

Algo-Rhythms



The (data) science
behind popular songs &
song recommenders on
Spotify

Ted Geh

01

Introduction

02

Spotify Data

03

Data Exploration

04

Modelling Popularity

05

Song Recommender

06

Conclusion

The Economics of Spotify



Users

345 Million Users
(Premium: 155 Million)



Market Share

2020: 32%
(Apple Music: 18%)



Track Count

70 million tracks



Demographics

55% of users
below age 35

Spotify Dataset

Consisting of ~175,000 Tracks



Spotify Dataset

13 Key Features

Musical Features (9)

- Acousticness
- Danceability
- Energy
- Instrumentalness
- Liveness
- Loudness
- Speechiness
- Tempo (BPM)
- Valence (How happy a song sounds)



Song Features (4)

- Popularity (Target Variable)
- Duration
- Song ID / Name / Artist
- Year

Modern Problems

- What are the ideal blend of features characteristic of popular songs?
- Song Recommender Based on Musical Features of a genre

01

Introduction

02

Spotify Data

03

Data Exploration

04

Modelling Popularity

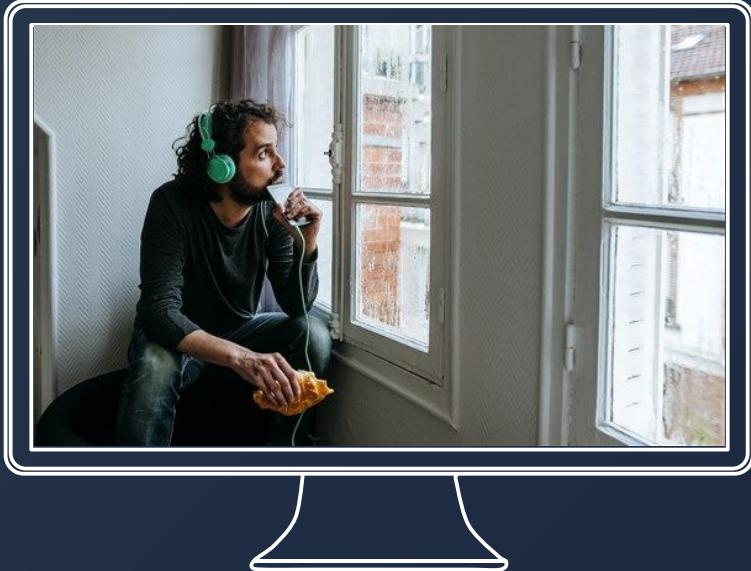
05

Song Recommender

06

Conclusion

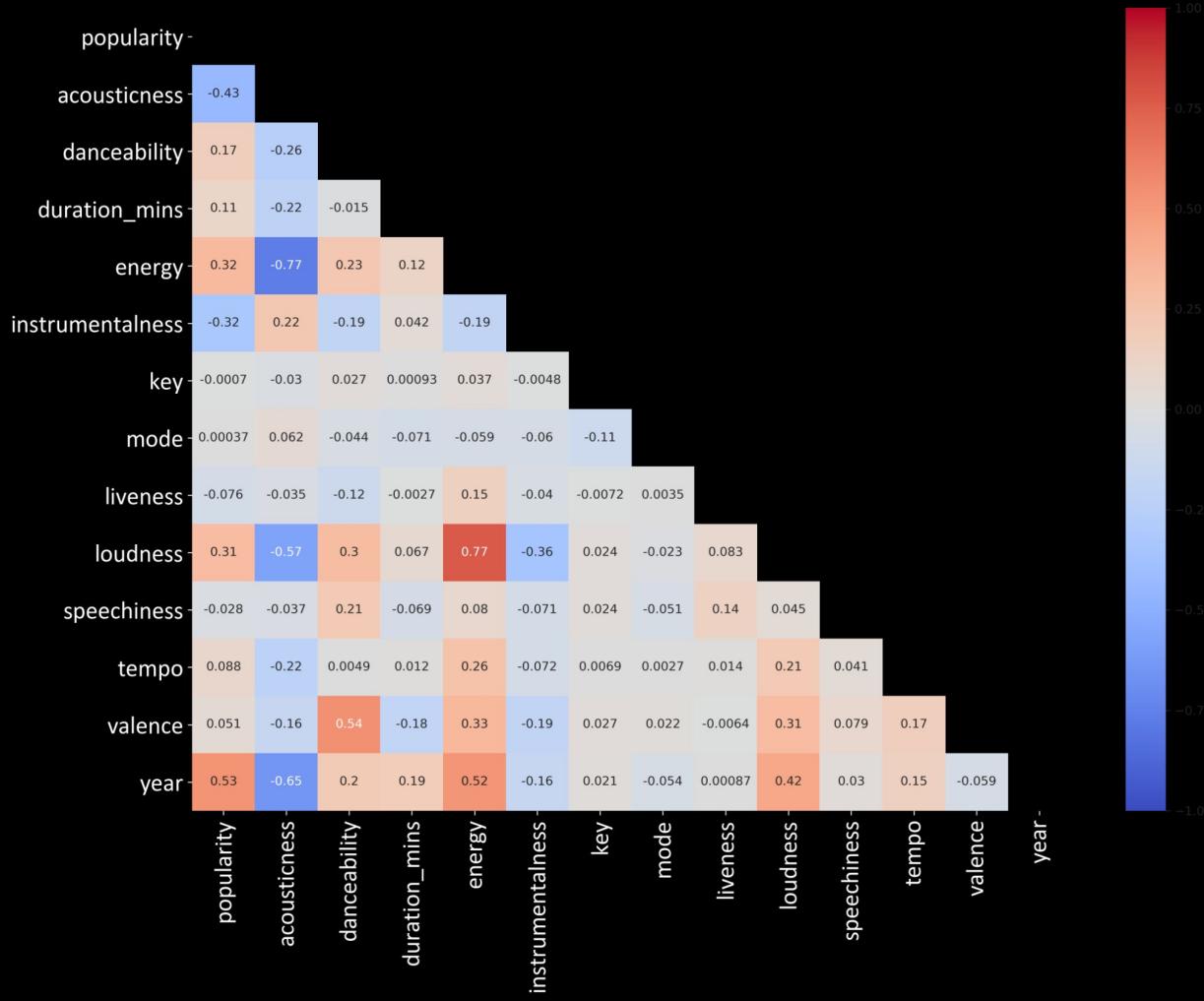
Challenges #1



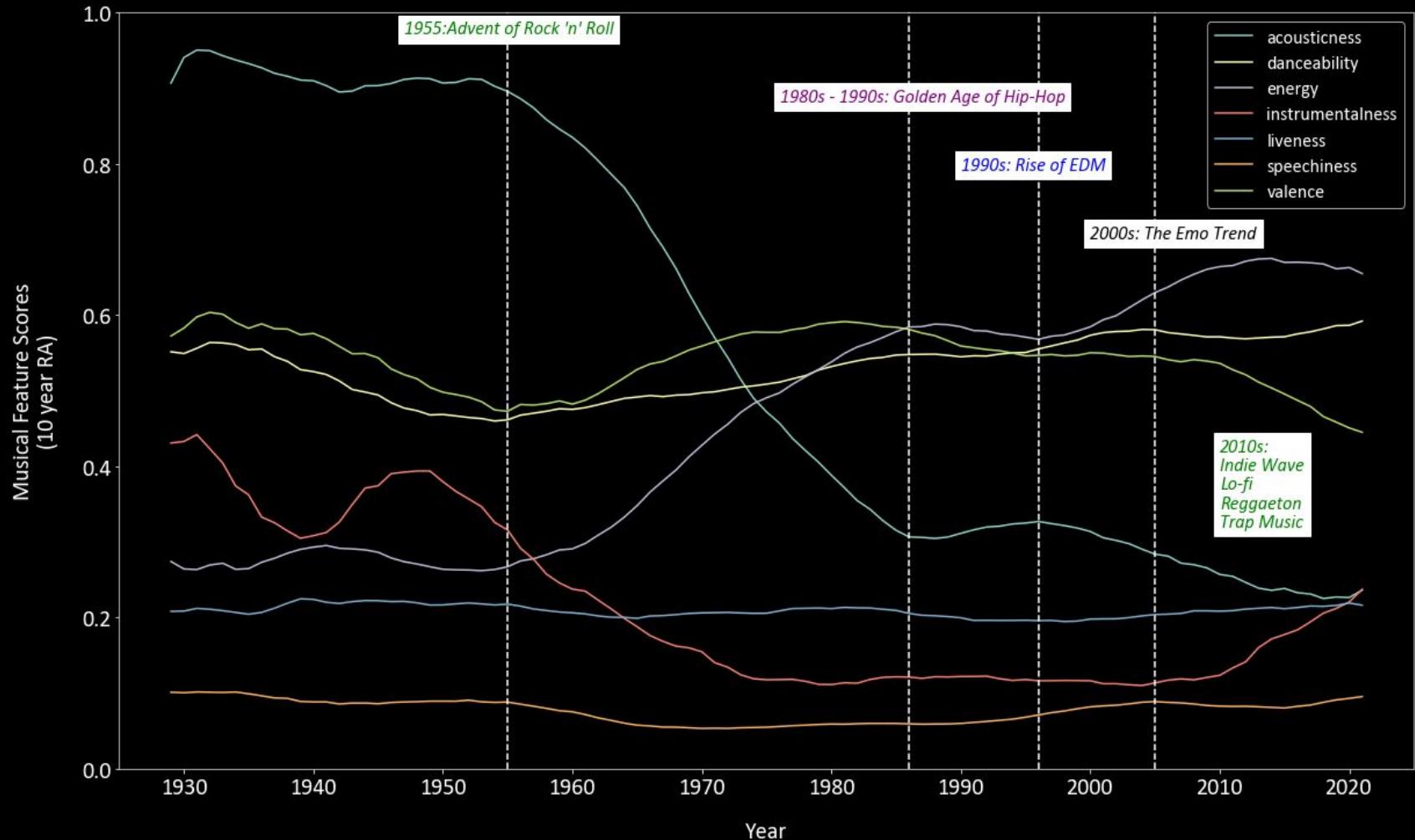
Challenges:

