Universitat Pompeu Fabra Barcelona

Analysis of Musical and Lyrical Trends

Ozan Güner
Miguel Habana

Intelligent Interactive Systems (MIIS)

Agenda

1. Introduction
2. Methodology
   a. Data Collection
   b. Data Cleaning
3. Analysis
   a. Sentiment Analysis
   b. Topic Modeling
   c. Prompt Engineering
4. Conclusion
5. Future Work

Introduction

Domination of Music Listening Habits Over Time

Since 1990, the way people interact with music has evolved significantly. This period marks a transition from the tangible, physical media of CDs and vinyl to the intangible, yet infinitely accessible world of digital music.

- **How digitalization and streaming have influenced listeners?**
- **How the listening habits have changed over time?**

We'll examine how our music listening habits have been affected from duration of the songs to the way sentiment of the lyrics.

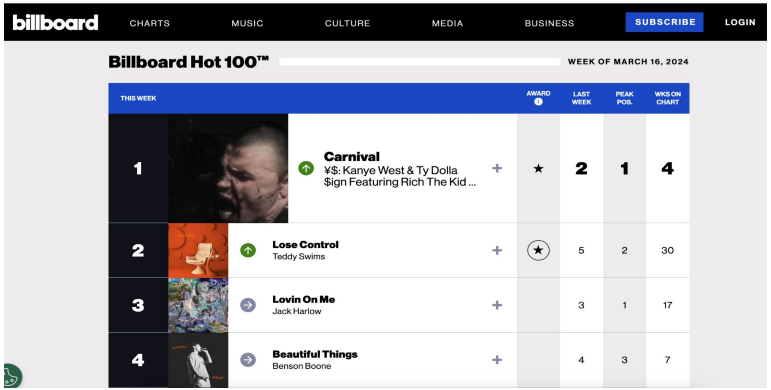
Methodology

Methodology (Data Collection)

Music Samples

Used BeautifulSoup to scrape and collect the most popular songs from every week from January 1990 to February 2024.

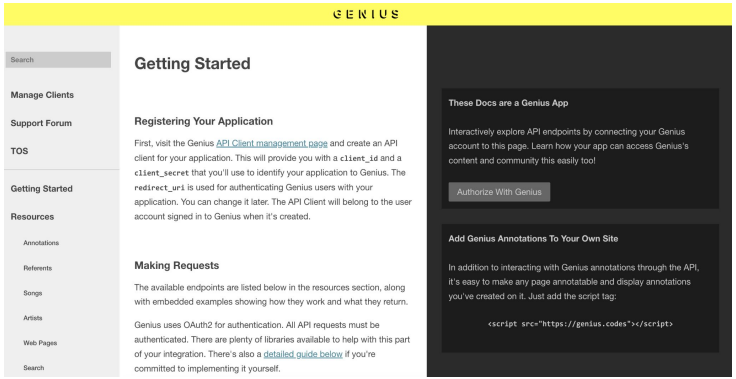
Collected 14000+ unique songs and artists from Billboard Hot 100 Charts.



Web Source: Billboard Hot 100™

Methodology (Data Collection) - 2

Collecting Lyrics



API Source: Genius API

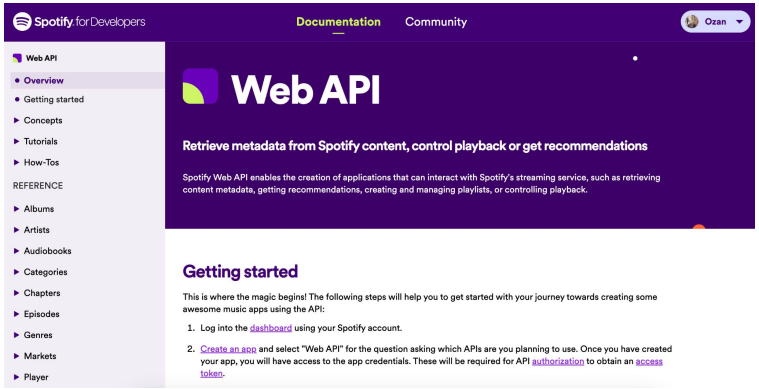
Used LyricsGenius, a Python wrapper of the Genius.com API to collect the lyrics from the Billboard Hot 100 songs.

Collected 12000+ unique song lyrics.

