

Disrupting the Music Industry through Machine Learning

Brian Cho

AGENDA

1. Project Overview
2. Process
 - a. Data Collection
 - b. Data Cleaning and EDA
 - c. Modelling
 - d. Evaluation
3. Conclusions and Recommendations

PROJECT OVERVIEW: Background

Who owns the licensing to music content?

- Record Labels
- Independent Artists

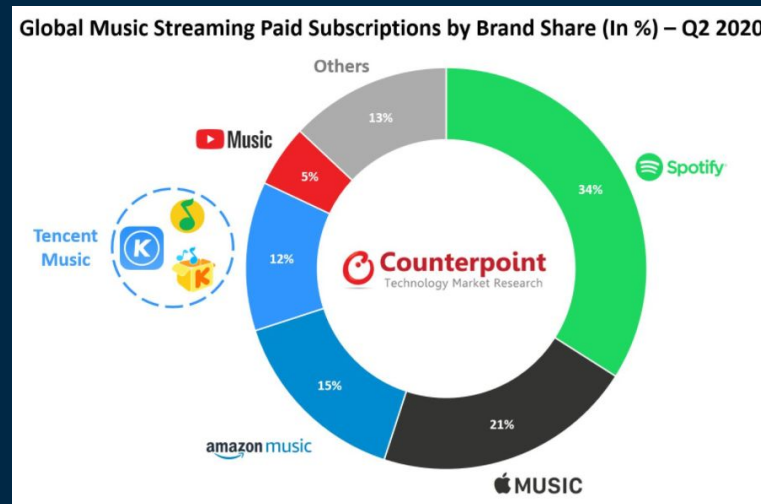
What makes a hit song?

- Artist paired with the right song
- Recording Quality
- Marketing/Promoting
- Timing
- Luck



SPOTIFY MISSION STATEMENT

“Our mission is to unlock the potential of human creativity – by giving a million creative artists the opportunity to live off their art and billions of fans the opportunity to enjoy and be inspired by it.”



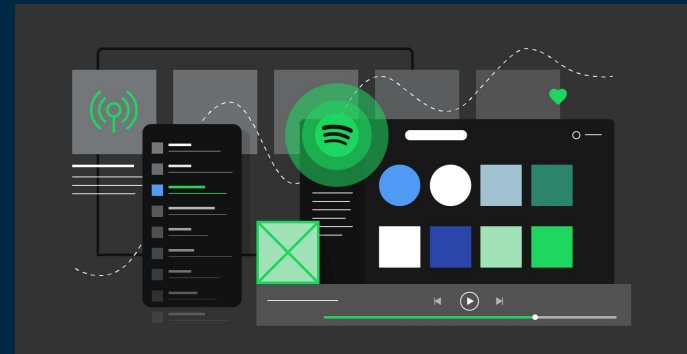
“UNLOCKING THE POTENTIAL OF HUMAN CREATIVITY”

Problem Statement: Create a machine learning model that would be capable of classifying whether a song will be a hit.

- Model will not factor in the artist
- Model will be built with only audio features/metrics
- Hit will be classified if a track made the Billboard Top 100

Success Metrics:

- Accuracy
- F1 Score



PROCESS

1

Collecting the Data

2

Data Cleaning & EDA

3

Modelling

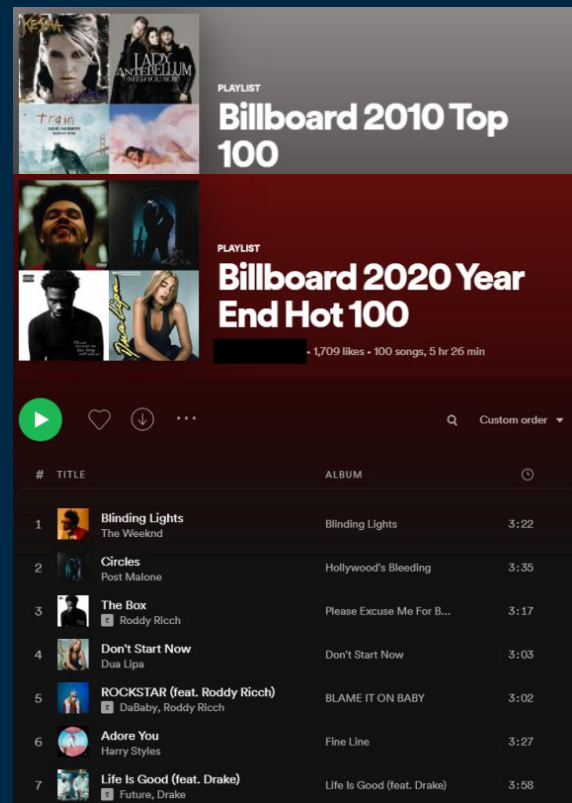
4

Evaluations

PROCESS: Data Collection

- Spotify API → Audio Features
- 14,000+ songs
- Hits from 2010–2020
- >9% of the songs were hits

Assumption:
Music trends have changed over time: Utilize
relevant data



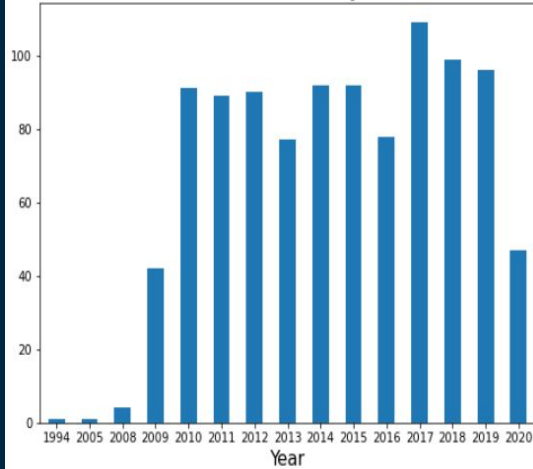
PROCESS: Data Cleaning and EDA

Cleaned Dataset:

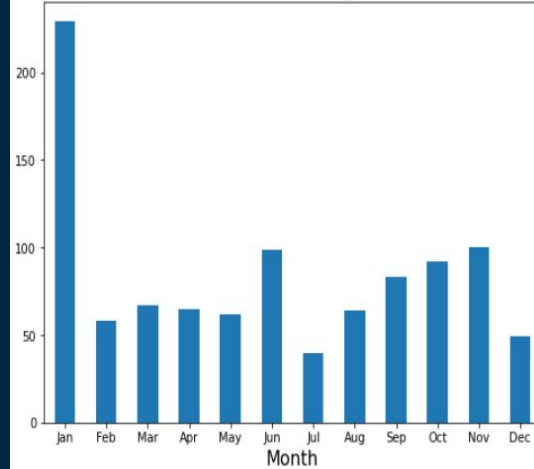
- Removed duplicate songs
- Removed hit songs that were not in the designated time frame (2010-2020)
- 10800 non-hits
- 975 hits

