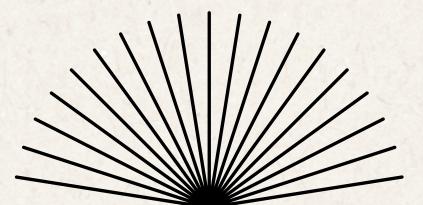


DECODING THE SOUNDTRACK OF SOCIETY

**Analyzing Cultural and Emotional Trends in Music Lyrics
Using NLP**

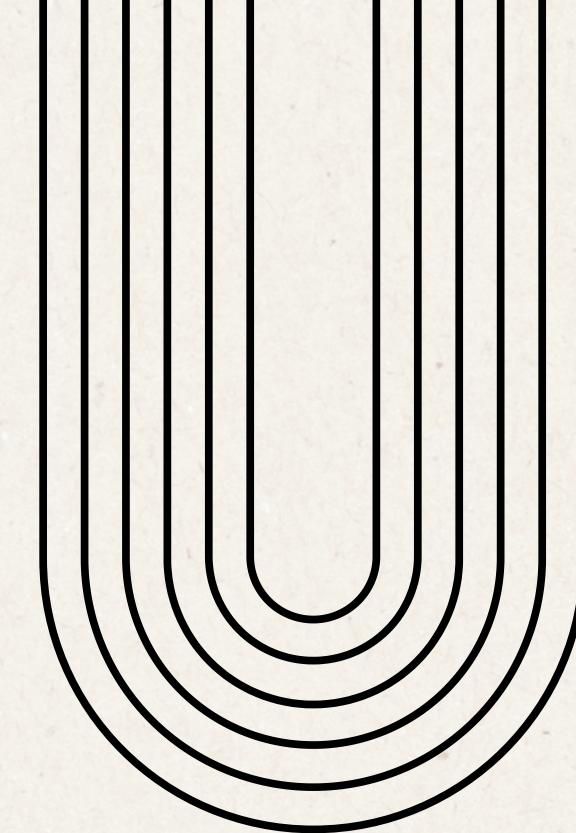
PRESENTED BY:

Prasanna Dengale
<https://www.linkedin.com/in/prasanna-dengale-07c>
<https://github.com/D-Prasanna>



Objectives

Clarifying the project's main overall objectives and goals.



Quantify Genre Trends



Decode Linguistic Shifts

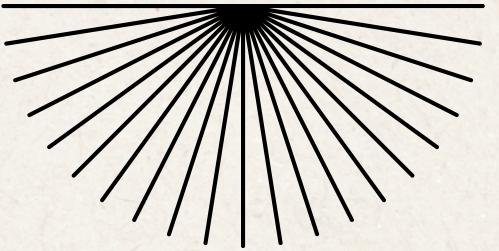


Highlight Genre Nuances

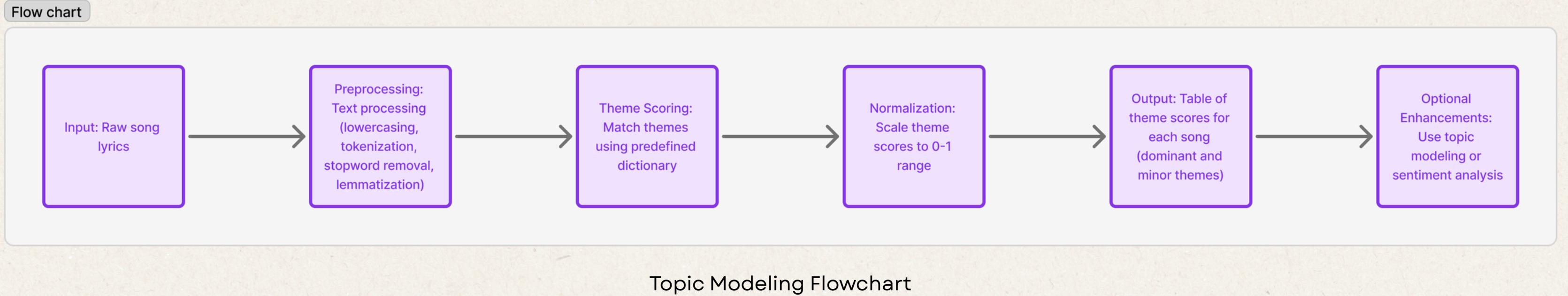


Connect Art to Historic Context

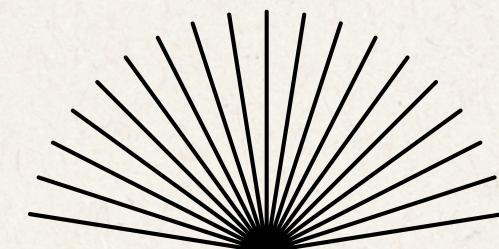
Cognitive computing technique of iterative hypothesis has been used.

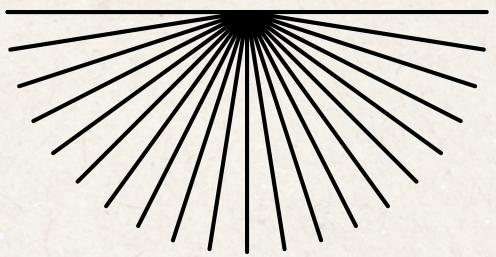


DATASET PREPARATION



Source: <https://data.mendeley.com/datasets/3t9vbwxgr5/3>





Key Features of the Dataset

1. Artist and Track Metadata

eg . Name of the artist

2. Lyrics and Text Features

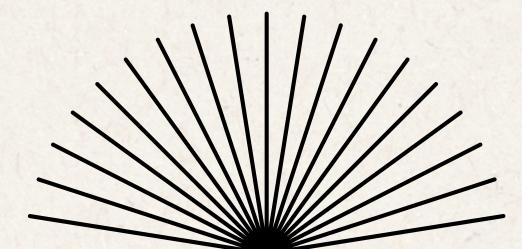
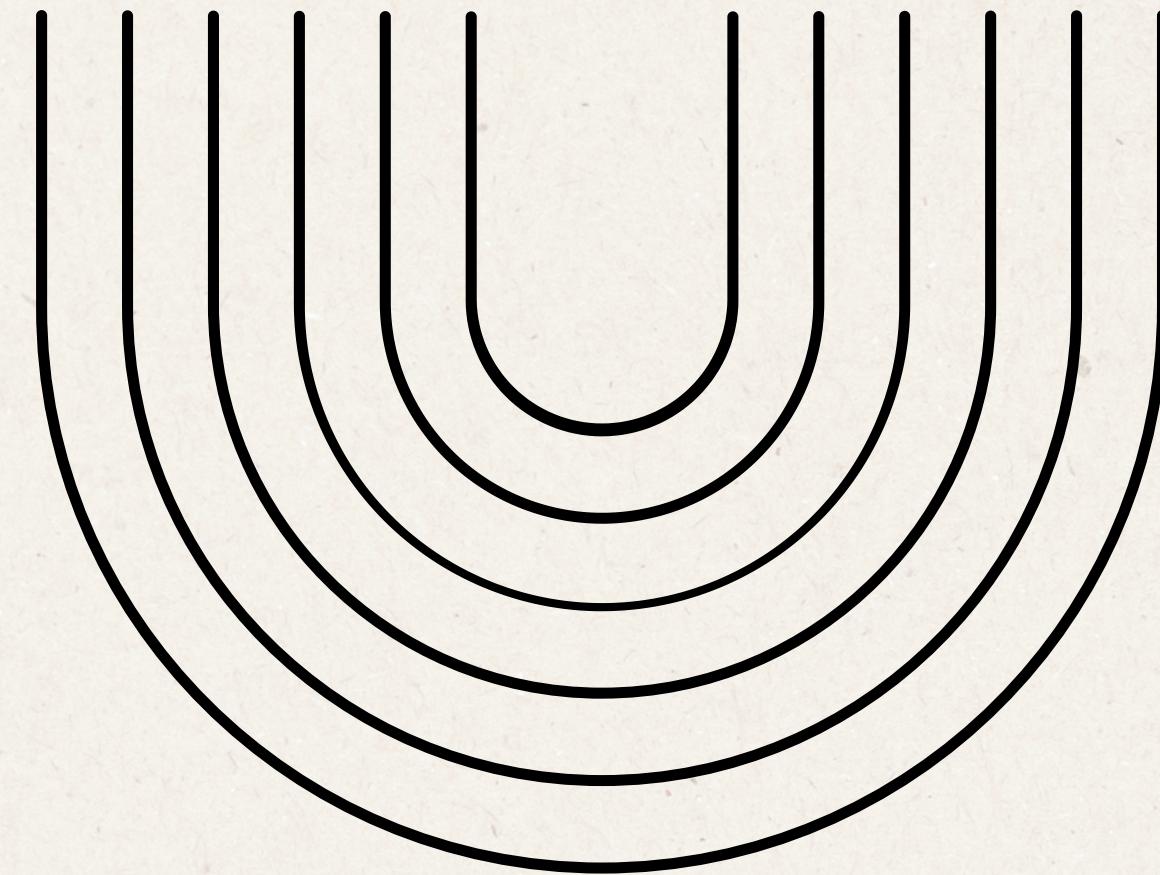
eg. Full lyrics of the song

