Machine Learning Classification Project

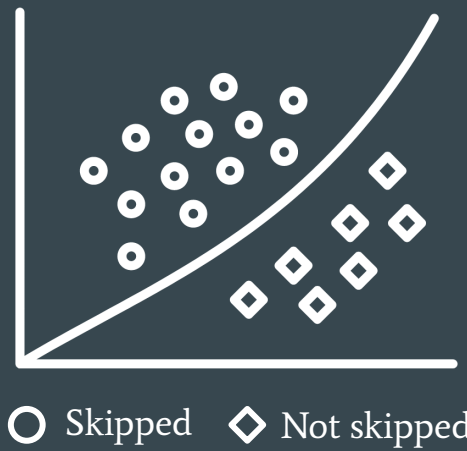
July 12, 2019

Chelsea Power

Project objective:
Determine if a song will be skipped or fully played in order to guide the business on how to negotiate upcoming licensing renewal fees.

Overview

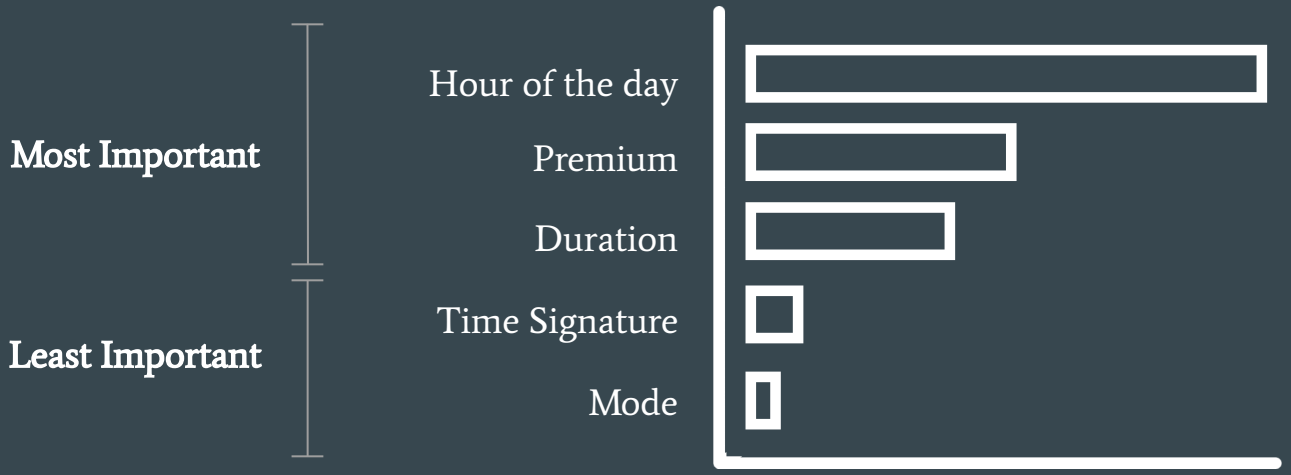
Based on Spotify 2018 streaming session data across 167,880 songs, this machine learning classification model determines with 91% accuracy whether a track will be skipped or fully played (not skipped).



