QMBE 1320 – Final Project: Stage 2

Nov 25, 2024

By: Anisa Mohamed

Spotify Songs

Introduction

Through playlists, customized recommendations, and genre-specific trends influencing the listening experience, Spotify has completely changed how people listen to music. Understanding the factors that influence a songs popularity is useful not only for streaming services, but also for artists and producers looking to create music that resonates with their target audience. To understand these aspects, i aim to uncover 2 questions in this project: 1) How do audio features like tempo, energy, and danceability affect song popularity across genres? 2) Are there seasonal trends in music popularity based on genre? By exploring this, it allows us to identify patterns that reflect changing cultural and musical interests to uncover patterns valuable to artists, producers, and streaming platforms.

To address the questions, I chose to explore the Spotify Songs dataset, which is an extensive dataset that contains **32,833** songs in well-known genres such as pop, EDM, Rap, Rock, Latin, and R&B. I used statistical methods like regression analysis to examine the connections between audio features and popularity, as well as visual tools like bar charts, and line charts to identify trends. I also created some new variables, most notably a season category, to find seasonal trends in popularity throughout the year.

This analysis, which looks at the relationship between audio features and song popularity as well as seasonal trends, can provide valuable information to all stakeholders in the music industry, from artists and producers to streaming platforms and consumers. It will be especially helpful for adjusting musical aspects to audience preferences and optimizing release timing. For instance, understanding how energy levels

affect popularity can help artists align their music with listener trends, and seasonal insights can help artists strategically release music to which season their song will be the most popular and generate the most sales. These findings not only improve our understanding of current patterns, but they may also impact how music is generated and produced in the future.

Data Preparation

Source of Data:

The basis for this analysis is the Spotify Songs dataset. With variables that capture important song characteristics like energy, danceability, tempo, and track popularity. It includes 32,833 tracks from genres such as Pop, EDM, R&B, Rock, Latin, and Rap. This dataset offers a unique angle for examining the relationship between song popularity and certain musical elements like tempo, energy, and danceability. It also enables us to identify trends that demonstrate how listeners' musical and cultural tastes are changing. However, a number of anomalies were found, such as songs having a popularity score of 0 and missing values.

Key variables include:

- **Track_popularity:** A measure of a song's popularity on Spotify.
- **Tempo:** The speed or pace of a song in beats per minute (BPM).
- **Energy:** A measure of intensity and activity in a track, ranging from 0 to 1.
- **Danceability:** A score indicating how suitable a track is for dancing, ranging from 0 to 1.
- **Playlist_genre:** The genre of the song, including pop, EDM, rap, rock, Latin, and R&B.
- **Track_album_release_date:** Date when album released
- **Track_popularity:** Song Popularity (0-100) where higher is better.

About Spotify Songs Data.docx

Cleaning Data:

I first Identified variables relevant to my questions and removed the ones that were not**:**

Question 1: How do audio features affect song popularity?

- **Variables:** danceability, energy, tempo, track_popularity, playlist_genre.

Question 2: Are there seasonal trends?

- **Variables:** track_album_release_date, playlist_genre, track_popularity.

Then, I standardized text-based columns like (playlist_genre) to ensure consistent formatting and bold them, i also expanded the rows so the text can viewed. I Checked for missing values in relevant variable columns. I Identified and removed tracks with exceptionally low popularity (track_popularity = 0,1,2) as they may represent anomalies or unplayed songs.

	Nahaze	2	Carillon (with Achille Lauro)

"Example of anomalies and missing values"

I Examined the dataset for missing values, particularly in columns like track_name, track_artist, and track_album_name. I also removed rows where both track_name and track_artist were missing, as these entries were deemed unimportant for the analysis. New variables were included such as song duration in minutes (converted from milliseconds for easier interpretation) and a seasonal category based on release dates. These changes resulted in a clean and suitable dataset.

