Capstone 2

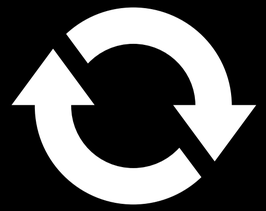
Applying Machine Learning to Predict Churn for Music Streaming Service KKBox

Bradley Mensah
Springboard Nov 2nd 2020 Cohort

The Problem

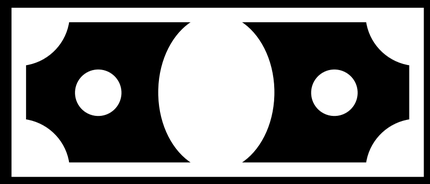
- The majority of KKBox subscriptions last only 30 days
- No single factor affects a subscribers decision to cancel or renew their membership.
- In order to be able to accurately forecast revenue and plan a budget, subscription-based services like KKBox must be able to predict how many subscribers will continue their memberships with reasonable accuracy.

What opportunities exist for KKBox to report a positive percent change in revenue by the end of the current quarter through:



subscriber retention

attracting new subscribers



increasing prices

Data

citation:

KKBOX Group. (2017, September). WSDM - KKBox's Churn Prediction Challenge, Version 2. Retrieved March 3, 2021 from https://www.kaggle.com/c/kkbox-churn-prediction-challenge/overview/evaluation.

- Four .csv files: train_v2.csv, members_v3.csv, transactions_v2.csv, and user_logs_v2.csv
  - Data selected from subscribers whose memberships are set to expire in March 2017
  - train_v2.csv: This dataset has our target value, is_churn, and a unique identifier for each customer, msno.
  - members_v3.csv, transactions_v2.csv, and user_logs_v2.csv are feature columns

members_v3.csv

msno: unique identifier

city: user's city

bd: user's age

gender: user's gender

registered_via: registration method

registration_init_time: date the user registered, format %Y%m%d

transactions_v2.csv

This dataset is a record of each customer's transactions.

msno: user id

payment_method_id: payment method

payment_plan_days: length of membership plan in days

plan_list_price: in New Taiwan Dollar (NTD)

actual_amount_paid: in New Taiwan Dollar (NTD)

is_auto_renew: whether or not the user signed up to have their membership renew automatically

transaction_date: format %Y%m%d

membership_expire_date: format %Y%m%d

is_cancel: whether or not the user canceled the membership in this transaction

user_logs_v2.csv

This dataset is a log of a user's activity.

msno: user id

date: format %Y%m%d

num_25: number of songs played less than 25% of the song length

num_50: number of songs played between 25% to 50% of the song length

num_75: number of songs played between 50% to 75% of of the song length

num_985: number of songs played between 75% to 98.5% of the song length

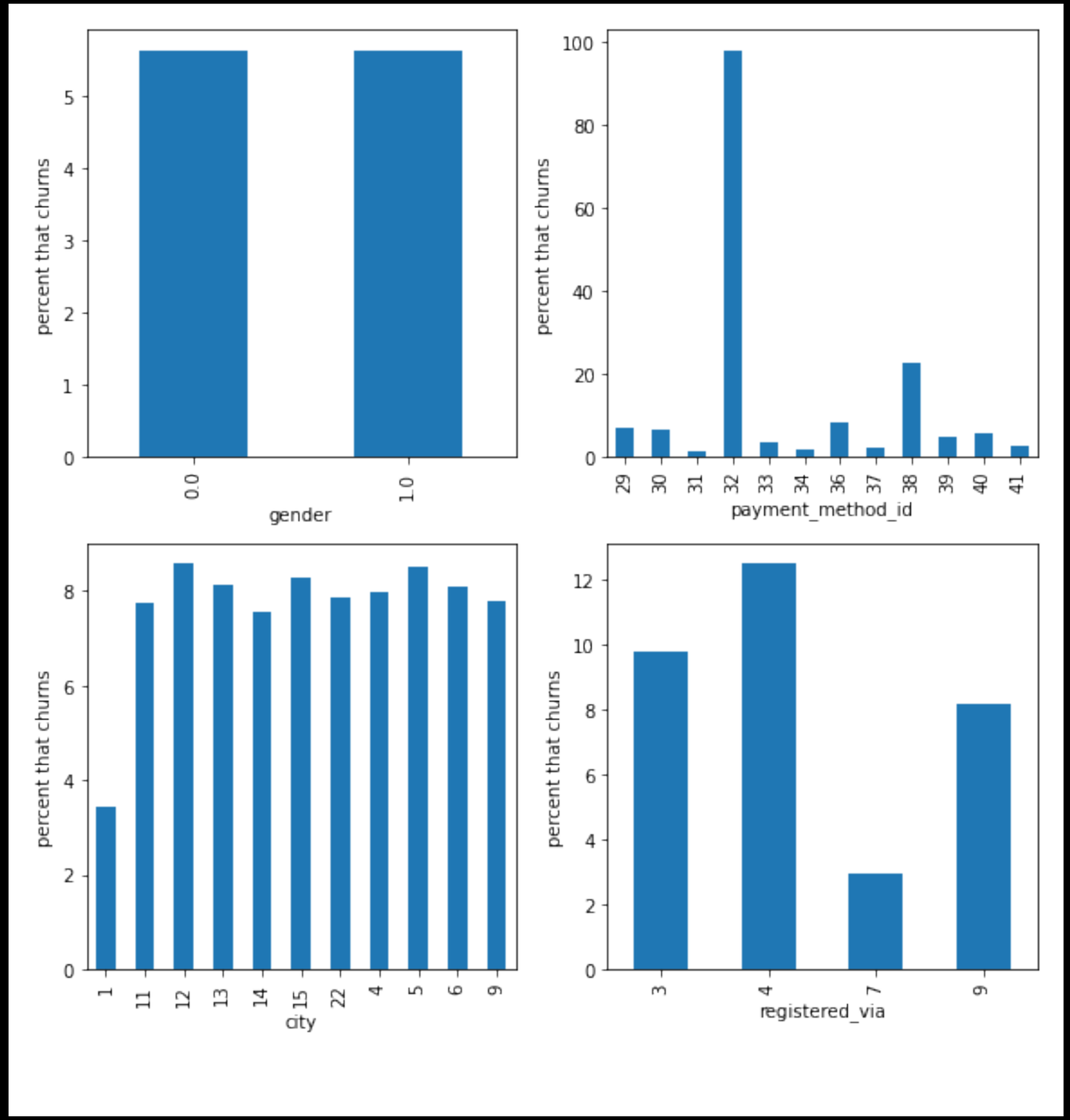
num_100: number of songs played over 98.5% of the song length

