Evolution of the music industry in India

Group 6

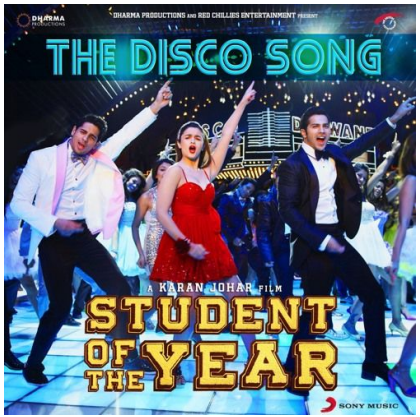
Madhushree

Prakhar Rathi

Rohit James

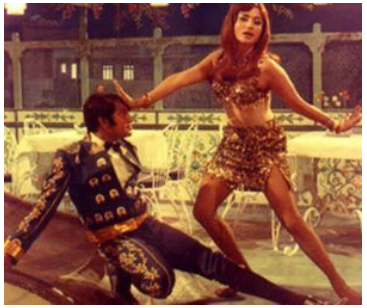
Vinayak Jha

Geetika (TA)



Background

Trends in music keep changing and India has experienced similar trends over decades.



Bollywood is more "an extended commodity function of the 'high profile, export oriented Bombay film,' which is about branding of India rather than a presentation of an aesthetic form, and that its use as a term of reference should be historicized to the early 1990"

(Vasudevan 2011: 6)

Discussion

- More foreign words
- Disappearance of Urdu words
- A new protagonist who represented the self-image of its elite and the aspirations of other classes

Urdu sounds are disappearing from Bollywood songs

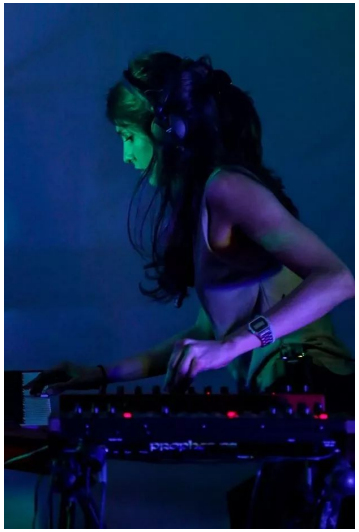
RIZWAN AHMAD 30 September, 2018 10:07 am IST

Research Questions

1. How has the music consumption of popular songs in India changed between 1960-2022?
2. Have the prominent themes of popular songs changed in the same period?

Data Sources

1. Spotify API
2. Genius Lyrics



Methodology

1. We made a list of top playlists on Spotify to collect the data on the most popular songs in each decade.

2. Iterating through each playlist, we collected the songs and the features of a track provided by the Spotify API as follows:

- Mood: Danceability, Valence, Energy, Tempo

- Properties: Loudness, Speechiness, Instrumentalness

- Context: Liveness, Acousticness

3. These features help us divide the music into many different properties and by comparing these properties, we can see how music changed in India from 1960-2020.

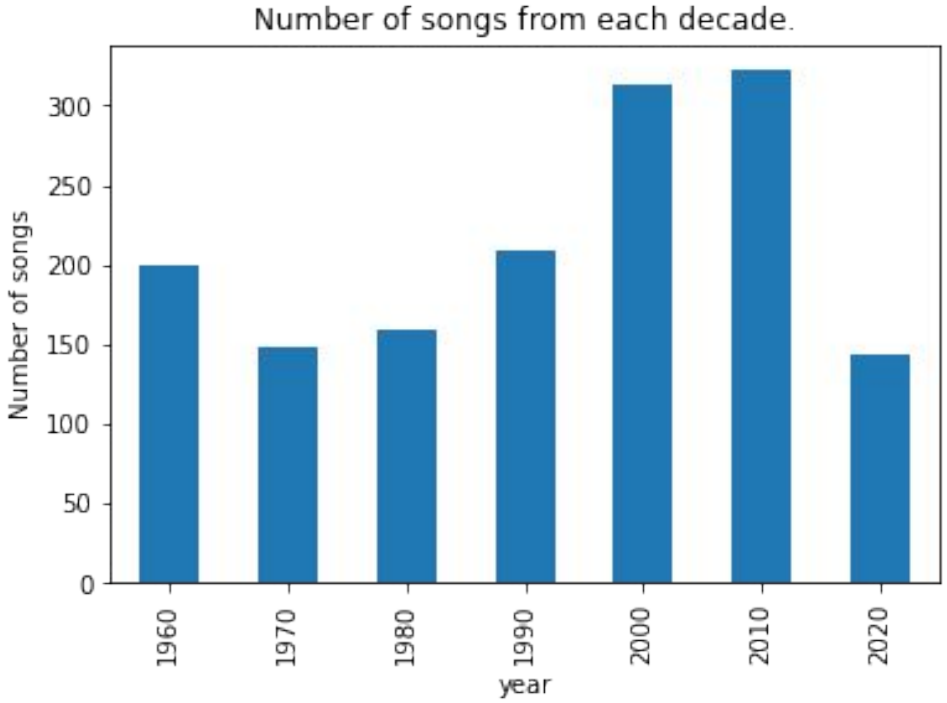
4. For the lyrics, we use the web scraping and manual data collection to get the Hindi and Hinglish Lyrics.

Data Description

Spotify API Data

- Spotify provides a rich API for collecting music data from their platform which is publicly available.
- We can collect things like track details, artist details and public playlist details.
- We collected a total of **1497** songs from the API and collected the data for the following features of the songs:
  - Mood: Danceability, Valence, Energy, Tempo
  - Properties: Loudness, Speechiness, Instrumentalness
  - Context: Liveness, Acousticness
  - Meta: Song Duration, year, decade
- The data was collected between 1960 and 2020 using the API and Python Spotipy library.