EDA: Release Date Analysis

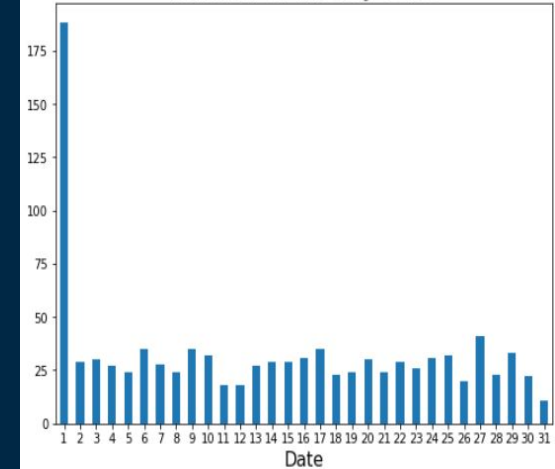
Number of Hits By Year



Number of Hits By Month



Number of Hits By Date

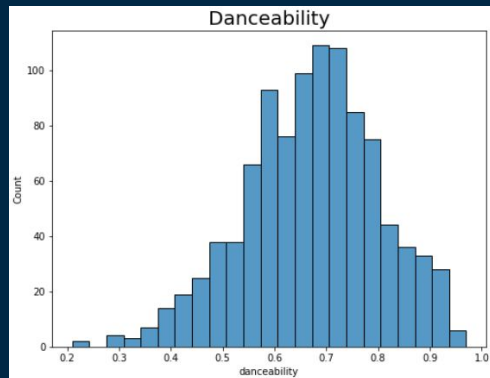


Audio Feature Analysis

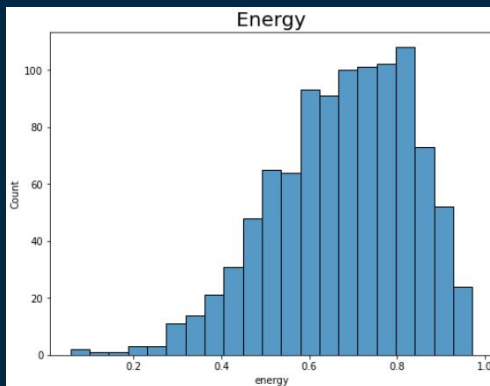
Features:

danceability, energy, key, loudness, mode,
speechiness, instrumentalness, liveness, valence,
tempo, duration, time signature, release date

Danceability/Energy

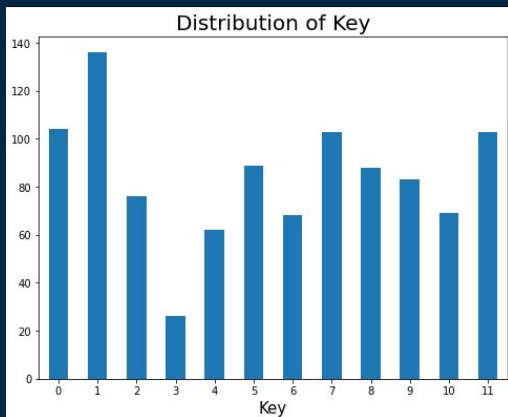


Danceability portrays how fit a track is for dancing.
0.0 being the least danceable
1.0 being the most danceable

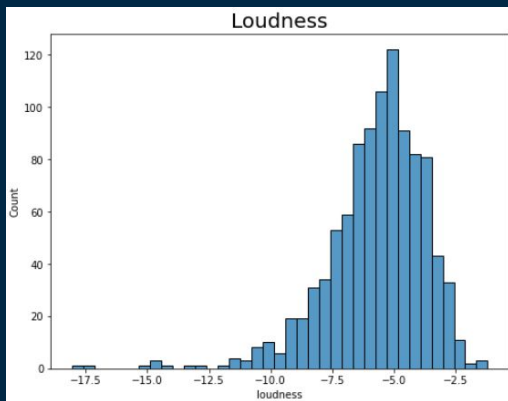


Energy measures how fast, loud, and noisy a track is.

Key/Loudness

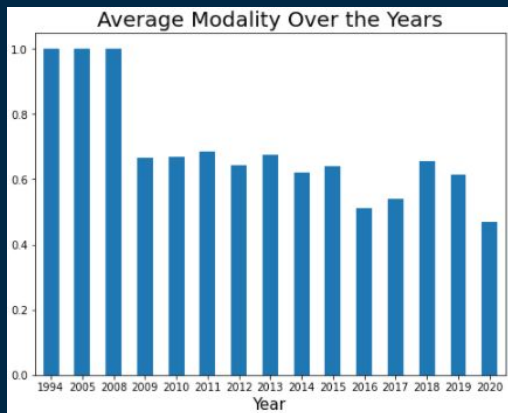


Key is the state measure of pitch in a song

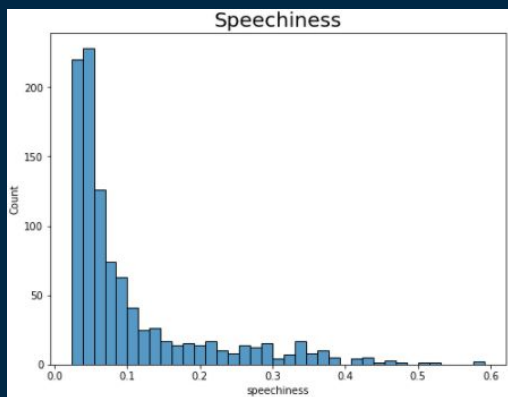


Loudness is measured in decibels (dB).

