

By: Jasmine Castañeda, Nayma Kim, Matilda Sorić

# OUTLINE

## OUR DISCUSSION POINTS

Goals

Data Overview

Features

Analysis

Linear Regression Model

Recommendation Model

Suggestions

# GOALS

## HELPING RECORD PRODUCERS MAKE INFORMED DECISIONS

Sign artists that have a high chance of being popular

## FIND VARIABLES THAT CONTRIBUTE TO POPULARITY

What features tend to make songs more popular?

## PREDICT POPULARITY

Model that will predict the popularity based on their features

## BUILD PLAYLIST BASED ON LISTENERS TASTE

Given a song, what are songs the user would also like

# Data Overview

---

## *Observations:*

- 🎵 30% did not have month or day of release
- 🎵 16% of data had a popularity of 0 (27,296)
- 🎵 79.5% only had one artists

## *Specifics:*

- 🎵 Mean is 31.5
- 🎵 Median is 34
- 🎵 Data Splitting

170,653 rows and 19 columns



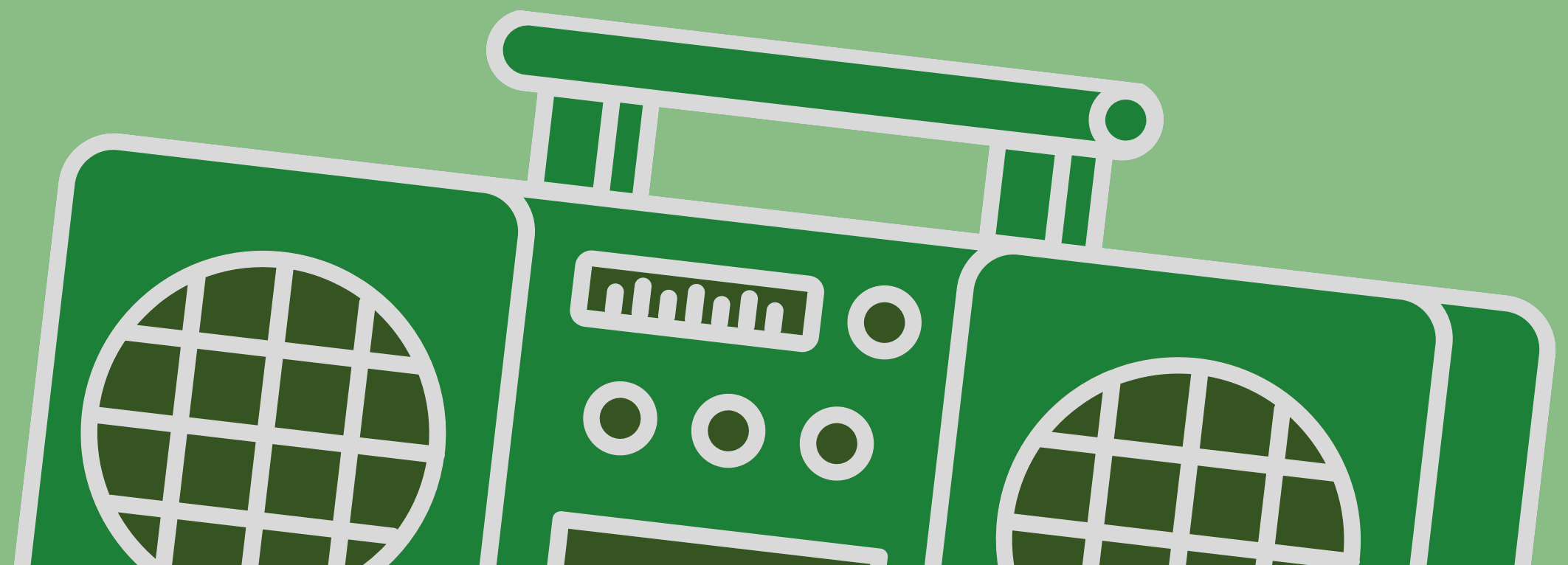
168,694 rows and 17 columns

*R: duplicates*

*C: Id and Key*

# FEATURES

HOW ARE SONGS MEASURED?