Track Name	Track Artist	Track Popularity	Track Album Name	Track Album Release Date	Seasons	Playlist Genre	Danceability	Energy	Valence	Tempo	Duration ms	Duration Minutes
I Don't Care (with Justin Bieber) - Loud Luxury Remix	Ed Sheeran	66	I Don't Care (with Justin Bieber) [Loud Luxury Remix]	6/14/19	spring	pop	0.748	0.916	0.518	122.036	194754	3.2459
Memories - Dillon Francis Remix	Maroon 5	67	Memories - Dillon Francis Remix	12/13/19	winter	pop	0.726	0.815	0.693	99.972	162600	2.71
All the Time - Don Diablo Remix	Zara Larsson	70	All the Time (Don Diablo Remix)	7/5/19	summer	pop	0.675	0.931	0.613	124.008	176616	2.9436
Call You Mine - Keanu Silva Remix	The Chainsmokers	60	Call You Mine - The Remixes	7/19/19	Summer	pop	0.718	0.93	0.277	121.956	169093	2.818216667
Someone You Loved - Future Humans Remix	Lewis Capaldi	69	Someone You Loved (Future Humans Remix)	3/5/19	Winter	pop	0.65	0.833	0.725	123.976	189052	3.150866667
Beautiful People (feat. Khalid) - Jack Wins Remix	Ed Sheeran	67	Beautiful People (feat. Khalid) [Jack Wins Remix]	7/11/19	Summer	pop	0.675	0.919	0.585	124.982	163049	2.717483333
Never Really Over - R3HAB Remix	Katy Perry	62	Never Really Over (R3HAB Remix)	7/26/19	Summer	pop	0.449	0.856	0.152	112.648	187675	3.127916667
Post Malone (feat. RANI) - GATTUSO Remix	Sam Feldt	69	Post Malone (feat. RANI) [GATTÅœSO Remix]	8/29/19	Summer	pop	0.542	0.903	0.367	127.936	207619	3.460316667
Tough Love - Tiesto Remix / Radio Edit	Avicii	68	Tough Love (TiÃ«sto Remix)	6/14/19	spring	pop	0.594	0.935	0.366	127.015	193187	3.219783333
If I Can't Have You - Gryffin Remix	Shawn Mendes	67	If I Can't Have You (Gryffin Remix)	6/20/19	summer	pop	0.642	0.818	0.59	124.957	253040	4.217333333
Cross Me (feat. Chance the Rapper & PnB Rock) - M-22 Remix	Ed Sheeran	58	Cross Me (feat. Chance the Rapper & PnB Rock) [M-22 Remix]	6/21/19	summer	pop	0.679	0.923	0.752	121.984	207894	3.4649
Hate Me - R3HAB Remix	Ellie Goulding	67	Hate Me (R3HAB Remix)	8/16/19	summer	pop	0.437	0.774	0.329	123.125	203733	3.39555
Body On My	Loud Luxury	67	Body On My	3/29/19	spring	pop	0.744	0.726	0.687	121.985	192507	3.20845
SOS - Laidback Luke Tribute Remix / Radio Edit	Avicii	68	SOS (Laidback Luke Tribute Remix)	5/17/19	spring	pop	0.572	0.915	0.678	123.919	164516	2.741933333
Summer Days (feat. Macklemore & Patrick Stump of Fall Out Boy) - Tiesto Remix	Martin Garrix	63	Summer Days (feat. Macklemore & Patrick Stump of Fall Out Boy) [Remixes	7/12/19	summer	pop	0.69	0.78	0.238	126.07	255238	4.253966667
South of the Border (feat. Camila Cabello & Cardi B) - Andy Jarvis Remix	Ed Sheeran	66	South of the Border (feat. Camila Cabello & Cardi B) [Andy Jarvis Remix]	12/11/19	winter	pop	0.805	0.835	0.722	125.028	188230	3.137166667
All My Friends - Eden Prince Remix	AJ Mitchell	60	All My Friends (Eden Prince Remix)	5/3/19	spring	pop	0.694	0.901	0.368	118.051	173548	2.892466667
Say My Name (feat. Bebe Rexha & J Balvin) - Lucas & Steve Remix	David Guetta	65	Say My Name (feat. Bebe Rexha & J Balvin) [Lucas & Steve Remix]	12/28/18	Winter	pop	0.678	0.747	0.516	120.002	189375	3.15625
Dancing With A Stranger (With Normani) - Cheat Codes Remix	Sam Smith	69	Dancing With A Stranger (With Normani) [Cheat Codes Remix]	3/8/19	Winter	pop	0.746	0.557	0.324	111.961	159404	2.656733333
Let It Be Me - Sondr Remix	Steve Aoki	35	Let It Be Me (Remixes)	10/18/19	fall	pop	0.467	0.821	0.232	122.676	185366	3.089433333
Heaven - David Guetta & MORTEN Remix	Avicii	70	Heaven (David Guetta & MORTEN Remix)	8/23/19	fall	pop	0.572	0.934	0.204	125.948	215238	3.5873
All Around The World (La La La) - Marnik Remix	R3HAB	64	All Around The World (La La La) [Marnik Remix]	8/23/19	fall	pop	0.708	0.913	0.707	135.016	170667	2.84445
Don't Leave Me Alone (feat. Anne-Marie) - EDX's Indian Summer Remix	David Guetta	62	Don't Leave Me Alone (feat. Anne-Marie) [EDX's Indian Summer Remix]	5/8/19	spring	pop	0.684	0.818	0.336	123.938	210968	3.516133333