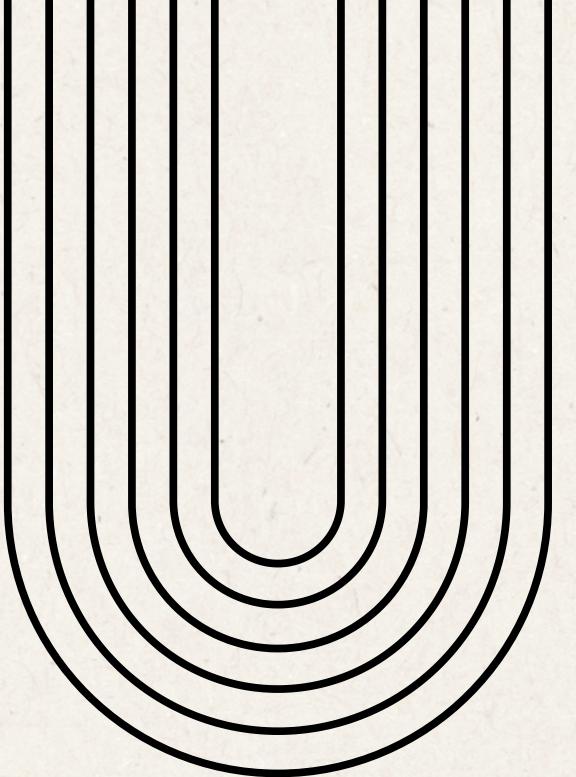
3. Thematic Scores

eg. Love and relationships,
Violence

4. Audio Features

eg .Positivity/musical happiness of the song





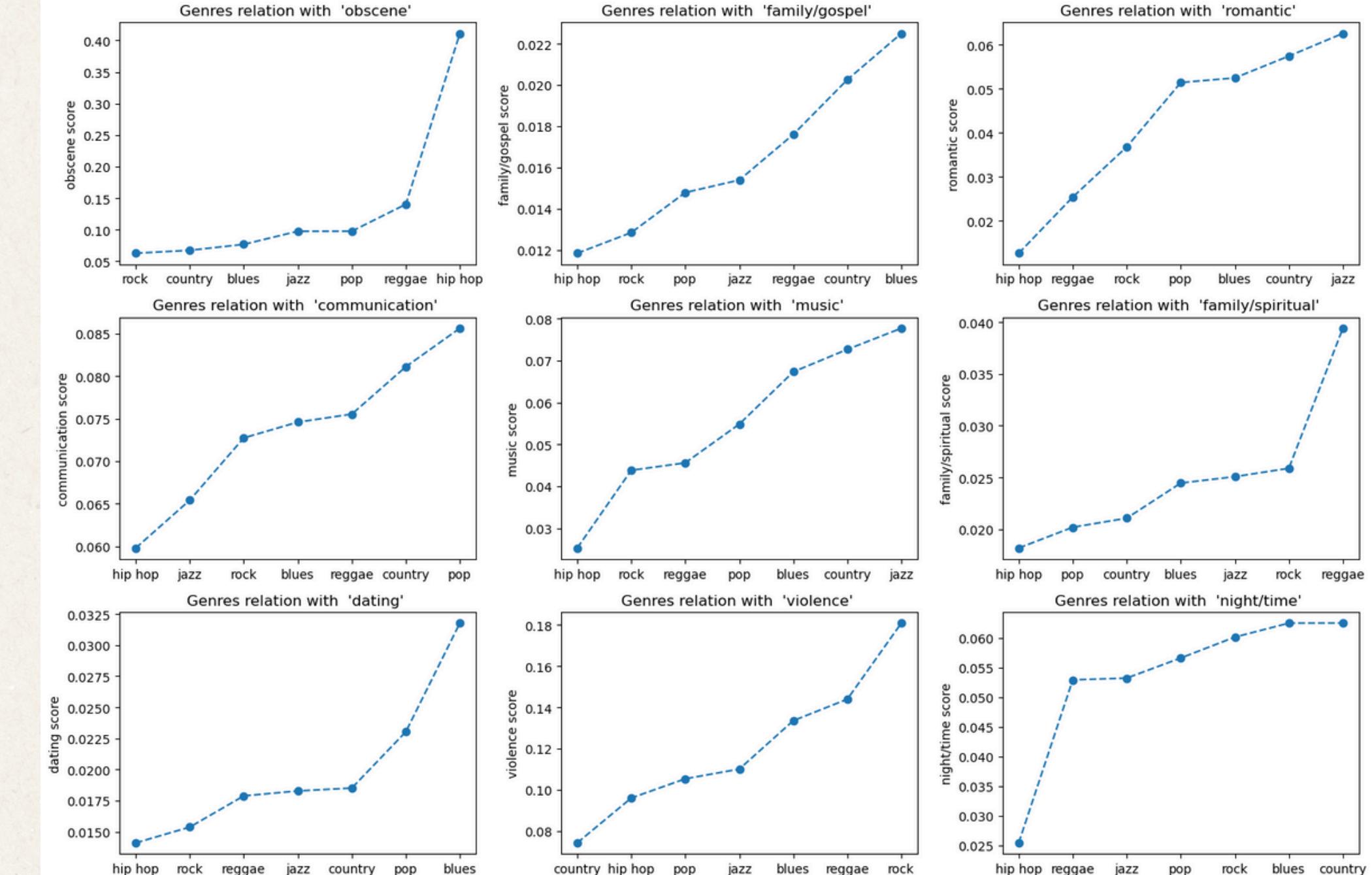
Hypothesis 1:

Top charts reflect the cultural and emotional stance of the society.

