HOME

TABLE OF CONTENTS

PLAYLIST

THANKS!

Spotify DATA Analysis

Team 6: Jewel Ling, Camilla Zhao, Katherine Wang, Tsubasa Lin, Ethan Liu, Hanway Chang

Mars Is a Cold Place
The 15th Planet

2:54

3:49

Meet Our Team!

Nicho Lin

Hanwei Chang

Katherine Wang

Ethan Liu

Jewel Ling

Camilla Zhao

Dataset

Spotify 1

A dataset spanning 2010-2020 of ~26,000 rows of ‘popular’ and ‘unpopular’ songs released up to March 2020; the target variable is “popular”

Mars Is a Cold Place
The 15th Planet

2:54

3:49

HOME

TABLE OF CONTENTS

PLAYLIST

THANKS!

Table of contents

Mars Is a Cold Place
The 15th Planet

2:54 3:49

HOME

TABLE OF CONTENTS

PLAYLIST

THANKS!

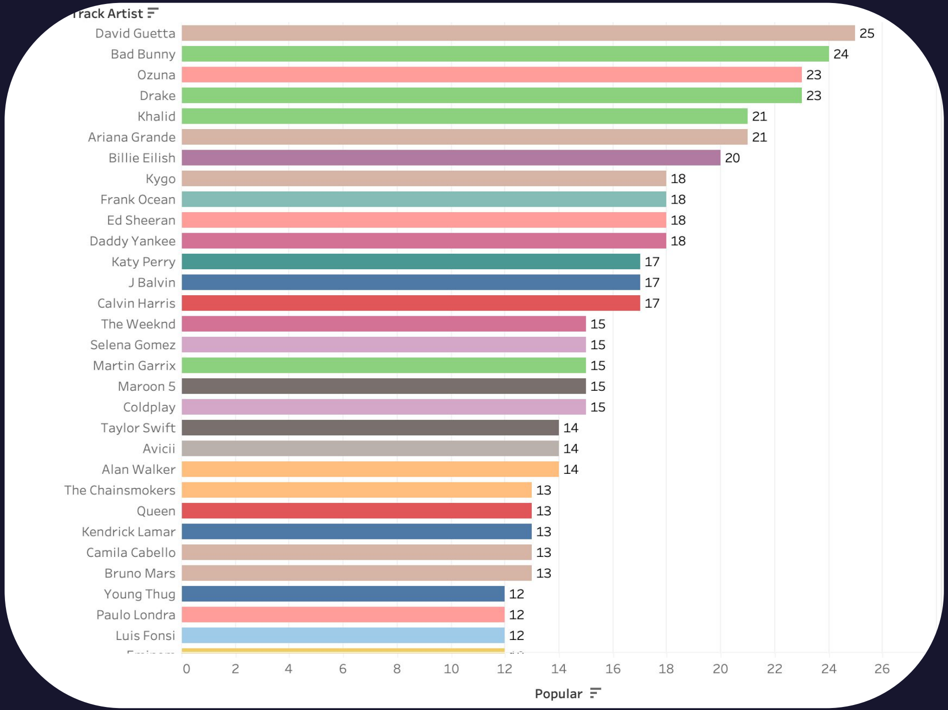
01

Popular Trends

Mars Is a Cold Place
The 15th Planet

2:54 3:49

Popularity by Artists

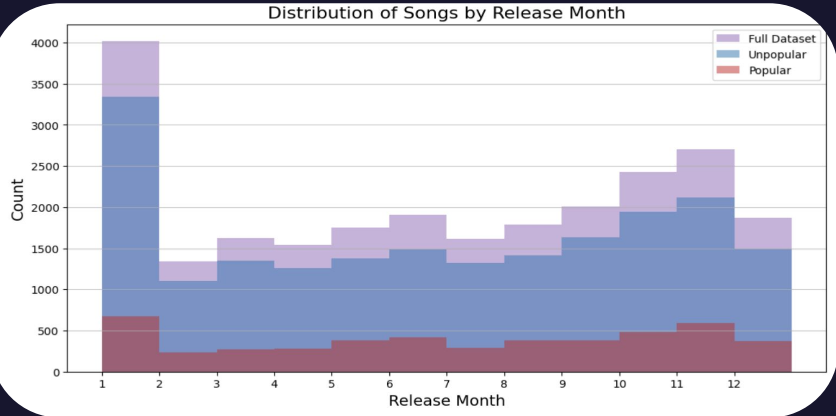


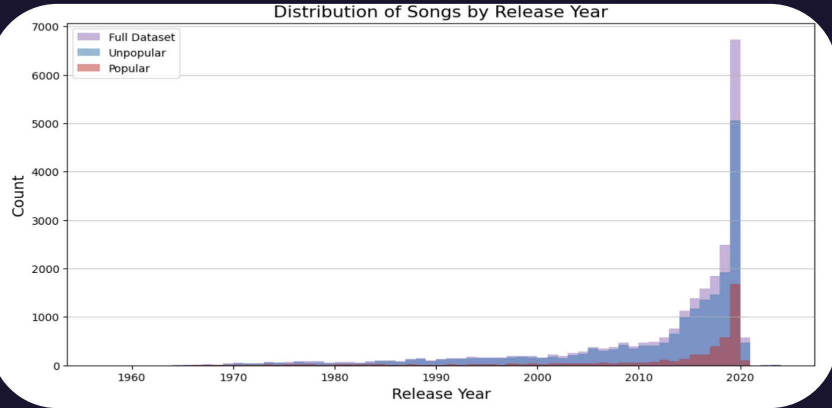
Mars Is a Cold Place
The 15th Planet

2:54

3:49

Popularity by Release Month & Year





Mars Is a Cold Place
The 15th Planet
2:54
3:49

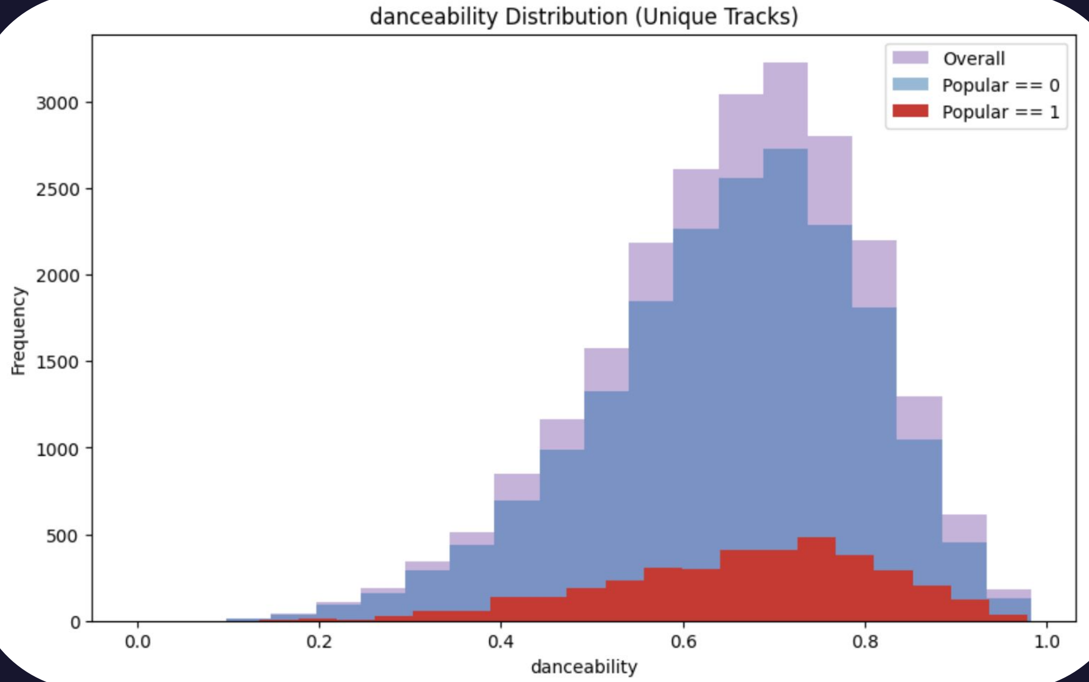
Popularity by Genre

Genre	Total	Popular	Unpopular	Popular Ratio
Pop	4099	1124	2975	27.42%
Latin	3756	911	2845	24.25%
R&B	4098	804	3294	19.62%
Rock	3567	693	2874	19.43%
Rap	4415	723	3692	16.38%
EDM	4434	455	3979	10.26%

