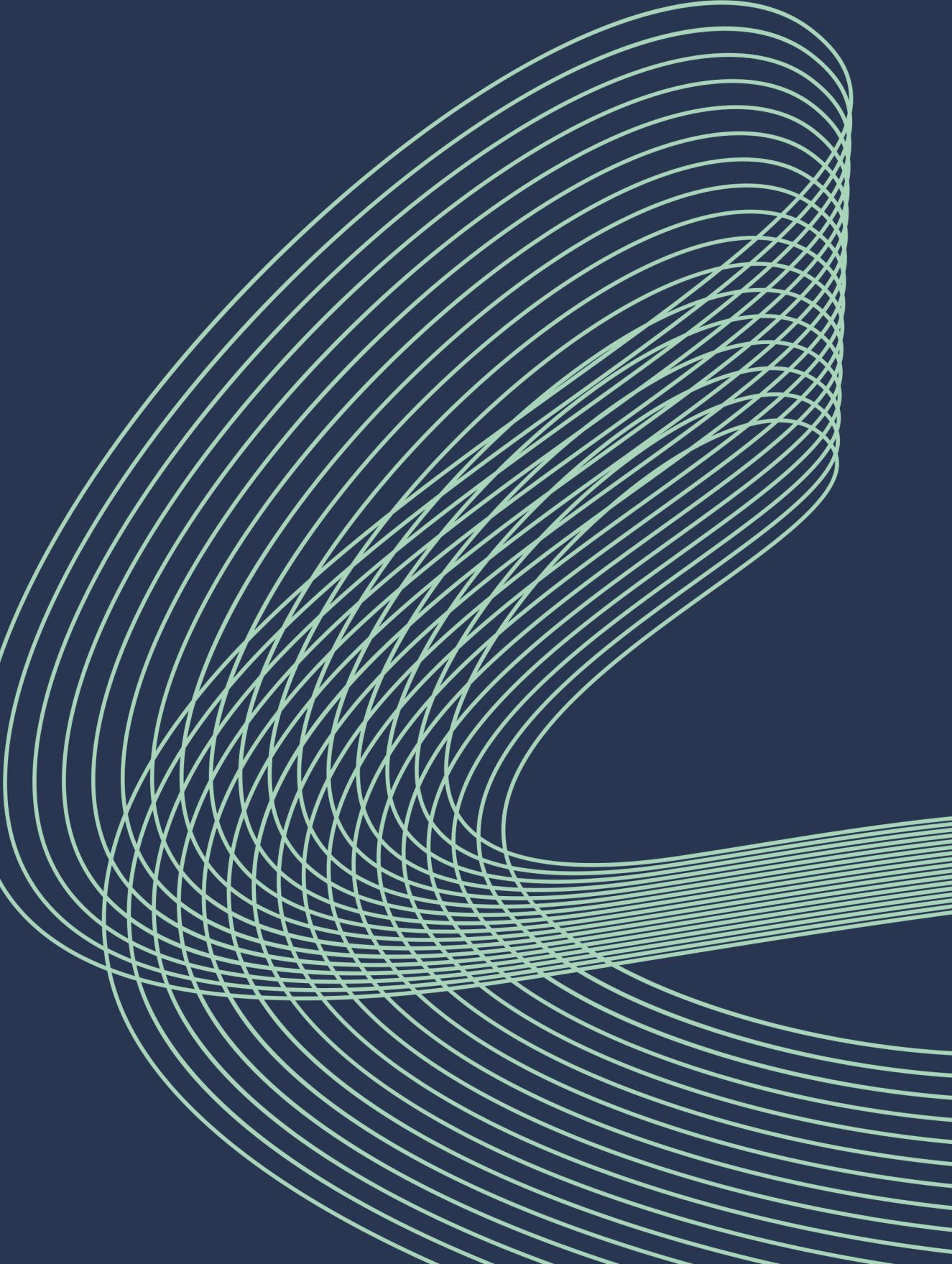


# SPOTIFY REVENUE

Understanding factors that influence stream revenue

PRESENTED BY

Wilson Qin, Kimberly Rodriguez,  
German Romero, Kayla Hoffman



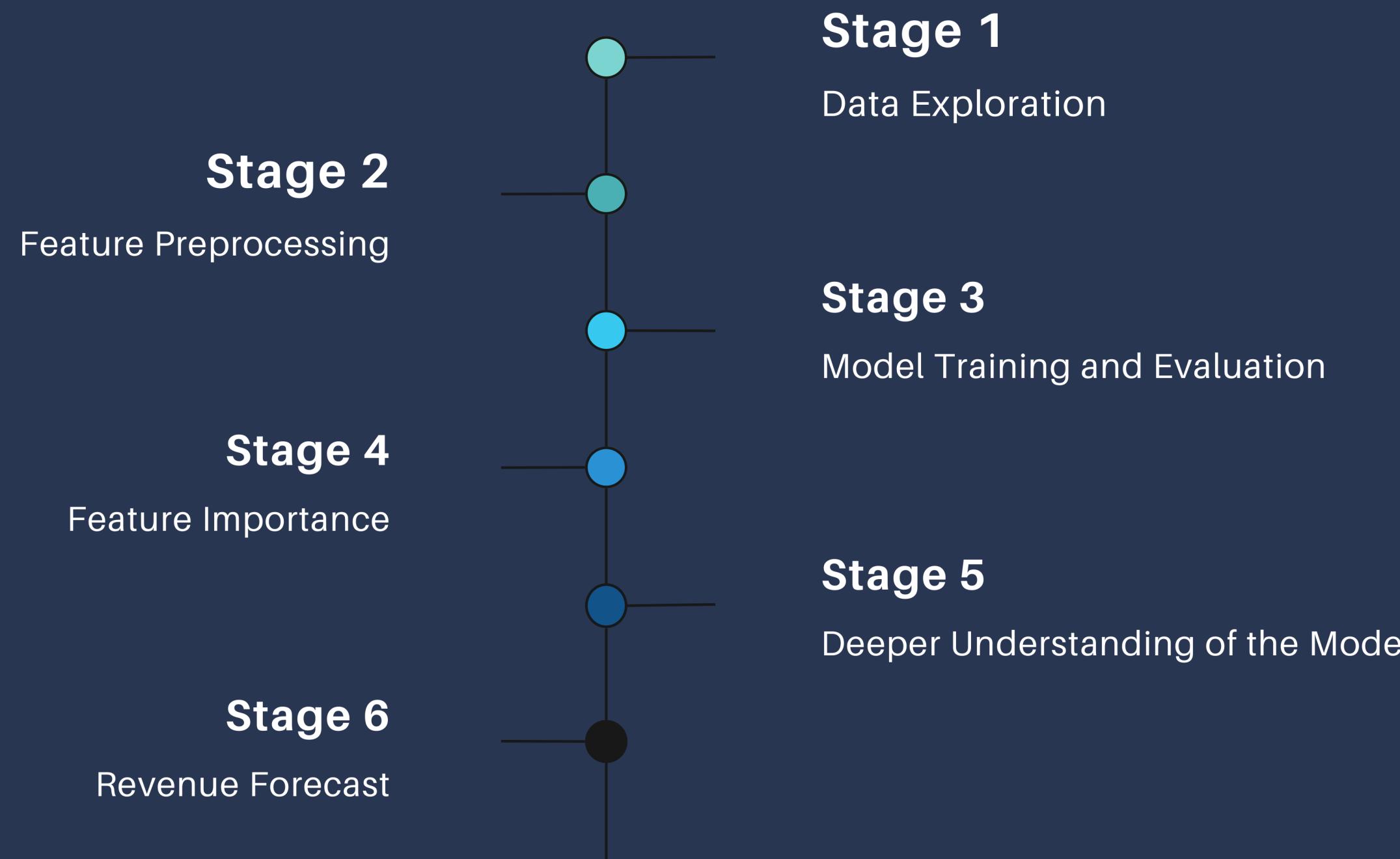
# PROJECT OBJECTIVE

Make predictions about how popular a music genre might be given different features. Predicting genre stream count can give insight on revenue generation

## HOW DOES SPOTIFY STREAMING WORK?

1. Spotify's total streaming revenue pool
2. The negotiated global payout as a percentage of that revenue
3. Total number of streams on the platform
4. The number of streams on your platform

