PREDICTING SONG POPULARITY

Duygu Göksu
April 12, 2024
Final Capstone Presentation

Spotify Popularity Scores

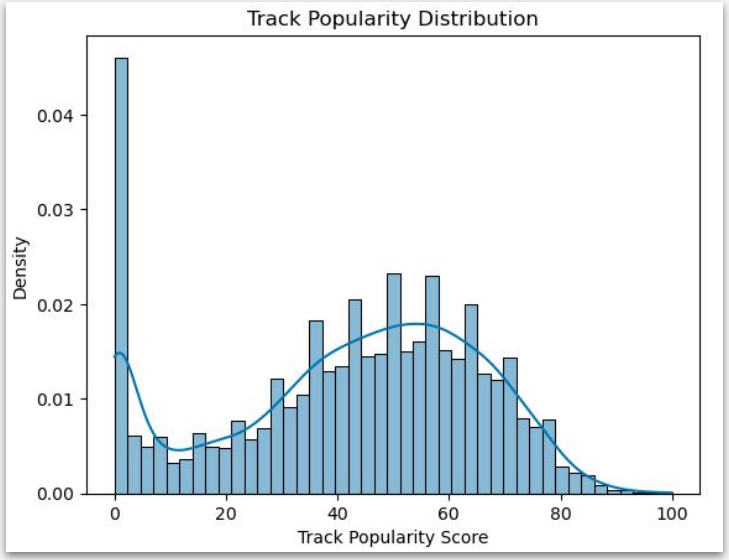
- Range from 0-100 tracking an artist's current popularity across Spotify
- Determined by recent stream count, save/skip rate, number of playlists*
- Artists with a popularity score closer to 50 and above are more likely to be in official Spotify playlists.**
- Very important for both artists and music label companies.

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

What has changed in the data?

- ~20k rows and ~70 columns
- 15 new features from lyrics
  - Out of 69 languages, one-hot encoded top 3:
    - English (77%),
    - Spanish(12%), and
    - Portuguese (2%)
- Target value distribution remains similar →



- 12 new features via API calls to Open AI using chat completion(GPT-3.5 Turbo) to analyze lyrics

“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:

