Beats & Bytes: Decoding The Secret Sauce of Chartbuster

Rachel Burgess, Sammy Cayo,
Bryce Loomis, Laura Lubben

Berkeley
UNIVERSITY OF CALIFORNIA

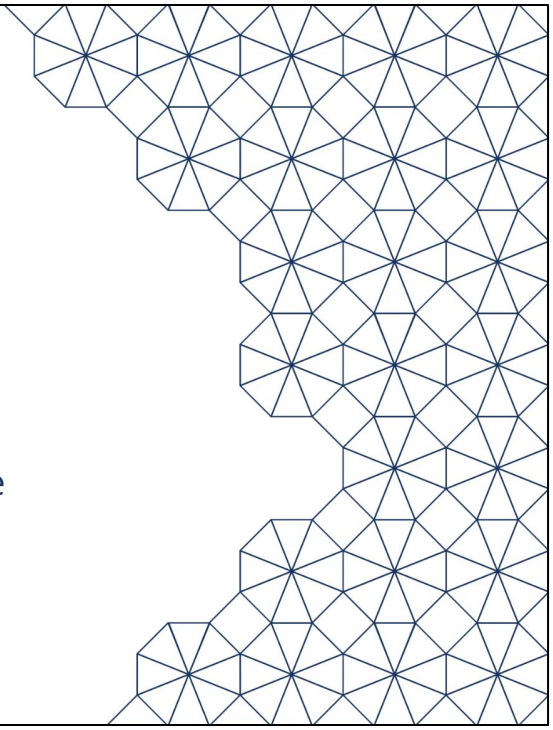
Research Questions

- **Musical Attributes:** How do musical characteristics influence a track's popularity?
- **Genre and Trends:** How does popularity vary across different genres and subgenres? Are there observable trends over time?

Berkeley
UNIVERSITY OF CALIFORNIA

Data Overview

- 32,000+ songs (tracks)
- 23 Columns
- Genres
- Musical Characteristics
- Longest Playlists by (sub-)Genre
- Shortest Playlists by (sub-)Genre
- Single Track Examination



11 Track Descriptions

Track ID	Track Name
Track Artist	Track Popularity: Score (0 - 100)
Track Album ID	Track Album Name
Track Album Release Date	Playlist Name
Playlist ID	Playlist Genre & Playlist Subgenre

Data Overview

Dance Monkey Tones and I

12 Musical Characteristics

Danceability	Acoustiness
Energy	Instrumentalness
Key	Tempo
Loudness	Liveness
Mode	Valence
Speechiness	Duration_ms

Data Overview

Type of Genres

6 Genres

```
array(['pop', 'rap', 'rock', 'latin', 'r&b', 'edm'], dtype=object)
```

24 Subgenres

```
array(['dance pop', 'post-teen pop', 'electropop', 'indie poptimism',
       'hip hop', 'southern hip hop', 'gangster rap', 'trap',
       'album rock', 'classic rock', 'permanent wave', 'hard rock',
       'tropical', 'latin pop', 'reggaeton', 'latin hip hop',
       'urban contemporary', 'hip pop', 'new jack swing', 'neo soul',
       'electro house', 'big room', 'pop edm',
       'progressive electro house'], dtype=object)
```

Longest & Shortest Playlists Across Genres

Longest

- Indie Poptimism

```
playlist_name
Indie Poptimism                                                                  112.516467
Permanent Wave                                                                    98.979413
Fitness Workout Electro | House | Dance | Progressive House                       86.133680
2020 Hits & 2019  Hits – Top Global Tracks 🔥🔥🔥                                    83.392242
Hard Rock Workout                                                                 81.842715
Ultimate Indie Presents... Best Indie Tracks of the 2010s                         81.787073
Southern Hip Hop                                                                  80.767408
Urban Contemporary                                                                69.588855
Classic Rock 70s 80s 90s, Rock Classics – 70s Rock, 80s Rock, 90s Rock Rock  Classicos  68.737590
Underground Party | Hypnotic | Minimal | Acid | Big Room | Tech | Liquid          63.857932
Name: duration_minutes, dtype: float64
```