Mars Is a Cold Place
The 15th Planet

2:54 3:49

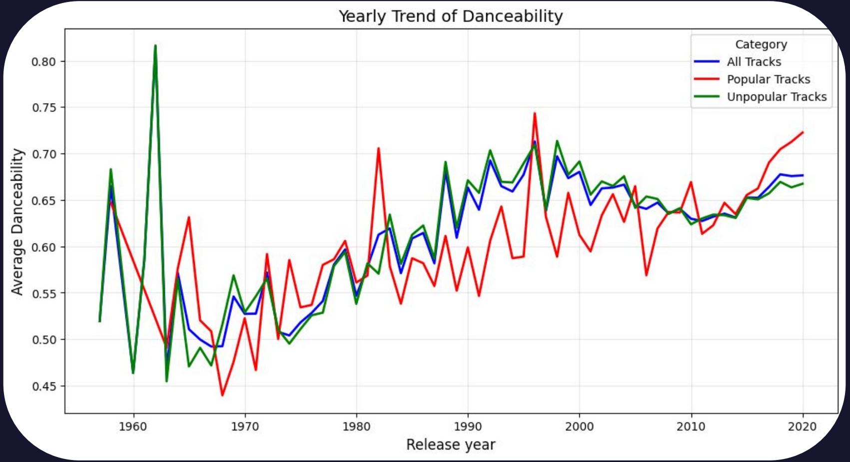
Popularity by Danceability



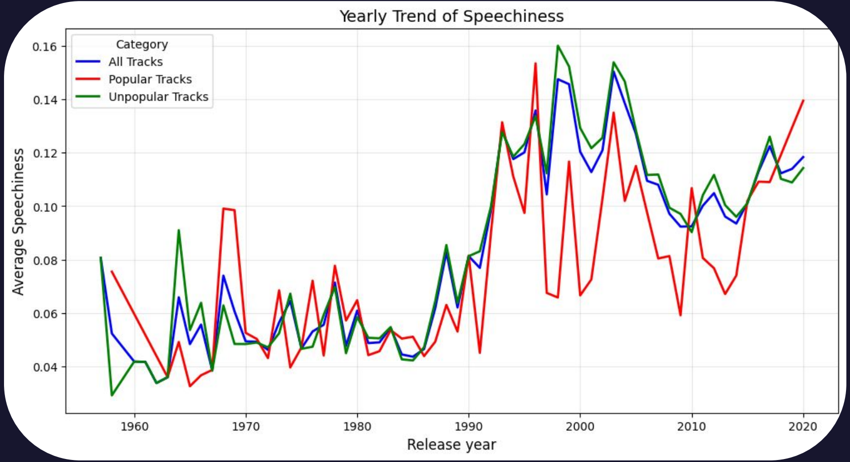
Mars Is a Cold Place
The 15th Planet
2:54
3:49

Other Exploration

Trend of Danceability



Trend of Speechiness



Mars Is a Cold Place
The 15th Planet

2:54

3:49

HOME

TABLE OF CONTENTS

PLAYLIST

THANKS!

02

Our Models

Mars Is a Cold Place
The 15th Planet

2:54

3:49

Data Cleaning

Findings

Duplicate Tracks_id

- Observation: Songs with multiple genres represented as individual rows.

Release Year Errors

- Observation: Large chunk of data set to 1905 and a well has value 0
- Likely Explanation: a placeholder for missing release years or data error.

Release Month Anomalies

- Observation: Disproportionate concentration in January.
- Likely Explanation: Placeholder for missing month data.

Other Missing Values

- Minimal single-digit missing values in a few columns.

release_year	count
0	21
1905	1360
1957	1
1958	1
1961	1

Mars Is a Cold Place
The 15th Planet

2:54 3:49

Data Cleaning

Correcting Data with Spotify API

- **Identify Errors:** Flag invalid info (e.g. release years == 1905)
- **Set Up API:** Create Spotify Developer account. Authenticate using Spotify API credentials (client_id and client_secret).
- **Fetch Data:** Query track info via **track_id** or search by **track name and artist.**
- **Validate Matches:** Ensure track and artist names align between API and dataset.
- **Update Dataset:** Correct release_year and release_month using API results.
- **Save Results:** Export corrected data to a CSV to prevent redundant API calls.