Methodology (Data Collection) - 3

Audio & Track Features

Used Spotipy, a Python wrapper for the Spotify API to collect additional audio and track information from Billboard Hot 100 songs



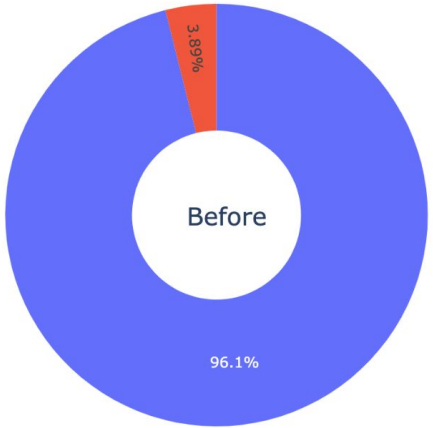
API Source: Spotify API

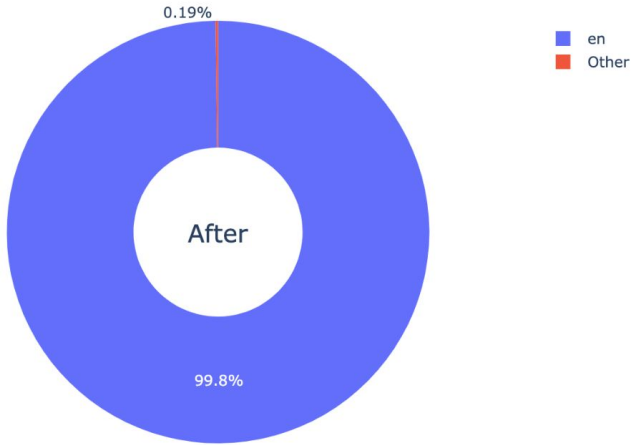
Methodology (Data Cleaning)

Issues with Genius API Lyrics - Translation

Used langdetect for language detection

Used deep translator for translation





Methodology (Data Cleaning)

Issues with Genius API Lyrics - Unrelated Lyrics

Artist Name: Farruko, Nicki Minaj, Bad Bunny, 21 Savage & Rvssian

Song Name: Krippy Kush

Lyrics: 4 ContributorsSpanglish Songs - List of songs in English and Spanish / Lista de canciones en español e ingles Lyrics• "#Freetag" by Kidd Keo • "1, 2, 3" by Sofia Reyes ft. Jason Derulo & De La Ghetto • "1-800-273-8255 (Remix)" by Logic & Juanes ft. Alessia Cara & Khalid • "24" by Duki & Kidd Keo • "Adrenalina (Spanglish Version)" by Ricky Martin ft. Jennifer Lopez & Wisin • "A.N.A." by Kidd Keo • "Aquí Yo Mando" by Kali Uchis & Rico Nasty • "Arms Around You" by XXXTENTACION & Lil Pump ft. Maluma & Swae Lee • "AWAKE_001" by Kidd Keo • "Ayy Macarena" by Tyga & Ozuna • "Back It Up" by Prince Royce ft. Jennifer Lopez & Pitbull • "Back to the Future" by Kidd Keo ft. MAD BASS • "Back to the Future II" by Kidd Keo • "Bailando" by Enrique Iglesias ft. Sean Paul, Gente De Zona & Descemer Bueno • "Baller" by Kidd Keo • "Bando Boyz Free" by Kidd Keo • "Bando Girl" by Kidd Keo • "Barcelona" by Ed Sheeran • "Barefoot in the Park" by James Blake ft. ROSALÍA • "Beautiful Liar (Spanglish Version)" by Beyoncé • "Better Days" by OneRepublic & KHEA • "Bigger Love (Remix)" John Legend ft. Mau & Ricky • "BOUNCE" by Kidd Keo • "Box Boy" by Berner ft. Pablo Chill-E • "BRRUM FAST & FURIOUS" by Kidd Keo • "Buscando" by Yung Sarria & Kidd Keo • "Calypso" by Luis Fonsi ft. Stefflon Don • "Can't Tell Me Nada" by Alemán ft. Chucho Ortega, Cozy Cuz & Schmidt • "Caress Me Down" by Sublime • "CELEBRATE" by Black Eyed Peas • "Chapa" by Kidd Keo ft. MAD BASS & Elegvngster • "Chica Me Tipo" by Sublime • "Colors" by Jason Derulo & Maluma • "Con Calma (Remix)" by Daddy Yankee & Katy Perry • "Contra la Pared" by Sean

Methodology (Data Cleaning)

Text Preprocessing

- Selecting english lyrics.
- Removing song title part in the lyrics.
- Removing numbers or special characters.
- Removing stopwords.
- Extracting only chorus parts in the songs for topic modeling part.
- Chunking to obtain the appropriate sequence length to the pretrained sentence embeddings model.
- Applying average pooling for the lyrics which having more than 1 chunk.
- Filtering to remove the outliers by taking the lyrics between the 5th and 95th quantile based on their sentence length.

upf. Universitat Pompeu Fabra Barcelona

Sentiment Analysis

Two Aspects of Sentiment in Music

Tonal Sentiment

- Spotify measures the tonal sentiment of the track through a metric they call *valence*.

"A measure from 0.0 to 1.0 describing the musical positiveness conveyed by a track. Tracks with high valence sound more positive (e.g. happy, cheerful, euphoric), while tracks with low valence sound more negative (e.g. sad, depressed, angry)" (Spotify, n.d.)

Lyrical Sentiment

- We generated the lyrical sentiment by applying TweetNLP to the lyrics of each song on the Billboard Hot 100.

