project 1

The impact of global events on music choice

Team 2 to Tango

Niharika, Hazel, Azlan, Peter and Enrique

executive summary

We wanted to **investigate** the correlations -if any- between key events in the US and abroad, and the music at the top of the weekly Billboard charts from **January 2022 to August 2024.**

Do global events impact the mood of a nation?
Can music tell us how?

By fetching data from multiple sources including **Billboard's top charts**, **Spotify Web API**, **Geopolitical Risk Index**, and **Wikipedia**, the team cleaned, transformed, analyzed, and plotted the data.

Creating a **'mood'** variable based on certain audio attributes for the top 10, then the **top 3 songs** over time, allowed us to question the data, visualize correlations, and hypothesized trends.

data fetching : music + events

data set

billboard

top ten songs in the US
from **01-01-2022**
to **08-08-2024**
136 .csv

Fetched 2022-01-01.csv
Fetched 2022-01-08.csv
Fetched 2022-01-15.csv

⋮

Fetched 2024-07-20.csv
Fetched 2024-07-27.csv
Fetched 2024-08-03.csv

data attributes

music

fetched **14** audio features for

205 unique songs

performed by **132** artists

over a period of **136** weeks

- acousticness
- danceability
- duration_ms
- energy
- id
- instrumentalness
- key
- liveness
- loudness
- mode
- speechiness
- tempo
- time_signature
- valence/positivity

data extraction : cleanup

data extraction

music data

- Billboard Hot 100 - top 10 songs from each week
- Spotify Web API - track_id -> audio features
- 11 spotify audio features were added to every billboard song

chart_week	rank	song	performer	danceability	energy	loudness	speechiness	acousticness	liveness	tempo	valence	time_signature	duration_ms
2022-01-01	1	All I Want For Christmas Is You	Mariah Carey	0.336	0.627	-7.463	0.0384	0.164	0.0708	150.273	0.350	4	241107

events data

- gpr data was extracted for the USA only

0.05 GPRC_SWE	Country GPR: Percent of articles (Sweden)
0.08 GPRC_THA	Country GPR: Percent of articles (Thailand)
0.02 GPRC_TUN	Country GPR: Percent of articles (Tunisia)
0.00 GPRC_TUR	Country GPR: Percent of articles (Turkey)
0.00 GPRC_TWN	Country GPR: Percent of articles (Taiwan)
0.05 GPRC_UKR	Country GPR: Percent of articles (Ukraine)
0.00 GPRC_USA	Country GPR: Percent of articles (United States)
0.03 GPRC_VEN	Country GPR: Percent of articles (Venezuela)
0.05 GPRC_VNM	Country GPR: Percent of articles (Vietnam)
0.06 GPRC_ZAF	Country GPR: Percent of articles (South Africa)
0.00 GPRHC_ARG	Country GPR Historical: Percent of articles (Argentina)
0.00 GPRHC_AUS	Country GPR Historical: Percent of articles (Australia)
0.01 GPRHC_BEL	Country GPR Historical: Percent of articles (Belgium)
0.01 GPRHC_BRA	Country GPR Historical: Percent of articles (Brazil)

- a clean .csv file was generated from the wikipedia extraction
- 143 events were tagged

