STEVENS INSTITUTE of TECHNOLOGY
School of Business

DECEMBER 20, 2022

SPOTIFY ANALYSIS

FINAL PROJECT M10.7

YASHKUMAR CHAUHAN
BIA-610 APPLIED ANALYTICS
Stevens Institute of Technology – Fall 2022

Revision: 1.0

Table of Contents

Table of Tables

Table of Figures

1. Introduction

1.1 Problem Statement

Music has been an indispensable element of our everyday lives for hundreds of years after Shakespeare. There isn't a day that goes by when you aren't singing or listening to music. Our wellspring of paying attention to them has also evolved with time. From rare Gramophone mix recordings to radio, our listening medium has progressed to internet music streaming. The most well-known of them include Spotify, Apple Music, Google Play, and others.

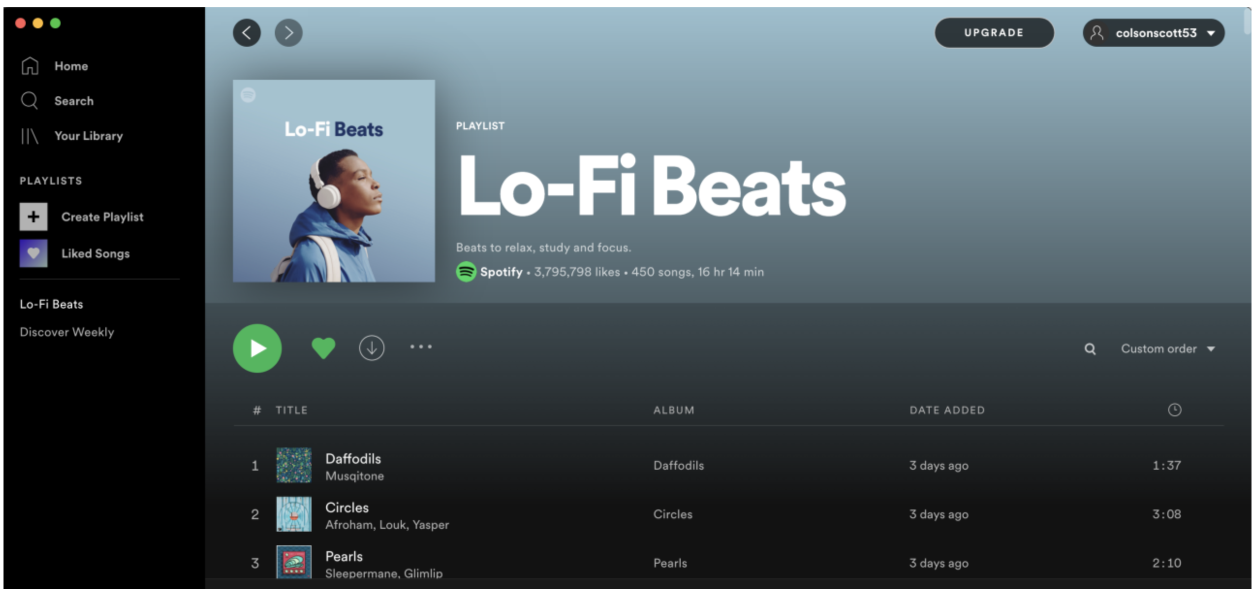

Figure 1 – Spotify UI

Spotify has the greatest position, with more than a 36% share of the whole business among online music endorsers and a fan following of more than 100 million.

I wanted to look at another facet of a song's hype. It is difficult to determine what makes a song trendy, regardless of its quality or the skill involved in its composition and performance. As a musician, I've always been baffled about what makes a song appealing to a broad audience.

So, one problem that is noticed is that is worth investigating and its: ***"Improve Spotify users' ability to locate songs to listen to."***

1.2 Plan of Execution

I genuinely love music thus I thought this would be a silly and exciting test to deal with. After spending some time investigating what material was available for download from data provided by professor, I realized that I needed to investigate how music tastes had evolved through time. The Spotify API allows you to get Audio Features (danceability, energy, rhythm) for each music, so I thought it would be interesting to see how these have evolved over time.

1.3 Summary

For the analysis part, the respective procedure is followed here:

- An overview of the genres and the artists
- An examination of the track's popularity and how it varies between subgenres.
- The Evolution of the Audio Feature

2. Data Preparation

2.1 Data Source

- Data source for this analysis and report is from file provided by professor: https://sit.instructure.com/courses/61293/pages/m9-dot-9-final-project?module_item_id=1566896
- Download spotify_songs.csv

2.2 Explanation of Data Source

From 1957 to 2020, the initial data set was compiled from the Spotify API of several distinct data points for visualization and modeling. Genres were also picked from Every Noise, an intriguing depiction of the Spotify class space maintained by a kind of taxonomist. The finest four sub-categories for each were used to ask Spotify for 20 playlists each, resulting in about 5000 melodies for each kind, spread over a different sub-type of space. Making the dataset public allows people to examine the data and derive meaning from the numbers.