Label: {neutral, positive, negative}

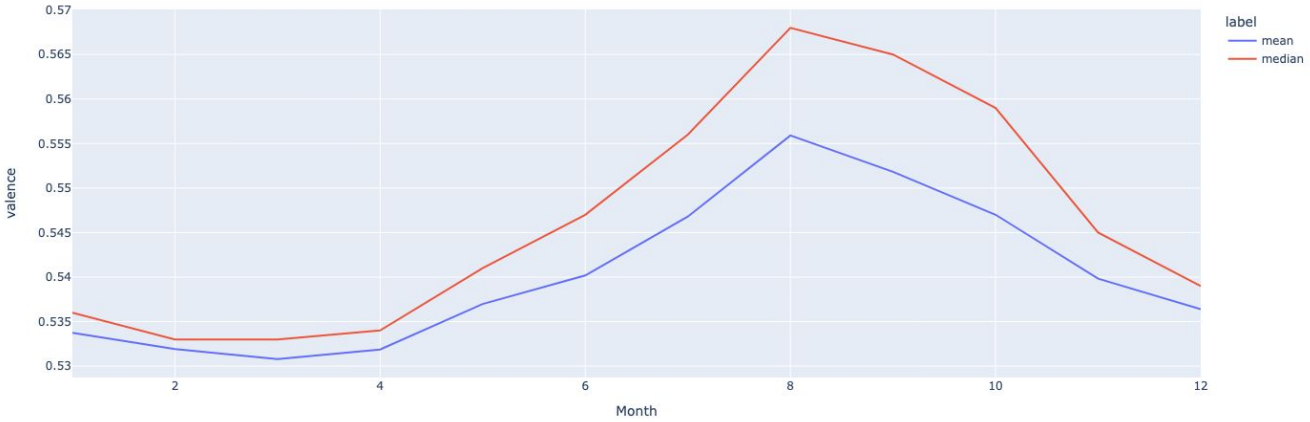
Probabilities for each sentiment (neutral, positive, negative)

Music is becoming progressively sadder over time

	p-value	Test Statistic	Critical Value (5%)
ADF	0.0000	-55.2960	-2.8616
KPSS	0.0100	14.3195	0.4630

The yearly time-series is trend stationary

Seasonality in Musical Mood

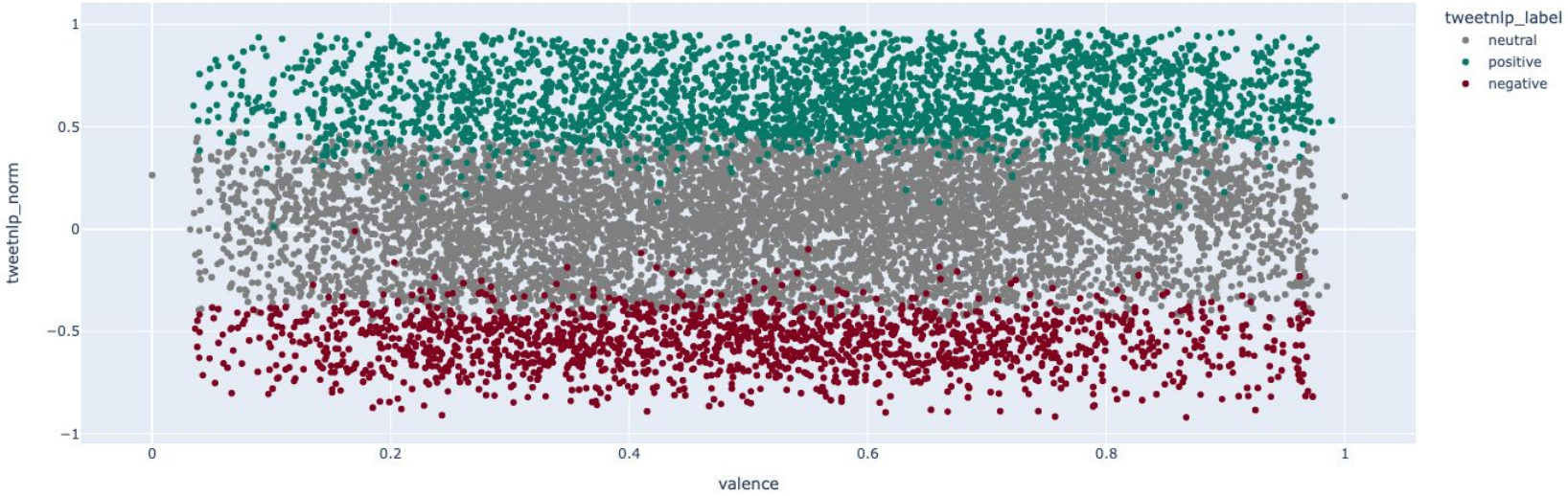


Happier songs tend to become more popular around the end of Spring to the middle of Autumn

	p-value	Test Statistic	Critical Value (5%)
ADF	0.0000	-55.2960	-2.8616
KPSS	0.0100	14.3195	0.4630

