Capstone 2

Applying Machine Learning to Predict Churn for Music Streaming Service KKBox

Bradley Mensah
Springboard Nov 2nd 2020 Cohort

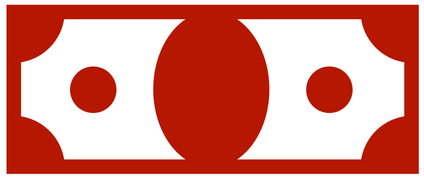
The Problem

- 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

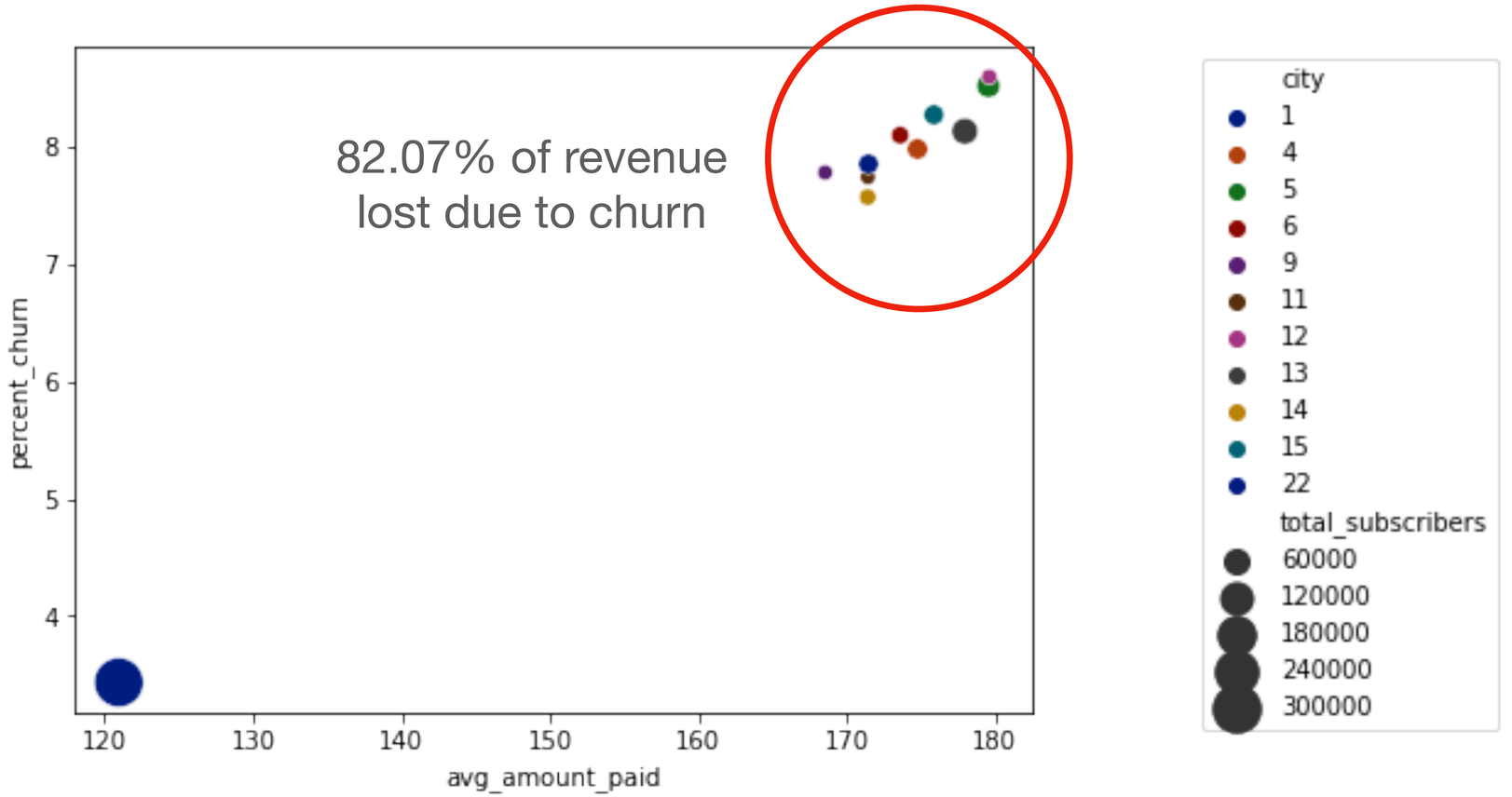