Iteration1 INFERENCE

Correlation analysis

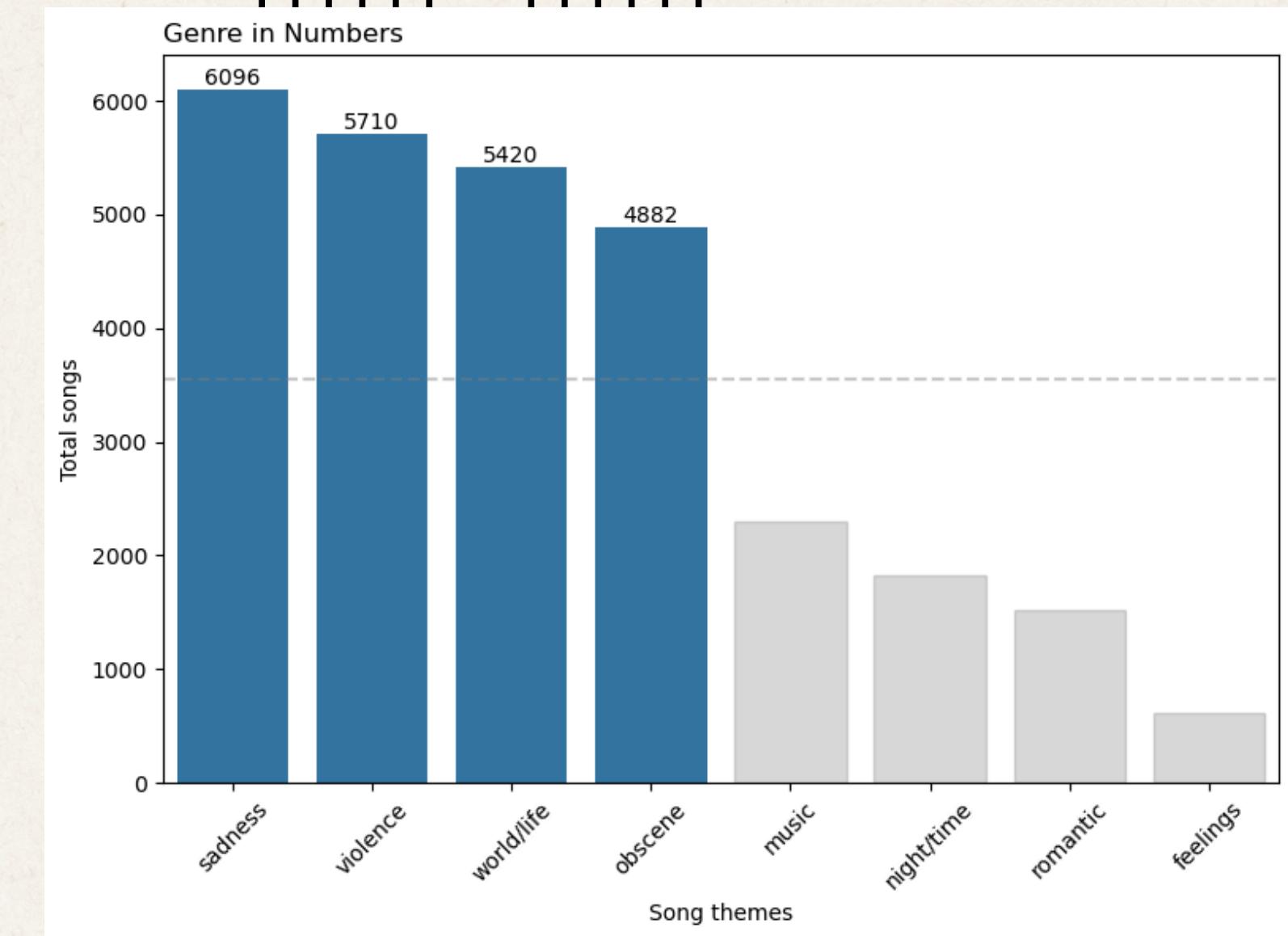
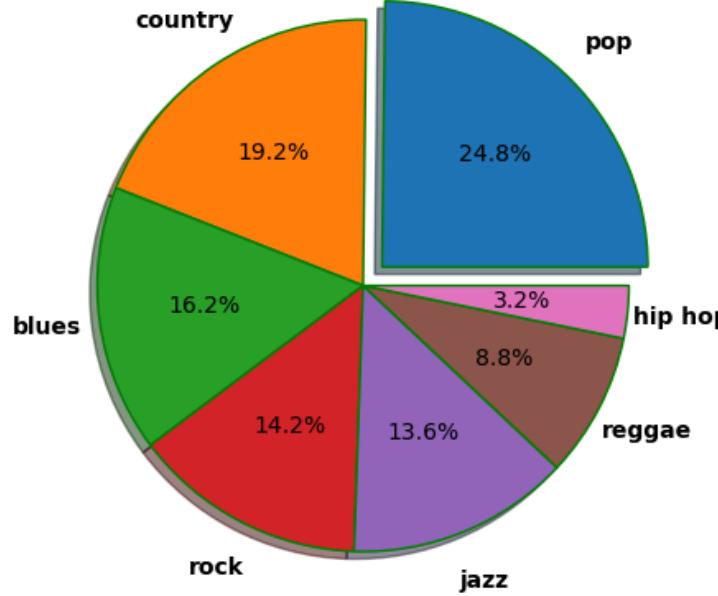
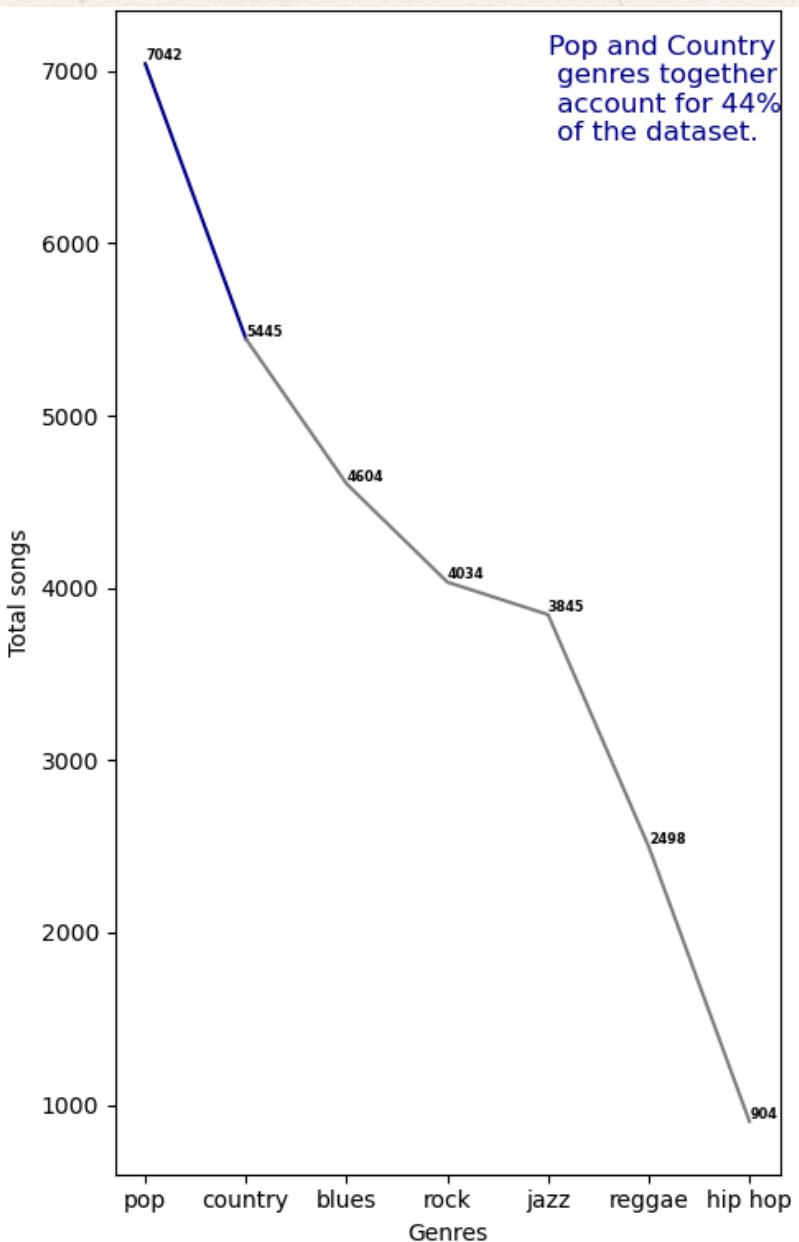
Genre associations reveal thematic contrasts rooted in cultural and historical contexts. **Hip-hop** leads in "**Obscene**" themes due to its raw reflection of urban struggles, while scoring lowest in "**Family/Gospel**," "**Romantic**," highlighting its secular, self-expressive focus. **Country** consistently ranks high in themes like "**Family**," "**Romantic**," "**Dating**," and "**Night/Time**," echoing its narrative, rural origins. **Jazz** dominates "**Romantic**" and "**Music**," reflecting its emotional and musical sophistication. **Reggae** excels in "**Spiritual**" and "**Violence**" themes—combining spirituality with political resistance. **Pop** scores highest on "**Communication**," aligning with its global, mainstream appeal. **Rock** uses violent imagery metaphorically, drawing from countercultural roots. **Blues**, like **country**, leans heavily on emotive storytelling, especially in "**Dating**" and "**Night/Time**."



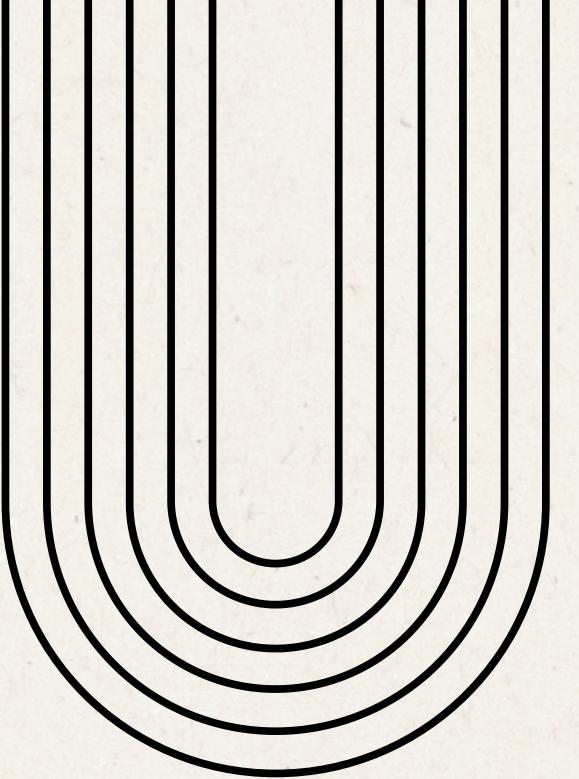
Overall, urban genres (hip-hop, reggae) foreground societal expression, while rural and traditional genres (country, blues) emphasize personal and relational themes.

