SPOTIFY POPULARITY ANALYSIS

INTRODUCTION

"Music is changing so quickly, and the landscape of the music industry itself is changing so quickly, that everything new, like Spotify, all feels to me a bit like a grand experiment" - Taylor Swift (excerpt from the Rolling Stone article from 2014, "Taylor Swift Shuns 'Grand Experiment' of Streaming Music" by Kory Grow).

The music industry has been completely upended by digital technologies. The value of the worldwide music industry was $25.2 billion in 1999, but in 2017 it was just $17.3 billion, or little under 75% of that amount (IFPI Global Music Report 2018). Records, CDs, and other physical items in the music industry saw a steep fall as digital technology spread throughout the world. Customers started using downloads as a more practical and frequently less expensive way to get music. After widespread concern about the state of the music business following the introduction of Napster, a number of music streaming services—most notably, Spotify—appeared in the late 2000s.
As of 2018, streaming platforms represent the largest revenue stream for the global music industry, making up 38% of total revenue (IFPI Global Music Report 2018). Thus, the economic relevance of music streaming platforms and their effects on the market has grown considerably.

Spotify dominates the music streaming market and serves as one of the main avenues through which over 200 million global users find new music. Users can follow artists, playlists, look at their country's top 200 daily hits, and hunt for music via a search function. Rather than having access to a few thousand records at a local music store, streaming platform users have access to millions of songs right at their fingertips. Considering that users can access nearly any song, the following question arises: what song attributes affect a track's popularity and how does that differ between countries? More specifically, which factors affect popularity on Spotify?

Previous studies in this field have examined variables influencing song popularity based

on data obtained from the Billboard Hot 100 or other popular music charts, but none have looked specifically at streaming platforms. Further, no previous studies have compared determinants of song popularity between different countries. I fill this gap by applying previous studies' methodologies to Spotify-specific data that is segmented by country.

ABSTRACT

Music is the most accessible and widely popular form of artistic expression, affecting our the lives of millions everyday. While music dates back as far as the modern human, it has also had mass effect on society in the last 10 years - making it vital for us to understand its trends and their significance going into the future. Many statistical analyses have been conducted on

musical trends, such as Shane Snow's analysis of key verbiage and content in songs since 1965. These works generally focus on the qualitative attributes such as verbiage of musical pieces. While powerful, this doesn't provide a analysis of songs in their purest form - their generalized attributes. Our approach aims to analyze the way songs are, not the "certain qualities a song holds". We hope to utilize this different approach to not just define historical music trends,

LITERATURE REVIEW

Song popularity and its determinants have been studied extensively across differing academic fields. Sociologists have examined the social circumstances that influence a song's success as well as the optimal level of differentiation in popular music. Data scientists and economists, however, have diverged in the subjects of their research. Most data scientists have examined the impact of a song's features on their success in charts such as the Billboard Hot 100, whereas economists have focused on the impact of streaming platforms on the music industry. Much of the existing research in this field ignores one of the main platforms from which people source their music – Spotify. A key determinant of a track's success is social influence. In an experimental study of an artificial music market, Salganik et al. (2006) found that consumer choice is largely a function of how many times a song has been downloaded by other users as well as their own music preferences. Participants of this study were randomly sorted into two groups that differed only in the amount of information available to the consumer. In the independent group, consumers only had access to the name of the band and song. Participants in the treatment group, however, had the same information in addition to the number of times a song had been previously downloaded, thereby serving as a proxy for popularity. Salganik et al. (2006) found not only that the number of downloads affects how highly people rate their enjoyment of a song but also that the degree to which a consumer can determine how other users view a song affects the magnitude of its impact on popularity. Songs with a higher number of downloads (or streams) breed more success because consumers assume that they are better than other songs, thereby contributing to a superstar effect in the music industry. but identify trends to predict the how music will be in the coming years. We believe that our key results could have great value to the music industry by giving an early glimpse of comprehensive music trends. This would allow artists, producers, and writers alike be able to know their audience on a greater level, and create songs that are predictably successful.

PROBLEM FORMULATION

Our projects' goal is to analyze quantifiable trends in music over the past many years. We hoped that through our analysis we could identify these quantitative trends in past music to predict both quantitative and qualitative qualities of music to the future. To do this, we needed to find a large dataset of songs that are evenly distributed over many years, genres, and qualities. Only then could we ensure that we are correctly analyzing past musical qualities in an evenly distributed fashion. As well, the dataset must be able to be used in-line with the Spotify API.

CONCLUSION

Over the course of finding and collecting our dataset, we have learned a number of lessons. One of the biggest lessons is how inconvenient data rate limiting can be. In our initial search through the Spotify API, we were only allowed one query at a time, which ended up causing our first script to run for 48 hours. In our second query of the Spotify API, we were able to take advantage of a different API request method that allowed us to query 50 song ID's at a time, which greatly reduced the run time. We also learned that data cleaning is a much more important step in the process of retrieving a dataset than we first anticipated. Throughout multiple steps in the process, there were many extra bits of unnecessary or redundant information that needed to be filtered out. Had we not caught these redundancies, it is likely our final analysis would be dramatically skewed. It was particularly important to make sure this step was done correctly in order to make sure results could actually be pulled from the Spotify search.