How can we increase revenue from Catch the Pink Flamingo?

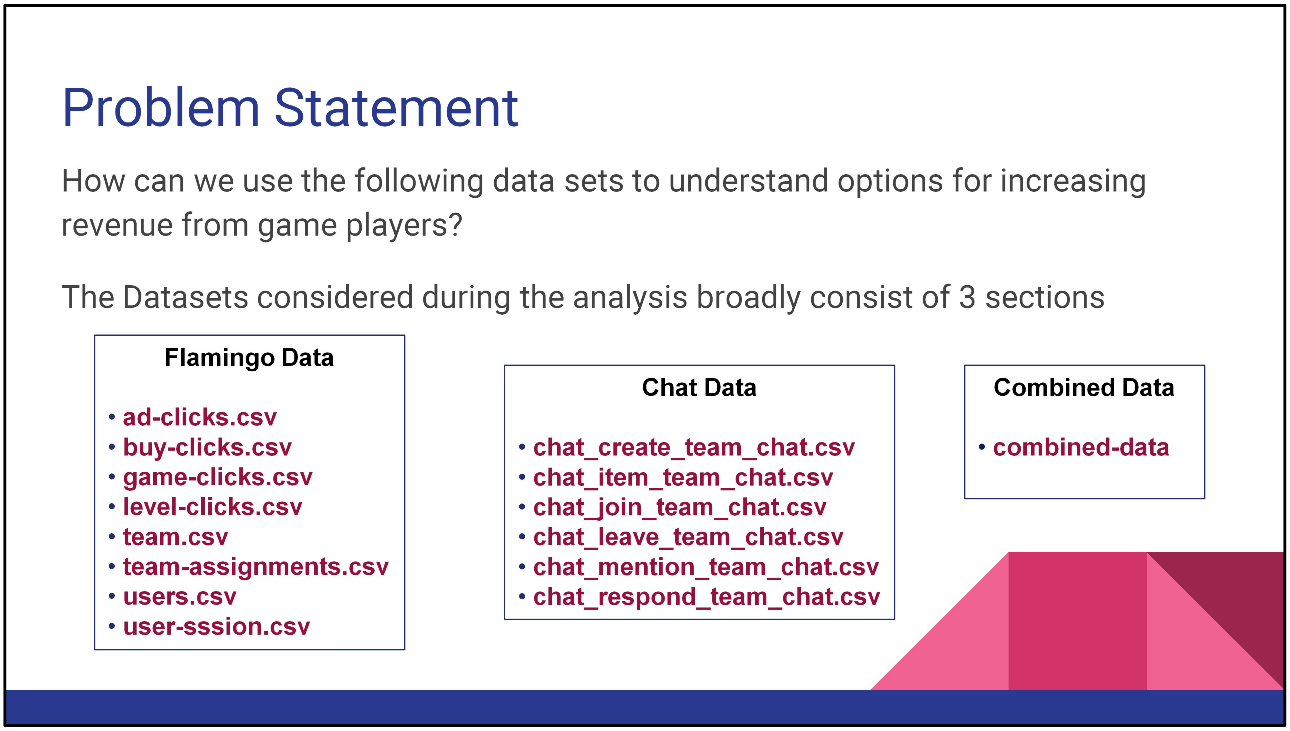
SUVARNA P

Good morning and a warm welcome to one and all.

My name is Suvarna. I am the chief analytics officer at Eglence Inc.

Today, I will be discussing on how we can increase revenue from our famous online game "Catch the Pink Flamingo".

In the next few slides I will swiftly go through the analysis procedures our analytics team has carried out focusing solely on the revenue enhancement from the game. I will also put forward some of the recommendations we have devised that can be implemented for further development of the game.



To realize our problem statement we made use of four different analytics methods such as Data exploration, Classification, Clustering and Graphic analytics techniques. The tools used for these were Splunk, KNIME, Pyspark – Jupyter Notebook and neo4j respectively.

The analysis process was done using the predefined datasets which were generated by our development team. These datasets are as shown here,

The Flamingo data contains 8 CSV files containing simulated game data and log data for Catch the Pink Flamingo. This set of data was made use while Exploring Data Sets and for clustering analysis methods.

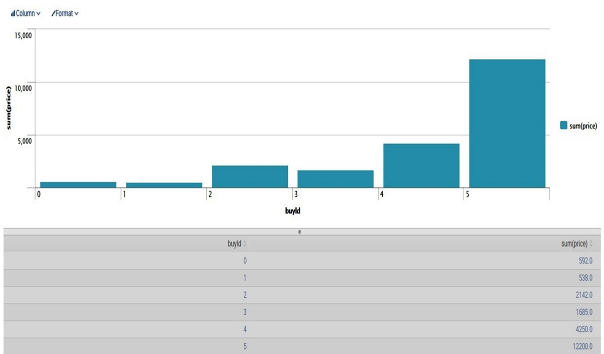
The combined data contains a single CSV file created by aggregating data from game data files like user-session.csv, buy-clicks.csv, and game-clicks.csv. The combined-data.csv file was made use while performing Classification and Clustering analysis.