The monthly time-series is difference stationary

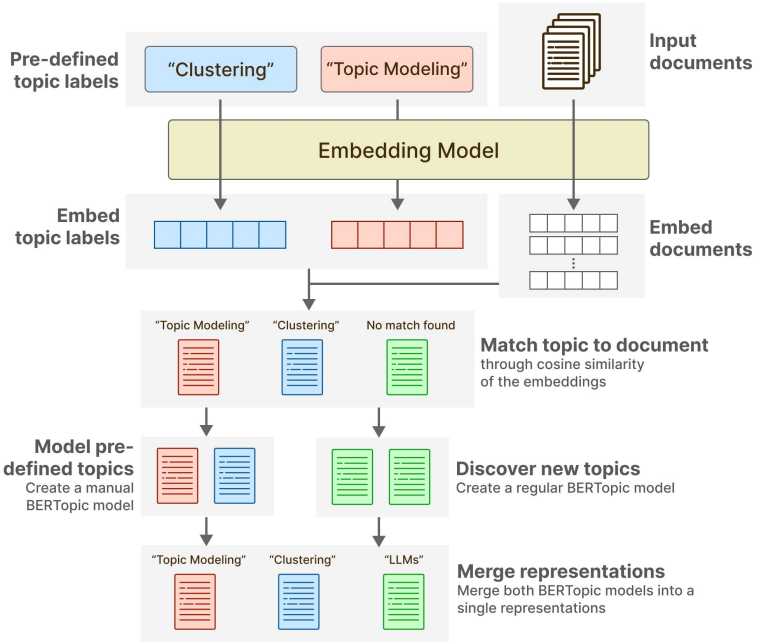
Lyrical Sentiment is not indicative of Musical Mood



upf. Universitat Pompeu Fabra Barcelona

Topic Modeling

Logic Behind BERTopic Technique



Source: [BERTopic: What Is So Special About v0.16?]

Steps of Topic Modeling

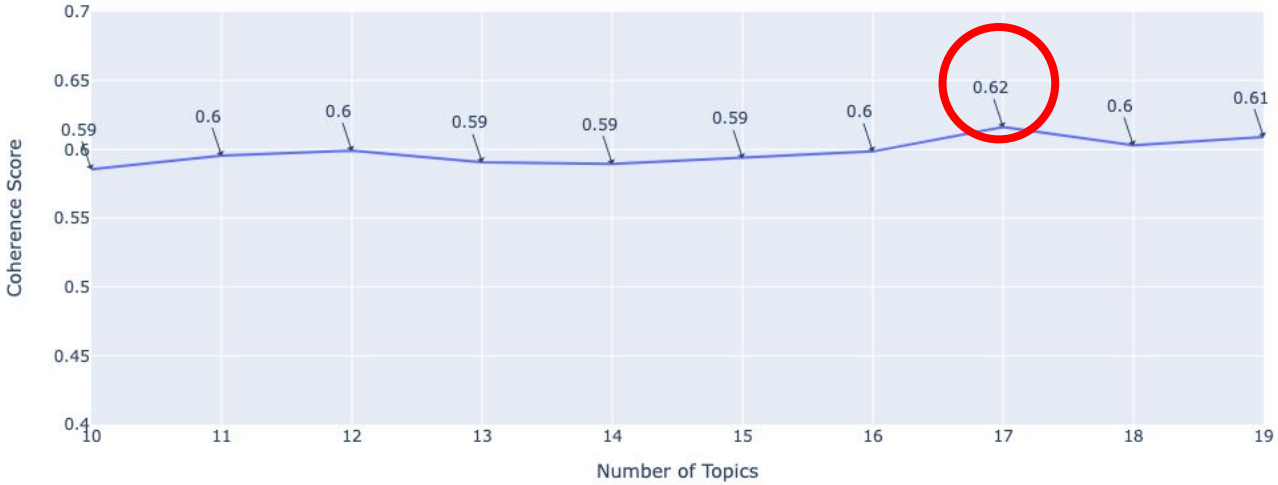
Used songs' choruses in topic modeling.

EMBED DOCUMENTS	Sentence Transformers (all-MiniLM-L6-v2)
DIMENSIONALITY REDUCTION	UMAP (Uniform Manifold Approximation and Projection for Dimension Reduction)
CLUSTERING	HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise)
TOPIC REPRESENTATION	CountVectorizer KeyBERT

Maximum sequence length: 256
Dimension: 384

Coherence Score for Optimization

Coherence metric evaluates documents similarity based on semantic similarity and co-occurance in assigned topics. Coherence score ranges between 0 and 1 (Higher is better). In addition, human-evaluation is also recommended.



Number of topics parameter in BERTopic is crucial for topic modeling performance.

Hardness of interpretation of topics in the topic model

Used ChatGPT 4 to find appropriate topic names to handle this issue.

