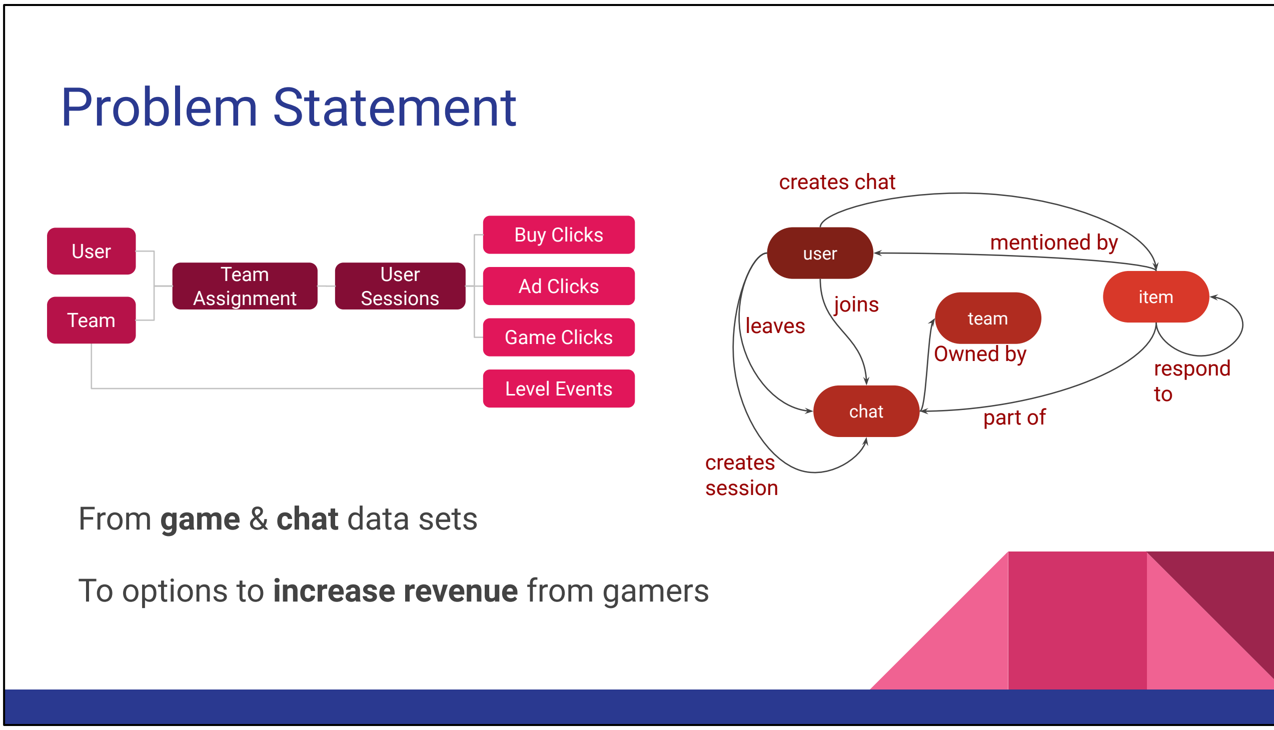
How can we increase revenue from Catch the Pink Flamingo?