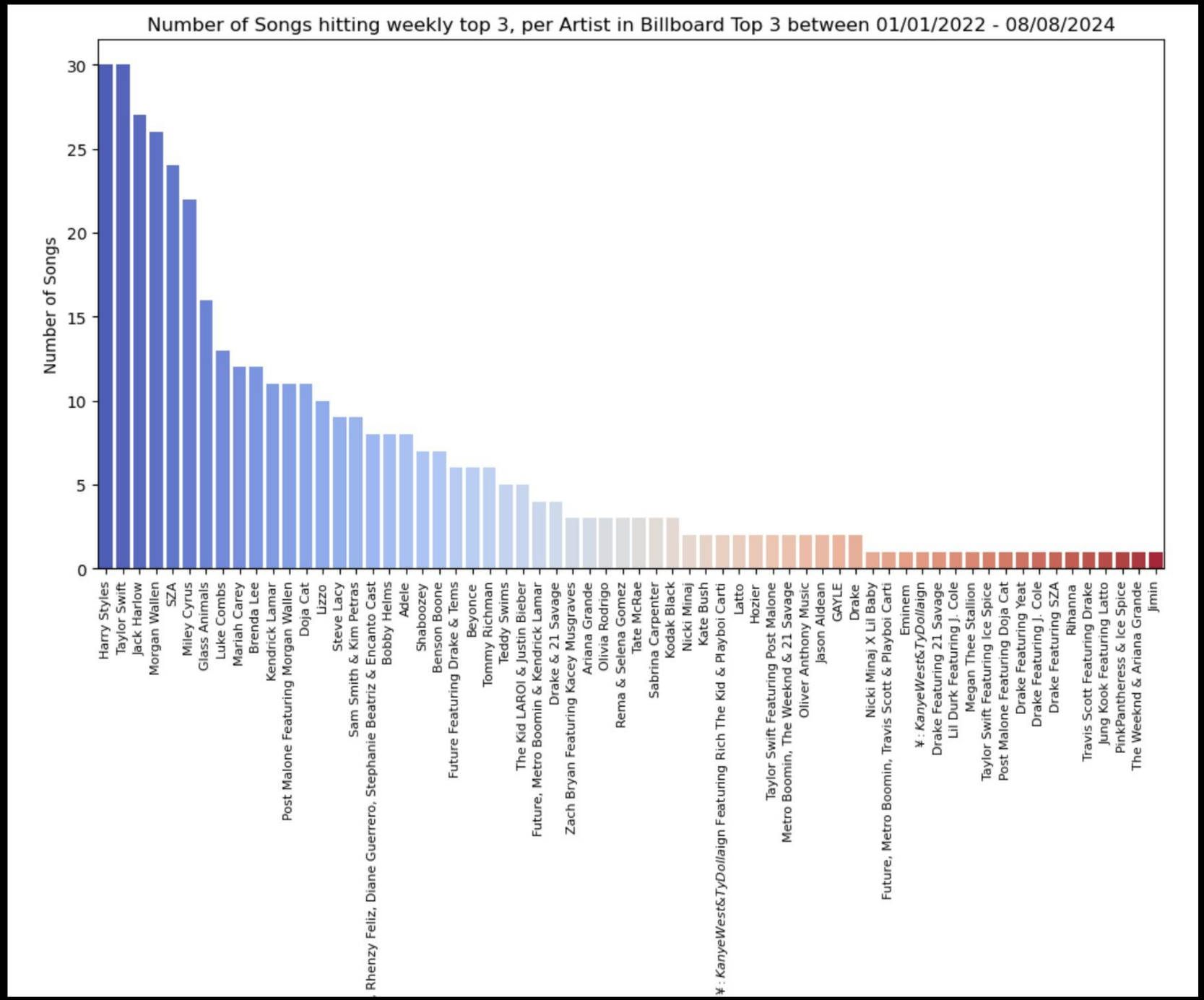
data exploration

music

	chart_week	rank	song	performer	danceability	energy	loudness	speechiness	acousticness	liveness	tempo	valence
0	2022-01-01	1	All I Want For Christmas Is You	Mariah Carey	0.336	0.627	-7.463	0.0384	0.1640	0.0708	150.273	0.350
1	2022-01-01	2	Rockin' Around The Christmas Tree	Brenda Lee	0.598	0.470	-8.744	0.0496	0.6170	0.5050	67.086	0.879
2	2022-01-01	3	Jingle Bell Rock	Bobby Helms	0.754	0.424	-8.463	0.0363	0.6430	0.0652	119.705	0.806
3	2022-01-01	4	A Holly Jolly Christmas	Burl Ives	0.682	0.375	-13.056	0.0303	0.5790	0.0760	140.453	0.888
4	2022-01-01	5	Easy On Me	Adele	0.604	0.366	-7.519	0.0282	0.5780	0.1330	141.981	0.130
...	...	...	...	...	...	...	...	...	...	...	...	...
1355	2024-08-03	6	Too Sweet	Hozier	0.741	0.620	-5.505	0.0412	0.0295	0.0398	117.038	0.934
1356	2024-08-03	7	Please Please Please	Sabrina Carpenter	0.669	0.586	-6.073	0.0540	0.2740	0.1040	107.071	0.579
1357	2024-08-03	8	Lose Control	Teddy Swims	0.561	0.604	-4.409	0.0337	0.1990	0.1040	159.920	0.242
1358	2024-08-03	9	Beautiful Things	Benson Boone	0.472	0.471	-5.692	0.0603	0.1510	0.1400	105.029	0.219
1359	2024-08-03	10	Good Luck, Babe!	Chappell Roan	0.700	0.582	-5.960	0.0356	0.0502	0.0881	116.712	0.785

consolidated data frame with all the audio features for 1360 entries needed for our analysis

data analysis

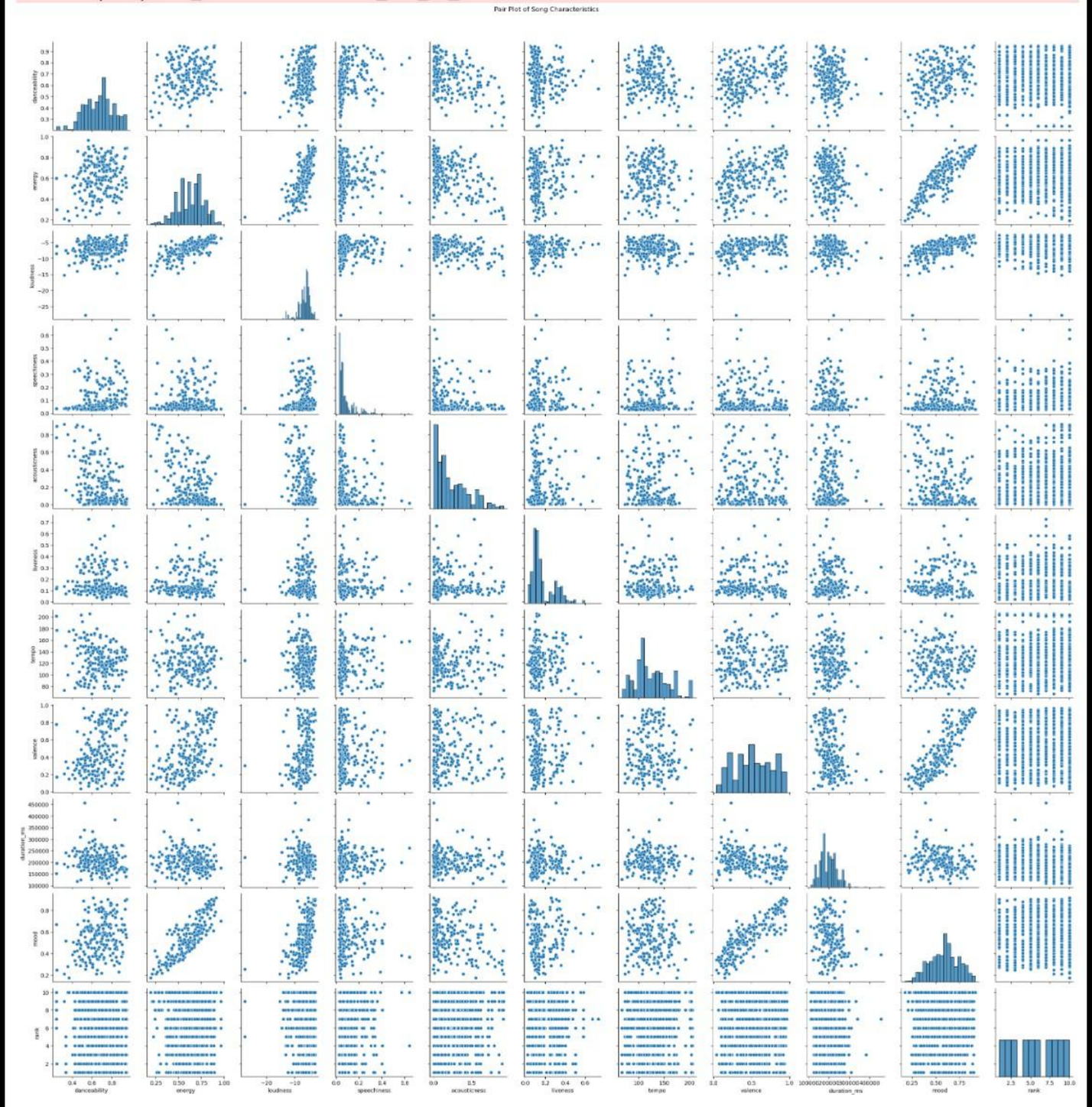


music

question #1

who are the top ranking artists?