Data

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 hold feature columns

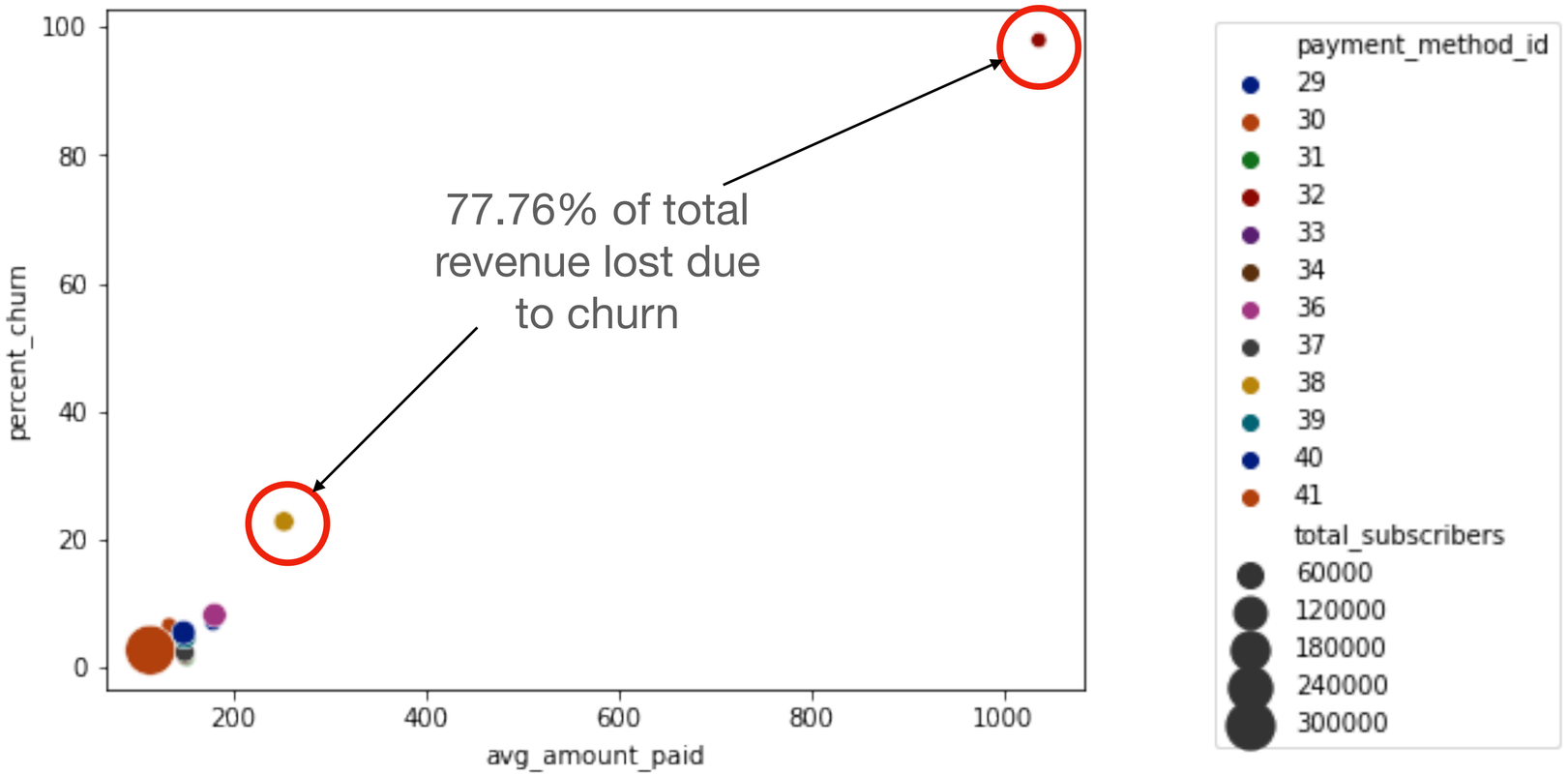
Visualizing Churn by City and Average Amount Paid



Insights:

KKBox could increase revenue by focusing on decreasing churn in cities where the service is less popular.

