Finding the Missing Link

Simplifying the Search for Music Rights

Stephanie Caress

June 2020

Quick Overview of Sync Licensing

- When a song is used in Film, TV, Ads, Trailers, etc., it is synchronized with the picture and requires a **Sync License**
- Rights need to be cleared/secured with the MASTER and PUBLISHING
  - MASTER refers to the actual recording, usually owned by the **Record Label**
  - PUBLISHING refers to the underlying composition, often represented by **Publishers**
    - i.e. Covers are different masters but the same composition
- Sometimes, a dozen or more parties are involved with the use of a single song
  - **<u>If there is an issue with any stakeholder, the entire deal could fall through</u>**

Research is Key!

- Doing extensive research online can help avoid problematic songs
- There are dozens of sites to check and cross reference to build the complete picture of the song
- But thoroughly checking each song is time consuming

Example

You just received word an editor wants to use **“Milkshake - Kelis”** in a cut of the episode to be released tomorrow.

TL;DR - less than 24 hours to find out if this song is ok

Check Spotify - Copyrights from Elektra (WMG), Virgin (UMG), and Arista (Sony)

- Hmmmm all three labels? Suspicion level: 50%

Check publishing: One source says 4 writers (and 4 publishers), others say only 2

- Another discrepancy, suspicion raised to 70%

BUT according to Tunefind, it was used in Looking for Alaska (2019)

- That helps! We might hit road bumps but this seems ok

You tell the editor, all good! But he says now they don’t like it and want similar options. So the cycle continues.....

How can we speed up the process of finding and analyzing song ownership to ultimately decide if a song is clearable?

Data Gathering

- Gathering the data was the most difficult part.
- Each new source required looping through the dataframe to match observation
    1. Random song list to start
    2. Scraped **What-Song** to have synced (in dataset) vs not synced (not in dataset)
    3. **Deezer** API for master/recording info
    4. Scraped **LyricsFreak** for publishing/composition info
    5. **Spotify** for audio features (i.e. Danceability & Liveness scores)

Cleaning and Calculating Accuracy

- The scraped sources required lot of text cleaning
- Since each row was from multiple sources, I created a “fuzzy” match score to remove songs that were matched incorrectly