Valence



Acousticness



Energy



Liveness



Explicit



Danceability



Instrumentalness



Liveness



Tempo



Mode

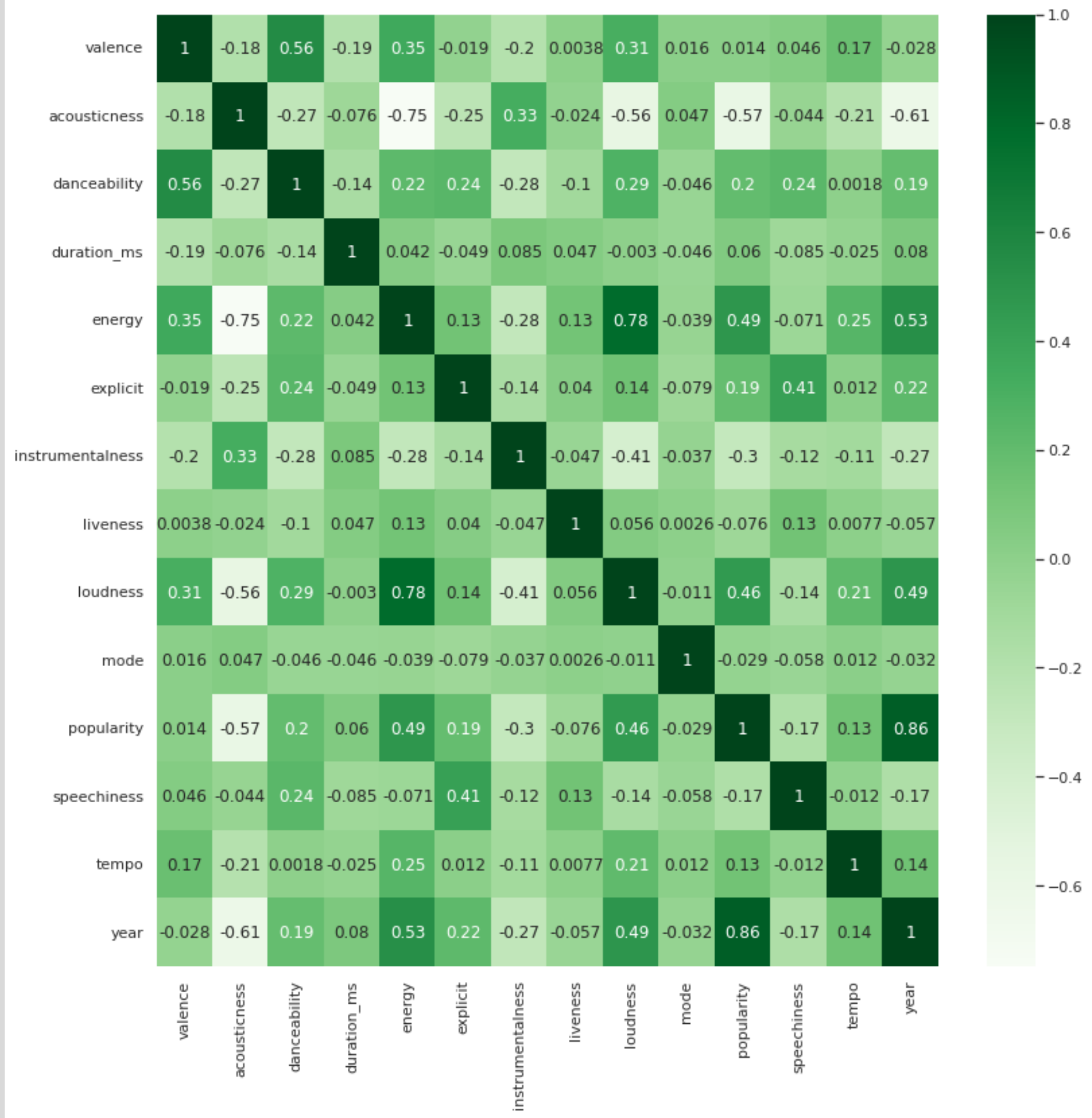


Loudness



Speechiness

# Correlation



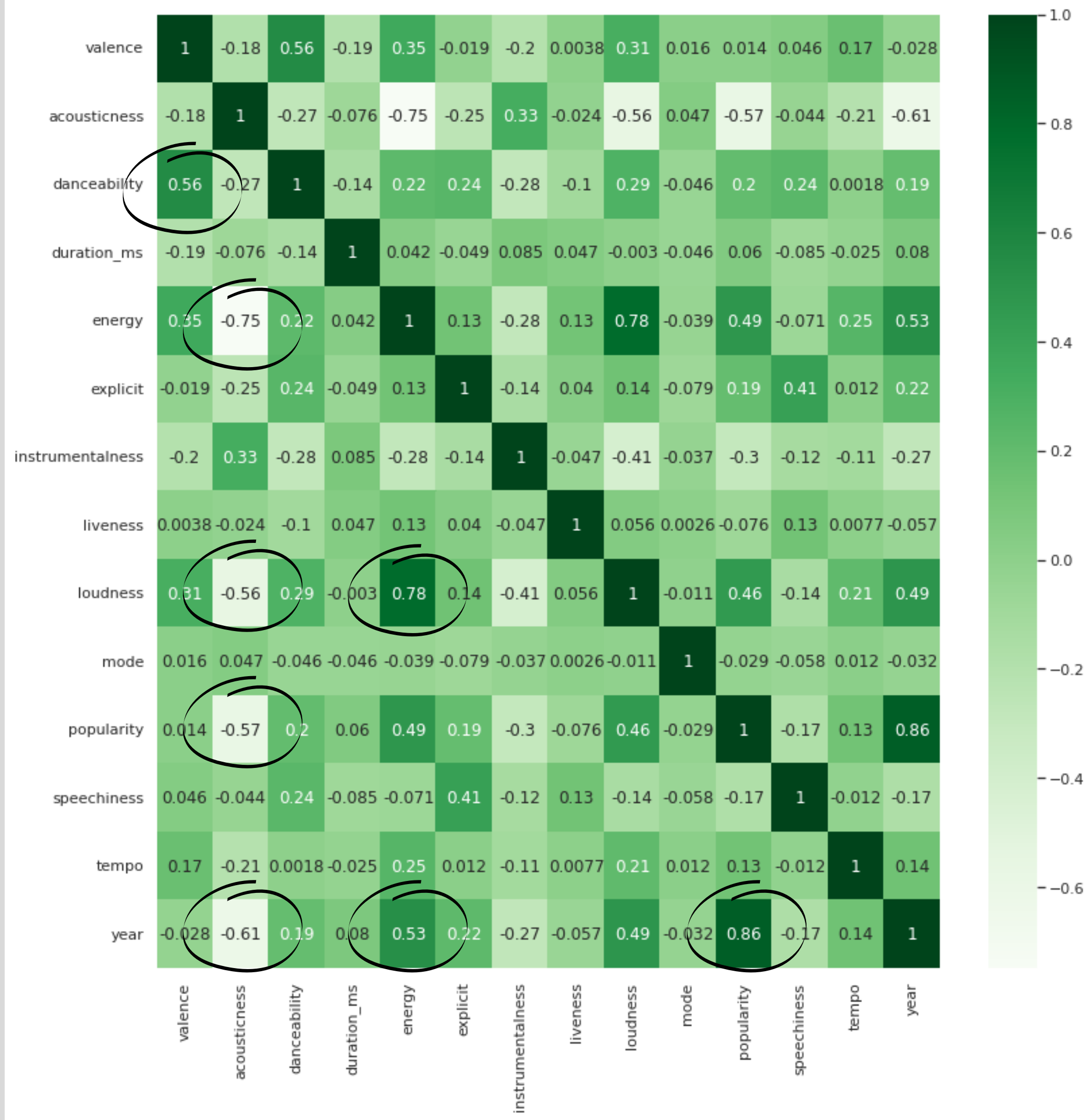


# Negative