Spotify for Developers

Mars Is a Cold Place
The 15th Planet
2:54
3:49

Data Cleaning

Create New Columns

Combine Year and Month into a New Column (date)

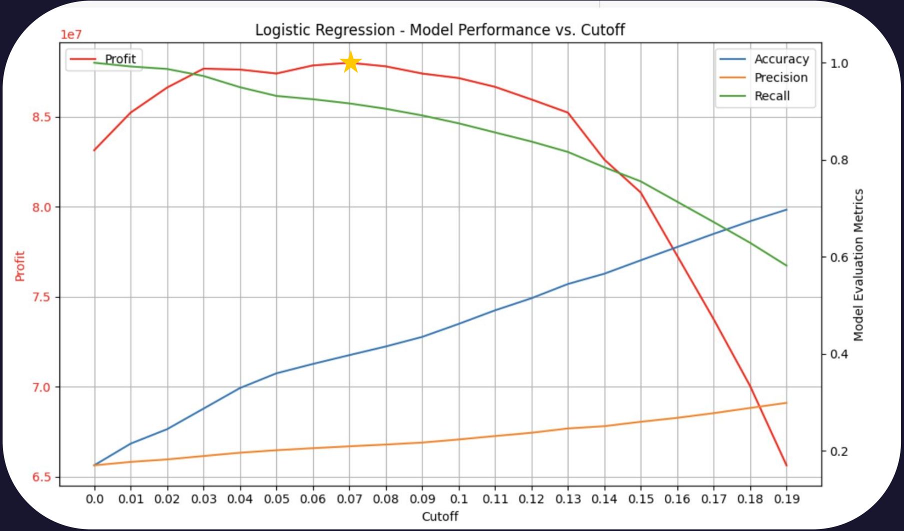
- Purpose: Enables time-series analysis to identify trends, such as popularity during specific periods.

Create a New Column (key_mode)

- Combines key and mode columns.
- In music theory, a key in a specific mode conveys unique meaning (e.g., C Major vs. C Minor), making it logical to combine them into a single entity.

Mars Is a Cold Place
The 15th Planet
2:54
3:49

Logistic Model Performance



Logistic Regression as a baseline model

Profit Equation = ($120K * TP) - ($10K * FP)

The highest profit, **$87,990,000** was achieved at a cutoff of 0.07, where accuracy = 39.70%, precision = 20.90%, and recall = 91.62%

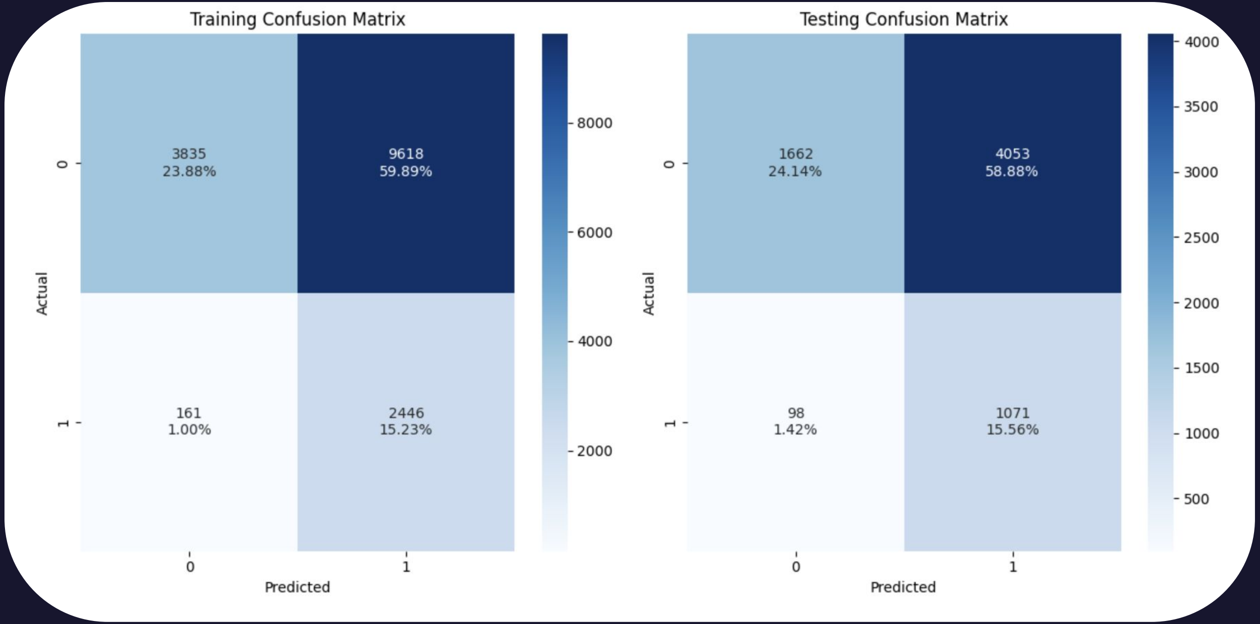
The graph pattern suggests that **recall** is the key metric we should be focusing on.