Iteration 1 INFERENCE

Genre distribution



Pop and Country genres constitute 44% of the analyzed corpus. Pop emphasizes catchy, repetitive, and colloquial lyrics for broad appeal, while Country favors narrative structures and vivid, grounded imagery. Historically, Country thrived regionally (1950s–70s), while Pop globalized from the 1980s onward, with both genres diversifying and expanding their reach.



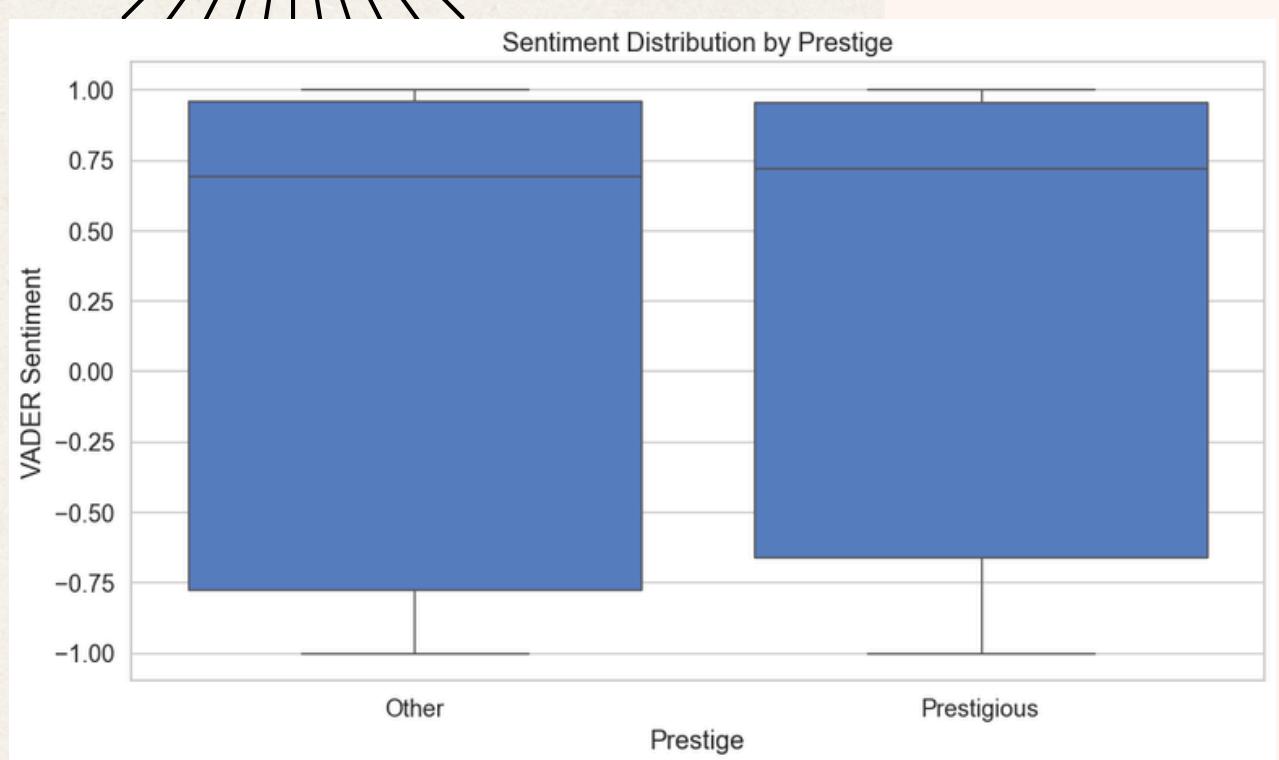
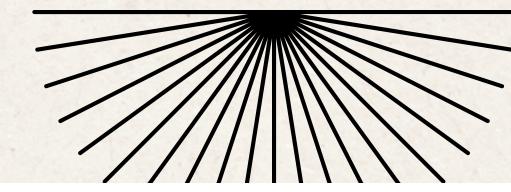
Hypothesis 2:

Top charts are influenced by bias based on audience's bias: a linguistic approach.

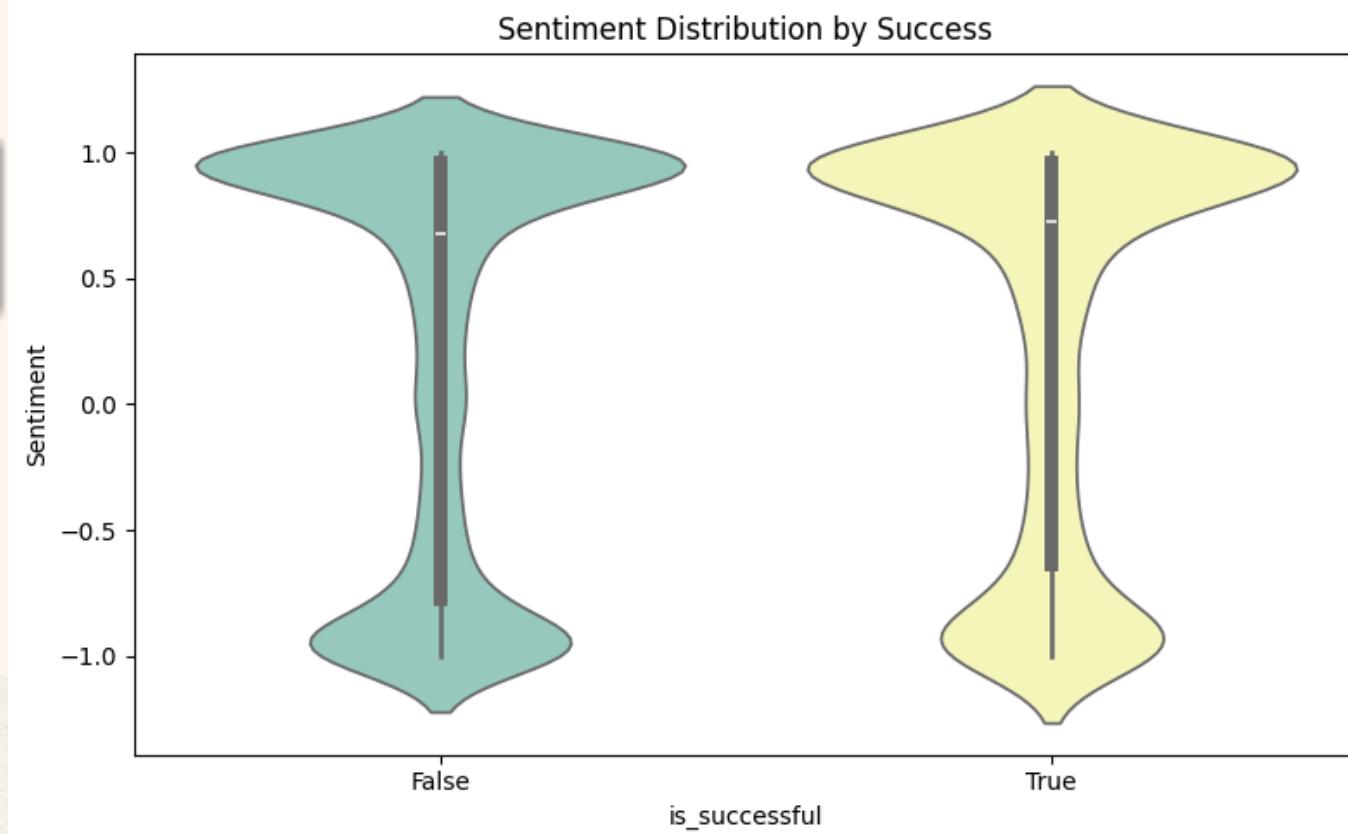
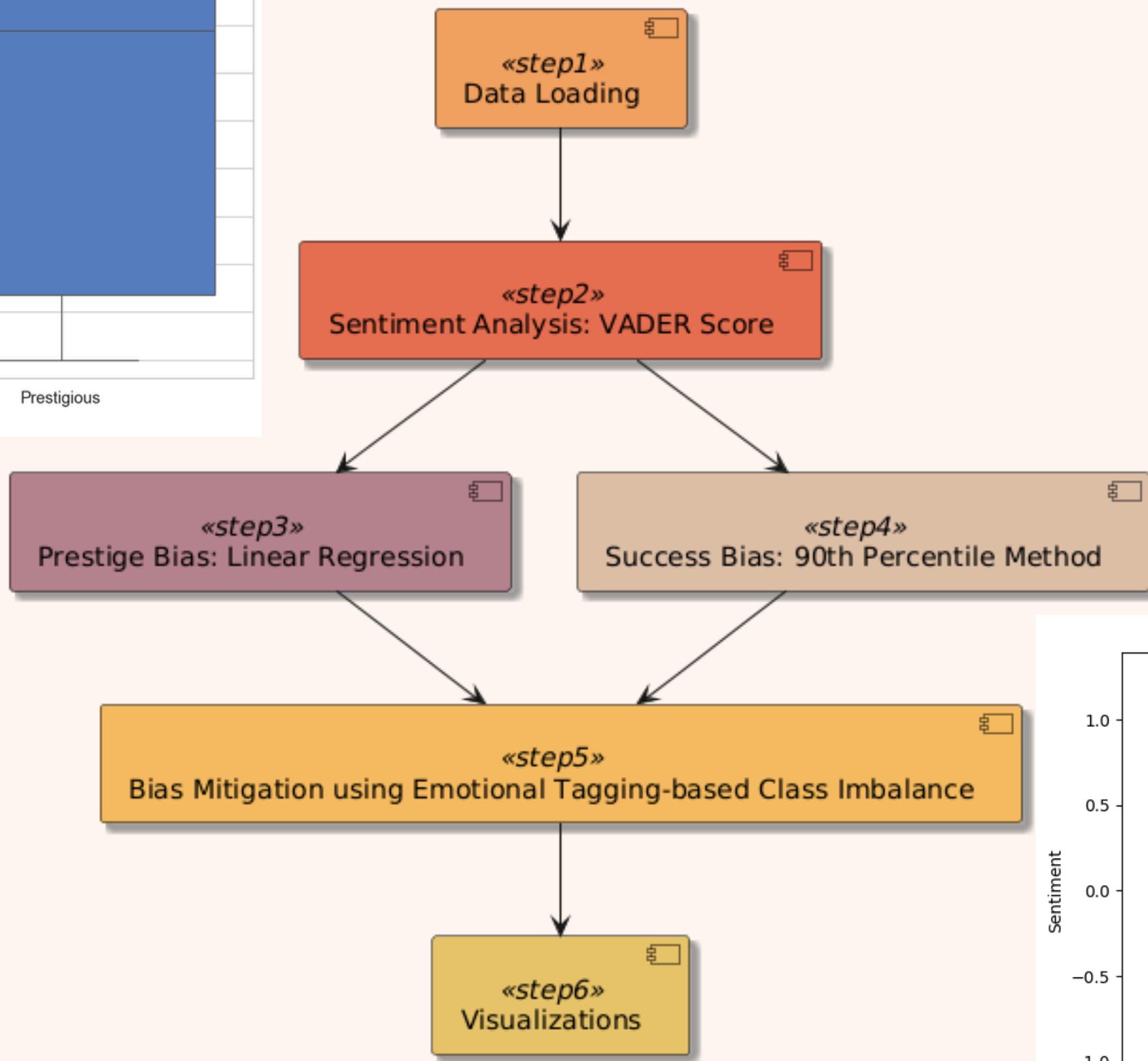
Prestige bias: best-selling artists top the charts