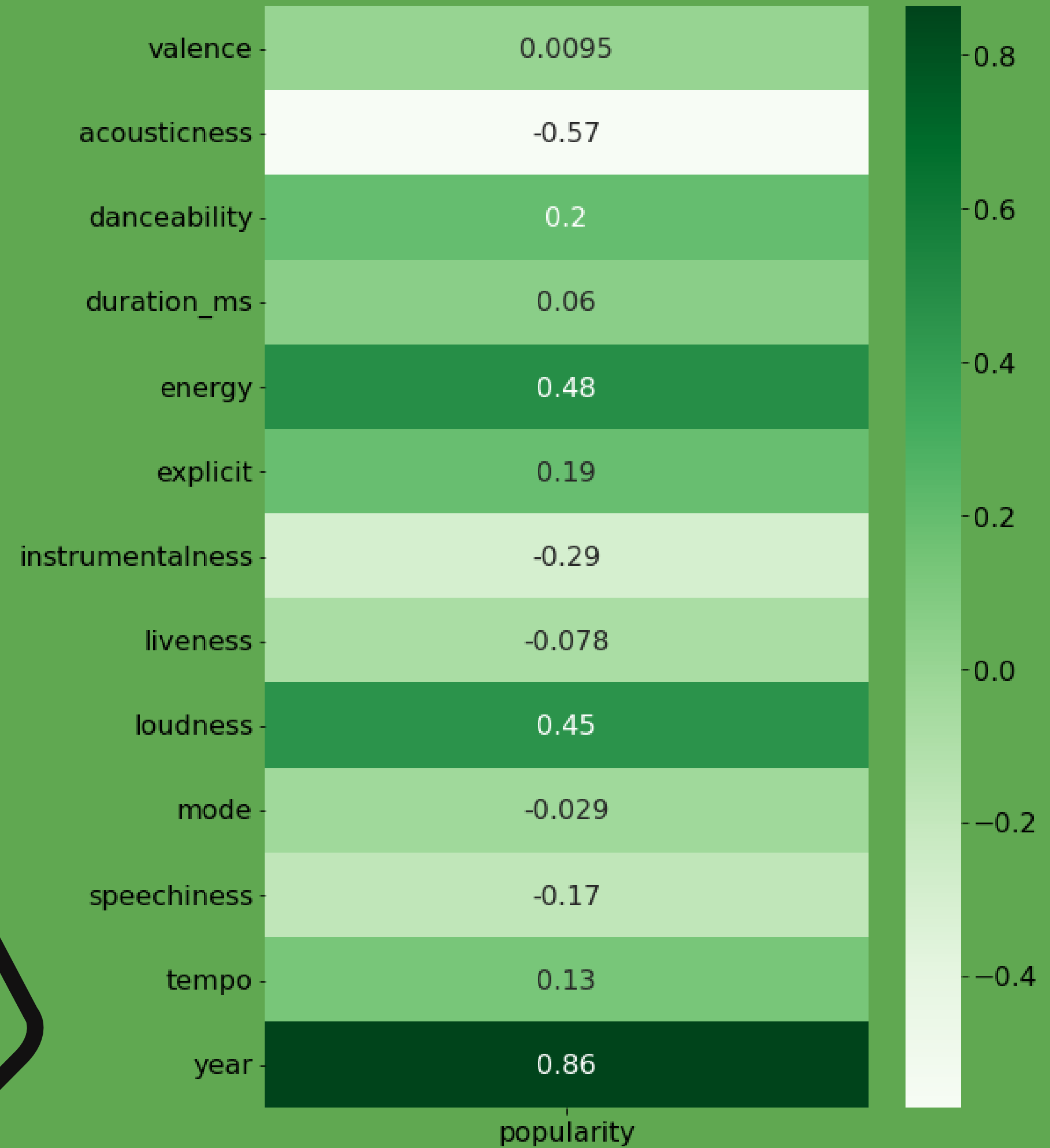
- Energy and Acousticness
- Loudness and Acousticness
- Popularity and Acousticness
- Acousticness and year

