

Capstone Project #1: KKBOX churn prediction

1. Motivation:

Any subscription-based business is committed to keep their customer happy in exchange for their loyalty. Despite their efforts, some customers will NOT renew their subscription. In this latter case, if a customer has not renewed within a time window after its subscription expiration date, this customer is said to have churned. Although each service provider offers several subscription options (monthly, yearly, basic, premium,...), the time window to consider a customer has churned varies quite a lot. It seems that each company sets their own.

This is a common issue for businesses where machine learning has proven to be effective. It is much costlier to attract new users rather than keeping your current subscribers. Therefore, customer behavioral analysis through data analytics is an investment that pays off. Becoming familiar with this type of dataset is highly valuable.

2. Description

This project is based on a [Kaggle competition](#) where KKBOX, an Asian leading music streaming service, is interested in predicting their churn rate on a monthly basis. The time window is 30 days in order to consider a customer has churned after its current membership ends.

3. Data

Most of the time, data from Kaggle competition comes pre-cleaned and arranged in an analysis ready format. This is not the case. There are multiple files that contains different information and in the data description it clearly states there are some unrealistic entries. The largest file size is nearly 30 GB for a total of about 33GB. More details can be found [here](#).