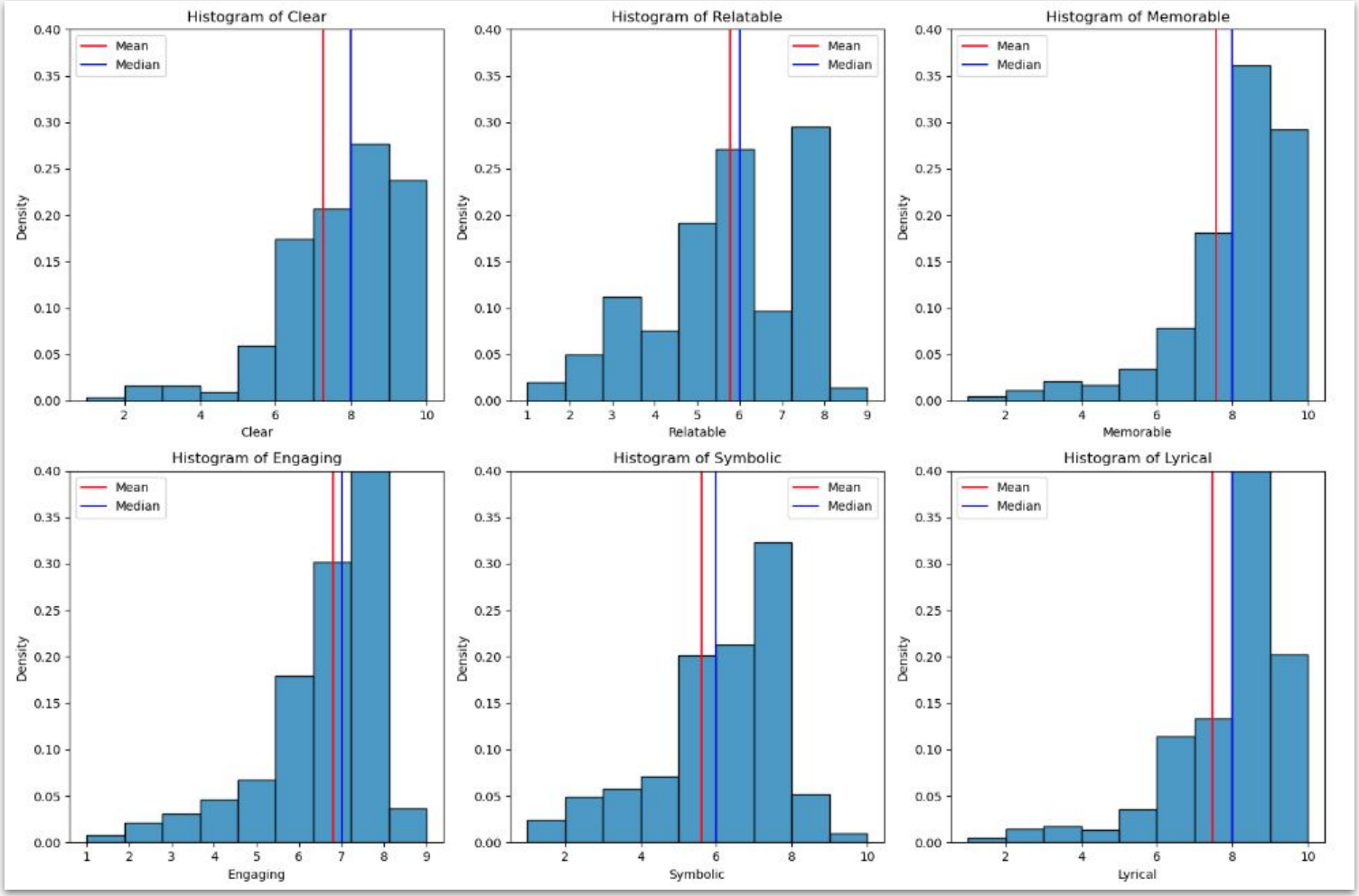
[I will give these in a second]

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

★ Also used an exemplar (Billie Eilish - ‘What was I made for?’)

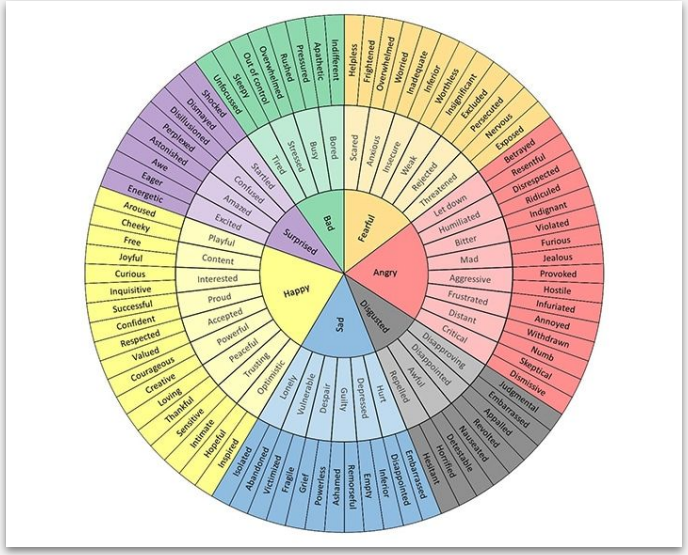
Criteria: suggested by Chat-GPT, simplified by me