# Positive

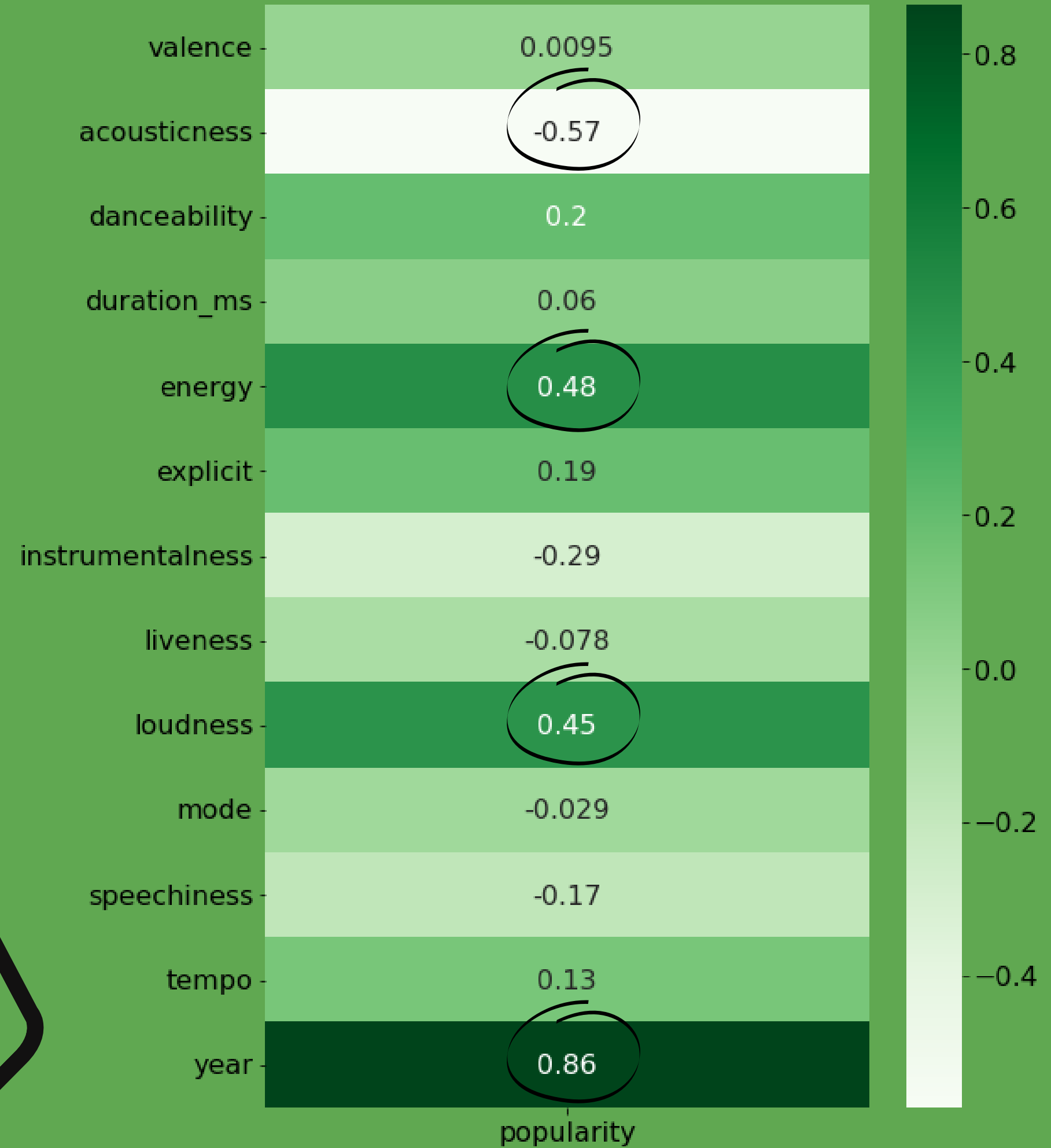
- Danceability and Valence
- Loudness and Energy
- Year and Popularity
- Energy and Year