Data Distribution



How does Spotify compute the track features?

- **Popularity** is an important metric that we use - The popularity is calculated by algorithm and is based, in the most part, on the total number of plays the track has had and how recent those plays are. Generally speaking, songs that are being played a lot now will have a higher popularity than songs that were played a lot in the past.
- The rest of the metrics have been defined in detail in the documentation but their calculations methods have not been made publicly available.
- The details can be found in the documentation - https://developer.spotify.com/documentation/web-api/reference/#/operations/get-audio-features

Lyrics Data

- Selected the top 10 songs from each decade and collected the lyrics for those songs.
- Genius API does not have the complete lyric data, especially for old Hindi songs.
- We collected the lyrics manually for these songs in Devanagri script and code-mixed English text.
- We used the Bing Translation API to convert the Hindi lyrics into English words so that we can use the existing SOTA models for various text analysis methods.

Top Songs for each Decade - 1960s

- Likhe Jo Khat Tujhe
- Abhi Na Jao Chhod Kar
- Roop Tera Mastana
- Ae Mere Zohra Jabeen
- Aane Se Uske Aaye Bahar
- Beqarar Karke Hamen Yun Na Jaiye
- Tum Agar Saath Dene Ka Vada Karo
- Isharon Isharon Men Dil Lenewale
- Tere Mere Sapne Ab Ek Rang Hain
- Aap Ki Nazron Ne Samjha

1970s

- Kya Khoob Lagti Ho
- Yunhi Tum Mujhse
- Kehdoon Tumhen
- Kabhi Kabhi Mere Dil Mein
- Main Pal Do Pal Ka Shair Hoon
- Main Shair To Nahin
- Mehbooba Mehbooba
- Kahin Door Jab Din Dhal Jaye
- Kabhi Kabhi Mere Dil Mein
- Zindagi Kaisi Hai Paheli

1980s

- Neele Neele Ambar Par
- Ae Mere Humsafar
- Dekha Ek Khwab
- Main Duniya Bhula Doonga
- Tum Ko Dekha To Yeh Khayal Aaya
- Dheere Dheere Se Meri Zindagi Mein Aana
- Tujhse Naraz Nahin Zindagi
- Janu Meri Jaan
- Jaan - E - Jigar Jaaneman
- Tera Naam Liya

1990s

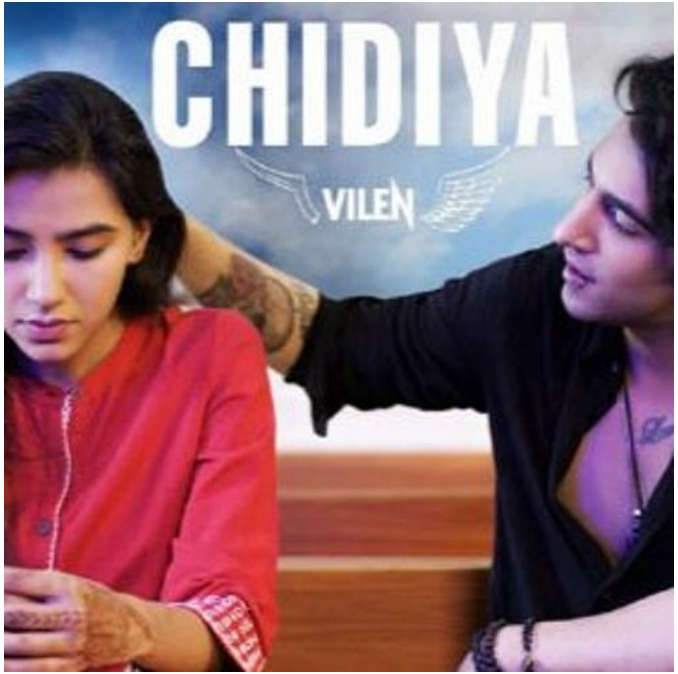
- Yeh Haseen Vadiyan Yeh Khula Aasman
- Kuch Kuch Hota Hai
- Pehla Nasha
- Tadap Tadap
- Ladki Badi Anjani Hai
- Koi Mil Gaya
- Taal Se Taal
- Chaiyya Chaiyya
- Ae Kash Ke Hum
- Baazigar O Baazigar

2000s

- Tum Se Hi
- Khuda Jaane
- Zara Sa
- Tera Hone Laga Hoon
- Chand Sifarish
- Tujh Mein Rab Dikhta Hai
- Tu Jaane Na
- Kabhi Kabhi Aditi
- Jashn-E-Bahaaraa
- Ajab Si

2010s

- Chidiya
- Khairiyat
- Duniyaa
- Kabira
- Agar Tum Saath Ho
- Makhna
- Saibo
- Baarishein
- Kalank (Title Track)
- Kaise Hua