# MACHINE LEARNING MODEL PIPELINE



# MACHINE LEARNING MODELS

Essential Machine Learning Models

**PYTORCH**

Training Neural  
Networks

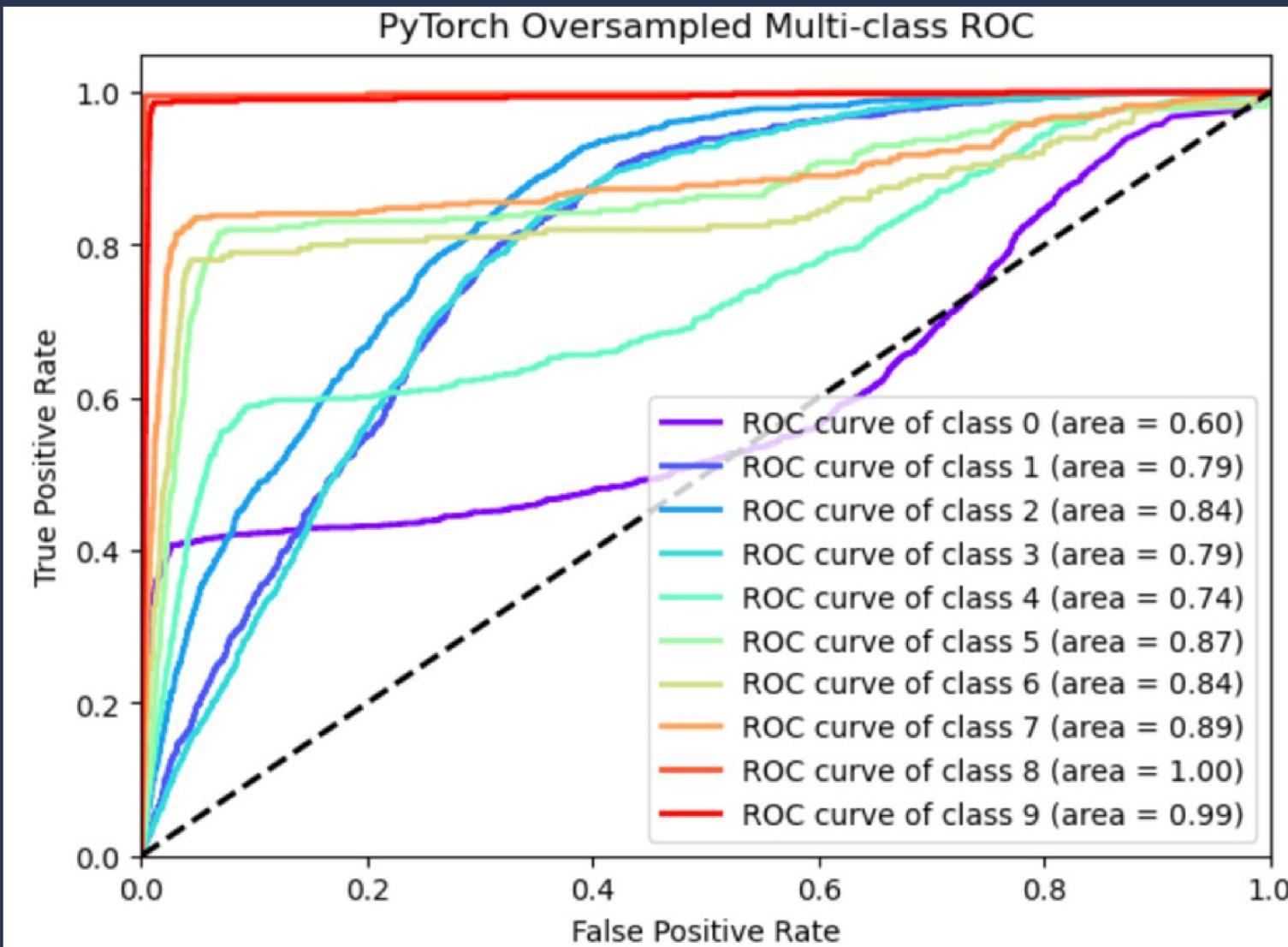
**RANDOM FOREST**

Builds Multiple  
Decision Trees

**XGBOOST**

Improves speed and  
performance of decision trees

# PyTorch



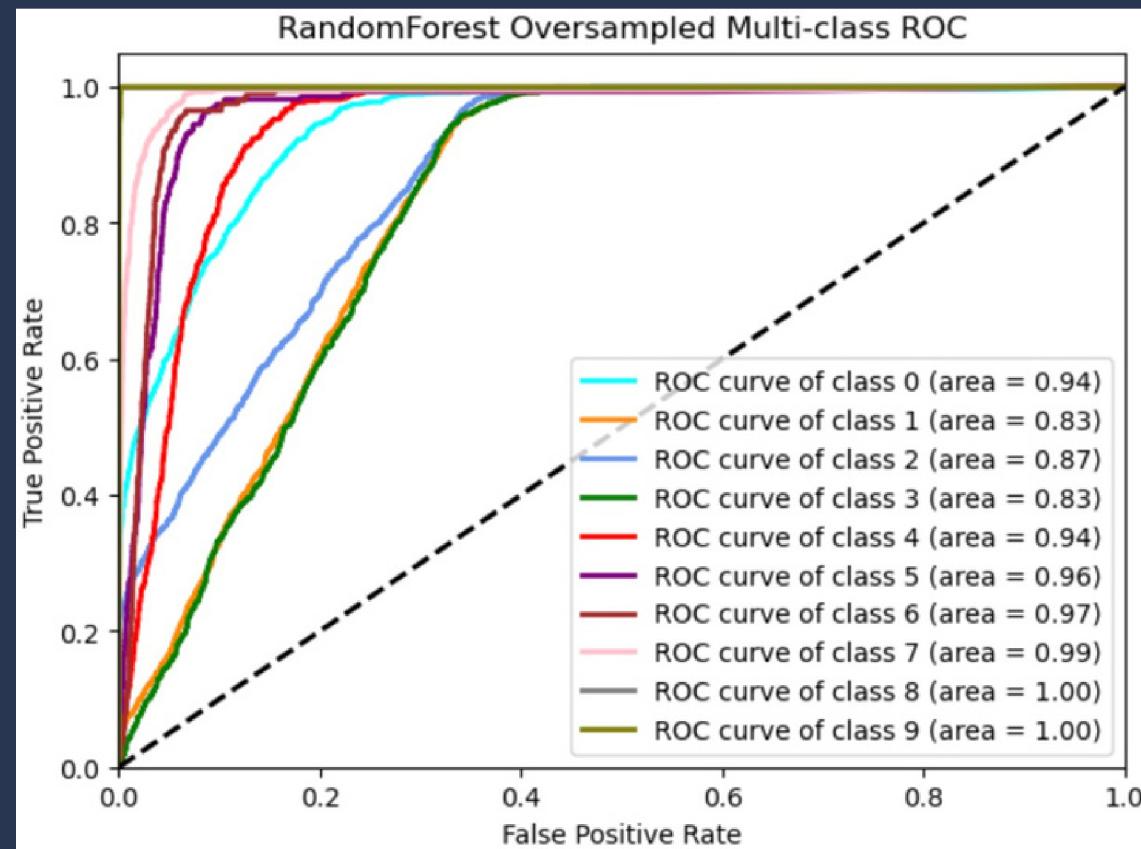
# ROC OVERSAMPLLED

Oversampling helped solve class imbalance by increasing the minority class

