How can we increase revenue from Catch the Pink Flamingo?

Max Pan Ziyuan

Problem Statement

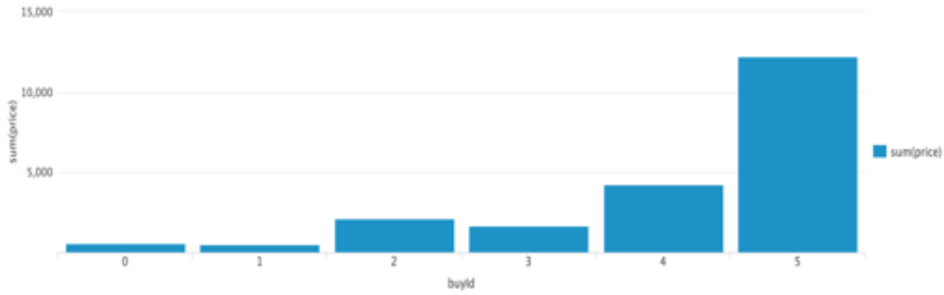
How can we use the following data sets to understand options for increasing revenue from game players?

- Flamingo dataset: Analyze the behavior of the users.
- Combined dataset: Categorize the users into different groups based on their behavior.
- Chat dataset: Social network analysis, targeting advertisement

The company has three main datasets. Flamingo dataset includes the detailed information about the user behaviors. With this dataset we can understand the association between the clicks events and the events like in-app purchase. Combined dataset contains users' information and allows us to categorize the users. Chat dataset is the data related to all the chat group events. We can get the information about the connection between users and teams with this dataset.

Data Exploration Overview

- Total revenue: $21407
- Most profitable item



We used splunk to explore the data. Splunk is a very convenient tool for data analysis. We run the queries in splunk to get more familiar with the dataset. And we are able to find the information like total revenue, most profitable items and etc.

What have we learned from classification?

By classifying the users, we learned the following things:

- What is the feature that differentiate the users? platform
- Users from which platform are more likely to buy items? iPhone

We trained a decision tree to classify the users based on the features including platform, team level, game clicks. And it turns out that the decision tree only need the information about the platform to classify the users. And iPhone users more likely to spend more than the users on the other platforms.

What have we learned from clustering?

Clustering helps us understand how many types of users there are.

Cluster #	Cluster Center
1	[-0.0967, 0.02876, 0.8414]]
2	[2.322, 0.0707, 0.8612]
3	[-0.0656, 2.5974, 0.1903]
4	[-0.4427, -0.5243, -0.9249]

These clusters can be differentiated from each other as follows:

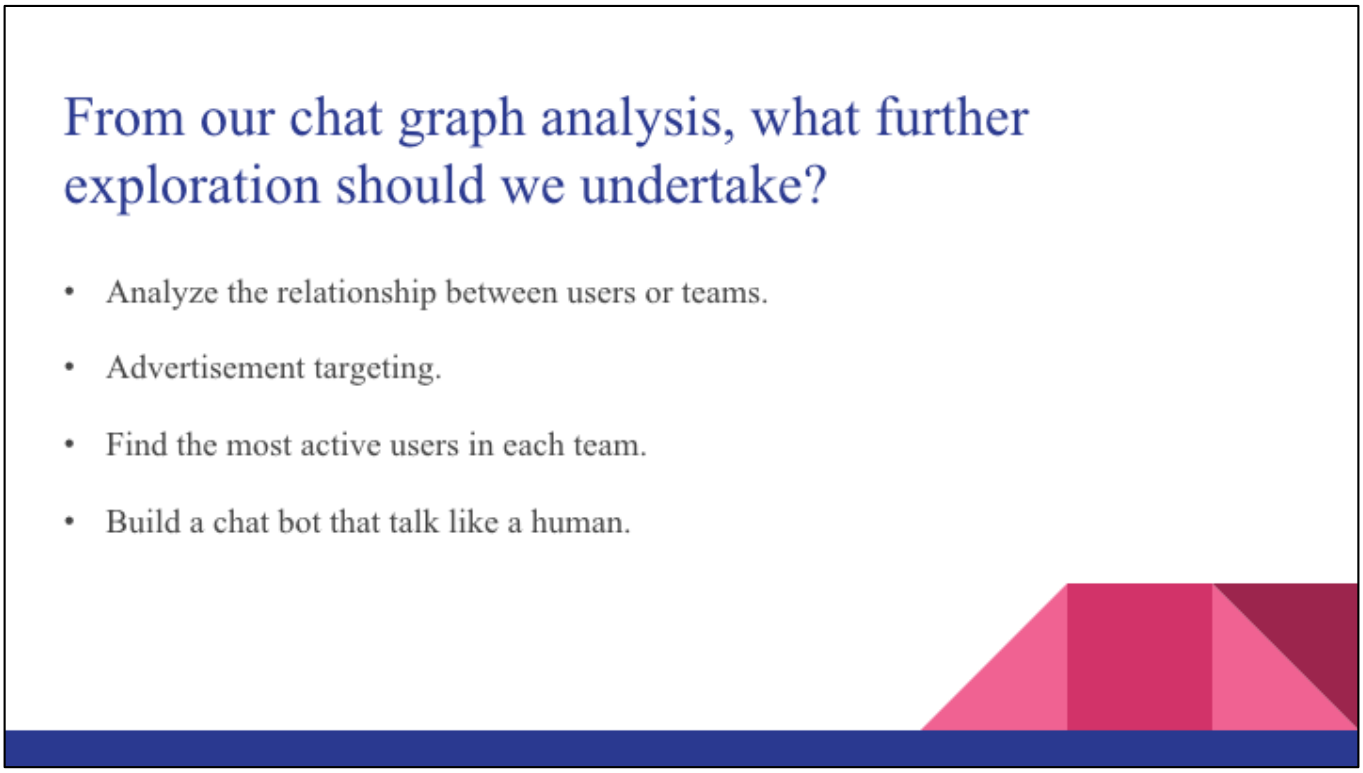
Cluster 1 is different from the others in that the users click on the ads very often but seldom pay for the items.