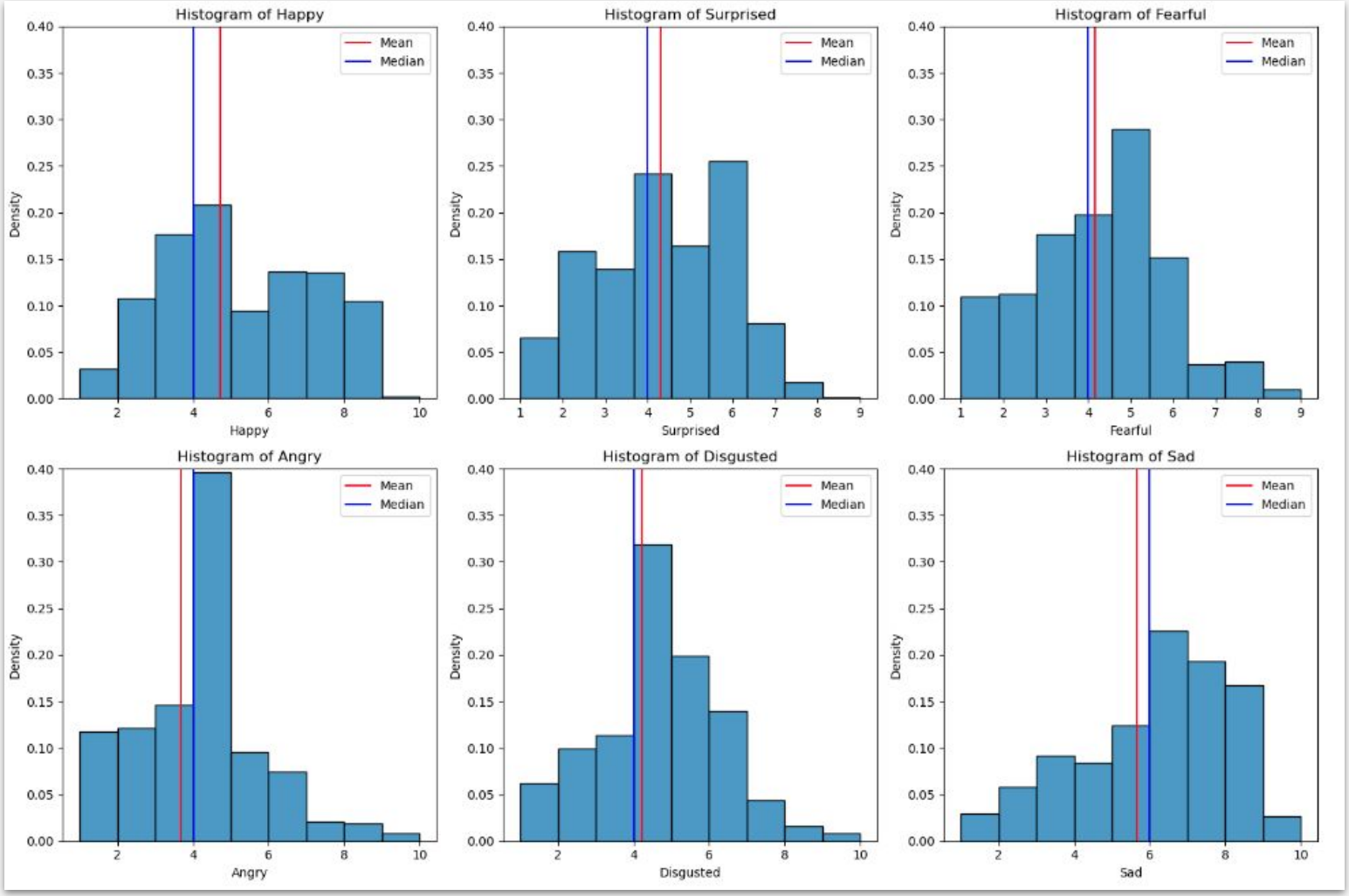
- **Clear:** Lyrics use simple and straightforward language that is easy to understand.
- **Relatable:** Lyrics use universal themes that resonate with a wide audience.
- **Memorable:** Lyrics contain catchy language that is easy to remember.
- **Engaging:** Lyrics weave an engaging narrative with vivid 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.