The chat data contains 6 CSV files representing simulated chat data related to the Catch the Pink Flamingo game. This set of data was made use while performing Graph Analytics.

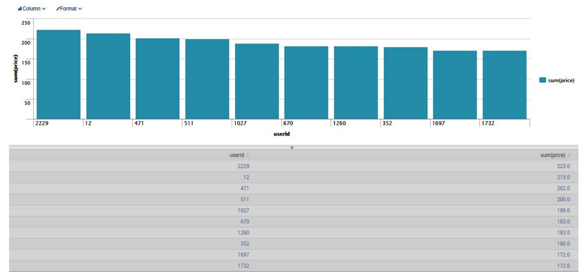
Data Exploration Overview

AGGREGATION

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



FILTERING



Rank	User Id	Platform	Hit-Ratio (%)
1	2229	iPhone	11.5970
2	12	iPhone	13.0682
3	471	iPhone	14.5038

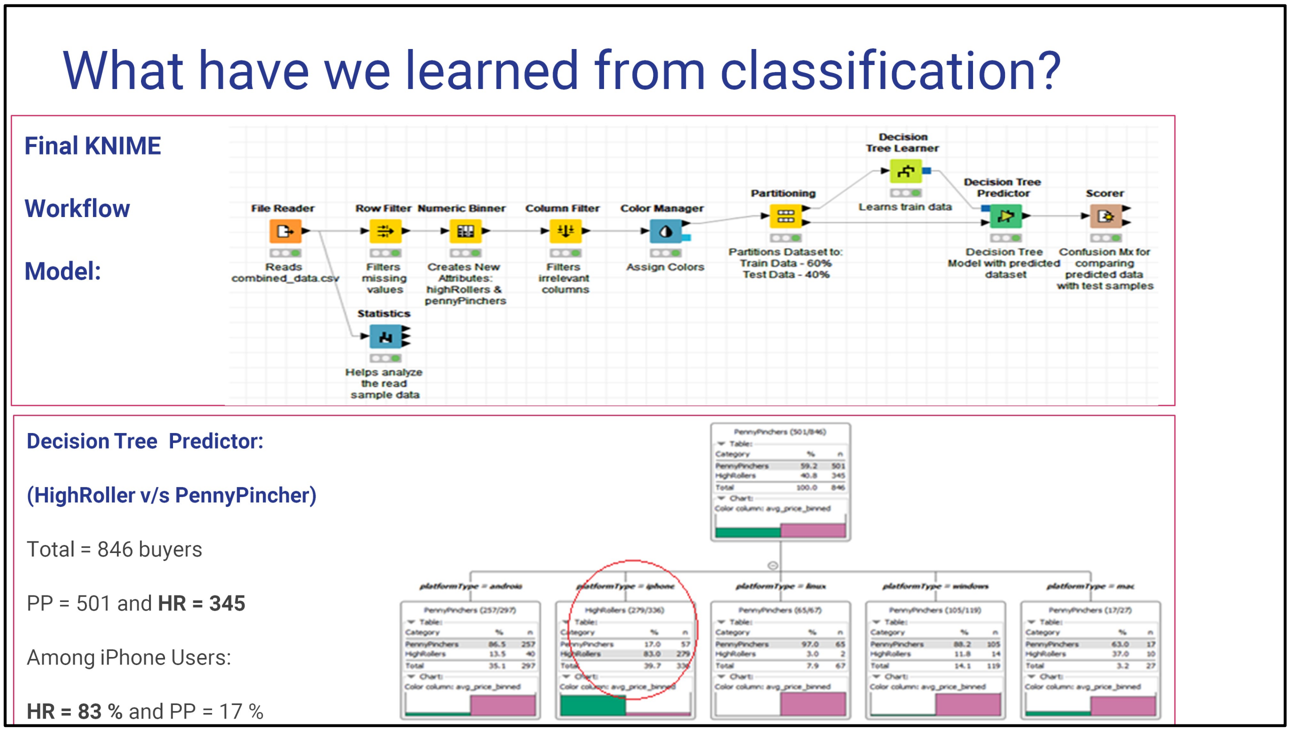
"Catch the Pink Flamingo" Data Exploration process was carried out using Splunk's high speed indexing and search platform.
Some of the simulated game data files generated by the "Catch the Pink Flamingo" python scripts lend themselves to analysis with Splunk.