num_unq: number of unique songs played

total_secs: total seconds played

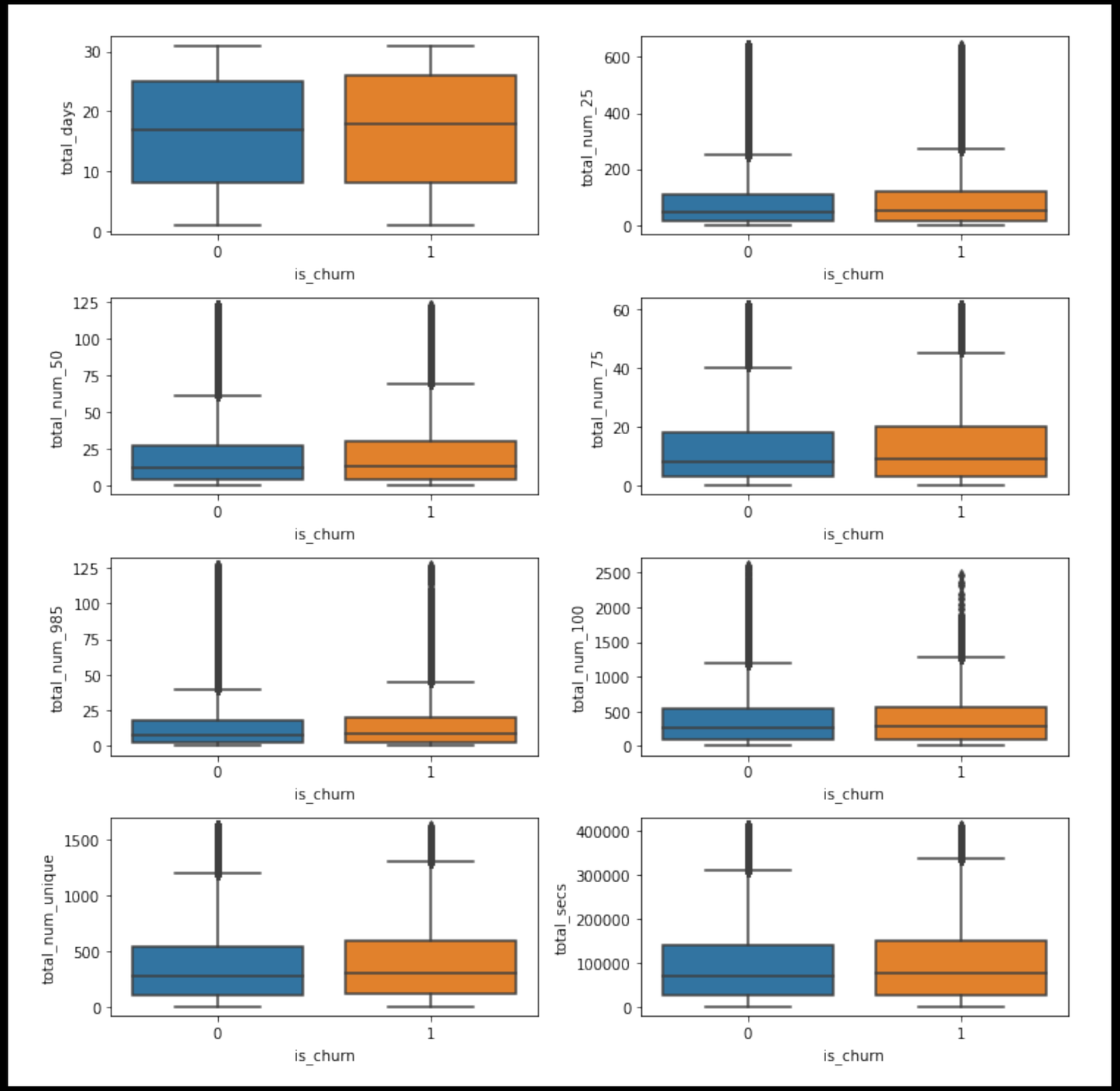
Visualizing Churn

Churn for categorical features



