



# TikTok Trend - Viral Background Music Analysis **Business & Analytics Plan**

Shiyi "Ashley" Yue, Huixin "Blessia" Li, Siwei "David" Ran, Zhihan "Jimmy" Jiang

# Agenda

01

## **Analytics overview**

Business Objective, Metrics  
of Success

02

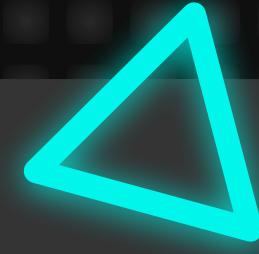
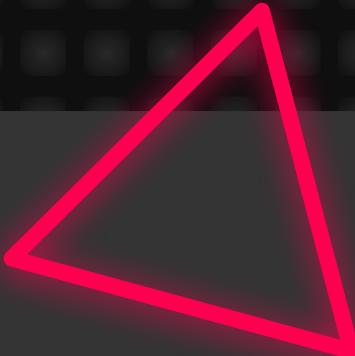
## **Analytics Methodology**

EDA, Model Building

03

## **Analytics results**

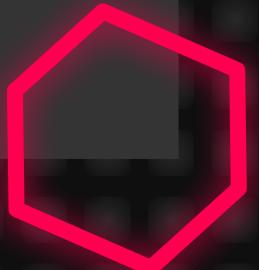
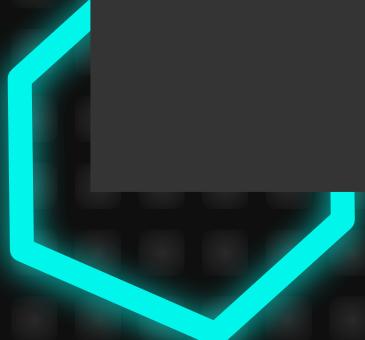
Analytics Insights



01

# Analytics overview

Business Objective, Metrics of Success



# Business Objective

01

Aim at enhancing users content strategy and engagement by identifying and capitalizing on emerging music trends

02

Aim to grow our user base and increase our market share in the influencer tool market, while ensuring a high level of customer satisfaction and retention

# Metrics of Success/KPI



**Prediction  
Accuracy**



**User  
Engagement**



**Customer  
Retention**

02

# Analytics Methodology

EDA, Model Building

# Data Source

Kaggle:

## TikTok Trending Tracks

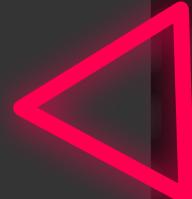
7000 Music Tracks, Technical Information

Data Card    Code (4)    Discussion (2)

---

### About Dataset

Usability ⓘ  
10.00



<https://www.kaggle.com/datasets/yamqwe/tiktok-trending-tracks?datasetId=1780769>

# Data Description

**Number of Rows?**

6747

**Number of Variables**

23

**Target Outcomes**

"Popularity"

**Explanatory Variable**

"Danceability", "energy", "key", "speechiness",  
"acousticness", "instrumentalness", "liveness",  
"valence", "tempo", "duration\_mins", "mode", "genre"



# Predictive

We are primarily using predictive analytics, as our goal is to predict the popularity of a song based on various musical and track features.