2.3 Cleaning of Data

Given such a vast dataset, it is critical to ensure that you can retrieve the data required while avoiding the inclusion of extraneous data in the research. After accessing the CSV file, I imported the data into RapidMiner Studio, that automatically delete duplicates for redundancy, then filtering each column for null values or incorrectly named columns, and then updating the date column in a standard format. After cleaning data can be seen such as *Figure 2.*

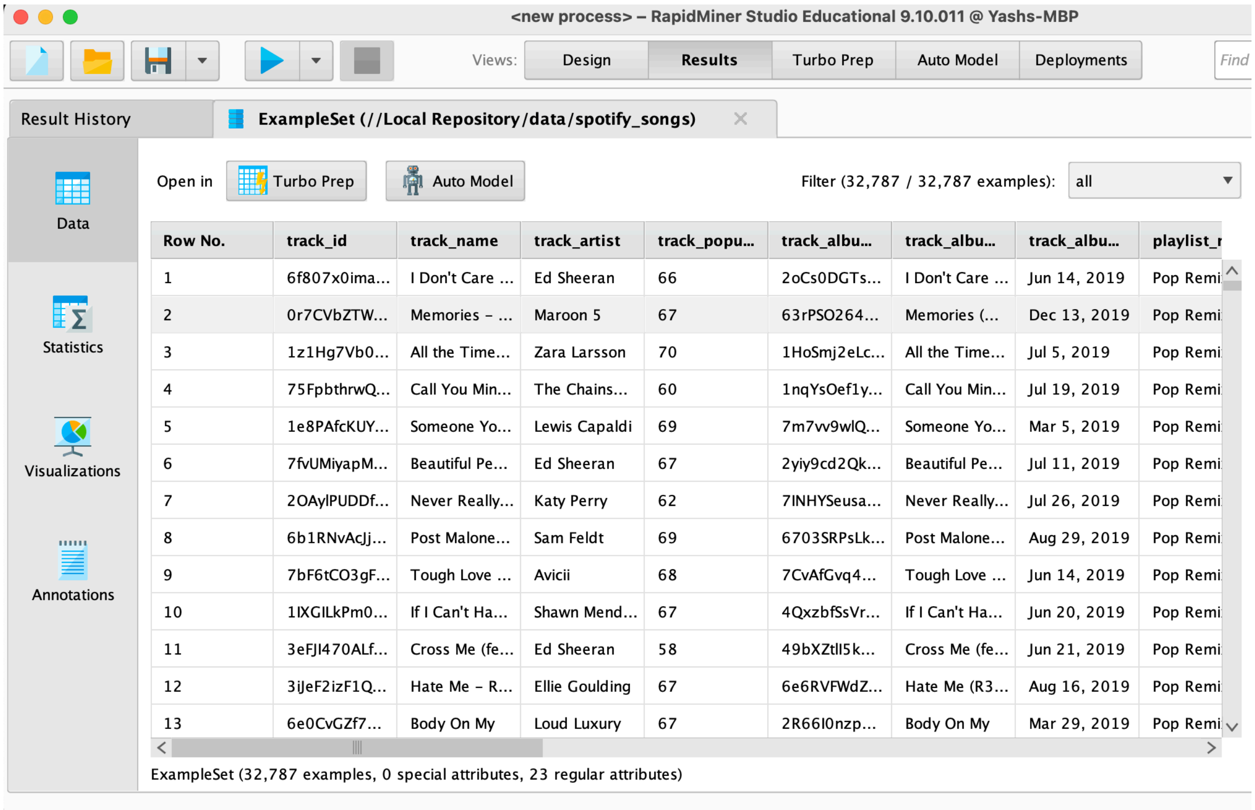

Figure 2 – Data Cleaning in RapidMiner

There were, however, some duplicates that were eliminated. Some ambiguous names in the playlist column were removed, and the remainder of the data was ready for visualization.

2.4 Data Dictionary

Track_id	Song ID
Track_name	Song Title
Track_artist	Song Artist
Track_popularity	Song Popularity (0-100) when higher is preferable
Track_album_id	Album unique ID
Track_album_name	Song album name
Track_album_release_date	Album release date
Playlist_name	Playlist name
Playlist_id	Playlist ID
Playlist_genre	Playlist genre