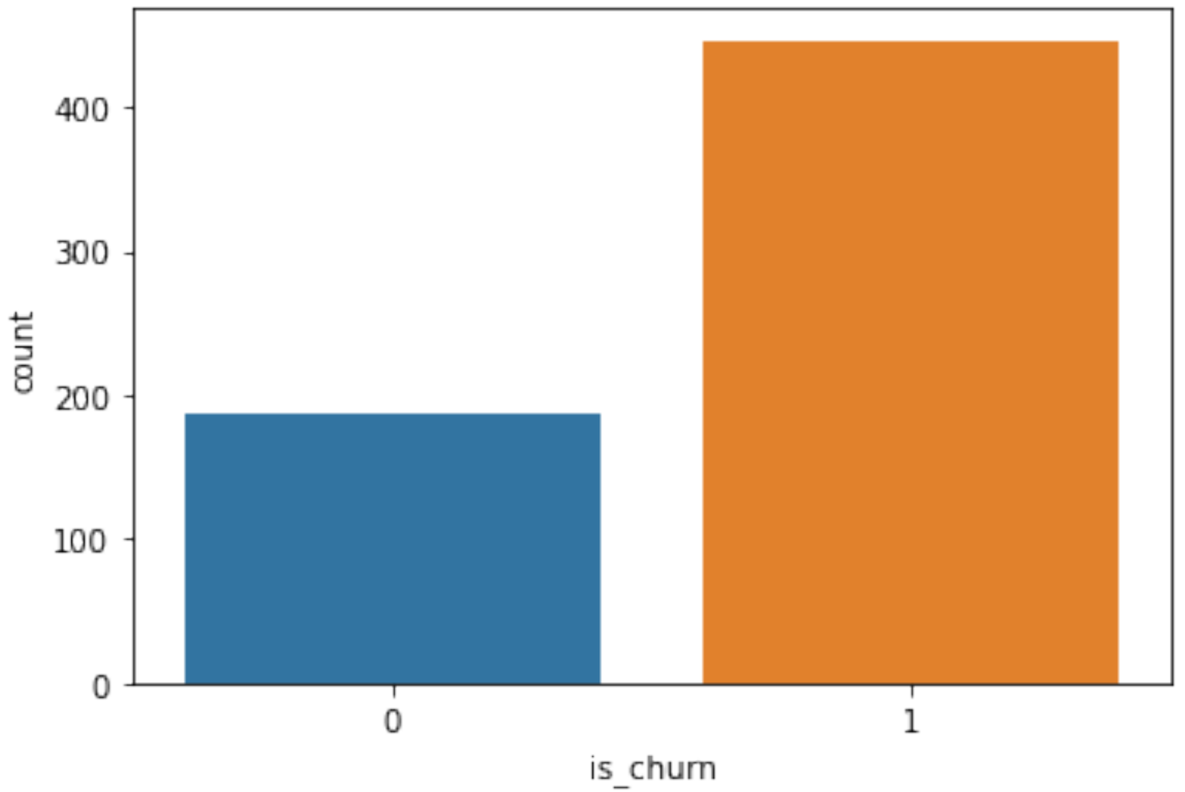
Visualizing Churn

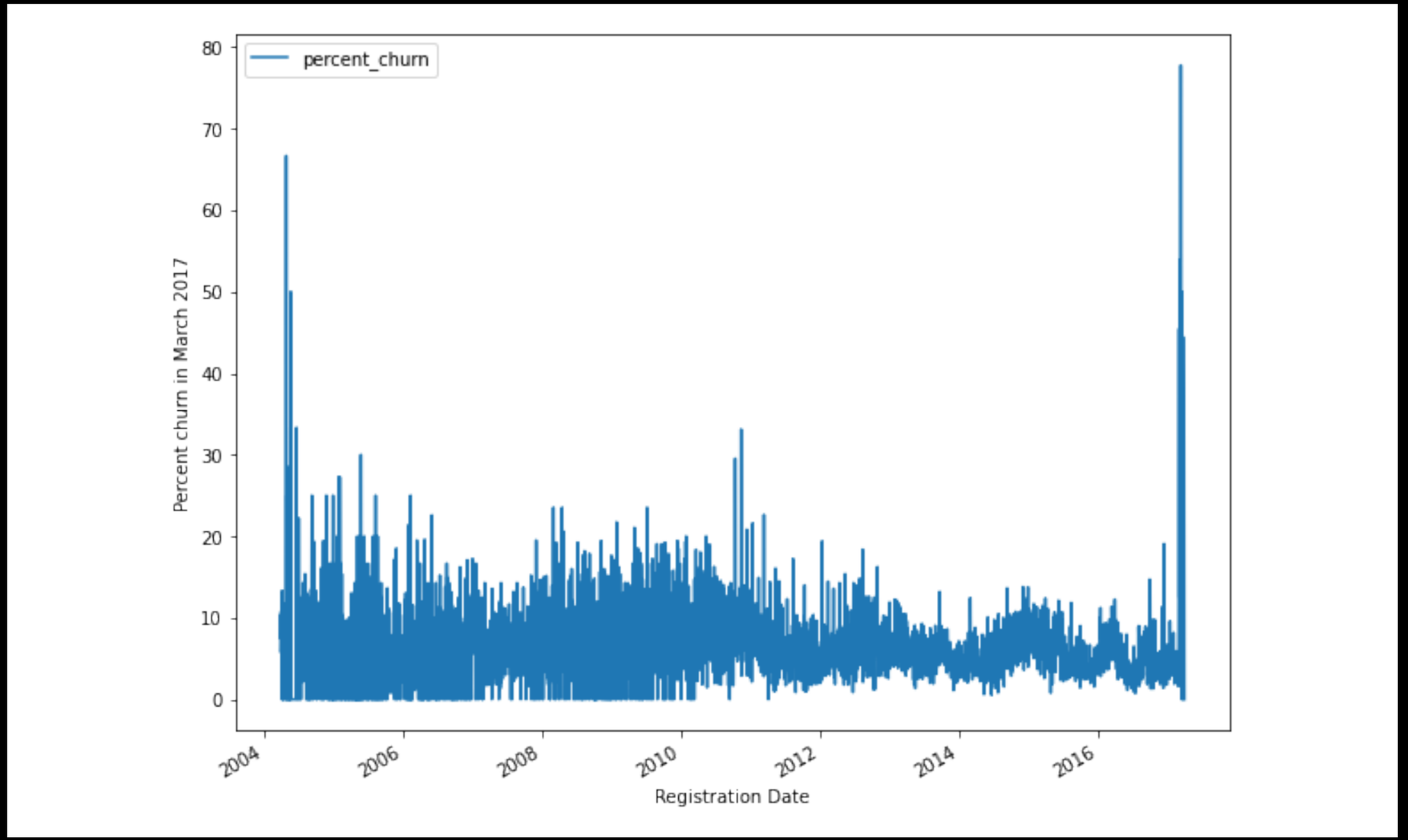
Are users who listen to more music less likely to churn?