data analysis



music

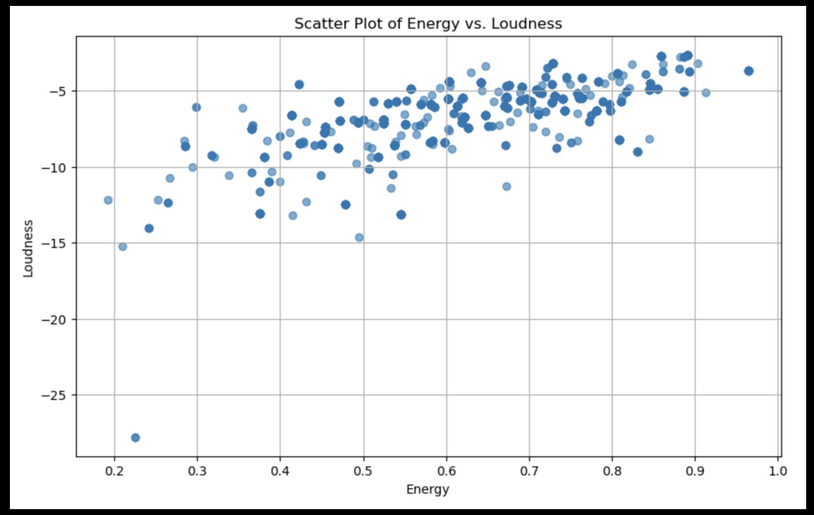
question #2

are there positive or negative correlations amongst the attributes of top songs

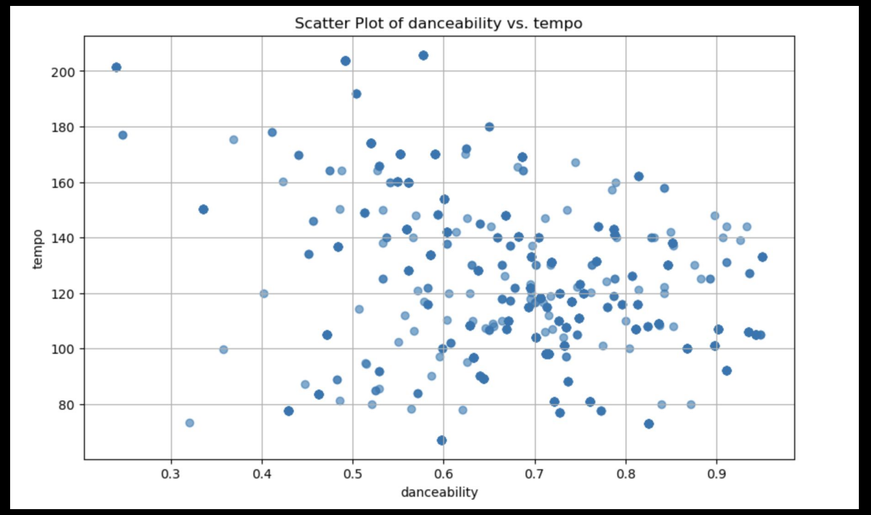
.pairplot of song characteristics allowed us to visualize a myriad relationships across audio features

data analysis

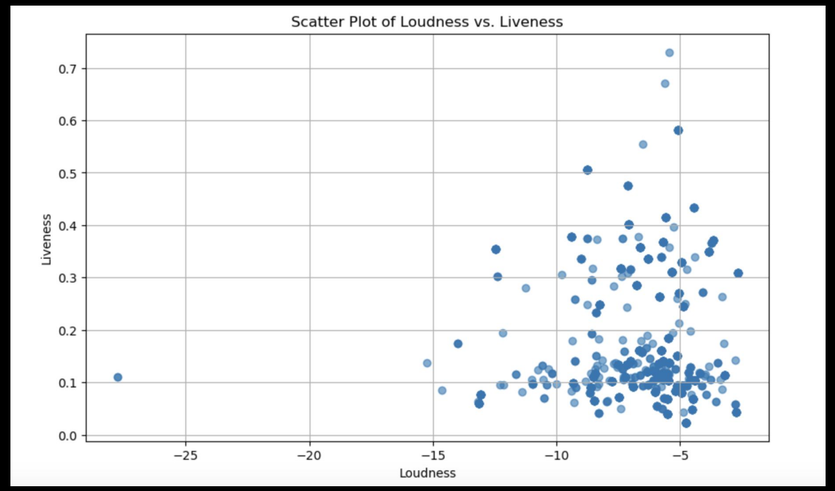
positive correlation:
energy and loudness
(0.632)

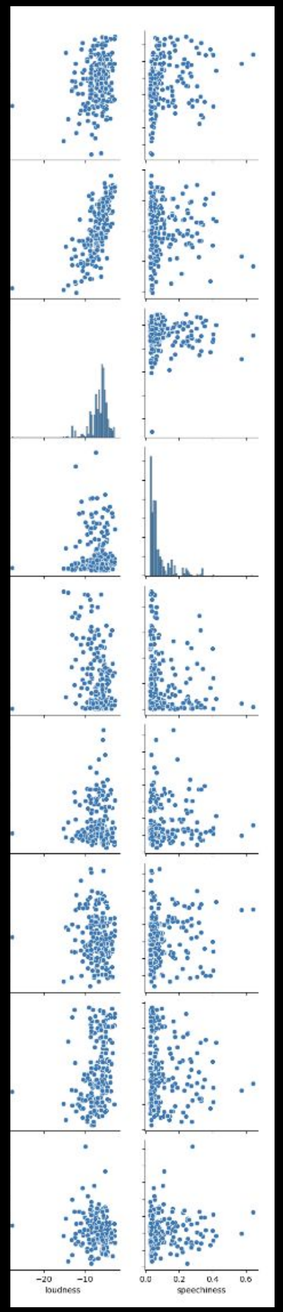


negative correlation:
danceability and tempo
(-0.398)



no correlation:
loudness and
liveness (0)