Histogram of Clear
Mean
Median
Density
Clear
Histogram of Relatable
Relatable
Histogram of Memorable
Memorable
Histogram of Engaging
Engaging
Histogram of Symbolic
Symbolic
Histogram of Lyrical
Lyrical

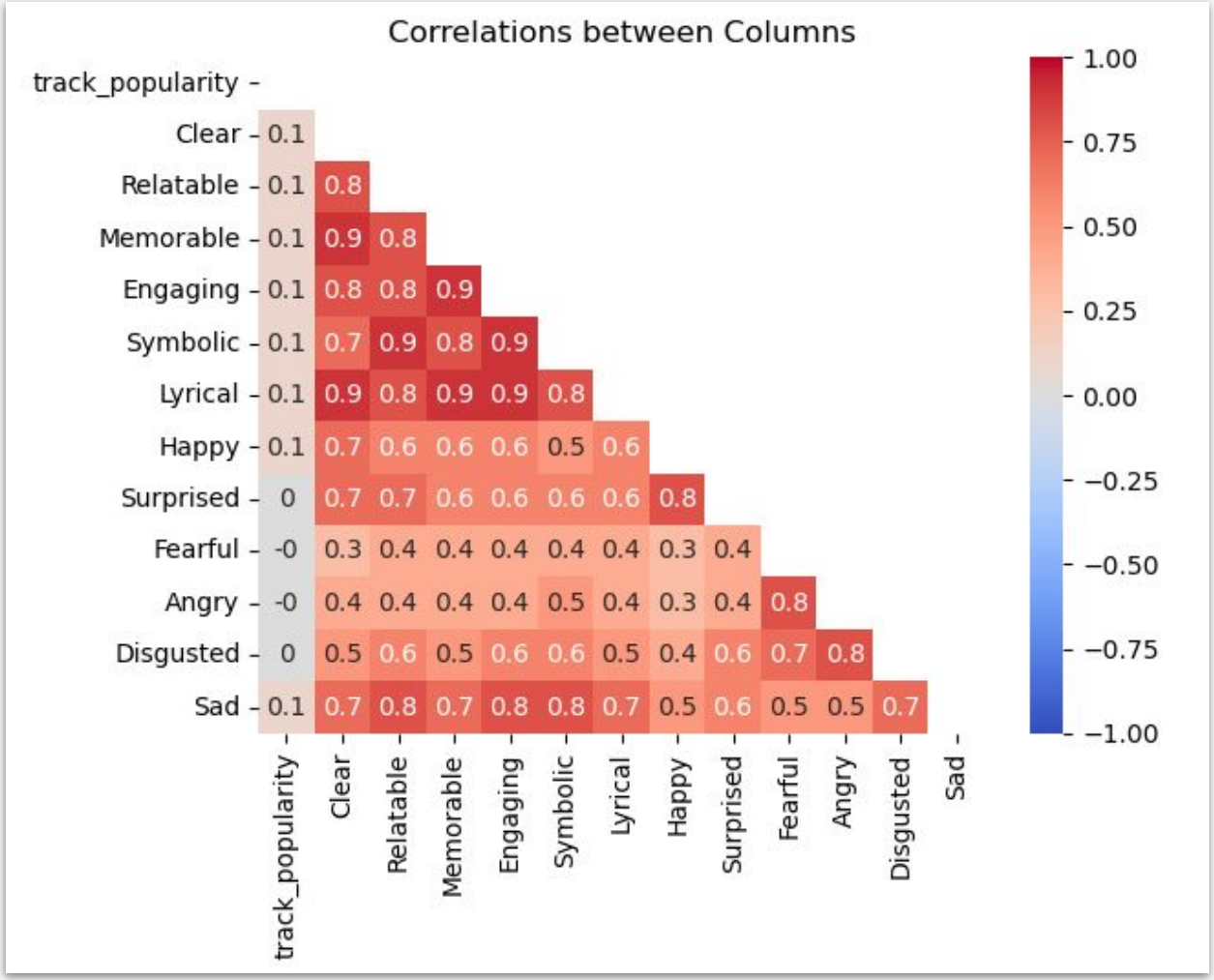
Criteria: sentiment analysis with a focus on emotions