Success bias: best-selling songs are copied

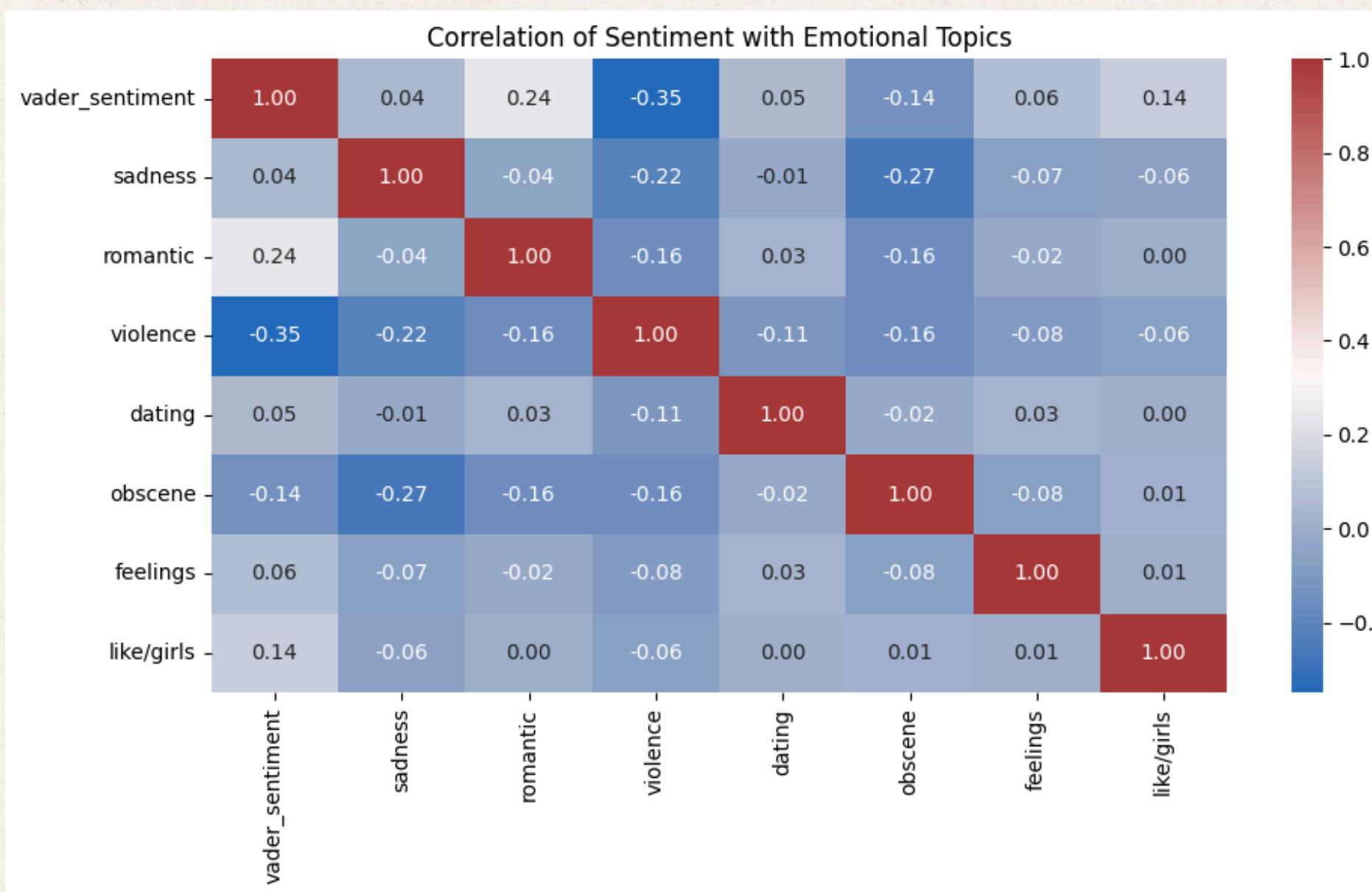
Bias Detection Module



NLP Bias Detection & Mitigation Flow

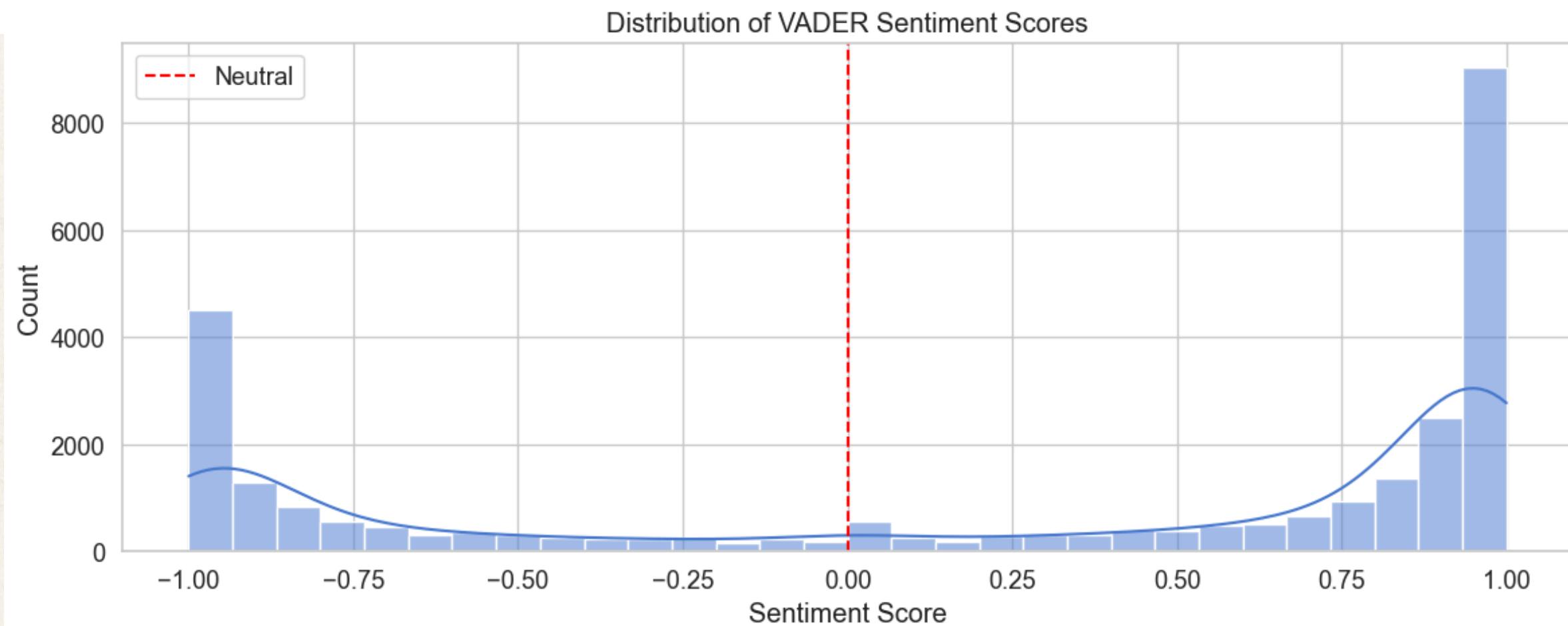


Iteration 2 INFERENCE

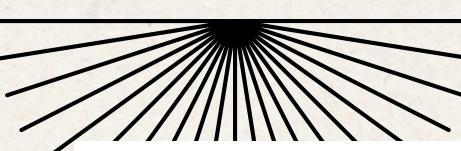


On mitigating biases, we can explain the two peaks in preferences:

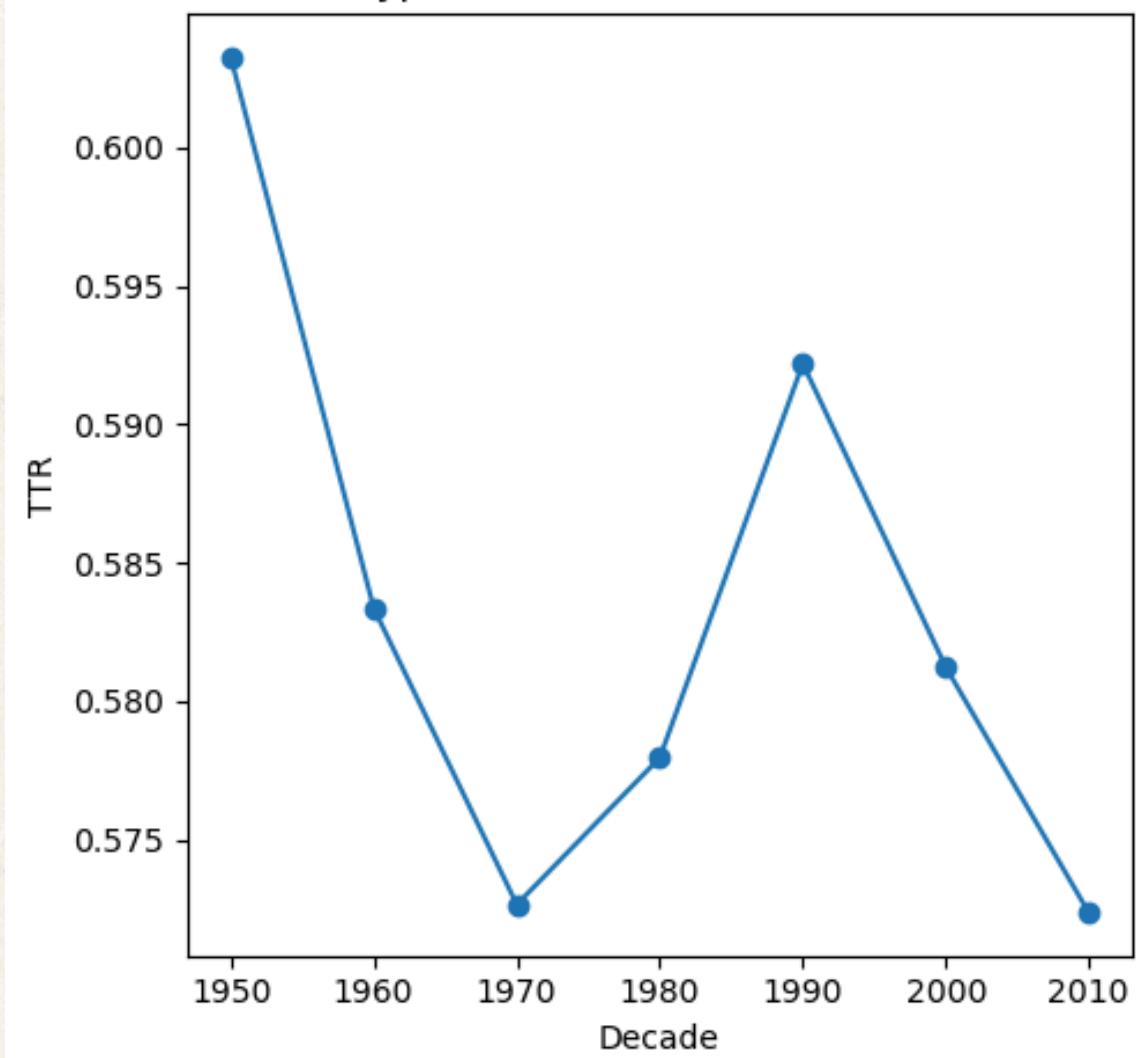
- Popularity influencing the positive sentiment scores.** Thus, people enjoy pop songs simply because of their popularity and rich presence
- Negative sentiment ≠ unpopular – in fact, negative or emotionally intense themes may be a hallmark of impactful music.**



