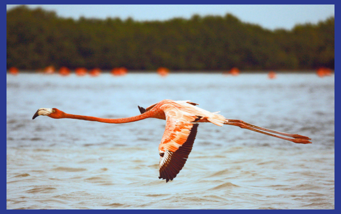
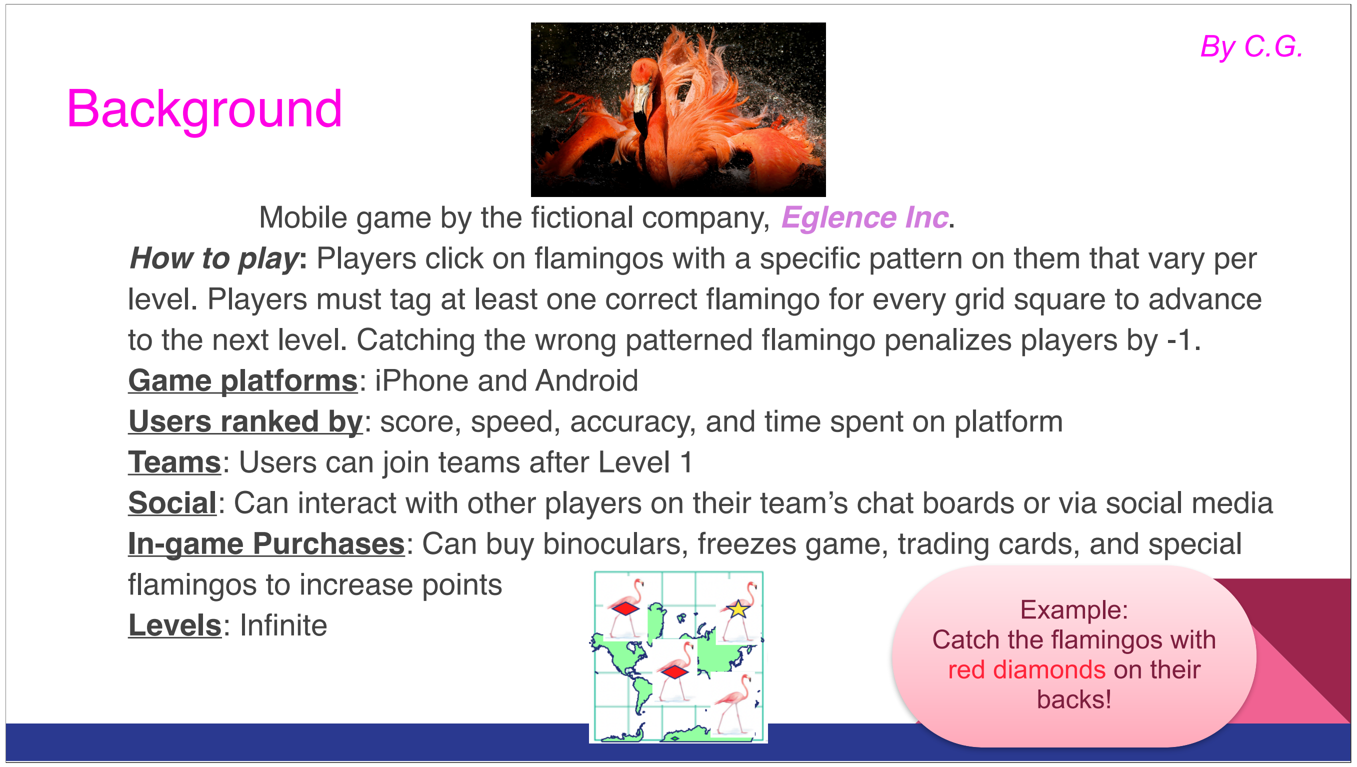
How can we increase revenue from *Catch the Pink Flamingo*?

By C.G.



“American Flamingo.” Audubon, October 20, 2021. https://www.audubon.org/field-guide/bird/american-flamingo.