Mode/Speechiness

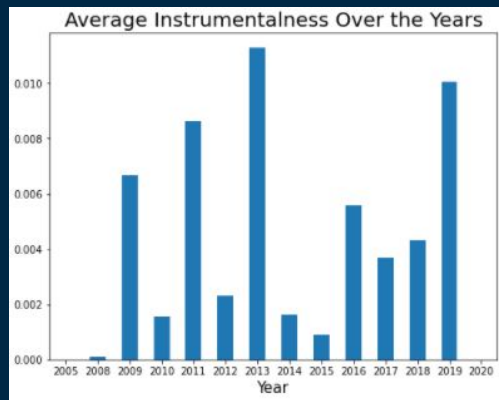


Mode is the measure of modality (major or minor) of a track. Major is represented by a value of 1 and minor is represented by a value of 0.

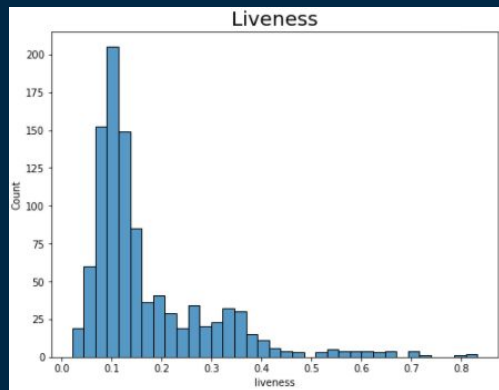


Speechiness portrays the measure of spoken words

Instrumentalness/Liveness

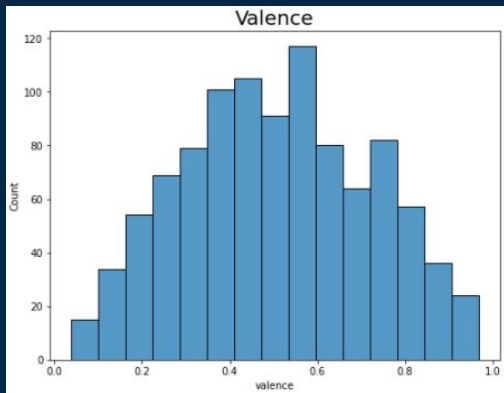


Instrumentalness measures whether a track contains no vocals. Attribute ranges between 0.0 and 1.0.
1 - no vocal content
0 - instrumental track

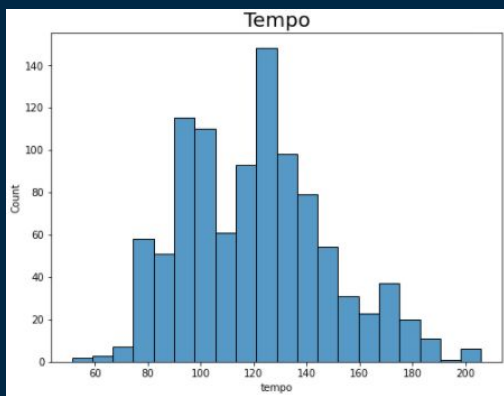


Liveness detects the presence of an audience in the recording. Higher values represent an increased probability the track was performed live.

Valence/Tempo

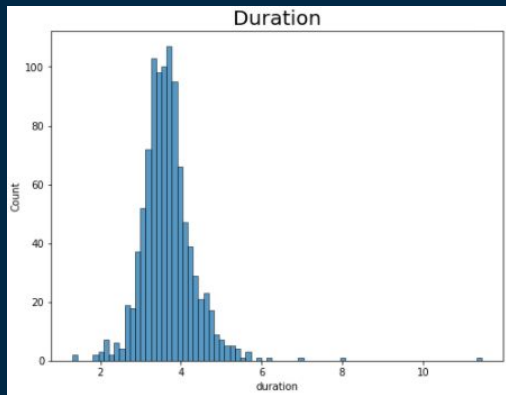


Valence portrays the musical positiveness conveyed by a track. Tracks with high valence values sound more positive.

