Marc Stern
M.S. in Data Science, Montclair State University

CSIT 555 - Theory Paper
Dr. Katherine Herbert-Berger
December 2, 2020

The Data Takeover in the Music Industry

1. Introduction

Data persists as a main resource in diverse industries, globally. As consumer applications continue to develop, and users engage and continue to share data, data sources have become so large that it has taken on a new term, Big Data, which is the collection and analysis of vast quantities of data. The techniques used to analyze Big Data sources, including artificial intelligence, have enabled some of the most popular companies, and widely used applications and websites to learn from their user data. For the music domain, the move into the digital universe has transformed the music itself, and the music business. The growth of digital resources in the music industry has led to the development and use of data science approaches that help companies to gain knowledge and insights about their users, and to drive decision-making to improve their products. This paper examines how the music industry stores and uses Big Data and artificial intelligence approaches in different business contexts, including music streaming and personalization, music databases for royalties and copyright, and music distribution, Big Data collection for music insights and music labels, and music analysis.

2. Music Streaming and User Personalization

In the music industry, music streaming is a prominent example where we see the application that applies artificial intelligence to Big Data. Music streaming accounts for 85% of the industry revenue share in 2020 [1]. There are numerous streaming services, hundreds, in fact, each with their own positive and negative features. However, what they all have in common is that they store vast amounts of data containing information, for example, about songs, artists, and musical genres, and the relationships between them. The leading streaming service is Spotify. This service accounted for 36% of the market share in 2019 [2]. Not only does Spotify store a large catalog of music in its database, but it also gathers massive quantities of user data collected in real time. The main purpose of this for Spotify is to learn about its users in great detail, and gather insights about how their users navigate through with the application, as well as what they like to listen to.

Spotify's dedication to user personalization is what has propelled the company into the technological spotlight. The most common and traditional approach to personalized experience

is exploitative recommender systems. These systems are designed to recommend items or content that have the highest probability of user selection based on the data it collects about the user engagement with the app, and the current context. A newer approach, referred to as the *bandit method*, uses a balance of exploitation and exploration, where the system recommends content it is unsure the user will select in order to learn more about the user. The purpose of this approach is to further evolve the personalized experience. Spotify researchers explain that, “recommendation methods that aim to maximize engagement without regard for model certainty or exploration sometimes unnecessarily ignore highly relevant items” [3]. As of 2018, Spotify researchers created a system called *BAndits for Recsplanations as Treatments (Bart)* that has shown significant improvement over previous recommendation methods of personalization [3]. This is accomplished through implementing a third key concept of recommendation, explanation. In [4], the authors note that, “explanations help the user better understand and interpret the rationale of the recommender system, thereby making it more trustworthy and engaging” [3]. Spotify’s goal with Bart “is to learn how items and explanations interact within any given context to predict user satisfaction” [3]. This method of user recommendation is used throughout the Spotify application.

Users first interact with Bart on the home page. When opening Spotify, users are given personalized home pages that contain exploitative categories, such as *Shortcuts*, *Recently Played*, and *Your Top Shows*. These include playlists such as, *Jump Back In*, and *On Repeat*. The idea is to draw in users based on what they have chosen previously. A user even has the choice to save their own library of music and playlists on the left hand side of the application, ignoring any new possibilities. However, Spotify’s personalization engine still captures this. Explorative options include categories like, *New releases for you*, *Album picks*, and *Shows you might like.* These categories try to suggest new music and podcast shows based on what a user has chosen in the past. The recommender system can then nudge a user to select new data points, from which more user data will be collected and used to adjust future personalized experiences. This is beneficial not only to learn about a user’s preferences, but also to adapt with the user as their preferences change over time. Furthermore, Spotify will also suggest categories and playlists that are popular in order to engage and study the social behavior of a user. This also helps Bart to learn as much about the user as possible.

The order of exploitative and explorative options is personalized as well. This ordering is an aspect of the explanatory process and engages user confidence in the recommender system. For instance, users who only listen to Classical music are less likely to want to engage with the personalized experience if the first recommended playlist they see contains currently popular Hip-Hop music. Likewise, users who listen to a wide variety of music and podcasts will likely be frustrated with too many exploitative recommendations on their home pages.