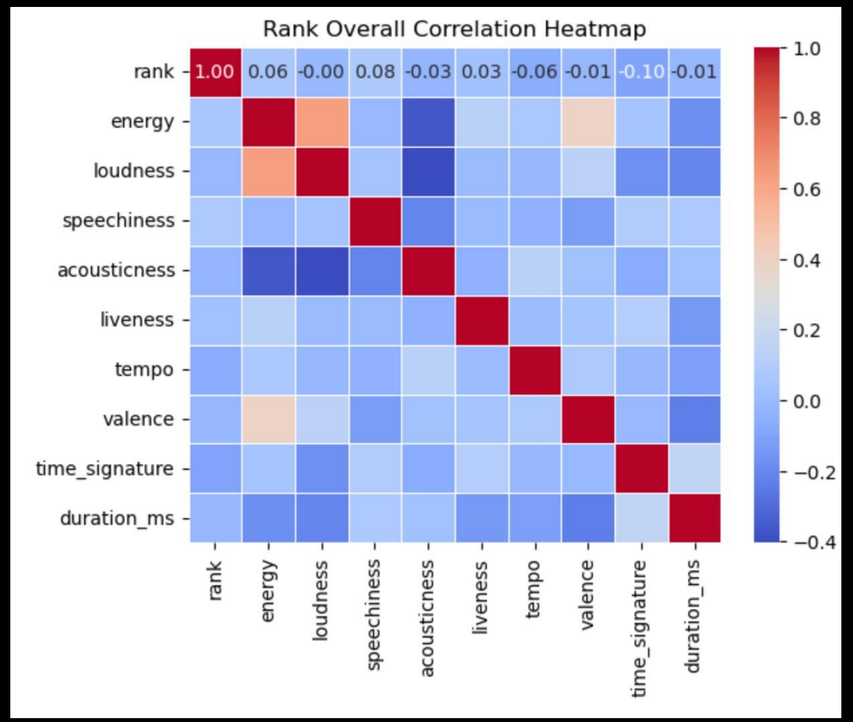
loudness - speechiness

data analysis

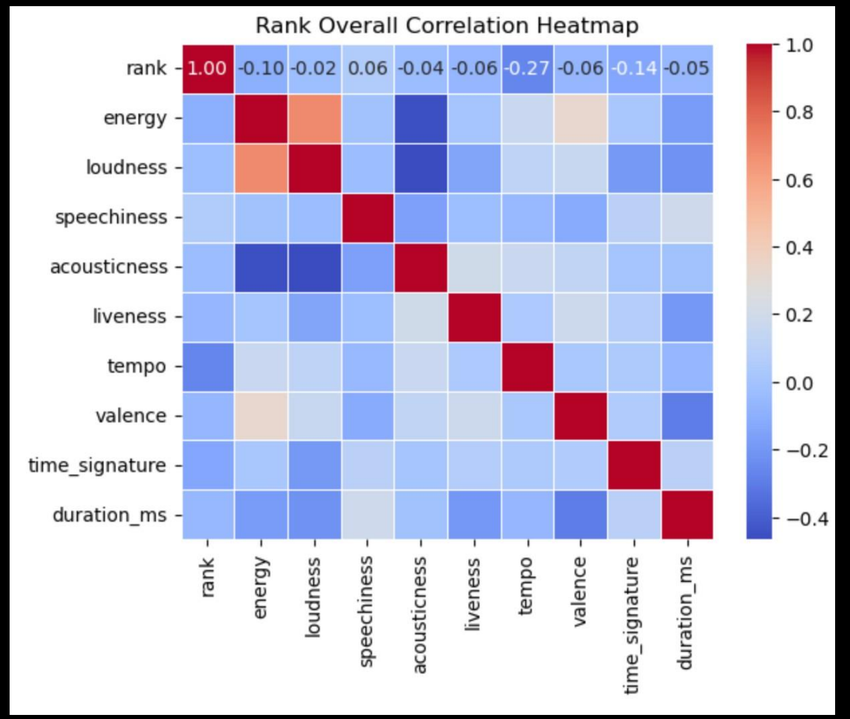
question #3

is there a correlation between the attributes of a song and its ranking on the charts?

top 10



top 3



conclusion

there is a relatively **high negative correlation between energy/tempo** and rank i.e. the **top songs generally have high energy** (as they have the lowest ranks) while not highly correlated, **top ranked songs are generally high in positivity** (identified by the negative correlation with valence = positivity) and **lower in speechiness** (positive correlation i.e. the lower the rank the lesser the speechiness)

defining mood

energy and **valence** is what we can use for mood. We added weights to the variables, it seems to be an intuitive way we can define happiness.

60% energy
intensity & pace?

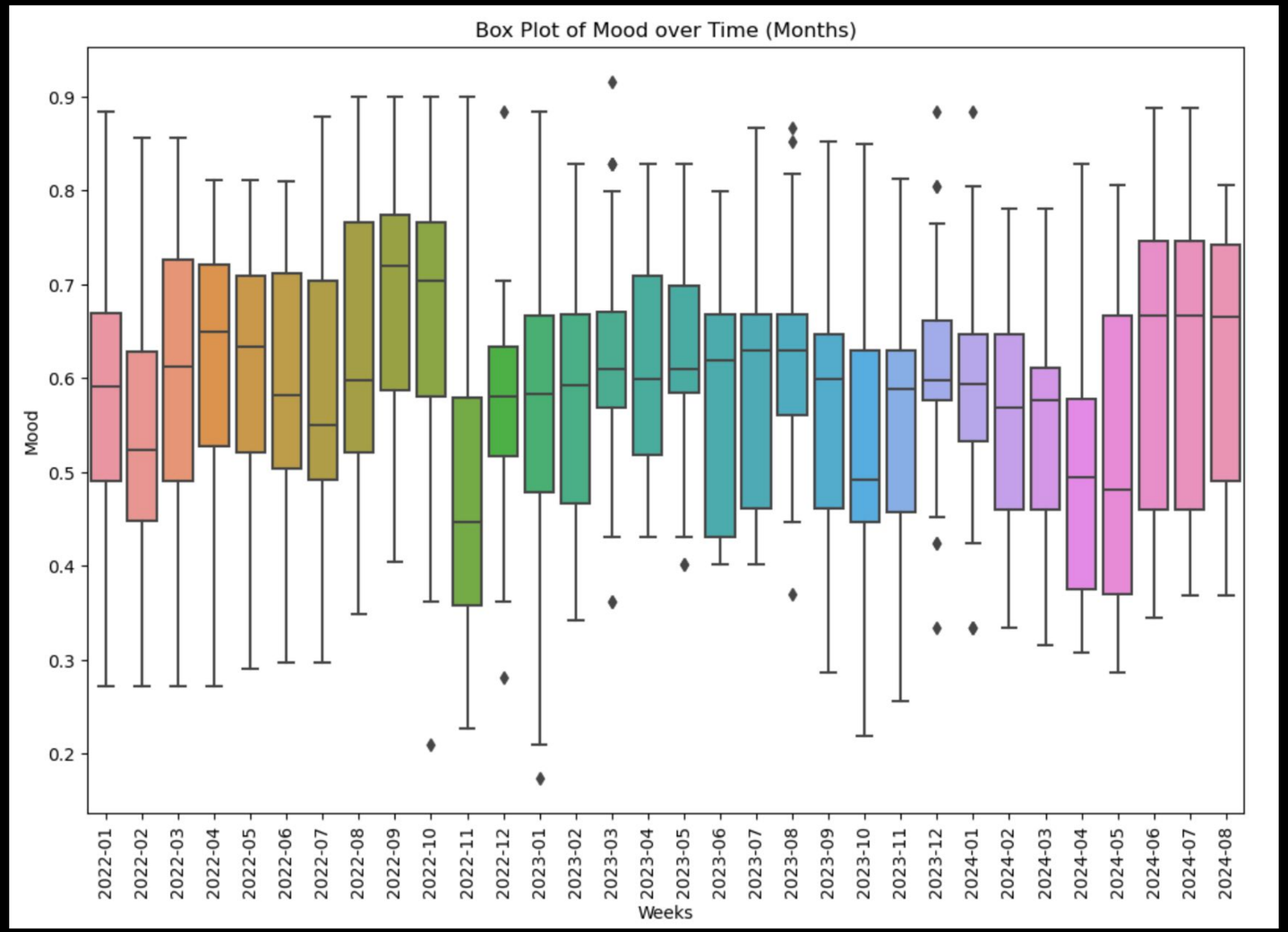
+

40% valence
positivity?