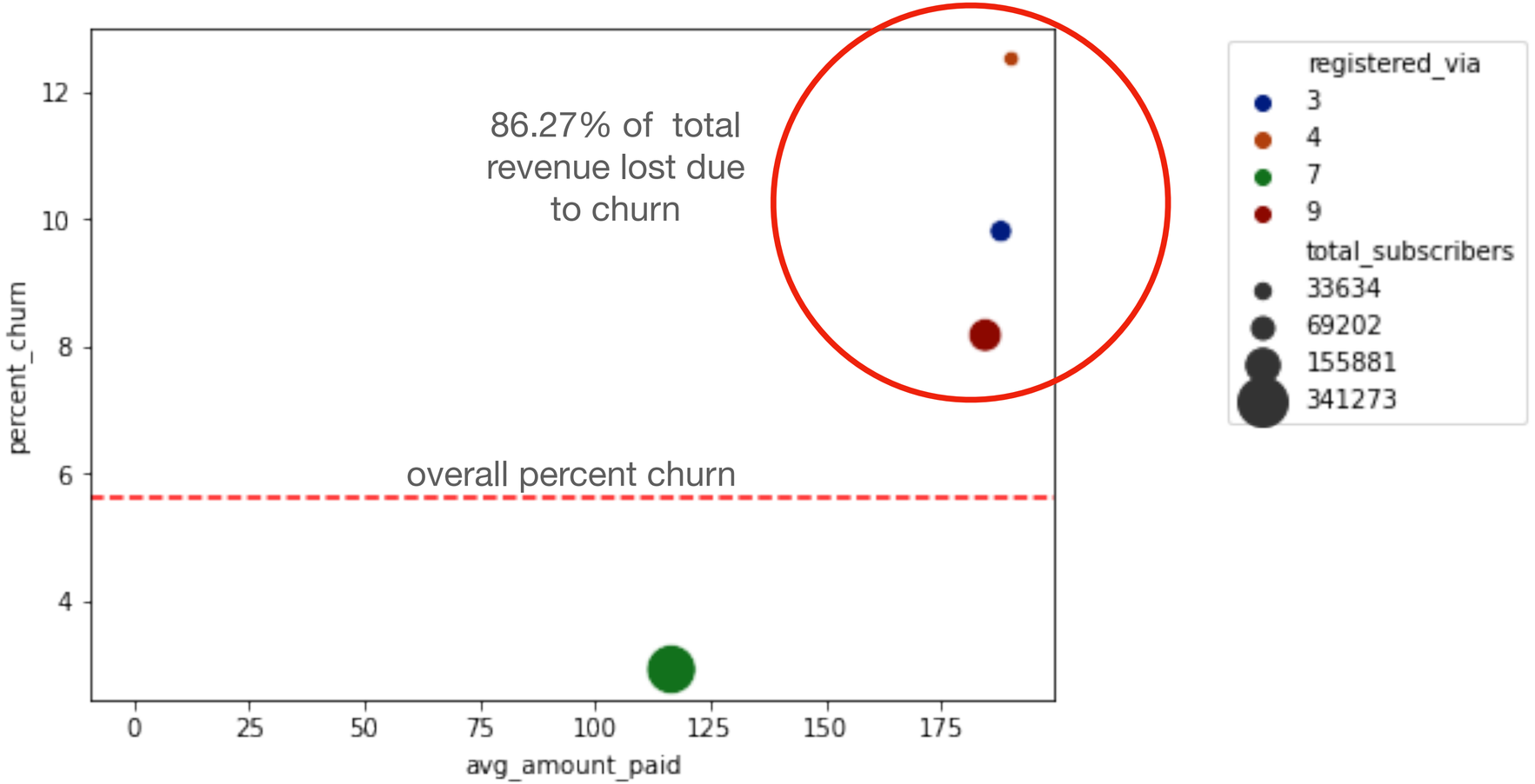
Visualizing Churn by Payment Method and Average Amount Paid



Insights:

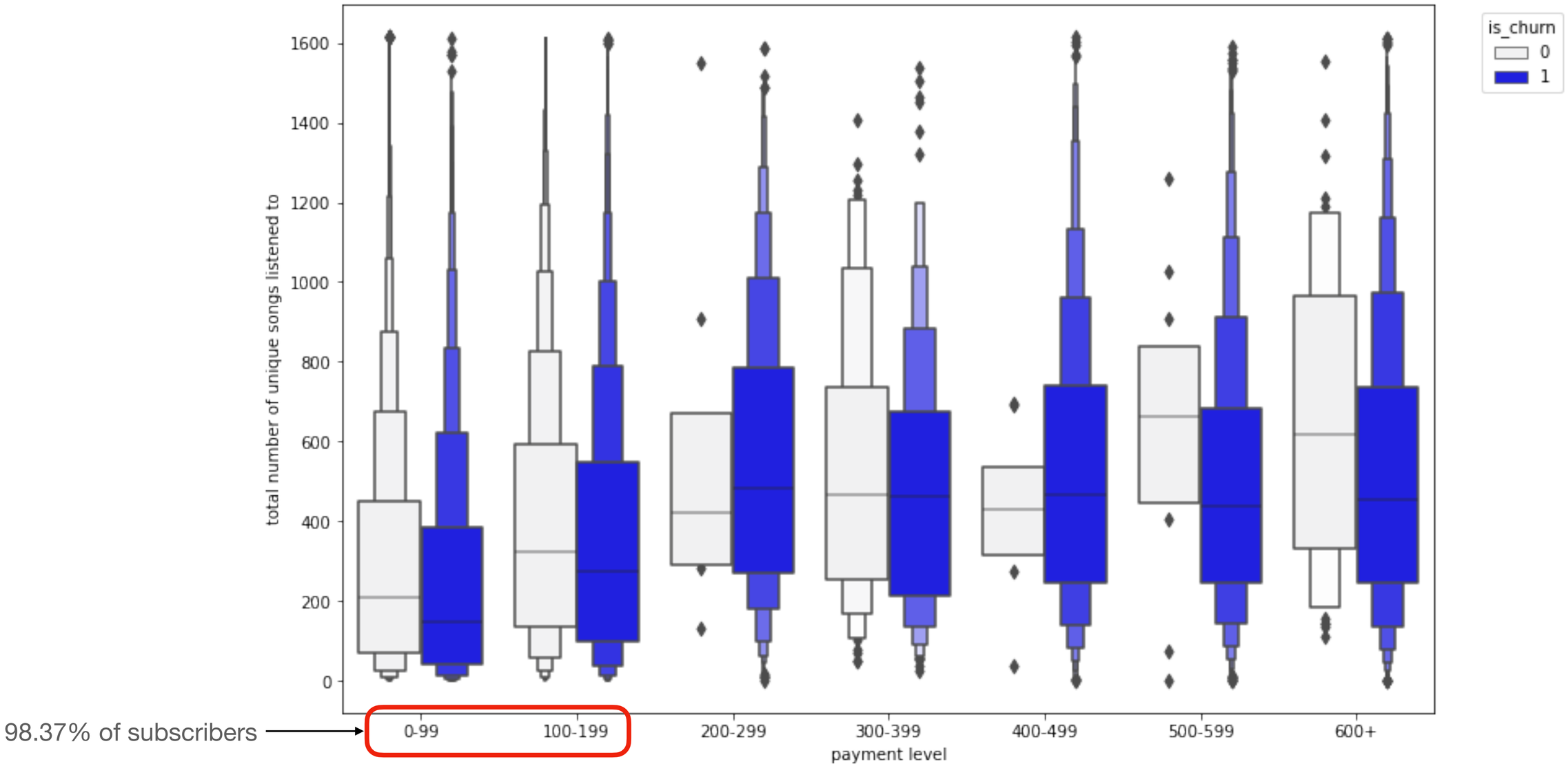
Subscribers who used payment methods 38 and 32 churned at unusual levels in March 2017. Features of these payment methods could be contributing to churn.