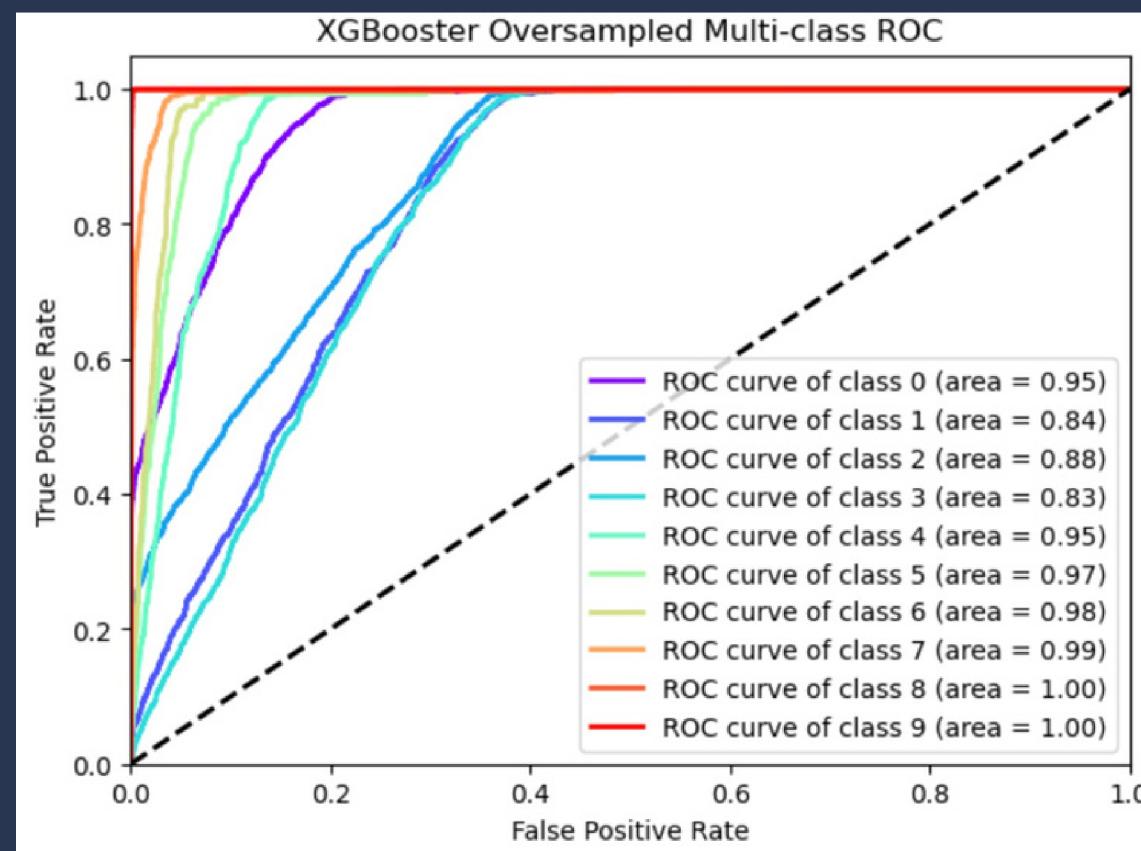
## PyTorch

- ROC Curve is the trade off between True Positive and False Positive Rate
- Class 8:  $AUC = 1.00$  is perfectly classified by the model
- Class 0:  $AUC = 0.60$  is the worst performing class because its closest to the "no skill" line
- Model Discrimination Ability: Varying across classes
  - Correctly differentiate classes