91% Accuracy

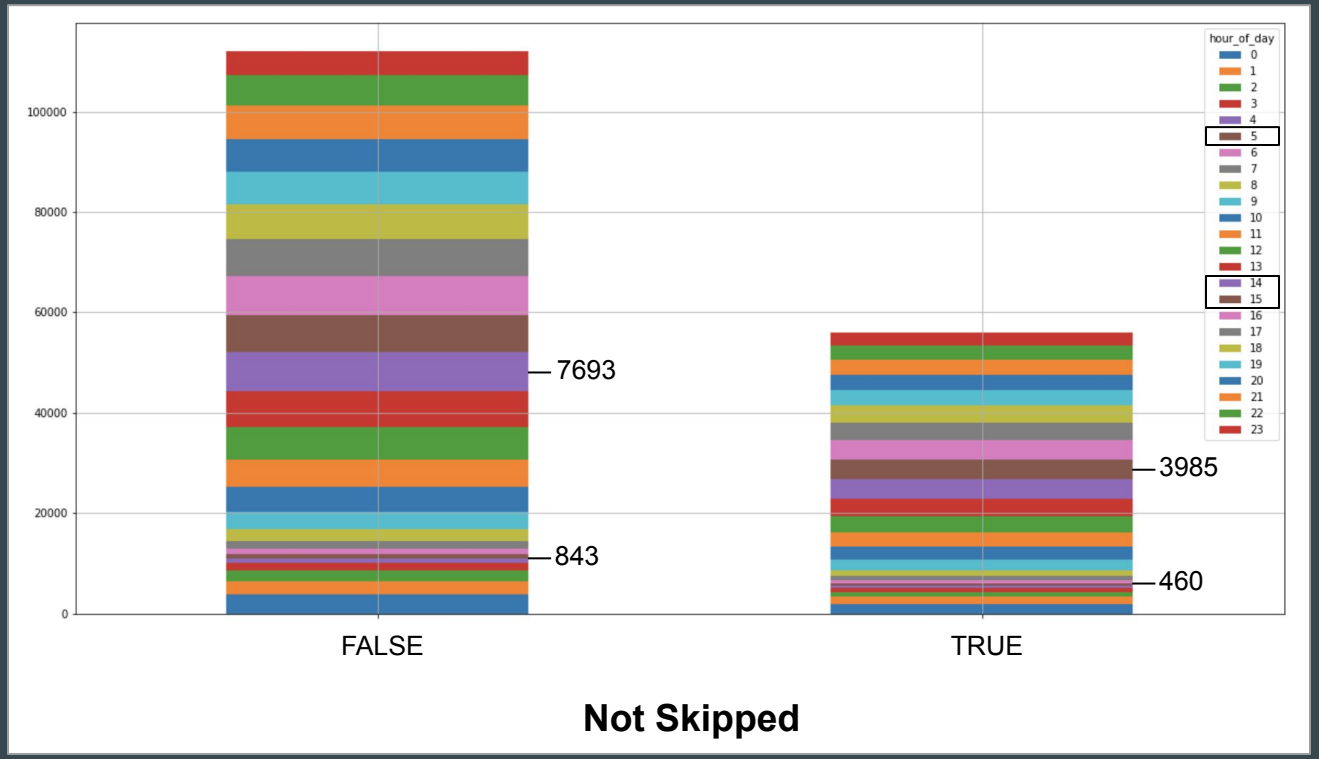
Song Attributes

The streaming session data started with 50 attributes and it was reduced to the 33. Of the 33 attributes, the chart below shows the three most important and two least important attributes of a song that impacts it being fully played.