Visualizing Churn by Registration Method and Average Amount Paid



Insights:

Subscribers who registered via method 7 pay less for KKBox's service, but are also significantly less likely to churn. Registering new subscribers using method 7 could lead to less churn and therefore increase in revenue.



Visualizing Churn by Total Unique Songs Listened to and Amount Paid

Visualizing Churn by Total Listening Time and Amount Paid

Insights

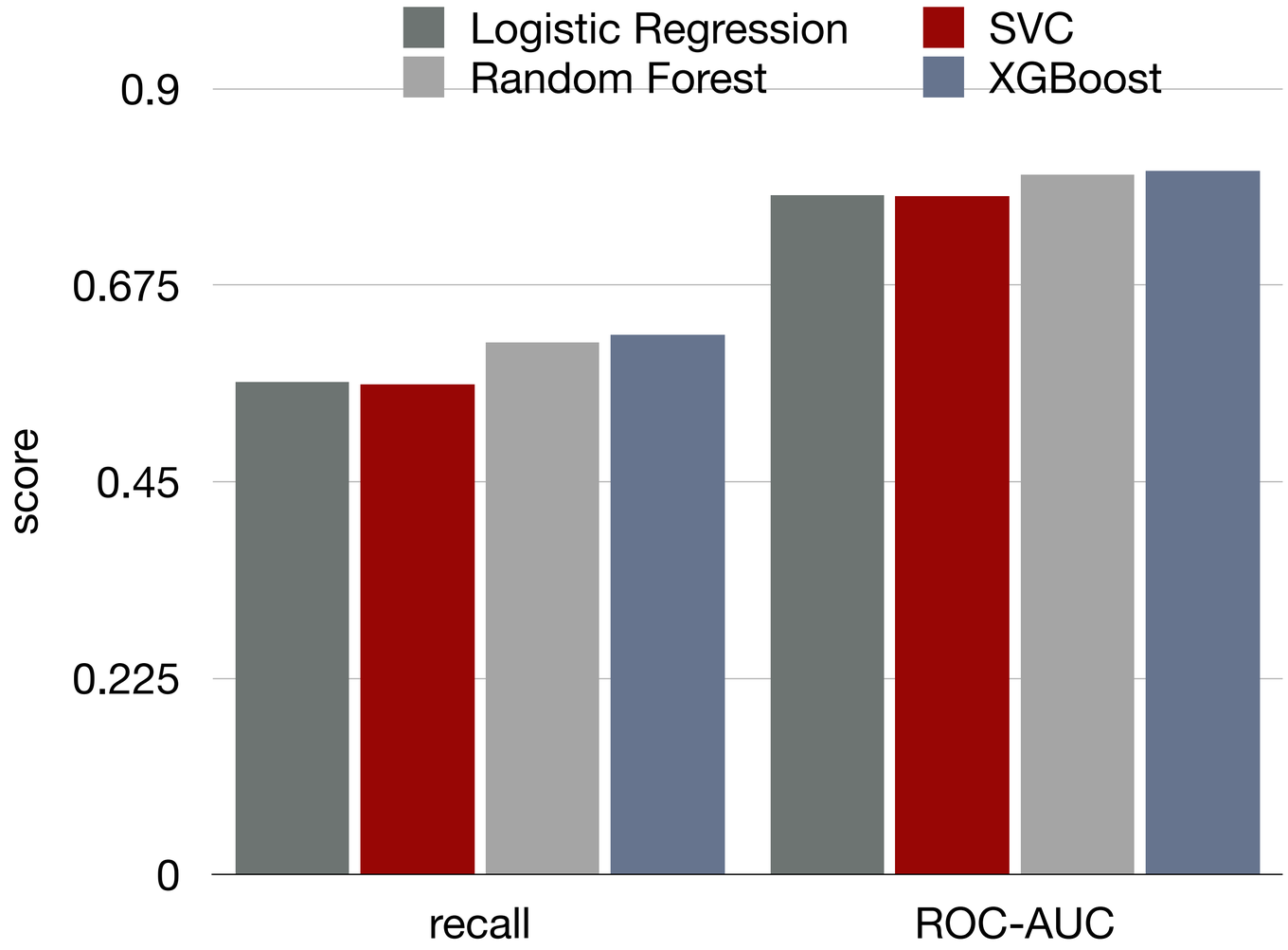
- Most subscribers (98.37%) listened to 200-300 unique songs and paid 99 to 199 NT$ in March 2017.
- The median amount of unique songs listened to tends to increase as the amount paid increases for subscribers who renewed in March 2017.
- In the majority of payment levels, subscribers who renewed in March 2017 listened to more unique songs than those who churned.
- Adding features to KKBox's app that encourage subscribers to listen to more music may decrease churn.