Cedric's Profile

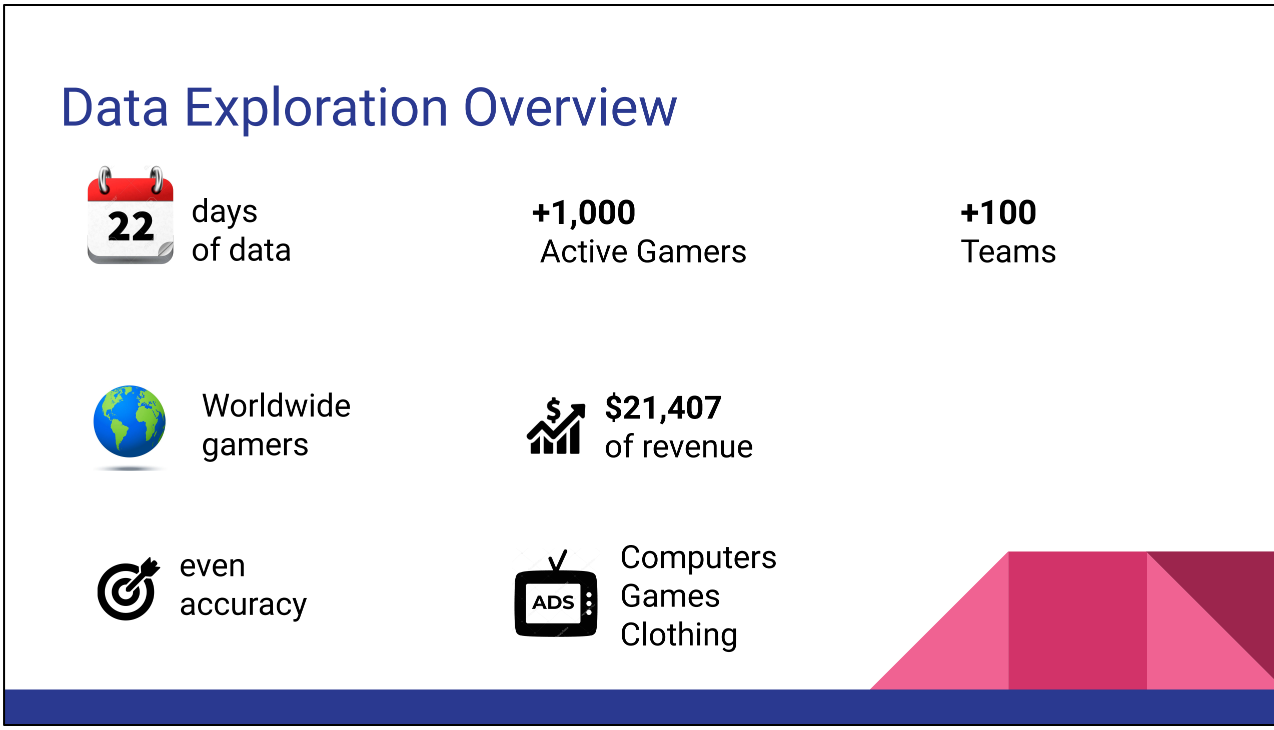
Cédric MEMBREZ, Switzerland

Good day ladies and gentlemen,
I am Cédric Membrez from Switzerland and I will be presenting my research on the game Catch the Pink Flamingo.
While analyzing the data, I keep in mind the goal to increase revenue for Eglence Inc.
Also, feel free to interrupt me if you have any question.



Thanks to our data-team colleagues whom acquired the relevant data, extensive **information on our users** are available. Information regarding game data (the left diagram) will enable us to gain **valuable insights** through **scientific methods** such as classification and clustering (that is grouping our users on specific characteristics). Additionally, through an **analytical tool** called graph analytics, the chat data (the right graph) will extend the **knowledge on the strength and direction of users' relationships**.
This presentation and its technical appendix clearly show that whether the **kinds of data** is categorical or numerical, or whether we face **data sourced** from social networks or in-game, we have the techniques and tools to **prepare** the data through an in-depth exploration and pre-processing. Afterwards, we **analyze** the data through selected models and methods to find powerful results. We communicate these results to other teams at Eglence Inc. such as the management and our sales teams. Through this data science story, we gain insights on our users to ultimately find options to increase revenue from them.
The left diagram is about game data. Specifically, it contains details on users: such as date of birth and the country they are living in. We know the strength and level events of the teams. Through team assignment, we know which player joins which team and when. The user sessions give information about how long users are playing and on which platform. Finally, we have clicks of the users in-game which is ideal to find accuracy for example. Also, we have the clicks on advertisements in the app to understand which category of ad is preferred, and also the clicks for actual purchases of items to analyze what items and what prices are preferred by whom.