=

Mood

data analysis



question #4

what is the 'mood' of top ranked songs over time?

data fetching : events and risk

data set

Chicago Tribune
Daily Telegraph
Financial Times
The Globe and Mail
The Guardian
Los Angeles Times
The New York Times
USA Today
The Wall Street Journal
The Washington Post

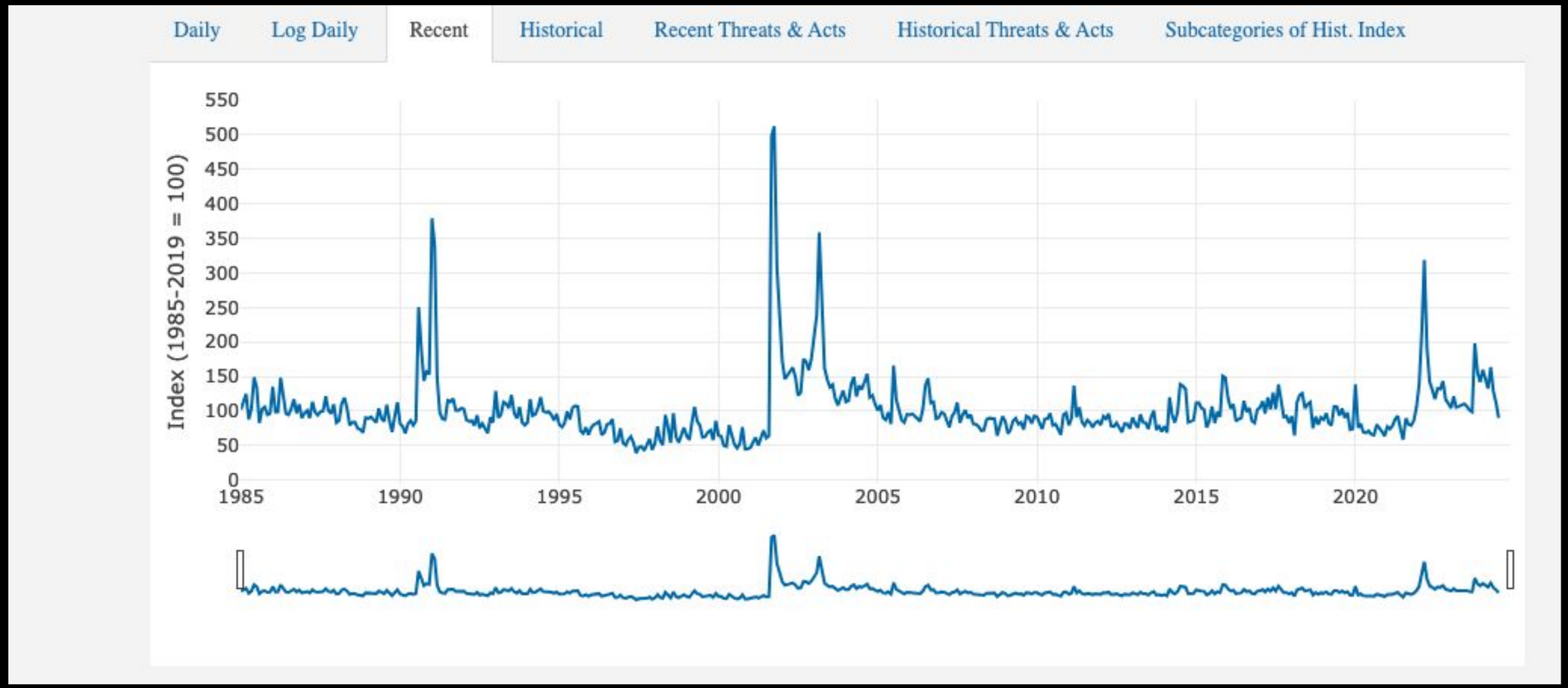
geopolitical risk (gpr) index

a measure of **adverse geopolitical events** and associated **risks** based on a tally of newspaper articles by Dario Caldara and Matteo Iacoviello

data source

geopolitical risk (gpr) index

measures how risk moves based on news and is an indicator of the nations mood



data source

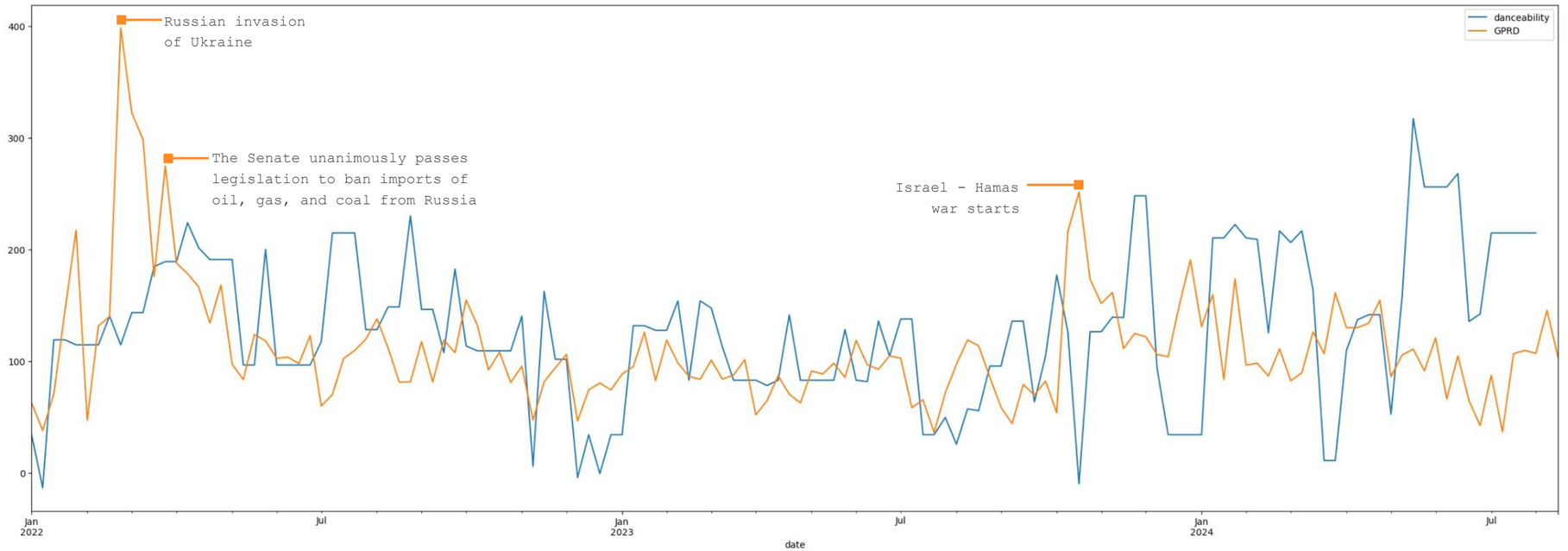
events

fetched and tagged 143 events from wikipedia
between **01-27-2022** and **08-06-2024**