Topic	Count	Name	Representation	Representative_Docs
-1	4605	-1_nigga_niggas_baby_wanna	['nigga', 'niggas', 'baby', 'wanna', 'ain', 'stay', 'girl', 'talk', 'bout', 'night']	['girl you know i can provide yeah you can get whatever you want whatever you need i m givin you whatever you want but i got you open now so i m bout to show you now finally got a rich nigga baby you made it we ain t fuc
0	2807	0_love_wanna_stay_don	['love', 'wanna', 'stay', 'don', 'babe', 'leave', 'ain', 'baby', 'say', 'want']	['you might also like because your love because your love said i wanna said i wanna said i wanna said i wanna know and walk with you down that aisle will you believe in love i wanna love you for life cause your love is why i l
1	1639	1_niggas_nigga_lil_ain	['niggas', 'nigga', 'lil', 'ain', 'yo', 'gang', 'ride', 'nothin', 'like', 'gettin']	['i put one leg on the headboard and leave the other leg on the mattress baby baby oh oh oh oh baby baby baby baby baby real hot girl shit i make him spend it all this boy done fell in love baby baby baby baby baby oh oh b
2	368	2_dreamer_dreams_dream_heaven	['dreamer', 'dreams', 'dream', 'heaven', 'pray', 'life', 'die', 'live', 'praying', 'survive']	['dream on dreamer life gets in your way then you can dream on dreamer life gets in your way you know you gotta dream on dreamer life gets in your way if the life you live is a spinning top of pain if you live to learn you ll be l
3	270	3_nigga_bitch_molly_pussy	['nigga', 'bitch', 'molly', 'pussy', 'girl', 'ain', 'say', 'wanna', 'girls', 'tellin']	['baby it s your world ain t it uh baby it s your world ain t it put it wherever you want she love me like a brother but fuck me like a husband pussy like a oven too hot to put my tongue in you might also like she tell me that s it s
4	33	4_rich_riches_money_cash	['rich', 'riches', 'money', 'cash', 'spend', 'dollars', 'payin', 'sell', 'work', 'buy']	['big white mansion in my habitat habitat smoke a lot of trees need a weed plant weed plant got rick owen jeans cost a cool rack cool rack big white mansion is my habitat habitat aim a lot of sticks like it s laser tag laser tag t
5	31	5_angel_angels_hallelujah_sing	['angel', 'angels', 'hallelujah', 'sing', 'choirs', 'singing', 'darling', 'bless', 'preacher', 'church']	['you re all i need to know tonight you re my angel my angel give me shelter from the rain you re the answer to my prayer you breathe life in me again and you re with me everywhere you might also likeoh you re my angel aha
6	29	6_santa_christmas_rudolph_reindeer	['santa', 'christmas', 'rudolph', 'reindeer', 'claus', 'sleighs', 'caroling', 'merry', 'holly', 'december']	['but last christmas i gave you my heart this year to save me from tears you got it last christmas i gave you my heart but the very next day you gave it away gave it away this year to save me from tears this year oh baby to sav
7	22	7_girls_girl_wanna_shorty	['girls', 'girl', 'wanna', 'shorty', 'tatted', 'laid', 'wants', 'want', 'gimme', 'boy']	['i m tryna pick the right one tryna pick the right one tryna pick the right one tryna pick the right one let s go there s a lot of pretty girls in this city hey there s a lot of pretty girls in this city there s a lot of pretty girls in this city i
8	20	8_pitches_pitching_braves_bat	['pitches', 'pitching', 'braves', 'bat', 'play', 'run', 'preach', 'win', 'talkin', 'swung']	['this game is different you only get one shot when niggas go foul on you man fuck em all man we want it all i do not chase girls but they run a mile for me better be comin with no strings we need some really big rings man w
9	19	9_love_real_loving_heart	['love', 'real', 'loving', 'heart', 'feels', 'feel', 'reality', 'lie', 'loved', 'livin']	['let s just be honest we all know the deal let s just be honest let s just be real yeah let s just be honest let s just be real so let s just be honest let s just be let s just be real yeah let s just be honest let s just be real so let s just
10	17	10_smile_singing_happy_crying	['smile', 'singing', 'happy', 'crying', 'laughin', 'say', 'laughing', 'breath', 'feel', 'tears']	['no i just couldn t help myself help myself at first when i see you cry i was quite unwell at worst i feel bad for a while but then i just smile but then i just smile but then i just smile and i tell you it don t mean jack it don t mean j
11	16	11_blame_blaming_fault_mistakes	['blame', 'blaming', 'fault', 'mistakes', 'guilty', 'missin', 'suffer', 'ain', 'hurt', 'watch']	['don t blame it on me blame it on the ni i i i ight don t blame it on me e e e e so blame it on the ni i i i ight don t blame it on me don t blame it on', 'blame it on me ayy you can t blame it on me ayy every day s the same yeah i
12	8	12_fight_fightin_youth_screamin	['fight', 'fightin', 'youth', 'screamin', 'tears', 'thunder', 'yell', 'burnin', 'kids', 'sacrifice']	['well only the good die young woah woah woah woah woah darlin only the good die young that s what i said only the good die young you know that only the good die young tell you baby oh oh ho baby only the good i tell ya
13	7	13_lights_shining_light_bright	['lights', 'shining', 'light', 'bright', 'headlights', 'dark', 'night', 'fade', 'backyard', 'runway']	['extra bright i want y all to see this want you to see all of the lights turn up the lights in here baby you know what i need want you to see everything', 'as i remember i know you love to show off but what do i know shining ligh
14	7	14_daddy_pills_mama_dessert	['daddy', 'pills', 'mama', 'dessert', 'addicted', 'body', 'wanna', 'baby', 'taste', 'dad']	['love so strong then you moved on wo oh oh oh ooh wo oh oh oh ooh girl s got a love like woe i kinda feel like it don t make sense cause you re bringing me in and now you re kicking me out again wo oh oh oh ooh la da dah
15	5	15_unknown_deep_sail_wind	['unknown', 'deep', 'sail', 'wind', 'line', 'sea', 'knows', 'train', 'cross', 'victory']	['into the unknown into the unknown', 'delicate in every way but one the swordplay glory and gore go hand in hand you might also like chance is the only game i play with baby oh oh everyone a rager oh oh but secretly they