Shortest

- Post-Teen Pop

```
playlist_name
Post-Teen Pop                                  0.283445
ALPAS Music Festival                           0.929517
TOP 50 GLOBAL 2020 UPDATED WEEKLY 🌍🎶 WORLDWIDE   2.073210
Big White Room-Jessie-J                        2.833353
URBAN NATION                                   2.902115
RADIO POP CHARTS                               3.035135
CSR 103:9 (GTA: SA)                            3.054287
ELETRIC POP & DANCE                            3.054405
Intro to Permanent Wave                        3.439643
EDM Pop                                        3.579055
Name: duration_minutes, dtype: float64
```

Data Overview

449 playlists 308 songs to 1 song

Top Track by Characteristic

Max tempo:
Song - Dope's Gotta Hold On Me (feat. Ese Rich Roc),
Artist - Spanish F.L.Y.,
Subgenre - latin hip hop

Max valence:
Song - Low Rider,
Artist - War,
Subgenre - classic rock

Max danceability:
Song - If Only I Could (feat. Steve Lucas) - Liem Remix,
Artist - Fusion Groove Orchestra,
Subgenre - progressive electro house

Max energy:
Song - Rain Forest and Tropical Beach Sound,
Artist - Nature Sounds Nature Music,
Subgenre - tropical

Max acousticness:
Song - I'm Going to Live My Life,
Artist - DJ Screw & The Screwed Up Click,
Subgenre - southern hip hop

Max instrumentalness:
Song - Chill Waves & Wind in Leaves,
Artist - Pinetree Way,
Subgenre - tropical

Max liveness:
Song - Knockin' On Heaven's Door - Live In London / 1992,
Artist - Guns N' Roses,
Subgenre - album rock

Max speechiness:
Song - I'M DEAD,
Artist - Duckwrth,
Subgenre - neo soul

Data Overview

Data Cleaning & Manipulation

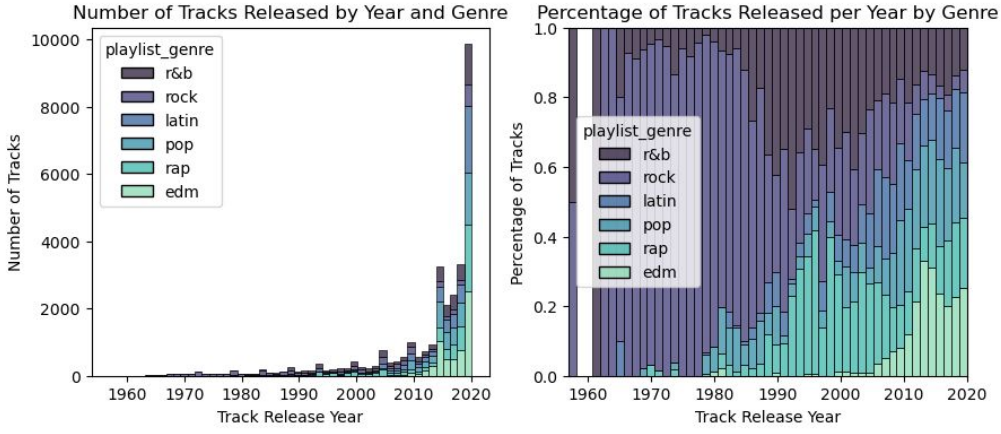
- Removing Nulls and Duplicates
- Converting Release Date to Date Format
- Creating Year and Decade Columns

Trends over Time

- Genre
- Musical Characteristics

After looking through our dataset, we analyzed some of the trends in genre, musical characteristics, and popularity overtime. As might be expected, different genres increase and decrease throughout the decades. While our analysis is complicated by the fact that a majority of the tracks in our dataset were released in the past ten years as you can see in our graphs, there are some definite trends outlined below:

Trends over Time: Genre



On the left you can see the absolute number of tracks per year is heavily skewed towards more recent years, broken down by share of play list. On the right is the relative share of tracks fom that year by genre.