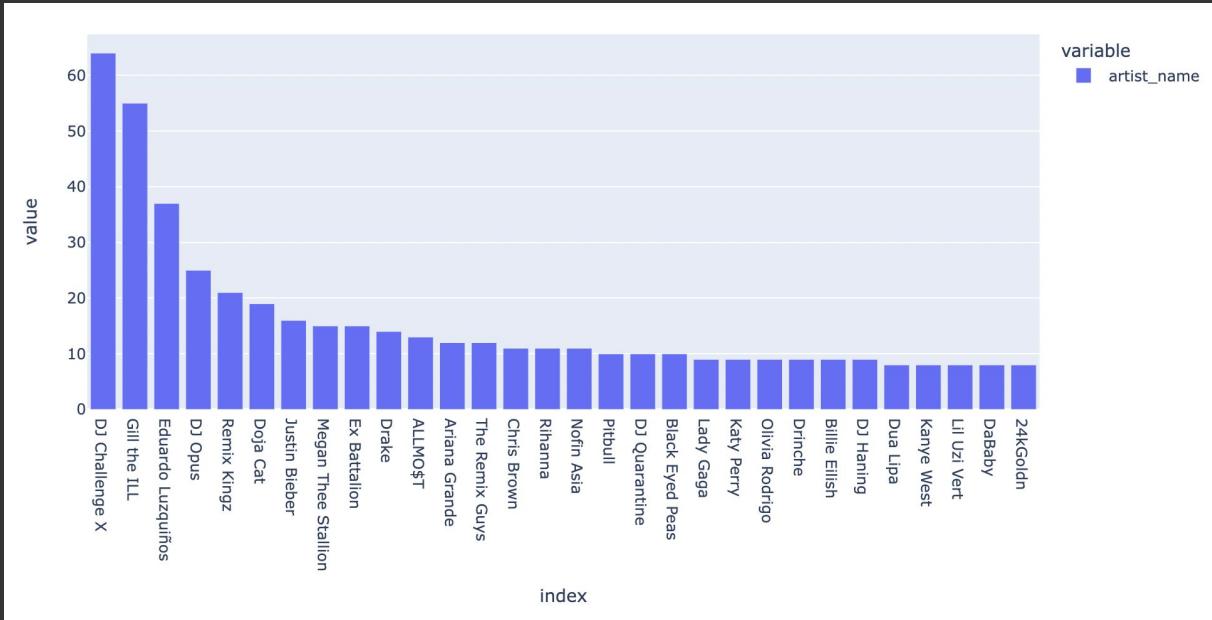


# Prescriptive

We cannot strictly interpret these relationships in a causal manner. Features are usually artist's creative choices.