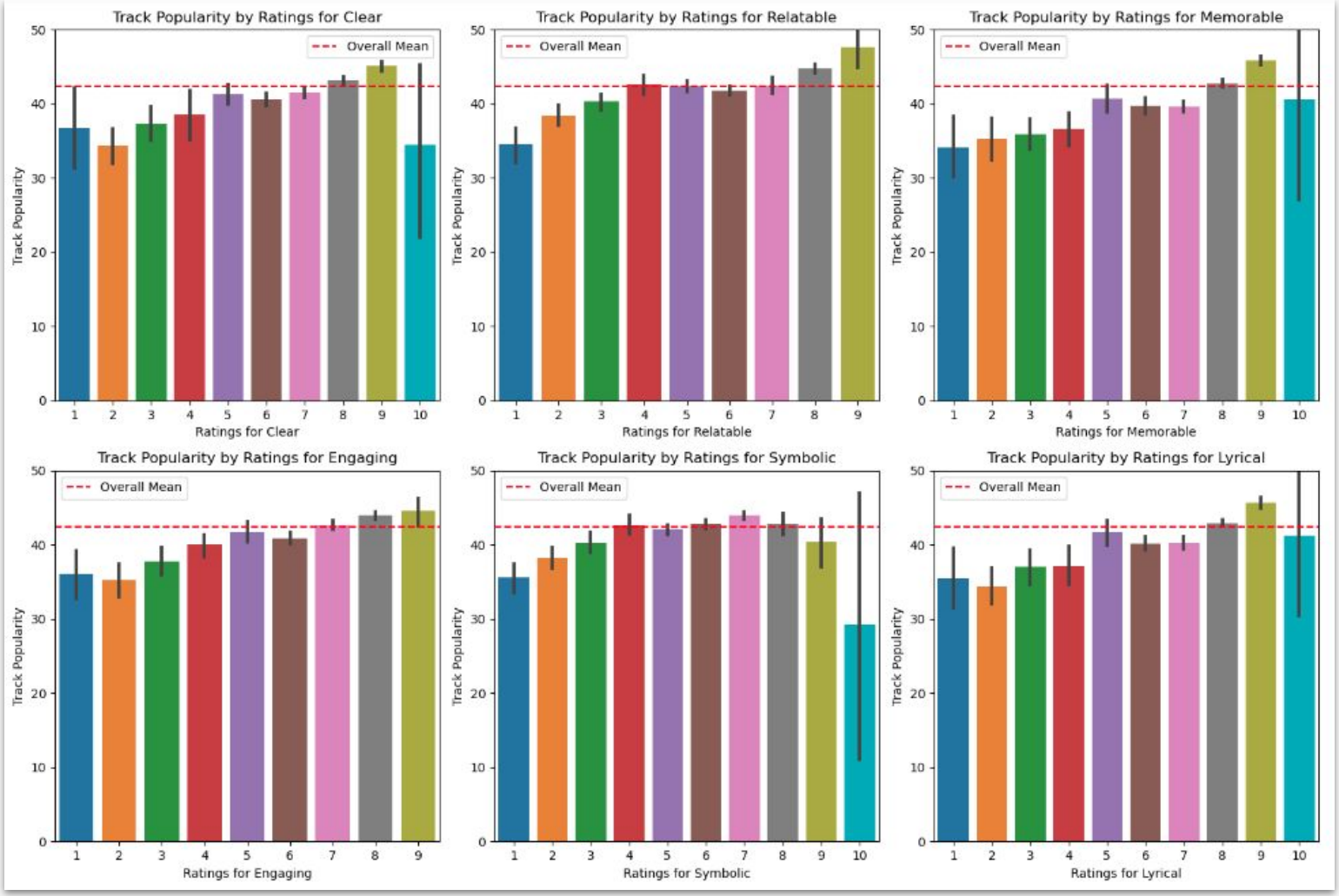
- **Happy:** Lyrics convey feelings of happiness, playfulness, contentment, acceptance, empowerment, serenity, trust, optimism.
- **Surprised:** Lyrics convey emotions of surprise, confusion, amazement.
- **Fearful:** Lyrics convey emotions of fear, anxiety, insecurity, rejection, danger.
- **Angry:** 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.

