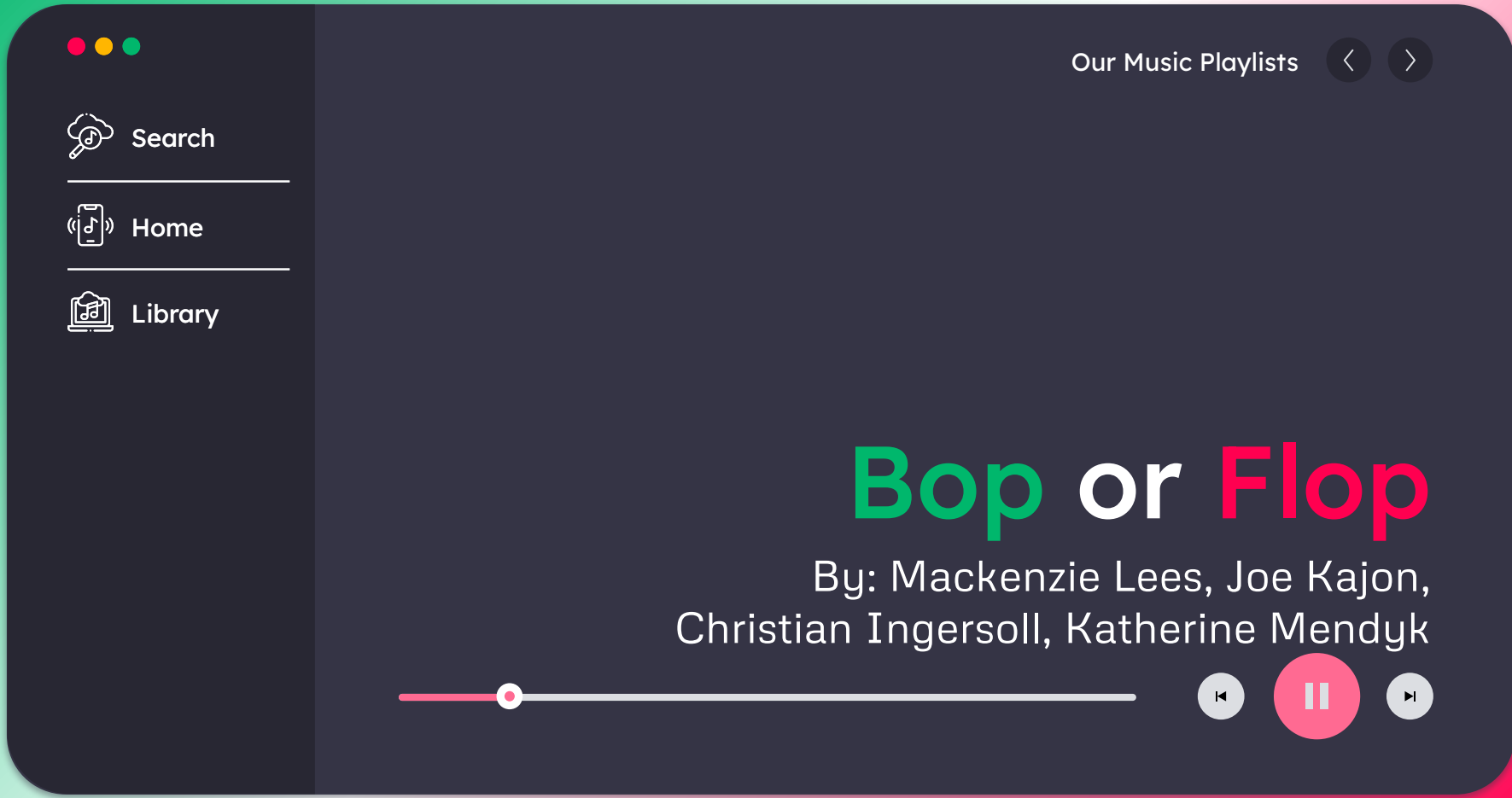

Search
Home
Library
Our Music Playlists
Bop or Flop
By: Mackenzie Lees, Joe Kajon,
Christian Ingersoll, Katherine Mendyk

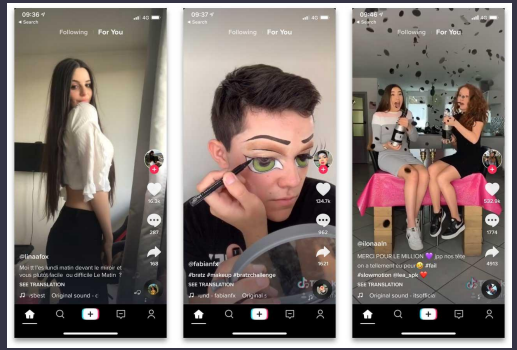
Agenda

- Scope
- Data
- Methods & Models
  - Sentiment analysis, Topic modeling, Predictive modeling
- Results
- Conclusion
- Impact & Recommendation