# Random Forest



# XGBooster



# ROC OVERSAMPLED

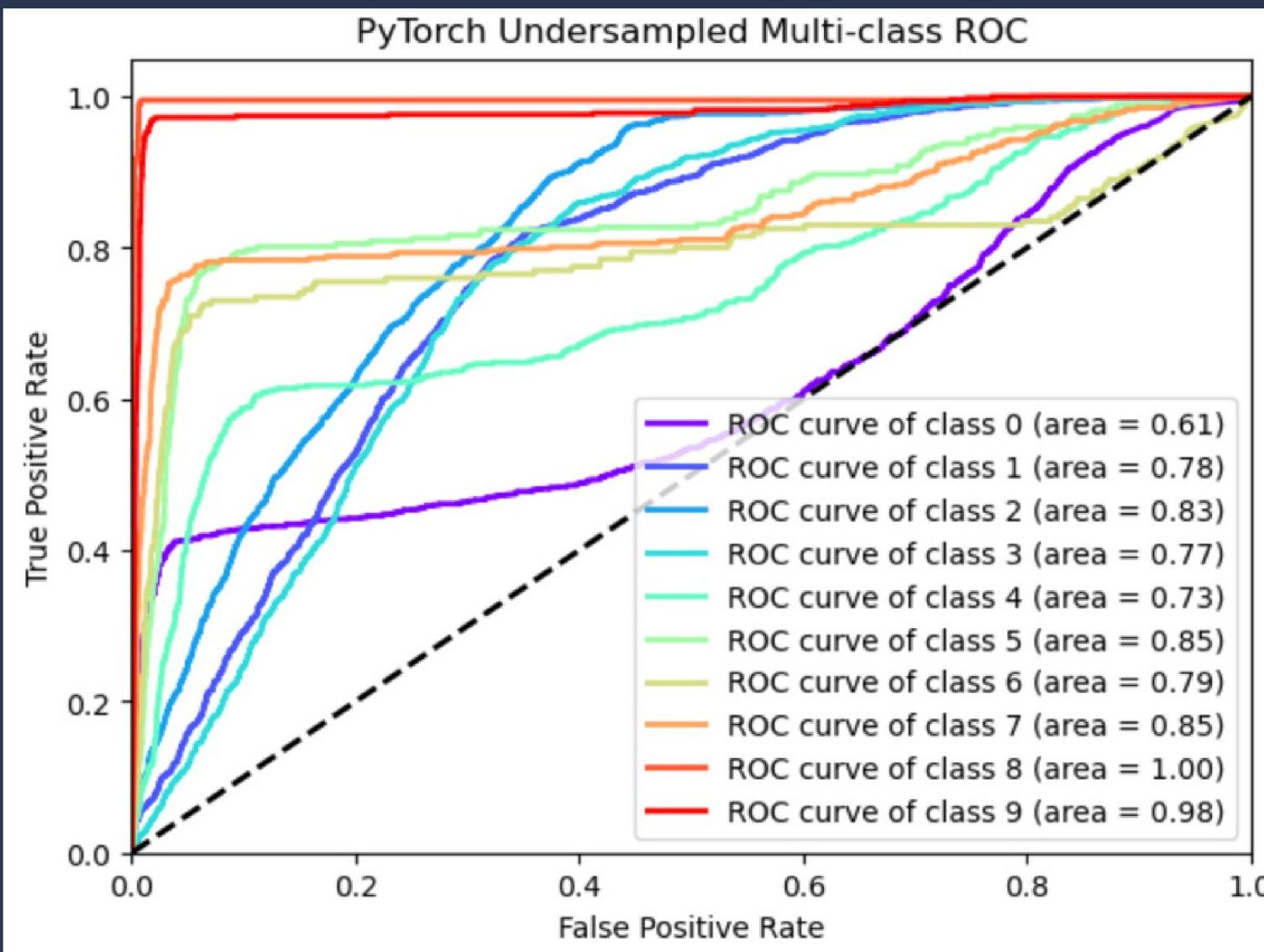
## Random Forest and XGBoost

- All classes have AUC values above 0.80
- Both models are consistently strong in distinguishing different classes
- Perfect scores for classes 8 and 9
- Random Forest and XGBoost performed better than PyTorch

## Reclassification Report

	precision	recall	f1-score	support
0	0.92	0.45	0.61	1644
1	0.38	0.43	0.40	1218
2	0.65	0.41	0.50	1670
3	0.35	0.47	0.40	1176
4	0.43	0.64	0.52	526
5	0.38	0.77	0.50	272
6	0.37	0.82	0.51	200
7	0.75	0.87	0.80	402
8	1.00	1.00	1.00	417
9	0.97	1.00	0.99	505
accuracy			0.56	8030
macro avg	0.62	0.69	0.62	8030
weighted avg	0.63	0.56	0.57	8030

# PyTorch



# ROC UNDERSAMPLED

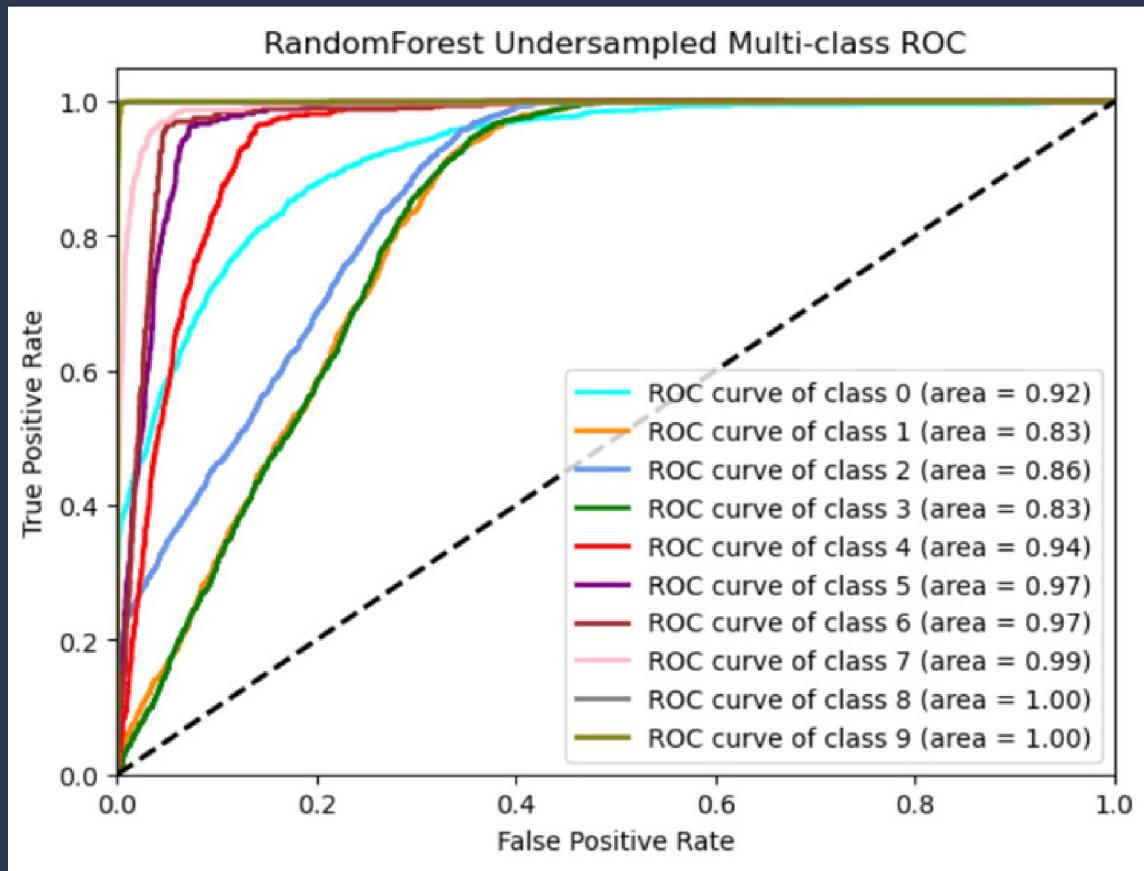
Undersampling helped solve class imbalance by decreasing the majority class