Visualizing Churn

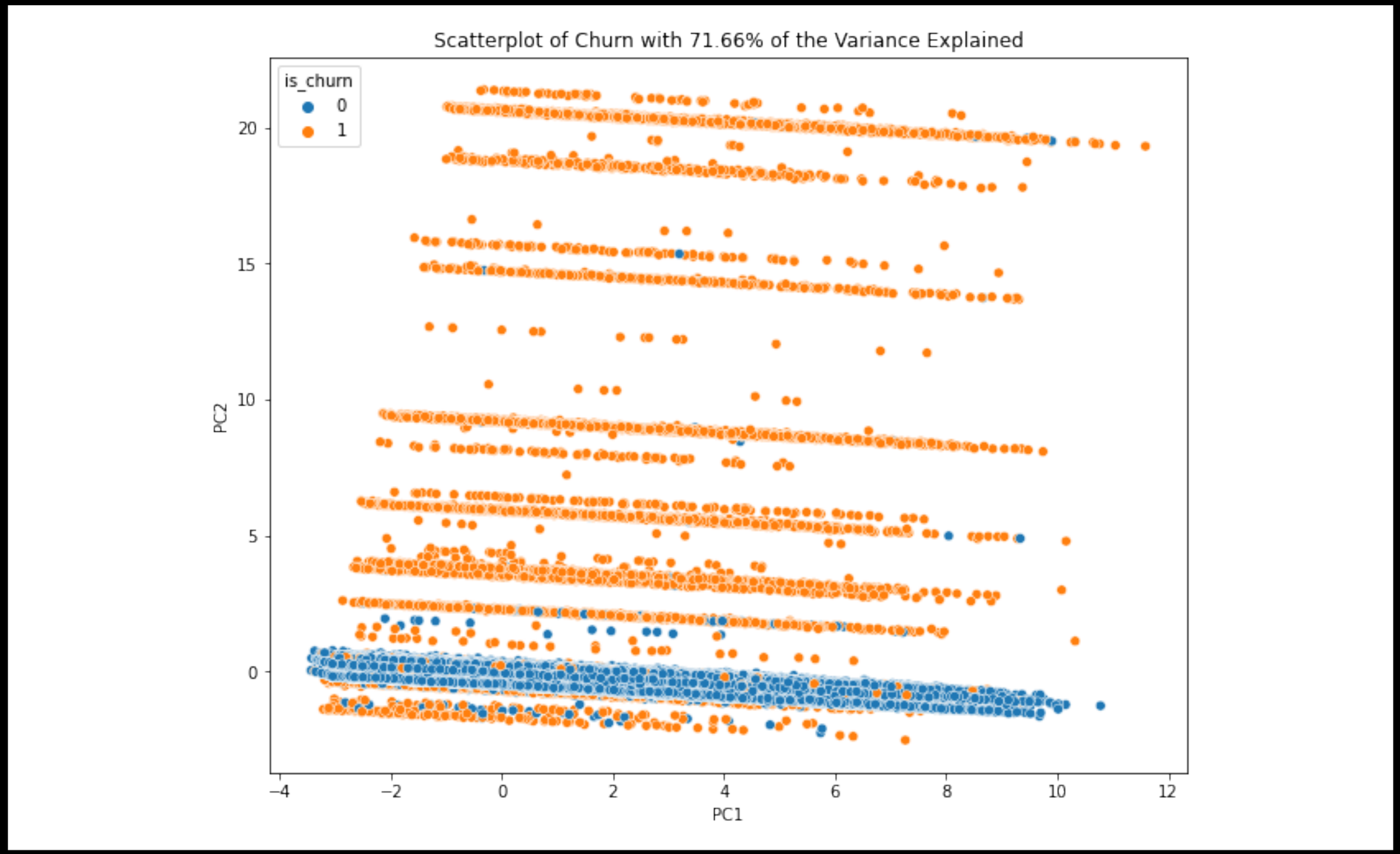
Does getting a discount discourage users from churning?