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)
- There is a **class imbalance**: only 5.63% of subscribers churn.
- **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

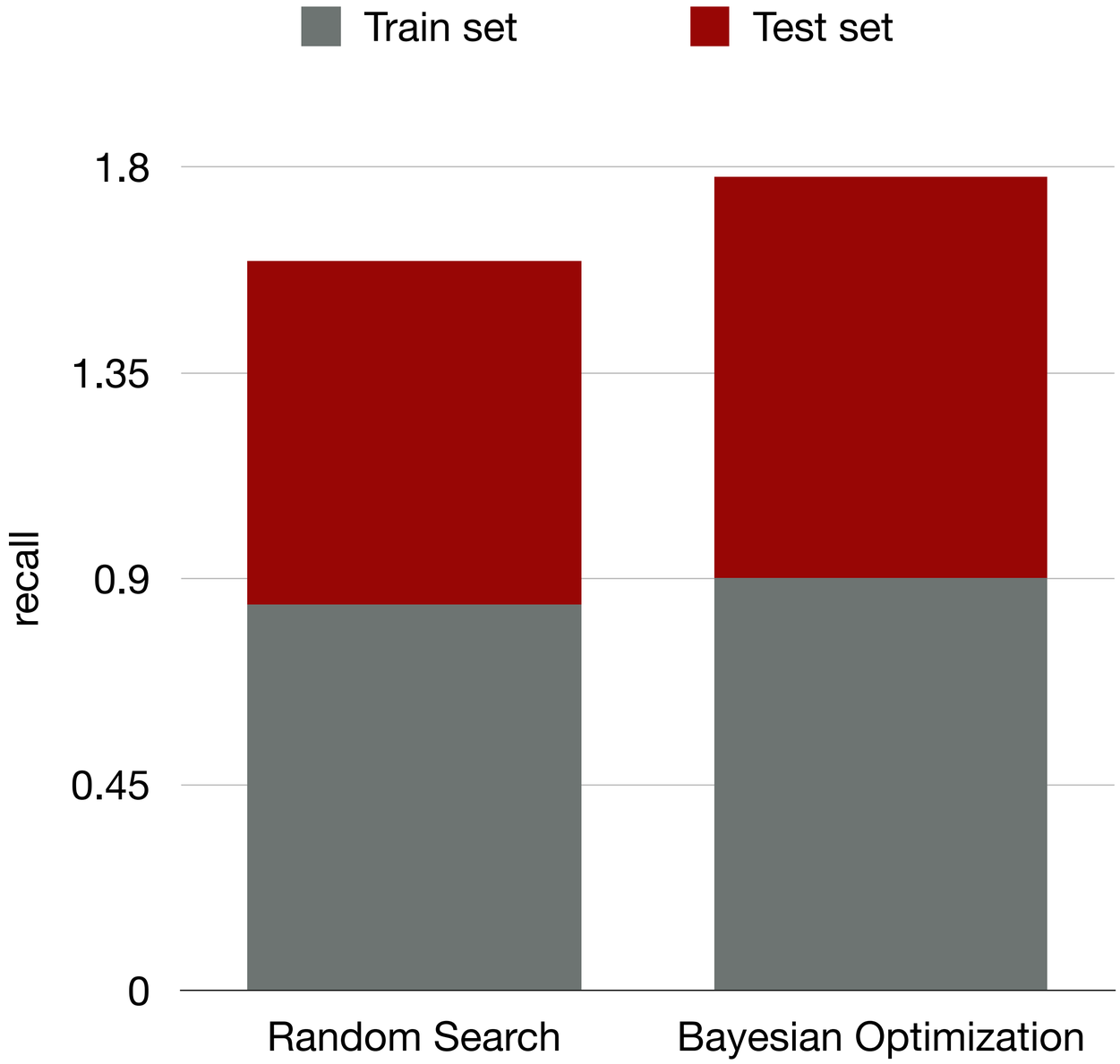
XGBoost performs the best on both recall and ROC-AUC metrics.