playlist_subgenre	Playlist subgenre
danceability	Danceability measures how good a tune is for dancing based on a mix of musical qualities like pace; overall regularity, rhythm consistency, and beat strength. Several 0.0 is the least danceable and a value of 1.0 is the most danceable.
Energy	Energy is a perceptual measure of intensity and activity that ranges from 0.0 to 1.0. Death metal, for example, has a high energy level, but a Bach prelude has a low energy level. This trait is influenced by nature of the evidence like as dynamic range, perceived loudness, timbre, onset rate, and general entropy.
Key	The track's projected overall key. Pitch Class notation is used to relate integers to pitches. E.g., 0 Equals C, 1 = C♯/D♭, 2 = D, and so on. The value is -1 if no key was found.
loudness	The total volume of a track measured in decibels (dB). Loudness ratings are averaged throughout the whole track and may be used to compare the relative loudness of different tracks. The major psychological correlate of physical strength is loudness, which is the quality of a sound (amplitude). Typical values vary from -60 to 0 dB.
Mode	A track's mode (major or minor) denotes the sort of scale from which its melodic material is formed. The major is represented by 1 and the minor by 0.
Speechiness	The presence of spoken words in a track is detected by speckiness. The closer to 1.0 the property value, the more solely speech-like the recording (e.g., talk show, audio book, poetry). Tracks with values greater than 0.66 are entirely composed of spoken words. Values between 0.33 and 0.66 reflect songs that may contain both music and voice, either in parts or layered, including such situations as rap music. Music and other non-speech-like recordings are represented by values less than 0.33.
Icousticness	A confidence level ranging from 0.0 to 1.0 indicating if the track is acoustic. 1.0 indicates a high level of certainty that the track is acoustic.
Instrumentalness	Determines whether or not a music has no vocals. In this context, "ooh" and "aah" noises are viewed as instrumental. Rap and spoken word are definitely "vocal" songs. The closer the instrumentalness score is near 1.0, the more likely the track does not contain any vocal content. Values greater than 0.5 are meant to signify instrumental recordings, although confidence grows as the value approaches 1.0.

liveness	The existence of an audience in the tape is detected. Higher liveness numbers indicate a greater likelihood that the track was played live. A number exceeding 0.8 implies a substantial possibility that the music is active.
valence	A scale from 0.0 to 1.0 that describes the melodic positivity given by a tune. Tracks with a high valence sound more positive (for example, pleasant, cheery, and euphoric), whereas tracks with a low valence sound more negative (e.g., sad, depressed, angry).
tempo	The estimated overall pace of a track in beats per minute (BPM). Tempo is the speed or pace of a composition in musical terms, and it is derived directly from the average beat length.
Duration_ms	Song length in milliseconds

Table 1 – Data Dictionary

3. Exploratory Data Analysis

3.1 Track Popularity v/s Track Subgenre

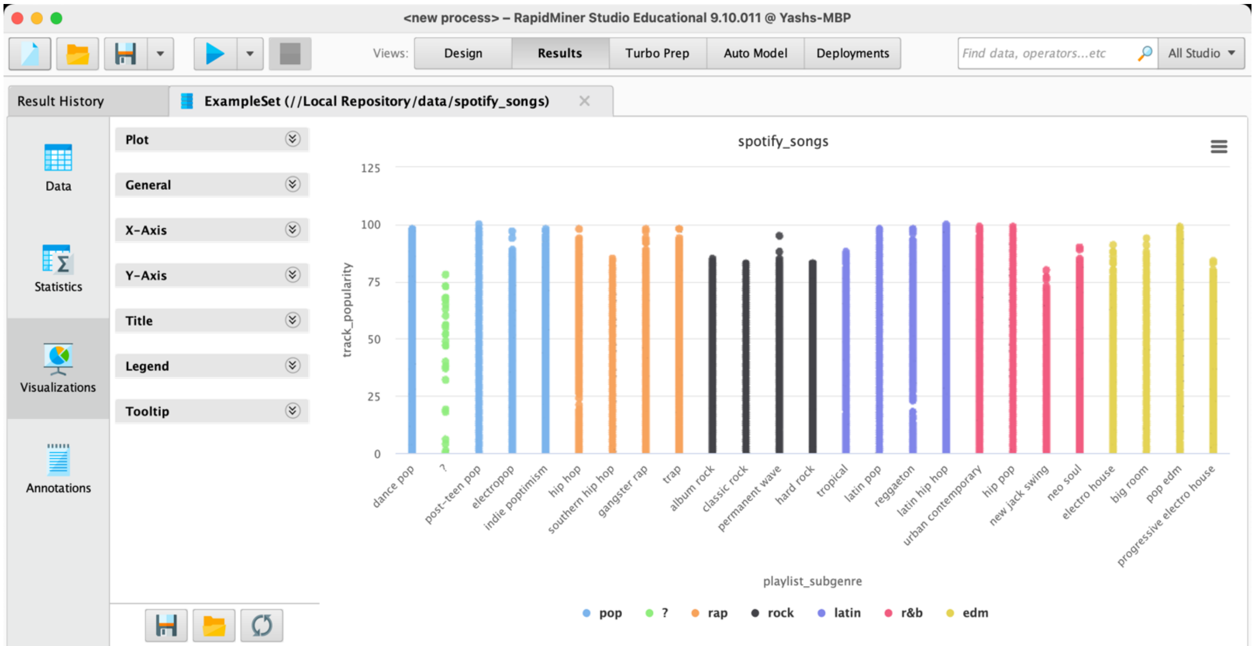

Figure 3 – Track popularity by year v/s subgenre

I began evaluating some of the columns, and it occurred to me that I had come across the various subgenres in the dataset. All the many subgenres considered comparing it to track popularity to see what kinds of tracks became successful based on their genre also put track album release dates by year so that we can see how the popularity of the music affected the year by subgenre.