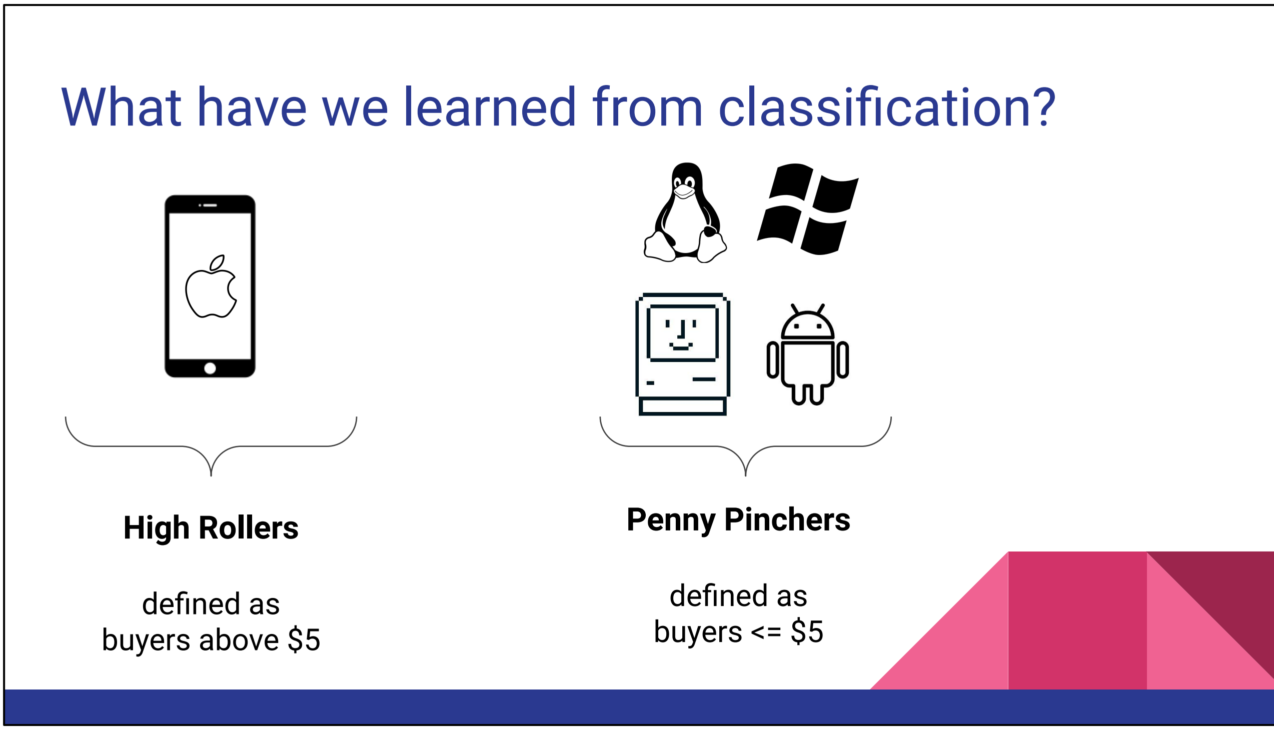
The graph on the right is about community of users in chat and includes only numeric values, no actual text or messages are available: however, we can analyze the interactions between users and better understand their relationships. We can now move to the data exploration.



The data covers 22 days between May 26th and June 16, 2016 and it counts 1,093 active players while over 2,400 registered. They are spread among ~100 teams.
Also, I account for registrations from all around the world included the seven seas with 79 gamers and the Antarctica with 10 gamers. Africa with 515 players is the top continent followed by Asia (487) and Europe (475).

Over the period, the active users spent in total $21,407 over the six items available. Their prices range from $1 to $20, and the most expensive item at $20 generated more than half of the total revenue. This would be interesting to investigate further. **Here the focus is on the price, but in future studies, I highly recommend to understand what are the characteristics of the items, besides their price**.

The accuracy of a gamer actually clicking on a flamingo in the game, also called hit ratio, is relatively even at 11% on average both among the users and the teams. Only one player peaks at 50% and one team at 20%. For an initial analysis, it might not be a priority to explore it in length. For the data on advertisements, computers, games and clothing categories are the top three. Automotive, electronics and hardware are the lowest ones. Interestingly, each category has its peak (that is, highest count of clicks) on either the 14th or 16th of July. There is not details on the monetization - but it could be of interest in future analyses. A final point on the time played. 41% of the gamers played less than 200 hours and 32% of them played over 400 hours. That leaves 27% playing between 200 and 400 hours. A user session lasts on average 2.5 days (!), and ranges from 0.5 to 98 hours. *At that rate, we might suggest food delivery in the items…*More seriously, I will present now the classification.

