Note: This presentation was made as a quick pitch presentation to the client, with more emphasis on value projection, less on technicalities
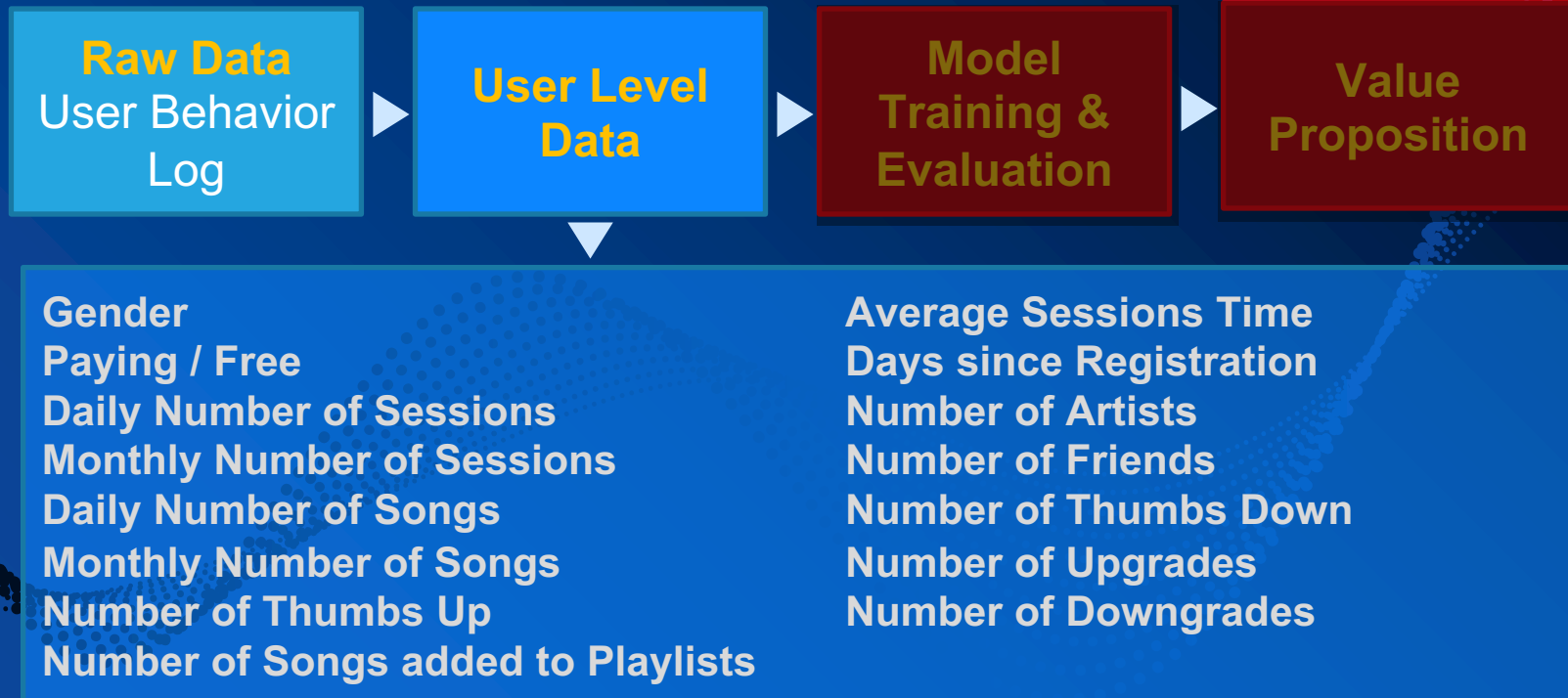
Business Problem & Objective