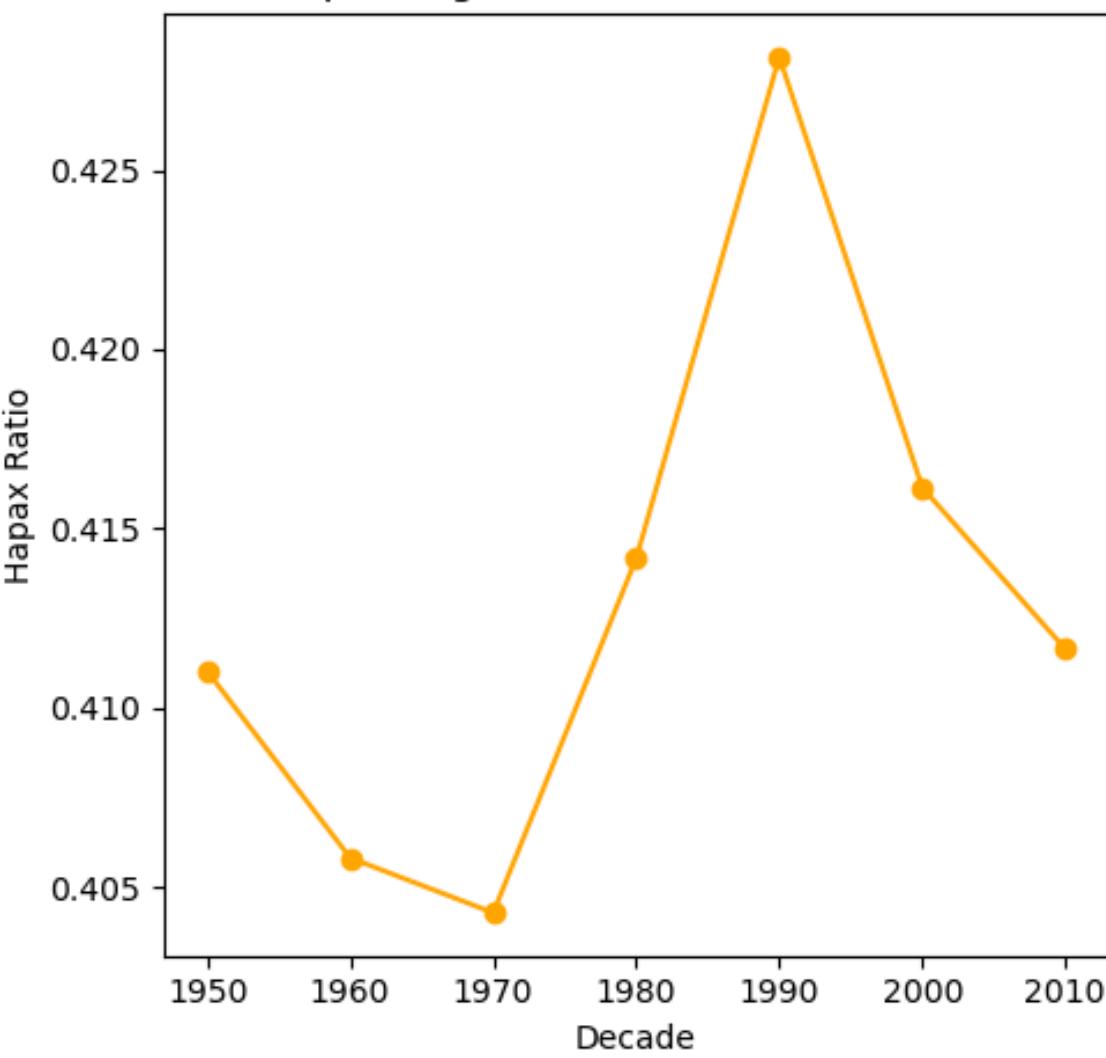
We Also found that...



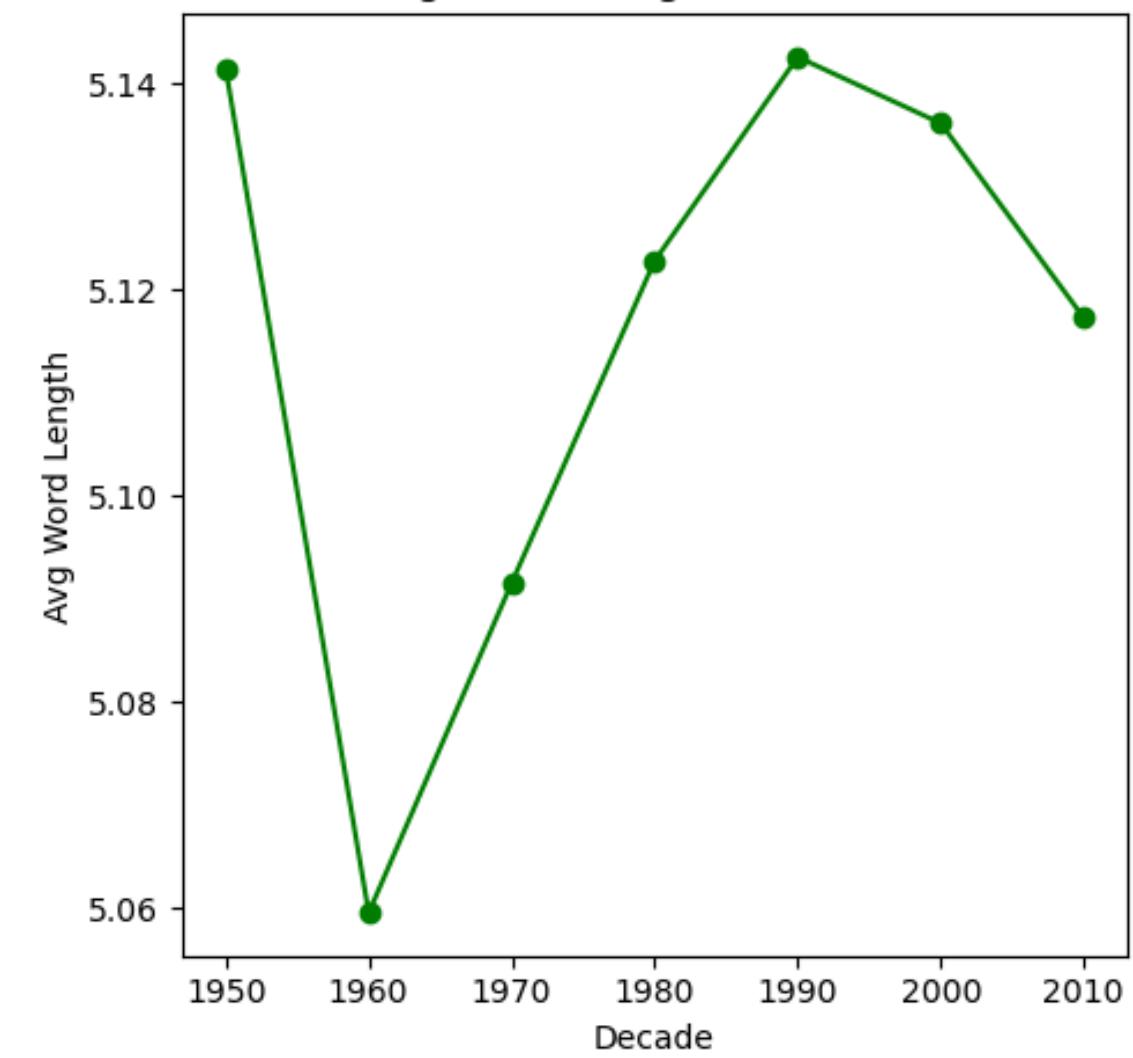
Type-Token Ratio Over Decades



Hapax Legomena Ratio Over Decades



Average Word Length Over Decades



1. Type-Token Ratio (TTR) Over Decades

TTR measures lexical diversity: high TTR indicates varied vocabulary, low TTR suggests repetition. In the 1950s-60s, TTR was moderate to high with poetic, metaphorical lyrics. In the 1970s-80s, TTR dipped due to simpler, rhythmic genres like disco and punk. From the 1990s onwards, TTR rose with genres like indie rock and hip hop focusing on narrative and diverse vocabulary.

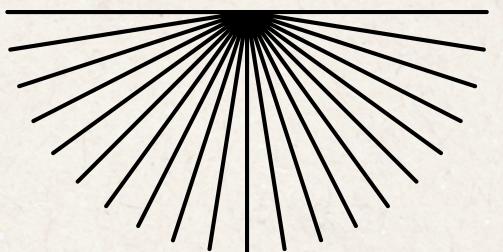
2. Hapax Legomena Ratio Over Decades

Hapax Legomena Ratio measures unique words used once, with high ratios indicating originality. Mid-century music had low ratios due to repetitive chorus structures. Modern genres like indie and rap have higher ratios, reflecting a shift towards lyrical originality and novelty in the digital era.

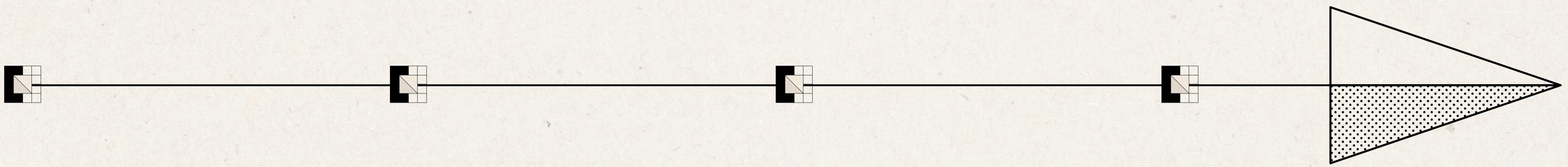
3. Average Word Length Over Decades

Shorter words in earlier decades indicated simpler, emotional language, while longer words in the 90s-2000s reflected complex storytelling in hip hop and indie rock. Post-2010,





Future Scope



Broaden Definition of "Prestige"