2020s

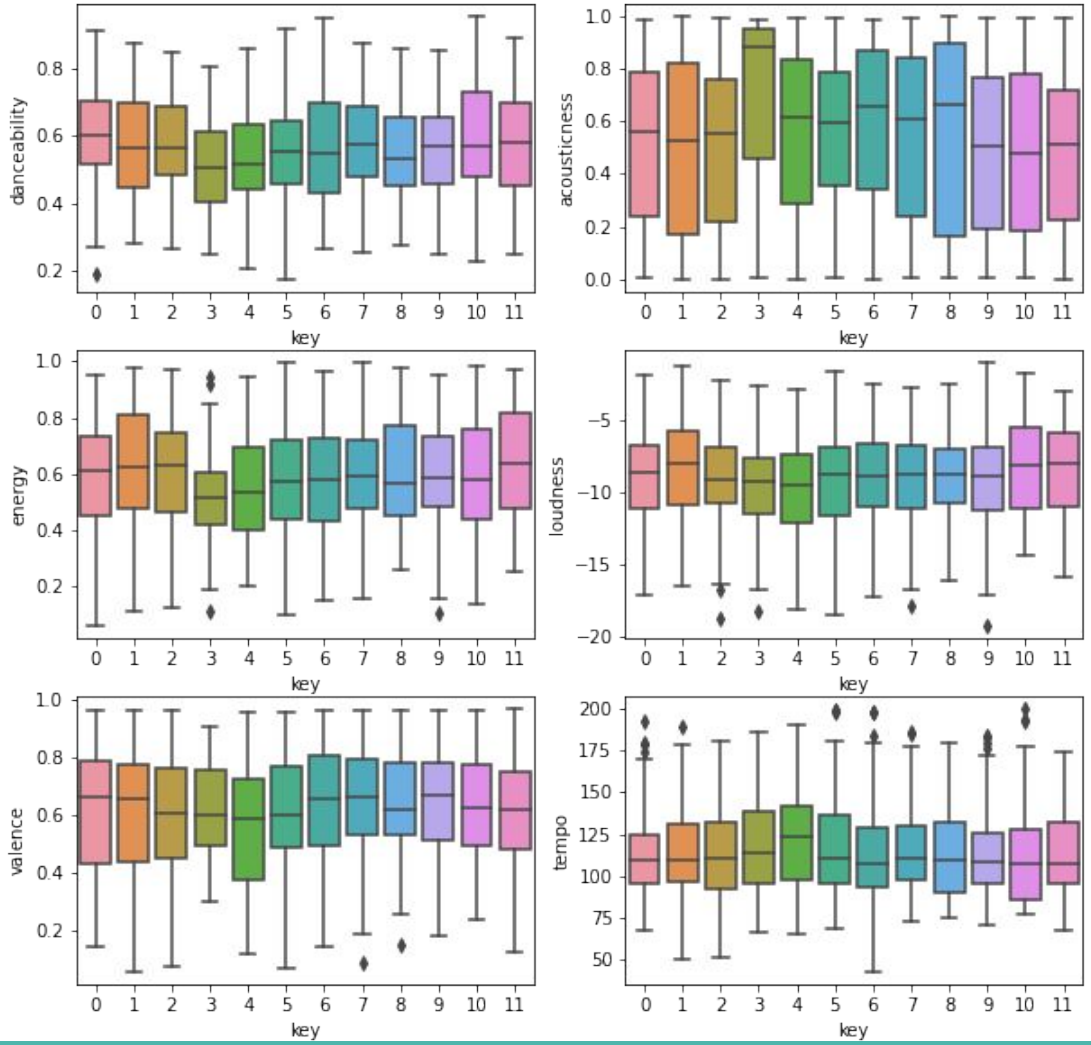
- Chaand Baaliyan
- Tu Aake Dekhle
- Raataan Lambiyan
- Kesariya
- Ranjha
- Bhool Bhulaiyaa 2 Title Track
- Akhiyaan
- Dil Ko Karaar Aaya
- Meri Jaan
- Atak Gaya