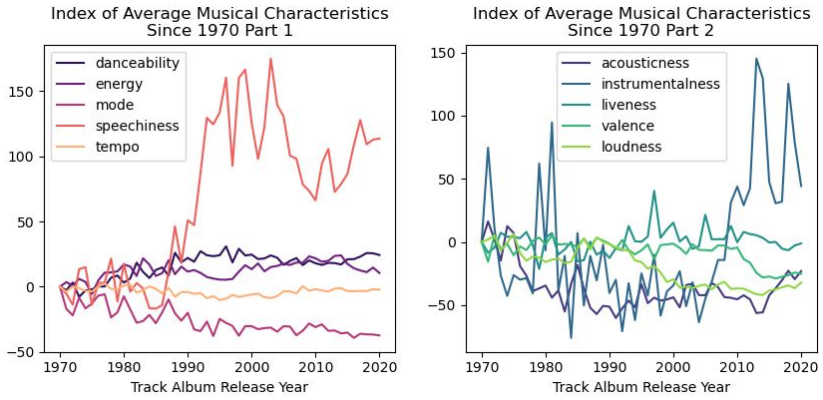
Genre: as you can see Rock and R&B dominate the early decades of the dataset before gradually being supplanted by rap and pop music. And EDM, which was the most popular genre in the early 2010s, hardly existed before 2000. For tracks selected from 2020, the genre composition is extremely diverse, with the largest share of tracks by anyone genre only about 20%

Trends over Time: Genre

track_album_release_decade	1950	1960	1970	1980	1990	2000	2010	2020
playlist_genre								
edm	0	0	4	5	9	150	4596	113
latin	0	1	3	51	146	637	3205	94
pop	0	1	18	114	121	441	4270	165
r&b	1	23	52	186	695	903	2580	61
rap	0	3	9	51	588	741	3856	152
rock	1	99	681	680	522	935	1327	41

Corroborating that graph, we can see the permanency of genres like Rock and R&B in this table, which appear consistently in each decade. Meanwhile, EDM has very few tracks until the 2010s.

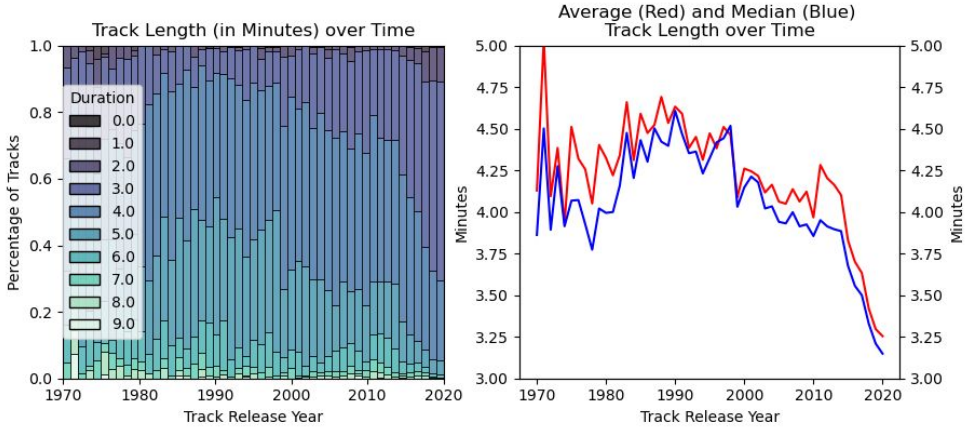
Trends over Time: Musical Characteristics



As well as genre We also tracked the development of musical characteristics over time. Here you can see the average of each year's musical characteristics indexed to 1970. Theyre split into separate charts just to make them easier to read.

The proliferation of rap especially has caused the average track "speechiness" to double since 1970. Tracks have also gotten quieter and sadder, as loudness and mode (a measure of whether the song was written in a major or minor key with) both decreased–potentially as rock songs become less common in the dataset.