	Date	Event	Category
0	2022-01-27	The Northeast experiences a major blizzard whi...	Weather
1	2022-02-01	February 2022 North American winter storm: A m...	Weather
2	2022-02-03	The share price of Meta falls by 26.4%, with F...	Economic
3	2022-02-24	Russian invasion of Ukraine	National Security
4	2022-02-24	The Dow Jones, Nasdaq, and S&P 500 fall sharpl...	Economic
...	...	...	...
138	2024-07-11	At least 11 people are killed in Texas and Lou...	Weather
139	2024-07-13	Trump is shot in an assassination attempt at a...	Political
140	2024-07-21	Vice President Kamala Harris launches her camp...	Political
141	2024-08-02	Kamala Harris is formally chosen as the Democr...	Political
142	2024-08-06	Vice President Harris announces her selection ...	Political

data analysis

```
danceability_group = top_3_dfs.groupby("chart_week")["danceability"].mean() * 950 - 500

fig = plt.figure(figsize=(30, 10))
danceability_group.plot()
gpr_df["GPRD"].plot()

plt.legend()
plt.show()
```

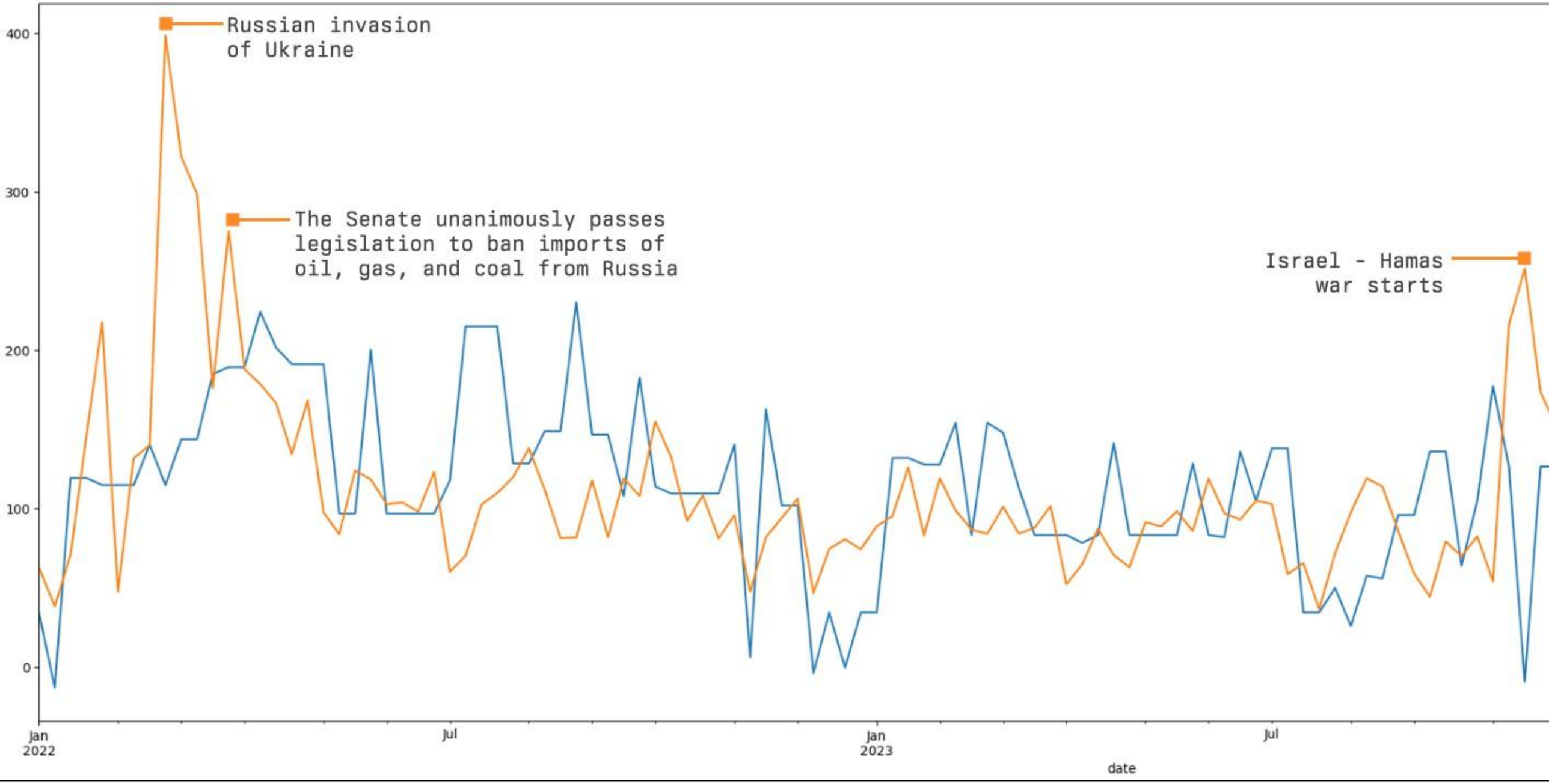


question #5

is there a direct relationship between the danceability of a top 3 song and the USA gpr based on events?

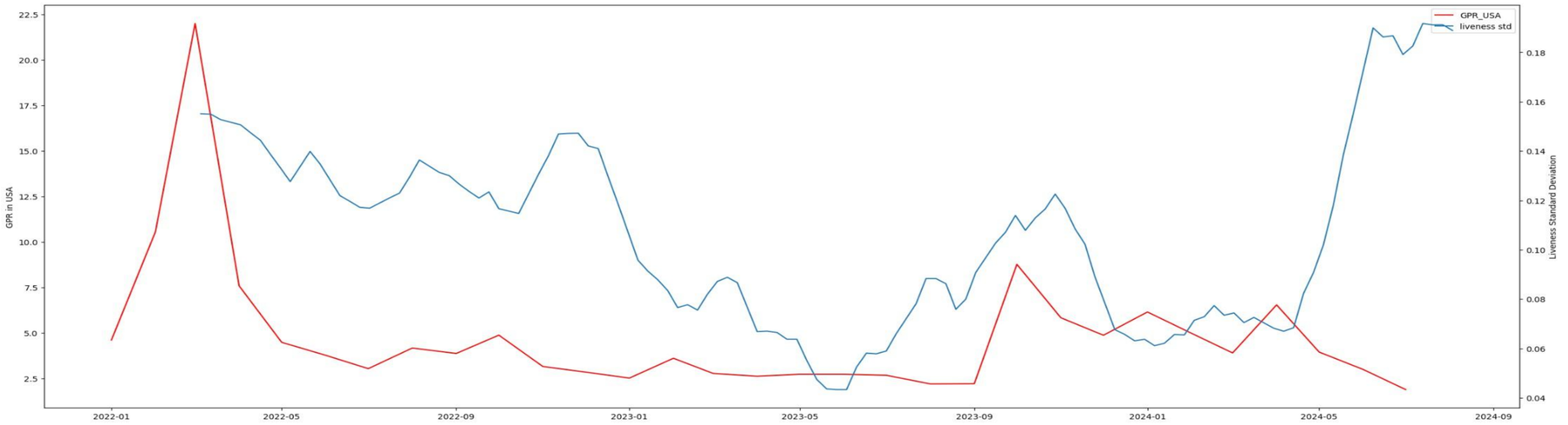
answer

A small number of really impactful events contrasted to the popularity of danceable music and, while we understand we are comparing apples and oranges, further analysis with more data could find a strong relationship between GPR and audio features like danceability.