- Undersampling:
  - Reduced bias towards the majority class
  - Potentially prevents overfitting

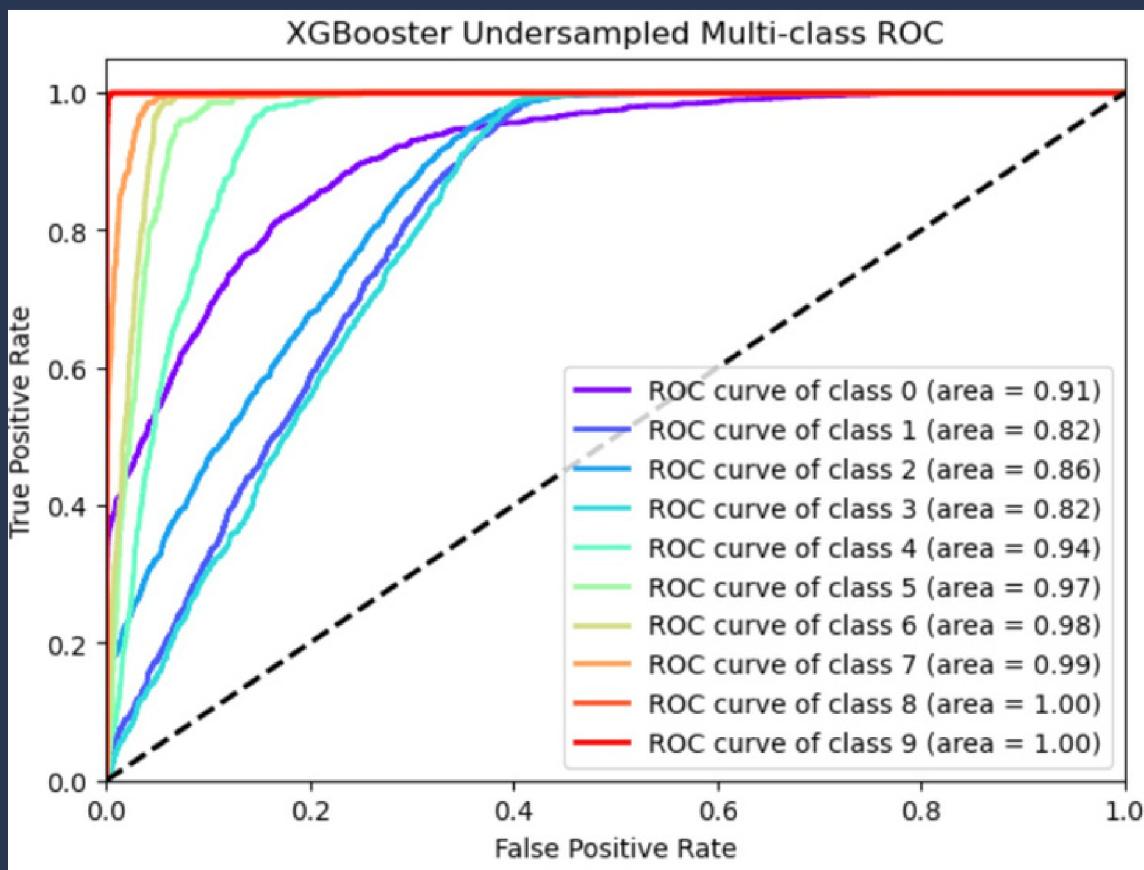
## PyTorch

- Most classes have AUC values above 0.70
- Class 8: AUC = 1.00 is perfectly classified by the model
- Class 0: AUC = 0.61 is the worst performing