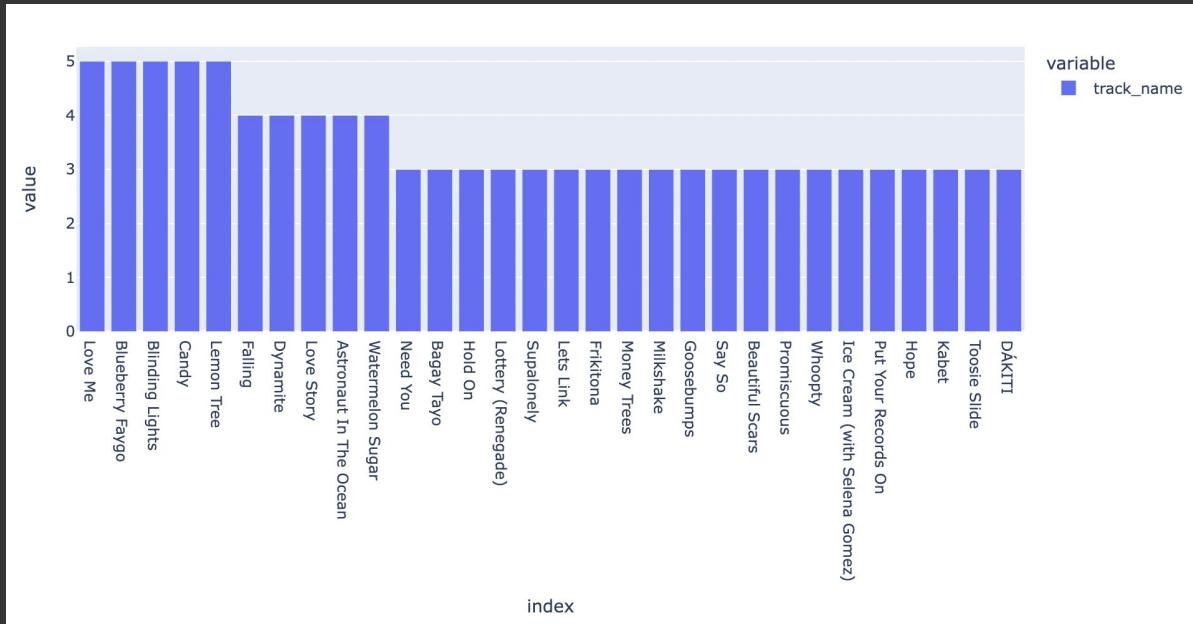
# Exploratory Data Analysis

## Top Artists



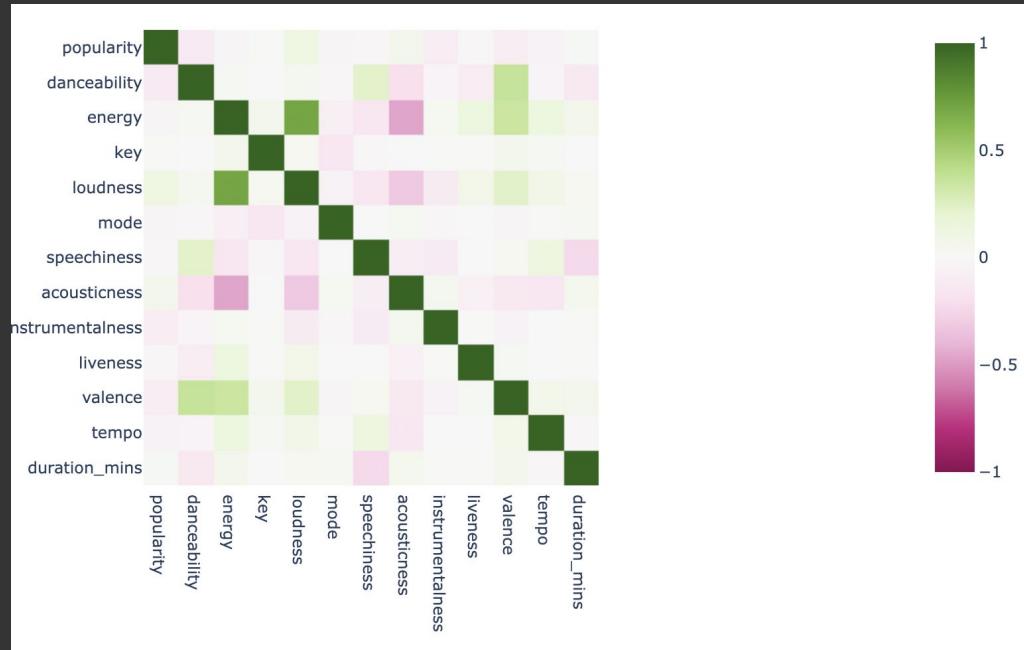
# Exploratory Data Analysis

## Top Tracks



# Exploratory Data Analysis

## Correlation



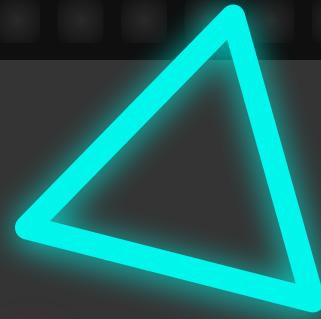
# Exploratory Data Analysis

## Variation inflation factor

	features	VIF
0	popularity	4.962520
1	danceability	21.607047
2	energy	17.543266
3	key	3.065895
4	mode	2.457687
5	speechiness	2.511407
6	acousticness	2.068911
7	instrumentalness	1.102366
8	liveness	2.770190
9	valence	8.258001
10	tempo	19.077360
11	duration_mins	10.536003