Search
Home
Library
Scope/Data

Background

- Features contributing to a song's success across different platforms and genres
- Genres: Pop, Rock, Latin, Rap, EDM, and R&B
- Focus on identifying the distinguishing elements that resonate with listeners on Spotify and TikTok
- TikTok: Social Media app



Scope

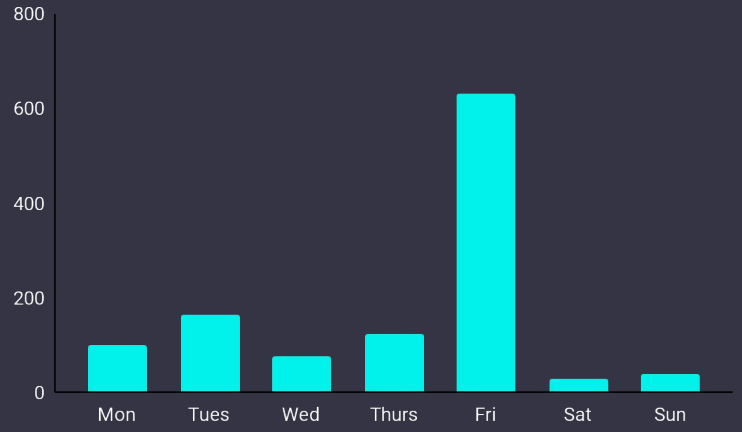
Identify key features that predict song's success and compare how these features differ across platforms (Spotify and TikTok) and by genre.

Data

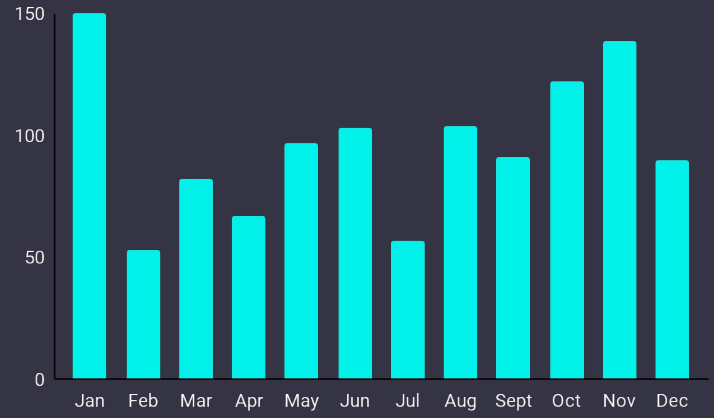
- Genre
- Characteristics of song
  - Ex: danceability, speechiness, mode
- Artist characteristics
  - Ex: # of streams, # of tracks
- Temporal data from release date
- No artist data for Tiktok