Visualizing Churn

Does the amount of time a subscriber has been registered for KKBox affect churn?



Visualizing Churn

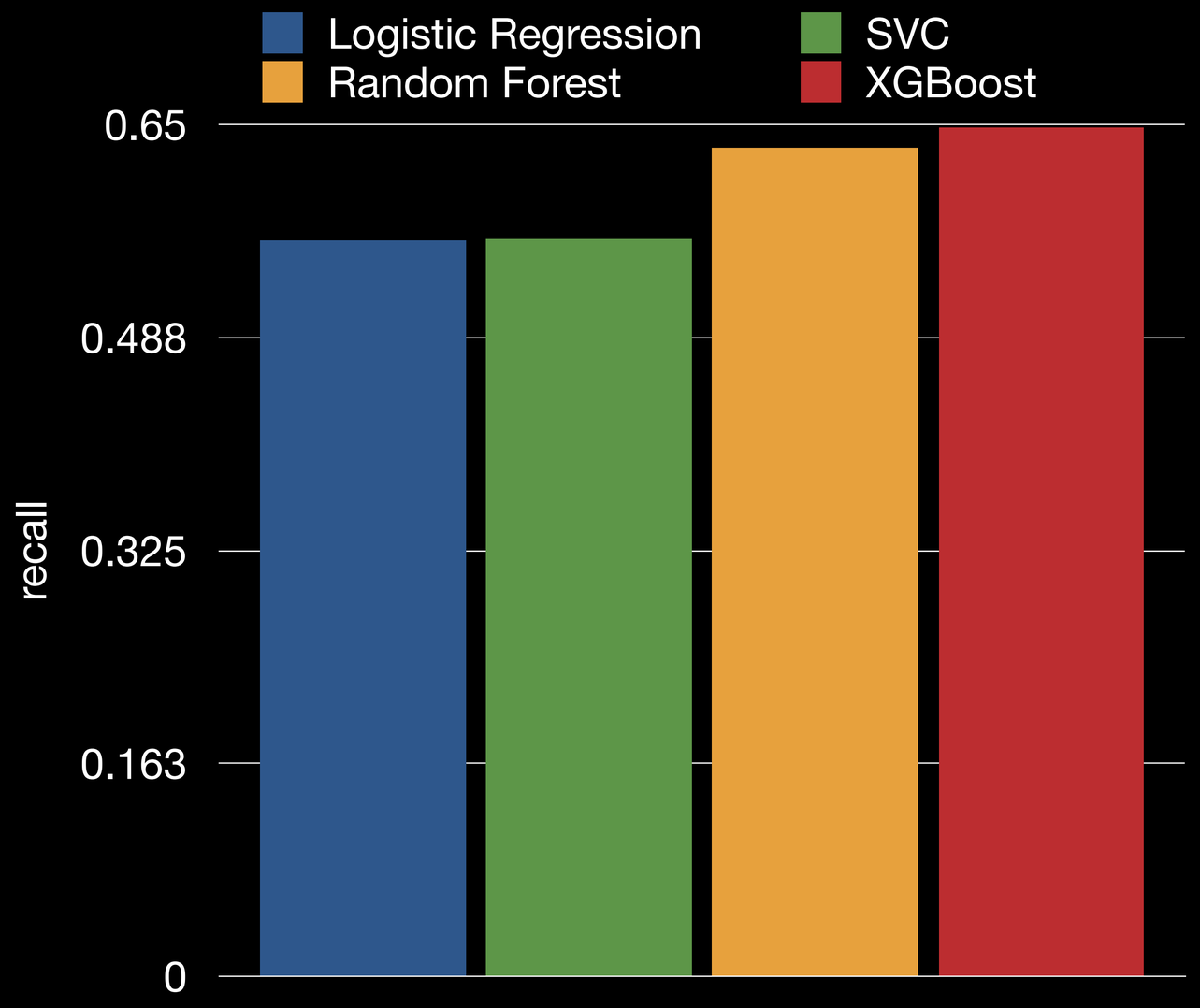
Exploring the distribution of churn using PCA components

Modeling Overview

- Since our data is labelled, this is a **supervised learning** problem.
- **Binary classification** will determine whether a subscriber will renew (0) or churn (1)
- **False negatives** are more important than false positives: missing a subscriber at risk of churn is worse than mislabeling a subscriber who is not at risk of churn
    - Therefore **recall** will be used as the primary metric to determining model efficacy.

Comparing Models

XGBoost is the best performing model on recall.



Hyperparameter Tuning

Random search and Bayesian optimization were used to to tune hyperparameters.

