**Capstone Project I: Customer Churning**

**Introduction and Problem**

Customer churn occurs when a customer or subscriber decides to stop doing business with a company or service. Customer churn is a critical metric for a company. Retaining a customer is not only more cost efficient than acquiring a new customer, it is also important for the growth of a business. For this project, the dataset is provided by KKBOX, Asia’s leading music streaming service. This goal of this project is to predict the probability that a customer will churn within 30 days after their current membership expiration date. With this information, a company can make better decisions in order to prevent a customer from switching to a competitor.

**Client**

The main client will be KKBOX. However, the results of this project may be applied to other subscription services. KKBOX is a music streaming service available in Asia. They have a freemium business model. Their revenue is gained through advertisements and subscriptions. Slight variations in churn can have a major effect on profit for a subscription business. Thus, accurately predicting churn will help them remain successful in the future.

**Data**

The data will come from a Kaggle competition. According to Kaggle, this dataset has been donated from KKBOX. The link is provided below.

https://www.kaggle.com/c/kkbox-churn-prediction-challenge/data

**Approach**

1. Data Wrangling - The data will be merged together into a single Pandas DataFrame object. The data will also be inspected and cleaned.
2. Exploratory Data Analysis - The dataset will be explored and visuals/graphs will be plotted.
3. Statistical Analysis - Hypothesis testing will be applied.
4. Machine Learning - The data will be split into a training set for the machine learning model and the testing set. The machine learning model will then be applied.
5. Documentation - All steps will be documented, and results will be presented.

**Deliverables**

1. Final report
2. Slide deck
3. Code
4. Machine Learning model
5. Summary of the performance of the model