As a result, Hip Hop was not widely popular in the 1990s. It peaked in the 2000s and has since risen to become the most popular genre of all.

Only urban contemporary has declined since 2011, because to the changing diversity of music. - While Latin pop existed in the 1990s and is still on the rise, dance pop and pop edm score approximately 40-45 on a scale of 0-100, with higher being better.

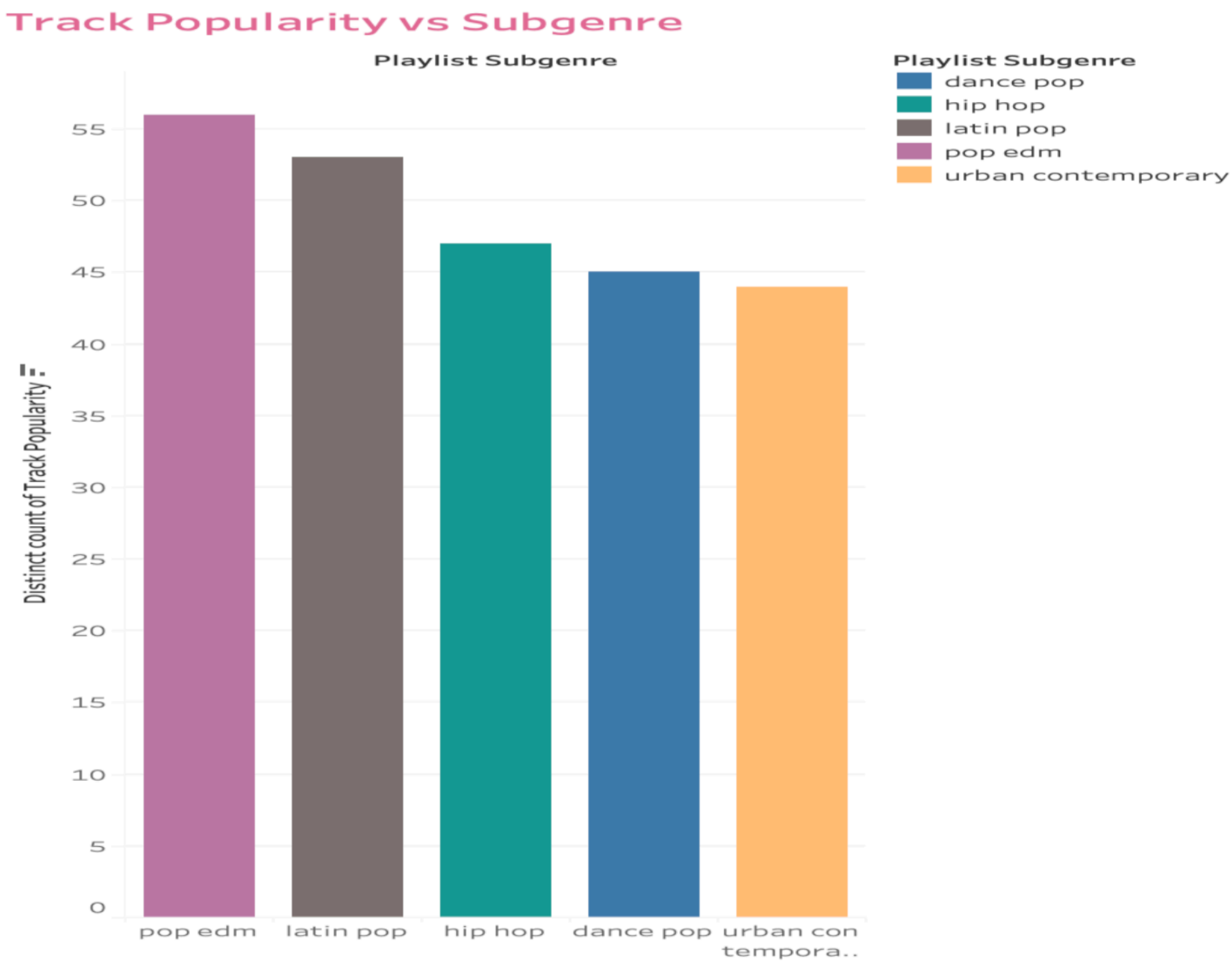

Figure 4 – Track popularity

The graph generated in Tableau above depicts the fundamental relationship between track popularity and subgenre. This was chosen because it has a distinct connotation than Figure 3.