"Snapshot of Cleaned data"

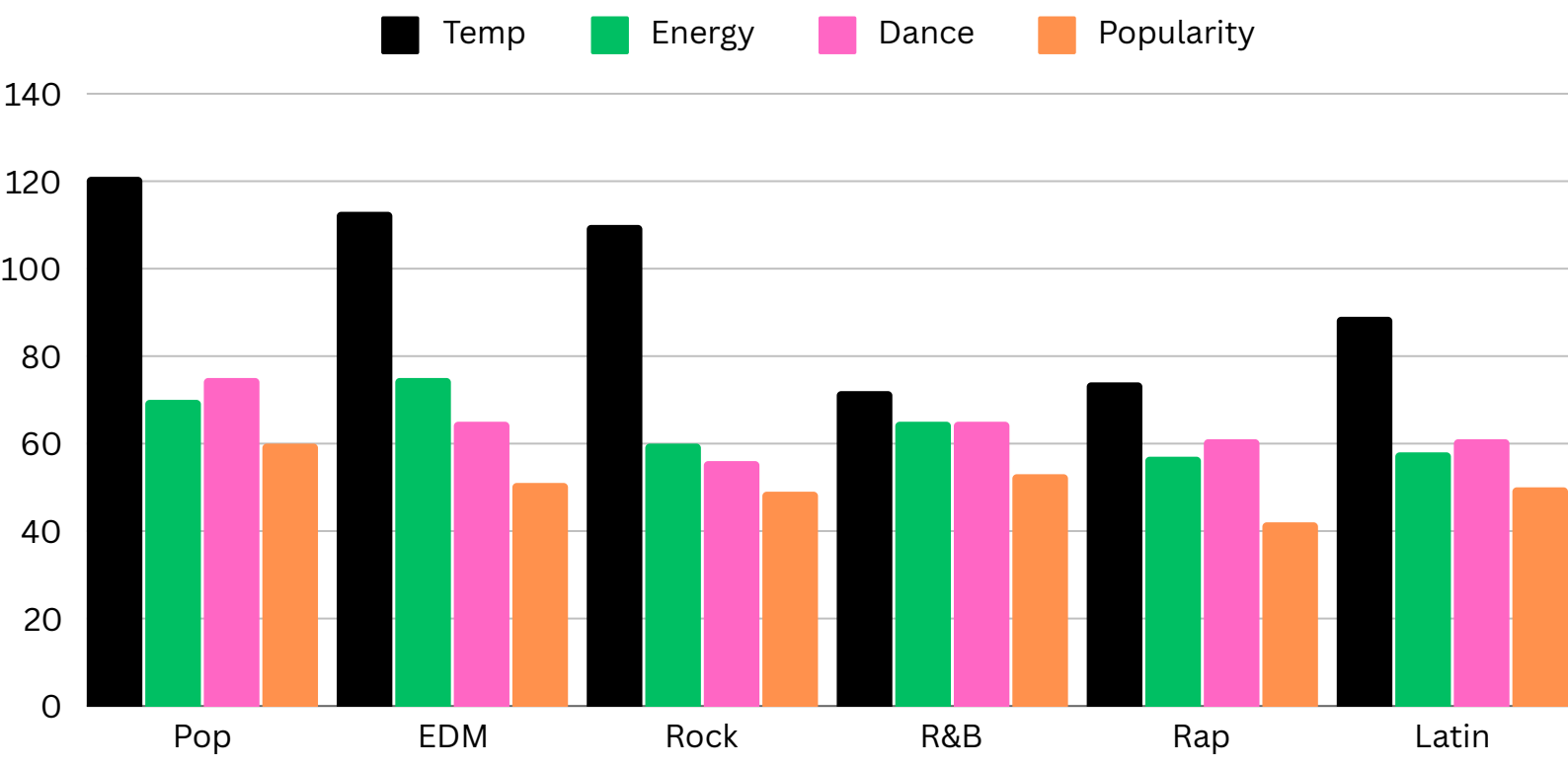
Exploratory Data Analysis:

The analysis provided insights into how audio features and seasonal trends influence song popularity. First, The data suggests that songs with higher energy levels, like pop and EDM, are more likely to be popular. This indicates that a song's success may be significantly affected by its energy.

Genre	Avg Tempo	Avg Energy	Avg Dance.	Avg Popularity
Pop	121	0.70	0.75	60
EDM	113	0.75	0.65	51
Rock	110	0.60	0.56	49
R&B	72	0.55	0.65	53
Rap	74	0.57	0.61	42
Latin	89	0.58	0.61	50

"Chart illustrating the average of audio features and popularity"

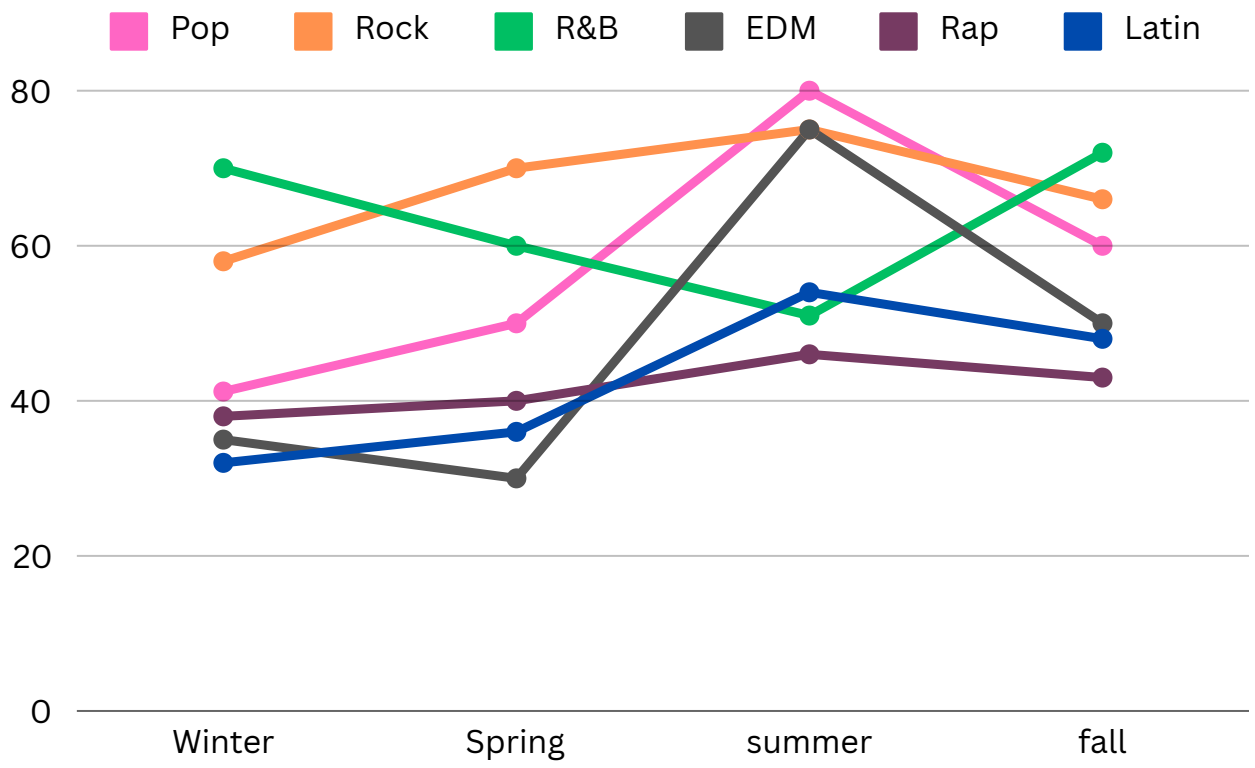
Higher energy levels were associated with greater popularity, suggesting that audiences are drawn to lively and upbeat tracks. Danceability showed similar patterns as well. Perhaps because they are more played at social gatherings and dancing environments, songs with a higher danceability have a tendency to be well-liked.



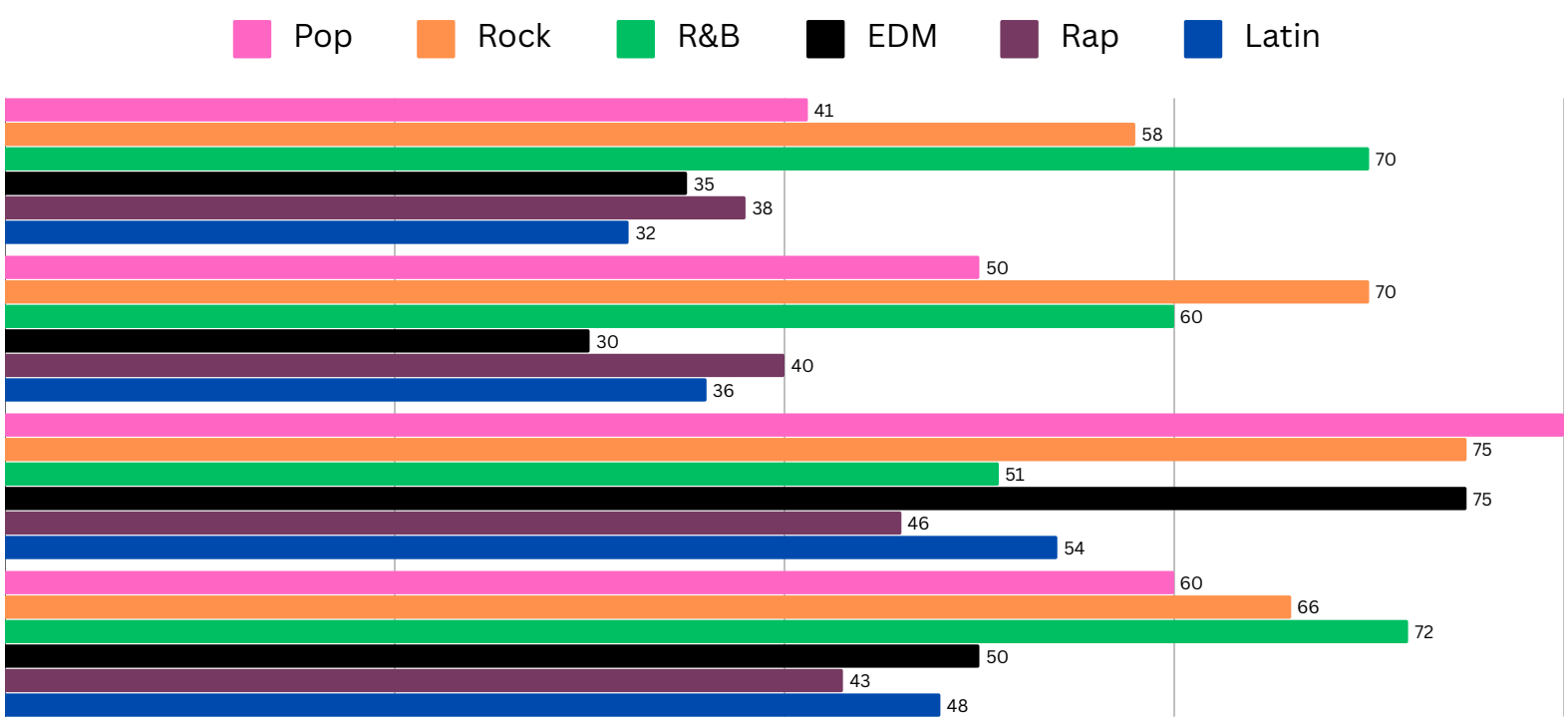
“Bar chart showing the relationship between audio features and popularity by genre”