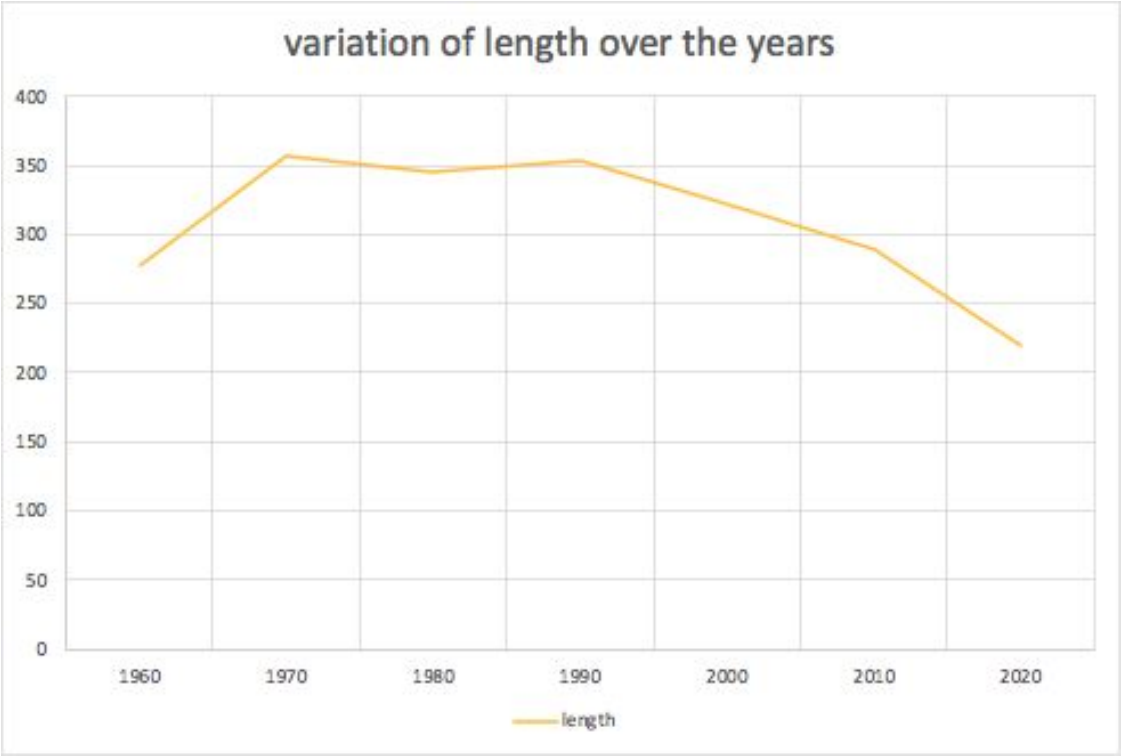
Preliminary Analysis

Evolution - through the features

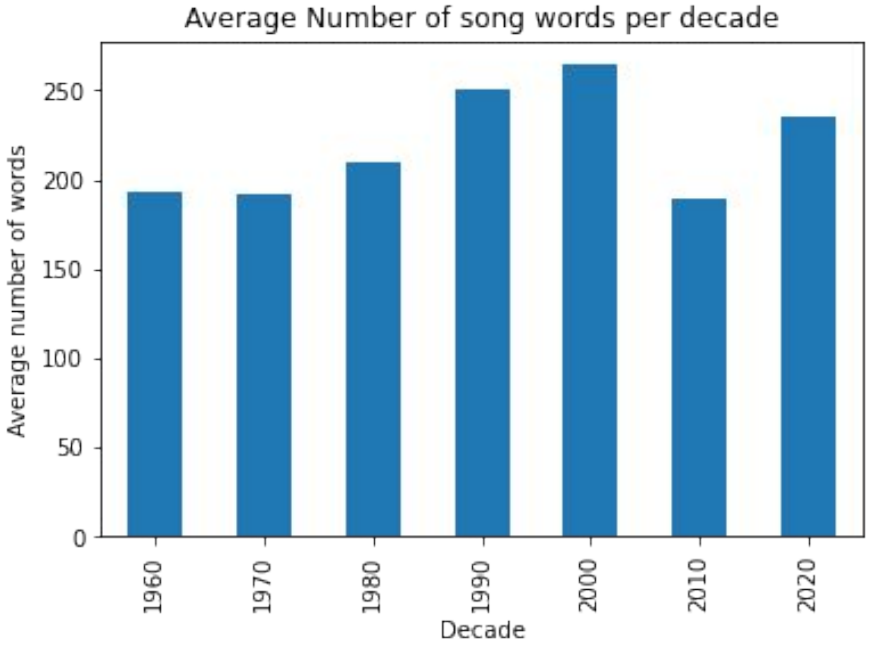
	danceability	acousticness	energy	loudness	speechiness	valence	tempo	time_signature	mode
0	0.472366	0.890345	0.471379	-10.129060	0.069909	0.660549	110.948962	3.778723	0.706383
1	0.451934	0.804987	0.516815	-9.717166	0.110026	0.573988	113.019735	3.794702	0.655629
2	0.507029	0.636761	0.563727	-9.890977	0.067581	0.617696	112.181983	3.920000	0.634286
3	0.579108	0.476888	0.609386	-10.888695	0.060376	0.690887	118.277453	3.905830	0.511211
4	0.648818	0.303538	0.677611	-8.180149	0.071457	0.625036	114.597644	3.972644	0.613982
5	0.609680	0.398887	0.666148	-7.501742	0.071436	0.618220	117.815217	3.881306	0.602374
6	0.619593	0.488323	0.549883	-8.031241	0.064481	0.452657	114.529641	3.875862	0.703448

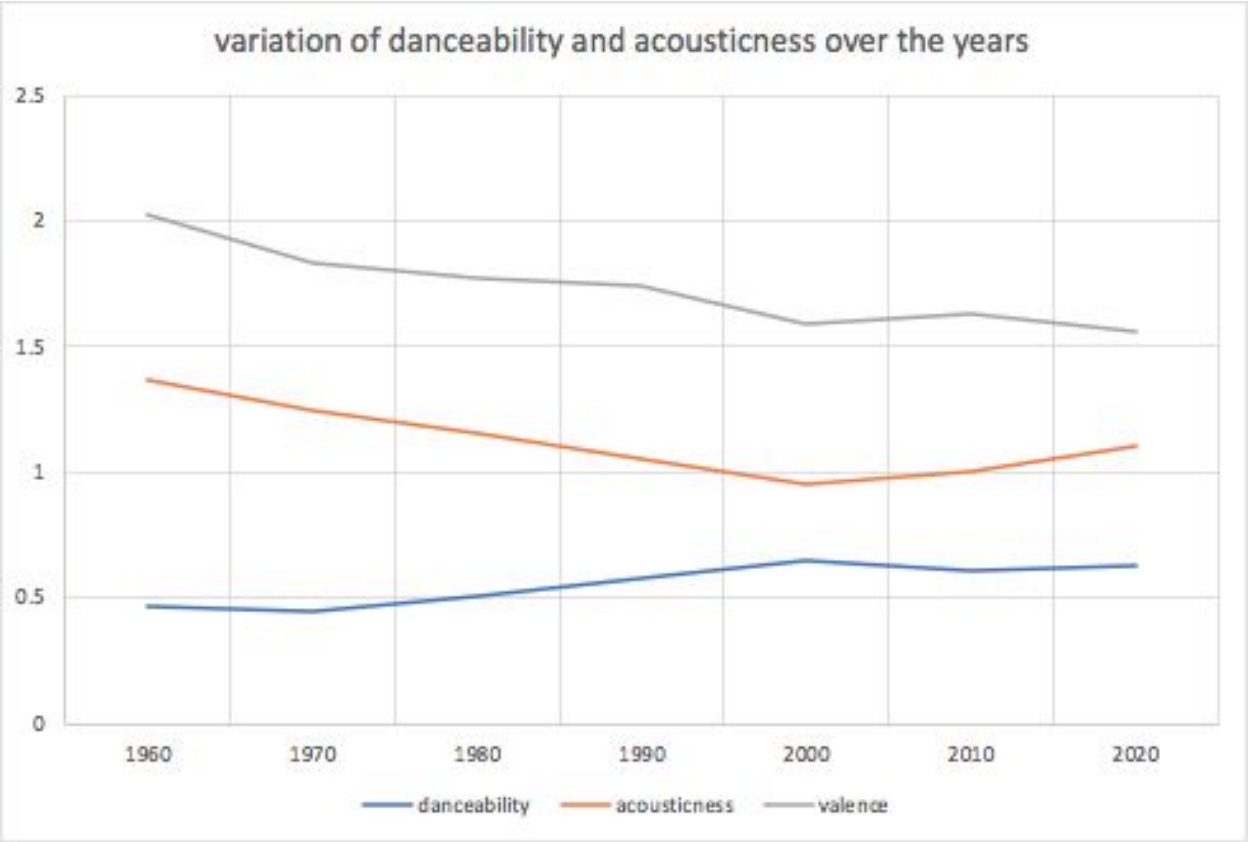
- Significant fall in track valence since 2020 - tracks sound more negative (e.g. sad, depressed, angry).
- However, danceability of tracks has changed little since 2000s.


