1. Goals & Results

1.1 Goals

User Churn Prediction

✓ Validate and clean up log-based raw datasets

✓ Data imputation, feature engineering, remove bots and outliers, downsample if possible

✓ Exploratory data analysis

✓ Build user churn prediction model based on user behavior

✓ Perform analysis, draw conclusion and give advices based on model results

1.2 Results

User Churn Prediction

✓ Pipeline and grid search were used to search for the model with best hyperparameters.

✓ All the models give great results. AUC of test is between 90.5% to 92.2%.

✓ According to feature importance, the recency of play is important. In order to prevent user from churning, to boost the user activity is critical. And according to recency figure in section 3.1.4, day 7 seems to be the equilibrium point which determines whether a user is more likely to churn or not.

✓ Moreover, device type is also an important feature. Please refer to section 3.2.1, Android phone has higher churn rate than iphone or other phones.

✓ In addition, churn rate in users who enjoy different types of songs is also significantly different.

Therefore, here are advices to decrease user churn.

a. Send reminders to user who does not have activities in 7 days. It can be mobile pop-up notifications.

b. Check the user experience in Andriod phone and make some improvements.

c. Improve the recommender system. And more researches should be done in analysis user behaviors who enjoy song-type-1.

1.3 Conclusion and suggestions

User churn prediction

- ✓ In order to decrease user churn, a better recommender system should be developed.
- ✓ When a user stop using our music box, a gentle remider should be send.
- ✓ User experience on Android phone may need to be improved.

2. Definitions

2.1 Churn prediction features and label window definition

All features is generated in the feature window and label as churn or not churn user is generated in label window. Churn will be generated as label 1 and not churn as 0.

✓ Feature window:
- 2017-03-30 ~ 2017-04-28 days: 30

✓ Label window:
- 2017-04-29 ~ 2017-05-12 days: 14

2.2 Population definition

Useful population in this project is considered as active users during feature windows. And their actives in the following two weeks as label window are used to generate labels.

✓ Incluse: all active users during feature time window

✓ Excluse: inactive user during feature time window and bots/outliers

2.3 Churn prediction features generation

✓ User active frequency

Defined as number of events over time windows.

✓ User active recency

Defined as number of days bewteen last event from snapshot date.

✓ User average play length percentage

Defined as average play time percentage of songs for each users.