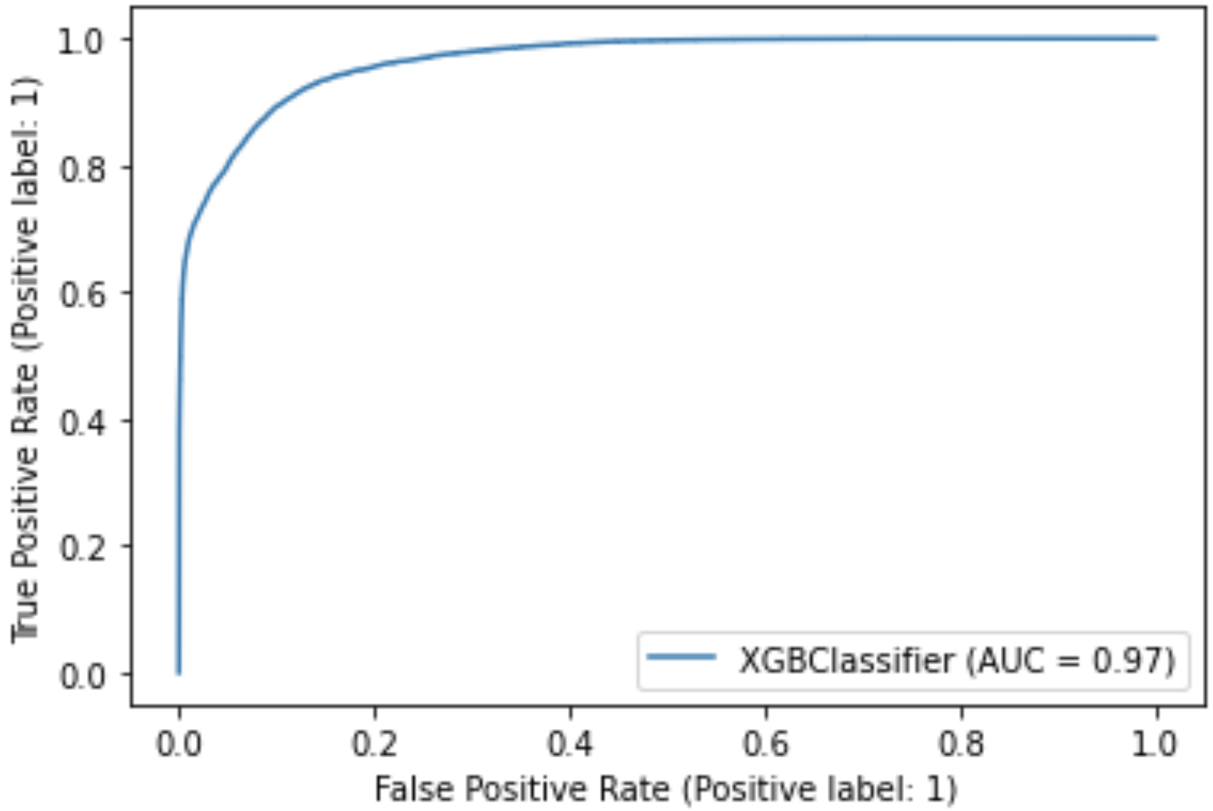
The following hyperparameter ranges were set:

- eta (learning_rate): between 0.3 and 0.9
- max_depth: between 1 and 9
- min_child_weight: between 1 and 7
- colsample_bytree: between 0.1 and 0.8
- gamma: between 0.1 and 0.5

Hyperparameter Tuning

Random Search

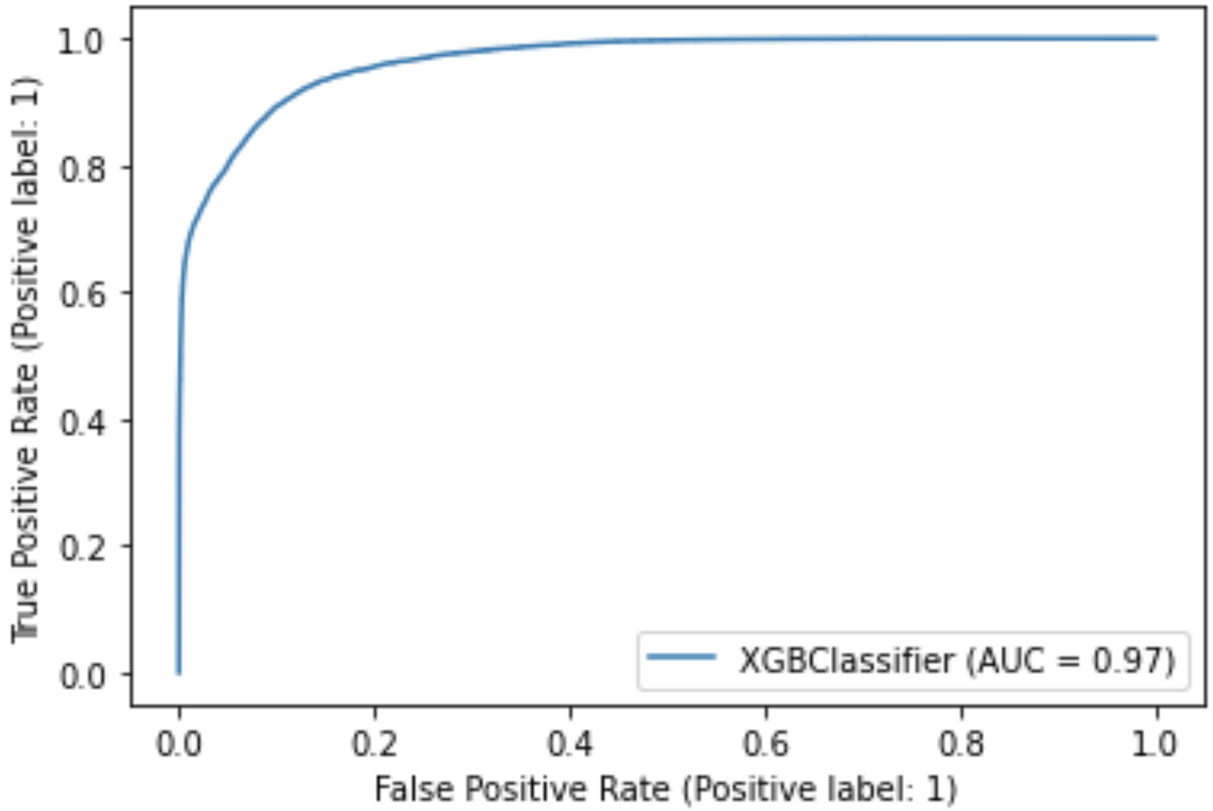
- eta = 0.6340410
- max_depth = 6
- min_child_weight = 6
- colsample_bytree = 0.6101575
- gamma = 0.4770061