Hyperparameter Tuning

Comparing Random Search and Bayesian Optimization

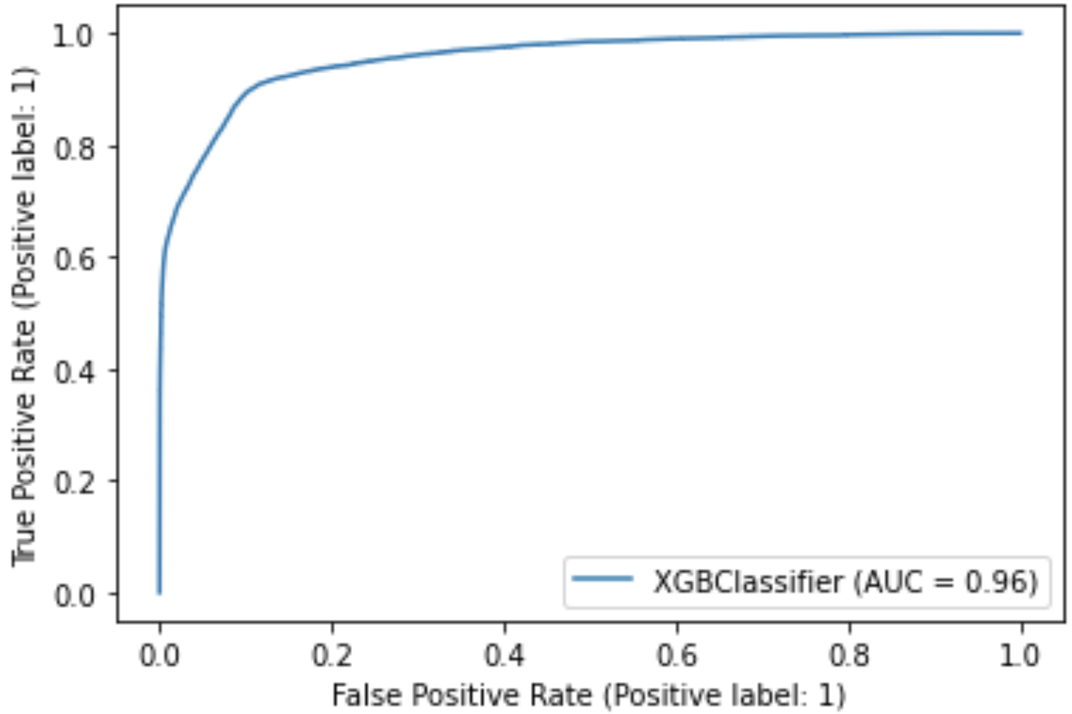
The model tuned with the hyper parameters chosen by bayesian optimization performs better on both the the train set and the test set.



Model Metrics

Hyperparameters and Performance of Final Model

- eta (learning_rate) : 0.600
- max_depth : 5
- min_child_weight : 4
- colsample_bytree : 0.450
- gamma : 0.300
- scale_pos_weight : 20.0