# Random Forest



# XGBooster



# ROC UNDERSAMPLED

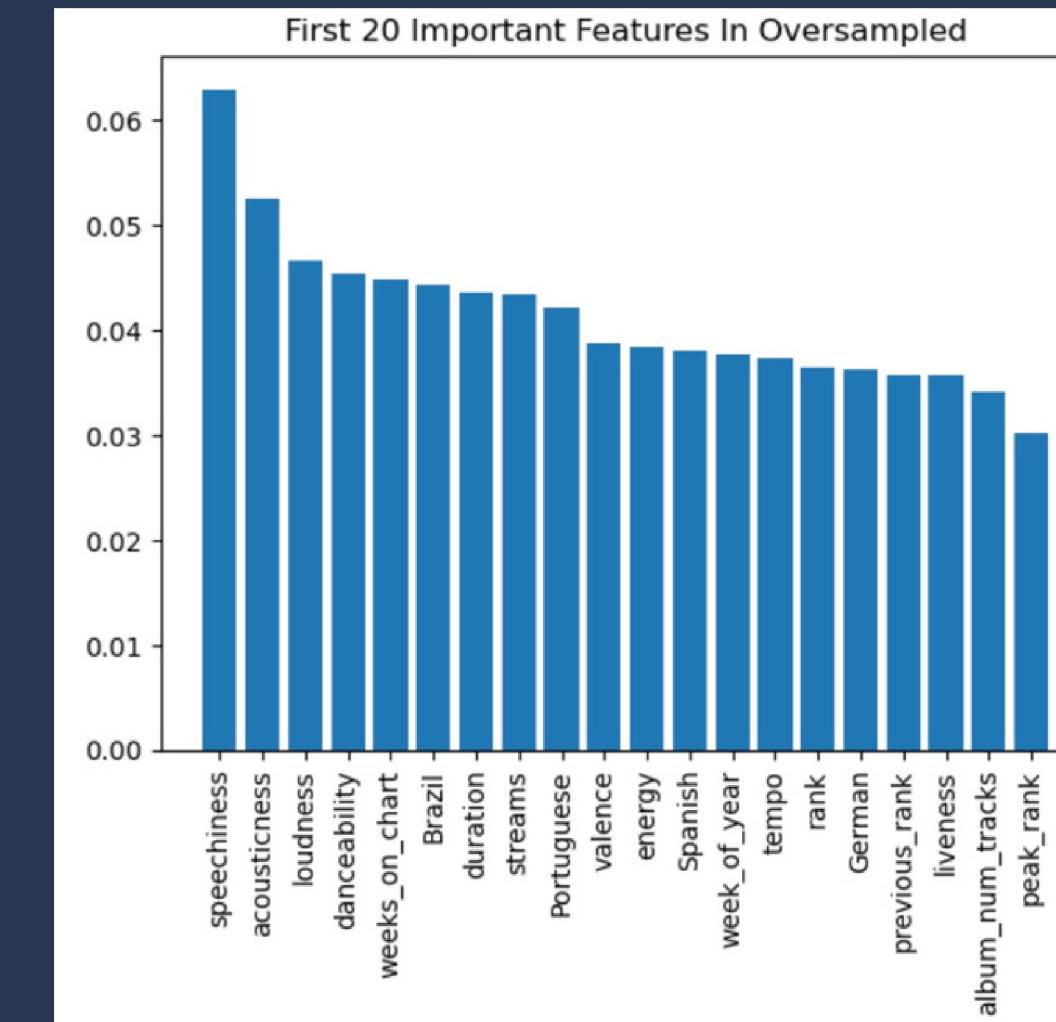
## Random Forest and XGBooster

- All classes have AUC values above 0.80
- Strong distinguishing between classes
- Perfect scores for classes 8 and 9 could be strong correlation features
- Random Forest and XGBoosters performed better than PyTorch

# FEATURE IMPORTANCE

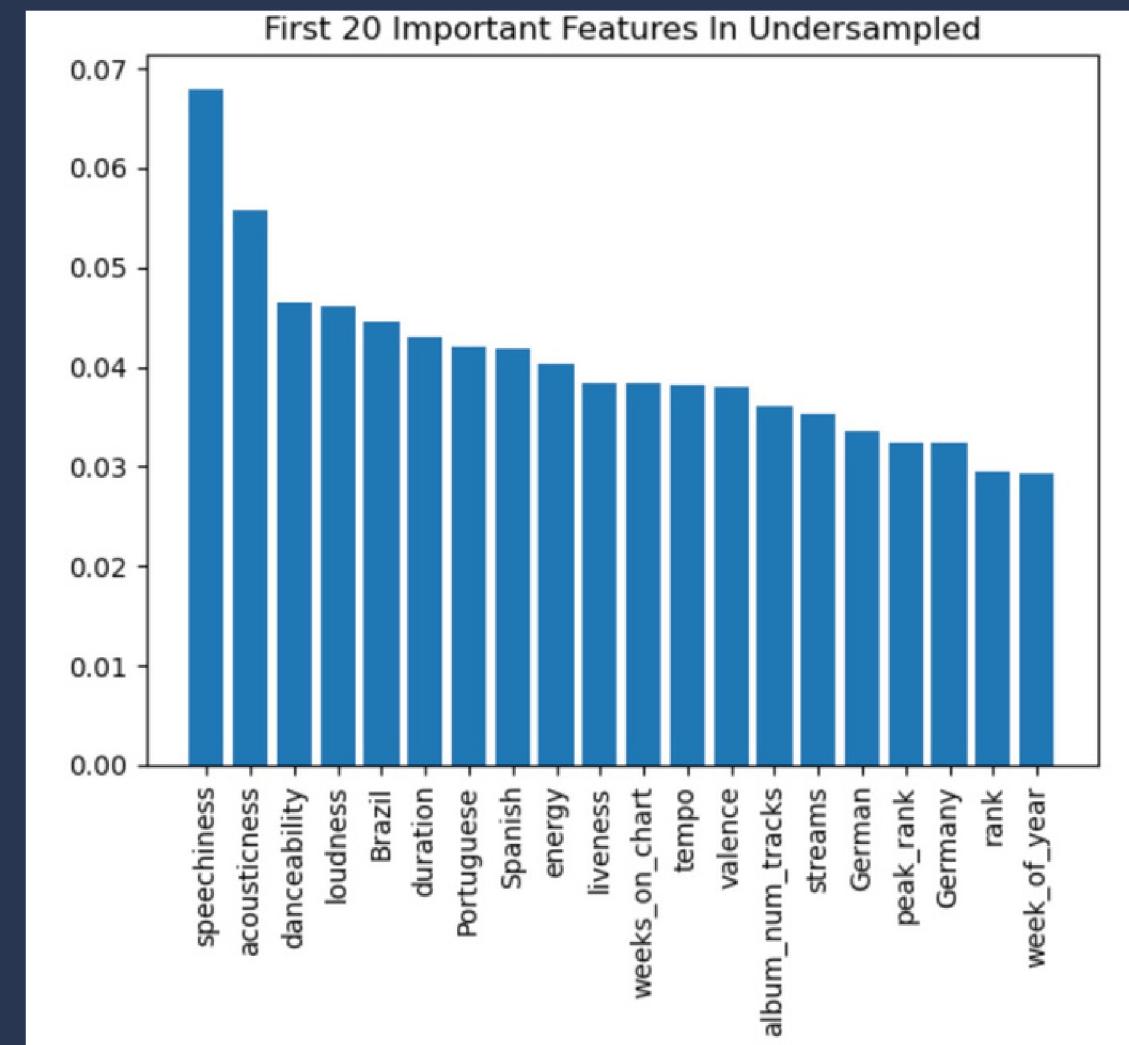
Which features have the highest influence on model output?