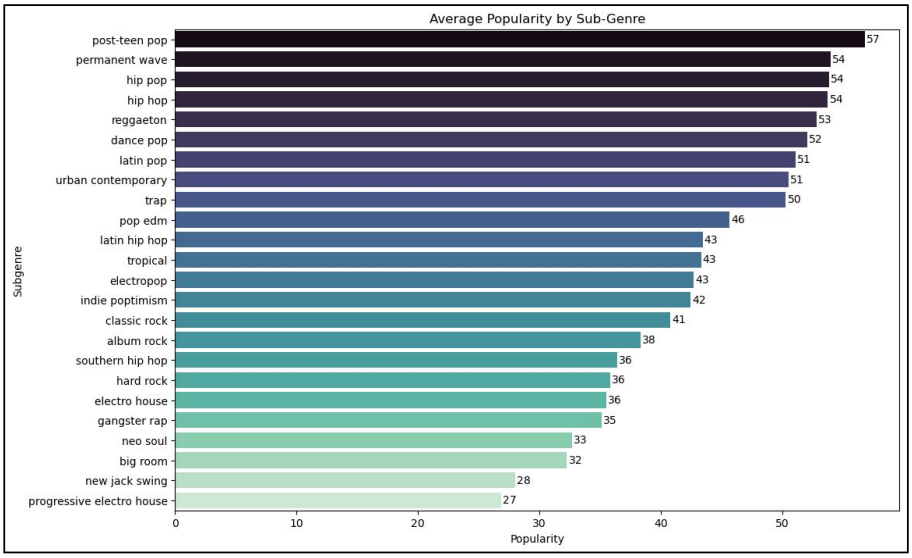
Trends over Time: Musical Characteristics



Songs generally grew in length through the seventies and eighties, reaching a peak near 1990. Since then, track duration has decreased steadily. In the figure on the left below, we can see how tracks of five or more minutes in duration, made up almost half of all tracks in the 1980s and 90s, are less than 10 percent of the share in the past few years. Those longer songs have been increasingly replaced by two- or three-minute tracks. On the right is the average and median track length over time