Mars Is a Cold Place
The 15th Planet

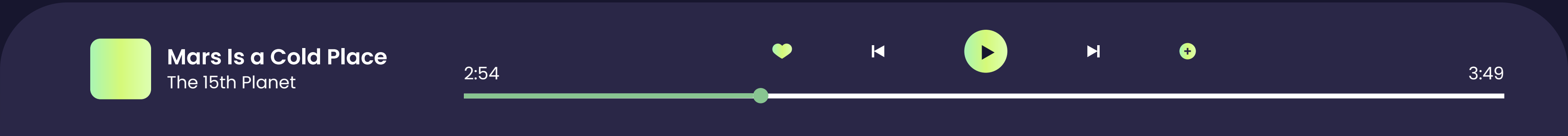
2:54

3:49

Logistic - Confusion Matrix



Confusion Matrix at 0.07 cutoff



Model Summary

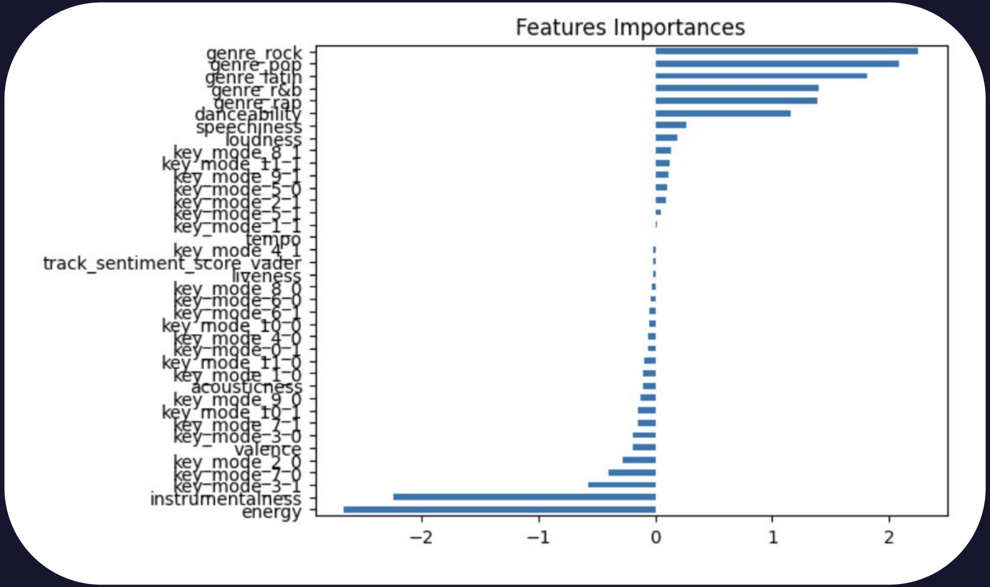
Model	Cutoff	Accuracy	Precision	Recall	Profit
Logistic Regression	0.07	0.397008	0.209016	0.916168	87990000
Decision Tree	0	0.169814	0.169814	1	83130000
Random Forest	0.09	0.425189	0.217356	0.917023	90040000
Bagging	0.05	0.344858	0.200788	0.958939	89900000
XGBoost	**0.08**	**0.429256**	**0.220195**	**0.928999**	**91860000**
Neural Network	0.06	0.444073	0.219739	0.89136	88040000

*Performance was obtained with optimal hyperparameters after tuning

Mars Is a Cold Place
The 15th Planet

2:54 3:49

Logistic – Inspecting the Features



Some prominent features: genres, danceability, energy, instrumentalness

Mars Is a Cold Place
The 15th Planet

2:54

3:49

HOME
TABLE OF CONTENTS
PLAYLIST

THANKS!