Another popular feature where the Bart system is prominent is with Spotify Radio. This concept is where a user can select an artist or a song, and from there a personalized playlist is created. Spotify Radio also kicks in automatically after a queue of selected songs ends. As Spotify Radio plays, users interact with Bart by skipping, liking, sharing, saving, and playlisting songs. Through its balance of exploitation and exploration, the personalized experience improves and adapts. The more a user interacts with the application, the better the

recommendations should become. Ultimately, this should lead to the user's greater satisfaction interacting with the Bart system. It becomes easier to understand how a satisfied user might use the application more, thus improving the personalized engine, and perpetuating a cycle of satisfaction. This is Spotify's goal with Bart.

A third important feature is the *Discover Weekly* playlist -- a weekly playlist of mostly new music customized for every user. This is the ultimate test for the exploration part of the Bart system as it has to put all the pieces together to suggest a full playlist of music that will satisfy the user. *Discover Weekly* is one of Spotify's strongest tools for data collection, and is crucial to keep its users coming back to listen every week. The benefit of Bart is that it does not try to be perfect, rather it attempts to entice the user into selecting items where the system is unsure; the Bart system can then learn from this selection. However, with this playlist, it is necessary to strike a balance between songs the user likes and songs they skip, too many skips and the user could lose confidence in the algorithm. This could deter them from listening to the playlist next week when it updates. For some users, too many songs they have already heard can be disappointing for a discovery-focused playlist. This juggling act of exploitation and exploration is the essence of the Bart system.

In addition to Spotify's user personalization, what makes the company an interesting case study on data usage in the music industry is their dedication to sharing data and integrating with other technologies. Spotify for Developers is a website that allows users to query the Spotify database through the Web API for their own research. Documented on their website, the company shows just how much data they make available. Categories of data include, *Albums*, *Artists*, *Episodes*, as well as *Browse*, *Search*, and *Playlists.* Spotify's new Application Program Interface (API), currently in beta, will include access to user playback, enabling users with stop, start, skip, and access more playback functionality [5].

Spotify's API is also widely used for application development. Specifically, many applications seek to implement Spotify's catalog of music as a feature. Spotify's online documentation states, "Web API endpoints enable external applications to access the Spotify catalog and user data. The endpoints are arranged in a structure defined by an object model" [5]. Instagram allows users to add music to their stories, while Spotify returns the favor and gives users the ability to share directly to Instagram Stories, as well as to other social media applications, such as FaceBook Messenger, and Snapchat. While Spotify's popularity has led to its integration in many popular applications, the Web API is also compatible with libraries from a variety of computer programming languages, including, but not limited to, Python, Java, and PHP, allowing a diverse user base to develop new and creative applications.

Spotify for Artists is another data-driven resource for the creators of Spotify's content. Artists are the users who distribute their music on Spotify. These users are likely to be different from users who frequent Spotify for Developers. Therefore, it makes sense that they can access a different view of the database. Spotify for Artists serves three main functions for artists. The first is the ability to update artist profiles with pictures, descriptions, artist playlists, and also a bi-weekly Artist's Pick, where an artist can link any song in the catalog to their Spotify page. The second is to provide artists with links to additional resources for independent artists (versus music-label-backed artists). And the third, and most important function, is to show artists a

variety of meaningful statistics and visualizations about their listening data and fan engagement. This serves as an essential tool for artists to gain important insights about the success of their music. This information can potentially help artists to make critical decisions about marketing campaigns, tours, and other business decisions to build out their fan base. Tools like these help level the playing field between independent artists who often have lower budgets than label-backed artists, and are crucial in Spotify's effort to satisfy its users, and its overall success in the streaming industry.

A final note on Spotify for Artists is that the application allows users to submit up to one song at a time for Spotify editorial playlist consideration. This is significant in an industry where playlists are gaining popularity, and an application, where, in 2019, about two thirds of the total minutes spent listening was from playlists [7]. In order for Spotify's team to organize the countless submissions they receive, they require the user to fill out a form, answering questions about song genre, style, lyrics, instrumentation, moods, and descriptors. By collecting this data, they are tagging the submitted songs with meta-data, or self-describing data, and further allowing their Big Data analytics processes to sift through and organize the songs, filtering them to the various teams of editors who ultimately choose and order the music to fit a highly specific mood, genre, or theme. When an artist finds success through this path, they will feel more satisfied with their user experience, and confident that Spotify can play a beneficial and direct role in their personal success. This is yet another example of user interaction, engagement, and satisfaction.