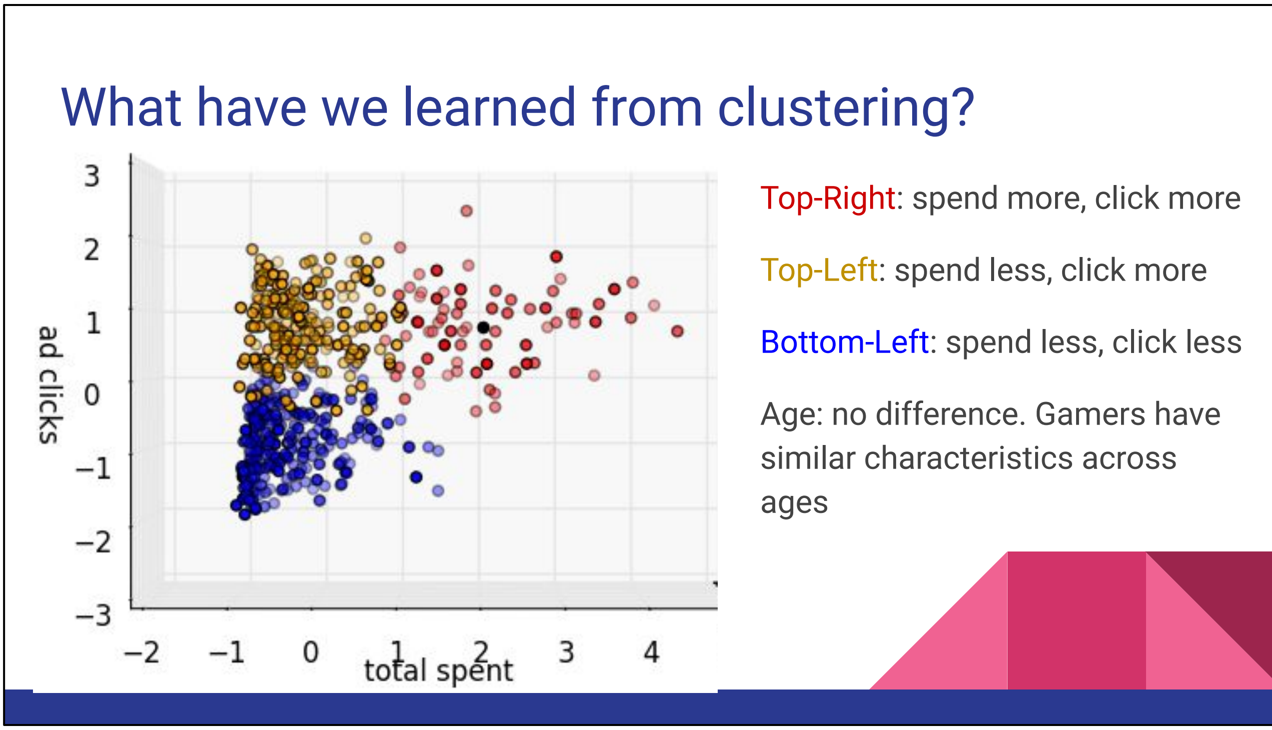


From the data exploration, and because we seek to increase revenue, it was surprising that the most expensive item actually brought more than half of the revenue. Because of that, I used a decision tree algorithm to classify our users. In addition to our data, I created a new categorical data based on the average price spent by our users. For average prices strictly above $5, the user is tagged as a **high roller**, and if the average price is below or equal to $5, the user is called a **penny pincher**. The goal of our classification is to know what defines a gamer that spend a lot (or so-called high roller) and what defines a gamer that spend little (or so-called penny pincher) on items.

The characteristics available at the beginning of the classification were the platform on which the gamer played, the team level, the number of in-game clicks, number of hit-clicks and purchase-clicks. The decision tree was set up in such a way that characteristics that have little power of explanation are discarded. The goal is to have a lean tree that classify well new players and that is relatively easy to understand.

