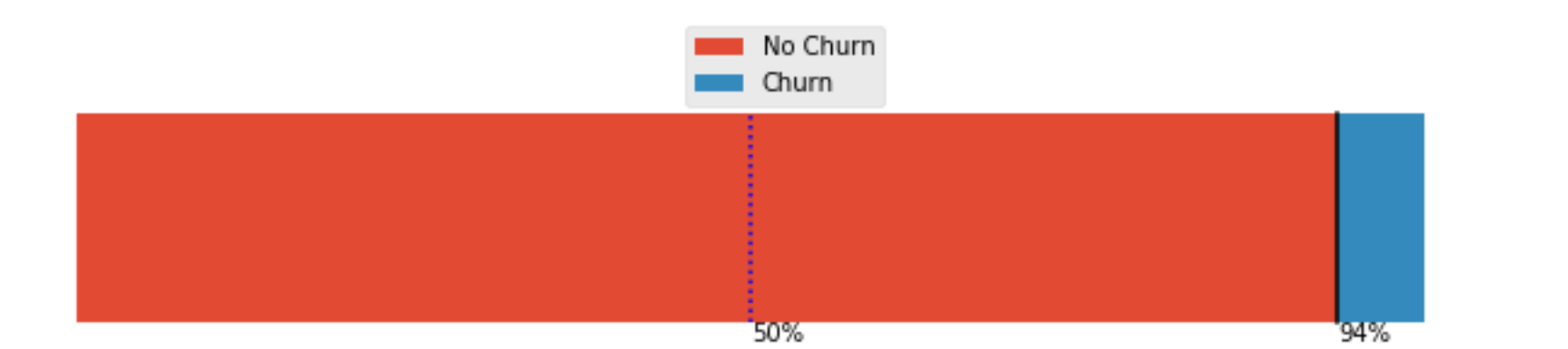
MUSIC STREAMING PROVIDER CHURN PREDICTION

THIS PROJECT IS BASED ON A KAGGLE COMPETITION WHERE KKBOX, AN ASIAN LEADING MUSIC STREAMING COMPANY, IS INTERESTED IN LEARNING FROM THEIR CUSTOMER BEHAVIOR TO PREVENT THEM FROM CHURNING.