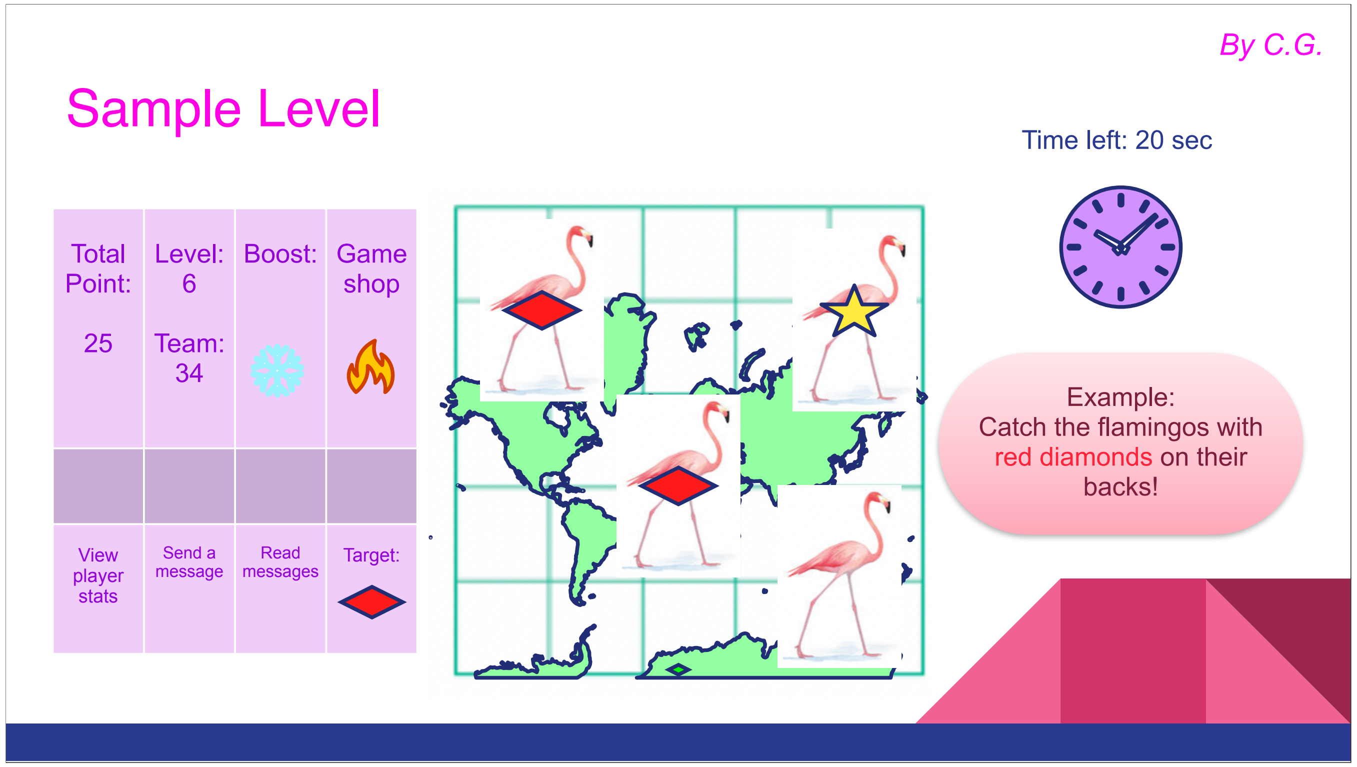


We will be using data from the fictional mobile game, **Catch the Pink Flamingo by Eglence Inc.**, to learn more about Big Data platforms.

As a background, Catch the Pink Flamingo has a world grid with flamingos on it. Players have to click flamingos with a specific pattern on them to earn points. The game's grid and flamingo pattern will change in each level.

After the first level of the game, players can join teams. Teams are a fun way to join multi-user missions and interact with other players on chat boards or social media. The game has infinite levels, each more challenging than the level beforehand.

Eglence earns money from the data it collects on its users. Ads are shown to players, and in-game profits directly increase the game's revenue.



Here’s an example of how the game works. In the sample level above, a player must click on the flamingos with the red diamonds to earn points. If a user clicked on the flamingo without a shape, for instance, the player would lose a point.

Players can visit the game shop to buy boosts to help them earn more points. In this case, the player can click on the freeze boost they purchased. A freeze boost will temporarily freeze the clock and allow the player extra time in this level.