3. Music Databases: Royalties and Copyright, and Distribution

In the music industry, a recorded song contains two separate copyrights, one for the mechanical phonograph, and the other for the songwriter. The main function of Performance Rights Organizations (PROs) is to keep a record of copyrighted works, and to make sure that the individuals who are responsible for the songwriting of these works are properly compensated. There are many instances where recorded music is performed, such as licensing to a TV show, a DJ set, or a radio station playing in the mall. These are cases where copyrighted works are being broadcast as part of an environment designed to make money, hence why the copyright holder is due a royalty. In order to accomplish this, PROs have to keep a robust database of music, and are responsible for collecting data about the various methods of recorded music performance and reporting it such that artists can see their rightful monetary gains.

In 2017, the two leading PROs, ASCAP and BMI, representing more than 22 million copyrighted works, announced that they would merge their databases to create a more robust, complete, and unified database in order to reset the industry to "offer greater transparency to music users and the industry" [6]. When dealing with finances and the people who are obligated to them, it's important that the system is clear and transparent. With a unified database, there is far less room for error and confusion, and more capability to cross reference copyrighted works. It is essential that these organizations not only track how often recorded works are performed,

but also how those individual works get divided between different entities, including both individuals and corporations.

The joint database effort is also ongoing and only the beginning for this new era of data transparency. The joint announcement also highlighted that “future phases will explore customizable, interactive API solutions and the potential inclusion of other databases” [6]. ASCAP and BMI recognize that data is an essential part of the music industry now and moving forward, and with the refinement of Big Data collection and analysis techniques over recent years, the two competitors have decided to actually work together in sight of leading their industry to a better place. In an industry that has historically taken advantage of its content creators, it is refreshing to see new leadership emerge in favor of the artist.

Another important role in the music industry is that of the music distributor. Serving as a liaison between artists and digital stores, including streaming services like Spotify, music distributors serve to do the work to actually make an artist’s music available online. While websites like Youtube and SoundCloud allow users to upload music directly for free, they also do not provide a direct route for artists to collect royalties, not to mention their notoriously low royalty payout rates. Paying to upload new music through distributors has many benefits, notably that the distributor will get approval to submit an artist’s music on every available store, and also that it will collect and pay out one’s deserved royalties.

Distrokid has emerged as a leader in its field due to its easy-to-use user interface for uploading music, reviewing streaming and sales statistics, and collecting royalties. The latter is as simple as connecting a PayPal or bank account. As far as uploading music, Distrokid makes it easy for users to upload the proper data, tag the correct meta-data, and send it on its way. Songs that have copyright issues will be flagged before they are live in stores. In the same vein, cover songs and heavily sampled works will go through a clearance process without the user having to do anything. It’s simple to use, and helpful for independent artists, keeping them away from legal copyright disputes. As of 2017, Distrokid also backs up numerous copies of music and the correlating upload data “across multiple data centers using the highest grade of Amazon AWS S3 storage” [8] in what they call the Archive. Founder Philip Caplan also writes this archive is cloud-based, so the backup files can be downloaded from any device by the appropriate user [8].

In 2018, Spotify took note of Distrokid and decided to purchase a minority stake in the company [9]. In doing so, the application has been able to expand and integrate its features directly into Distrokid. A new artist uploading music for the first time, for example, can now create a Spotify for Artists account as soon as their music has been processed for release by Spotify from within the Distrokid interface. This pairing between music distributor and streaming service has enabled the industry to continue progressing toward a transparent and unified digitized and data-driven approach.

With regard to royalties, Distrokid automatically assigns each song an International Sound Recording Code (ISRC). This is a unique identifier that helps Distrokid, as well as all digital technologies that employ sound recordings, properly organize the songs, so that the proper corresponding streaming and sales data can be collected and converted into royalty

payments [10]. Distrokid also keeps a record of the royalty percentage splits per song among users, and shows its users a breakdown of where their music royalties are coming from.