danceability
key
acousticness
key
energy
key
loudness
key
valence
key
tempo
key


variation of length over the years
400
350
300
250
200
150
100
50
0
1960
1970
1980
1990
2000
2010
2020
length

Number of words in songs




variation of danceability and acousticness over the years
2.5
2
1.5
1
0.5
0
1960
1970
1980
1990
2000
2010
2020
danceability
acousticness
valence

Similarity Analysis

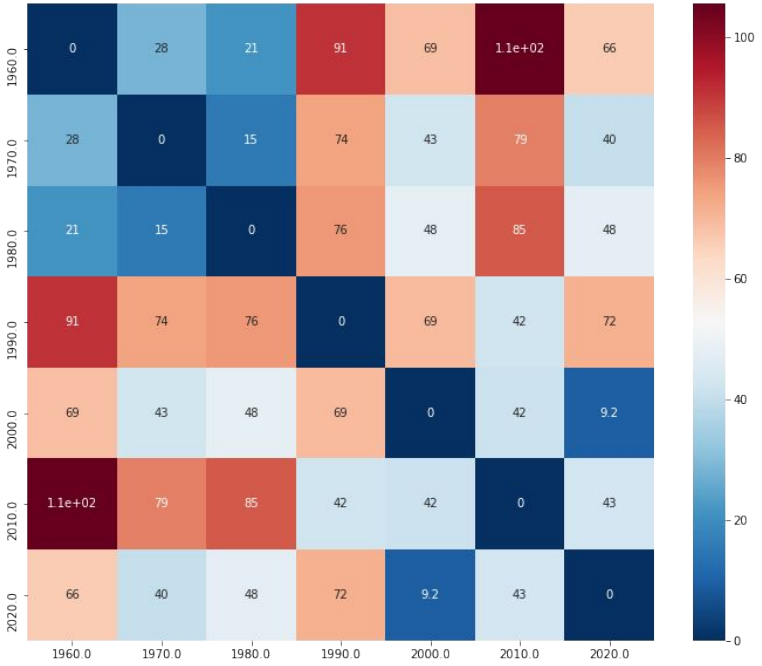
Similarity Analysis

- Aggregated the decade-on-decade data to analyse how similar (or different) the features of songs were in each decade.

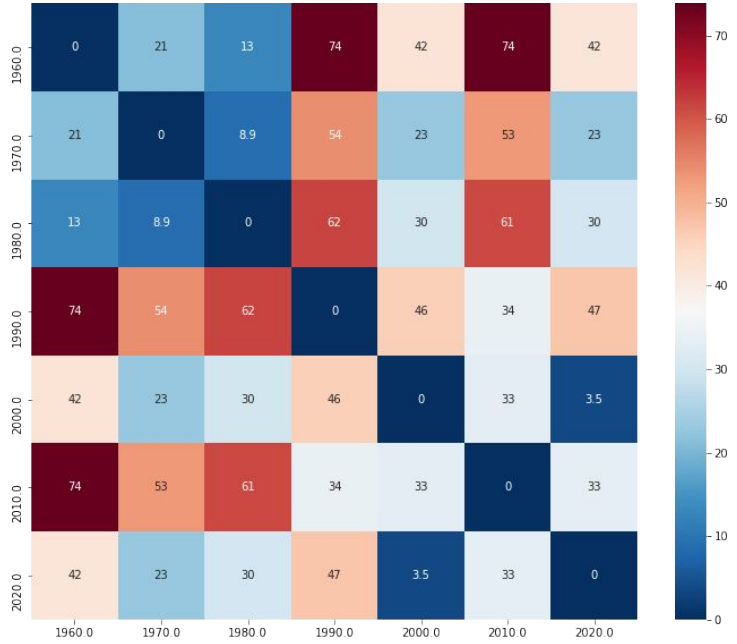
- Further, we calculated the Euclidean Distances and Manhattan Distances to see which decades have similar songs.

- The graphs are presented here.