There were two parts in data exploration:
Aggregation & Filtering.

Aggregation: Splunk allows to perform aggregation operations such as sum and average on our dataset and the results could be represented as histograms for better understanding. Using these operations we analyzed the following:
1. Total amount spent by all the users for buying items (Shown in slide)
2. Number of unique items available for purchase (Shown in slide)
3. How many times each item is purchased?
4. How much money was made from each item? (Shown in slide)

Filtering: Filtering helped to calculate the aggregate operations on a subset of the data. Using filtering methods we analyzed the following
1. Total amount of money spent by the top ten users (ranked by how much money they spent). (Shown in slide)
2. Extract the user id, platform, and hit-ratio percentage for the top three buying users (Shown in slide)

Examples from each of the above are demonstrated in the slide.



Classification was performed using Decision Tree Predictor in KNIME workflow software.

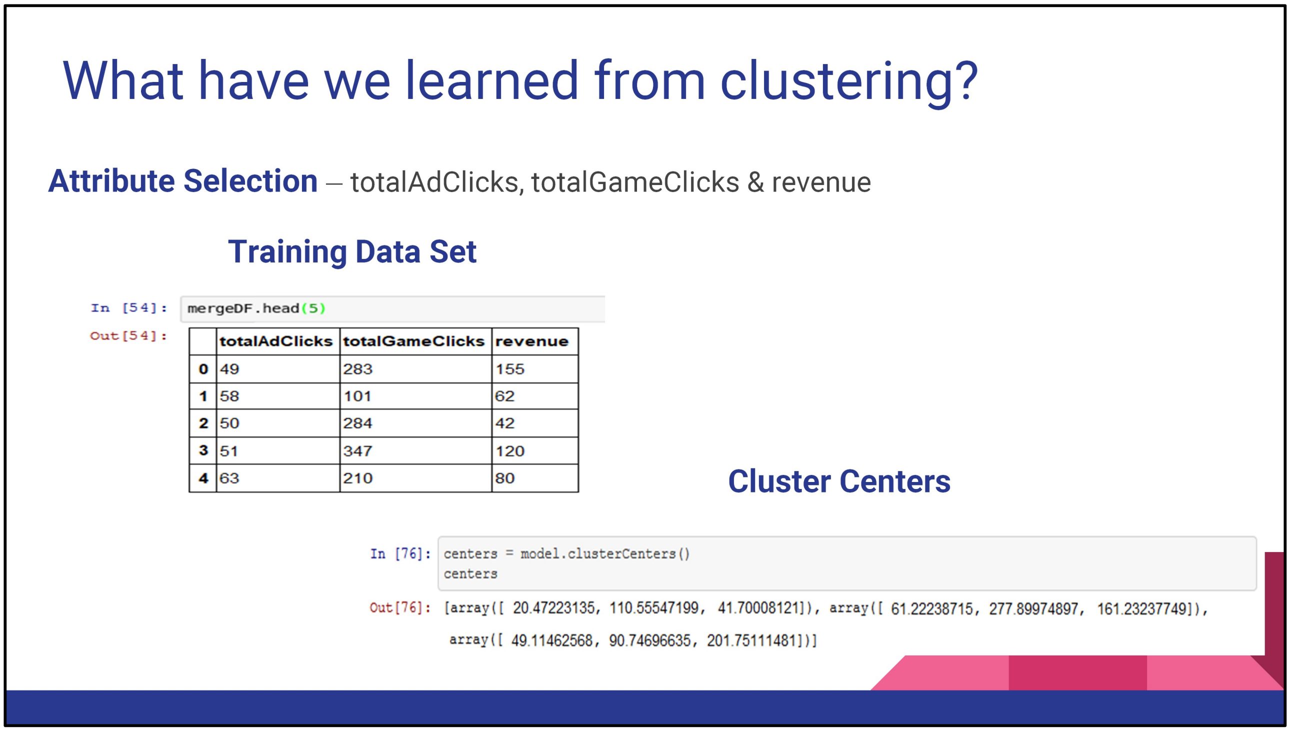
Predicting which user is likely to purchase big-ticket items while playing Catch the Pink Flamingo is valuable knowledge to have for Eglence since in-app purchases are a major source of revenue. Here we analyzed available data to classify users as buyers of big-ticket items (price > $5.00) as "HighRollers" and buyers of inexpensive items(price < $5.00) as "PennyPinchers".

There were 4 parts in our analysis:

1. Data Preparation, 2. Data Partitioning & Modelling, 3. Data Evaluation & 4. Analysis Conclusion

A KNIME workflow was build using the combined-data.csv file to perform analysis for this classification problem.
The combined data file is an aggregate of user-session.csv, buy-clicks.csv, and game-clicks.csv.
The data file "user-session.csv" contains a column of data for "platformType". This column is 'enumerated' with five values: windows, mac, android, iphone, linux. The relative distribution of these operating systems was analyzed for our dataset.