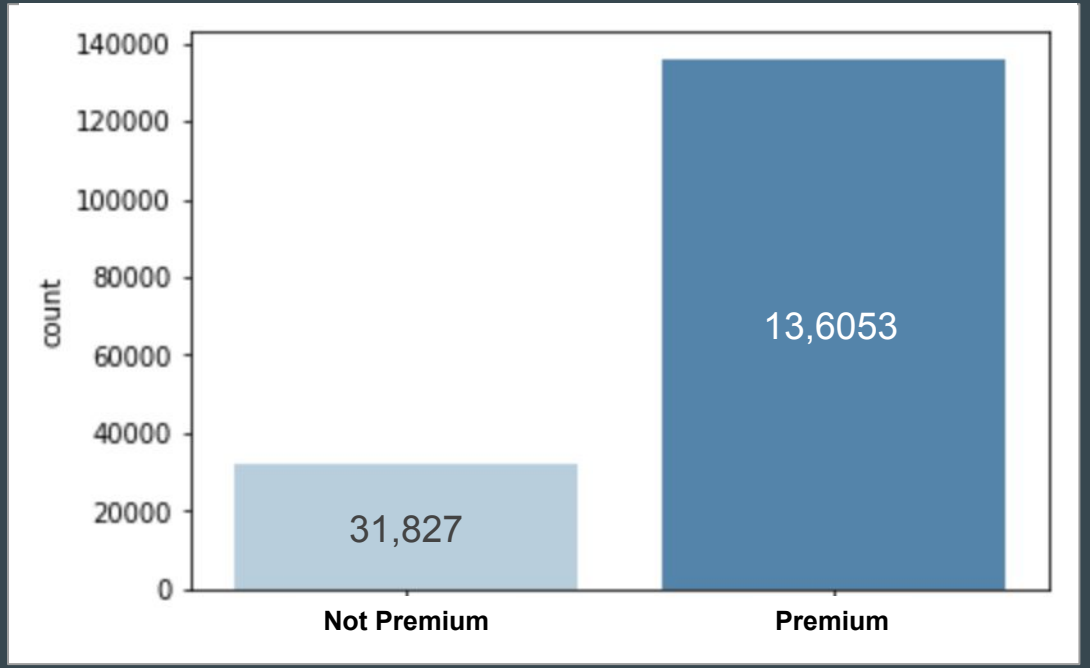
Hour of the Day

The chart to the right shows the amount of skipped or fully played songs by hour of day. The most skipped songs happen at **hour 14**. The least skipped songs happen at **hour 5**.



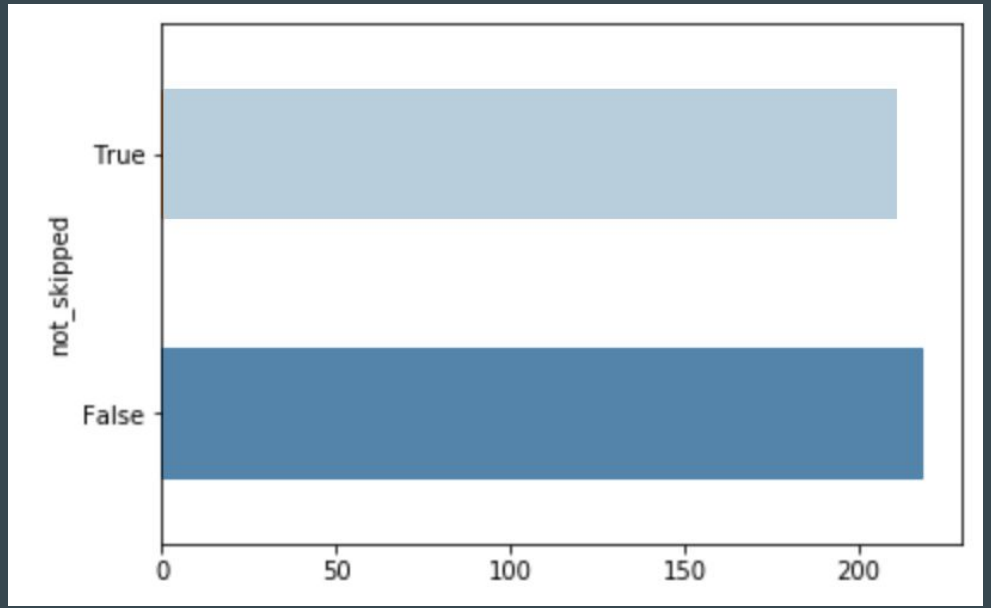
Premium Subscriptions

The chart to the right shows how many premium vs. non-premium accounts played the songs.



Duration

The chart below shows the average duration of the songs that were skipped vs. fully played.



Next Steps

Since first Random Forest classifier model produced **91% accuracy,** it would be interesting to know if the model can be further improved by running **Principal Component Analysis (PCA)**.

Another consideration would be to re-run this project using a different **predictor variable** to see if the accuracy increases further.

Thank You