upf. Universitat Pompeu Fabra Barcelona

Prompt Engineering

Prompt Template

Your role: You are an expert Customer Service Lead with 30 years of experience in customer service. Your wisdom is coming from your unique experience. You are the best about extracting useful insights from the data and also detecting the main theme in the data.

ROLE

Your task: I am sharing a csv file with you attached. This file contains several information about songs which is clustered. It has a column that is called 'Representation'. In this column there are a lot of keywords. These keywords represents the most frequent words of the song lyrics in each clusters. Your task is to find the most comprehensive, meaningful and appropriate cluster name based on the keywords i each row. Each row MUST represent one unique cluster name. You MUST NOT use the same cluster name You SHOULD create 17 different cluster names. The cluster names SHOULD contain maximum 3 words.

TASK

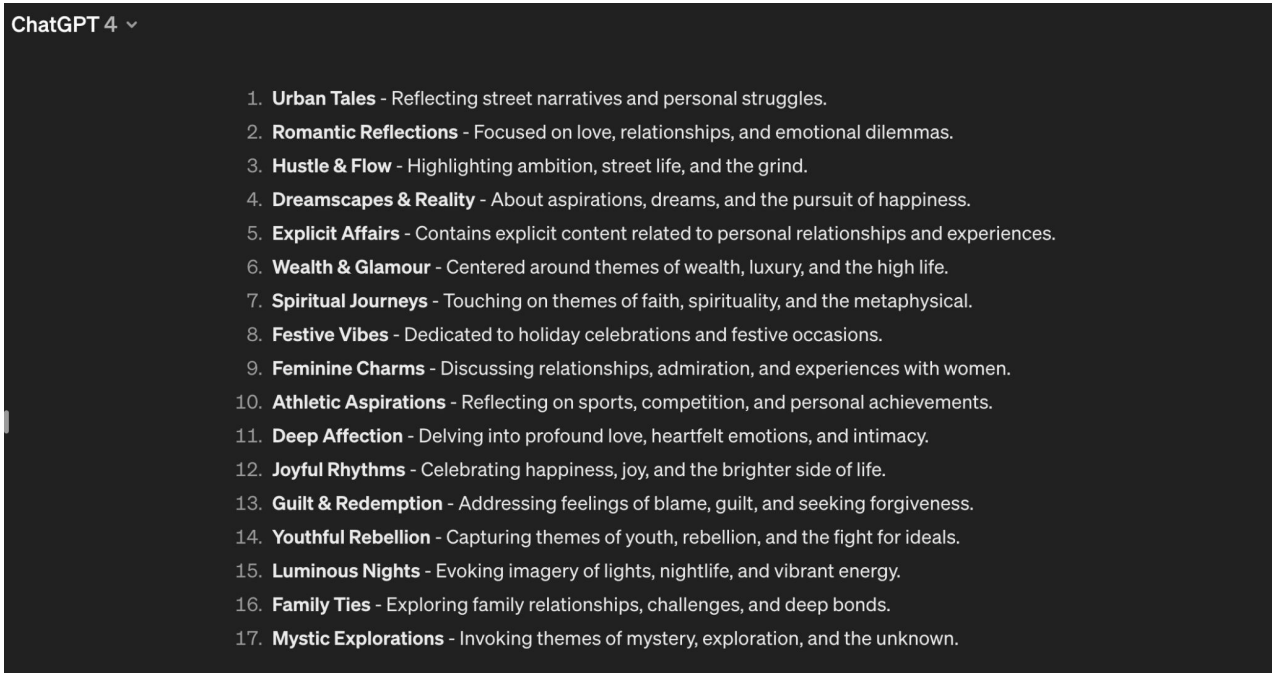
The steps you must follow:
1. Examine the "Representation" column to understand the context and grouping of keywords.
2. If you need to understand more about the concept of each cluster, you can also examine 'Representative_Docs' column to find a better cluster name.
3. After examining and understanding these keywords you SHOULD think about what topic might be appropriate for this cluster to represent it best.
4. You MUST be sure that you don't give the same cluster name to any other clusters.
5. The cluster name that you create MUST relate with the lyrics not the music style.
6. You just need to focus on what the lyrics are about.
7. If the lyrics contain explicit, racist or pornographic words you can create a cluster name using prefix as explicit, racist or pornographic.
8. You SHOULD be sure to follow all these steps above before giving a response.