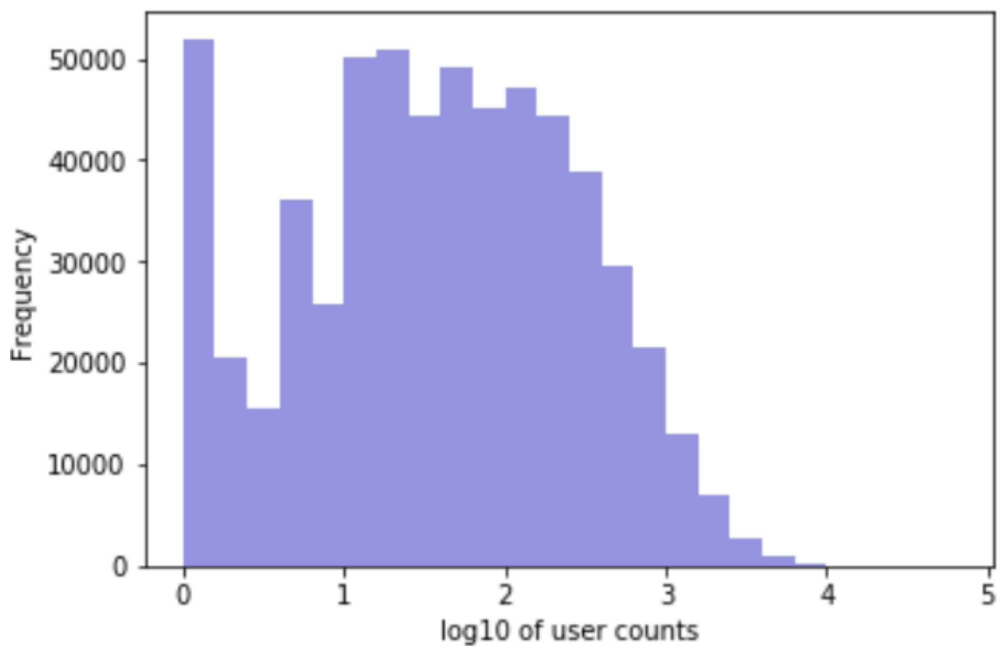
✓ User devic

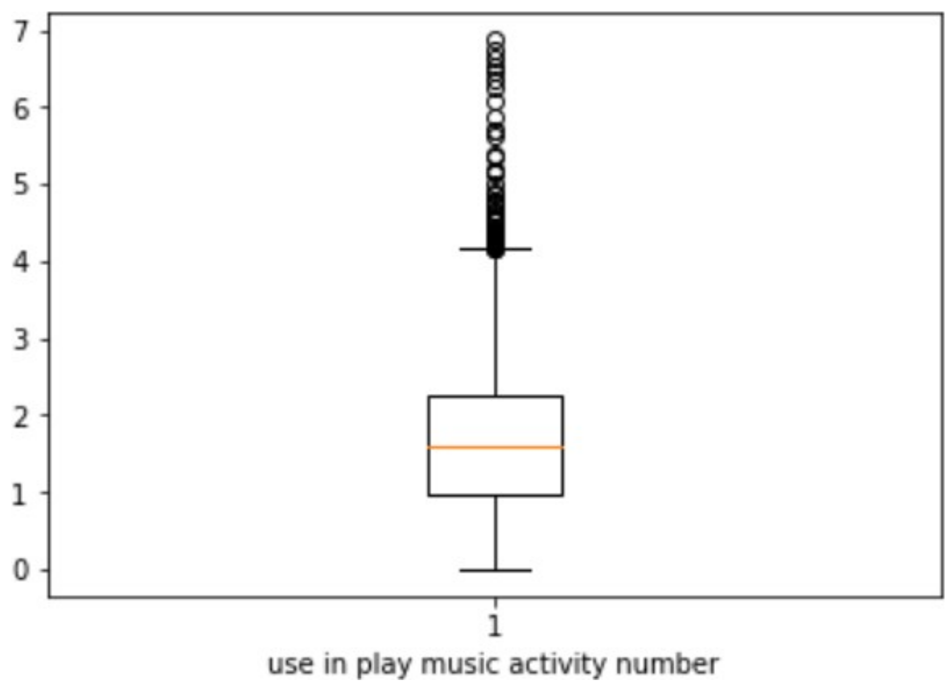
3. Exploratory data analysis

3.1 User Activity Analysis

From 2017-03-30 to 2017-05-12, the total number of users on file is 59,4720, who have activity of play/search/download music on the music box platform. However, the dataset also contains bot user, which should be removed. In addition, in order to handle large dataset calculation, downsample was applied. And the total number of users after down sample is 11,8247.