Cluster 2 is different from the others in that the users click on the ads a lot and also like to buy the items.

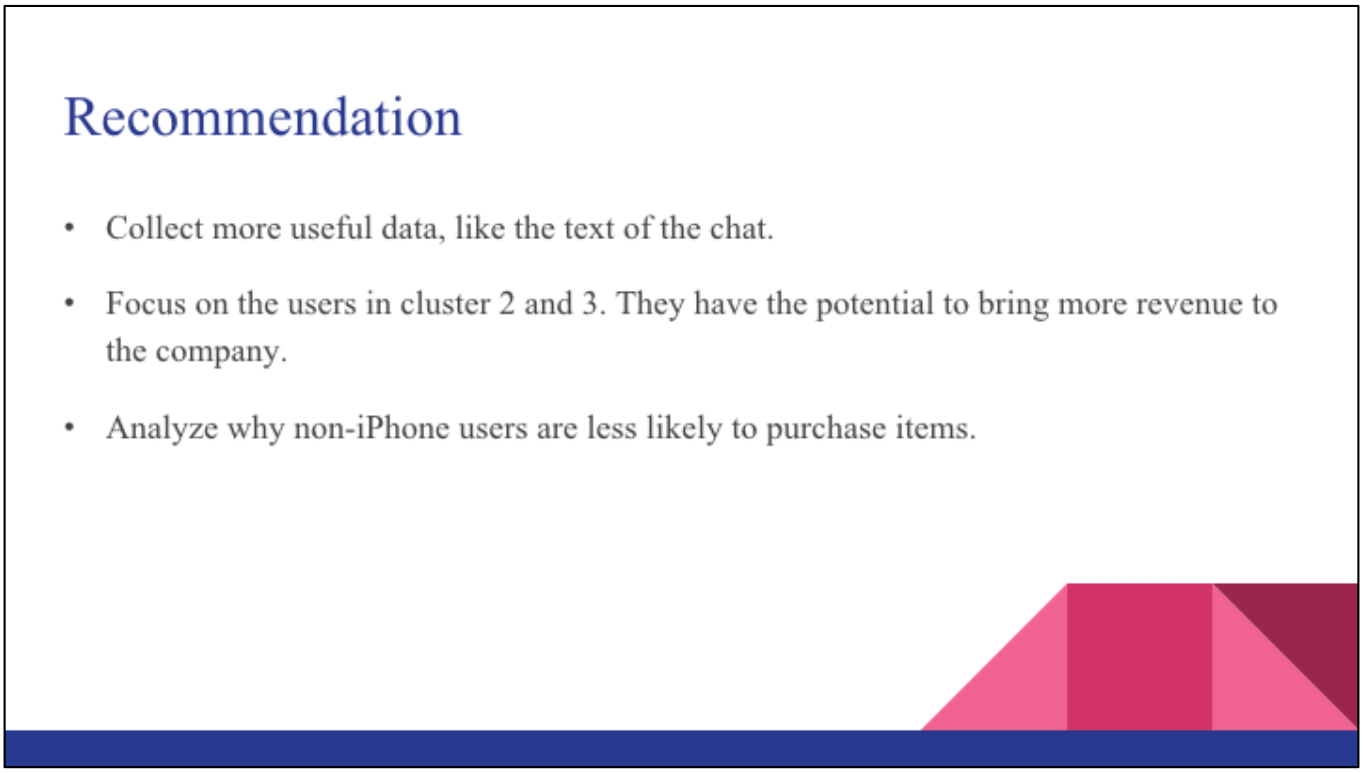
Cluster 3 is different from the others in that the users play the game very often.

Cluster 4 is different from the others in that the users is not active in playing the games, buying items and clicking the ads.

Clustering is an unsupervised learning approach. It allows us gain more insight from data without requiring the data to be labeled. From clustering, we can find the pattern of different types of users, and possibly develop the policy accordingly.



Chat data allows us to do a lot of things. We can dive deeper and analyze the relationship between users or teams. And we can even provide friend recommendation or team recommendation based on the chat data. We can do advertisement targeting. For example, if a user click on an ads, the user's friends may be likely to click as well. We can also award the most active users in the teams.



Based on our previous observation, we make the following recommendations. Basically, we recommend the company to focus on the users that can generate more revenue for the company, and collect more data so that more insight can be gained.