“Titanium” - David Guetta

“Titanium (feat. Sia)” - David Guetta

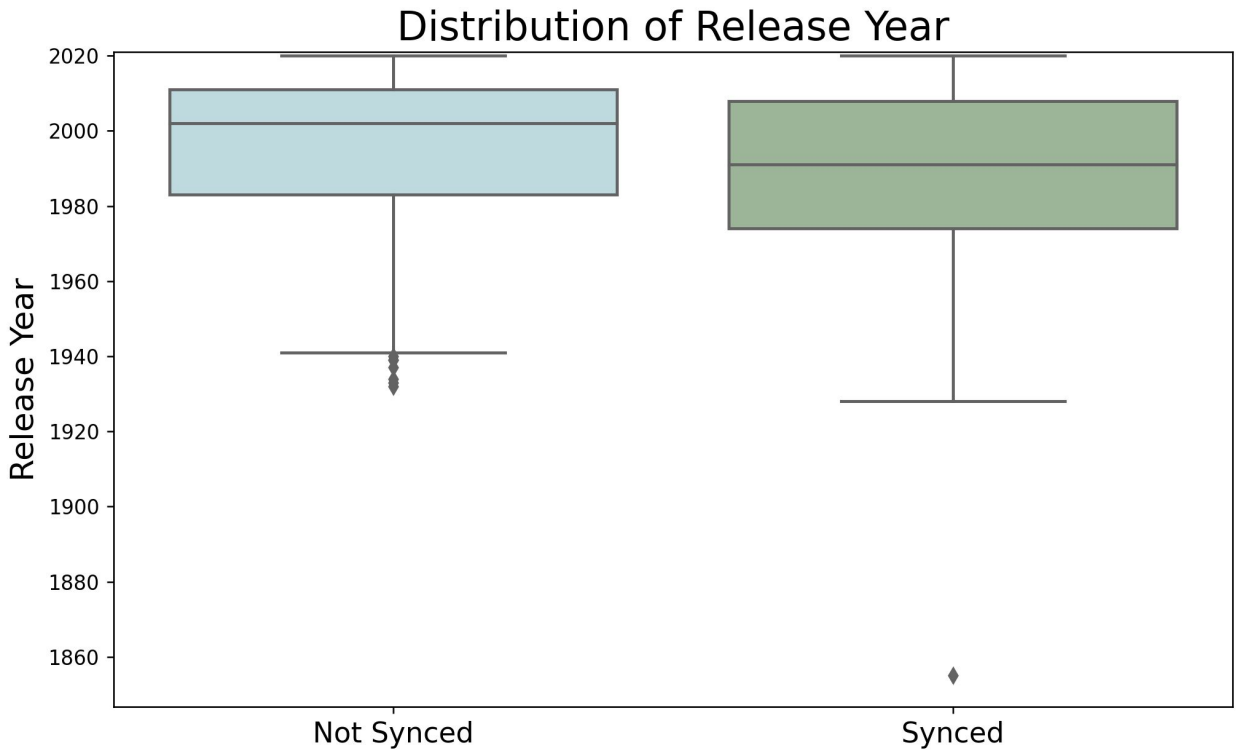
43%

“The Full Monty Medley” - Robbie Williams

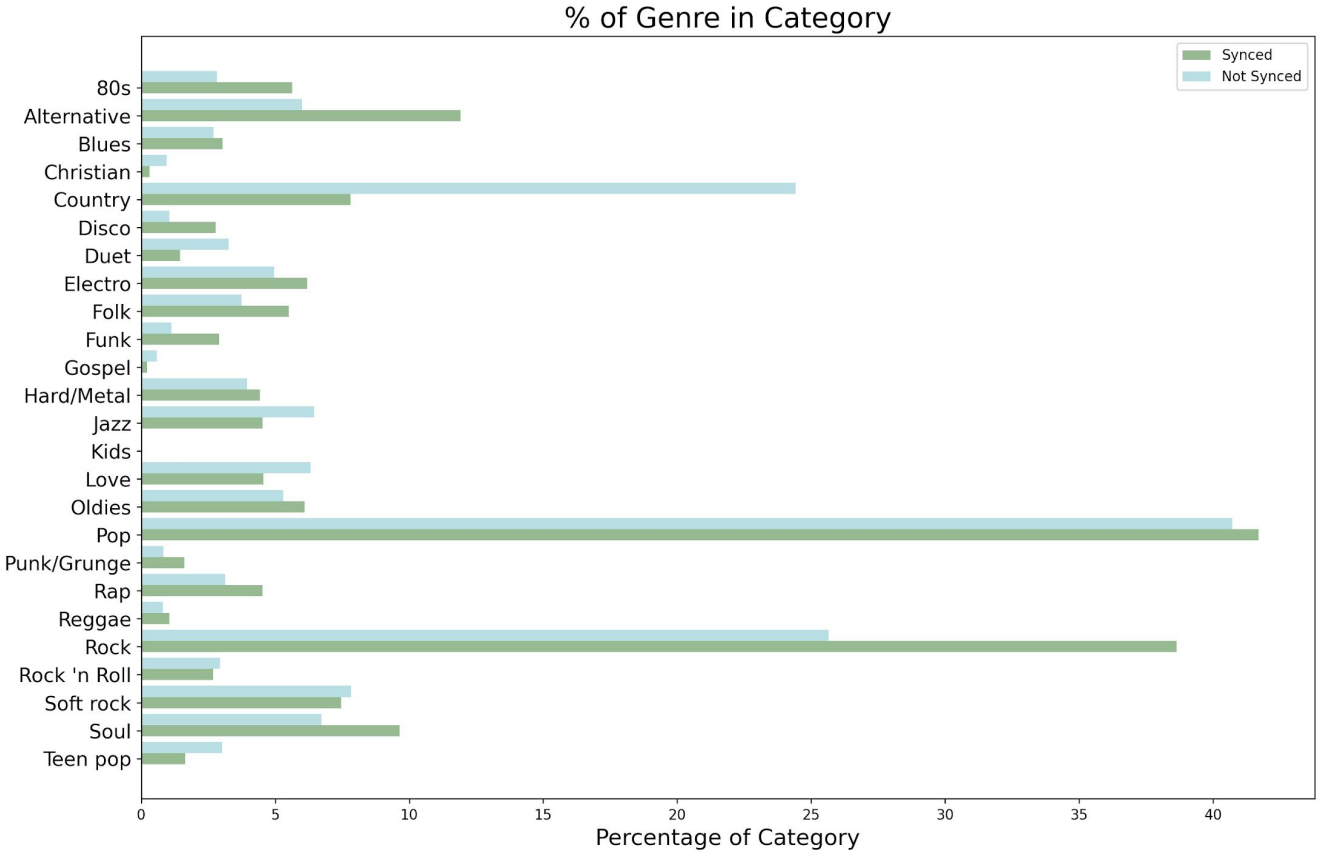
“The Full Monty Medley” - Tom Jones