- Top 2 features within both the Oversampled and Undersampled Data
  - Speechness
  - Acoustics
- Strong relationship between target variable and feature
- Strong stability across samples
- Independent from class distribution



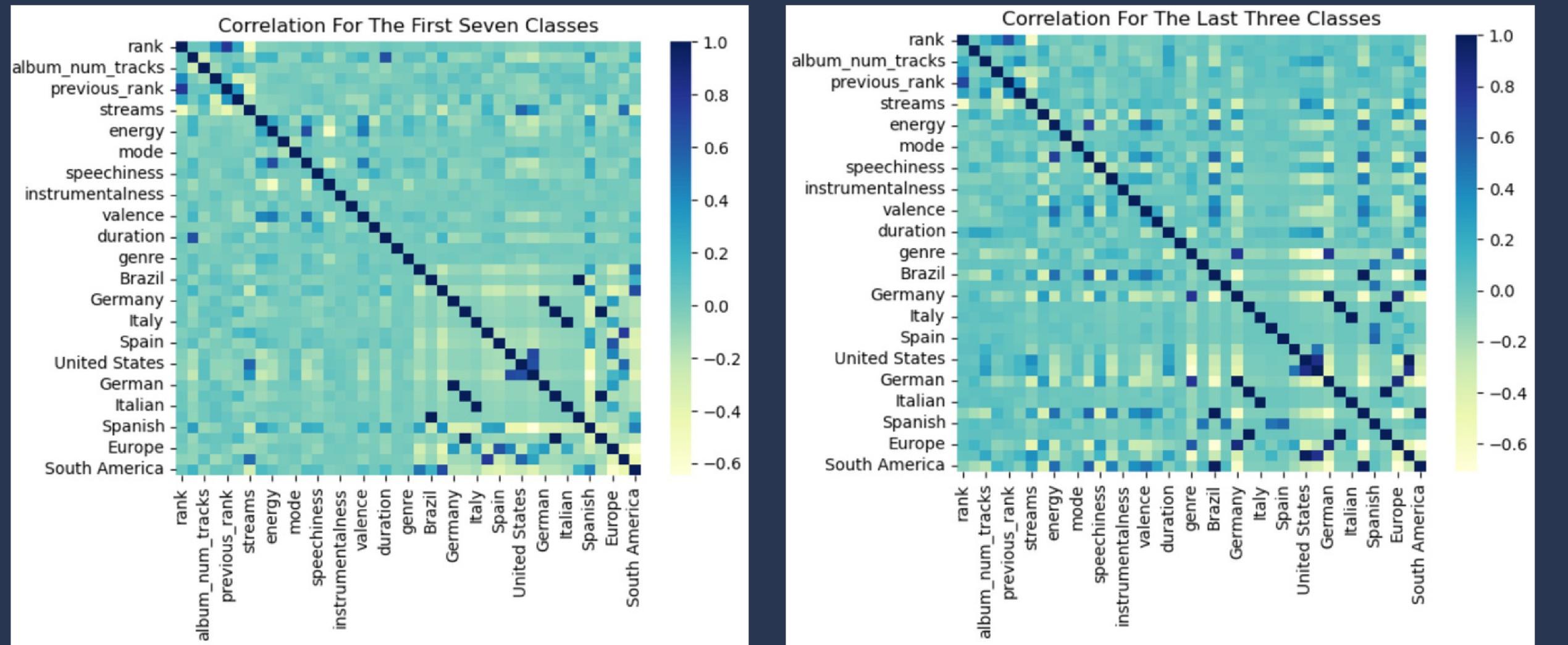
Our Top 20 Features  
Ranged from

- Acoustic Features
- Language
- Timeframe



# DEEPER UNDERSTANDING OF THE MODEL

Correlation Matrix of the first seven classes and last three classes



## Last 3 Classes

- Last 3 Classes = Highest Scores
- High correlation to specific features
- Easier to predict these classes

