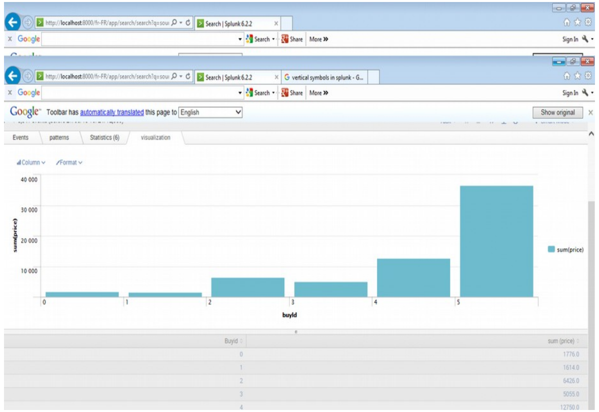
a)Flamingo dataset: These dataset is compose of 8 csv files with stimulated game and log data for the stimulated game data.these data shall be exploited for use in splunk.

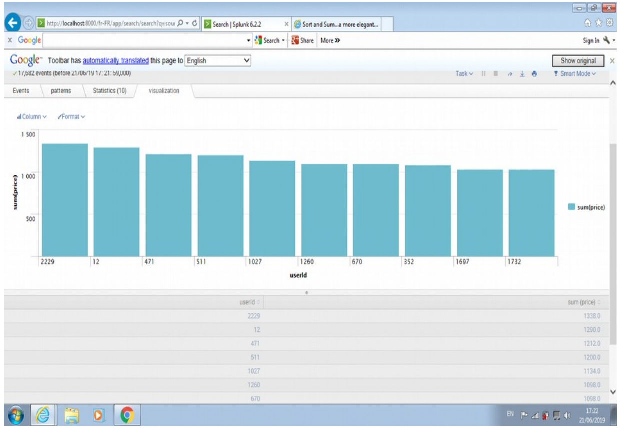
b) Combine Data : These data contains a single csv file created by aggregating data from several game data files. It is being used in the KNIME platform.

c) Chat data : The "chat-data" contains 6 CSV files representing simulated chat data related to the Catch the Pink Flamingo game.These data is for use in Graph analysis in the Neo4j database.

Data Exploration Overview





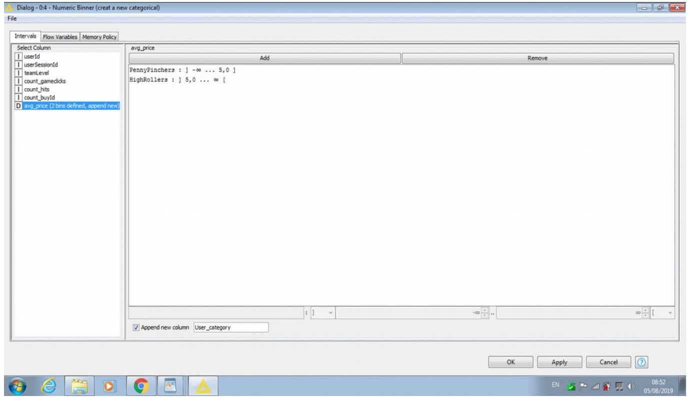


- The histogram above was being made using the 'Buy-Clicks.csv' file.From the histogram we can observe that item 2 was the most purchased, while item 1 was the least purchased.

- The histogram above shows the amount of money that was made on each item. item number "5" made the highest amount of money, whereas item number"1" made the least amount of money, which is also the least purchased .These, simply means that people are not much interested in item number"1" much as compare to other items. So,providers to find a way such that people could pay more attention on item number "1".

-The above histogram was made from the "buy-clicks.csv" file showing the total amount of money spent by the top 10 Users.From the observation on the histogram,the user with userId "2229" spends the most amount of money.

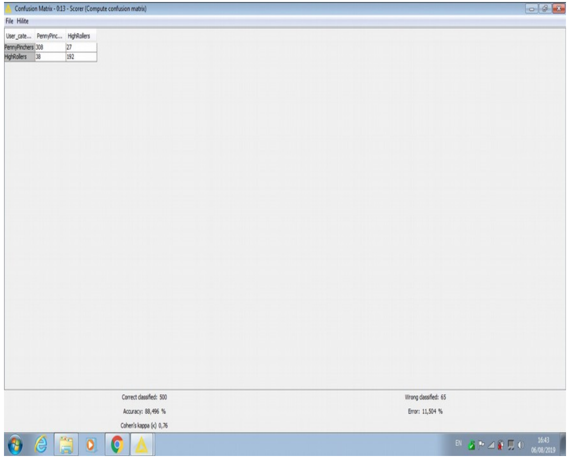
What have we learned from classification?