The various hues represent the various subgenres. I narrowed it down to five genres: pop edm, latin pop, hip hop, dance pop, and urban contemporary.

In Figure 3, we observed that hip hop was the most popular genre, but here, pop edm with a score of 56 is the most popular of all. Why is this the case? Figure 4 summed together all the songs since there was no dissection by year, therefore pop edm came out on top, but when we look at the increase of genre by year, we see a different picture.

3.2 Audio Feature with Top Artist

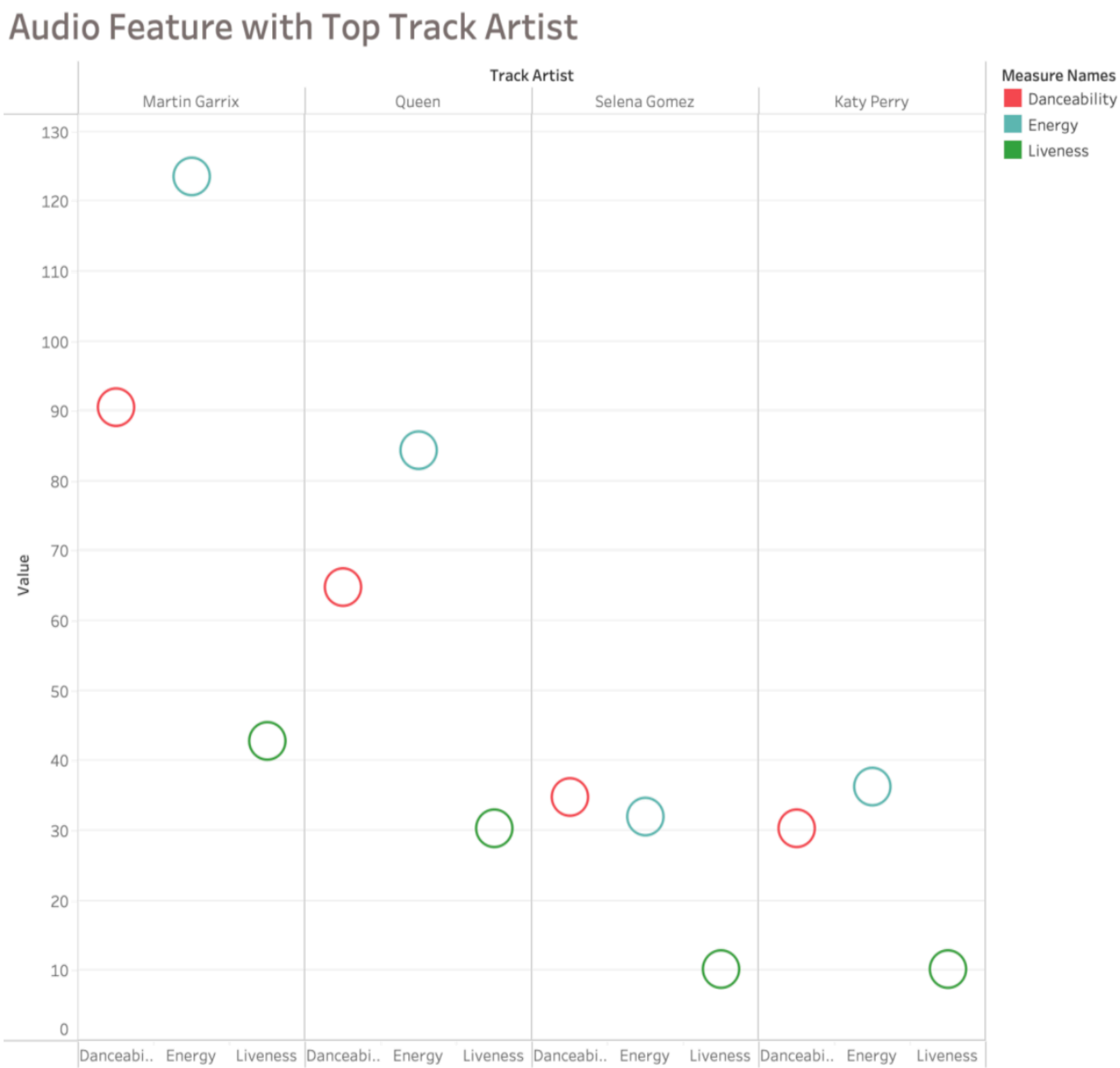

Figure 5 – Top grossing Artist

The image above highlights top artist audio features. I have narrowed down the top artists to four, which can be seen above. Also filtered the audio features to those that make music memorable, such as danceability, vitality, and liveliness.