- Audiobook & podcasts
- Duplicated songs
 - ID
 - Same Musical Features
- Track duration (up to 90 mins)

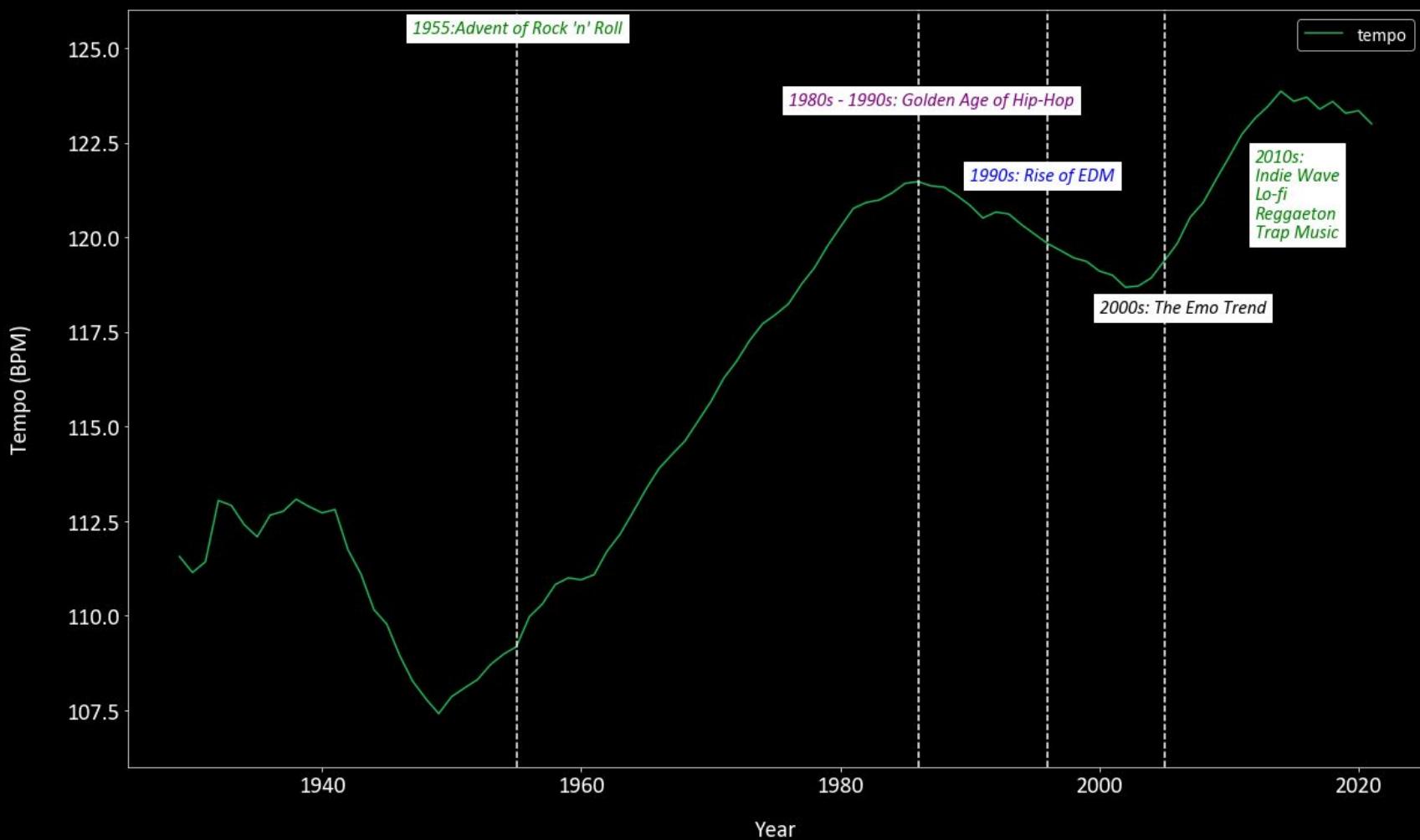
Spotify Features Correlation Analysis



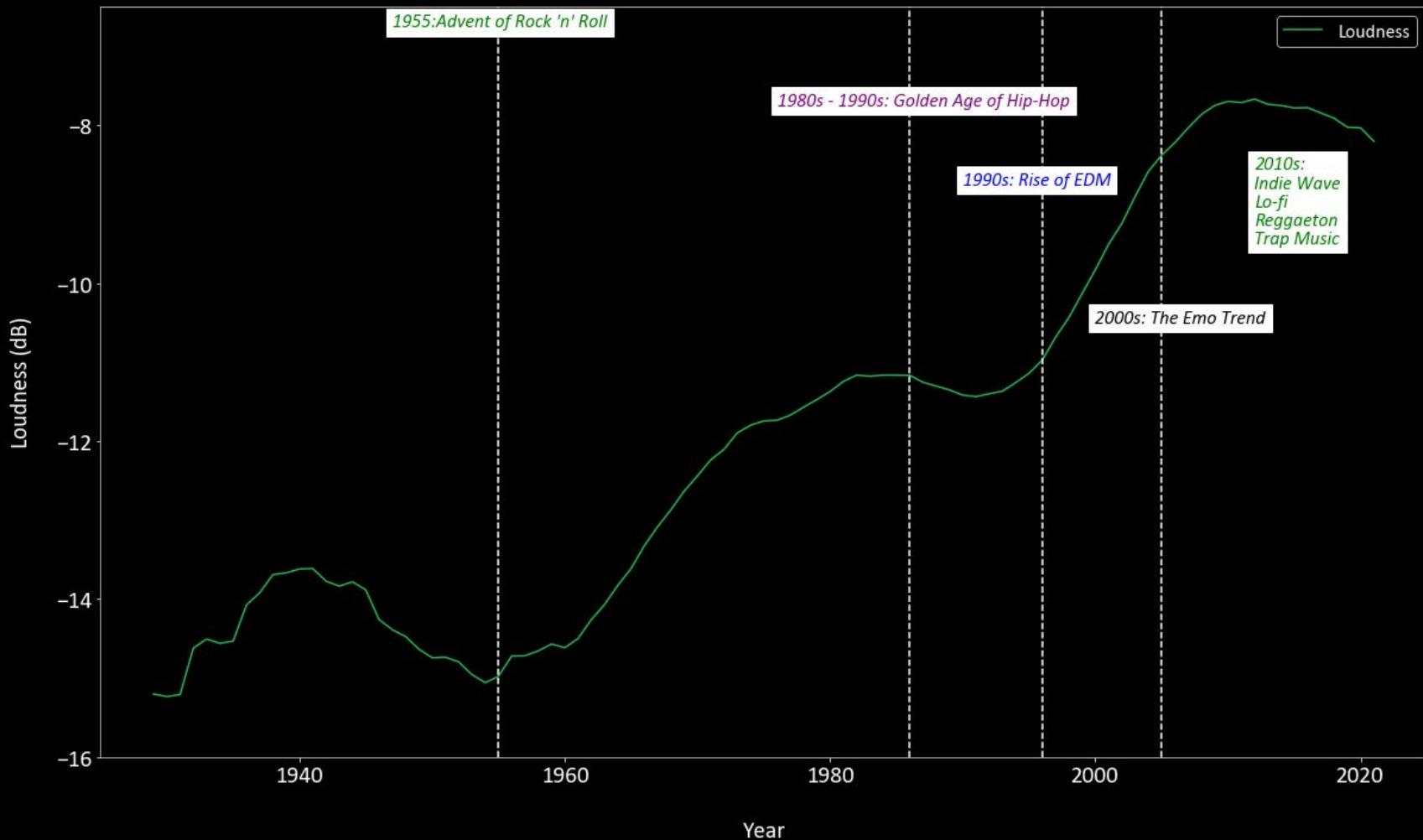
