Duygu Göksu

Data Scientist with a Linguistic Background

Data Science

PhD, Linguistics

University of Massachusetts Amherst

9+ Years

Experience Overseeing Research Projects

Volunteer

Coach at Women's Money Matters

Languages

Proficient in English, and Turkish

9+ Years

Experience Educating as a Teaching Assistant and Instructor

dg.duygu.goksu@gmail.com

https://www.linkedin.com/in/duygugoksu2023/

SONG POPULARITY PREDICTOR

- NLP analysis of lyrics by GPT 3.5
- Spotify song metrics

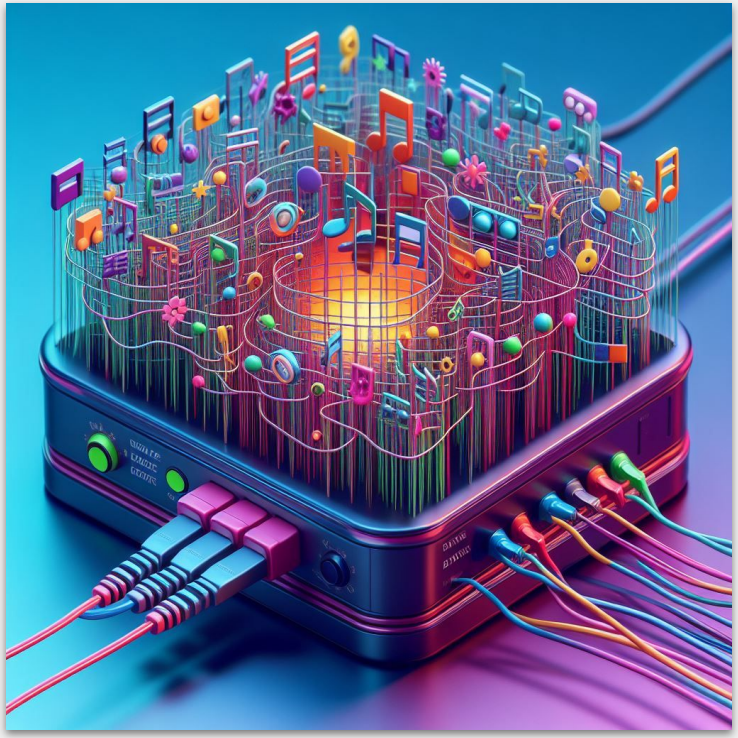
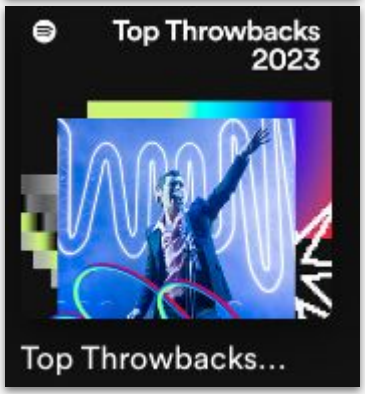


Image by Microsoft Copilot Designer

Spotify Track Popularity Scores



- Range from 0-100
- Determined by recent stream count, save/skip rate, number of playlists*
- Closer to 50 and above to be in official Spotify playlists **

Impact: 11 million artists get paid about $0.003 - $0.005 per stream.***

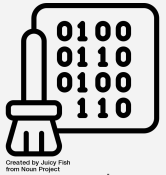
Sources: *https://www.loudlab.org **https://medium.com ***https://www.searchlogistics.com

Data Collection

~30,000 Spotify songs dataset

+ ~2000 lyrics from other Spotify datasets

API calls to Genius.com for more lyrics



~20,000 songs with lyrics and Spotify song metrics

“You are a helpful assistant who…

… knows what qualities in lyrics contribute to song popularity. You rate song lyrics in terms of these qualities. If there are swear words, ignore the swear words and rate the lyrics based on the rest. Give ratings out of 10, where 10 is the highest and 1 is the lowest for each criteria below:

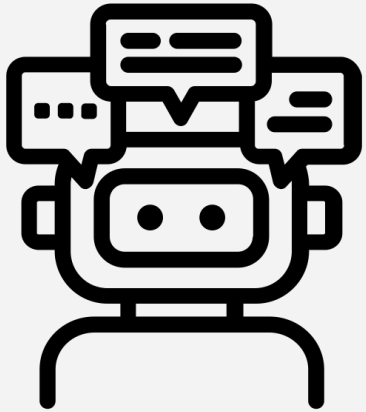
(Coming on the next slide)

Return only the 12 ratings as a space delimited list, in the same order as the criteria.”