Martin Garrix has a high quantity of energy audio feature with more dance music, making him one of the top performers, whilst Selena Gomez and Katy Perry have a reasonable amount of all three elements in their songs

3.3 Audio Feature Change by Year

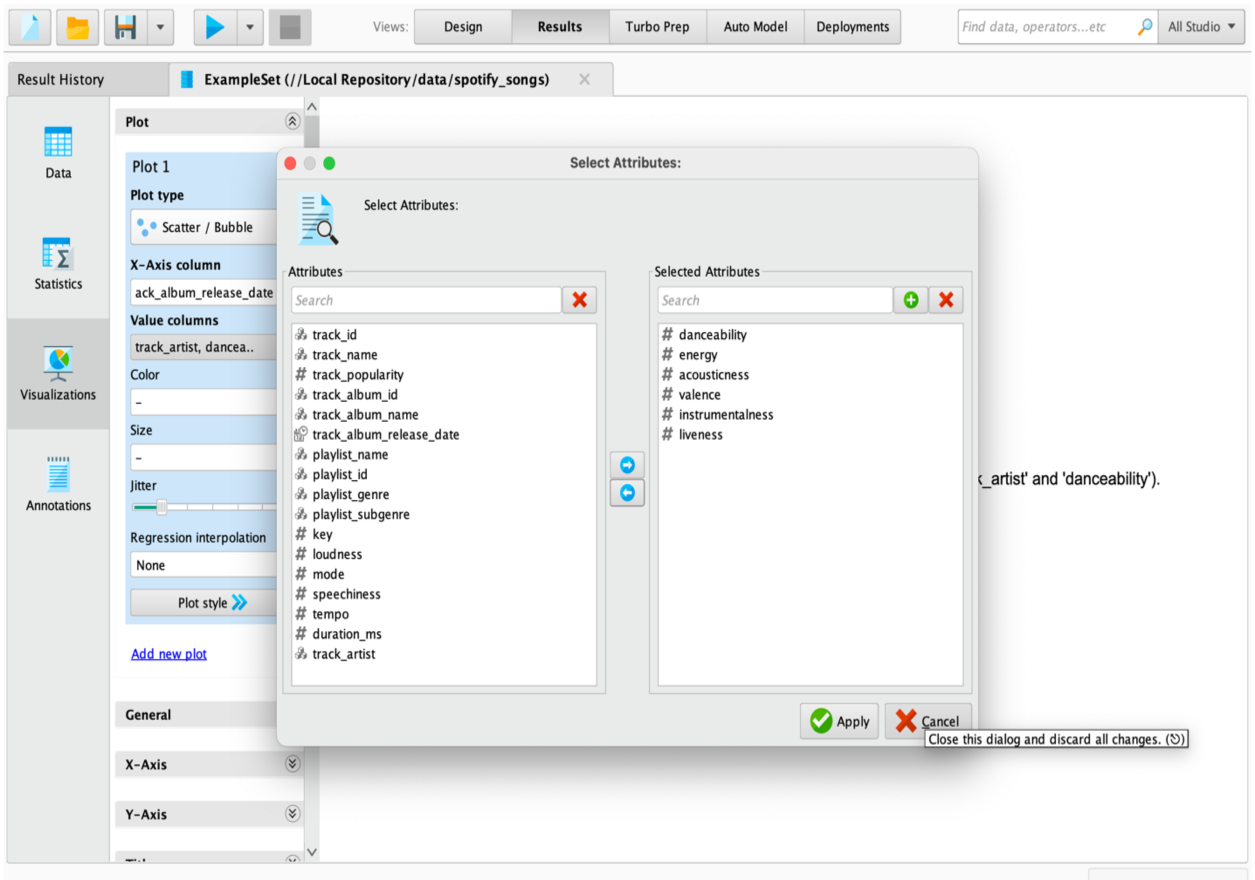

Figure 6 – Yearly Audio Feature attribute selection

As I was going over the data in RapidMiner Studios, it occurred to me that I should look at how the audio function has evolved over time. So here it is, I have not taken all the qualities and have filtered them down to the most important ones, from which we can see that every feature has surged steadily over the years.

It is amazing to watch energy and danceability taking a much bigger stride than others, that is reflected in below figure. It has all genre trends from 1960 to 2020. This analysis graph has been generated in RapidMiner.