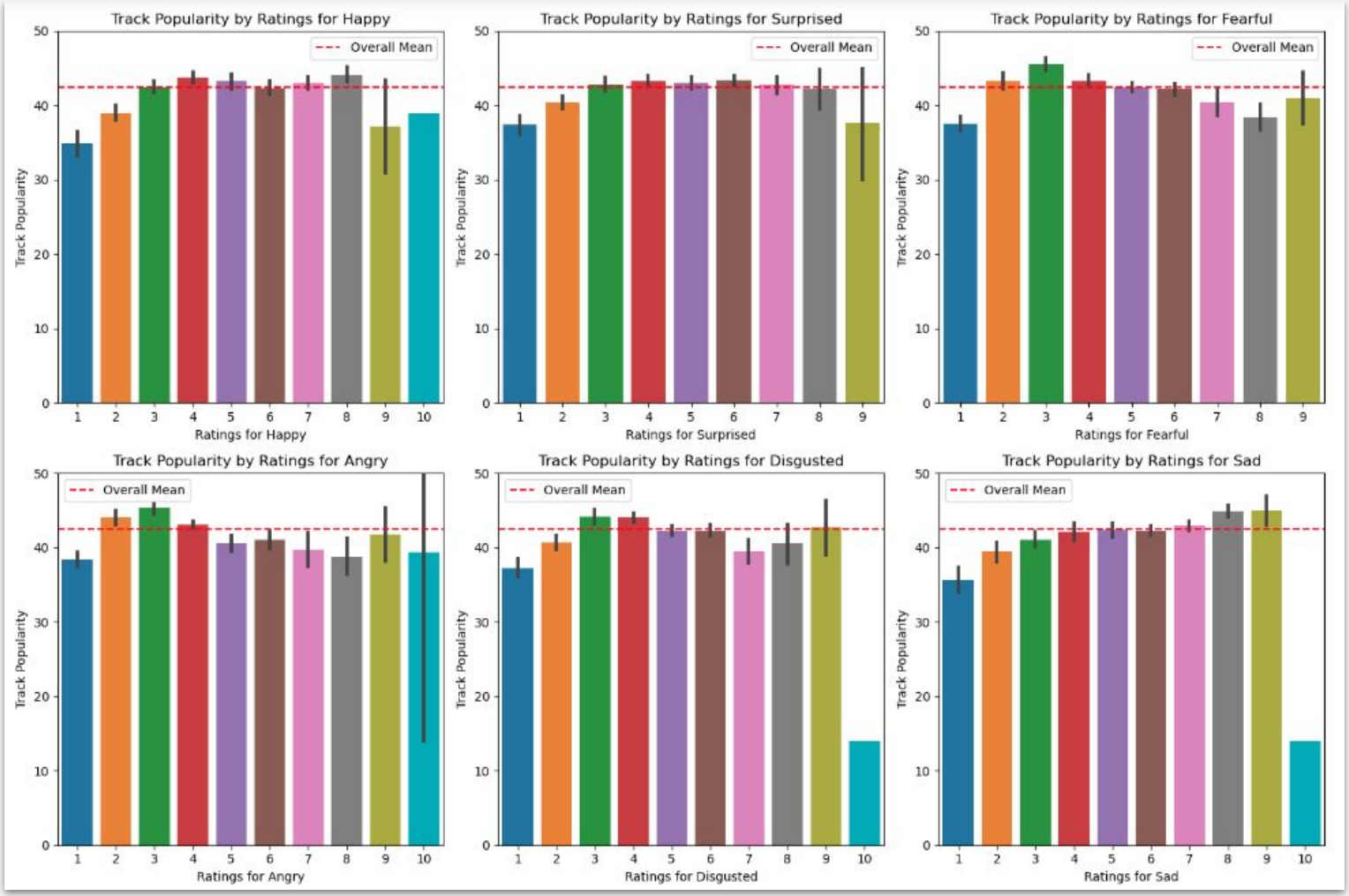



Histogram of Happy
Mean
Median
Density
Happy
Histogram of Surprised
Surprised
Histogram of Fearful
Fearful
Histogram of Angry
Angry
Histogram of Disgusted
Disgusted
Histogram of Sad
Sad

How does track popularity change by these scores?