Evolution of Musical Features Over Time



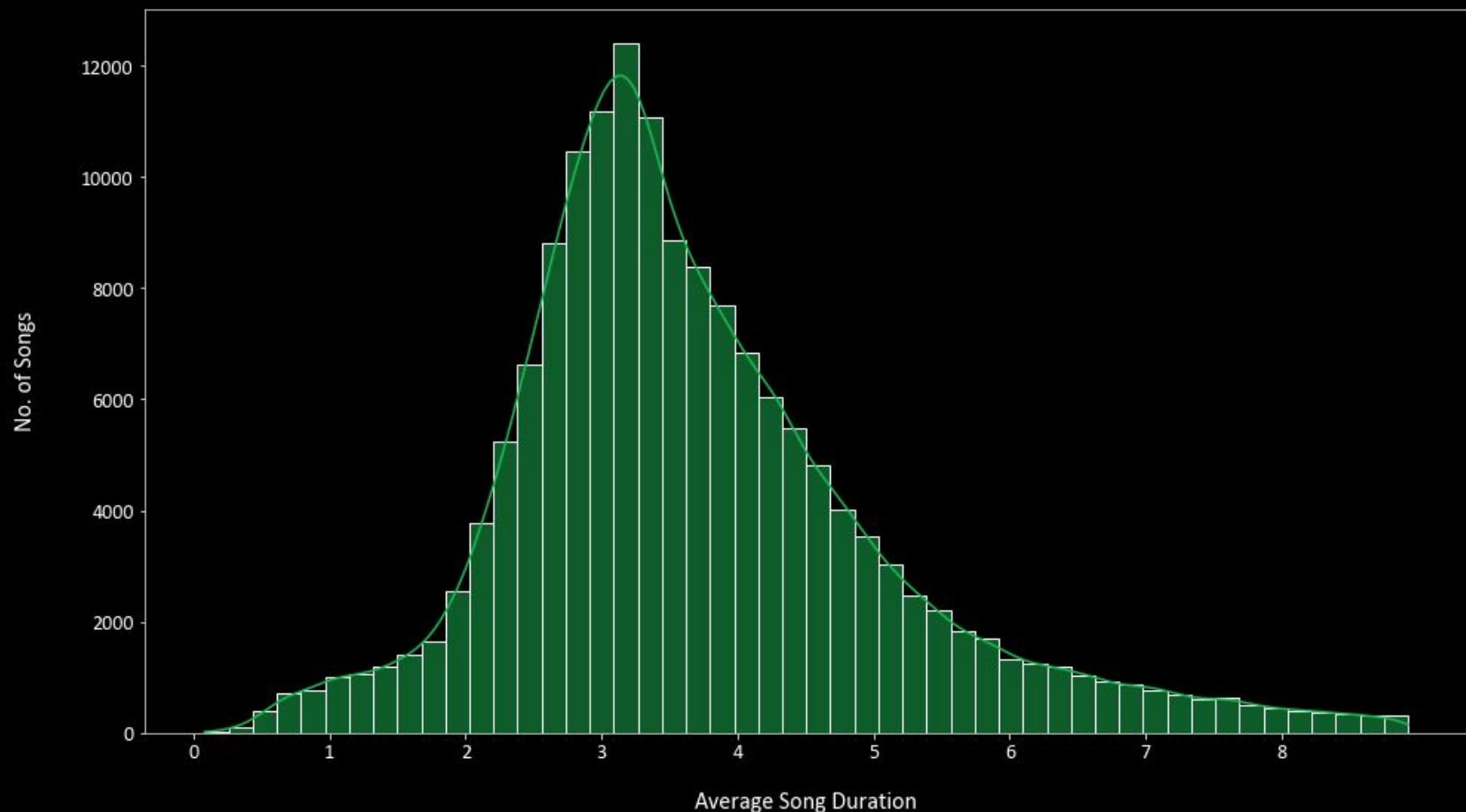
Tempo (BPM) Over Time



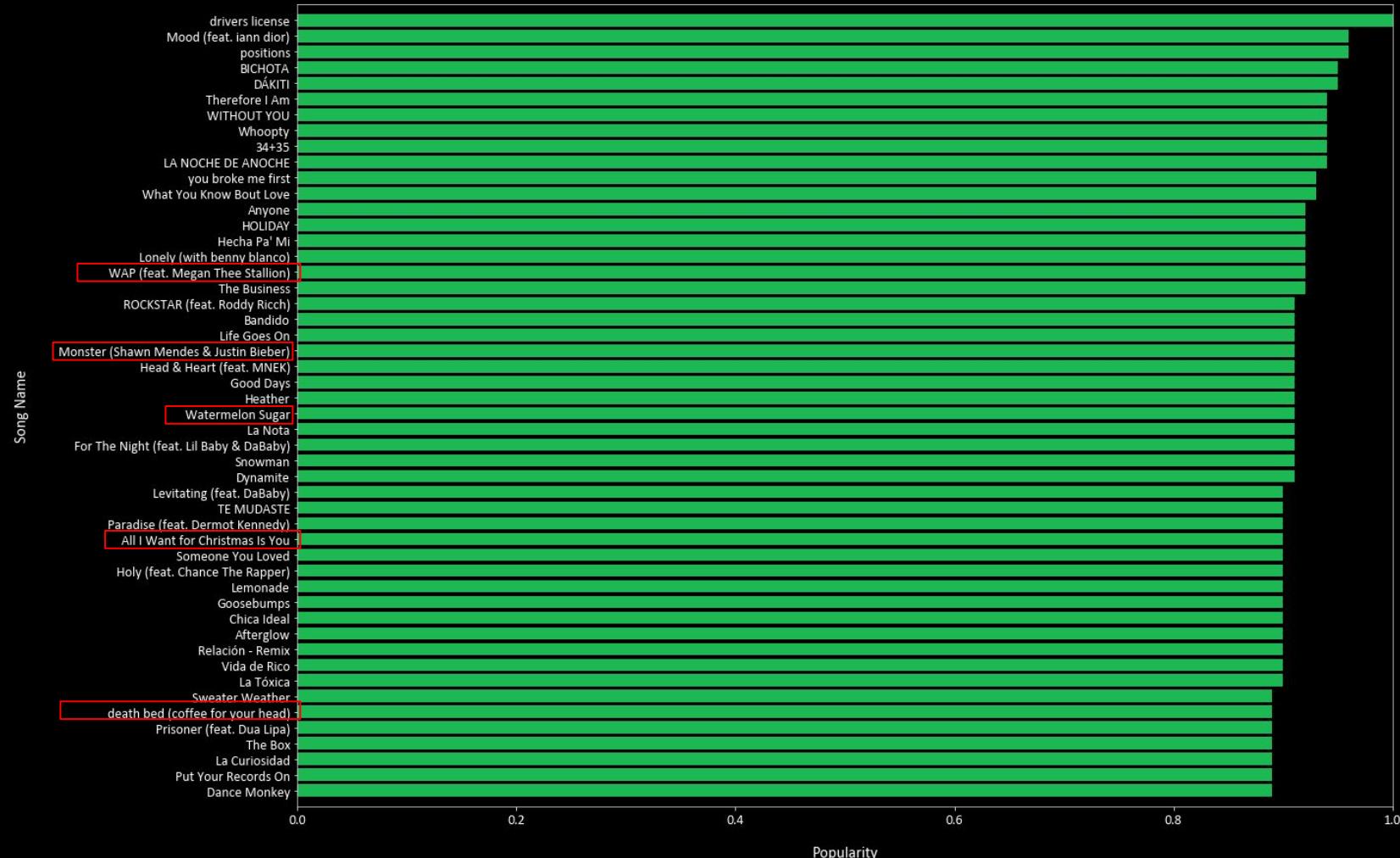
Loudness (dB) by Year



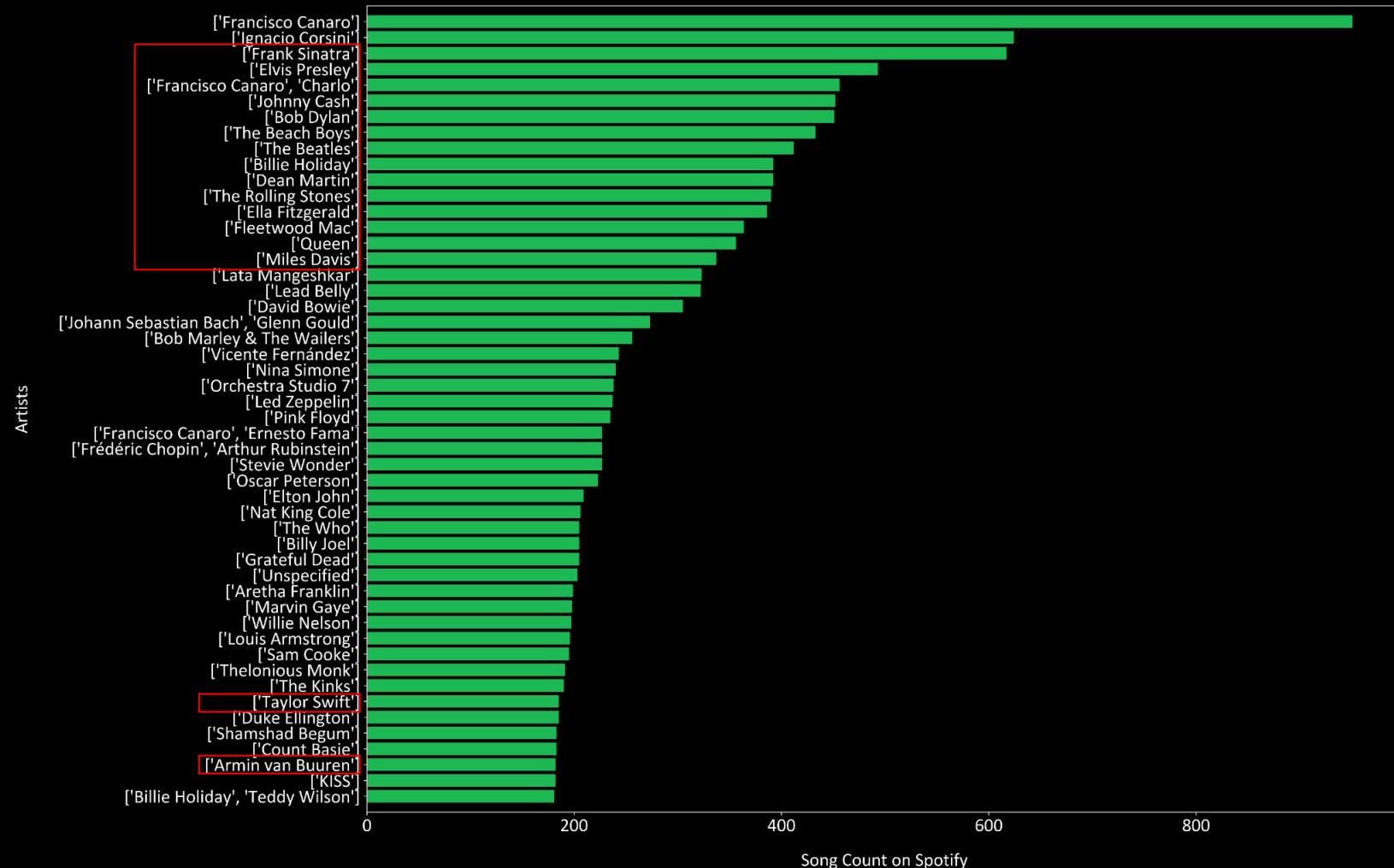
Distribution of Songs by Duration (Mins)



