



THE POWER OF LYRICS

ANALYZING THE
RELATIONSHIP BETWEEN
LYRICAL COMPLEXITY AND
SONG POPULARITY

SARAH DIAS BARRETO
NEHA CHEDE
ANGEL GEORGE

INTRODUCTION: THE EVOLVING MUSIC LANDSCAPE

- Rising Investments in Music
 - Annual surge of investments (~\$4.1 billion) by record labels due to increased accessibility
 - Exponential growth in available music tracks demands efficient content retrieval methods
- Music Information Retrieval (MIR)
 - Machine learning field analyzing music attributes (e.g., emotions, instruments, genres)
 - Enables predictions of songs with shared characteristics

BEYOND BEATS AND MELODY: THE NEED FOR LYRICAL ANALYSIS IN POPULARITY PREDICTIONS

■ Past Research

- Predicted music popularity using acoustic features, metadata, and temporal data
- Classifiers such as Logistic Regression (LR) and Naive Bayes (NBC) demonstrated mixed performance
- Studies achieved accuracy as high as 99% but lacked diverse performance metrics
- Limitations
 - Minimal focus on correlating lyrical complexity with popularity and overemphasis on acoustic features, neglecting lyrical content

■ Research Gap

- Focus on Lyrics - While musical features like melody, beat, and tempo are extensively studied, lyrical analysis remains underexplored
- **Hypothesis** - Listeners may subconsciously prefer songs with emotionally engaging or complex lyrics, influencing engagement levels and repeat listens

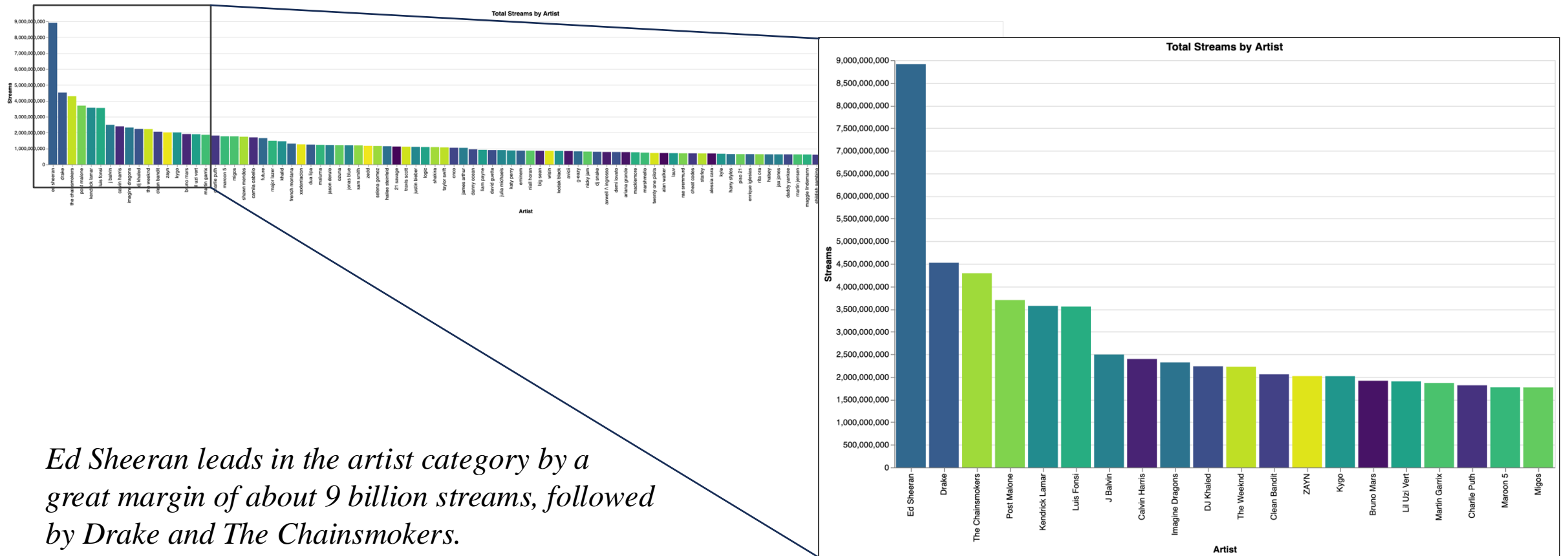
FOCUS AND OBJECTIVE

- Focus of this Study
 - Investigate whether **lyrical complexity** impacts song popularity on streaming platforms, with a focus on Spotify
 - Addresses a gap in research:
 - Prior studies explored the influence of genres, tempo, and artist popularity
 - **Lyrical depth and listener engagement** remain underexplored
- Objective
 - Explore the relationship between **lyrical complexity** and **song popularity** across 60,000 Spotify songs