-Due to the analysis made from the decision tree,it can be observed that the predicted Users_categories are not the same in all the platforms.The HighRollers Users mostly uses the Iphone Platform whereas the PennyPinchers Users uses the Platform Android,Linux,Windows,and Mac platforms.

- From the instructions we needed to creat two categorical attributes from the Numeric Binner called "User_category".These two categories has to be defined for the price that will be use then we can distinguish between HighRollers(buyers of items that cost more than $5.00) and PennyPinchers(i.e buyers of items that cost $5.00 or less).

- These categorical attribues were created so as to facilitate our classification of users and these have contributed to the following steps



- From the screenshot , we observed on the confusion matrix that the overall accuracy of the model is 88.496%.

PennyPicnhers/PennyPinchers: 308 actual PennyPinchers were correctly predicted as PennyPinchers.

-PennyPinchers/HighRollers: 27 PennyPinchers were incorrectly predicted as HighRollers.

-HighRollers/PennyPinchers: 38 HighRollers were incorrectly predicted as PennyPinchers.

-HighRollers/HighRollers: 192 HighRollers were correctly predicted as HighRollers.

What have we learned from clustering?

-In part 3 of these project i used spark ML to undergo clustering analysis.
- Firstly three partitioned attributes such as Total add-clicks,Total Game-clicks,and Total revenues are selected and a brief explanation of these attributes reflected the behaviour of each Users in the Game.
- The data was then created and trained using the three attributes.

Attribute Selection

Attribute	Relationale for selection
Total Game Clicks	This is simply the total number of times each User clikcs on the Game which showing how reactive the Users are.
The amount of add-clicking per Users	These simply shows us the behaviour of each user ,that is the amount of time a user might click on the game throughout the game when an add is being made.
revenue	These will give us the total amount of money spent by each User on an item,which also reflectsthe behaviour of each user in the game.

- These dataset was later also trained inother to obtain clusters centers below.

Cluster Centers

Cluster #	Cluster Center
1	(25.11037047, 352.50308542, 35.35802469)
2	(32.05, 2393.95, 41.2)
3	(36.47486034, 953.82122905, 46.16201117)

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

-From the graph, the data was being loaded into Neo4j and following analysis was being carried out.

-The longest conversation chain in the graph chat data is obtained using the ‘Respond to Edge Label,
To do these,we first find how many chats the participants were being involve ,and secondly how many Users were in the chain which gave the result 9.

- To obtain the top chattiest Users,we match all the Users with the CreatChatEdge,which inturn retun the Users ID and the count of the Users.

- We can futher explore the Neo4j graph by analysis the relationships and the interactions that exist between the Users in each teams such as RespondTo an item which is how the individual Users in their various teams behave when an avertisement is being made.Further analysis can be carried out using the following properties; like "Chat_start_team"which contains the ,the UserID,the TeamChatSessionId, and the TimeStamp showing when the event started ."The Chat_write_Team" which contains the UserId,TeamId, TextId,that's (a unique Id within Buying clicks for the purchase)."Level-events",which contains the Timestamp the time when the Team starts or finishes a level in the game,eventId;which is a unique Id of the event.The Team Id ,the Id of the Team,TeamLevel: The level of the team whether it started or has been completed.eventype: The type of event ,either (start or end).

-We also obtained the top chattiest teams by matching all ChatItems with a PartOf edge connecting them with a TeamChatSession node AND the TeamChatSession nodes must have an OwnedBy edge connecting them with any other node.


$ match (u)-[:CreateChat*]->(i) return u.id, count(i) order by count(i) desc limit 10

u.id	count(i)
394	115
2067	111
1087	109
209	109
554	107
999	105
516	105
1627	105
461	104
668	104

Started streaming 10 records after 745 ms and completed after 745 ms.

Recommendation

* Provide more products of different varieties to the High level spenders.Since the high level spenders purchase more produts providing them with new products can push them to buy more items and increase the revenue.

* Offer promotions and affordable prices to the low level spenders.We noticed that the low level spenders play more but spend less.So, by offering them with some promotions and affordable prices, can give them the opportunity in purchasing more products hence which could lead to an increase in the revenue .

*** The Eglence company ,should be capable in identifying the type of products the users are attracted too and which products the Users are likely to purchase for future marketing because from the decision tree we do observe that the Highrollers which uses mostly the Iphone platform should be offered more products.**