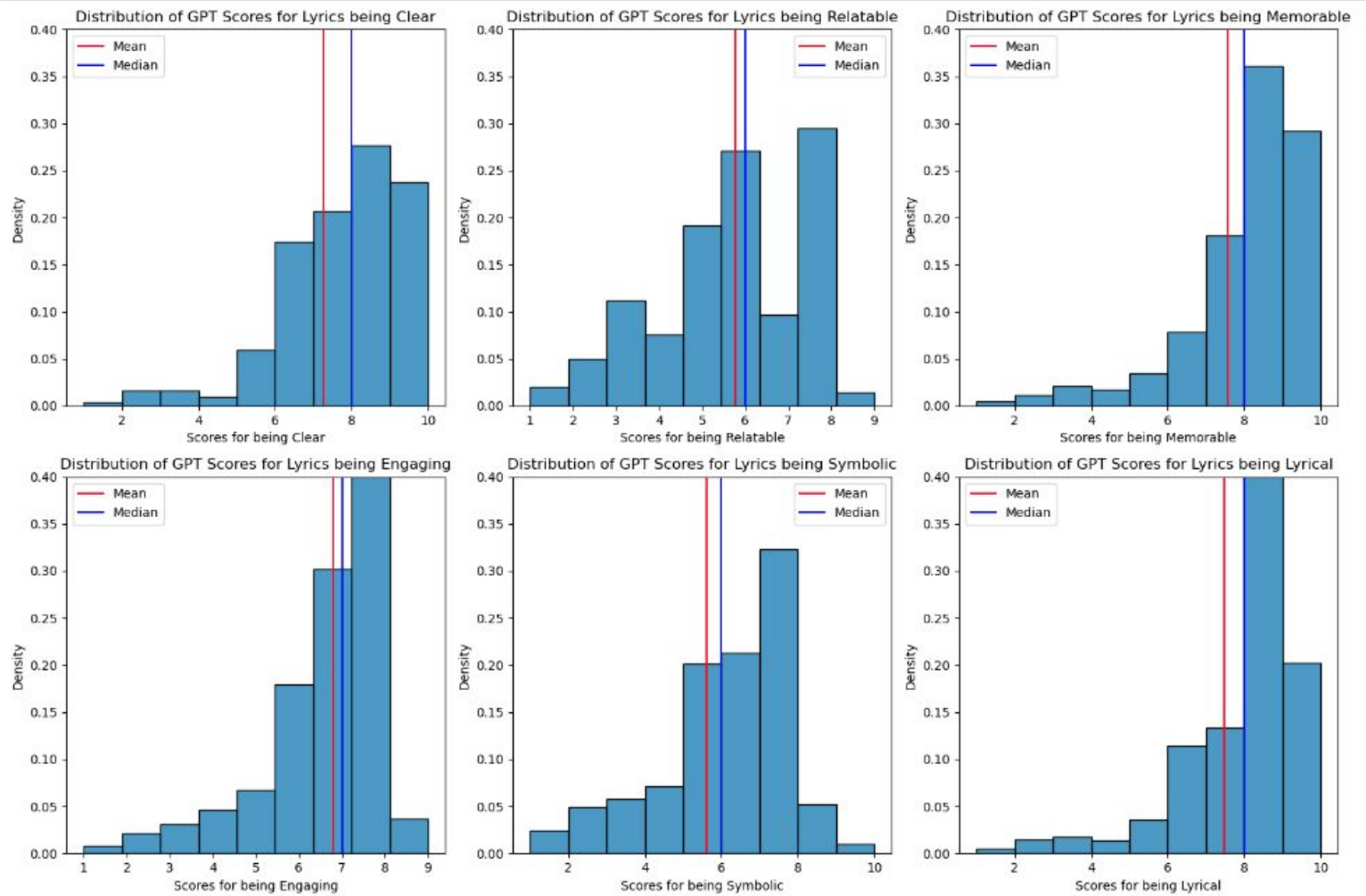
★ Exemplar: Billie Eilish - ‘What was I made for?’

Criteria: suggested by Chat-GPT, simplified by me

- **Clear:** Lyrics use simple and straightforward language that is easy to understand.
- **Relatable:** Lyrics use universal themes that resonate with a wide audience.
- **Memorable:** Lyrics contain catchy language that is easy to remember.
- **Engaging:** Lyrics weave an engaging narrative with vivid imagery.
- **Symbolic:** Lyrics use rich symbolism and metaphors to convey deeper meanings.
- **Lyrical:** Lyrics exhibit characteristics typical of song lyrics rather than other forms of content.