The results as seen in the slide shows that most HighRollers were using the iPhone platform, hence as a suggestion we are recommending Eglence Inc., to focus the pricey ads on these potential high revenue generating users.



Clustering can be an important tool for developing insights in many areas of a company's business.

Our Eglence users were clustered into groups based on characteristics such as their game playing behavior, purchase behavior, inclination to click on displayed ads, etc

The user data was used to create 3 clusters, and these were analyzed based on game-clicks, ad-clicks and total revenue generated by each user.

Our clustering process consisted of 3 parts: As shown in the slide)
Attribute Selection
Training Data Set Creation
Train to Create Cluster Centers

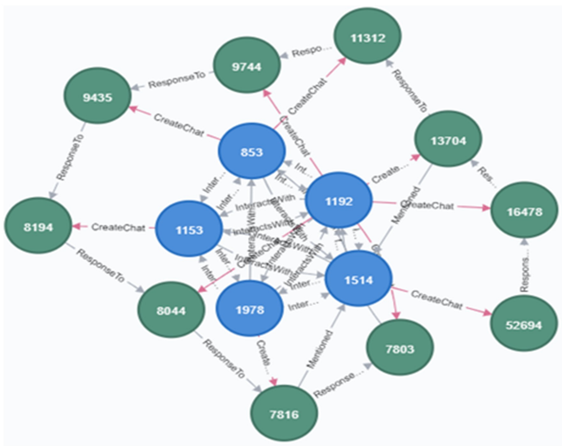
From the analysis performed using K-Means clustering cluster centers for the 3 clusters were formed.

These centers were analyzed as to which cluster had:
High/Low adclicks, High/Low game-clicks and High/Low Total revenue.

Based on these findings we have some recommendations to Eglence Inc. that will have a direct or indirect impact on their revenue. These will be discussed in the last slide.

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

Longest conversation chain with participating users



Top 3 Chattiest Users

Users	Number of Chats
{“id” : 394}	115
{“id” : 2067}	111
{“id” : 1087} & {“id”: 209	109

Most Active User

User with **User ID : 209**, as this user has the highest cluster coefficient

User ID	Coefficient
209	0.95
2067	0.5238
394	0.5
554	0.904
1087	0.8

The chat graph analysis was performed using the neo4j graph analytics tool.

The chat data set used consisted of 6 CSV files which contains information of chats between all the users. This simulated chat data which relates to the Catch the Pink Flamingo game, currently consist of chat data that is purely numeric, no text.
Analytically, it can still serve a useful purpose in revealing certain types of behaviors which can only be observed within a graph analytics context.

Using this graph analytics method lot of critical information was extracted which can be used by Eglence Inc., to further improve the game. Some of these are:

1. **Longest conversation chain in the chat data**
2. **How many chats are involved in it?**
3. **How many users participated in this chain?**
4. **Do the top 10 of the chattiest users belong to the top 10 chattiest teams?**
5. **How Active are Groups of Users?**
6. **Who is the most active user?**

Some of these are shown in the slide.
Further exploration that can be done from graph data analytics are:

1. **Which is the shortest conversation chain in the chat data?**
2. **Who is the least active user?**
3. **Does a top ten chattiest team contain a less active user?**

Recommendations

1. To decrease the number of PennyPinchers small discounts, coupons, promotions etc. can be added to High ticket products from time to time.
2. To prevent the iPhone users from becoming PennyPinchers an increase in the volume and variety of similar products can be included based on their purchase history.
3. Combo offers of inexpensive products can be introduced to iPhone users so that the sale of products with cost < $5 goes high.
4. Improve game strategies and include inbuilt game assisting tool purchases
5. Identify the adCategory a user might be interested in from their social media activities
6. Improve product varieties and introduce exciting offers to the existing high revenue generators

From Cluster Analysis we can point out the following recommendations:

Moderate players seem less interested in clicking the product ads or buying products outside the game.
To enhance their gaming frequency interesting strategies can be included in the game. Each mission can include purchasable inbuilt game assisting tools that can be used to reduce the game difficulty.

The most frequent players who are also frequent ad clickers do not make a lot of actual purchases. Such users might be clicking through the ads searching for those things that they really require.
Identify those products from their browsing history and social media activities and put up those product related ads in their game sessions.

Those players who are the least frequent players tend to produce high revenues, so by including exciting perks on completion of a set of levels, like special discounts on their product purchases, might enhance their playing frequency and also improve sales.

With that I am concluding my presentation. Thanks a lot to everybody who is present here for being patient with me. Hope Eglence Inc., will benefit from our analysis report.