03

KPIs & Results

Mars Is a Cold Place
The 15th Planet
2:54
3:49

Best Final Model: XGBoost

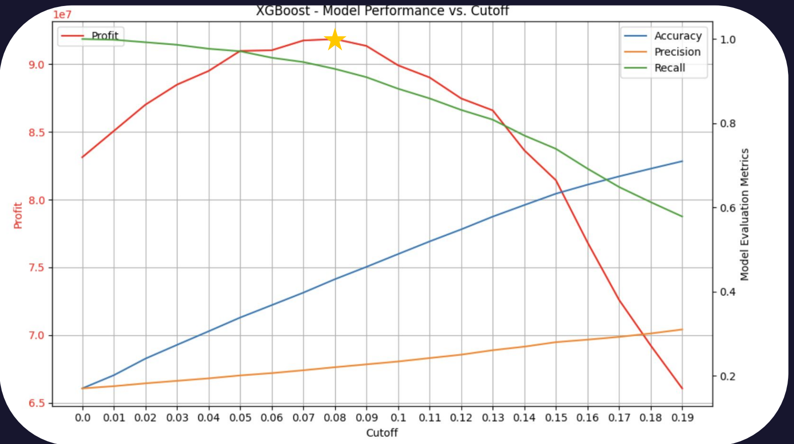
XGBoost stands out as the best final model, delivering optimal financial outcomes with balanced metrics.

Why?

- **Feature Handling:** Captured complex relationships in Spotify's dataset.
- **Optimized Threshold (0.08):** Balanced recall and precision, ensuring popular tracks were promoted while minimizing wasted resources.
- **Scalability:** Efficiently processed large data, optimizing predictions for financial outcomes.
- **Cost-Effective:** High recall avoided missing hit tracks, and precision ensured focused investments.

Best Score (Accuracy) for XGBoost: 0.8514321295143212

	Cutoff	Accuracy	Precision	Recall	AUC_ROC	Confusion Matrix	Profi
0	0.00	0.169814	0.169814	1.000000	0.738699	[[0, 5715], [0, 1169]]	83130000
1	0.01	0.201191	0.175120	0.998289	0.738699	[[218, 5497], [2, 1167]]	85070000
2	0.02	0.240558	0.181847	0.992301	0.738699	[[496, 5219], [9, 1160]]	87010000
3	0.03	0.273242	0.187785	0.986313	0.738699	[[728, 4987], [16, 1153]]	88490000
4	0.04	0.305491	0.193691	0.976903	0.738699	[[961, 4754], [27, 1142]]	89500000
5	0.05	0.338175	0.200636	0.970915	0.738699	[[1193, 4522], [34, 1135]]	90980000
6	0.06	0.367664	0.206165	0.955518	0.738699	[[1414, 4301], [52, 1117]]	91030000
7	0.07	0.397298	0.212909	0.945252	0.738699	[[1630, 4085], [64, 1105]]	91750000
8	0.08	0.429256	0.220195	0.928999	0.738699	[[1869, 3846], [83, 1086]]	91860000
9	0.09	0.458600	0.226943	0.909324	0.738699	[[2094, 3621], [106, 1063]]	91350000
10	0.10	0.488960	0.233734	0.881950	0.738699	[[2335, 3380], [138, 1031]]	89920000
11	0.11	0.519030	0.241928	0.858854	0.738699	[[2569, 3146], [165, 1004]]	89020000
12	0.12	0.547647	0.249936	0.831480	0.738699	[[2798, 2917], [197, 972]]	87470000
13	0.13	0.578007	0.260618	0.808383	0.738699	[[3034, 2681], [224, 945]]	86590000
14	0.14	0.605462	0.269036	0.770744	0.738699	[[3267, 2448], [268, 901]]	83640000
15	0.15	0.632481	0.279702	0.739093	0.738699	[[3490, 2225], [305, 864]]	81430000
16	0.16	0.653835	0.285664	0.692044	0.738699	[[3692, 2023], [360, 809]]	76850000
17	0.17	0.673591	0.292213	0.648417	0.738699	[[3879, 1836], [411, 758]]	7260



Question 2b

Old Profit Formula = ($120K * TP) - ($10K * FP)

What if a 20% chance of unpopular songs get popular after we promote it?

- What songs to promote?
  - **FPs**, because they were relatively more likely to be popular than the rest of the songs (i.e. songs classified as unpopular)

New Profit Formula = ($120K * TP) + (**$120K * 0.2** * FP - **$10K * 0.8** * FP) - (Promotion Cost/Song * FP)
= ($120K * TP) + ($16K * FP) - (Promotion Cost/Song * FP)

If Promotion Cost < $16K:
Lower cutoff to include more songs as expected return for each promoted FP would be positive.

Else if Promotion Cost >= $16K:
Raise the cutoff with caution to avoid losses and optimize profit.

Mars Is a Cold Place
The 15th Planet
2:54
3:49

HOME

TABLE OF CONTENTS

PLAYLIST

THANKS!