The player can view their player game stats, points, team name, and current level. Players can also send and receive messages from other players while they’re in the middle of a match.

Each player uses the platform differently. Eglence tracks this information on its players to make a profit.

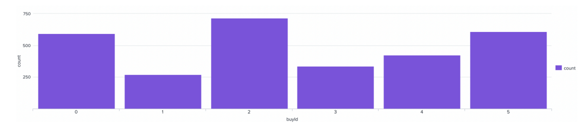
For instance, one player may use an iPhone while another player may use an Android. One player may also be more social on the chat boards than another player. Players are shown ads and may make purchases in the in-app store to help them earn more points. With all these choices on how to use the game, some players will be more profitable to Eglence than others.

This leads us to ask: How can we use the following data sets to understand options for increasing revenue from game players?

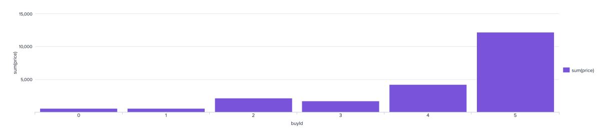
The variety of big data sources are crucial for the game to make a profit because they all shed light on different aspects of users' gaming behaviors. This analysis will use Splunk, KNIME, Python, and Neo4J to determine how data from players' gaming, spending, and social habits can improve the game's marketing strategy and increase revenue.

By C.G.

Data Exploration Overview



Number of Purchased Items



Profit from Purchased Items

Histograms created in Splunk, using queries written in Cypher

Of the six items available for purchase…

buyID 5 was the **most profitable**
buyID 1 was the **least profitable**

buyID 2 was the **most purchased** item
buyID 1 was the **least purchased** item

Rank (by Christina G)	User Id	Platform	Hit-Ratio (%)
1	2229	iPhone	11.60%
2	12	iPhone	13.07%
3	471	iPhone	14.50%

More purchases -> Lower Hit-Ratio
Top 3 spenders all have an iPhone

For data exploration Splunk was used to create the two histograms above and to write queries to fill in the table on the right.

There are six items available for purchase in the in-app store. The most profitable item was buyID 5. The least profitable item was buyID 1, closely followed behind by buyID 0.

The bottom histogram reveals that the most purchased item was buyID 2. The least purchased item was buyID 1.

After analyzing both histograms above, it is clear that buyID 1 is the least popular item. The game makes least profit from buyID 1, and the item is purchased the least. In contrast, buyID 0 is purchased frequently, likely because it may be the cheapest item in the game. Meanwhile, buyID 5 is a popular item despite its high price tag.

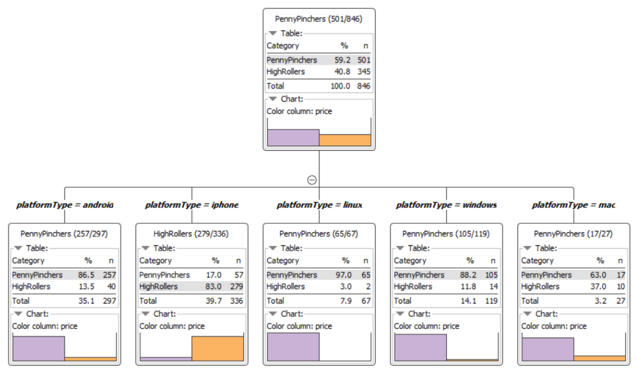
Although there are exceptions, the more popular an item is, the greater the revenue will be. Note that price doesn't necessarily correlate with popularity for all items. For instance, buyID 2 was the most purchased item, yet it made the game little profit. This may be because buyID 2 could correspond to a helpful game boost that players use frequently, but it is not necessarily an expensive item. Despite these exceptions, the most purchased items in

Based on the table above, the more money a player spends, the lower the Hit-Ratio the Player may have. This is likely because purchases may award game players an advantage in the game over other players. Also, the top three spenders players all have iPhones. This is likely because those who can afford an iPhone can also afford to make more game purchases compared to other platform users.

By C.G.

What have we learned from classification?

Decision tree in KNIME



Two types of spenders:

- **PennyPinchers** (**low** spenders)
- **HighRollers** (**high** spenders)

Mobile platforms received **3 times** more users than desktop platforms.