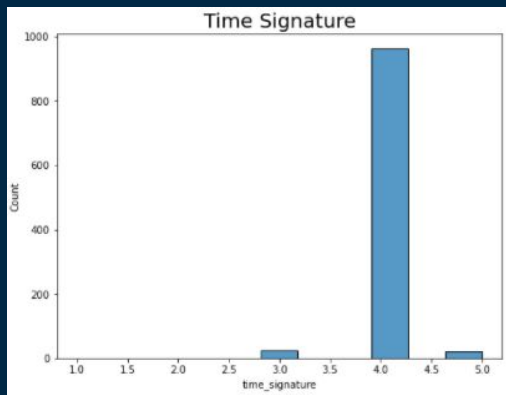


Tempo is the measure of beats per minute (BPM) of a track. It is the speed or pace of a track.

Duration/Time Signature

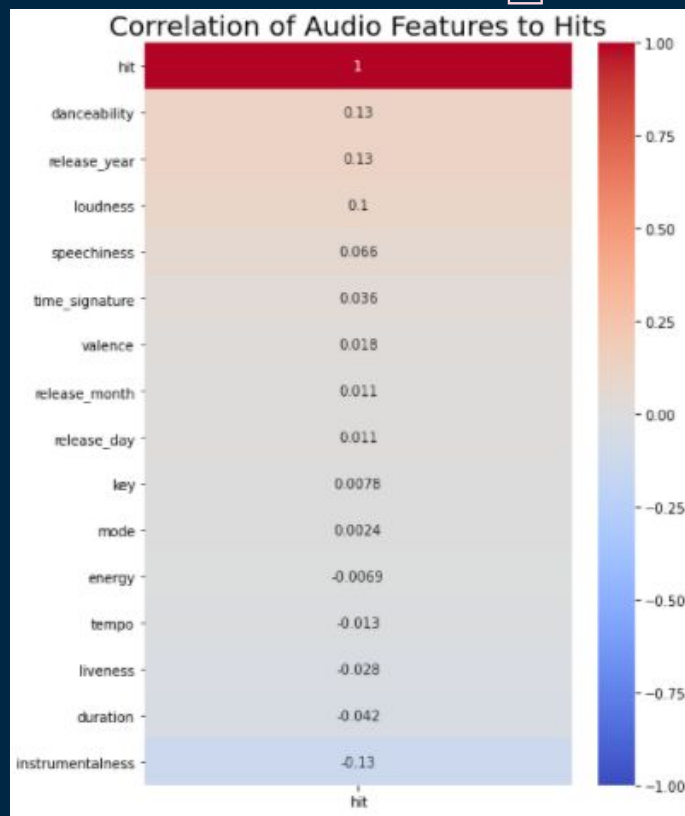
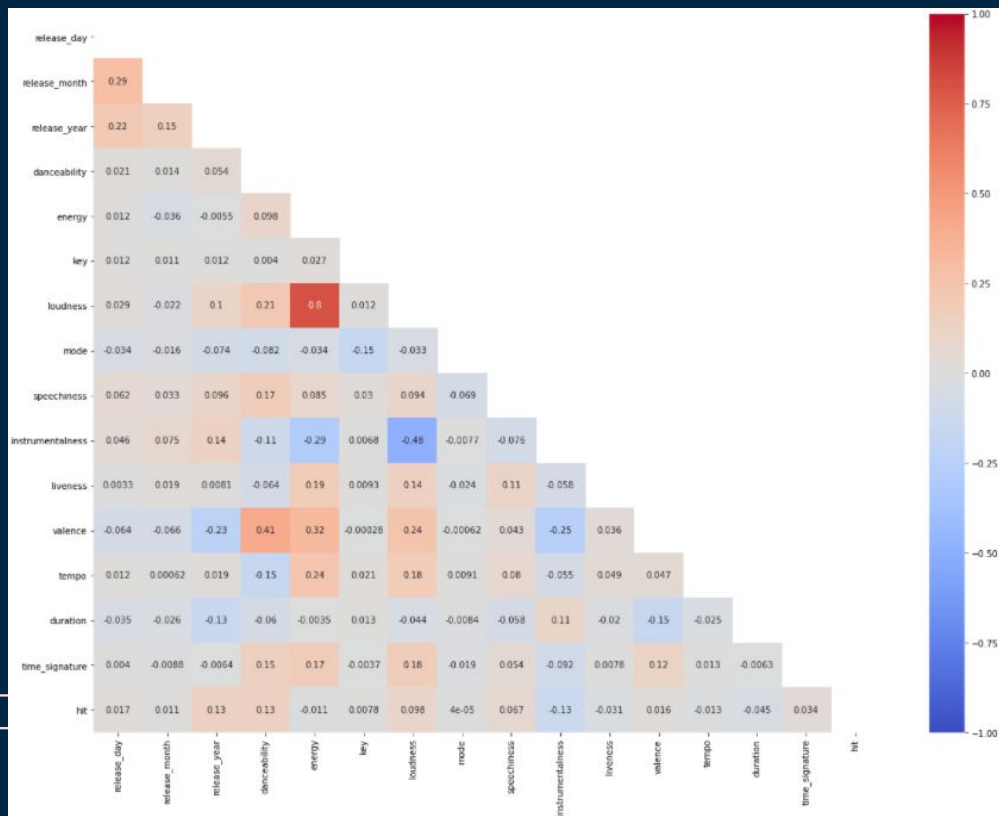