Top 50 Songs by Popularity



Artists with Most Songs (Top 50)



01

Introduction

02

Spotify Data

03

Data Exploration

04

Modelling Popularity

05

Song Recommender

06

Conclusion

Modelling Popularity

Classification Problem

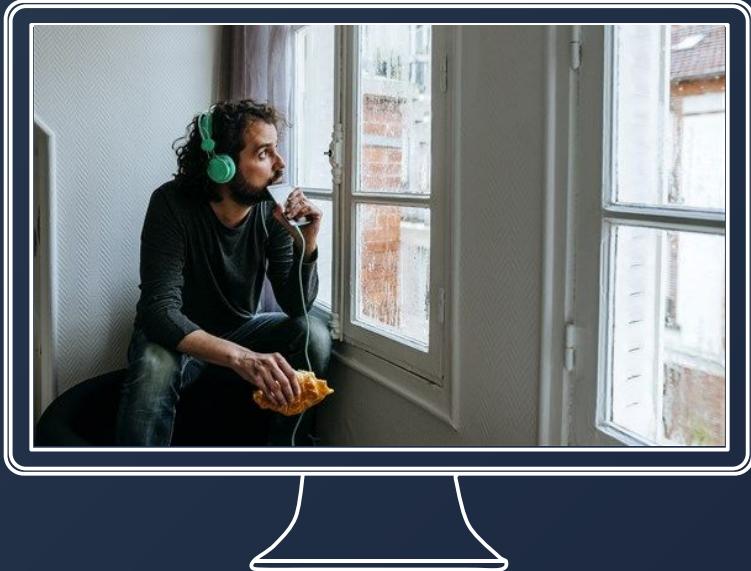
- Popularity range: 0 – 1.0
- Songs above 0.7 deemed as popular (arbitrary)

Models Tested

- Logistic Regression
- K-Nearest Neighbours
- Random Forest
- XGBoost



Challenges #2



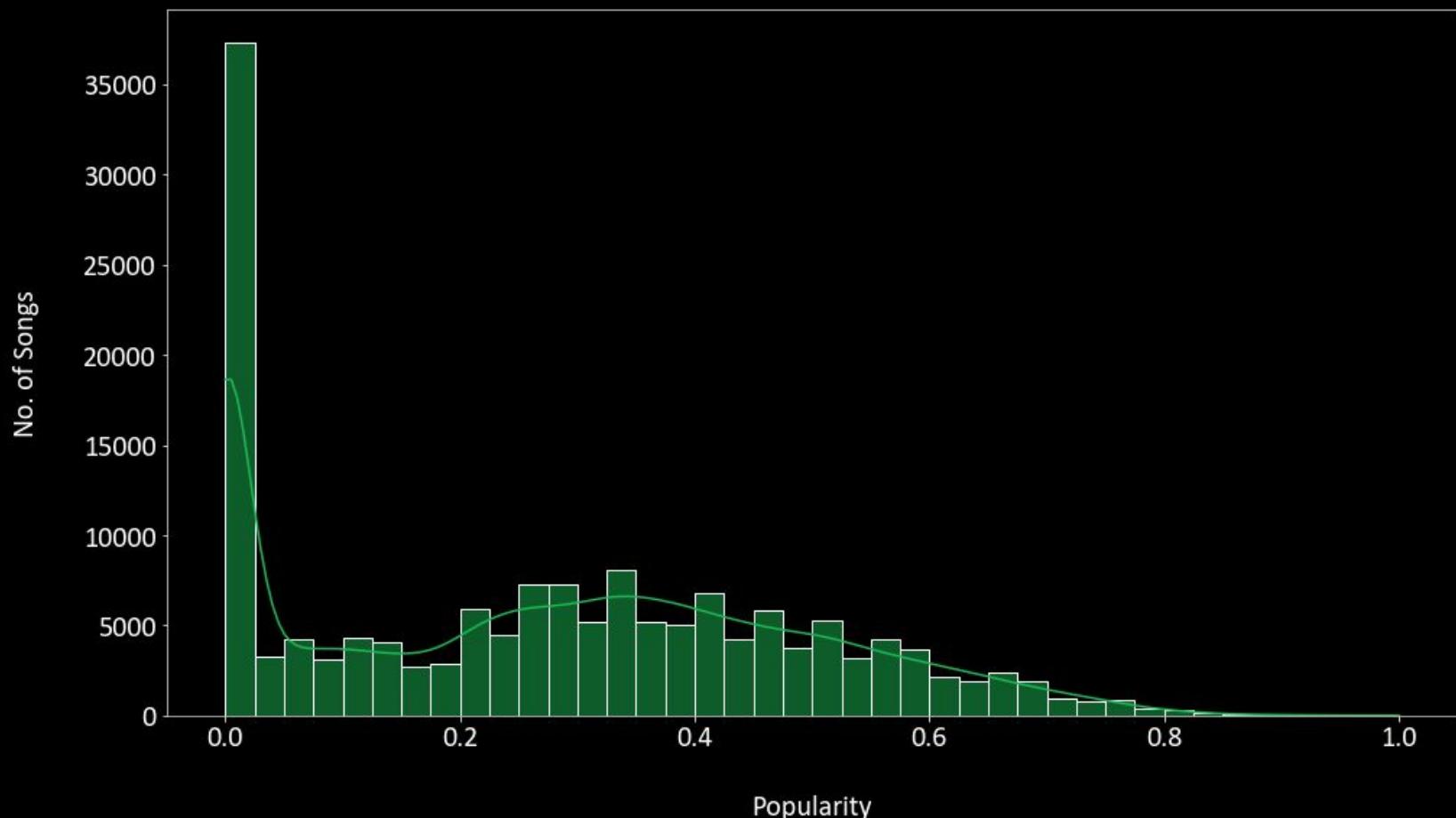
Challenges:

- Huge Dataset (160,000 songs)
- Feature Importance
- Imbalanced Classes

Solution:

- Model over subsets of data
- SMOTE

Popularity Distribution



Modelling Popularity

Models run over:

1. Full song data set (Appendix)
2. Songs from 2000 onwards
3. Songs from 1980 - 1999

Year 2000+ Results

	LR	KNN	RF	XGB
Train Acc	0.776	0.925	0.999	0.840
Cross Val	0.775	0.761	0.793	0.802
Test Acc	0.786	0.766	0.788	0.802
Train AUC	0.503	0.816	0.998	0.758
Test AUC	0.504	0.700	0.671	0.716
Train Recall	0.008	0.944	0.997	0.755
Test Recall	0.009	0.725	0.459	0.680
F1 Score	0.017	0.288	0.349	0.328

Year 2000+ Results

XGB with SMOTE best parameters:

```
'k_neighbors': 6,  
'sampling_strategy': 'minority',  
'xgb__alpha': 0.5,  
'xgb__lambda': 1,  
'xgb__learning_rate': 0.1,  
'xgb__max_depth': 4,  
'xgb__n_estimators': 100,  
'xgb__scale_pos_weight': 1,
```

2000 Onwards	Predicted Not Popular	Predicted Popular
Actually Not Popular	8671	2857
Actually Popular	370	786

1980 – 1999 Results

	LR	KNN	RF	XGB
Train Acc	0.707	1.00	0.692	0.850
Cross Val	0.696	0.620	0.682	0.669
Test Acc	0.702	0.677	0.684	0.677
Train AUC	0.656	0.899	0.645	0.767
Test AUC	0.631	0.610	0.630	0.599
Train Recall	0.717	1.00	0.779	0.764
Test Recall	0.659	0.435	0.739	0.435
F1 Score	0.034	0.040	0.032	0.037

1980 – 1999 Results

LR with SMOTE best parameters:

```
'k_neighbors': 6,  
'sampling_strategy': 'minority',  
'lr__C': 0.4,  
'lr__penalty': 'l2'
```

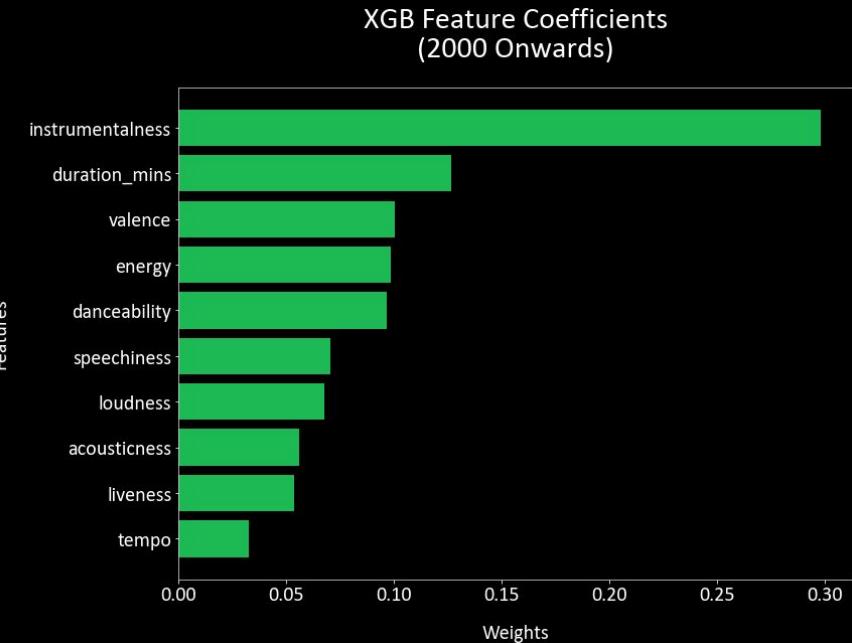
1980 – 1999	Predicted Not Popular	Predicted Popular
Actually Not Popular	7738	5123
Actually Popular	47	91