8%

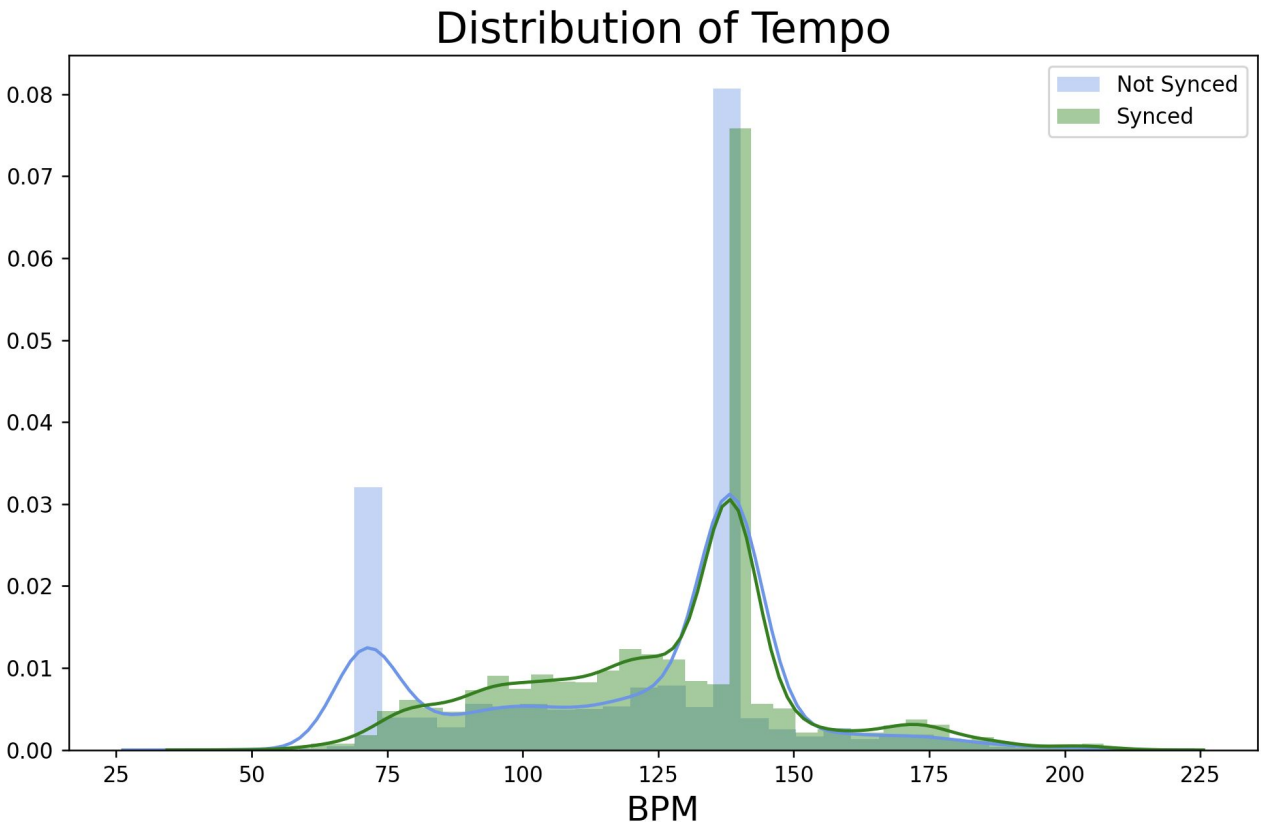
Exploring the Data: Release Year



Exploring the Data: Genres



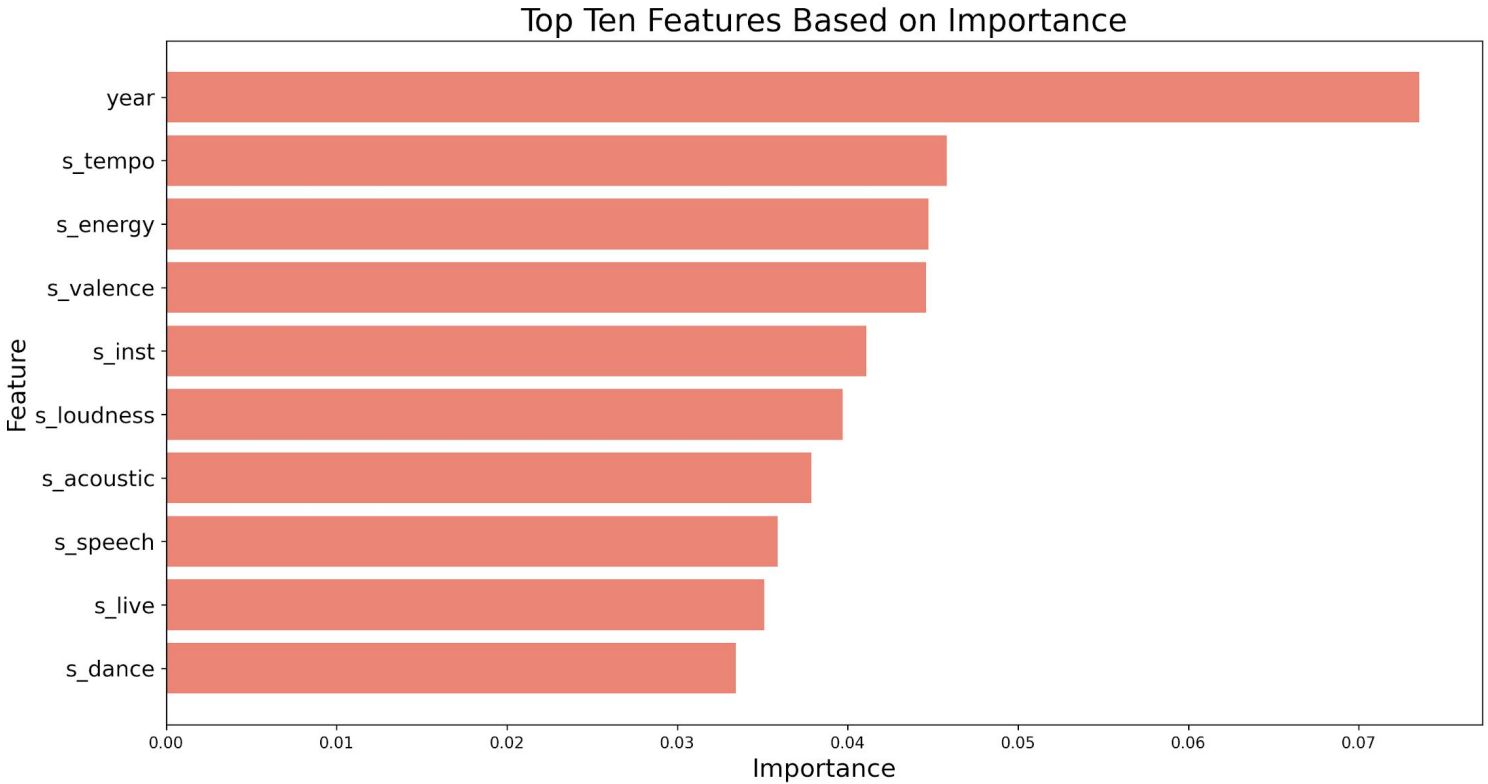
Exploring the Data: Tempo



Modeling

- Tried a variety of classification models and landed on Random Forests
- Scaled numeric features, Count Vectorized the artist, publishers, and writers
- Accuracy of 70%

Modeling: Feature Importance



How can we apply this information?