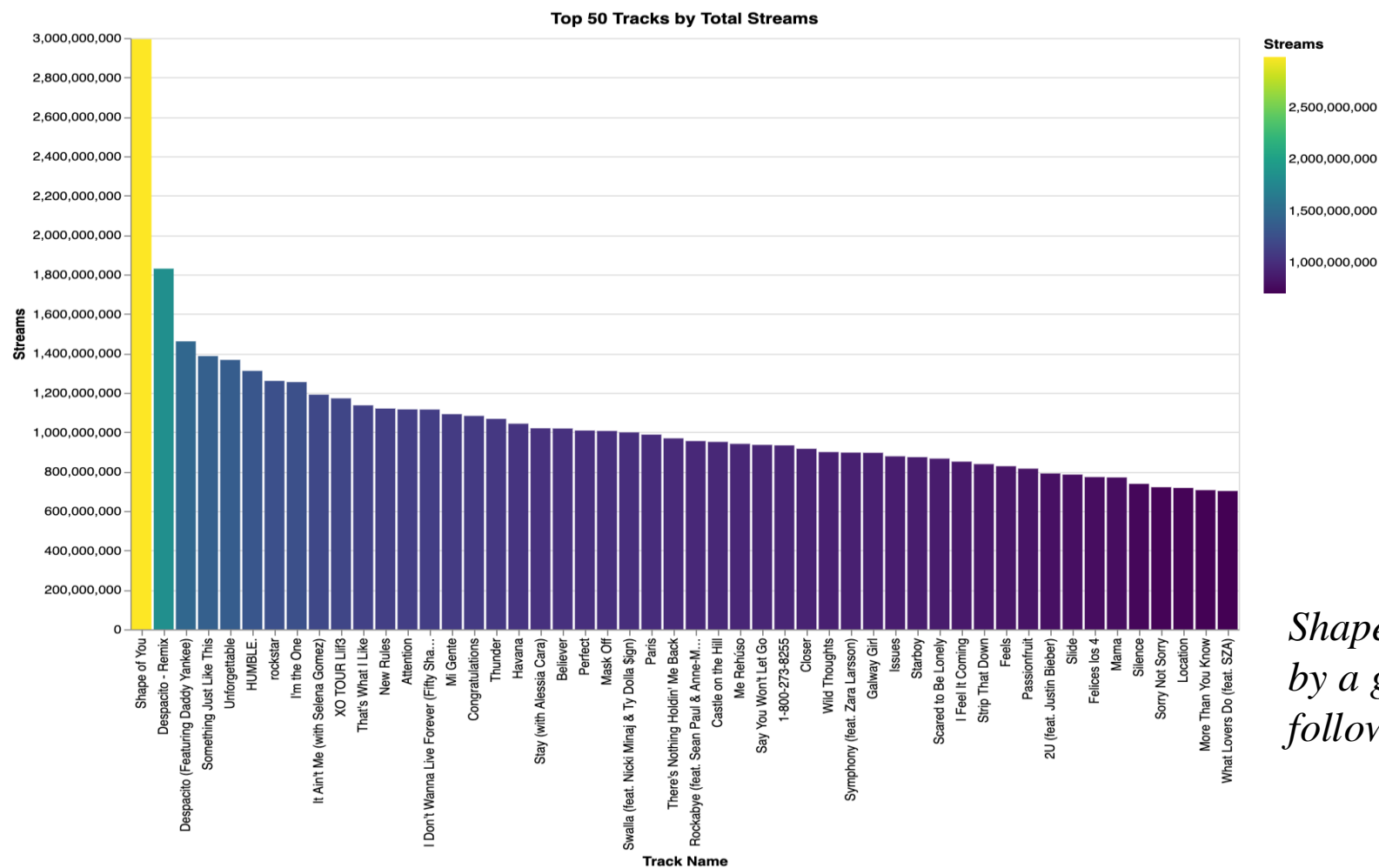
DATA DESCRIPTION

- Spotify's Worldwide Daily Song Ranking Dataset
 - Daily rankings of the top 200 most-streamed songs in **53 countries**
 - Covers data from **January 1, 2017, to January 9, 2018**
 - Over 3 million rows with about **6,629 artists** and **19,923 songs** totaling **178 billion streams**
 - **Metadata** - song popularity metrics such as stream counts, user ratings, and listener engagement over time