It is interesting to see that not only does the process through which royalties are generated and paid out depend heavily on Big Data storage and analytical techniques, but also that consumer level tools that are being built, namely, the Streaming Royalty Calculator. This is a free website that allows users to input a quantity of streams and select a popular streaming service, and it returns the expected amount of money owed in royalties. Streaming services have different royalty payout rates, and they change slightly every quarter, but this is a good estimate of what an artist can expect. There are also multiple websites that accomplish this function. This is another example of how data-driven tools and industries can provide more transparency during important processes such as music royalty distribution and collection.

4. Big Data Collection & Music Business Decision Making

4.1 Music Insights for Artists & Music Companies While there is significant competition in the market for Big Data collection in the music industry, Soundcharts is the largest music data company. The French company collects trillions of data points from social media applications and streaming services to analyze and display to their users. Soundcharts' users are typically music companies, such as music labels, seeking insights regarding online music trends. Using a variety of APIs from other companies such as Spotify and FaceBook, Soundcharts offers insights regarding social media engagement and streaming. They track nearly 1.5 million artists' followings on seven different social media platforms, including Instagram, Twitter, and Tik-Tok. They also record data from over 2.5 million playlists across five streaming services, and nearly 9,000 charts across seven streaming and social media platforms. Moreover, Soundcharts has servers throughout the world in order to include radio play data from 69 countries, and website mentions from 159 countries [17] [18].

The radio play technology is powered by a tool called ACRCloud. ACR stands for Automatic Content Recognition; it creates an acoustic footprint (see Section 5 for more detail). The general concept is that recorded music can have a unique identifying binary code that can be matched in a database where the song information is stored [20]. A song played on the radio can be analyzed by the technology and matched in the database to the information regarding its title or writer, for example. The acoustic footprint is based on the music itself. Because ACRCloud is a tool particularly helpful in identifying radio play, PROs across the world, such as Malaysia's Public Performance Malaysia, and Portugals' Pass Música, use the software for their business purposes, exemplifying the variety of ways a data-driven music tool can be implemented industry-wide, and even beyond (ACRCloud is used in advertising and tv industries as well).

Soundcharts' product is not particularly innovative, but it has become a staple for decision making for many industry-leading companies. Over 250 companies subscribe to Soundcharts, ranging from major music labels to independent publishers and management companies [18]. This dependence demonstrates the importance of user data capture in the

music industry. It is a basic tool that provides its users with the baseline level of information, so that a variety of knowledge can be learned and applied to to gain insights from different stakeholders in the music industry, from independent artists to large music companies.

Blog posts are another important metric for understanding the music listening base, but are often not represented in data collection. This makes sense as there is no unified online location or application storing a majority of blog posts regarding the music industry. However, with different techniques, it is a possibility. Hype Machine is a website that uses web scraping techniques to collect information about which songs are being written about in blogs, and what is being said about them. Blog posting is a common way for less-than-popular music to gain traction within a specific sub-genre's community of fans. Hype Machine aggregates their data from a network of blogs and compiles it into a ranking system, or a chart. The Hype Machine charts can be likened to Billboard's Top 40 or Hot 100 charts, except they feature charts of different genres, and the selected songs are based on a different set of criteria. While the website is not sensationally popular, Hype Machine has a unique role in the music industry. For many artists, it is a milestone to be on the Hype Machine charts, and it is also not uncommon for artists who top the Hype Machine *Popular* chart to break into the mainstream [19].

4.2 Music Labels & Data Insights One notable group who is relying heavily on data are music labels, such as Universal Music Group (UMG), or Atlantic Records. Traditionally, Artist and Repertoire (A&R) has been a highly subjective field. Before the internet and mass data collection, music labels had to rely upon professionals who used their networks to learn about new bands who were not affiliated with a music label by physically attending their mid-sized shows. If a music label A&R representative was impressed by what they heard and by the reception from the crowd, they might feel compelled to sign the artist to their label and have a stake in their music career. This style of working, however, was not rooted in science. Some A&R representatives were better at sensing an "it" factor than others, but there has yet to be a way to objectively measure the "it", and model it to predict "the next big artist." For major record labels especially, using resources such as Soundcharts is an effective way to track trends in order to make decisions about resource allocation, and to decide which new artists to sign. With the increased number of independent artists who release their music into the public through mediums like Spotify, YouTube, and SoundCloud, as a result, there are a large number of artists who are not yet signed with a label, but who have significant data suggesting the trajectory of their success. The introduction of data to this decision-making process removes a layer of subjectivity related to artist success. Labels have a tendency to steer their sound toward the mainstream as it evolves, but with data analytics so readily available, executives have more objective insights about new artists, enabling them to make more informed business decisions. Artists making genres of music that are not mainstream have a better chance of being discovered by a label now than prior to the availability of data analytics.