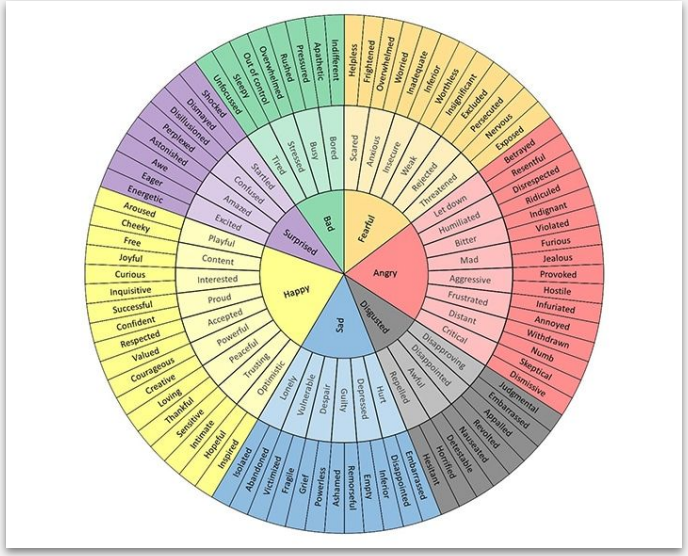


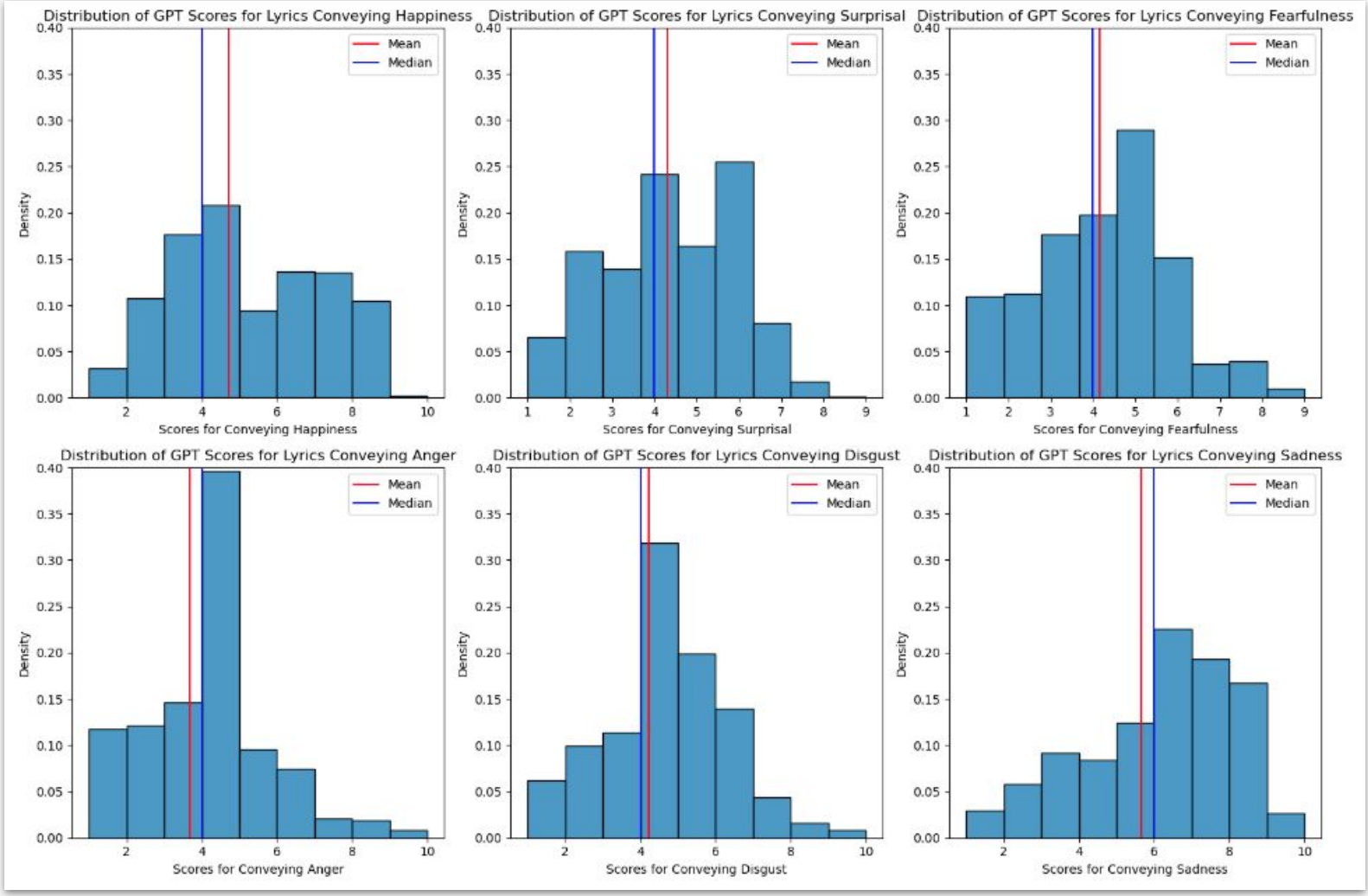
Created by 응가
from Noun Project


Distribution of GPT Scores for Lyrics being Clear
Mean
Median
Density
Scores for being Clear
Distribution of GPT Scores for Lyrics being Relatable
Mean
Median
Density
Scores for being Relatable
Distribution of GPT Scores for Lyrics being Memorable
Mean
Median
Density
Scores for being Memorable
Distribution of GPT Scores for Lyrics being Engaging
Mean
Median
Density
Scores for being Engaging
Distribution of GPT Scores for Lyrics being Symbolic
Mean
Median
Density
Scores for being Symbolic
Distribution of GPT Scores for Lyrics being Lyrical
Mean
Median
Density
Scores for being Lyrical

Criteria: sentiment analysis with a focus on emotions

- **Happy:** Lyrics convey feelings of happiness, playfulness, contentment, acceptance, empowerment, serenity, trust, optimism.
- **Surprised:** Lyrics convey emotions of surprise, confusion, amazement.
- **Fearful:** Lyrics convey emotions of fear, anxiety, insecurity, rejection, danger.
- **Angry:** Lyrics convey feelings of anger, disappointment, humiliation, bitterness, aggression, frustration, distance and criticism.
- **Disgusted:** Lyrics convey disgust, disapproval, disappointment, repulsion, and aversion.