# Feature Engineering



0 1 0  
1 0 1  
0 1 0



## Categorical Variable Preprocessing:

`StringIndexer +  
OneHotEncoder`

To transform into binary sparse vectors

## Target Variable Categorization:

`Bucketizer`

To improve user experience

## Feature Scaling:

`StandardScaler`

To prevent one feature from dominating others

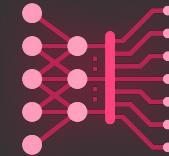
# Model Selection



**Multi-class  
Logistic Regression**



**Random Forest**



**One-vs-Rest (OVR)**  
**Base classifier: SVM**



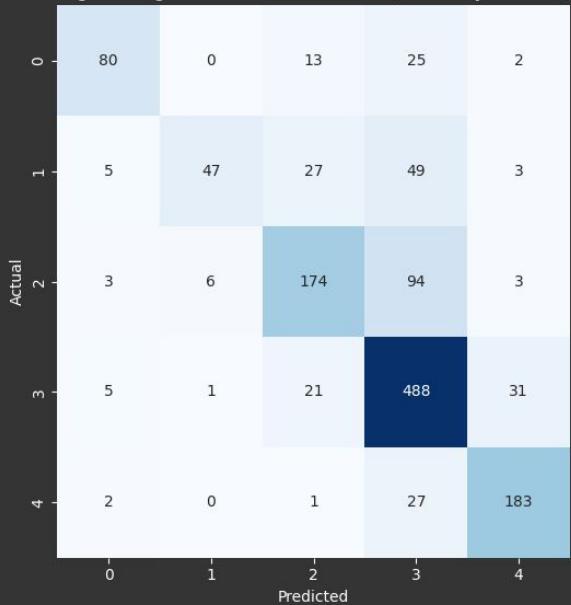
# 03

# Analytics results

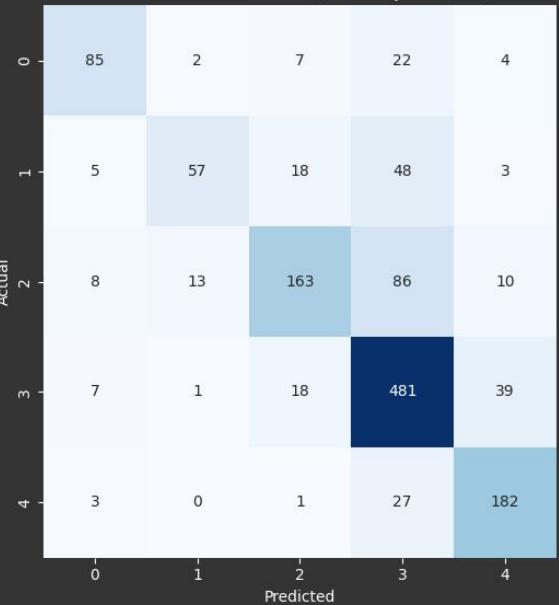
Analytics Insights

# Model Evaluation

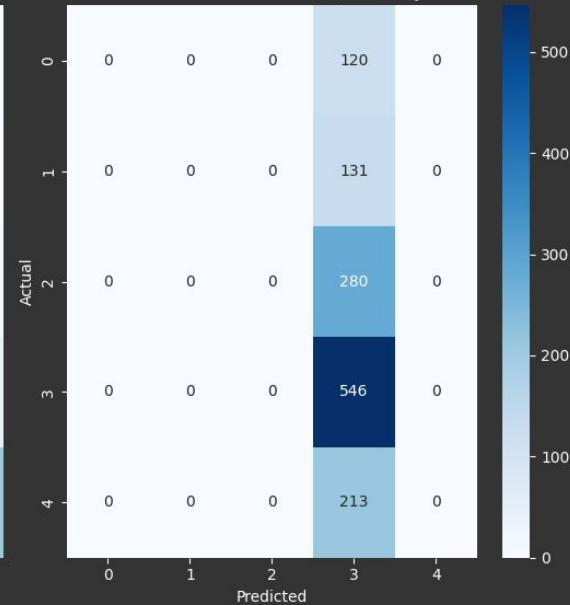
Logistic Regression Confusion Matrix (Accuracy: 75.35%)