04

Recommendations

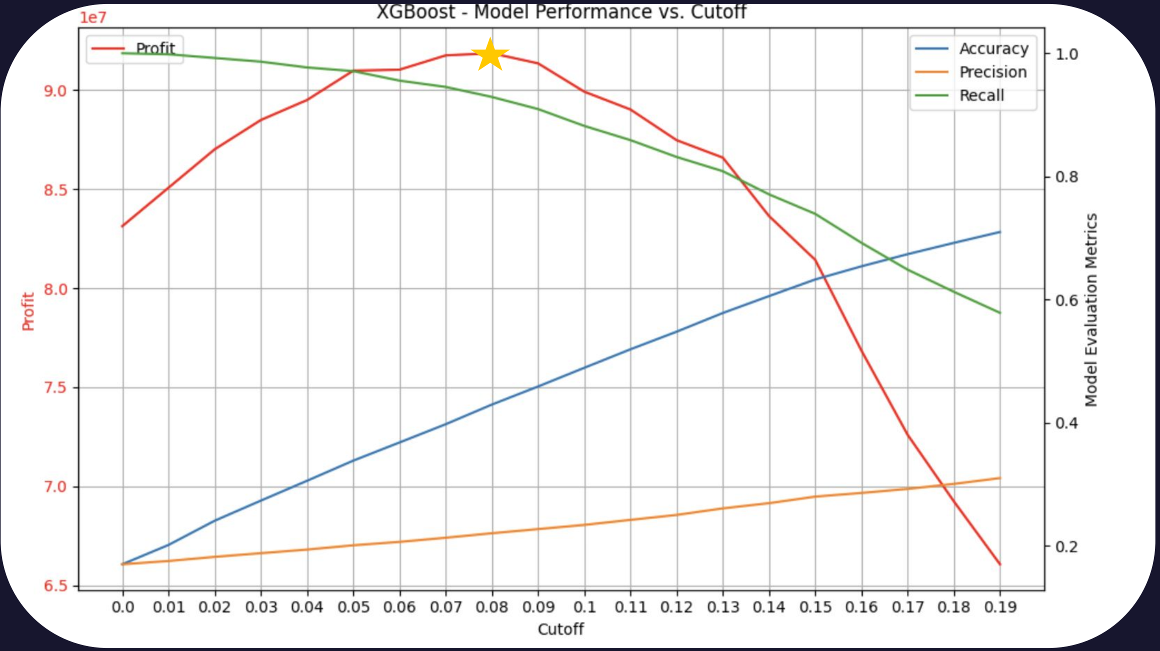
Mars Is a Cold Place

The 15th Planet

2:54

3:49

Recommendations Based on XGBoost Model



Real-Life Implementation for UMG

- **Feature-Based Predictions:** Use more comprehensive data like streaming metrics, fan engagement, and sentiment analysis to predict hit tracks.
- **Targeted Promotion:** Prioritize marketing for tracks predicted to succeed (e.g., regional campaigns for Olivia Rodrigo in emerging markets).
- **Resource Optimization:** Allocate higher budgets to high-probability hits, avoiding overpromotion of low-potential tracks.
- **Collaboration Strategy: I**dentify data-backed artist pairings (e.g., emerging talent with top performers like Billie Eilish)

Mars Is a Cold Place
The 15th Planet

2:54

3:49

Recommendations for Universal Music

Data-Driven Music Production

Employing the XGBoost model during the music creation process to **strategically balance song features**, maximizing the likelihood of producing hits.

Strategic Playlist Curation

Using the XGBoost model to **predict the popularity and profitability** of various **bundled songs** is an ideal approach for optimizing curated playlists.

Optimized Song Promotion

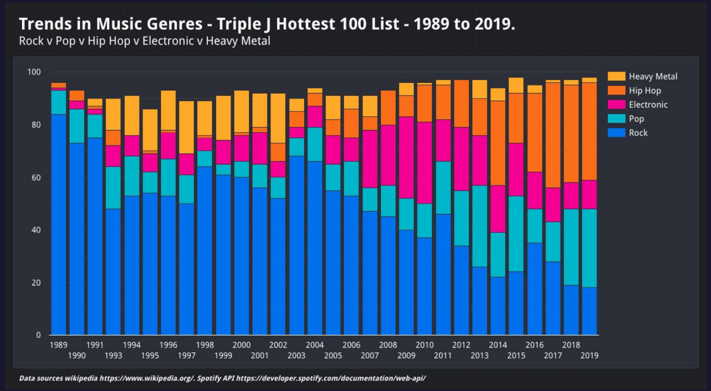
Prioritize tracks with high predicted **popularity** probabilities for promotional efforts;
Use **genre and feature** insights to design tailored campaigns.