Russian invasion of Ukraine
The Senate unanimously passes legislation to ban imports of oil, gas, and coal from Russia
Israel - Hamas war starts
400
300
200
100
0
Jan 2022
Jul
Jan 2023
Jul
date

data analysis

```
data = top_3_dfs.groupby("chart_week")["mood"].agg(["std"])

window_size = 10
smoothed_data = data.rolling(window=window_size).mean()

fig = plt.figure(figsize=(30,10))
ax1 = fig.add_subplot(111)
line_1 = ax1.plot(monthly_gpr_df.index, monthly_gpr_df["GPR_USA"], "-r", label="GPR_USA")
ax2 = ax1.twinx()
line_2 = ax2.plot(smoothed_data.index, smoothed_data["std"], label="liveness std")

lns = line_1+line_2
labs = [l.get_label() for l in lns]
ax1.legend(lns, labs, loc=0)

ax1.set_ylabel("GPR in USA")
ax2.set_ylabel("Liveness Standard Deviation")

plt.show()
```

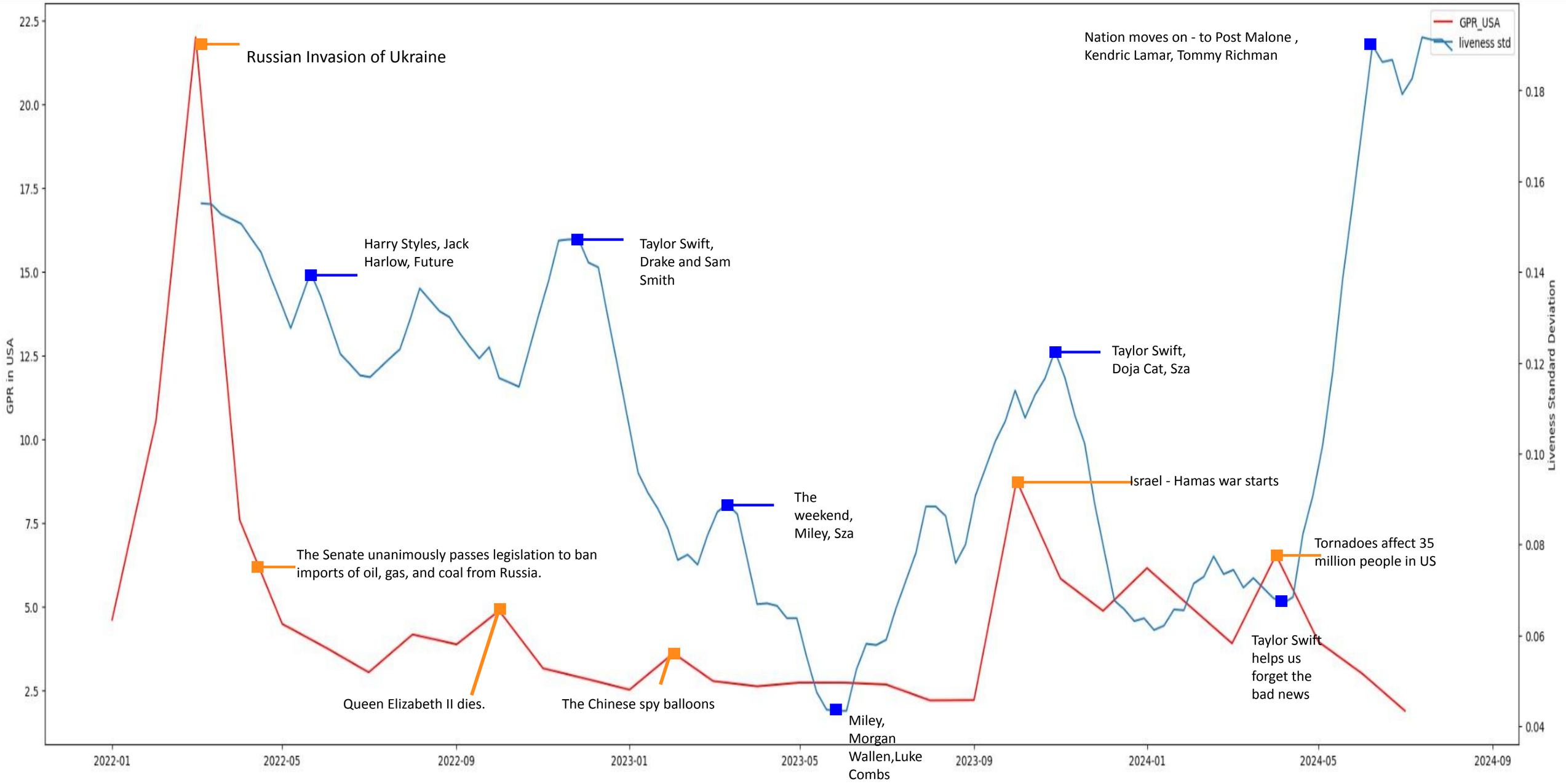


question #6

how does the standard deviation of the top 3 songs' mood compares to the gpr in the USA?

conclusion

The graph does show a downward trend in both the STD of the top 3's mood and the GPR_USA over time. Peaks in the GPR data seems to precede those in mood variability. This suggests a possible but very week relationship. The overall correlation between these two is not strong if there is any at all. This means there are more accurate factors that influence the variability in the top 3's mood


GPR_USA
liveness std
GPR in USA
Liveness Standard Deviation
Russian Invasion of Ukraine
Harry Styles, Jack Harlow, Future
Taylor Swift, Drake and Sam Smith
The weekend, Miley, Sza
Miley, Morgan Wallen,Luke Combs
Taylor Swift, Doja Cat, Sza
Israel - Hamas war starts
Nation moves on - to Post Malone , Kendric Lamar, Tommy Richman
Tornadoes affect 35 million people in US
Taylor Swift helps us forget the bad news
The Senate unanimously passes legislation to ban imports of oil, gas, and coal from Russia.
Queen Elizabeth II dies.
The Chinese spy balloons
2022-01
2022-05
2022-09
2023-01
2023-05
2023-09
2024-01
2024-05
2024-09

findings

do global events impact the mood of a nation?
can music tell us how?