Problem 🚶

Potential Revenue is lost from Users who cancel their subscription (churn). In particular, we have no insight into who these churn users might be, thus preventing follow-up retention strategies.
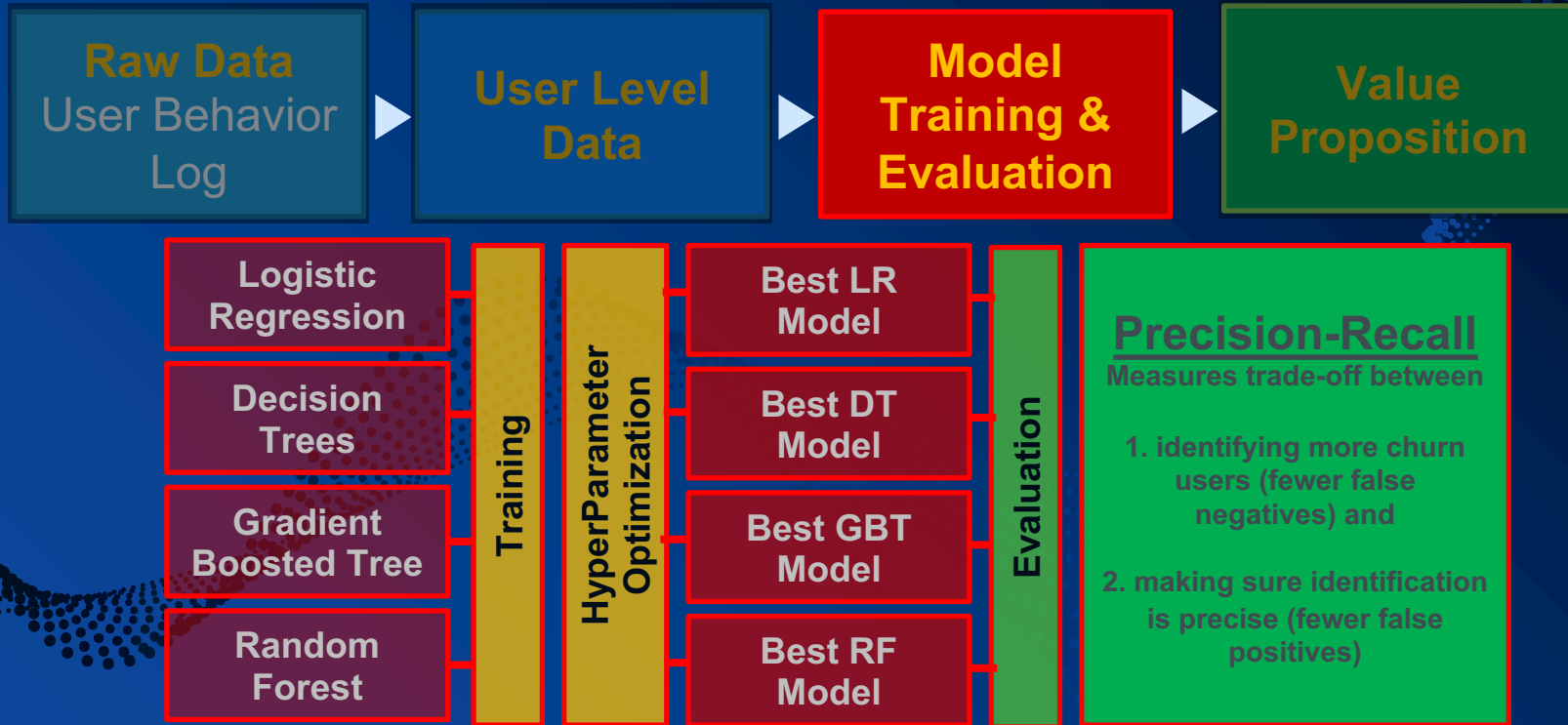
Objective 🔍

To identify potential churn users using individual and usage data, which will subsequently inform targeted user retention strategies

Approach : Feature Engineering

Raw Data
User Behavior Log

User Level Data

Model Training & Evaluation

Value Proposition

Gender
Paying / Free
Daily Number of Sessions
Monthly Number of Sessions
Daily Number of Songs
Monthly Number of Songs
Number of Thumbs Up
Number of Songs added to Playlists

Average Sessions Time
Days since Registration
Number of Artists
Number of Friends
Number of Thumbs Down
Number of Upgrades
Number of Downgrades

Approach : Modelling & Evaluation

Raw Data
User Behavior Log

User Level Data

Model Training & Evaluation

Value Proposition

Logistic Regression

Decision Trees

Gradient Boosted Tree

Random Forest

Training

HyperParameter Optimization

Best LR Model

Best DT Model

Best GBT Model

Best RF Model

Evaluation

Precision-Recall
Measures trade-off between

1. identifying more churn users (fewer false negatives) and

2. making sure identification is precise (fewer false positives)

Results : Value Proposition

| Raw Data
User Behavior Log | → | User Level Data | → | Model Training & Evaluation | → | Value Proposition |

	Predict Non-Churn	Predict Churn
Actual Non-Churn	True Neg 316	False Pos 33
Actual Churn	False Neg 9	True Pos 90

Value Proposition

	Predict Non-Churn	Predict Churn
Actual Non-Churn	True Neg 316	False Pos 33
Actual Churn	False Neg 9	True Pos 90

Out of 100 Churn users, we can identify 90 beforehand.

Given baseline assumptions[*], and 100% retention from identification, revenues will increase up to **26%**.

Assuming proportion of users by paying/free and churn/no-churn are as in the dataset, and a 1:3 revenue ratio for free:paying users (which turns out to not be a crucial assumption)

End of Presentation