STEPS

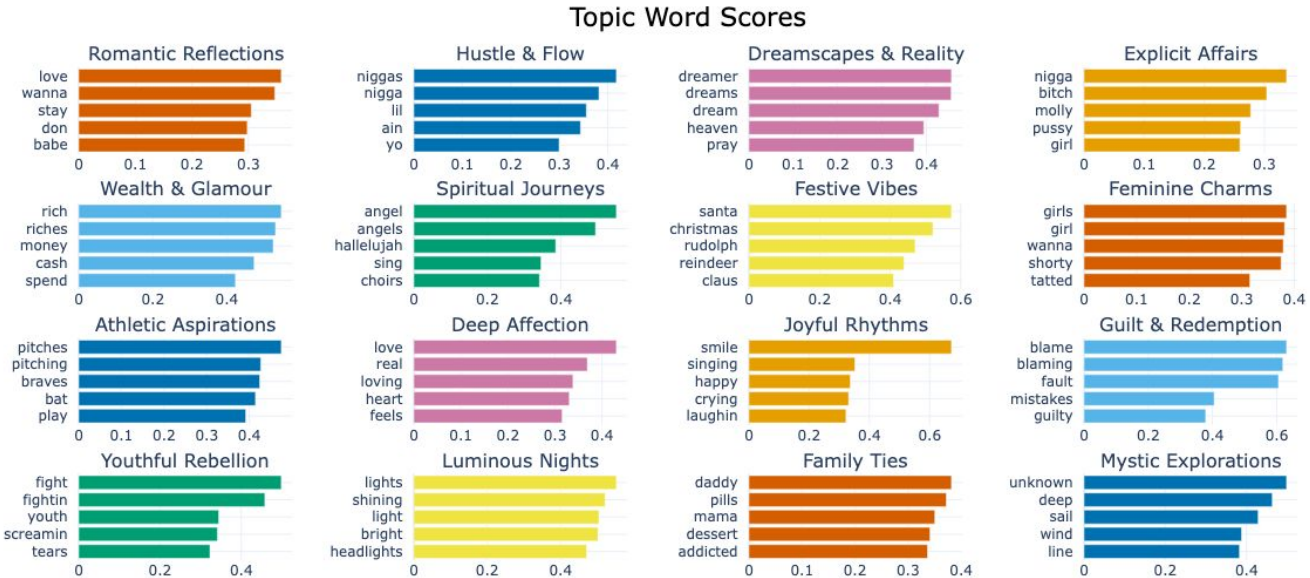
Output format: Share them with me as a spreadsheet by adding a new column at the end of existing file.The column name should be 'Topic Name'.

RESPONSE FORMAT

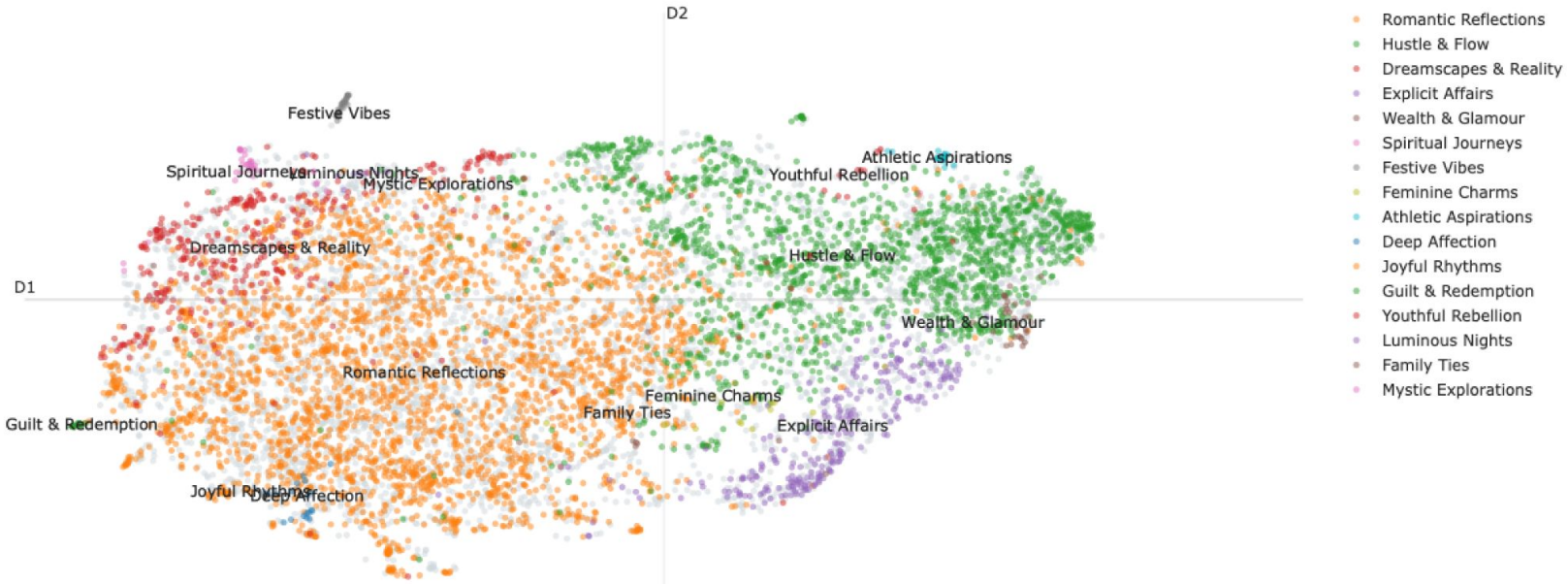
Topic Descriptions Assigned by ChatGPT 4