Labels have extended their decision making to social media data as well [21]. Soundcharts supplies its users with social media data, because the music labels and other music users are interested in how social media can correlate with an artist's success. It is

considered an industry standard for an artist to have at least 40,000 followers on social media before a label is interested in their music. Google search and Wikipedia viewing are also important indicators tracked to understand artist popularity. This shows the level of commitment major music labels are making towards a data-driven business mindset.

After signing artists, labels will use locational data to understand how to best use financial resources for an artist. Labels can learn about the region where their artists are most popular. They can then focus their spendings in those areas. As mentioned earlier regarding independent artists, tour routes and festival bookings are impacted by the locational data of the online listeners of an artist. Festival lineups are often based on genre and ordered by popularity -- a metric determined by sales, streaming, social media presence, and more.

Indify is a company who tries to help artists connect with music publishers, management groups, and even labels. Before the company transitioned to focus their attention on helping artists, they were merely a company who used Big Data to predict the future success of tens of thousands of artists they deemed “up-and-coming” [21]. Using the same data sources and techniques as previously mentioned, Indify gained the industry’s respect and attention when they featured an up-and-coming artist named Kai, who would a year later release his debut album under the name Khalid. Since this release, Khalid has become a global icon, and the ninth most streamed artist in 2020 [22].

The music market has become overwhelmingly saturated with talent, so it is no surprise that labels are heavily using streaming and social media data to make their decisions. It is also interesting to see them embrace predictive modelling in an industry that is rooted in creativity and human connection. However, the data are compelling, and businesses are always looking for a competitive edge. The time has come for Big Data tools to become that edge.

5. Music Analysis

Data analytics and computer algorithms can be used more than business purposes. They can also be used to analyze music. One of the most widely used music technologies is Shazam. This well-known application can capture a 20-second snippet of music and tell the user the song information with high accuracy, and in about 6 seconds on average [11]. To understand how Shazam works, we need to briefly enter the world of audio signal processing.

Upon opening Shazam, a mobile application, a user records the signal coming from the music using their built-in phone microphone. As with any microphone, the difference in air pressure level (what the human brain interprets as sound) as the song travels through the room is converted into a continuous electrical signal. From here an analog to digital converter analyzes the continuous signal and converts it one step further into a discrete binary number representing the amplitude of the signal at a given moment in time. Amplitude can be measured over time, and its output looks like that of an audio file. Now, that the audio file has a shape, there is still one important element missing that would help to identify a song: frequency. Songs not only have changes in volume, but they also have melodies and harmony. The Discrete Fourier Transform (DFT) method mathematically converts the amplitude measurements over

time (time domain) into frequency measurements over time (frequency domain), based on the Fourier Analysis, which states that any complex waveform is the sum of multiple simple, sinusoidal waveforms, meaning the wavelength represents a single, fundamental frequency. With this foundation, a more popular method of this conversion from the time domain to the frequency domain of the Fourier Analysis has been developed, called the Fast Fourier Analysis (FFT). A popular variant of FFT, the Cooley-Tukey algorithm, runs in O(n logn) time, much more efficiently than the $O(n^2)$ run time of DFT. It's worth noting that libraries for FFT exist for C, C++, Java, Python, and Ruby. An issue with FFT is that it can lose information about the timing of frequencies if it has to complete a time to frequency domain conversion over a long period of time such as a three-minute song. This is compensated for by analyzing tiny slices of music sequentially, instead of the whole song at once, called chunks. The Red Book standard, the specifications for CD, requires that digital recordings are made at 44,100 samples per second, and a bit depth of 16 bits [13]. This can be thought of as audio resolution, and dynamic amplitude range, respectively. A 4 kb size chunk of music allows an FFT analysis to convert 44 chunks per second at Red Book standard, and is large enough for proper audio identification. From here, small ranges of frequencies are selected in order to identify algorithmically the frequencies in a given song that are deemed foundational. For each chunk, the prominent frequency is noted and stored with the chunk. Piecing together all the chunks of a song in the end totals to an acoustic footprint of data that is stored in a hash table in the database corresponding to the song meta-data, such as title, name, album, year of release, etc. Shazam runs its audio catalog through these processes and stores them in its database for matching later on.