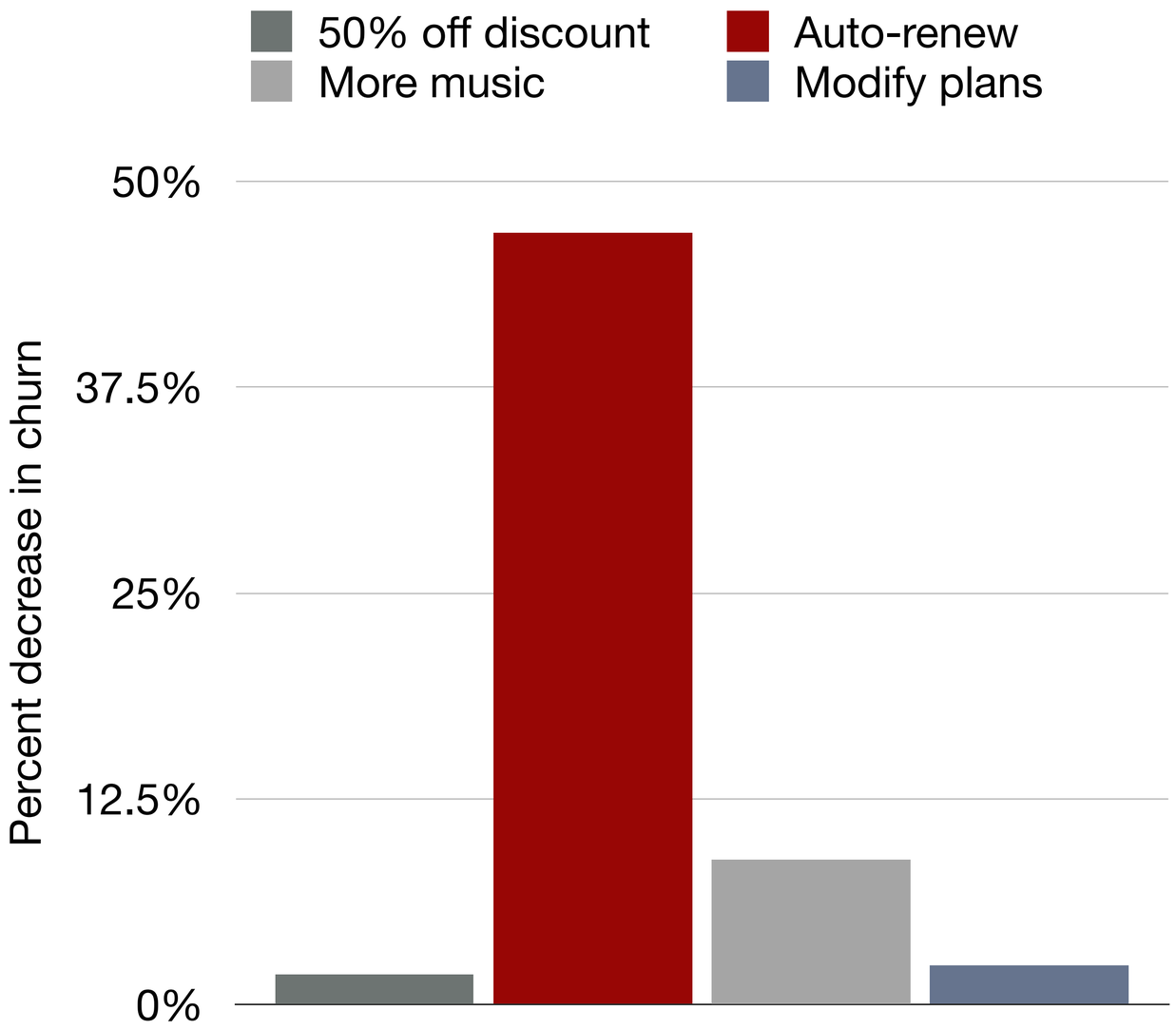
ROC-AUC: 0.96
recall: 0.877

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
  - increase time spent listening to music by 20%
  - transform plans lasting more than 30 days to monthly plans
  - encourage subscribers at risk of churn to sign up for auto-renew

Modeling Scenarios

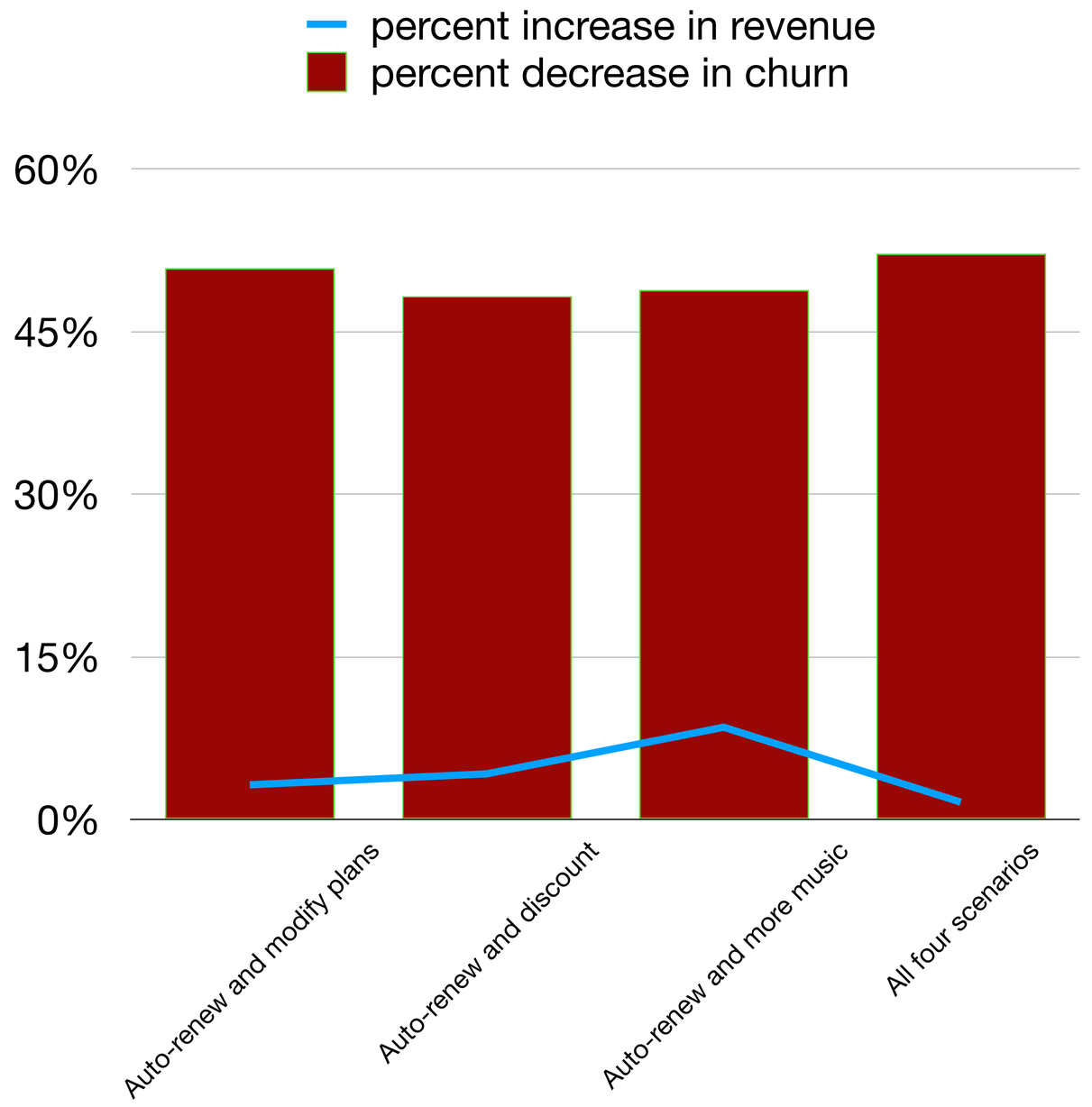
Signing up subscribers at risk of churn to auto-renew has an outsized effect on churn, reducing churn by 38.52% more than the next best scenario.