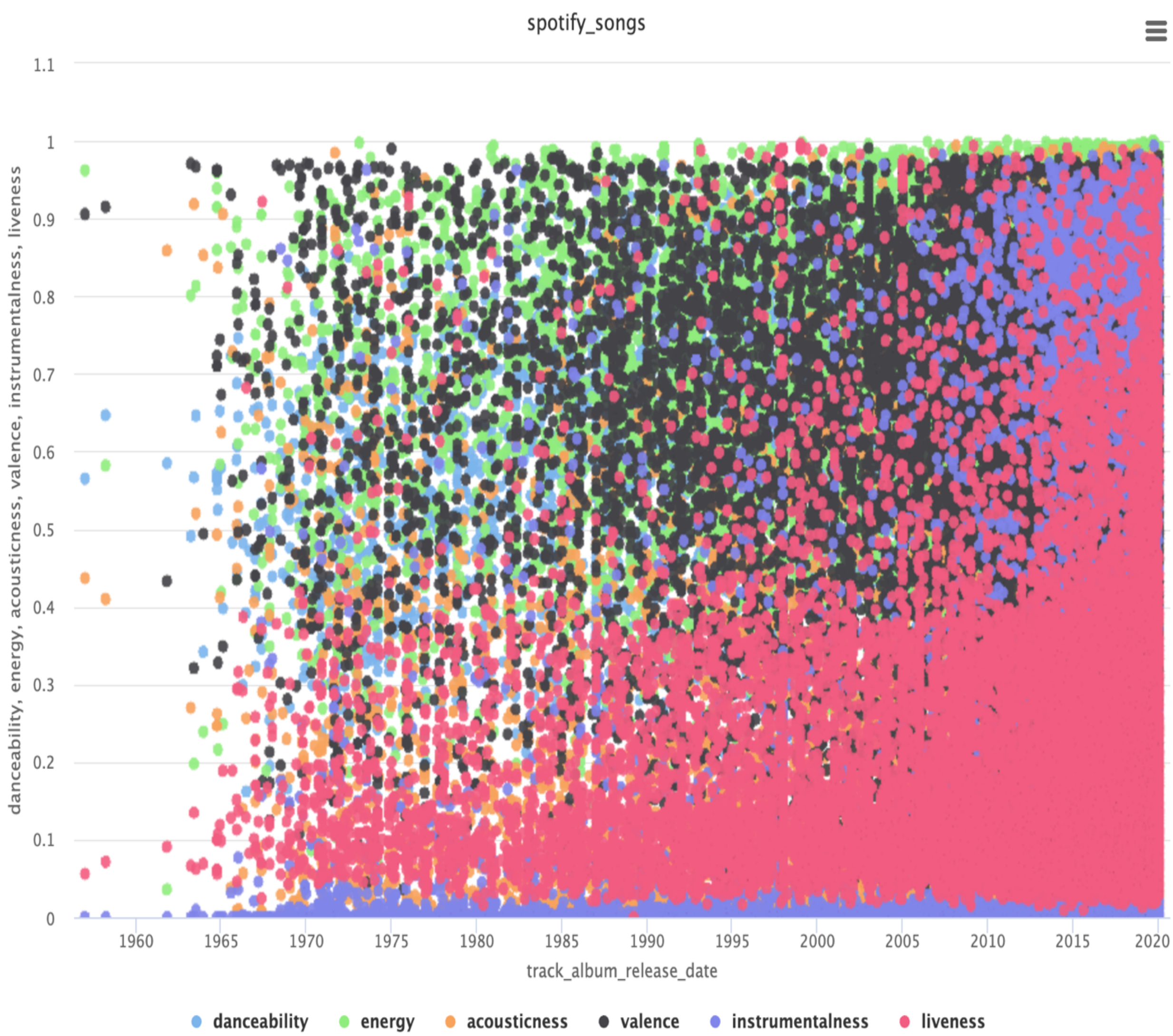

Figure 7 – Yearly Audio Feature

This indicates the taste in music has evolved throughout the years and thus the musician does have to come along with it, which will be further explored. More important thing to not down is as below:

- People tend to listen more liveness genre
- Valence has also increased over the years
- Energy and danceability has been consistent over the years but with a slight rise

To sum up for this graph that now, listeners have more interest in danceability and liveness of songs on Spotify.