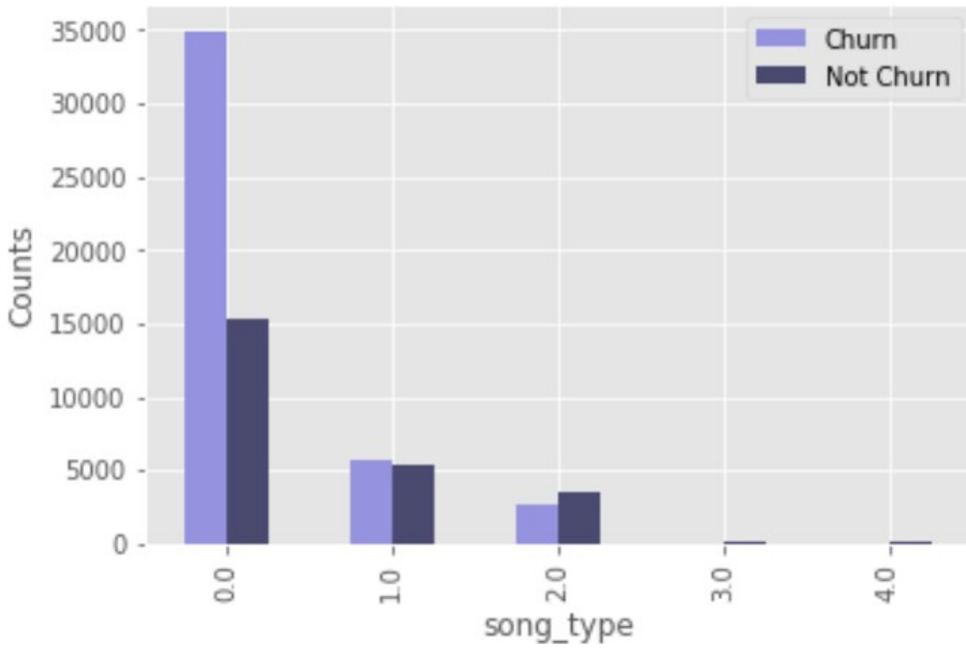
3.1.1 Distribution of number of user play during the window time





It can be easily shown that most of our users play roughly hundreds of songs during these periods of time. In average, 2~10 songs per day. Let's asume average 4 minutes per song. It means that on average, most of our customers spent less than an hour per day on our music box platform. And the lack of time spends on music box platform may lead to customer churn eventurelly.

3.1.2 Song type and churn



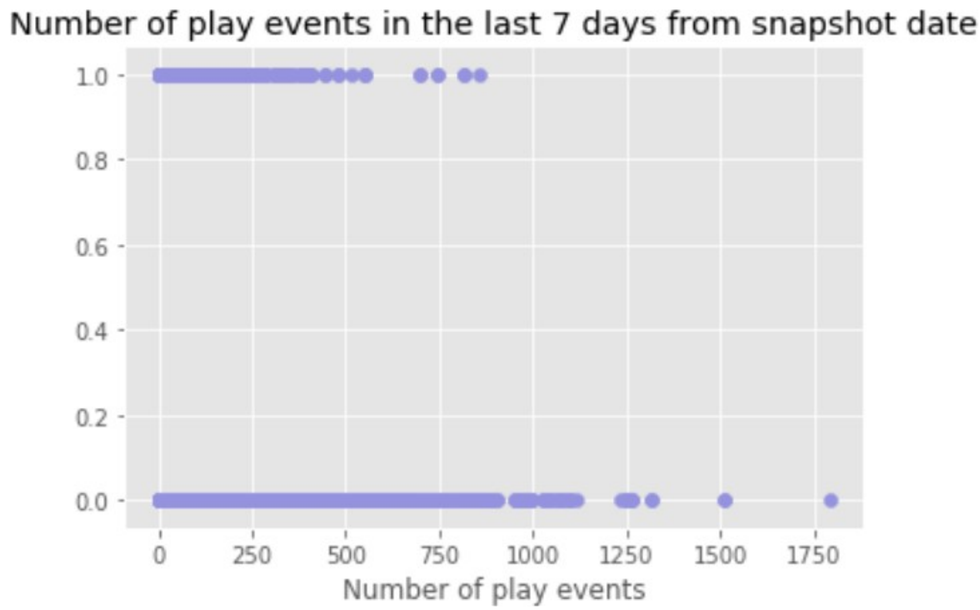
This figure interestinly shows the loyalty of our user by the type of song they like. Our music platform classifies the songs into 4 different types. Type 1 is the most popular. However, the churned users are much more than the unchurned users. There are several reasons may cause it.