Combining Scenarios

Providing a discount and modifying plans decrease prices.

Therefore, even though implementing all four scenarios leads to the largest decrease in churn, it would lead to the smallest increase in revenue.



48.83% ▼

projected decrease in churn when subscribers at risk of churn sign up for auto-renew and listen to 20% more music

8.51% ▲

estimated increase in subscription revenue with an 48.83% decrease in churn

Conclusions and Recommendations

Problem Statement:

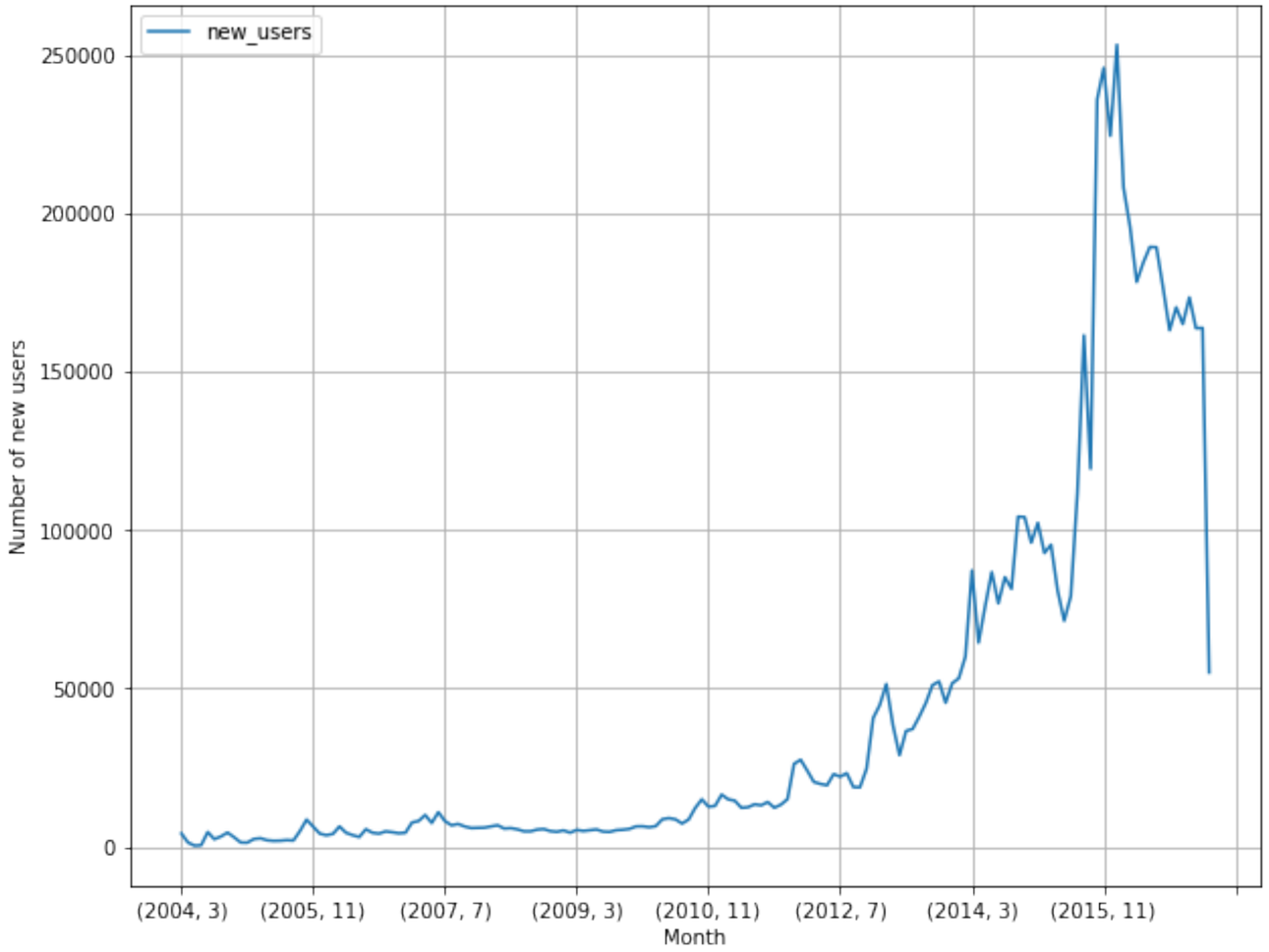
What opportunities exist for KKBox to report a positive percent change in revenue by the end of the current quarter through subscriber retention, attracting new subscribers, or increasing prices?

- Outreach to emerging markets in cities in which KKBox is less popular.
- Adaptations to operational procedures: examining potential challenges with payment and registration methods that make subscribers more likely to churn when using one method over another.
- Marketing campaigns targeting subscribers predicted to be at risk of churn which advertise the auto-renew feature and new music available on KKBox .

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 and Updating the Model

- The data used to train the model in this analysis is not suited to determine if raising prices would also lead to an increase in churn.
    - Further analysis using competing businesses' prices and features should be done in order to determine if KKBox can increase prices without increasing churn.
- The reasons why subscribers decide to churn are multi-faceted. The model should be continuously updated with new data as recommended solutions are implemented.