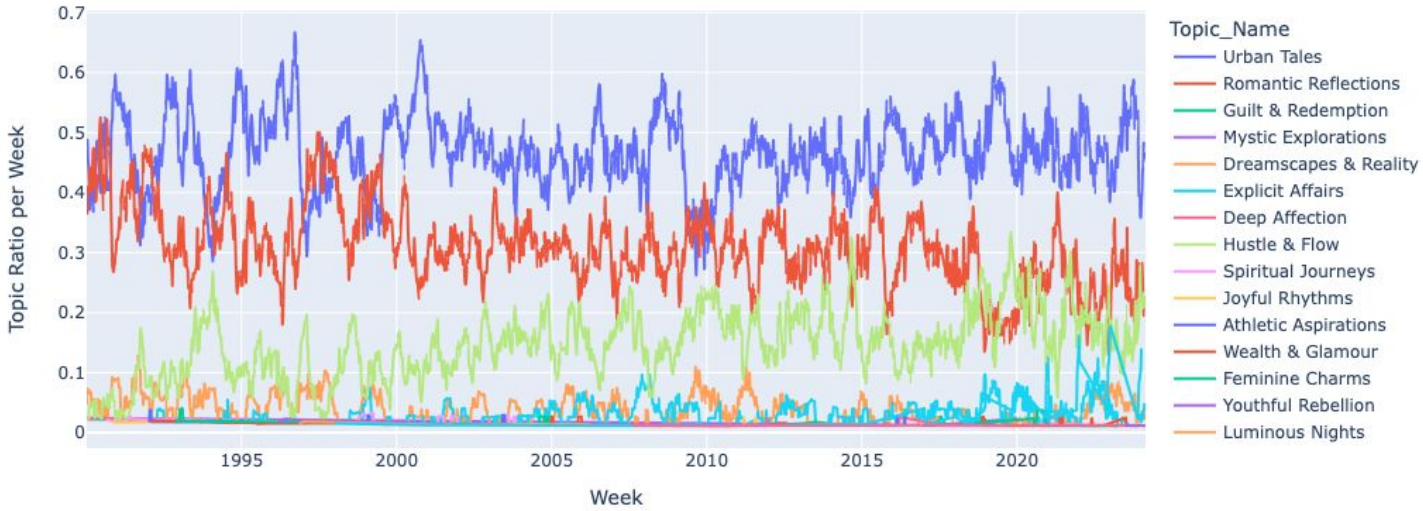
The Most Frequent Words Per Topic



Topic Clusters



While romanticism is decreasing, aggression is increasing in choruses over time



upf. Universitat Pompeu Fabra Barcelona

Additional Analysis

Songs are getting shorter over time

	p-value	Test Statistic	Critical Value (5%)
ADF	0.0000	-24.7878	-2.8616
KPSS	0.0100	53.7981	0.4630

The yearly time-series is trend stationary

Songs are getting shorter over time

Conclusion

Conclusion

1. Overall Music is becoming more somber in its tonal qualities
2. People's preference towards romantic lyrics and music is decreasing, while currently people prefer more aggressive and vulgar lyrics
3. There is an evident trend that shorter tracks are becoming more popular

Future Work

Future Work

- Additional Analysis on trends in music:
  - Changes in popular genres
- Look into external factors on popularity
  - Historical events
  - Social media trends
- Finding alternatives to Genius API
  - Could improve our topic and sentiment analysis
- Topic Modeling for the entire song
  - Obtaining compute power

References

Spotify. (n.d.). Retrieved on March 13, 2024, from https://developer.spotify.com/documentation/web-api/reference/get-audio-features.

BERTopic modeling technique details in

BERTopic (maartengr.github.io)