Correlations between Columns
track_popularity
Clear
Relatable
Memorable
Engaging
Symbolic
Lyrical
Happy
Surprised
Fearful
Angry
Disgusted
Sad
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00


Track Popularity by Ratings for Clear
Overall Mean
Track Popularity
Ratings for Clear
Track Popularity by Ratings for Relatable
Overall Mean
Track Popularity
Ratings for Relatable
Track Popularity by Ratings for Memorable
Overall Mean
Track Popularity
Ratings for Memorable
Track Popularity by Ratings for Engaging
Overall Mean
Track Popularity
Ratings for Engaging
Track Popularity by Ratings for Symbolic
Overall Mean
Track Popularity
Ratings for Symbolic
Track Popularity by Ratings for Lyrical
Overall Mean
Track Popularity
Ratings for Lyrical


Track Popularity by Ratings for Happy
Overall Mean
Track Popularity
Ratings for Happy
Track Popularity by Ratings for Surprised
Overall Mean
Track Popularity
Ratings for Surprised
Track Popularity by Ratings for Fearful
Overall Mean
Track Popularity
Ratings for Fearful
Track Popularity by Ratings for Angry
Overall Mean
Track Popularity
Ratings for Angry
Track Popularity by Ratings for Disgusted
Overall Mean
Track Popularity
Ratings for Disgusted
Track Popularity by Ratings for Sad
Overall Mean
Track Popularity
Ratings for Sad

Preprocessing

Additional at this step:

- MinMaxScaler on gpt scores

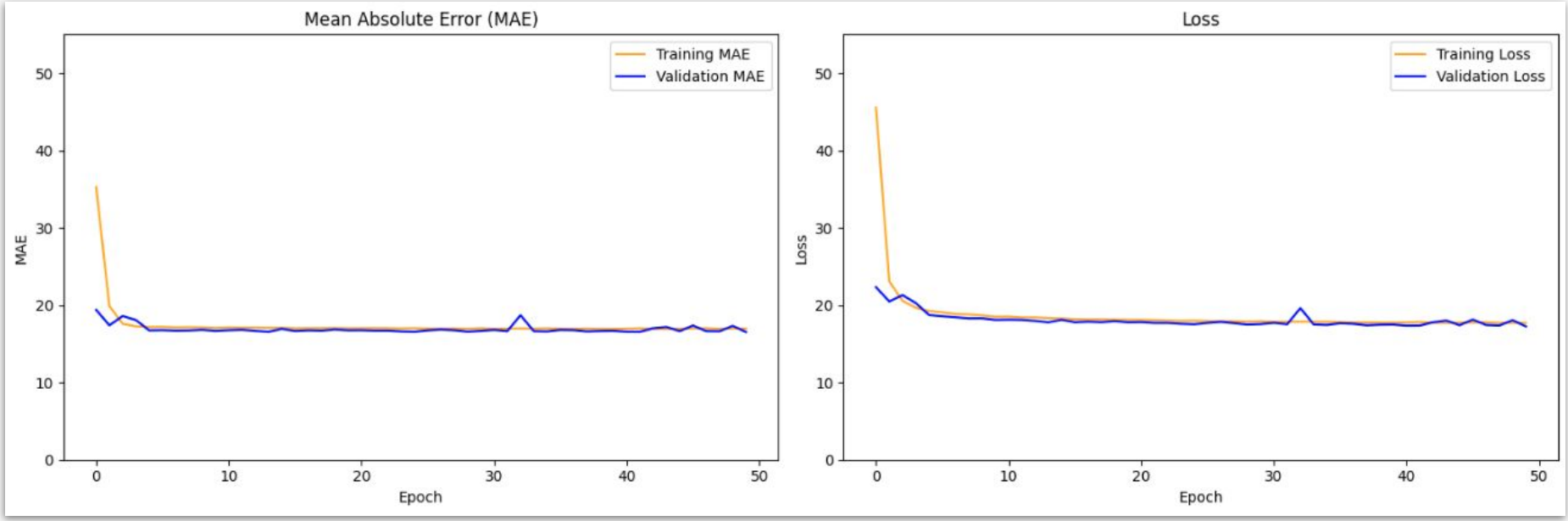
Same as before:

- MinMaxScaler on positive scores,
- StandardScaler on positive-negative scores.