Hyperparameter Tuning

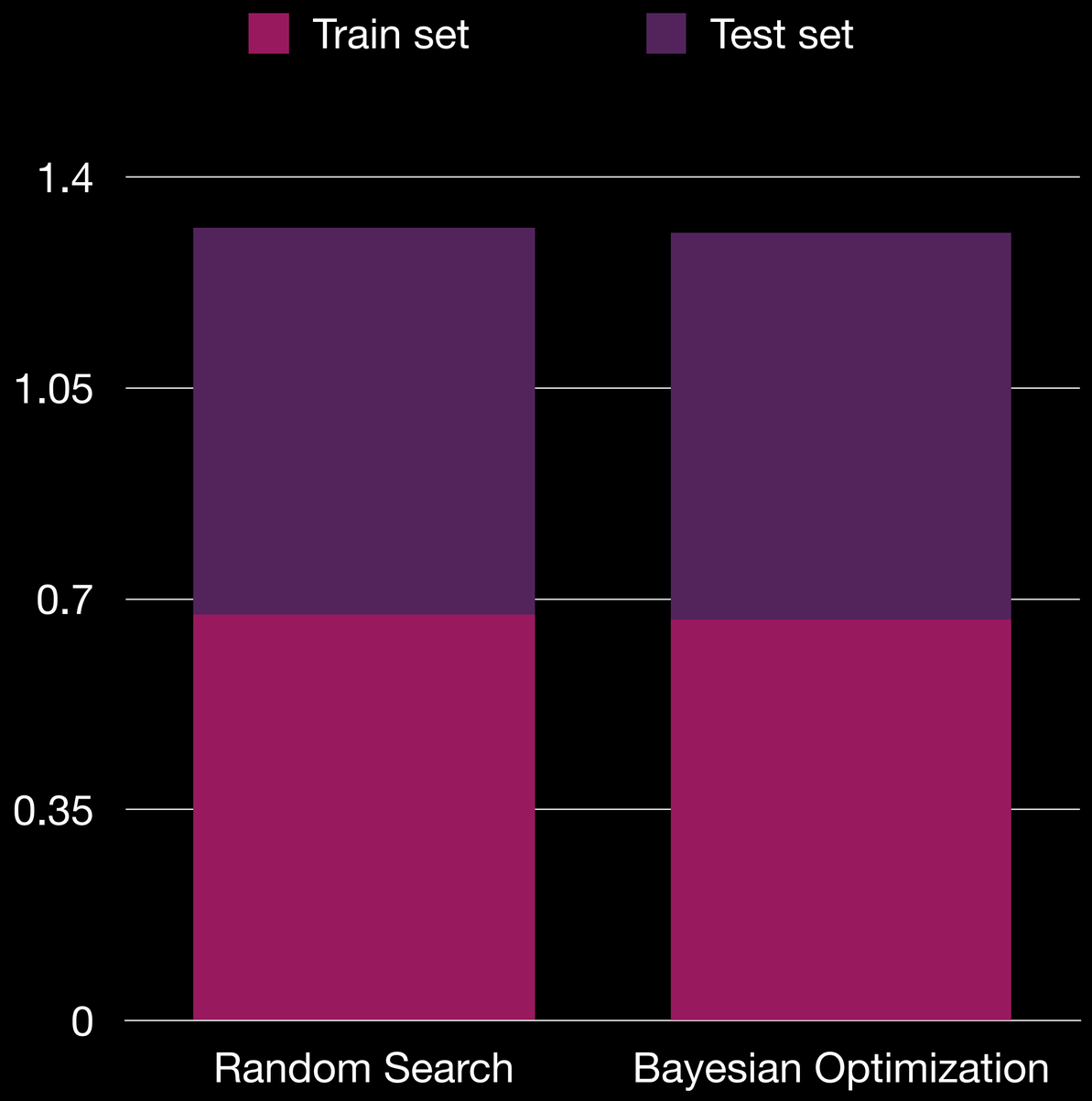
Bayesian Optimization

- eta = 0.60
- max_depth = 5
- min_child_weight = 4
- colsample_bytree = 0.45
- gamma = 0.30



Hyperparameter Tuning

Though disagreeing on the best hyperparameter values, both search methods produce a recall score of 0.642 on the test set.