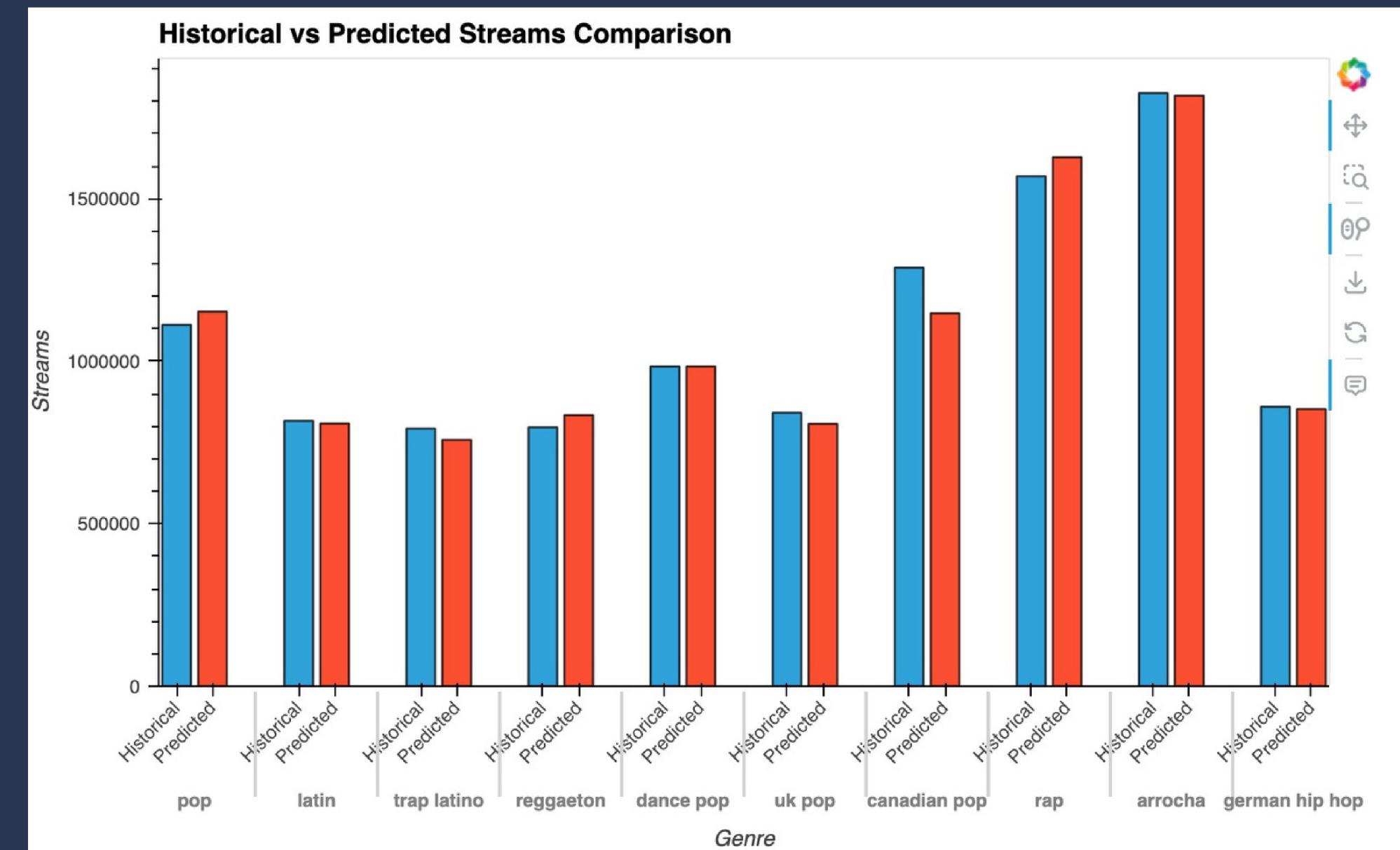
### Acoustic Features Include

- Energy
- Speechiness
- Instrumentalness
- Valence
- Duration

# Streams Performance

Comparing historical and predicted stream count

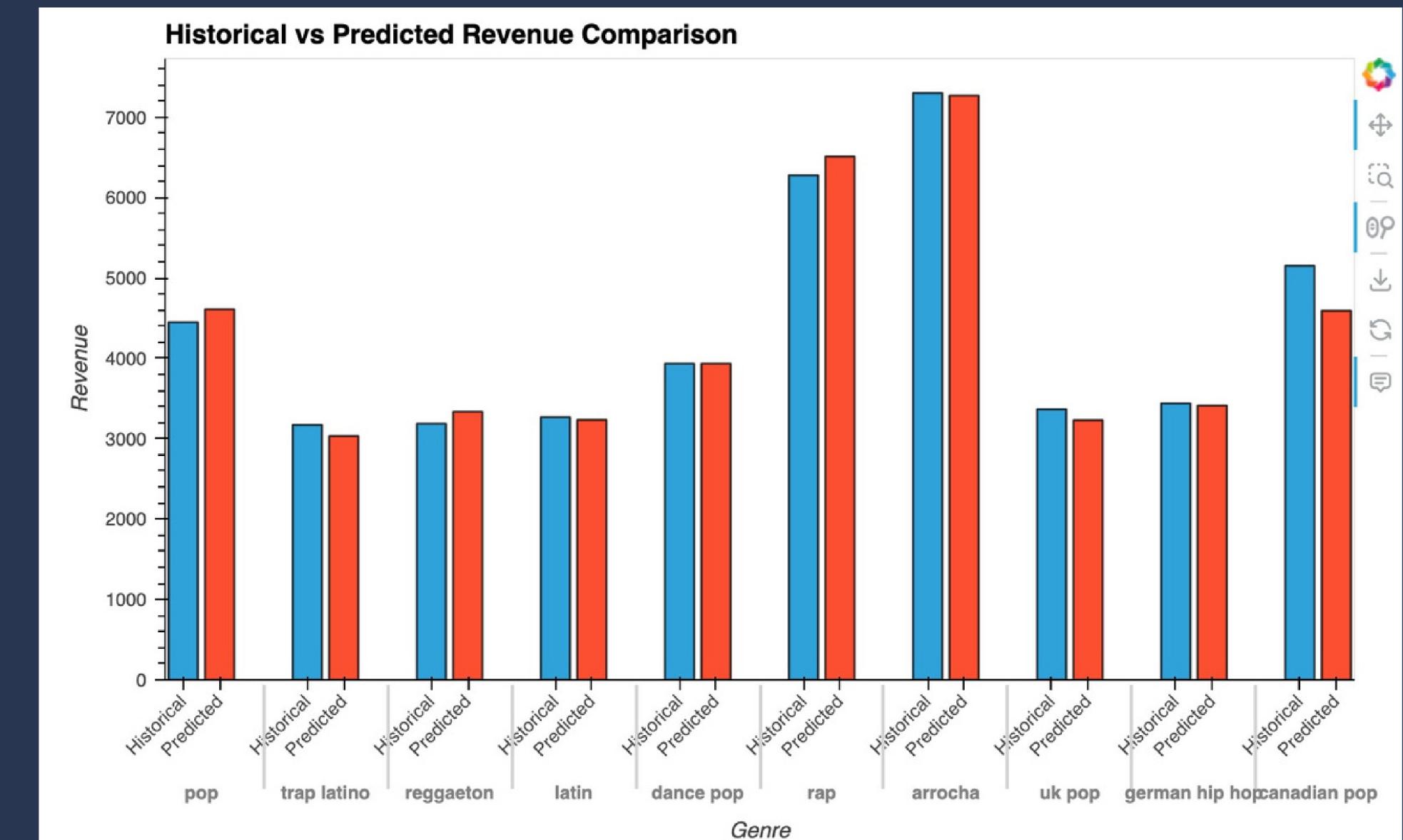
- Overall accurate predictions compared to historical data



# Revenue Performance

Comparing historical and predicted revenue

- Average pay per stream = \$0.003 to \$0.005
- Source : <https://dittomusic.com/en/blog/how-much-does-spotify-pay-per-stream>
- Successfully able to accurately predict genre revenue



# Concluding Remarks

## Overview

---

- Random Forest and XGBoost had similar performance in both under sampled and over sampled data
- Discover some important features regardless of class distribution
- Accurately predicted

## Future Improvements

---

- Analyze larger dataset
- Fix PyTorch problem
- Time Series Analysis

# THANK YOU

---



# RESOURCE PAGE

DI+O Music: <https://dittomusic.com/en/blog/how-much-does-spotify-pay-per-stream>

---

Click 'Go to recording studio,' where you can choose the video and audio source for your video presentation.

---

Feel free to choose the 'No camera' option and record your voice only.

---

Start recording, and press pause in between takes if you have to.

---

Once you're done, download your Canva Presentation in MP4 file format or get a link to your Talking Presentation and share it with others.

---

You can also record a video inside the editor! Go to 'Uploads' and click on 'Record yourself'.