Duration of the track converted into minutes



Time Signature is the estimated overall time signature of a track. Known as a notational convention to specify how many beats are in each bar/measure.

PROCESS: Data Cleaning and EDA



PROCESS: Modelling

1. Logistic Regression
2. KNN
3. Random Forest
4. Voting Tree Classifier
5. SVM
6. Neural Network

PROCESS: Evaluation

MODEL	ACCURACY	PRECISION	RECALL	F1 SCORE
LOG REGRESSION	.76	.728	.83	0.78
KNN	.831	.753	.986	.854
RANDOM FOREST	.837	.775	.95	.853
VOTING TREE	.894	.864	.935	.90
SVM	.853	.80	.948	.866
NEURAL NETWORK	.856	.822	.916	.867

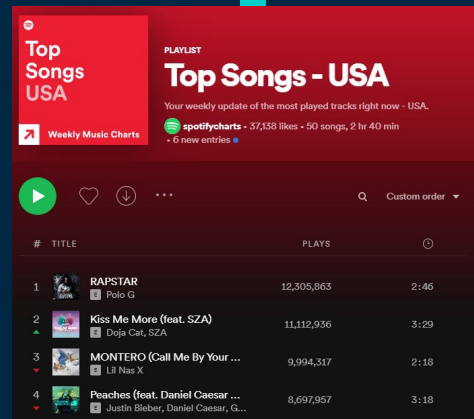
CONCLUSION

Dataset was constructed off of the Spotify API

- Improve the speed in filtering the search for the next hit song
- Help edit/produce music utilizing the trends in music
- Spotify is quietly disrupting the music industry
 - Spotify the next big Music Record Label?
 - Spotify allowing indie artists to bypass traditional record labels

RECOMMENDATIONS

- Data gap: Play Count
- Train a model for each genre
- Traditional Record Labels must be aware of the strides Spotify are making.
- New Spotify Initiative
 - Giving artists more freedom
 - Continue housing more data



“Licensing content does not make us a label, nor do we have any interest in becoming a label,” Daniel Ek, Spotify’s chief executive

Do you have any questions?

bcho91@gmail.com

THANKS



CREDITS: This presentation template was created by [Slidesgo](#),
including icons by [Flaticon](#), and infographics & images by [Freepik](#)
Please keep this slide for attribution