Correct classified: 500 Wrong classified: 65

Accuracy: 88.496 % Error: 11.504 %

Cohen's kappa (κ) 0.76

price \Pre...	PennyPinc...	HighRollers
PennyPinchers	308	27
HighRollers	38	192

In data exploration, we learned that not all of the most purchased items generated a profit for the game developers. In classification, we divided users into groups based on their spending habits. Then, we analyzed which types of users would buy items that would generate the greatest profit for the game.

KNIME was useful for creating a decision tree to predict users' spending habits. The decision tree was highly reliable with classifying users, with an approximately 88% accuracy. The lowest spenders were nicknamed PennyPinchers, because they spent less than $5 on an item. The highest spenders were nicknamed HighRollers and spent $5 or more on an item.

The decision tree reveals that approximately High-Rollers were predominately iPhone users. Most players used the mobile platform compared to the desktop platforms. Mobile platforms had roughly the equivalent amount of players. Over half of all desktop users had Windows, with hardly any users using a Mac.

Given the high price of Apple products, it isn't a surprise that most High-Rollers have Apple products. Marketing should likely focus on mobile users, given that roughly 75% of all users play the mobile version of the game.

By C.G.

What have we learned from clustering?

Machine Learning Model: KMeans

Cluster #	Center
1	[40.87037037, 9.75925926, 138.24074074]
2	[25.29532164, 4.28947368, 15.23684211]
3	[34.63945578, 6.45578231, 59.12244898]

The indices for each cluster are as follows: **[totalAdClicks, totalBuyClicks, revenue]**

Strong, positive correction among total amount of:

- **ads clicked on by a user**
- **game purchases made by a user**
- **revenue**

The tables below were created in Python:

```
trainingDF = combinedDF[["totalAdClicks", "totalBuyClicks", "revenue"]]
trainingDF.head(n=5)
```

	totalAdClicks	totalBuyClicks	revenue
0	44	9	21.0
1	10	5	53.0
2	37	6	80.0
3	19	10	11.0
4	46	13	215.0

```
trainingDF.describe()
```

	totalAdClicks	totalBuyClicks	revenue
count	543.000000	543.000000	543.000000
mean	29.373849	5.419890	39.349908
std	15.216343	3.244713	41.221737
min	1.000000	1.000000	1.000000
25%	16.000000	3.000000	10.000000
50%	30.000000	5.000000	25.000000
75%	42.000000	7.000000	55.000000
max	67.000000	16.000000	223.000000

In the previous decision tree, we learned that users' gaming platform affects their likelihood to spend more in the game. For the next phase of classification, we used K-Means clustering in Python to analyze how users' ad clicks and game purchases increase Eglence's revenue. K-Means is a type of unsupervised machine learning model that categorize players without explicitly assigning them into labeled groups.