 - Analyze song popularity trends globally and locally and capture metrics like stream counts, user ratings, and listener engagement over time
- Song Lyrics Dataset
 - Includes lyrics of **~60,000 songs** with metadata
 - Provides artist, song title, and hyperlinks to the original lyrics source
 - Perform text analysis on lyrics to explore their impact on song popularity and study linguistic and stylistic features in music trends

STREAM KINGS: ED SHEERAN LEADS



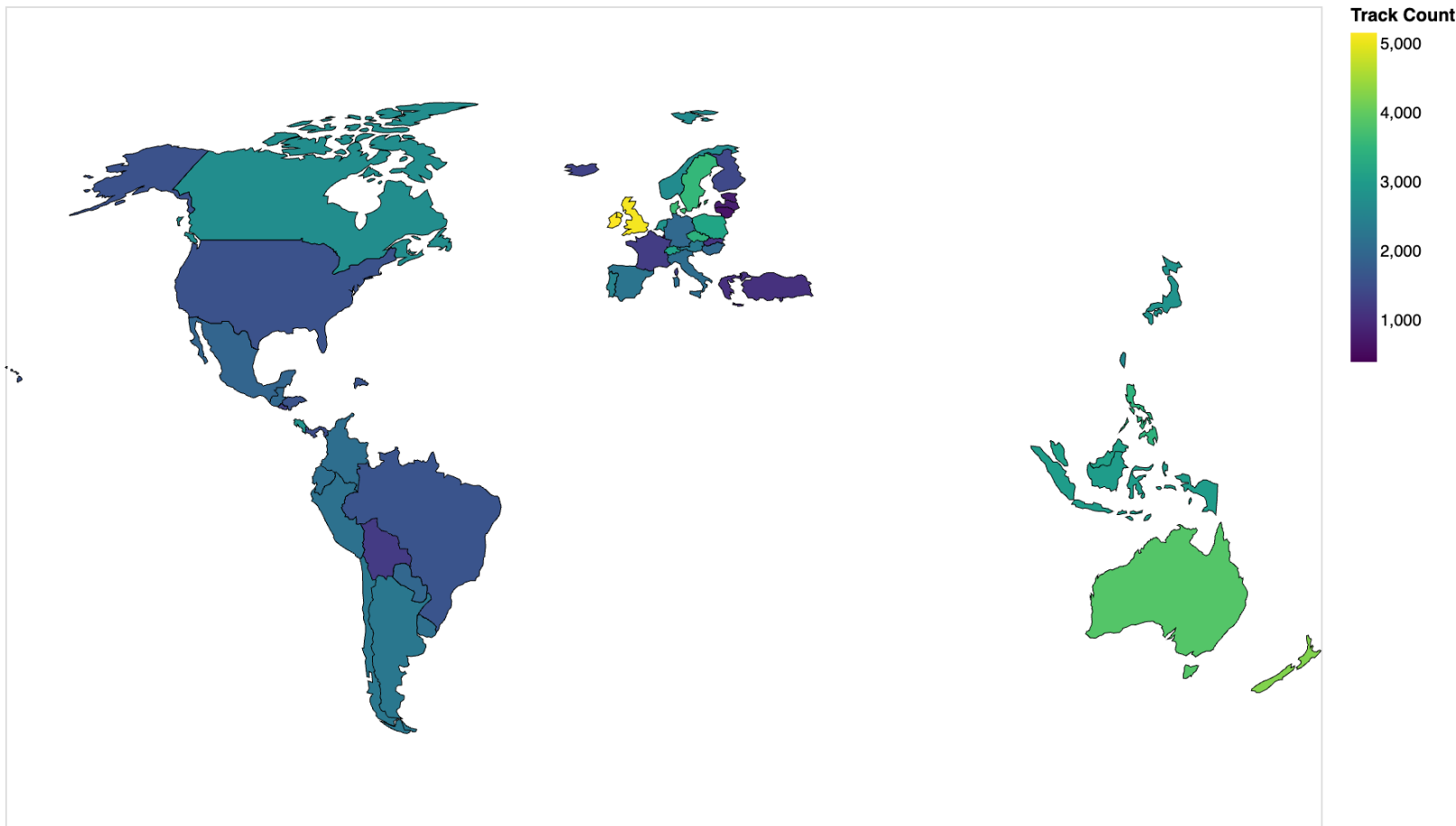
TOP 50 TRACKS BY TOTAL STREAMS: DOMINANCE OF 'SHAPE OF YOU'