Popularity and tempo have a weaker correlation, suggesting that other factors may be more important in a song's success.

Seasonal trends emerged as another significant factor influencing music popularity. By categorizing song based on their release dates into seasons, it revealed distinct patterns.



"Line chart illustrating seasonal popularity trends across genres."



"Row chart illustrating seasonal popularity trends across genres."

Summer:

Pop and EDM music are likely to reach its peak in the summer since their lively and upbeat aspects appeal to audiences during the warmer months. The music is played at festivals, vacations, and outdoor events. These genres are popular in the summer because of their ability to improve moods and liven up social gatherings.

Winter:
In contrast, R&B is more likely to reach its highest point in the winter. R&B's soulful and soothing sound goes well with the holiday season's quiet and somber vibe, and it frequently becomes a staple on playlists during this season.

Year-Round Genres:
Rock, Latin, and Rap are among the genres that show more consistent yearly trends without noticeable seasonal increases. Both of these genres are played all year round and appeal to a broad audience. They are also unrelated to any particular season. Due to its lively sounds, Latin music might see some rise during the summer months, but it doesn't peak like Pop and EDM.

Summary

This project analyzes the Spotify Songs dataset to examine the relationship between audio features and song popularity in addition to identifying seasonal trends across genres. The analysis focuses on two main

questions: (1) How do features like tempo, energy, and danceability affect a song's popularity? and (2) Are there seasonal patterns in popularity by genre?

Key Findings

Based on the findings, songs that are are high in energy— such as pop and EDM song—tend to be more popular, which is indicative of listeners' preference for lively and engaging music. The importance of context in music consumption is further shown by seasonal trends: pop and EDM are more popular in the summer, when outdoor gatherings, trips, and festivals are more common, while R&B's soulful tones are more prevalent during the winter holidays. While Latin music has a minor uptick in popularity over the summer, other genres, like Rock and Rap, remain comparatively prevalent throughout the year.

Insights

Using this information, artists and producers can arrange releases and marketing efforts around seasonal trends. For example, releasing energetic songs during the summer might boost the number of streams. Also, streaming services can boost customer satisfaction by refining recommendation algorithms to prioritize tracks appropriate for the current season.

Limitations

Limitations of this study include the dataset's lack of emerging artists, niche genres, regional preferences, or outside influences like advertising campaigns and promotions. For a more comprehensive analysis, listener demographics and regional patterns could be included in future research.