Side-by-side

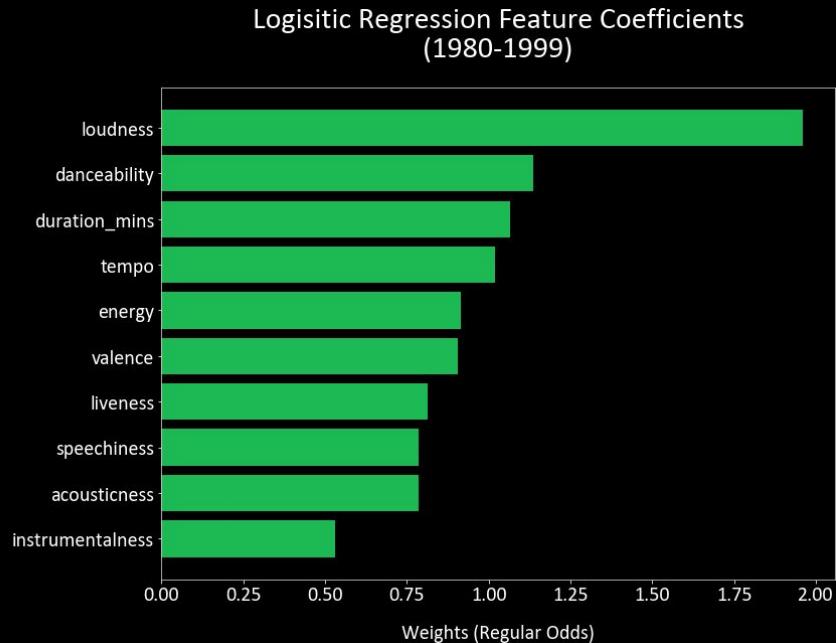
	LR (1980-1999)	XGB (2000+)
Train Acc	0.707	0.840
Cross Val	0.696	0.802
Test Acc	0.702	0.802
Train AUC	0.656	0.758
Test AUC	0.631	0.716
Train Recall	0.717	0.755
Test Recall	0.659	0.680
F1 Score	0.034	0.328

1980 - 1999	Predicted Not Popular	Predicted Popular
	Actually Not Popular	Actually Popular
2000 Onwards	Predicted Not Popular	Predicted Popular
Actually Not Popular	7738	5123
	47	91
Actually Popular	8671	2857
	370	786

Features



Features



Tips for writing a popular song now



People nowadays like songs with

1. Higher levels of instrumentalness (i.e less vocals, more instrumentals)
2. Songs with high energy and danceability
3. Songs from 2000 onwards



01

Introduction

02

Spotify Data

03

Data Exploration



04

Modelling Popularity

05

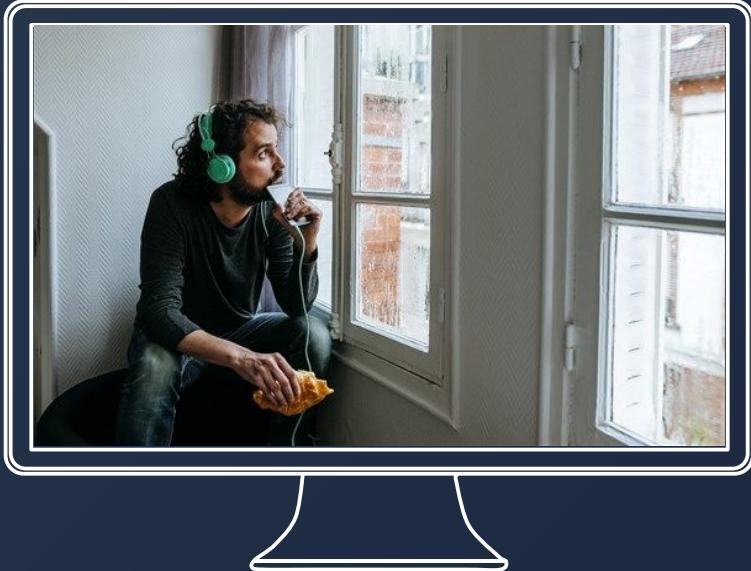
Song Recommender

06



Conclusion

Challenges #3



Challenges:

- Huge Dataset (160,000 songs)
- 213 GB RAM required

Solution:

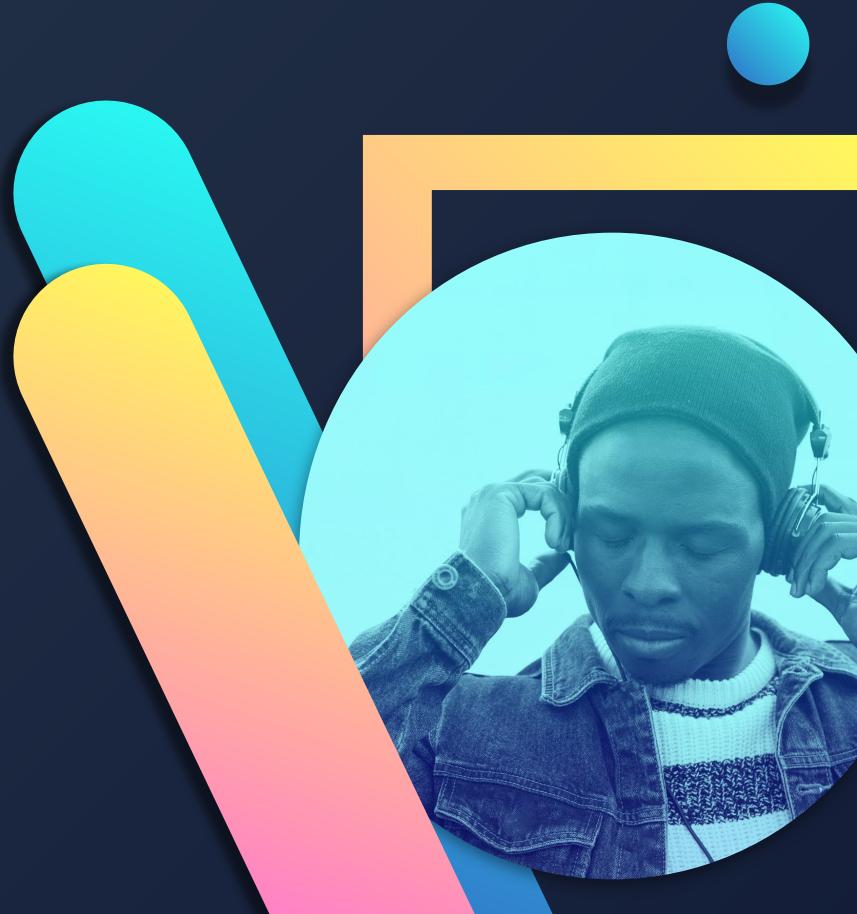
- Recommend songs from 2000

Music Funnel

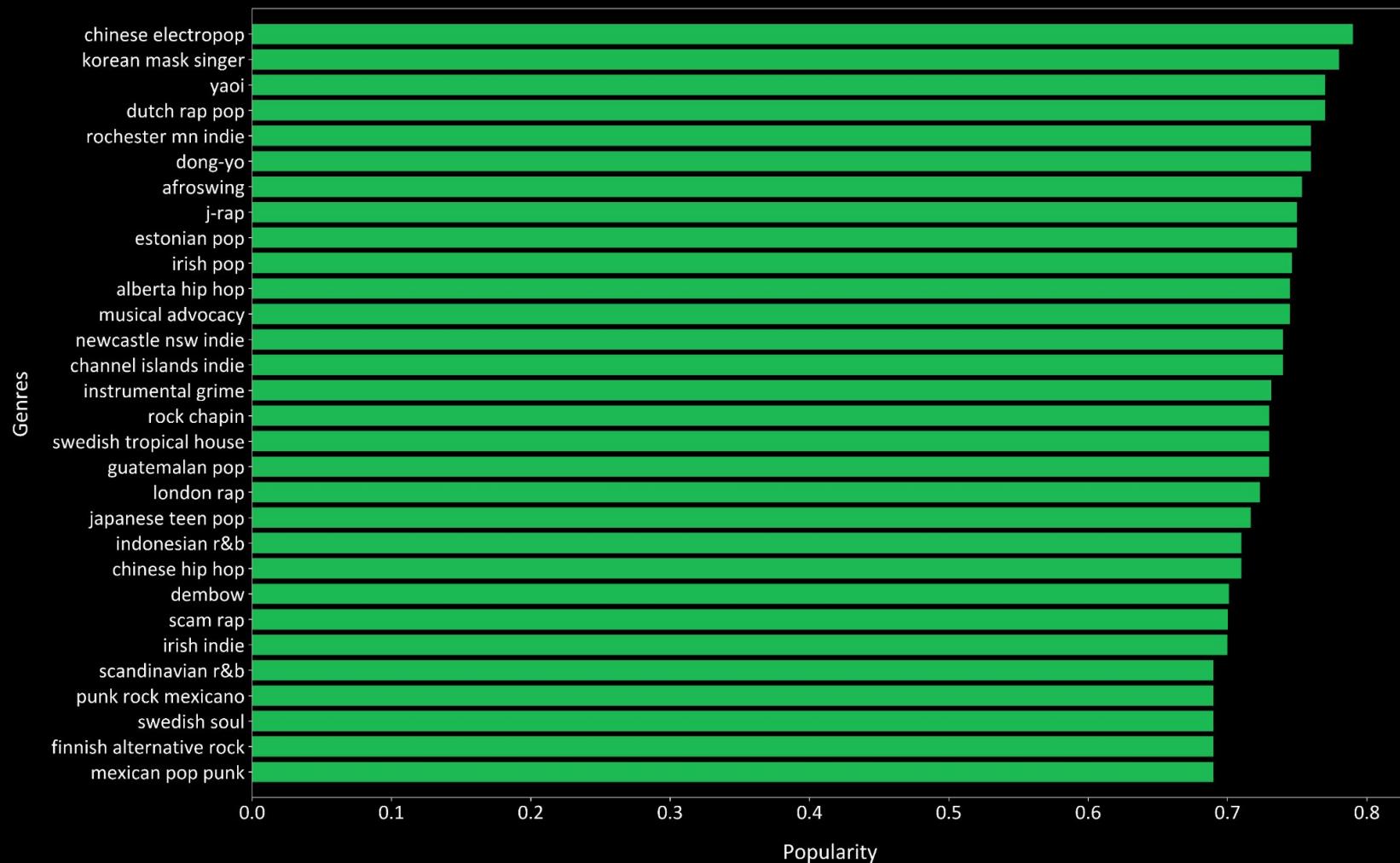


Song Recommender

- Cosine Similarity
- Build playlist of 30 songs
similar to a genre



Top 30 Genres



genres	acousticness	danceability	energy	instrumentalness	liveness	loudness	speechiness	tempo	valence	popularity	key	mode
chinese electropop	0.002570	0.660000	0.787000	0.000000	0.323000	-4.592000	0.032000	142.018000	0.199000	0.790000	1	0
korean mask singer	0.238000	0.658000	0.749000	0.000002	0.272000	-2.967000	0.063400	128.909000	0.442000	0.780000	9	0
yaoi	0.006550	0.603000	0.964000	0.000003	0.143000	-2.886000	0.048700	135.028000	0.796000	0.770000	8	0
dutch rap pop	0.216000	0.767000	0.643000	0.000104	0.080600	-5.225000	0.101000	169.990000	0.536000	0.770000	5	1
rochester mn indie	0.034700	0.655000	0.487000	0.897000	0.271000	-7.988000	0.033000	139.914000	0.045400	0.760000	6	1

...



Home



Browse



Radio

Algo-Rhythms



Search



Si Rong Geh

PLAYLIST FOLDER

Algo-Rhythms

Created by: Si Rong Geh • 30 playlists

PLAY



Sounds like Scandinavian R&B Playlist

Sounds like Mexican Pop Punk Playlist

Sounds like German Pop Rock Playlist

Sounds like Finnish Alternative Rock ...

Sounds like Future House Playlist

Sounds like Irish Indie Playlist

Sounds like Scam Rap Playlist

Sounds like Dembow Playlist

Sounds like Chinese Hip Hop Playlist

Sounds like Indonesian R&B Playlist

Sounds like Japanese Teen Pop Playlist

Sounds like London Rap Playlist

Sounds like Swedish Tropical House ...

Sounds like Rock Chapin Playlist

Sounds like Guatemalan Pop Playlist

Sounds like Instrumental Grime Playlist

Sounds like Newcastle Nsw Indie Pla...

Sounds like Channel Islands Indie Pla...

Sounds like Alberta Hip Hop Playlist

Sounds like Musical Advocacy Playlist

Sounds like Irish Pop Playlist

Sounds like Scandinavian
R&B Playlist
Si Rong GehSounds like Mexican Pop
Punk Playlist
Si Rong GehSounds like German Pop
Rock Playlist
Si Rong GehSounds like Finnish
Alternative Rock Playlist
Si Rong GehSounds like Future House
Playlist
Si Rong GehSounds like Irish Indie
Playlist
Si Rong GehSounds like Scam Rap
Playlist
Si Rong GehSounds like Dembow
Playlist
Si Rong GehSounds like Chinese Hip
Hop Playlist
Si Rong GehSounds like Indonesian R&B
Playlist
Si Rong GehSounds like Japanese Teen
Pop Playlist
Si Rong GehSounds like London Rap
Playlist
Si Rong Geh

New Playlist



Ted Recommends

April 28

<https://open.spotify.com/playlist/1DafxkKqRmBY91Hcisi7xdu3mb0R06np0206FE4zQ>

eh i updated it - try dis instead

Spotify
Sounds like Scandinavian R&B Playlist
Si Rong Geh · Playlist · 30 songs · 0 likes



Not bad leh 8:25 AM

Though idk what they singing but feels good to my ears 8:26 AM

Sounded like malay mix spanish mix thai at times lolol 8:26 AM

hmm just breezing through the finnish alternative rock playlist

2:54 PM

Lucas Lee
idk bout this algo man 2:55 PM

they're all quite differnt songs 2:54 PM
but actually 2:54 PM
have similar surface elements 2:55 PM

so interesting 2:55 PM
u used algorithms to put these playlists together eh 2:56 PM

Si Rong
<https://open.spotify.com/playlist/1DafxkKqRmBY91Hcisi7xdu3mb0R06np0206FE4zQ>
wa this playlist actually looks fire 2:57 PM

Gordon
But I randomly picked a song 6:25 PM

From indo rnb 6:25 PM

It's clearly rock lolol 6:25 PM

Search...
PLAYLIST
Sounds like Afroswing Playlist
Created by Si Rong Geh · 29 songs, 1 hr 39 mins

TITLE	ARTIST	ALBUM	DATE	TIME
Sign Extravagant	J Balvin	Ruta Resaca: Latino Mix	10 hours ago	3:23
Rated	James Arthur	Rated	10 hours ago	3:24
The Look Of Love - Pt. 1	ABC	The Lexicon Of Love	10 hours ago	3:29
Cravin (feat. G-Eazy)	Dave Luegh, G-Eazy	Cravin	10 hours ago	3:39
Tina Fliss	Druks	Dark Lane Demo Tapes	10 hours ago	3:43
Drop It Like It's Hot	Snoop Dogg, Pharrell Williams	Saturday Night	10 hours ago	3:47
SADI	XIXTENTACION	9	10 hours ago	3:47

got j balvin, which is latin pop 2:55 PM

james arthur which is vanilla bread acoustic 2:55 PM

then hiphop snoop dogg, drop it like its hot 2:56 PM

like the sounds n bpm not the same LOL 2:56 PM

Teach Me How to Dougie	Gill Swig District	The Kickback	10 hours ago	3:07
Year Body Is a Wonderland	John Mayer	Roon For Squares	10 hours ago	4:10