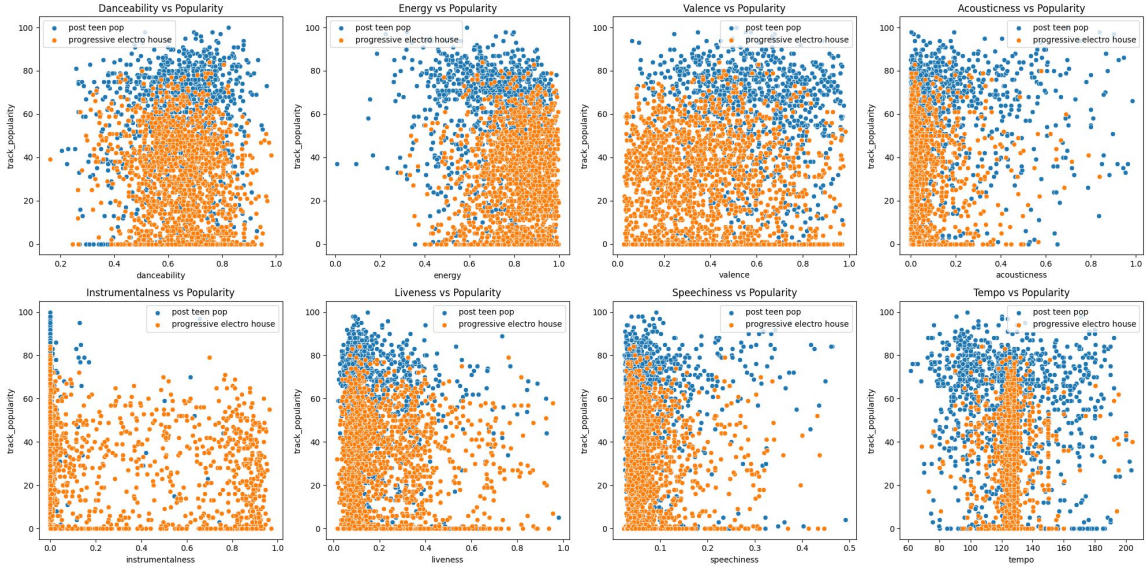
Zoom in on a Couple Genres

Berkeley
UNIVERSITY OF CALIFORNIA



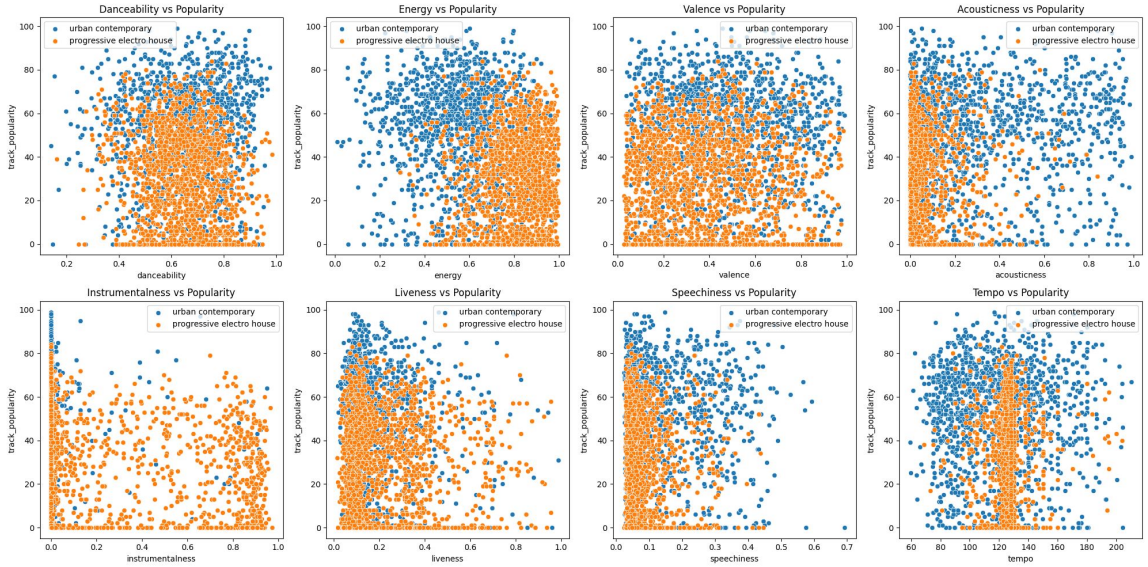
We started by segmenting our dataset by decade. This allowed us to see how musical preferences have evolved over time. By focusing on the top five songs from each era, we've uncovered fascinating patterns in what makes a song popular. Take a look at the graph here. We found that post teen-pop leads in popularity, closely followed by permanent wave. Surprisingly, new jack swing and progressive electric house trail behind. Each bar here represents a subgenre, with its height showing average popularity. We've also annotated each bar with the number of tracks in that subgenre.

Post-teen Pop vs Progressive electro house



Post-teen pop has more variance across attributes than progressive electro house. As an EDM genre, electro progressive house songs clusters very closely to each other on tempo.

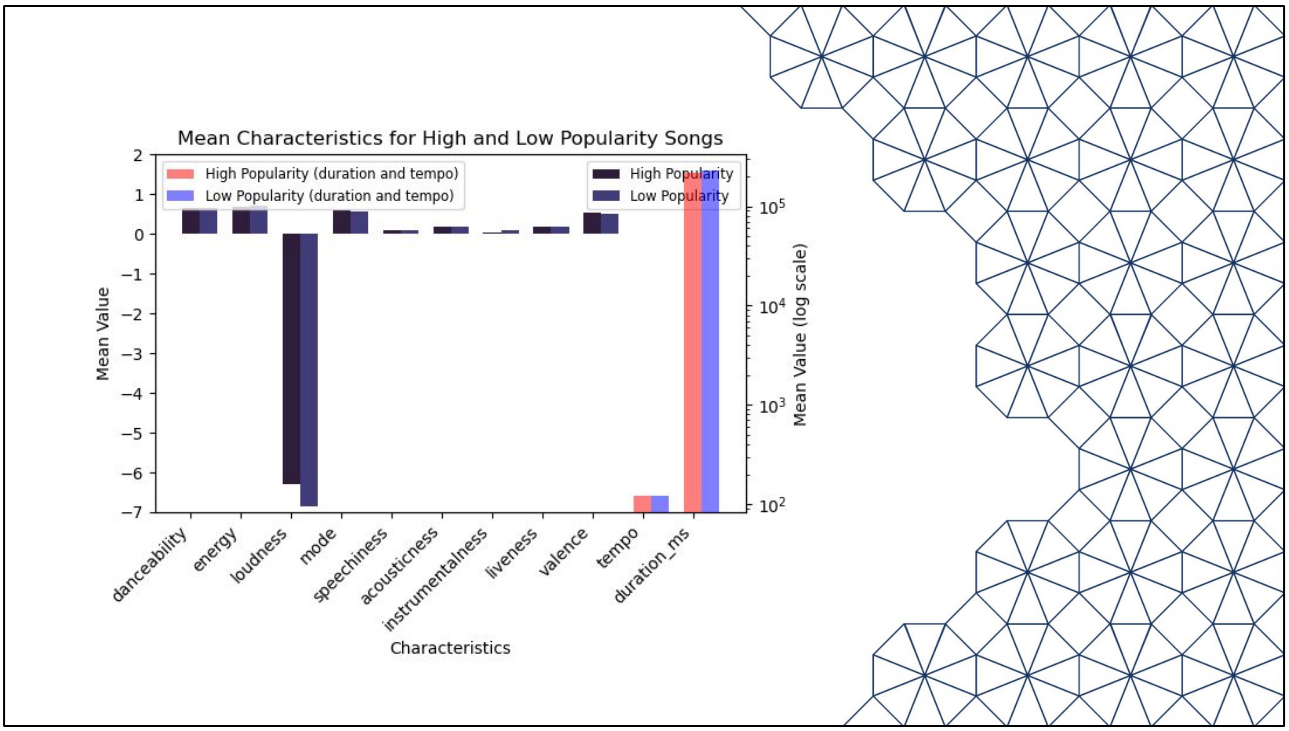
Urban Contemporary vs Progressive electro house