The results were relatively clear. The only characteristic left, to classify high roller against penny pincher, was the platform used. Gamers playing on an **iphone have a strong tendency to buy expensive items** (in our case the items that cost either $10 or $20). Gamers playing on other platforms (Linux, Windows, Mac, and Android) have a tendency to buy the cheaper items (in our case the items that cost either $1, $2, $3 or $5). Based on several ratios and measures, the decision tree has the potential to generalize well for new users.

The **accuracy was around 88%:** in other words, the model correctly classified 500 gamers while 65 were not. Two other measures, one of exactness and one of completeness, were around 90% and confirm the good quality of the model. I invite you to consult the technical appendix for details, page 18 for the ratios. Next, I will explain the clustering method.



In the slide before, I pre-defined the ‘high rollers’ and ‘penny pinchers’ classes in which gamers would be assigned based on their characteristics. That was **classification**. Now, I will seek similarities between our gamers based on three characteristics: age, number clicks on advertisements and total spent on items’ purchases. There is no pre-defined group to which the gamers are assigned. This is **clustering**.

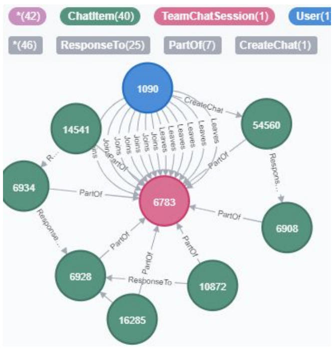
This is actually a 3D-plot where each dot represents a gamer. The third axis, the age of gamers, is hidden for clarity. Actually, the gamers are evenly spread-out across different ages. From 14-year old to 69-year old with a mean at 37-year old, no difference was visible. However, with the total spent on the horizontal axis (spending less on the left, more on the right) and the number of advertisement clicks on the vertical axis (spend less on the bottom, more on the bottom), interesting grouping (or cluster) is occuring.

The red cluster groups gamers that spend and click a lot more. By the density of the dots, we can see there are fewer of them. However, they are supposedly the gamers that contribute the most to our revenue (on a per-gamer basis of course). **Again, we should go a step further with ad-clicks to understand how it is actually monetized for Eglence, Inc..**

To reinforce this, the orange cluster in the top left corner regroups players that spend much less but are still clicking heavily on the advertisements. As a consequence, they could be an important source of revenue for Eglence, Inc..

Moving in the opposite direction, the blue dots in the bottom left corner represent gamers that spend and click less. That concludes the analysis on game data and we can move to the chat data.

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



cross-references

to game data

It was an important step to set up the analysis of our chat data. We can understand better the strength and directions of relationships among our gamers community. For example, it was possible to list the users that chat the most and that have the strongest relationships.

Interestingly, among those top users, a cross-reference with the game-data shows that one user (id 209) has not made any purchases nor did she click on any advertisement. However, she did play over 437 hours with an average hit ratio of 13%. Similarly, a second user (id 668) played 275 hours with a hit ratio at 14% on average. User 999 does not have any of these records.

Among the top 10 active users, user id 1087 did purchase seven items (1x #0, 3x #1, 2x #2, 1x #5) for a total of $33. In addition, she clicked on 33 advertisements across multiple categories such as computers, hardware and sports. Finally, she plays on a mac and evolved up to the level 6 currently. As six out of seven items are below the $5 threshold and that she plays on a mac, she is a penny pincher.

This cross-reference is for illustration purposes. If Eglence, Inc. wants to reward the most active users, it has to approach them with the most suited gift or promotion.

Recommendation

- Offer a wide selection of cheaper items to Linux, Windows, Mac and Android
- Provide a richer offering of luxury items to iphone users
- Wider offering & marginal price increase for users that spend and click a lot

Understand Items Characteristics to Improve Items Offering:

Increase Number of Items & Widen Price Range

As a final note, I want to focus on the items offering currently available.

Through the current analysis, it has been shown that Iphone users tend to have a preference for more expensive items. However, these gamers can only choose one item at $20 and another one at $10. They would probably benefit from intermediate choices or more expensive choices with added value of course.

Additionally, our penny pinchers, the gamers on Linux, Windows, Mac and Android platforms, are enjoying four items between $1 and $5. We can argue that they would have the potential to buy a few new items in the same prices range.

A wider offering targeted to the right platforms and users could be very valuable for Eglence, Inc..