Exploratory Data Analysis - Seasonality

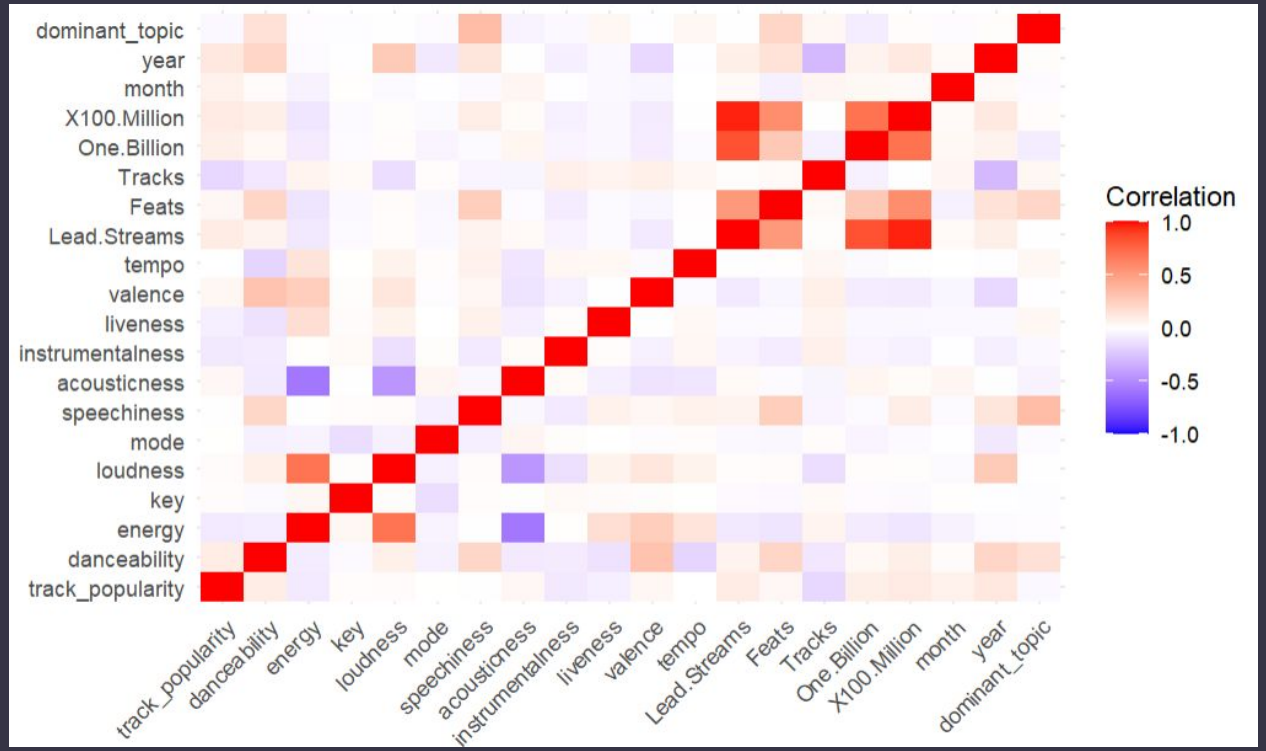
Popular Songs Released By Day of Week



Popular Songs Released By Month



Correlation Matrix



Assumptions/limitations

Popularity

Above the third quartile within a given genre

Repeated Songs

Keep max popularity score for repeated songs and remove other rows

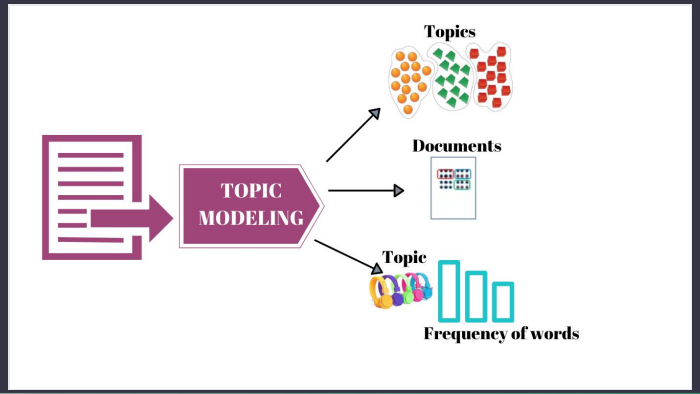
N/A rows

Removing rows with N/A in the release date