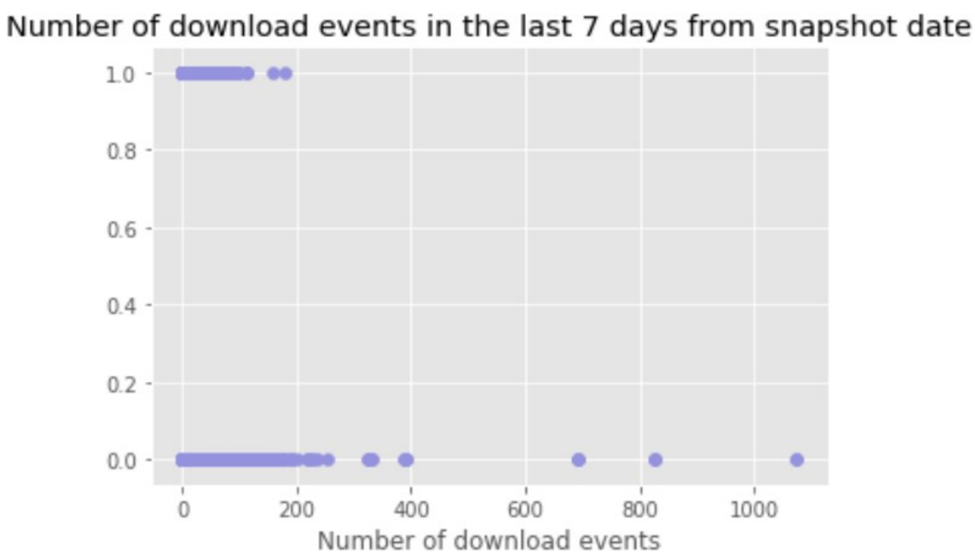
- a. Lack of type 1 songs or low quality of type 1 songs to users.
- b. Recommender system cannot provide what customers like

But on the other hand, users who enjoy type 1,2,3,4 are less likely to churn. It may be due to our music box can provide more of these types of songs than other platforms.

3.1.3 Event frequencies and churn

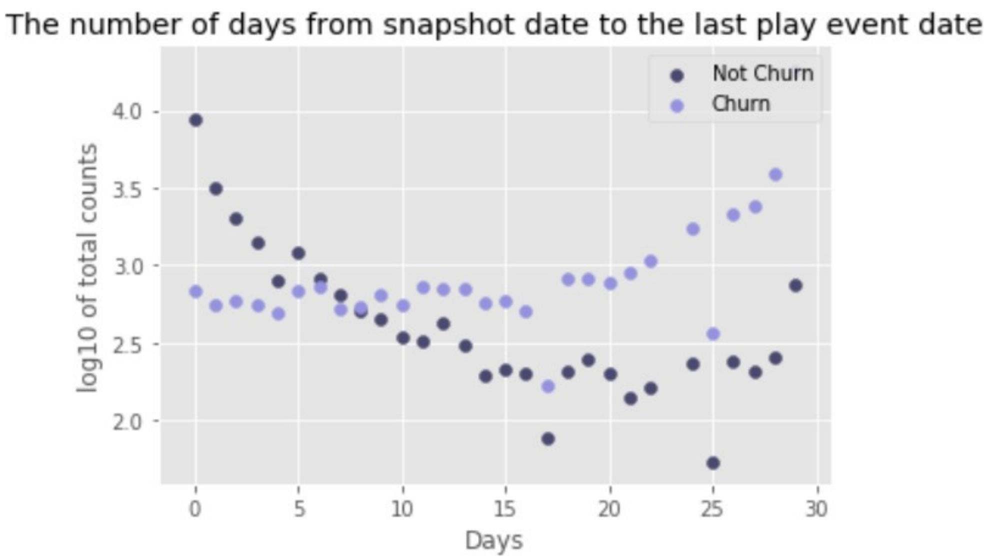
Event frequencies are another great indicators to user churn. Users who have more activities during a window time are less likely to churn and vice versa. Below shows the churn vs activities number in a 7 days window.