Shape of You leads in the tracks by total streams by a great margin of about 3 billion streams, followed by Despacito.

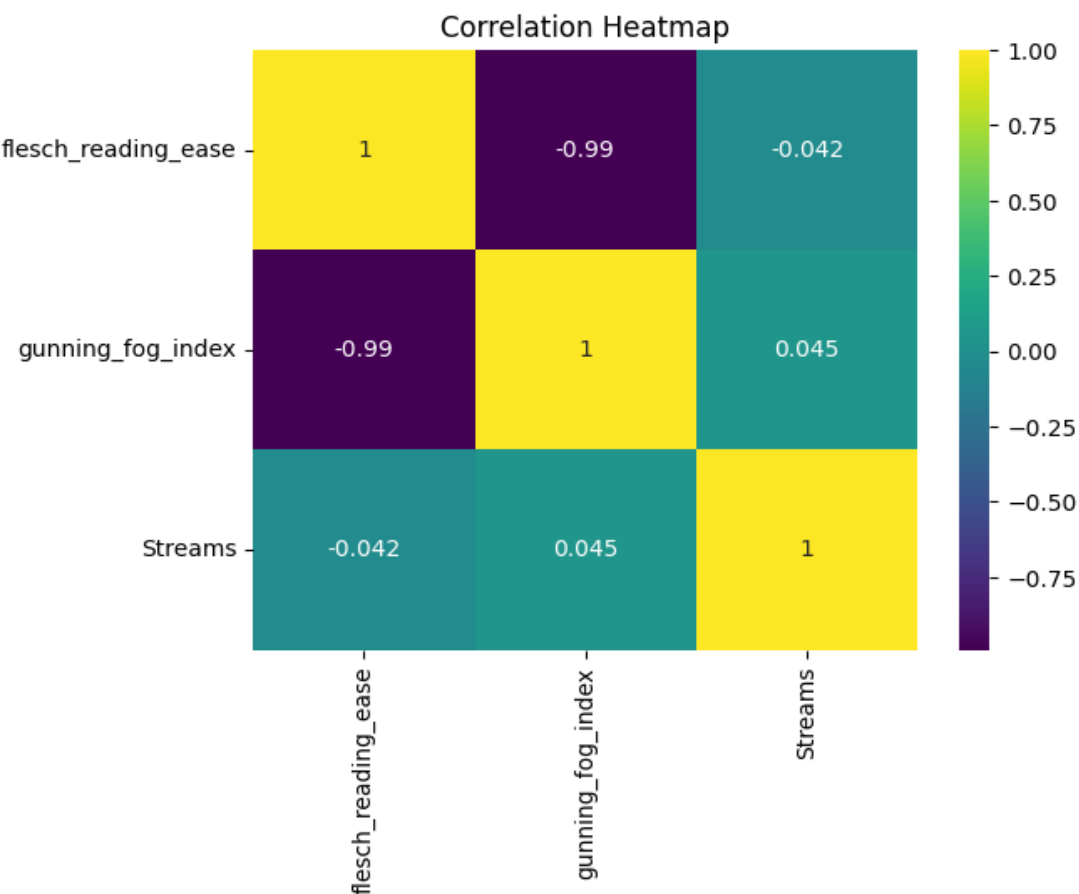
ED SHEERAN'S GLOBAL POPULARITY: A REGIONAL BREAKDOWN

Ed Sheeran Track Count by Region



The map shows Ed Sheeran's global reach, with the UK - his home country - leading at 5,000+ tracks, followed by Australia, and New Zealand (4,000+). European countries like Germany and France follow, while emerging markets in South America, Asia, and Africa have fewer tracks, under 2,000.