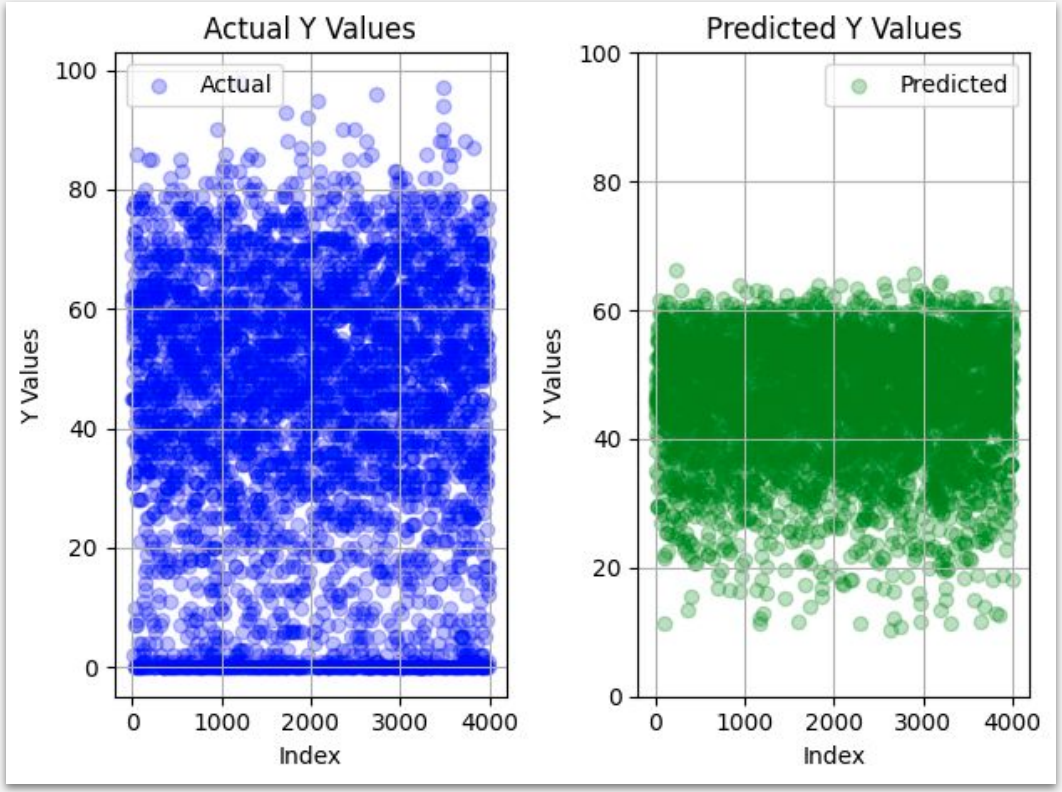
My Best Neural Net (so far)

```
Model: "sequential_6"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 dense_28 (Dense)            (None, 73)                5110

 batch_normalization_10 (Ba  (None, 73)                292
 tchNormalization)

 dense_29 (Dense)            (None, 36)                2664

 dropout_3 (Dropout)         (None, 36)                0

 dense_30 (Dense)            (None, 18)                666

 batch_normalization_11 (Ba  (None, 18)                72
 tchNormalization)

 dense_31 (Dense)            (None, 9)                 171

 dense_32 (Dense)            (None, 1)                 10

=================================================================
Total params: 8985 (35.10 KB)
Trainable params: 8803 (34.39 KB)
Non-trainable params: 182 (728.00 Byte)
_________________________________________________________________
```

Model Evaluation



R-squared: 0.15, MAE: 16.6 (Linear model was 0.17 and 17.3)


Actual Y Values
Actual
Y Values
Index
0
20
40
60
80
100
0
1000
2000
3000
4000
Predicted Y Values
Predicted
Y Values
Index

Next Steps

- ➢ Create an interactive 3D visual showing song popularity scores with PCA,
- ➢ Run a linear regression grid search with the updated data,
- ➢ Explore the coefficients for interpretation,
- ➢ Select features with high coefficients and re-build the models.

TEŞEKKÜRLER!