3.1.4 Event recencies and churn

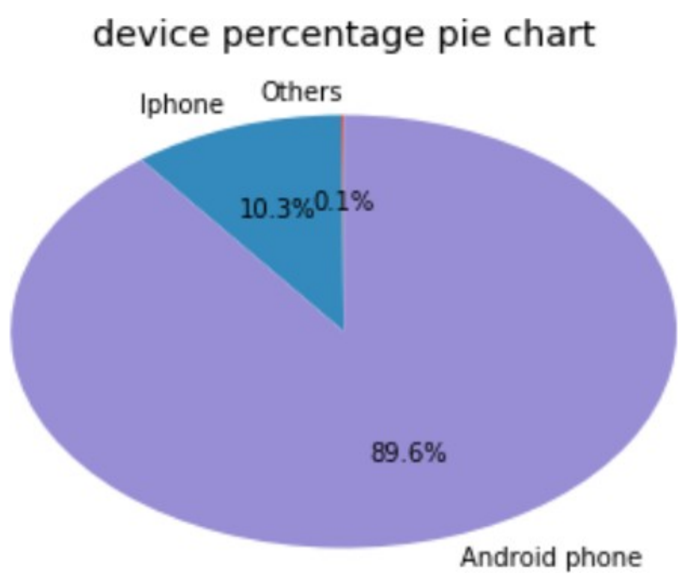
Event recencies represent the number of days from snapshot date to the last event occurred date. From the figure below, it is obvious that 7 day is a critical number. If users' last play activities date is whithin 7 days to the snapshot date, the not-churn probability of those users are higher.



3.2 User Profile Analysis

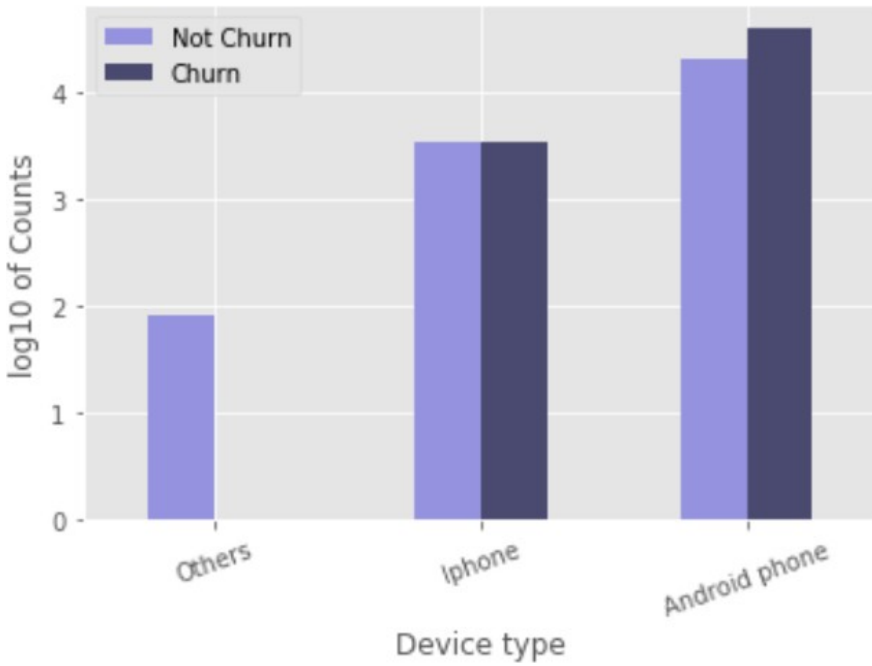
3.2.1 device type analysis

On our platform, nearly 90% of devices are with android system and 10% of IOS system. The rest includingmac and windows phone is about 0.1%.



However, according to the figure below, there are more user churns on android phone than the others. It may be due to the following reasons.

- a. The user experiences on android phone are not as good as some other music player platform.
- b. And there are more platform options on andriod phone system.


Not Churn
Churn
log10 of Counts
0
1
2
3
4
Others
Iphone
Android phone
Device type