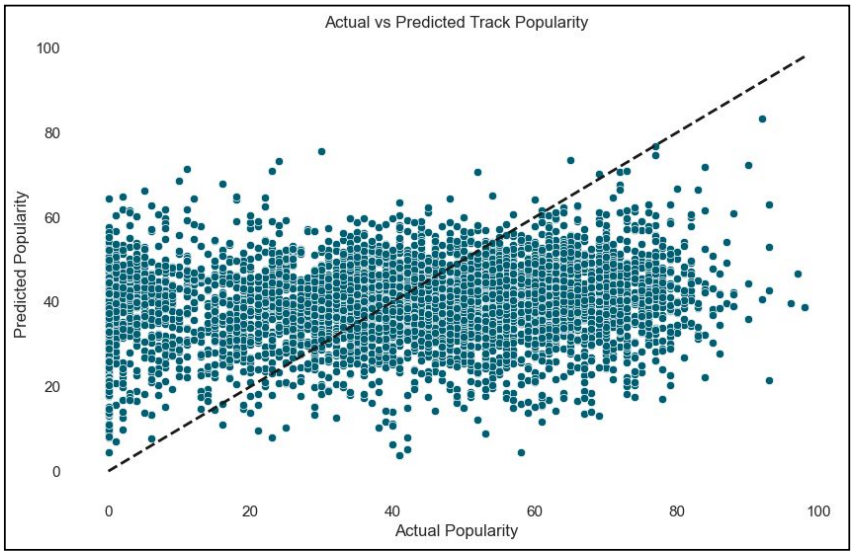
Urban contemporary, a pretty broad genre that can sometimes be a catch all, we see a lot of variance and is even more spread out, which highlights the clusters for where progressive electro house solidifies.

Estimating Track Popularity

Berkeley
UNIVERSITY OF CALIFORNIA



Interestingly, we discovered that the relationship between individual audio features and song popularity is quite weak. This highlights the complex nature of musical preferences and the importance of considering multiple factors. Our initial models, using linear regression, didn't quite hit the mark, showing high error rates and low fit..Next, we compared songs of high and low popularity, setting our high-popularity threshold at the 75th percentile of track popularity. This split our dataset into two: high-popularity songs and their less popular counterparts. We then analyzed various audio characteristics to see how they influence a song's success. We can see how loudness and duration per ms has an effect on songs.



So, we shifted gears and applied a feature selection approach using SelectKBest with f_regression. This helped us identify key features but still didn't perfect our predictions. Finally, we struck gold with the Random Forest Regressor model. It showed a significant improvement over the linear models, with lower error rates and a higher R-squared score. This model highlighted the top features like duration, acousticness, and danceability, providing us with a better understanding of what influences a track's popularity. while the Random Forest model offers valuable insights, there's still room for improvement, especially for tracks with very high popularity. Further refining our model and exploring new features could lead us to even more precise predictions. This research not only informs the music industry's marketing and playlist curation strategies but also gives us a deeper appreciation of how our listening preferences have evolved over time.

Q&A