- **Sad:** Lyrics convey emotions of sadness, loneliness, despair, guilt, depression, and hurt.



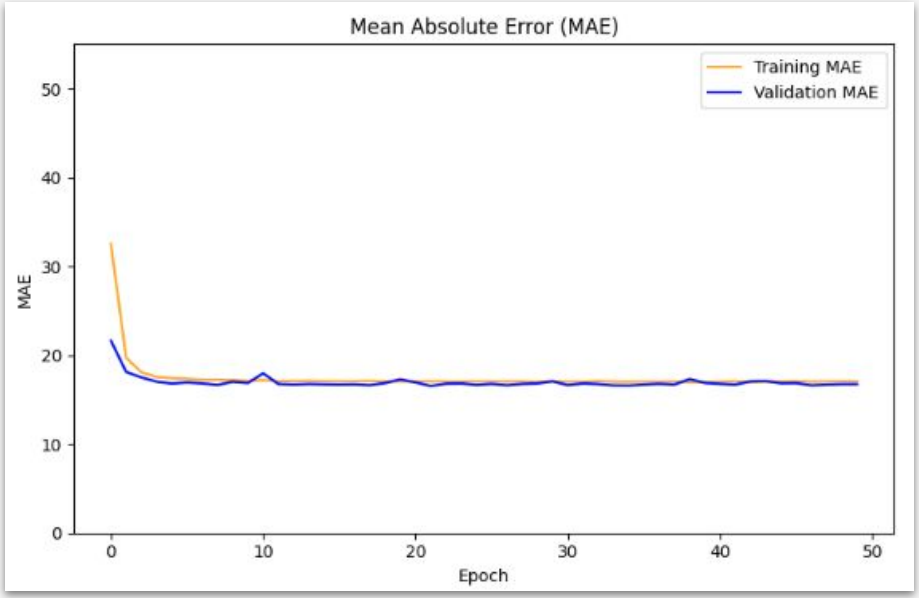

Distribution of GPT Scores for Lyrics Conveying Happiness
Mean
Median
Density
Scores for Conveying Happiness
Distribution of GPT Scores for Lyrics Conveying Surprisal
Mean
Median
Density
Scores for Conveying Surprisal
Distribution of GPT Scores for Lyrics Conveying Fearfulness
Mean
Median
Density
Scores for Conveying Fearfulness
Distribution of GPT Scores for Lyrics Conveying Anger
Mean
Median
Density
Scores for Conveying Anger
Distribution of GPT Scores for Lyrics Conveying Disgust
Mean
Median
Density
Scores for Conveying Disgust
Distribution of GPT Scores for Lyrics Conveying Sadness
Mean
Median
Density
Scores for Conveying Sadness

The Best Performing Model

- 4 dense layers (69 > 36 > 18 > 9) with L1 (0.05) regularization
- 2 batch normalizations
- 1 dropout layer (0.2)

Evaluation:

- MAE: 16.7
- R-squared: 0.15
- my_app



Next Steps

- Retrieve the data from Spotify directly with API calls
- Work on a larger dataset to improve results, and try other models

dg.duygu.goksu@gmail.com

https://www.linkedin.com/in/duygugoksu2023/

TEŞEKKÜRLER!