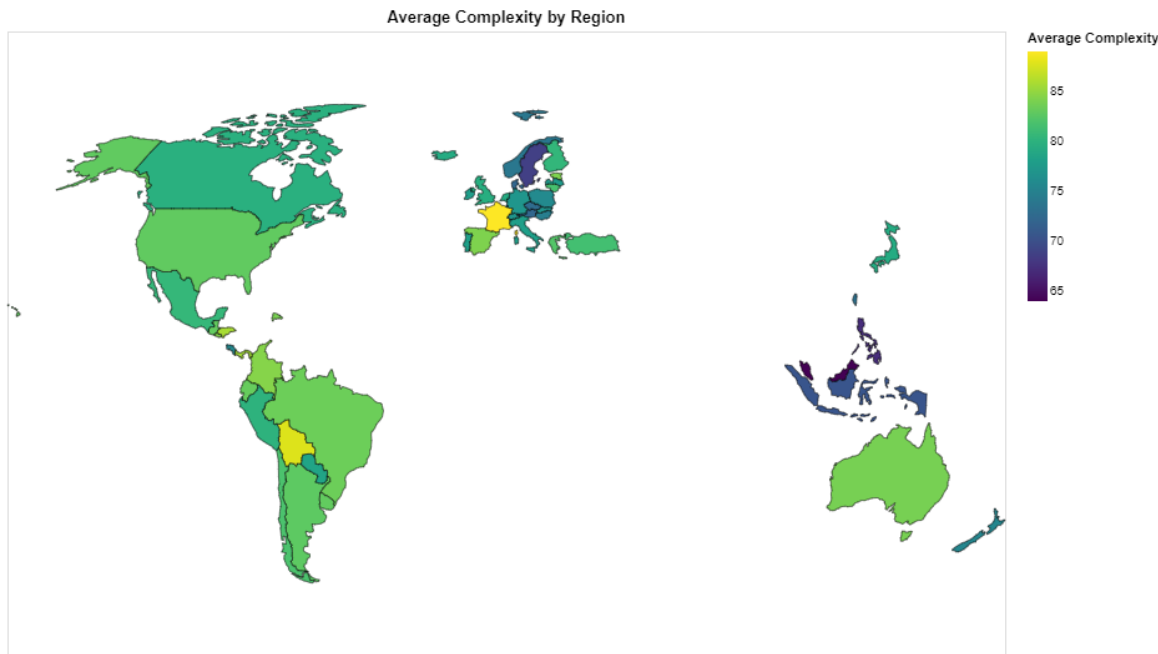
CORRELATION ANALYSIS: READABILITY AND STREAMS



	Flesch Reading Ease	Streams
Flesch Reading Ease	1.000000	-0.042006
Streams	-0.042006	1.000000