How can we apply this information?

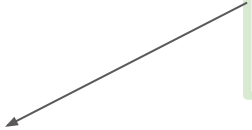
Sync Link Application

User Input:

- Song Title
- Artist

Sync Link Application

User Input:
- Song Title
- Artist



Master
- Searched on **Spotify**
- Matched to a master
- Outputs ISRC

Backup: Deezer

Sync Link Application

User Input:
- Song Title
- Artist

Master
- Searched on **Spotify**
- Matched to a master
- Outputs ISRC

Backup: Deezer

Publishing
- ISRC searched on **MusicBrainz**
- Matched to a composition

Backup: LyricsFreak

Sync Link Application

User Input:
- Song Title
- Artist

Master
- Searched on **Spotify**
- Matched to a master
- Outputs ISRC

Backup: Deezer

Publishing
- ISRC searched on **MusicBrainz**
- Matched to a composition

Backup: LyricsFreak

Additional
- Lyrics from **Genius**
- Artwork and audio from **Spotify**
- Sync Score generated with Random Forests **model**

Sync Link Application

User Input:
- Song Title
- Artist

Master
- Searched on **Spotify**
- Matched to a master
- Outputs ISRC

Backup: Deezer

Publishing
- ISRC searched on **MusicBrainz**
- Matched to a composition

Backup: LyricsFreak

Additional
- Lyrics from **Genius**
- Artwork and audio from **Spotify**
- Sync Score generated with Random Forests **model**

Sync Link displays: Master, Composition, Lyrics, Artwork, Audio & Sync Score

Demo

Sync Link Sources

SPOTIFY:

1. Artwork
2. Master Info
3. Embedded Audio

MUSICBRAINZ:

4. Publishing Info

DEEZER:

5. Additional Master Info

LYRICSFREAK:

6. Additional Publishing Info

GENIUS:

7. Lyrics

SyncLink:|

"Tennessee Whiskey" - Chris Stapleton

1

"Tennessee Whiskey" Lyrics

POWERED BY GENIUS 7

[Verse 1]
Used to spend my nights out in a barroom
Liquor was the only love I'd known
But you rescued me from reachin' for the bottom
And brought me back from bein' too far gone

[Chorus]
You're as smooth as Tennessee whiskey
You're as sweet as strawberry wine
You're as warm as a glass of brandy
And honey, I stay stoned on your love all the time

[Verse 2]
I've looked for love in all the same old places
Found the bottom of a bottle's always dry
But when you poured out your heart, I

Master	2 Spotify	5 Deezer
Song Title	Tennessee Whiskey	Tennessee Whiskey
Artist	Chris Stapleton	Chris Stapleton
Album	Traveller	Traveller
ISRC	USUM71418088	USUM71418088
Release Date	2015-05-04	2015-05-04
Is Explicit?	No	No
Track ID on Site	3fqwjXwUGN6vbzIwvyFMhx	98975170
Album UPC	00602547273123	602547273123
Label	Mercury Nashville	Mercury Nashville

Publishing	4 MusicBrainz	6 LyricsFreak
Composition Title	Tennessee Whiskey	Tennessee Whiskey
Writers	Dean Dillon, Linda Hargrove	Linda H Bartholomew, Dean Dillon
Publishers	EMI Algee Music Corp., Universal-Songs of PolyGram International, Inc.	Sony/ATV Music Publishing LLC, Universal Music Publishing Group
ISWC	-	-

Sync Score: Syncable

3

Next steps

- Build a larger and more accurate dataset
    - The “synced” vs “not synced” determination was based on one crowdsourced site. It doesn’t cover the scope of songs being used
- Add more sources to the app
    - Goal would be 4+ sources for master and publishing
- Deploy the app