teach me how to dougie is followed by john mayer 2:56 PM



01

Introduction

02

Spotify Data

03

Data Exploration



04

Modelling Popularity

05

Song Recommender

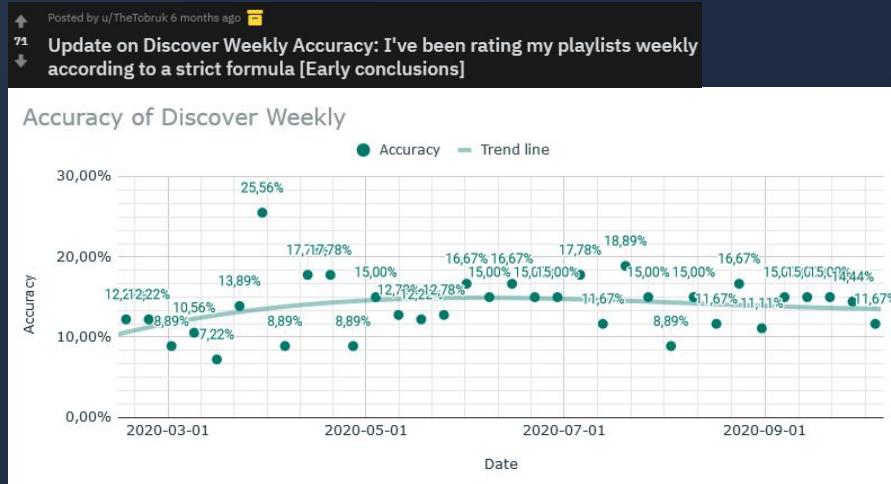
06

Conclusion



Recommender Conclusions

1. Recommendation systems are NOT perfect

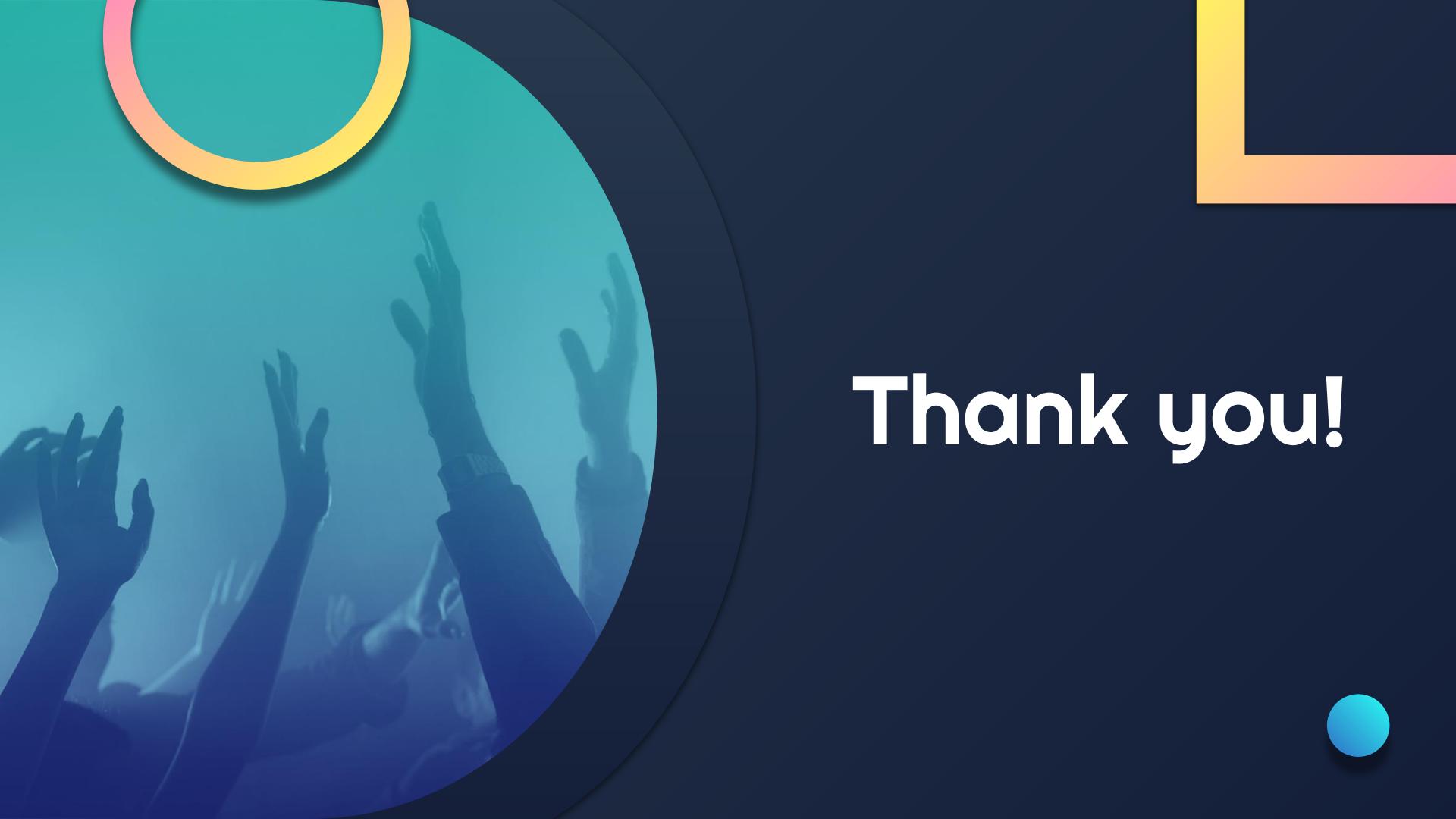


2. Songs with similar features may not be from the same genres
3. Music genres of our time overlap significantly (eg. Indie pop, country rock)



"Music transcends genres,
boundaries & languages. Music
is a feeling, an emotion, a
mood."

– Geh, 2021



Thank you!



Appendix

- Full Dataset
- EDA (Mode, Key)
- Modelling (Full Hyperparameters)
- Recommender (All playlists)

Spotify Dataset



19 features

Musical Features (Numerical)

- Acousticness
- Danceability
- Energy
- Instrumentalness
- Liveness
- Loudness
- Speechiness
- Tempo (BPM)
- Valence (How happy a song sounds)

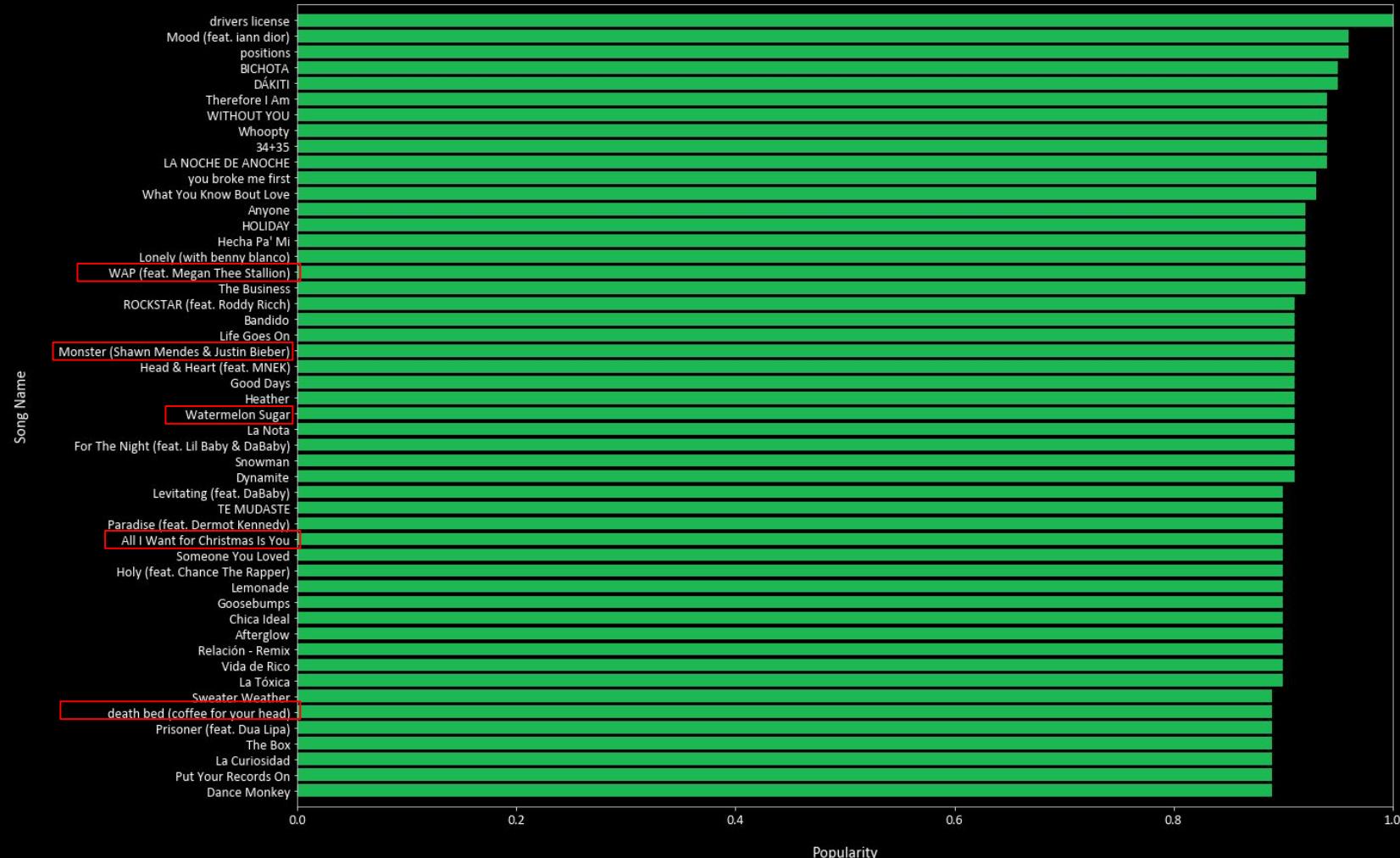
Categorical Features

- Explicit
- Key (Main chord of song)
- Mode (Minor or major)