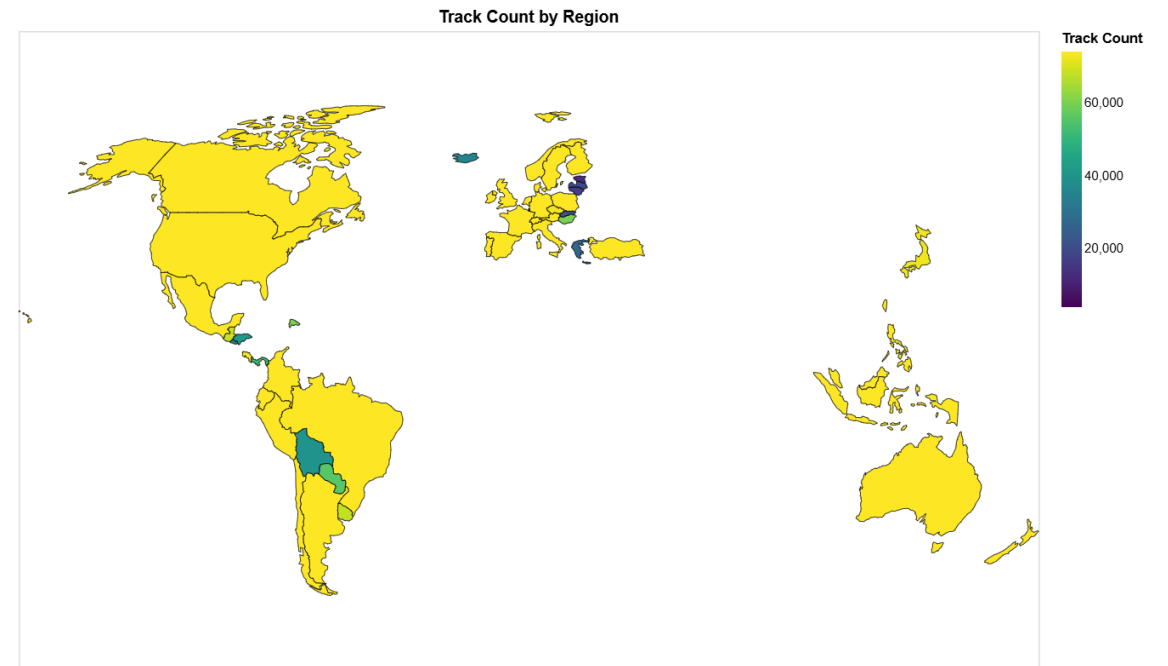
The Flesch score indicates extreme difficulty, while the Gunning Fog index suggests extreme simplicity. This discrepancy might indicate an issue with the content's structure, such as fragmented text or inconsistent sentence complexity.

CHOROPLETH



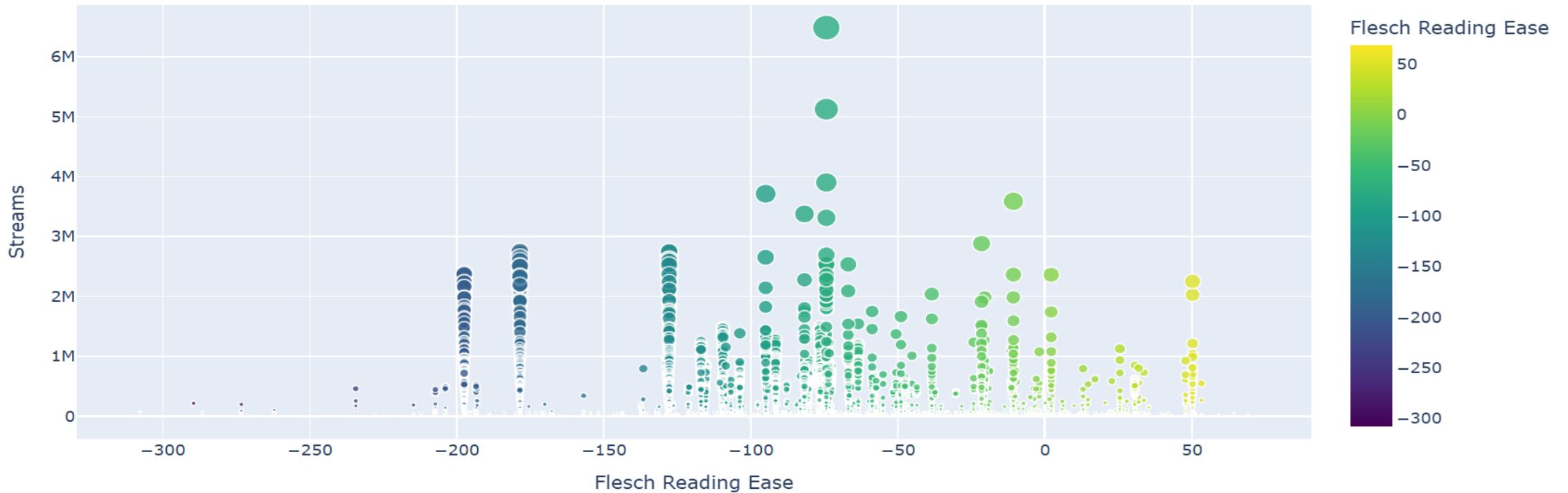
European and Southeast Asian regions show higher lyrical complexity, favoring intricate songwriting, while the Americas have moderate to lower complexity, with simpler, more accessible lyrics common in popular genres.