looking for unexpected patterns

how **homogeneous** the **music taste** of a nation becomes when a popular performer stays on the charts for a long time

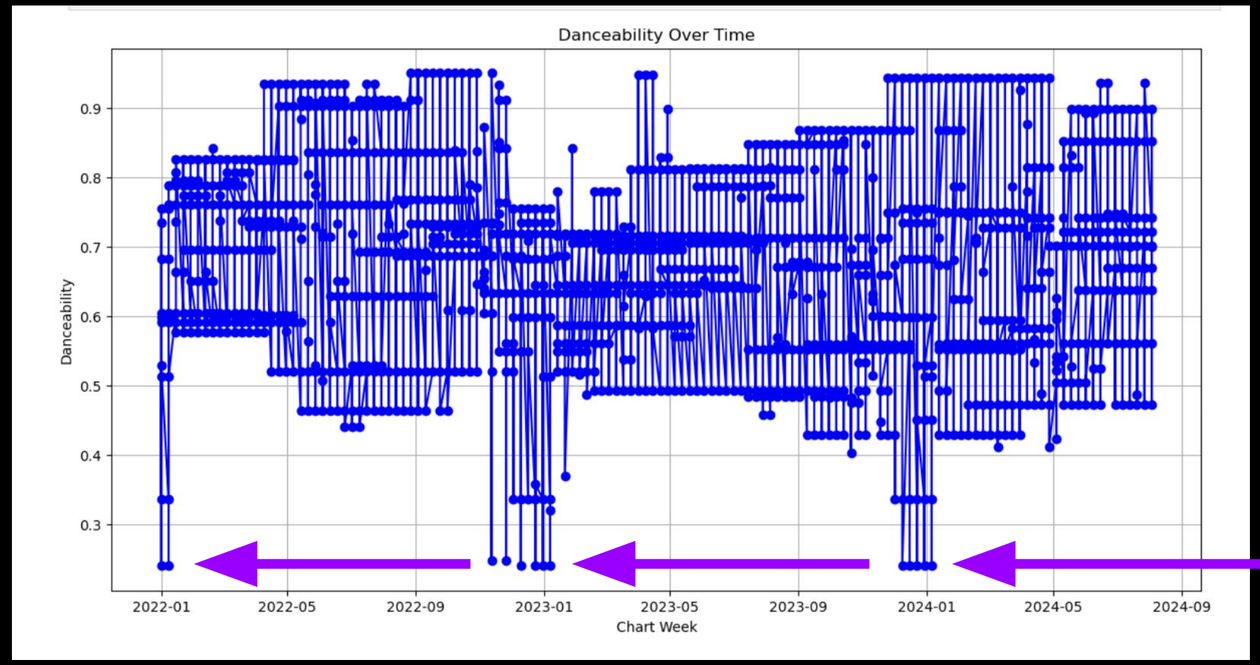
there are correlations between mood and gpr - but the peaks are delayed (**events happen** first, **music responds** later)

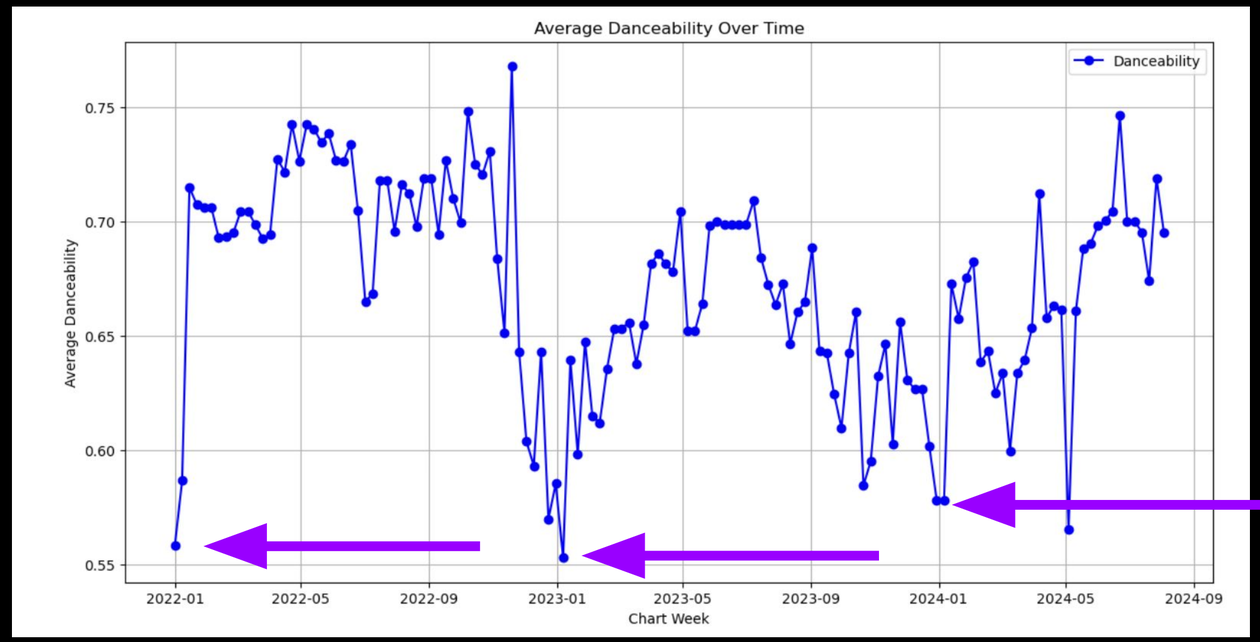
can we use music platforms such as spotify, apple music, and tidal to **understand the mood of a nation beyond the united states** - maybe?

are there seasonal mood trends? (recurring vs outliers) ie. **in the US people don't dance in December** (danceability decreases dec-jan every year)

how do relevant global/local events impact the mood and music listening habits?

looking for unexpected patterns





recurring event

but…

Skews could play a bigger role...

Spotify has 65 million active users in the USA compared to a total population of 259MM over the age of 18- hence Spotify represents ~25% of the adult population

Spotify data skews younger and more female than male. The mood index may have these inherent biases based on the data skews

'Popular Artists' and their planned release schedules could have a bigger influence on mood than global events

ultimately...

GPR and spotify 'mood' index are attempts to generalize the mood of the nation - both rely heavily on popular opinions, we need more data to strengthen the observations

future direction

what would make results more accurate over time?

add further **localization**
- breaking down of results by city
- include local, trustworthy news sources
- ingest events as they are happening

granular tagging
what kind of news impact musical preferences more than others?

expand the data fetching to other countries not just the US
- diversify the input by investigating music charts abroad