Similarity Matrices - Visualization



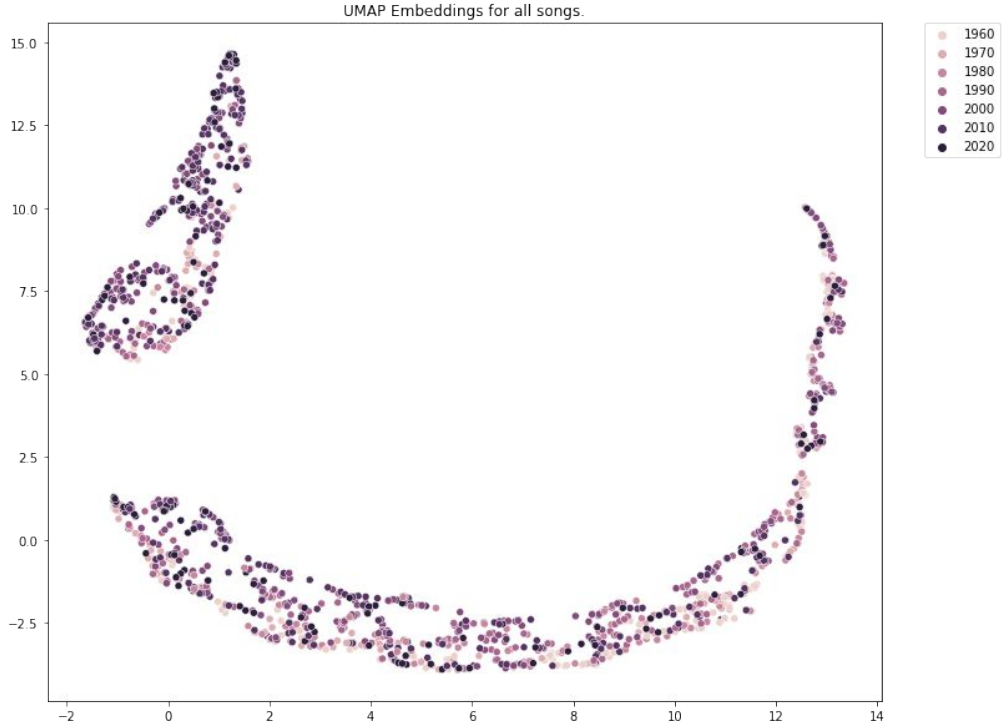
Manhattan Distance



Euclidean Distance

U-Map

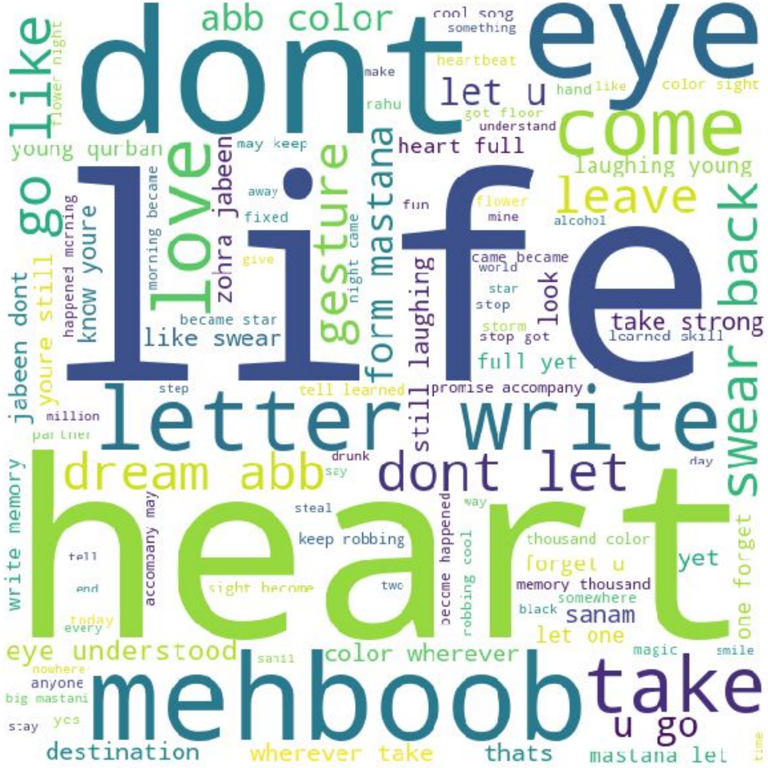
1. **Uniform Manifold Approximation and Projection (UMAP)** is a dimension reduction technique.
2. We wanted to see how individual songs are similar to each other and different to each other.
3. We also want to explore why **this (image on right)** division exists.
4. We further see that songs are not clustered on the basis of decades and we want to know what they are clustered on the basis of?

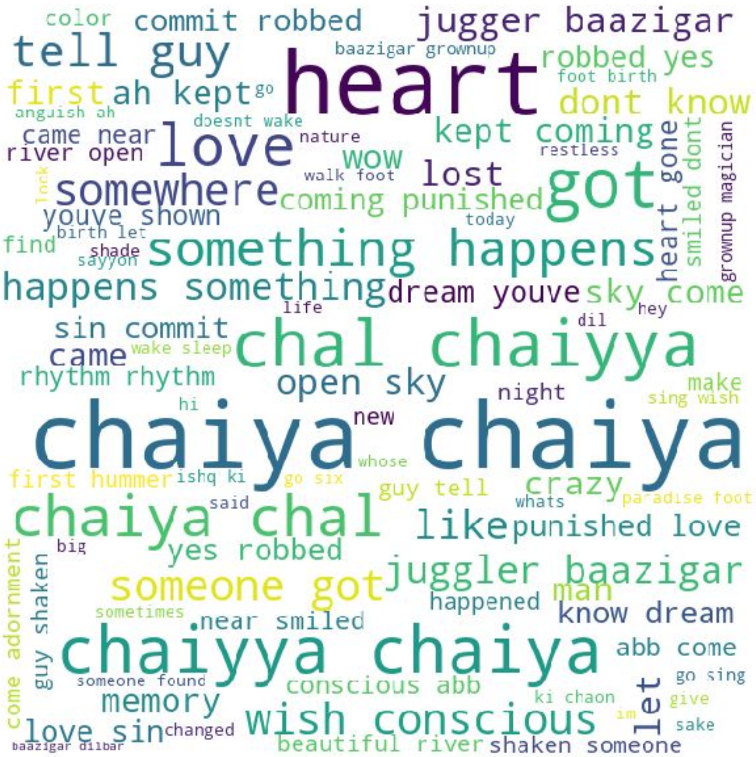


Lyric Analysis

- We collected the lyrics of the top 10 songs from each decade and translated them into English.
- We cleaned and lemmatized the text to identify patterns in the data.
- We made word clouds for each decade and looked at the most common bigrams.
- **Limitation -** We faced an issue because of the less number of lyrics which caused some songs to dominate certain decades due to their repetitiveness and length. This issue could be overcome by building a Selenium Web Driver for collecting data from the Genius.com website or some other website for old Bollywood songs.