Interestingly, Shazam accounts for what is known as a *fuzz factor*, which assumes that the conditions for audio recording into the application will not be ideal. Suppose a user is using Shazam in a noisy restaurant, the fuzz factor is an additional algorithm (or many algorithms) trying its best to isolate the music in the room. Of course, there is a threshold at which the fuzz factor is too great, and Shazam cannot properly convert the music into an identifiable match in the database. This complicated process that reads changes in air pressure level, and ultimately converts it into a digital stamp of data is an extraordinary way data is used to analyze music itself [12].

In addition to the functionality of the application itself, Shazam also collects data about its users, and their activity with regards to the music they hear and wish to learn more about. Due to its strong marketplace success, there is a notion that people who "Shazam" a song are enjoying it, or are interested in it in some way. The music industry values Shazam as a source of information from which it can learn about the listener base. Popular data collection companies like Soundcharts and Chartmetric organize and present Shazam data as a valuable metric for its users, which they can further use to help model business decisions.

While Shazam has been officially released for 16 years, a new startup, Secret Chord Labs, is working to release a beta of its music analysis software *dopr*. Using techniques of artificial intelligence and Big Data, Secret Chord uses models based on neuroscience research in order to predict the commercial success of unreleased music. As most of the industry focuses on data collection from streaming services and social media to make business decisions a

reactive process, Secret Chord's team is providing a proactive alternative for data-driven decision making.

Similarly to Shazam, dopr analyzes the audio waveform in tiny chunks to precisely measure the amplitude and frequency content of a moment of a song, and accurately correlate the data to a meaningful musical experience. However, Secret Chord Labs' technology is designed based on the key concept of harmonic surprise. This notion hypothesizes that a combination of increased harmonic complexity, and "tension" and "release," are enjoyable musical experiences. This surprise also triggers measurable changes in brain waves that would suggest a level of preference by a listener [15]. By measuring harmonic surprise, as it relates to musical preference, statistical predictions can be made about the likelihood of a target audience enjoying a particular song [14]. Music is interpreted differently between cultures, so the context of who the audience is, and where they are located are key variables in the formula for determining a listener's musical expectation, and consequently surprise and enjoyment. Secret Chord Labs relays this information to the user in the form of a score, along with various other visualizations to help the user understand their score. As with any technology rooted in artificial intelligence, dopr will continue to grow over time. As more data is analyzed through the model, the more accurate the model can become. Future updates for the model will add new algorithmic components based on additional elements of surprise, including, rhythm, melody, and timbre [16].

Secret Chord Labs represents innovation in the space of music technology. Their goal is for major record labels to depend on their software to drive business decisions like signing artists, and also to be integrated into a software such as Spotify, who can help spread its usage over a more diverse population of artists.

6. Conclusion

In its current state, many aspects of the music industry are now driven by data in some way. Spotify uses Big Data and artificial intelligence to install user personalization, and has been able to tilt the music industry towards streaming. Competing leading Performance Rights Organizations decided to unify their databases to improve transparency regarding royalty collection and distribution. Leaders across the music industry sectors are embracing the role that data is playing currently, and are seeking to leverage this trend of using data to support their users, and draw insights from their data to make business decisions.