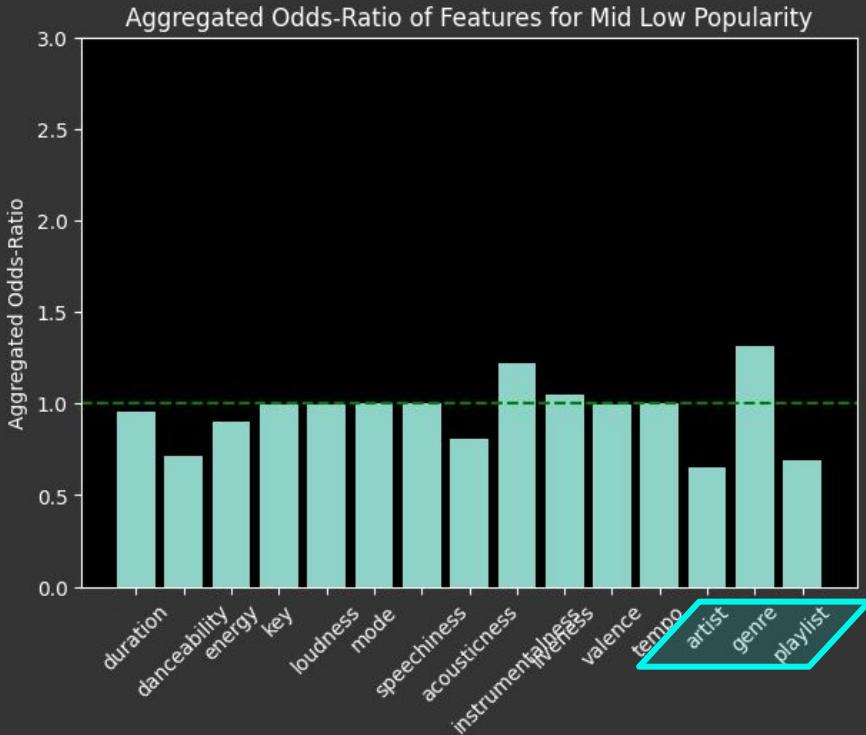
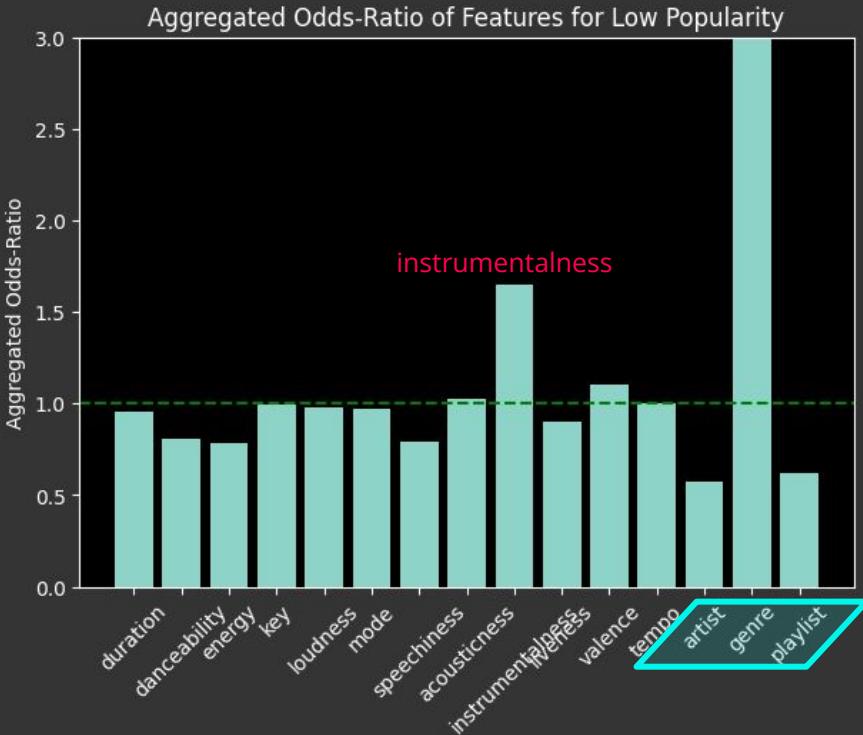
OVR Confusion Matrix (Accuracy: 75.04%)