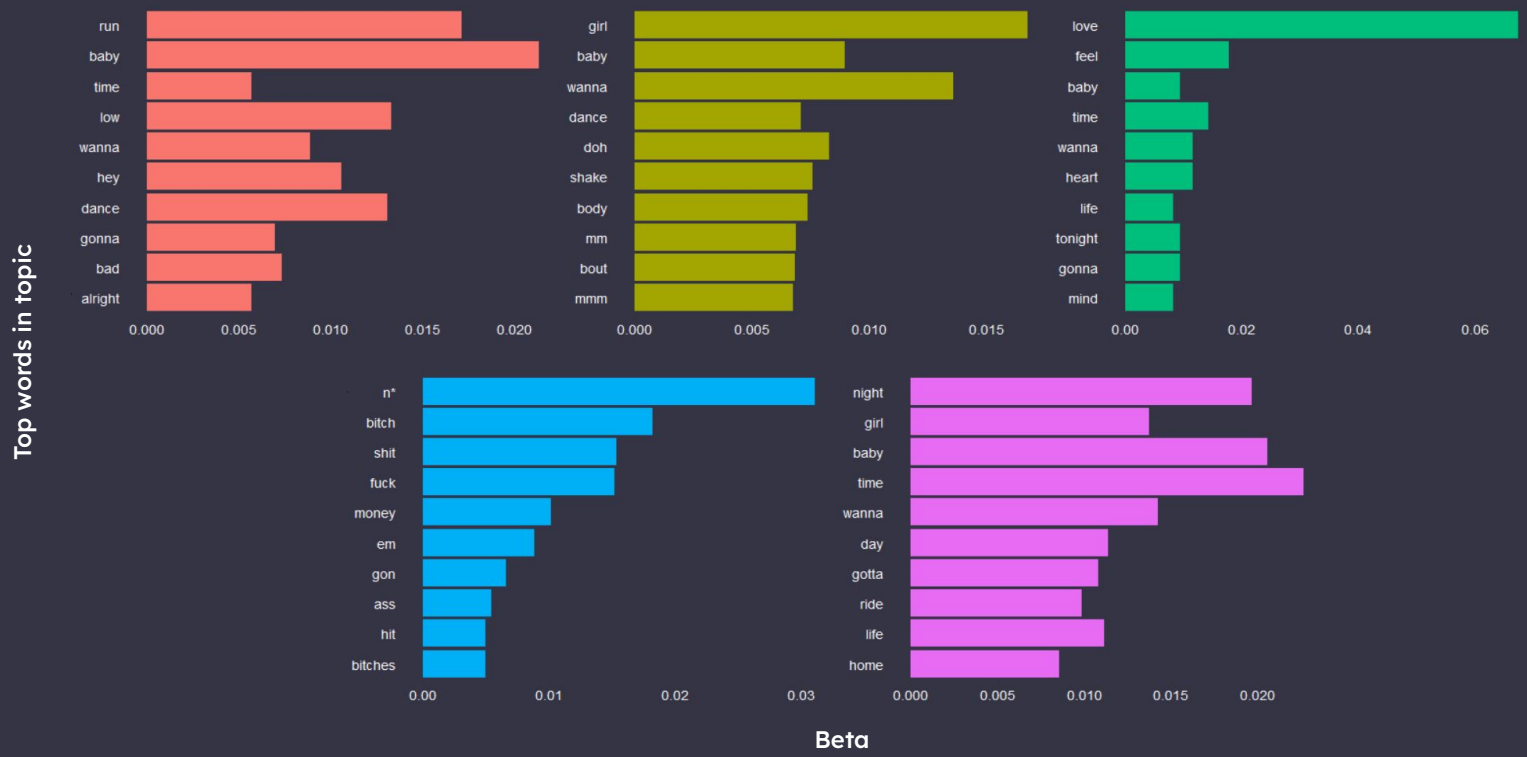
Methods and Models

Feature Engineering - LDA

- Latent Dirichlet Allocation (LDA)
  - Topic Modeling
  - Song lyrics into 5 main topics
  - Not necessarily significant features on both Spotify and TikTok



Feature Engineering - LDA



Feature Engineering - Bing Sentiment Analysis

Negative:

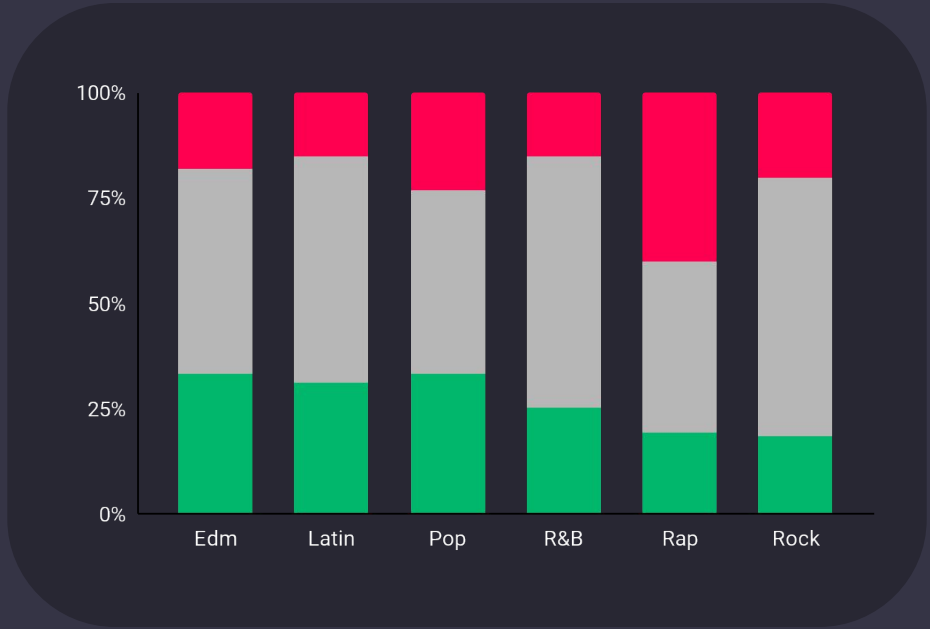
Rap, pop

Neutral:

Rock,R&B

Positive:

Edm, Pop



Feature Engineering - NRC Sentiment Analysis

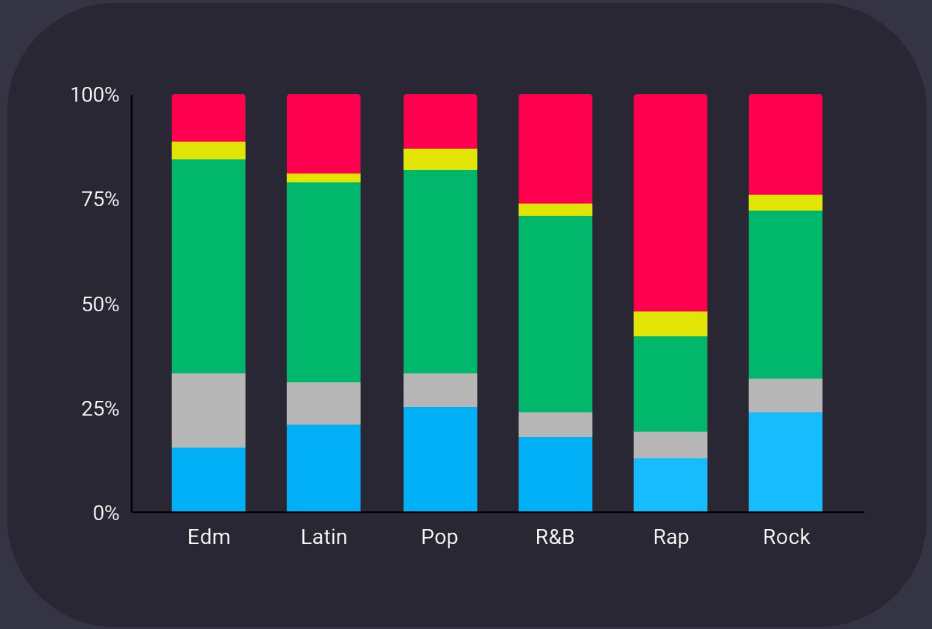
Anger
Rap, R&B

Disgust
Pop, Rap

Happy
EDM, Pop

Neutral
EDM, Rock

Sad
Pop, Rock



Predictive Models

01
Logistic Regression
Simplicity and Interpretability; Baseline

02
CART
Non-linear relationships and Interpretability

03
Random Forest
Robust; Feature Importance evaluation

04
XGBoost
High performance and Robust

Model Evaluation

- Initial Goal -> Balance interpretability and performance
- Initial metric of focus -> accuracy
- Due to imbalanced data, shifted focus to AUC

Selected model:

XGBoost

Increased AUC outweighed

lesser interpretability

Challenges

- Initially - > predict popularity overall
  - Improvement when split by genre
- Genres change over years
- Within one genre can be several subgenres
- EDM and Latin limited rows


Search
Home
Library
Results

Results

	Genre	Model	AUC	Top Variable Importance
1	Rap	XGBoost	0.8357	Tracks/X100.Million/energy/danceability/feats
2	R&B	XGBoost	0.7199	Tracks/speechiness/Lead.Streams/tempo/ danceability
3	Pop	XGBoost	0.7127	One.Billion/Lead.Streams/energy/Tracks/Feats
4	Rock	XGBoost	0.6487	Lead.Streams/Tracks/energy/danceability/ acousticness
5	EDM	XGBoost	0.6485	Loudness/speechiness/acousticness/Tracks/ valence
7	Latin	XGBoost	0.6018	Tracks/loudness/danceability/Feats/valence

TikTok

xgbTree variable importance

	Overall
loudness	0.16203
tempo	0.13839
speechiness	0.11569
acousticness	0.10986
valence	0.10140
liveness	0.10009
energy	0.09341
danceability	0.08158
instrumentalness	0.06439
mode	0.01681
key	0.01634

- Top predictors: loudness, tempo, speechiness
- XGBoost, like Spotify
- No Artist data
- AUC: 0.6676

Spotify versus Tikok

- Similar predictors with Spotify: Loudness, speechiness, acousticness
- Energy and danceability surprisingly low in variable importance

Results - Insights

- XGBoost was the top performing model
- Most Important variables were artist-specific statistics
  - More popular artists are more likely to have more popular songs
- Different genres/platforms emphasize different characteristics
  - EDM-> Loudness
  - Latin -> Danceability

Impact/ Recommendations

- Producers and Artists should focus on the most important predictors within their respective genres
  - Pop -> energy
  - Rock -> energy and danceability
- Tiktok targeted - > loudness and tempo

Search

Home

Library

Thank you

Any Questions?