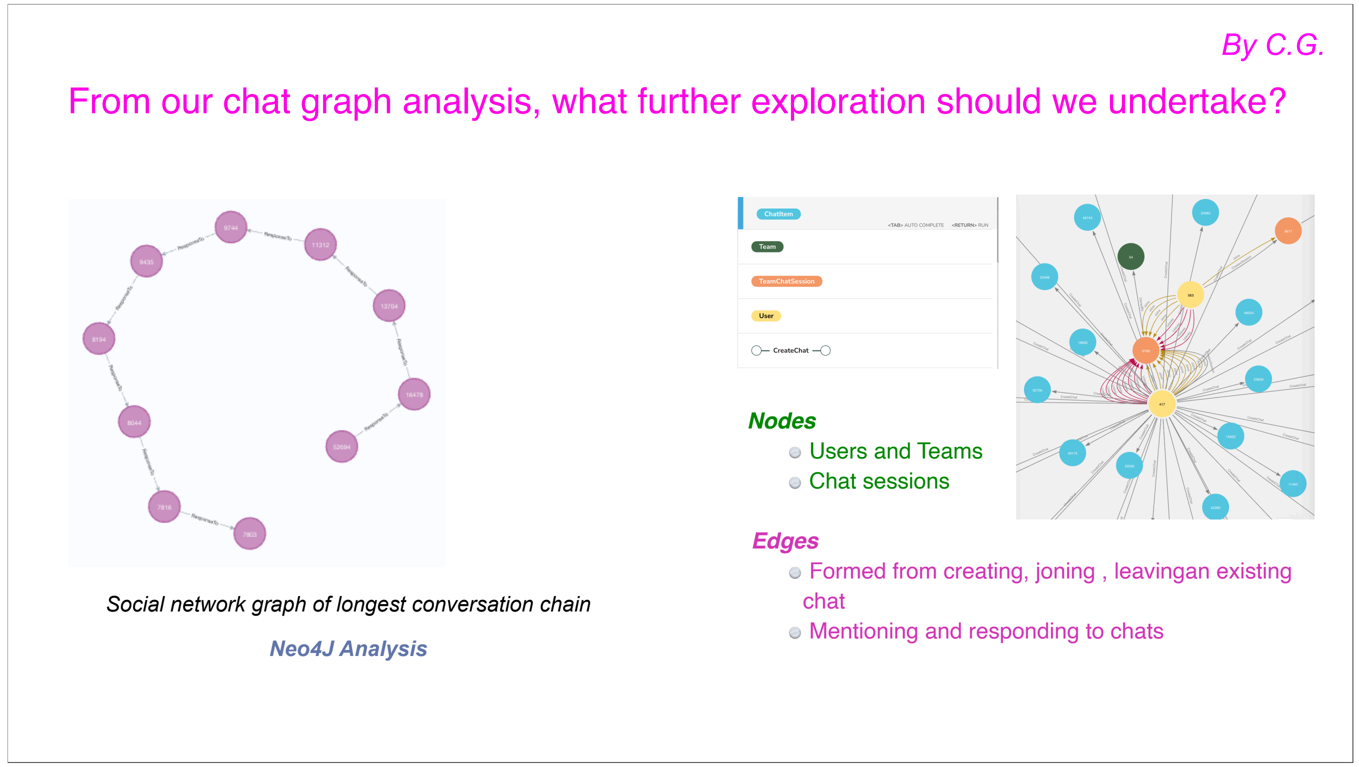
A sample of the training data is shown on the top right. The bottom right table reveals that on average, users click on 29 ads, buy 5 items in the in-app store, and generate $39 in revenue.

The first cluster shows that users who click on the most ads and make the most purchases also generate the highest revenue in the game.

In contrast, cluster 2 shows that users who click on the least ads and make the least purchases also generate the lowest revenue in the game. Cross-referencing with the bottom right table, Cluster 2 is essentially the average user.

Cluster 3 is a great comparison point among the other clusters and falls between Cluster 1 and 2 for ads, purchases, and revenue.

Analyzing all clusters together, we can conclude that there is a strong, positive correlation among the total amount of ads users clicked on, the total amount of game purchases, and the total amount of revenue that the game generates from ads and purchases.