Word Clouds


dont
eye
life
heart
mehboob
take
letter write
come
leave
love
gesture
form mastana
swear back
dream abb
dont let
go like
abb color
let u
heart full
laughing young
young qurban
zohra jabeen
eye understood
color wherever
destination
wherever take
thats
mastana let
u go
sanam
take strong
like swear
full yet
still laughing
look
know youre
jabeen dont
write memory
keep robbing
memory thousand
thousand color
yet
one forget
let one
promise accompany


heart
chaiya chaiya
chal chaiyya
chaiyya chaiya
chaiya chal
something happens
happens something
love
got
somewhere
wish conscious
someone got
juggler baazigar
jugger baazigar
commit robbed
tell guy
ah kept
first
robbed yes
dont know
kept coming
lost
wow
coming punished
youve shown
dream youve
sky come
sin commit
came
rhythm rhythm
open sky
night
like
punished love
yes robbed
happened
man
know dream
near smiled
abb come
memory
love sin
beautiful river
shaken someone
conscious abb
came near
river open
crazy
guy tell
heart gone
smiled dont
grownup magician
come adornment
guy shaken
let

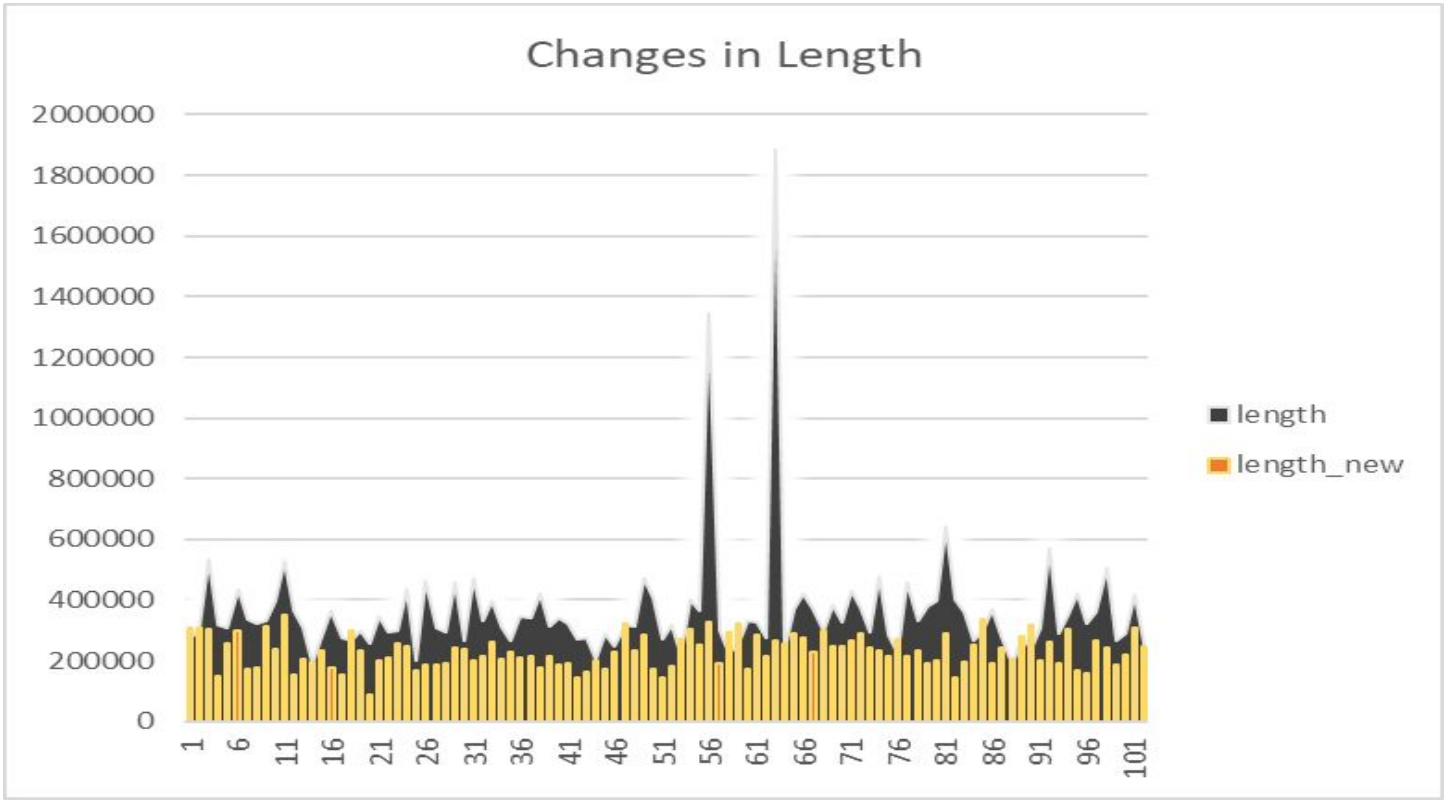
Remake Analysis

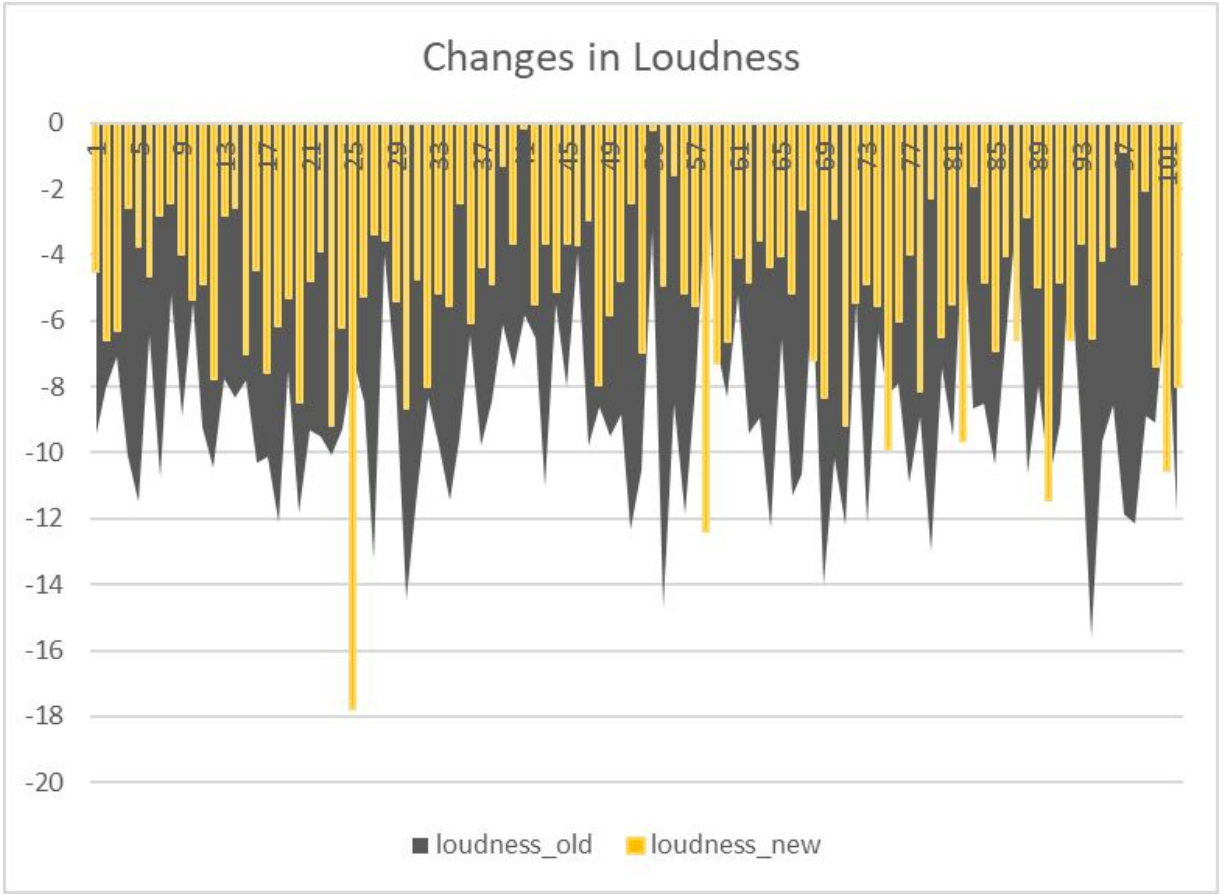
Remakes across the past 7 decades

- 200+ songs from 1950-2000s; from ‘Ghar Aaya Mera Pardesi’ to ‘Mera Dholna’
- Significant decrease in track length (~36%), but energy levels soar by over 13% in the remakes: Songs that are "catchy" and "short" do well in the context of consumption platforms that are driven by short-form content.
- A corresponding significant rise (16.5%) in danceability: Calculated using the tempo, rhythm stability, and beat strength, danceability increases when a song is consistent and uniform with little or no structural intricacies or experimentation.

Commerialisation of Art

