# Social Media and Success in the Music Industry

Client: Sony Music – Marc Fridson | Advisor: Michael Smith | Team: Zhexin Chen, Ghazal Erfani, Gaurij Hardikar, Daniel Lesser, Phani Krishna Pasumarthi

## Introduction

Social media music promotion is a necessity in today's music marketplace. Artists are building their own personal brand by engaging with their fans online.

Sony aims to better understand how to promote content on social media and support artist development.

## Purpose

1. How does **artist activity** on Twitter, Facebook, Instagram, and YouTube drive **social media engagement**?
2. How does activity **differ** by **platform** and **genre**?
3. Can we make recommendations around **when**, **where**, and **how** artists should engage in social media?

## Data Processing

**6** Genres

**141** Artists

**6** Data Sources

**13** Core Features

**1M** Observations

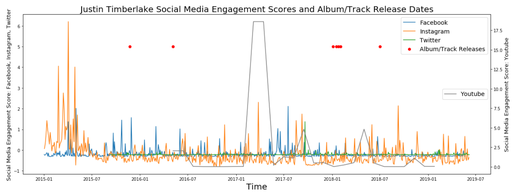
## Exploratory Analysis

High frequency of posts

↓

Difficult to visually discern relationship between social media activity and engagement

↓




**For a better understanding, we took 3 approaches:**

Social Engagement Success Prediction | Social Media Activity Clustering | Social Media Sentiment Analysis

*Quantitative* ⟷ *Qualitative*

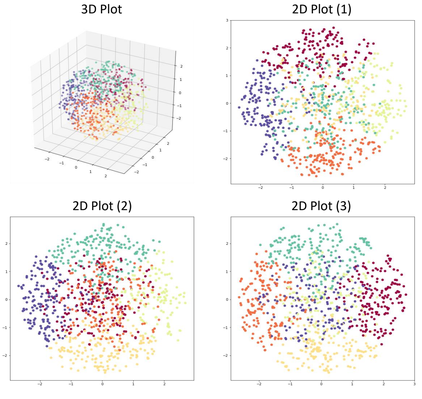
## Activity Clustering

### 1. Dimensionality Reduction

27 Features → PCA → t-SNE → 3 Features

### 2. Gaussian Mixture Modeling

Each post is probabilistically assigned to one of six clusters, represented here as collections of colored data points




**Cluster 3**

- High Rock & Electronic Engagement
- Early Morning Posts
- Close to Release-Date Posts
- Would benefit from more photo posts

**Cluster 5**

- Older Artists
- Afternoon Posts
- High Facebook Engagement
- Would benefit from early morning posts

**Cluster 4**

- Low Post Frequency
- Early Morning Posts
- High Instagram Engagement
- Frequent Twitter Posts
- Would benefit from more Instagram posts
- Would benefit from more photo posts
- Would benefit from lower-frequency posts

## Success Prediction

**Method**

Linear regression and feature selection for each social media

↓

Model selection with cross validation

↓

Best model to predict social engagement scores

**Example: Instagram**

A linear regression for 23 features and 214k observations with -0.85 $R^2$ score on the test set

↓

Cross validation with multiple models: Lasso, Ridge, Random Forest, XGBoost, etc

↓

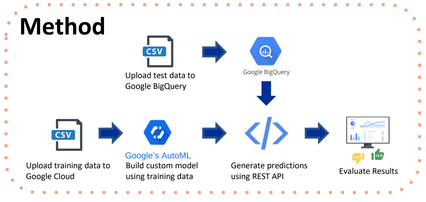
An XGBoost Regression with 300 estimators and 9 max depth results in 0.58 $R^2$ Score on the test set

**Key Takeaways:**

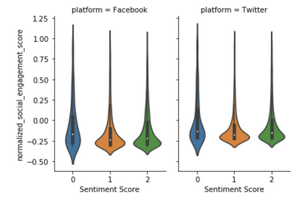
1. Lower post frequency and shorter length posts lead to higher engagement
2. Established artists have higher engagement scores than newer artists
3. Rock artists have higher Instagram engagement
   - **Positive** Coefficient for Rock Artists: **(+) 0.122**
   - **High** Feature importance of Rock Artists: **0.342**

## Sentiment Analysis

**Method**








**Key Takeaways:**

1. Most posts have positive sentiment
2. Facebook and Twitter have similar sentiment analysis distribution
3. No significant correlation between sentiment and engagement

## Future Work

1. **Beyond Social Media**
   Predict artist success on music streaming services based on social media activity across various platforms.

2. **Searching for New Talent**
   Leverage Social Media presence to help music managers identify up-and-coming artists that will contribute to the success of Sony Music.