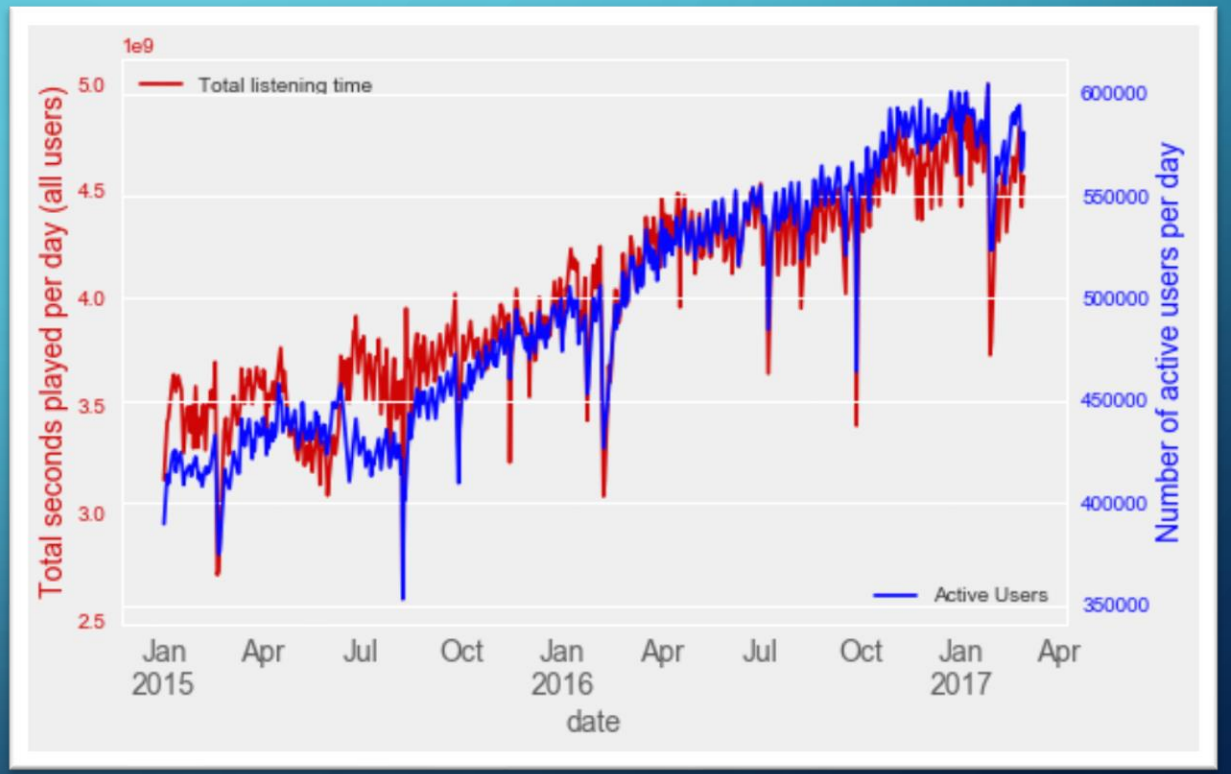
PROBLEM DESCRIPTION

- A customer has churned if he/she didn't renew his membership 30 days after it expired.
- Attracting new customer is costly.
- Customer Retention can be assessed through modeling.
- Class imbalanced ratio of 15:1 (Most customers maintain their subscription)



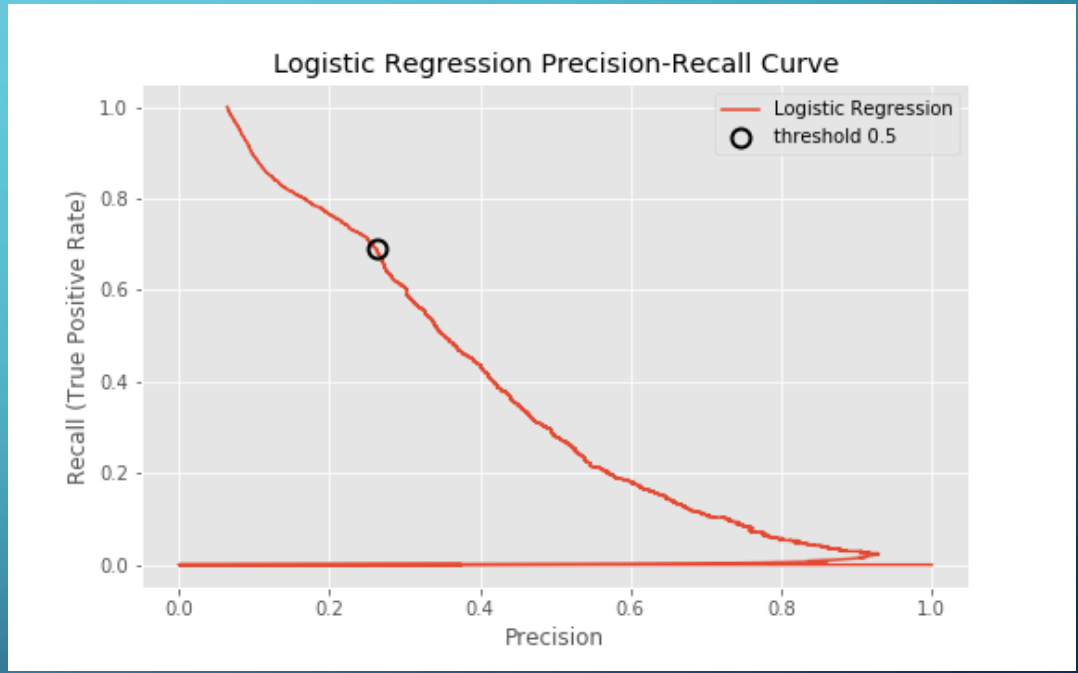
DATA FROM JAN-2015 TO MARCH-2017

- Transaction history (1.5GB):
  - Most customers opt for a monthly plan
  - Most customers chooses automatic renewal
  - Payment method 41 (unknown here) is the most popular
- Activity log (30 GB):
  - Customers tend to listen to music a few minutes a day or by 15 min increment.
  - Customer engagement drops during holidays (especially Chinese New Year, see figure to the right)



MODEL

- Model include 19 fundamental features based on transactions and listening activity.
- Identifying churn correctly is more important (less costly!) than label a no churn as churn (False Positive).
- Given our Logistic regression threshold of 0.5, we maximize recall=69% and comprise on precision=26%.



In order to limit the number of churn misidentification (high cost), we will have to include customers who were not going to churn in our results.