References

[1] J. Friedlander. (2020). *MID-YEAR 2020 RIAA REVENUE STATISTICS.* [Online]. Available: https://www.riaa.com/wp-content/uploads/2020/09/Mid-Year-2020-RIAA-Revenue-Statistics.pdf

[2] T4 Labs. (2020, August 16). *Music Streaming Market Share.* [Online]. Available: https://www.t4.ai/industry/music-streaming-market-share

[3] James McInerney, Benjamin Lacker, Samantha Hansen, Karl Higley, HuguesBouchard, Alois Gruson, Rishabh Mehrotra. 2018. Explore, Exploit, and Explain: Personalizing Explainable Recommendations with Bandits. In Twelfth ACM Conference on Recommender Systems (RecSys '18), October 2–7, 2018, Vancouver, BC, Canada. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3240323.3240354

[4] Pigi Kouki, James Schaffer, Jay Pujara, John O'Donovan, and Lise Getoor. 2017. User Preferences for Hybrid Explanations. In Proceedings of the Eleventh ACM Conference on Recommender Systems. ACM, 84–88.

[5] (2020). *Web API Reference.* [Online]. Available: https://developer.spotify.com/documentation/web-api/reference/

[6] ASCAP press. (2017, July 26). *ASCAP & BMI Announce Creation Of A New Comprehensive Musical Works Database To Increase Ownership Transparency In Performing Rights Licensing.* [Online]. Available: https://www.ascap.com/press/2017/07-26-ascap-bmi-database

[7] Mansoor Iqbal. (2020, October 30). *Spotify Usage and Revenue Statistics (2020).* [Online]. Available: https://www.businessofapps.com/data/spotify-statistics/#2

[8] Philip Caplan. (2017, June 12). *Introducing DistroKid Vault.* [Online]. Available: https://news.distrokid.com/introducing-distrokid-vault-89eb4d3474df

[9] Tim Ingham. (2018, October 17). *Spotify acquires minority stake in DistroKid, enabling artists to upload to multiple services… from Spotify.* [Online]. Available: https://www.musicbusinessworldwide.com/spotify-acquires-minority-stake-in-distrokid-offering-independent-artists-to-upload-to-multiple-services/

[10] (2020). *What is an ISRC?* [Online]. Available: https://www.usisrc.org/about/index.html

[11] Laurie J. Sampsel. "Beyond Popular Music: How Do Shazam and SoundHound Compare?" *Music Reference Services Quarterly.* Vol. 19, Issue 2. Mar 15, 2016. [Online]. Available: https://doi.org/10.1080/10588167.2016.1167426

[12] Jovan Jovanovic. *How does Shazam work? Music Recognition Algorithms, Fingerprinting, and Processing.* [Online]. Available: https://www.toptal.com/algorithms/shazam-it-music-processing-fingerprinting-and-recognition

[13] TravSonic Sound Studios. (2020). *Red Book CD Format*. [Online]. Available: https://www.travsonic.com/red-book-cd-format/

[14] Scott A. Miles, David S. Rosen, Norberto M. Grzywacz. "A Statistical Analysis of the Relationship between Harmonic Surprise and Preference in Popular Music," *Front. Hum. Neurosci.* May 18, 2017. [Online]. Available: https://doi.org/10.3389/fnhum.2017.00263

[15] David S. Rosen. "Behavioral Evidence of Harmonic Surprise Effects on Music Preference." 2017. unpublished.

[16] David S. Rosen. “SCL Deck.” Nov 30, 2020. unpublished.

[17] Soundcharts. (2020). *Data Resources.* [Online]. Available: https://soundcharts.com/datasources

[18] Romain Dillet. (2017, September 12). *Soundcharts is an analytics tool for music artists and labels.* [Online]. Available: https://techcrunch.com/2017/09/12/soundcharts-is-an-analytics-tools-for-music-artists-and-labels/

[19] Jeff Zeller. (2015, November 16). *HypeMachine: Using Data and Aggregation to Surface The Next #1 Song.* [Online]. Available: https://digital.hbs.edu/platform-digit/submission/hypemachine-using-data-and-aggregation-to-surface-the-next-1-song/

[20] ACRCLoud docs. (2020). *Audio Fingerprinting.* [Online]. Available: https://docs.acrcloud.com/docs/acrcloud/introduction/audio-fingerprinting/

[21] Shawn Setaro. (2019, September 5). *How Data Is Making Hits and Changing the Music Industry.* [Online]. Available: https://www.complex.com/music/2019/09/data-changing-music-industry

[22] Sophia Silva. (2020) *Top 20 most popular artists on Spotify in 2020!* [Online]. Available: https://www.apptuts.net/en/tutorial/android/most-popular-artists-on-spotify/