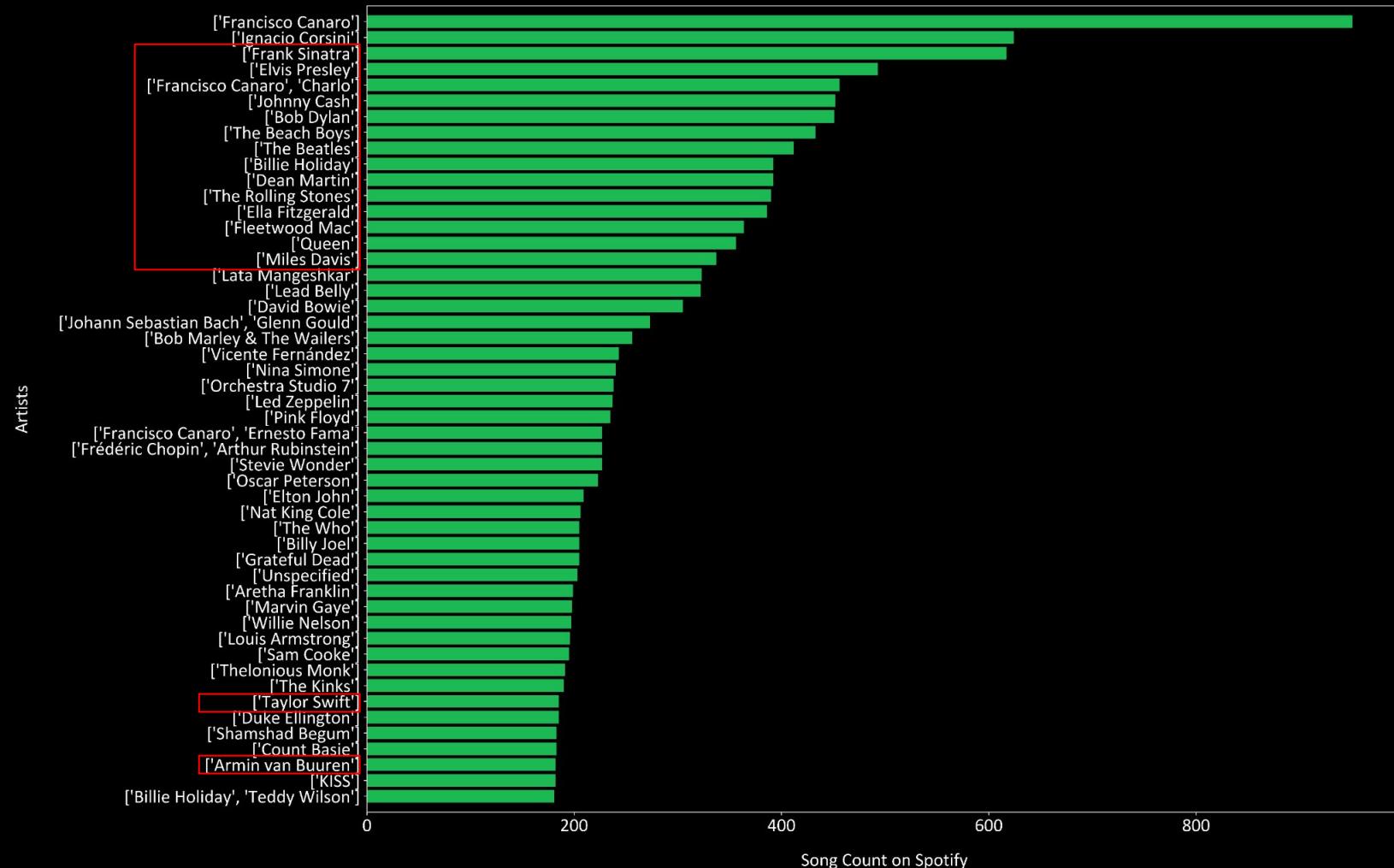
Song Features

- Popularity (Target Variable)
- Duration mins
- Song ID / Name / Artist
- Year / Year Released

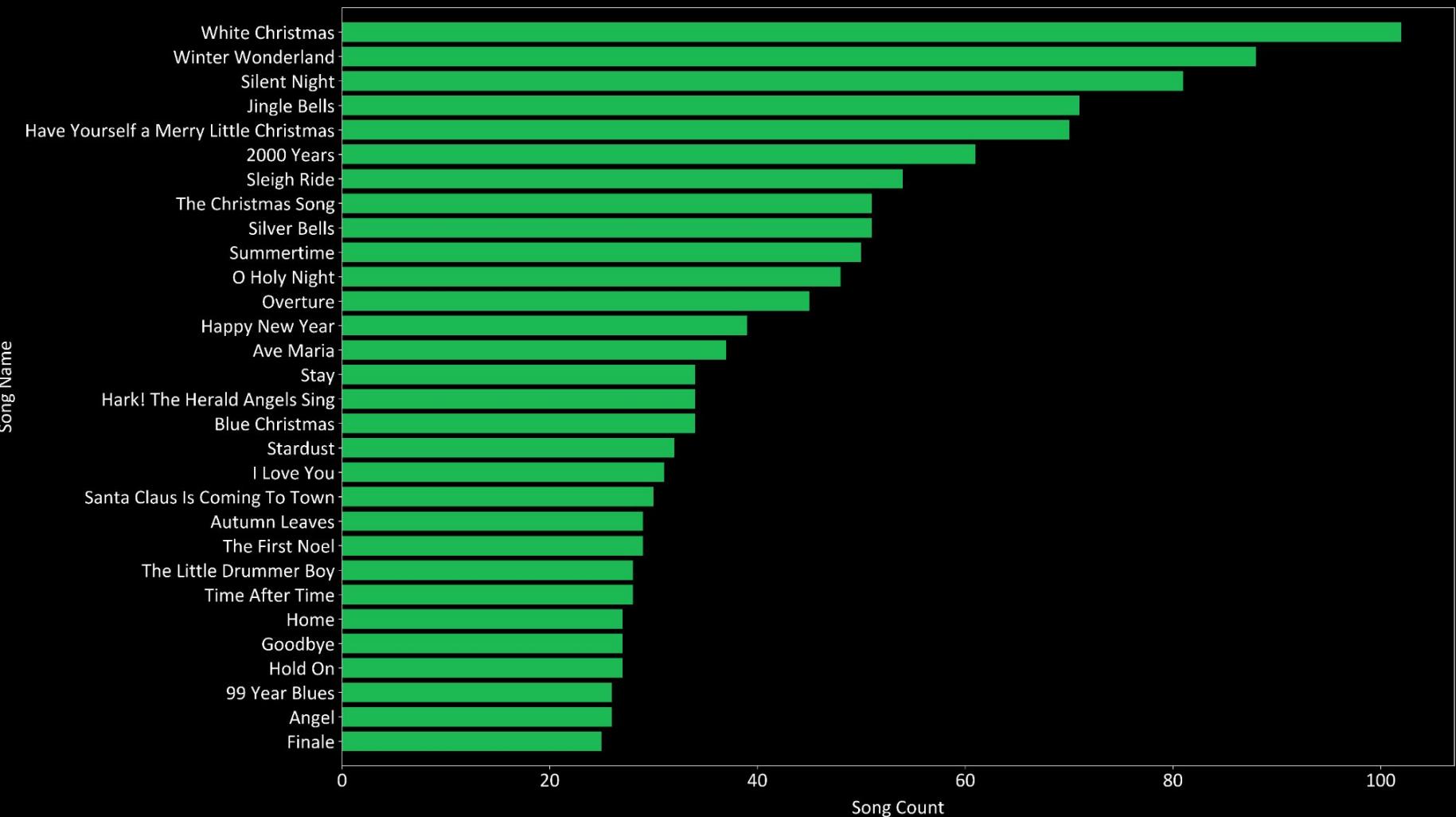
Top 50 Songs by Popularity



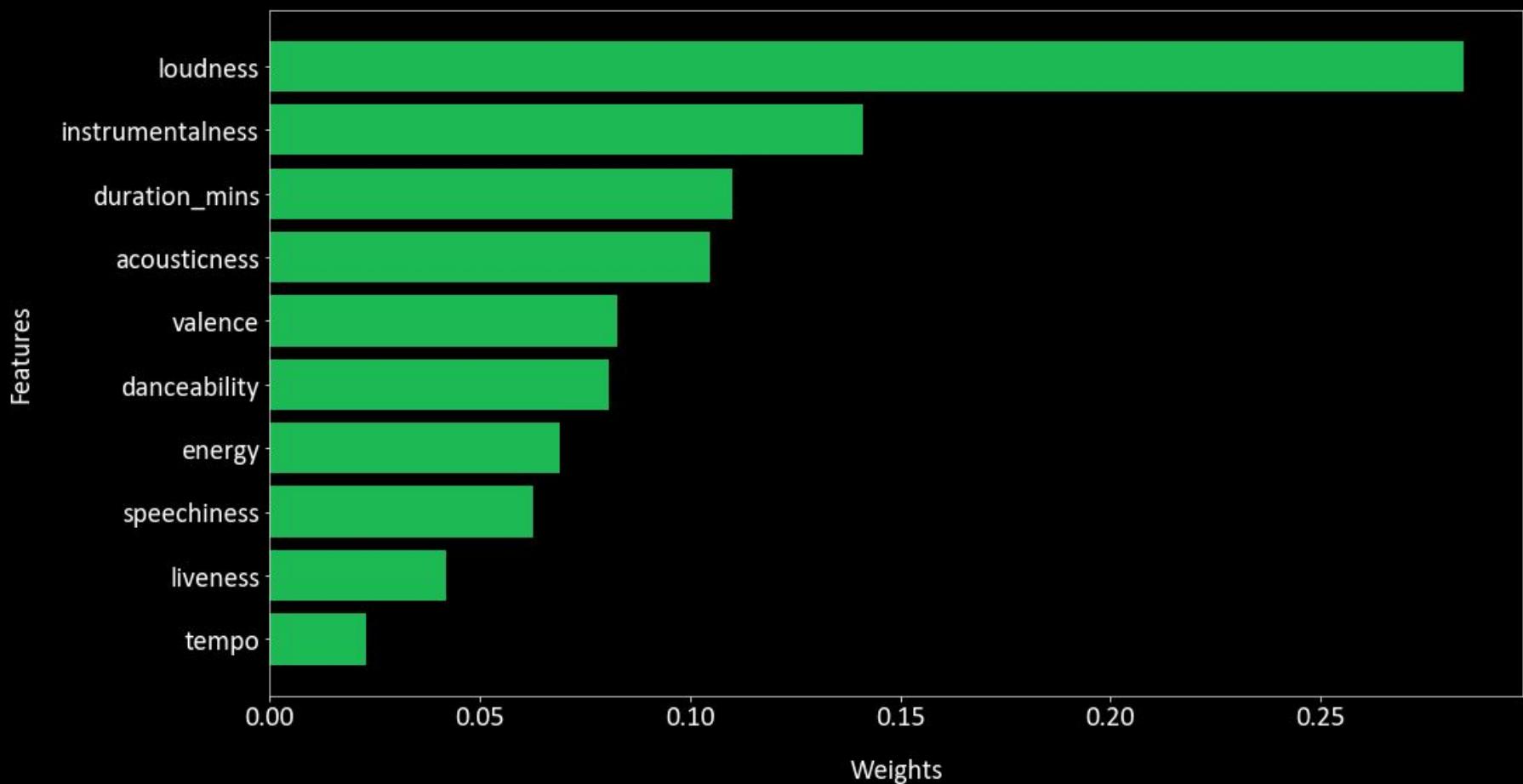
Artists with Most Songs (Top 50)



Top 30 Most Common Song Names



XGB Feature Coefficients (Overall Data)



Credits

CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon, and infographics & images by Freepik.

This slide was kept for attribution.