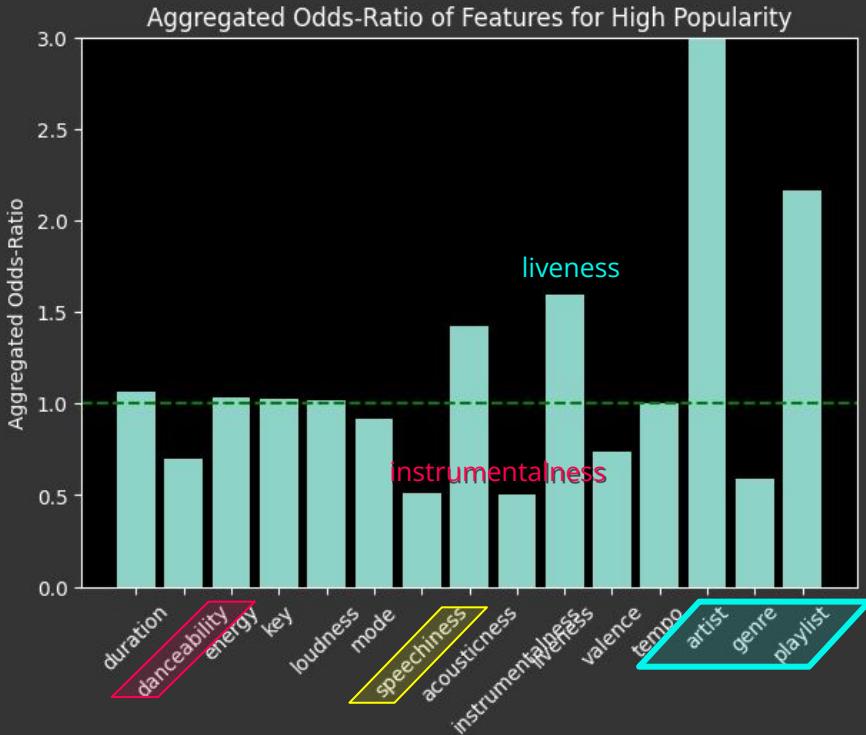
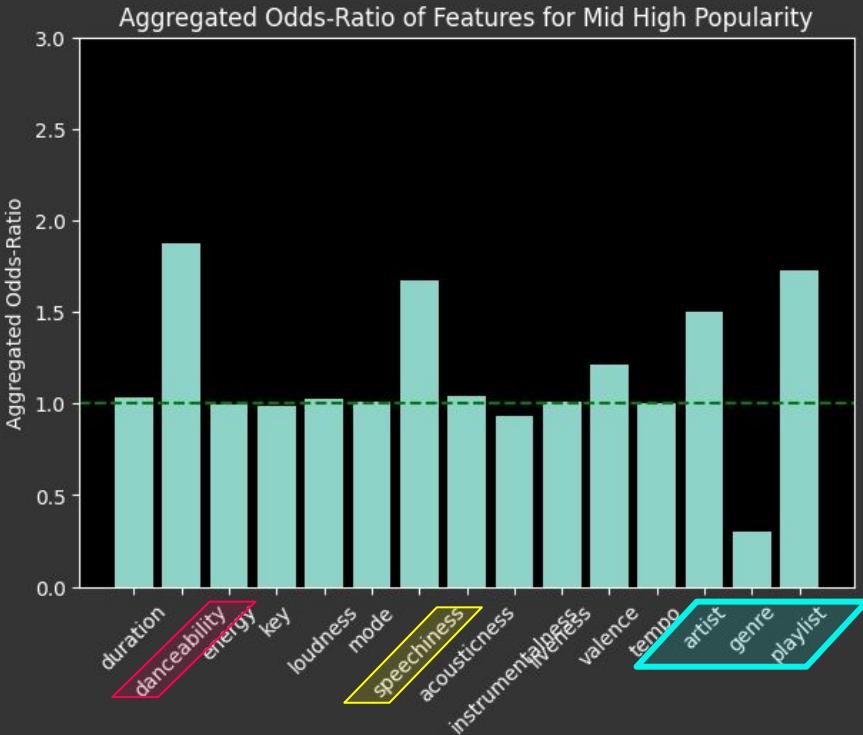
Random Forest Confusion Matrix (Accuracy: 42.33%)



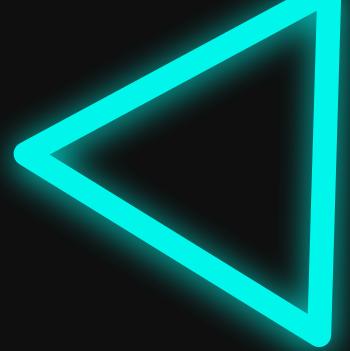
# Feature Exploration



# Feature Exploration



# Insights Summary



- Popularity is highly correlated with the genre, artist, and playlist of the song
- High Instrumentalness indicates low popularity and vice versa
- Danceability strongly correlates with Mid High Popularity, but not High Popularity
- Speechness is positively correlated with Mid High and High Popularity
- High Popularity indicates a positive correlation with Liveness



**Thank You**