Track counts vary widely by region: the US leads with 80,000+ tracks, followed by Western Europe. Latvia and Lithuania have under 20,000 tracks, while emerging markets like Brazil and South Korea are growing.



SCATTERPLOTS

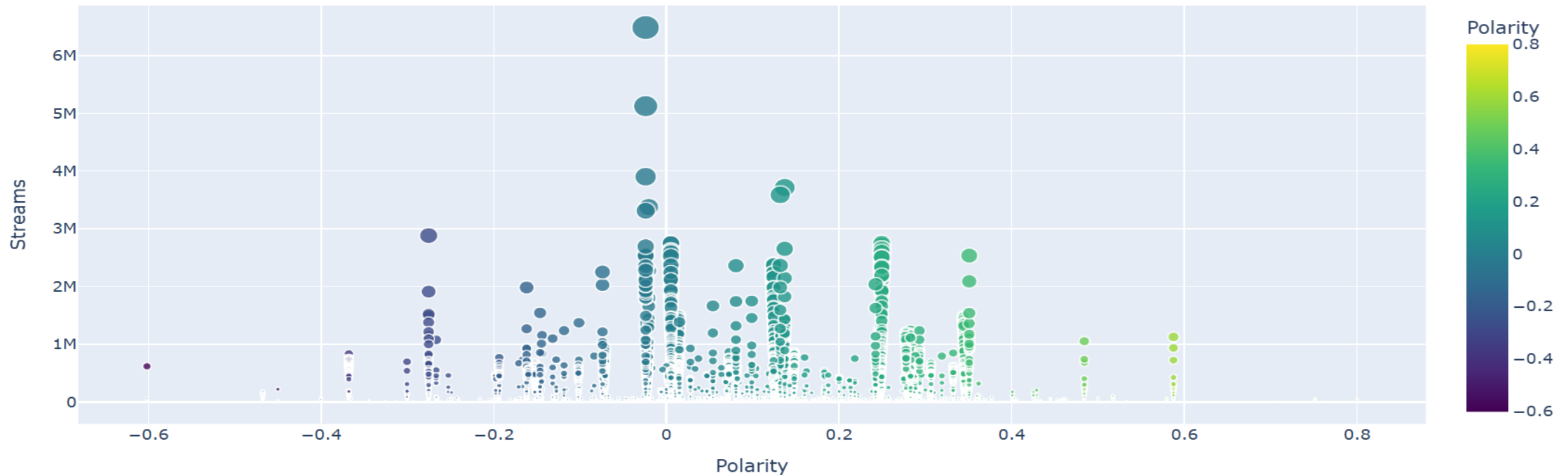
Popularity vs. Complexity (2D Scatter)



Most streamed content correlates with moderately negative Flesch Reading Ease scores (-200 to -100), suggesting that content with a certain level of complexity might be more appealing or relatable to the audience.

SCATTER PLOT INSIGHTS : UNDERSTANDING POPULARITY THROUGH SENTIMENT

Popularity vs. Number of Streams



Slightly negative sentiment (-0.3 to -0.1) and positive sentiment (0.1 to 0.3) drive high streams, as they strike a balance between emotional depth and relatability. Extreme sentiment (positive or negative) tends to lower popularity, possibly due to being too intense or niche.

CONCLUSION

Results

- English-speaking countries dominate the music industry leading in track and stream counts.
- Moderately complex and emotionally positive or negative lyrics resonate most with listeners.
- Regional differences highlight cultural and linguistic diversity in music preferences.
- Data limitations, such as missing genre classifications, impact analysis accuracy.

Future Scope

- Use detailed datasets with genre and lyrical features like rhyme patterns.
- Explore temporal trends and regional nuances in song preferences.
- Investigate social media engagement to understand user listening habits
- Deep dive into regional trends for specific regions like Latvia to understand local preferences and trends.



THANK YOU