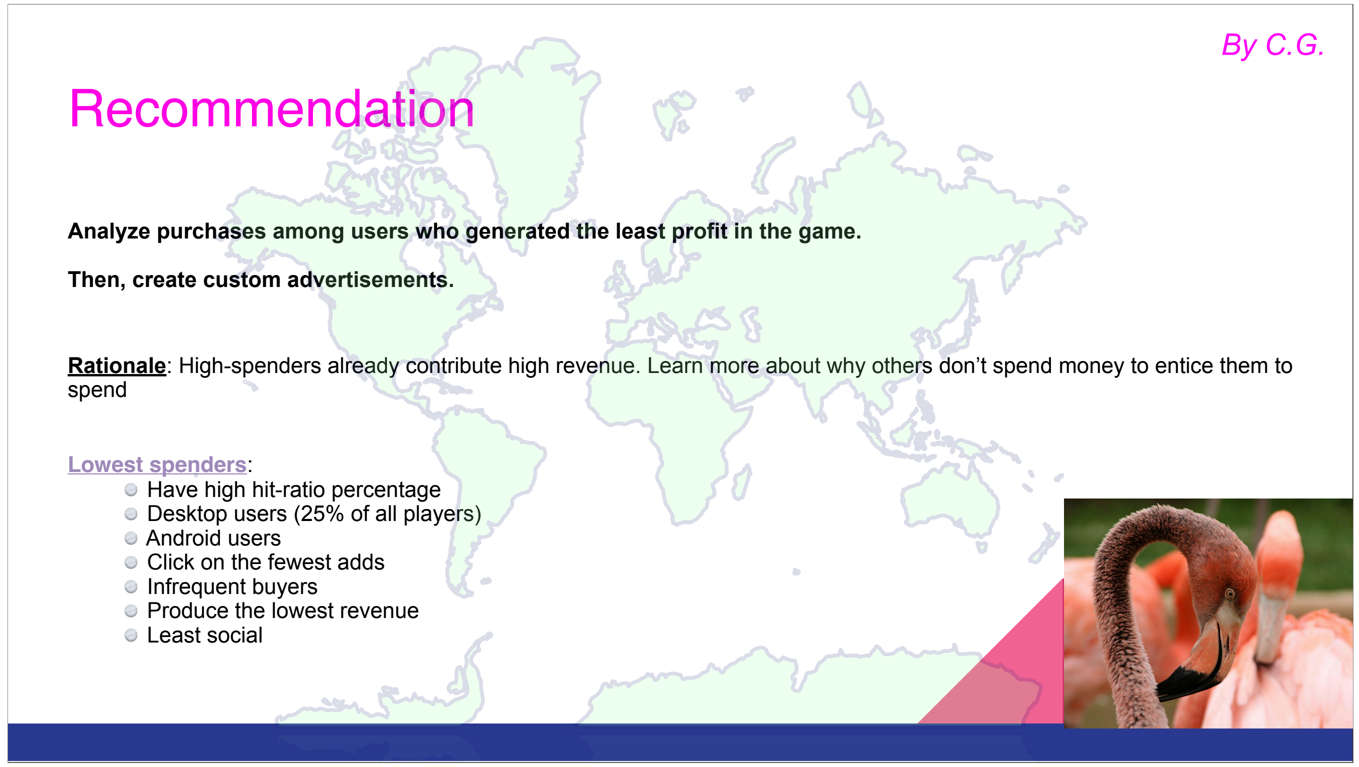


Neo4J was a helpful tool in capturing players' social interactions. Eglence collects information on the relationship formed between all users who interact with another use.

In the example social network graph on the right, we can see users, teams, and chat sessions are denoted as nodes. The interactions that these nodes have are denoted as edges. Edges are created whenever someone creates, joins an existing chat, and whenever a user mentions or responds to another user's chats. If a user, for instance, joins a team, an edge would be created between the user and their new team. Similarly, if a user writes in the chatroom about another user and another user responds back to the writer, these dialogue exchanges between players would be marked with edges and the nodes would be the users and that chat session.

Since interactions are often numerous, these graphs quickly become large. The complexity of these graphs are also indicative of the vast amount of knowledge we can gain by learning about these interactions. On the left, the social network graph was created to find the longest conversation chain among players in the game. The graph reveals that max 10 players have participated in a single conversation thread.

By learning more about how players interact, Eglence can revise its marketing scheme to appeal to its players. If players, for instance, mostly talk about their scores, Eglence can focus its advertisements on products that increase their scores.



Despite using different big data tools for our analysis, all of the tools reveal that there is a divide among the lowest and highest spenders in the game. To improve marketing strategy, the game can cater advertisements to infrequent buyers in order to prompt these players to make more in-app purchases. This can be accomplished by analyzing the products these users buys and marketing these products towards these users.

For instance, if the infrequent buyers only purchase game boosts, these users can be shown more game-boost advertisements rather than showing users other items they likely will never buy.

We have already learned a lot about the lowest spenders, such as from the classification models, clustering and the social network analysis that were displayed in the previous slides. he lowest spenders have a high error rate for targeting flamingos, losing them points in the game. Lowest spenders are desktop users in general, and Android users among mobile users. The lowest spenders click on the fewest ads, infrequently make purchases, and also produce the lowest revenue in the game. These users are also the least social on social networks.

Eglence should revise its marking strategy to these lowest spenders. Custom ads can be displayed to Penny-Pinchers that advertise game-boosts and may entice them to spend more in the game.