Move beyond chart frequency to include Awards, Critical Acclaim, etc

Time-Based Bias Evolution

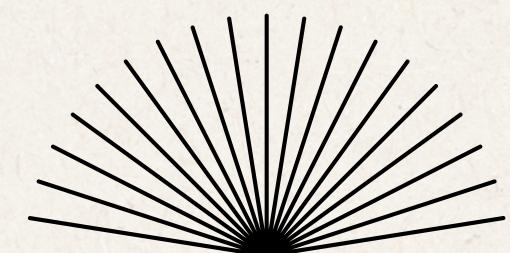
Study how prestige/sentiment bias has evolved by decade

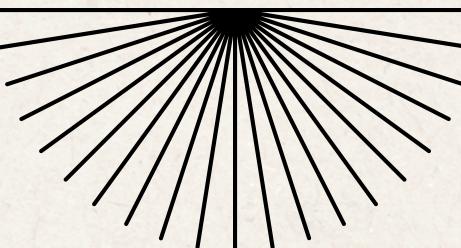
Create Predictive Models

Supporting with real-time data

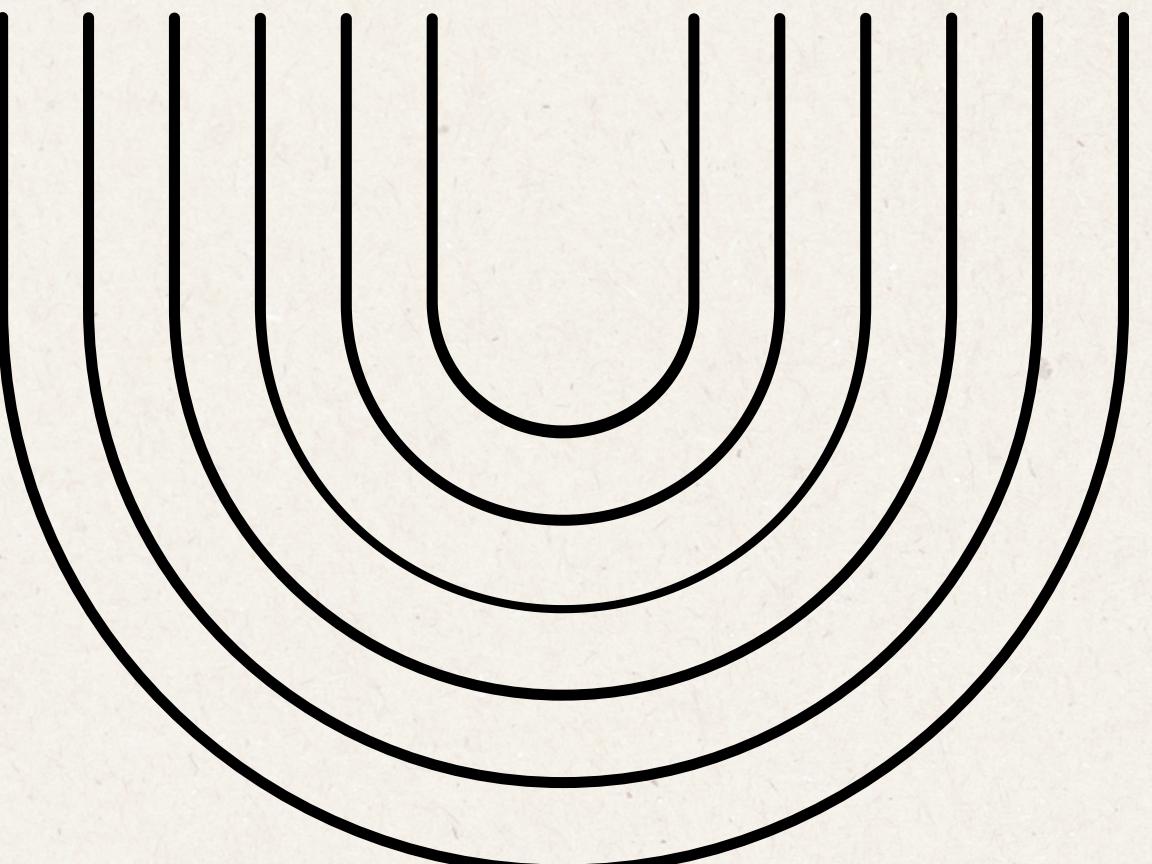
Language Inclusivity

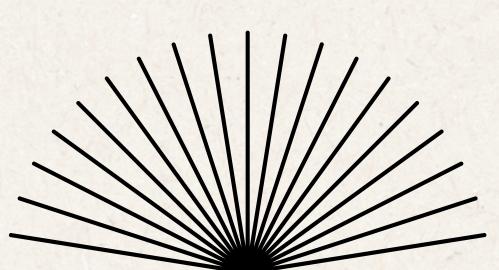
Including multilingual Charts, needs various models trained on native languages

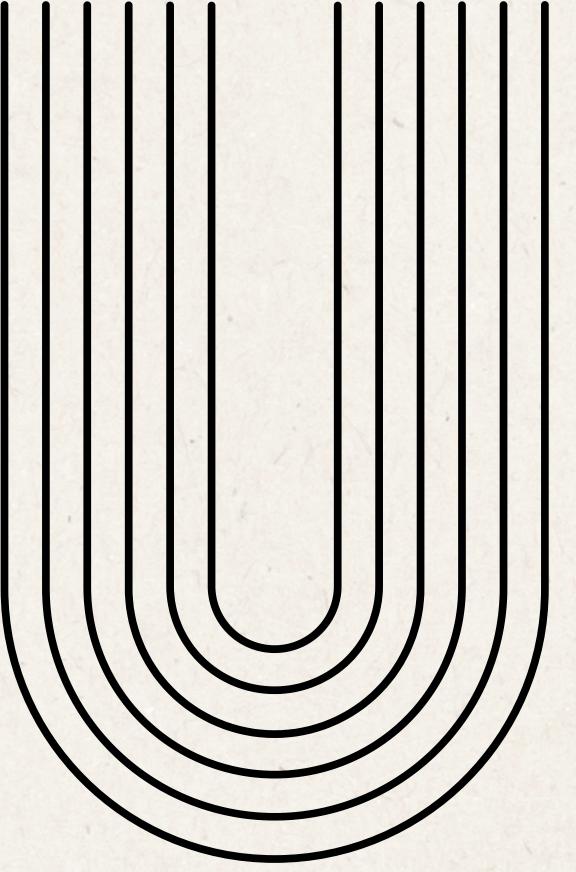
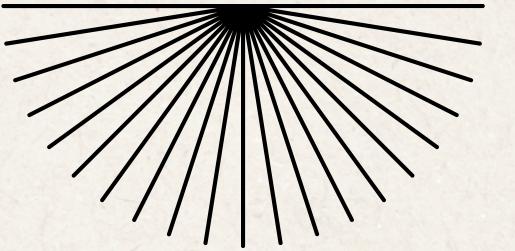




References



- [1] C. O. Brand, A. Acerbi, and A. Mesoudi, "Cultural evolution of emotional expression in 50 years of song lyrics," *Evolutionary Human Sciences*, vol. 1, e11, Nov. 2019. [Online]. Available: [eg_. Name of the artistPubMed](#)
- [2] K. Watanabe and M. Goto, "Lyrics Information Processing: Analysis, Generation, and Applications," in *Proceedings of the 1st Workshop on NLP for Music and Audio (NLP4MusA)*, Online, Oct. 2020, pp. 6–12. [Online]. Available: [eg_. Name of the artistACL Anthology](#)
- [3] M. R. Ghassemi, M. A. Heidari, and M. M. Ebrahimi, "Automatic Detection of Sentiment and Theme of English and Ukrainian Song Lyrics," in *2020 IEEE 14th International Conference on Application of Information and Communication Technologies (AICT)*, Tashkent, Uzbekistan, Oct. 2020, pp. 1–5. [Online]. Available: [eg_. Name of the artist](#)
- [4] M. A. Heidari, M. R. Ghassemi, and M. M. Ebrahimi, "Automatic Detection of Sentiment and Theme of English and Ukrainian Song Lyrics," in *2020 IEEE 14th International Conference on Application of Information and Communication Technologies (AICT)*, Tashkent, Uzbekistan, Oct. 2020, pp. 1–5. [Online]. Available: [eg_. Name of the artist](#)
- [5] S. Primack, M. N. Shensa, and J. Sidani, "Trends in Positive, Negative, and Neutral Themes of Popular Music From 1998 to 2018: Observational Study," *JMIR Pediatrics and Parenting*, vol. 4, no. 2, e26475, 2021. [Online]. Available: [eg_. Name of the artist](#)
- 



Thank you!