Collaboration Strategies

Identify **artist collaborations** based on complementary styles or shared audience demographics.

Mars Is a Cold Place
The 15th Planet
2:54
3:49

Limitations & Mitigations



01 Dynamic Market Trends

The model relies on historical Spotify data, which may not fully capture rapidly changing audience preferences.

Mitigation: Regularly retrain the model with new data to reflect evolving trends.

02 Limited Creativity

While data can guide production, artistic creativity and experimentation remain vital for breakthrough success.

Mitigation: Use insights as a supplement, not a substitute, for artistic intuition.

“What Was I Made For?” Bongo Cat Cover “Meow” ver. Became a hit

Mars Is a Cold Place
The 15th Planet
2:54
3:49

01

Incorporate Real-Time Metrics

Next Steps

Regional Customization

02

03

Feedback Loop

Mars Is a Cold Place
The 15th Planet
2:54
3:49

HOME

TABLE OF CONTENTS

PLAYLIST

THANKS!

Thank you!

Mars Is a Cold Place
The 15th Planet

2:54

3:49

05

Appendix

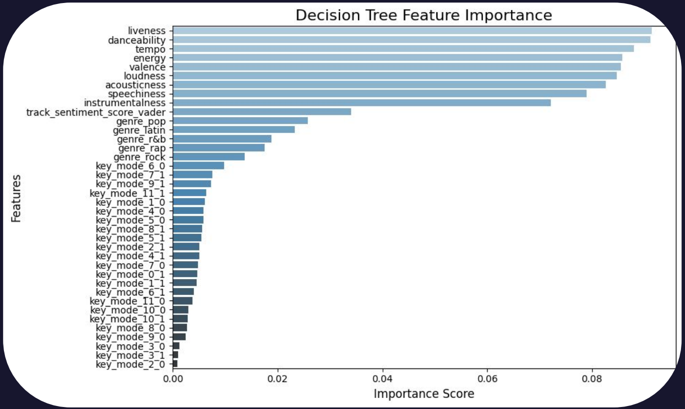
Mars Is a Cold Place

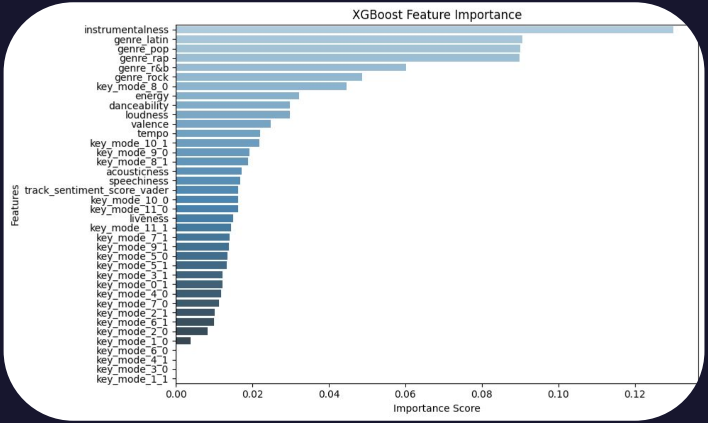
The 15th Planet

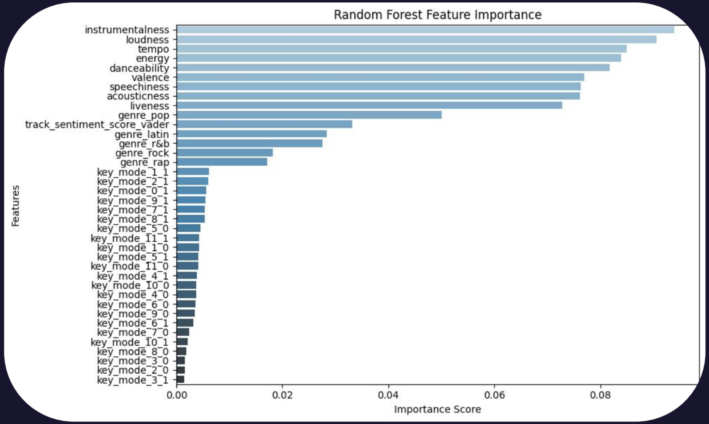
2:54

3:49

Feature Importance







Main takeaway: Key_mode combination has minimal predictive power on a song's popularity.