3.4 Top Artist Popularity by Year

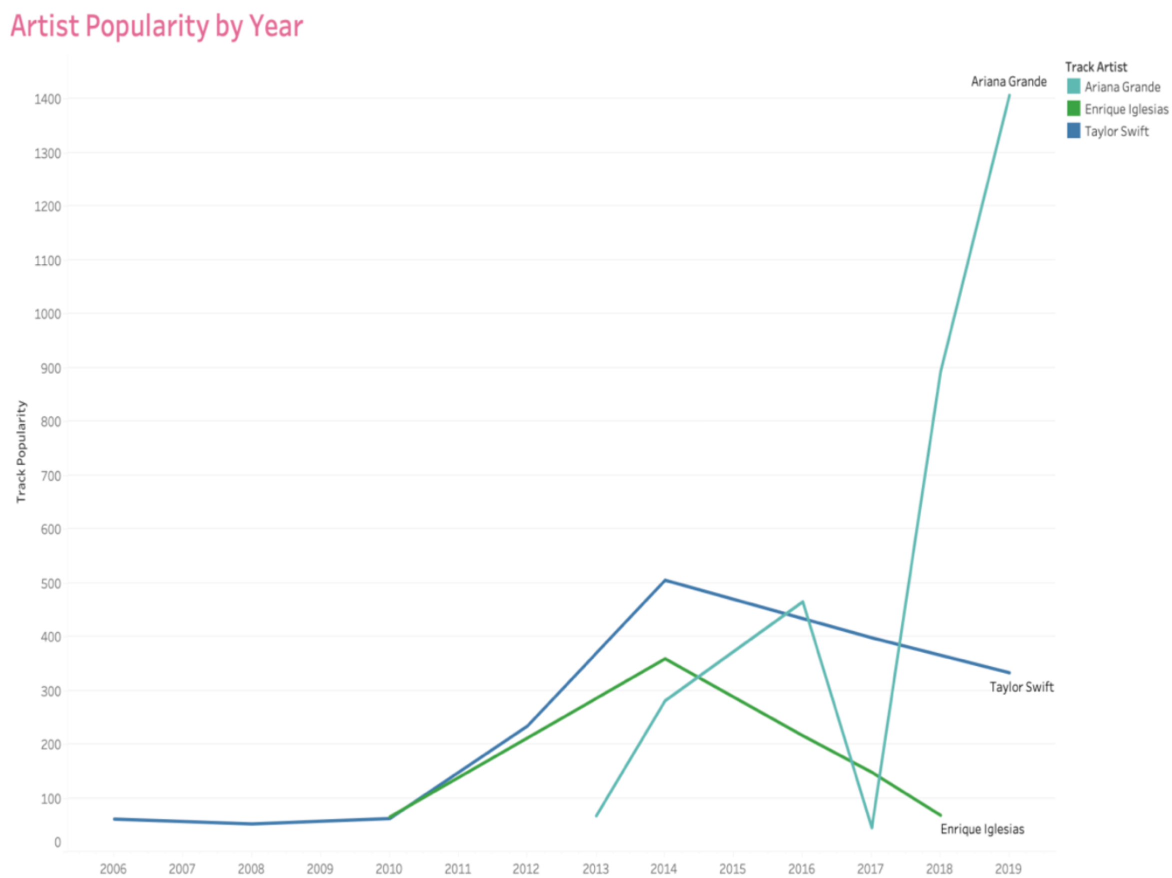

Figure 8 – Trend by Users

The graph above illustrates the top three I picked based on their popularity throughout time. Where we can see one of them has already already stepped down by 2018, with another one on the road but in later 2019. - Enrique and Taylor are well-known singers from the 2000s who rose to prominence by 2014 and whose songs were extremely popular among young people, whereas in 2013 we saw a surge of Ariana, which fell in 2017 but is now on the rise again because her music has more tempo and energy, which makes it more grooving.

3.5 Top Insights

The research performed using the data provided allowed us to decipher why and how tracks with their audio features, album name, and genre make it to the top 100 or billboard songs of the year. Do you want to write a hit song? Make it a high-energy, electronic dance tune to increase your chances! Keep things simple by sticking to a 4/4 time scheme.

The research includes all the other audio metrics on the dashboard, including Energy, Danceability, Popularity, Speechiness, and Tempo. That was the most valuable aspect of this assignment for me: understanding how to establish a Tableau parameter that would allow me to toggle between multiple variables inside the same charts. The technologies I used to carry out this analysis were RapidMiner Studio, Tableau and Excel.

4. Summary

4.1 Problem Statement

The analysis focuses on providing crucial trends and insights to music artists and music distributors that work in the arena of digital streaming services. The overall trends in track popularity were observed for several music genres and album release times. Furthermore, the popularity class of a track was predicted, as well as the primary factors of popularity.

4.2 Insights

When I first addressed the subject, I had no idea what to anticipate in terms of statistics. As a music/dance aficionado, I had an inkling of what the trends may be, but the level to which they leaned towards the subgenre side of playlists astonished me. The study's key takeaway for me was that, since the 2000s, music has changed to reflect what the generation wants, and it is altering year by year.

4.3 Implications

While the 'pop' genre contains the most tracks on Spotify, it also has the highest average popularity score among users. Because 'edm' and 'rap' have more tracks on Spotify, emerging artists in these genres face stronger competition in terms of track visibility. The popularity of a music diminishes somewhat as the song duration and instrumentality increase. Tracks with greater valence (positivity) ratings have higher danceability indexes. Rock music from earlier eras is more popular among users than recently released tunes. Tunes with titles such as 'dance,' 'hits,' 'hip,' and 'hop' are more common in extremely popular tracks than in less popular ones. "song duration," "instrumentalness," and "loudness" are important factors in determining a track's popularity on Spotify.

4.4 Limitations

The analysis is performed on a static data set. This might be improved further by having data scraped dynamically using a web API.

If accessible, user reviews might be utilized to further understand the essential components that improve a user's affinity for a music/podcast track.