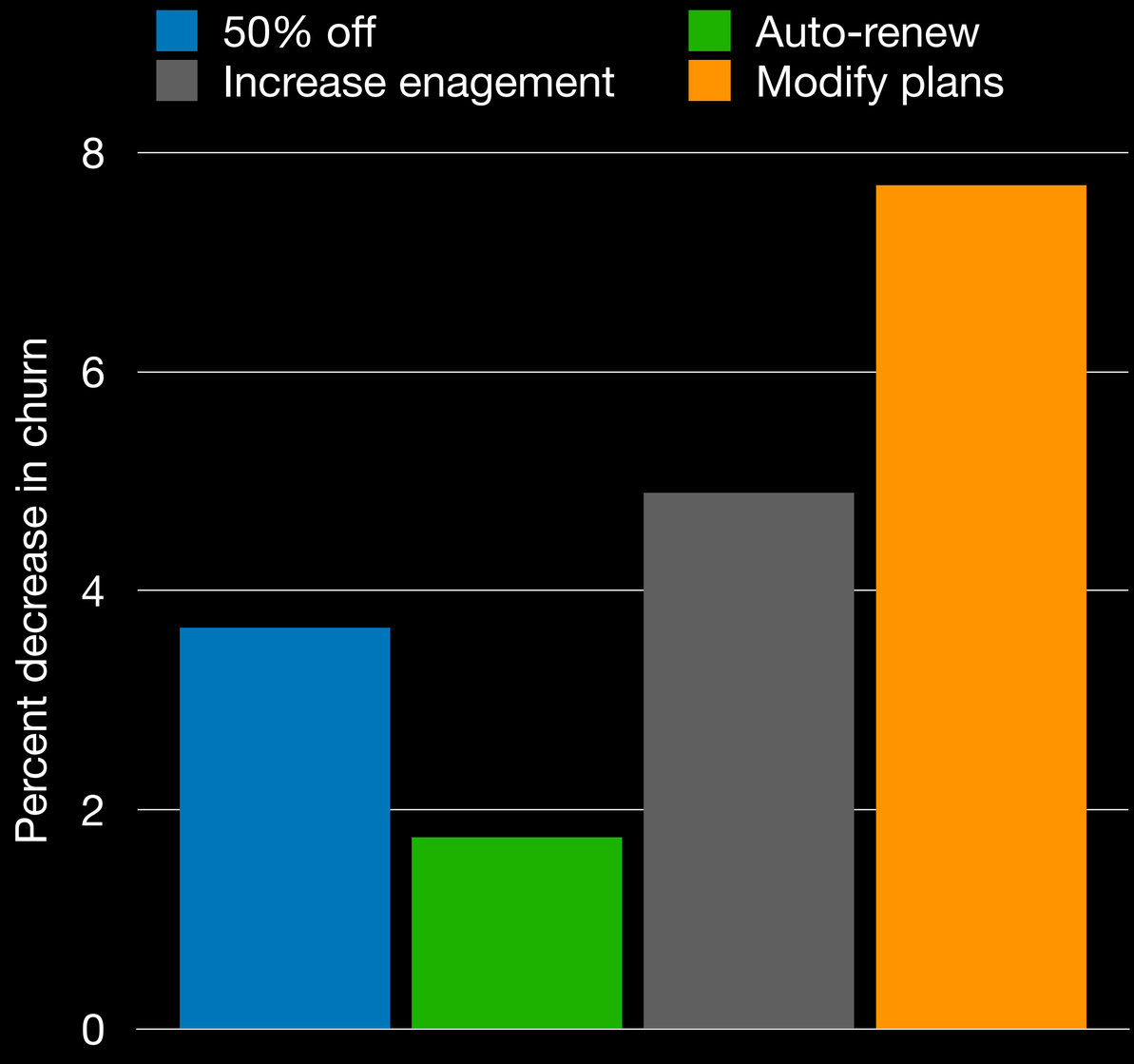
Modeling Scenarios

The following scenarios were modeled to discover what methods KKBox could employ to decrease churn:

- provide subscribers at risk of churn with a 50% off discount
- double user engagement with the app
- switch subscribers on plans lasting more than 30 days to monthly plans
- convince subscribers at risk of churn to sign up for auto-renew

Modeling Scenarios

Bar graph showing the predicted decrease in churn in each scenario



18.28% ▼

predicted decrease in churn with all four scenarios combined

0.56% ▲

predicted increase in subscription revenue with an 18.28% decrease in churn

Conclusions and Recommendations

- Payment method id 32 should no longer be used
- Targeted marketing to cities where KKBox has a smaller presence could help reduce churn.
- Registration method 7 should be used to sign up new subscribers.
- To reduce churn, KKBox should implement any or all of the following methods:
  - convince subscribers at risk of churn to sign up for auto-renew
  - restructure any subscriptions longer than 30 days to be monthly
  - provide subscribers at risk of churn with a 50% off discount for one month
  - invest in efforts to increase user time spent listening to music on the app

Further Analysis

Attracting New Subscribers

- There has been a downward trend in new subscribers since November 2015.
- More data analysis should be done to determine the cause of this downward trend and predict what options KKBox has to increase the number of new users added.

Further Analysis

Increasing Prices

- The model used in this analysis predicts that doubling prices would only increase churn by 4.25%. However, this figure is doubtful. The data this model was trained on is likely ill-suited to predict if increasing prices would increase churn.
- Further analysis using competing businesses' prices and features should be done in order to determine if KKBox can increase prices without increasing churn.