Theodor Adorno and Max Horkheimer, addressed the growing uniformity in art and cultural forms as the result of what they called the mechanisations of the ‘culture industry’ in contemporary Western societies, that is, the penetration of industrial mass-production methods into the realms of culture and art. The basic kernel of Adorno and Horkheimer’s analysis of the cultural industries lies in the latter’s systematic fabrication of modes of behaviour whereby the profit-seeking, technology-driven Western capitalist ethic extends beyond material mass consumption to include cultural items and even people’s perceptions of what is ‘good’ and what is ‘smart’.


Changes in Length
2000000
1800000
1600000
1400000
1200000
1000000
800000
600000
400000
200000
0
1
6
11
16
21
26
31
36
41
46
51
56
61
66
71
76
81
86
91
96
101
length
length_new


Changes in Loudness
0
-2
-4
-6
-8
-10
-12
-14
-16
-18
-20
1
5
9
13
17
21
25
29
33
37
41
45
49
53
57
61
65
69
73
77
81
85
89
93
97
101
loudness_old
loudness_new

Current Limitations

- We collected the data based on the playlist and some pre-existing knowledge. However, this has a caveat with backtesting.
  - Album sales are possible in the western countries but in India songs are found in Bollywood movies.
  - A method could be performing surveys and asking people from that decade how they would rank popular songs.
  - Another method would be to identify the top artists from a decade using a survey and then analyzing their songs.
- Are the methods of creating features derived from Hollywood songs?
  - We can compare the most danceable Hollywood songs to the most danceable Bollywood songs and compare their ranking.
  - If the ranking is Hollywood specific, then building a model for song danceability through Spotify would be a nice idea.

Future Work and Ideas

1. Work on ‘song repetitiveness’ - In 1977, the great computer scientist Donald Knuth published a paper called The Complexity of Songs, which is about the repetitive lyrics of newfangled music. Similar work has been done in the US using the Billboard Top 100 list which doesn’t exist in India.

Thank you!