# Correlation and Popularity

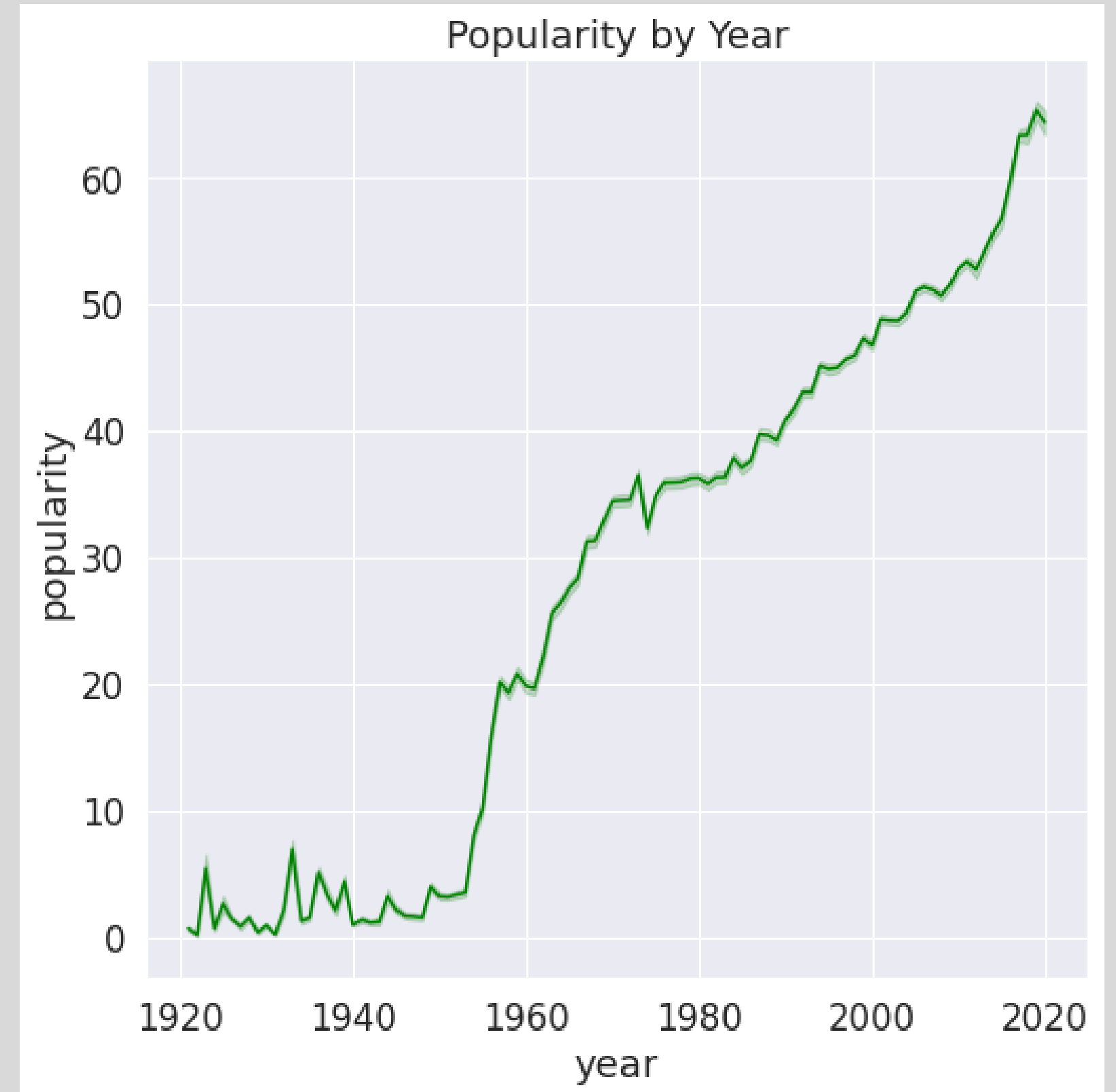


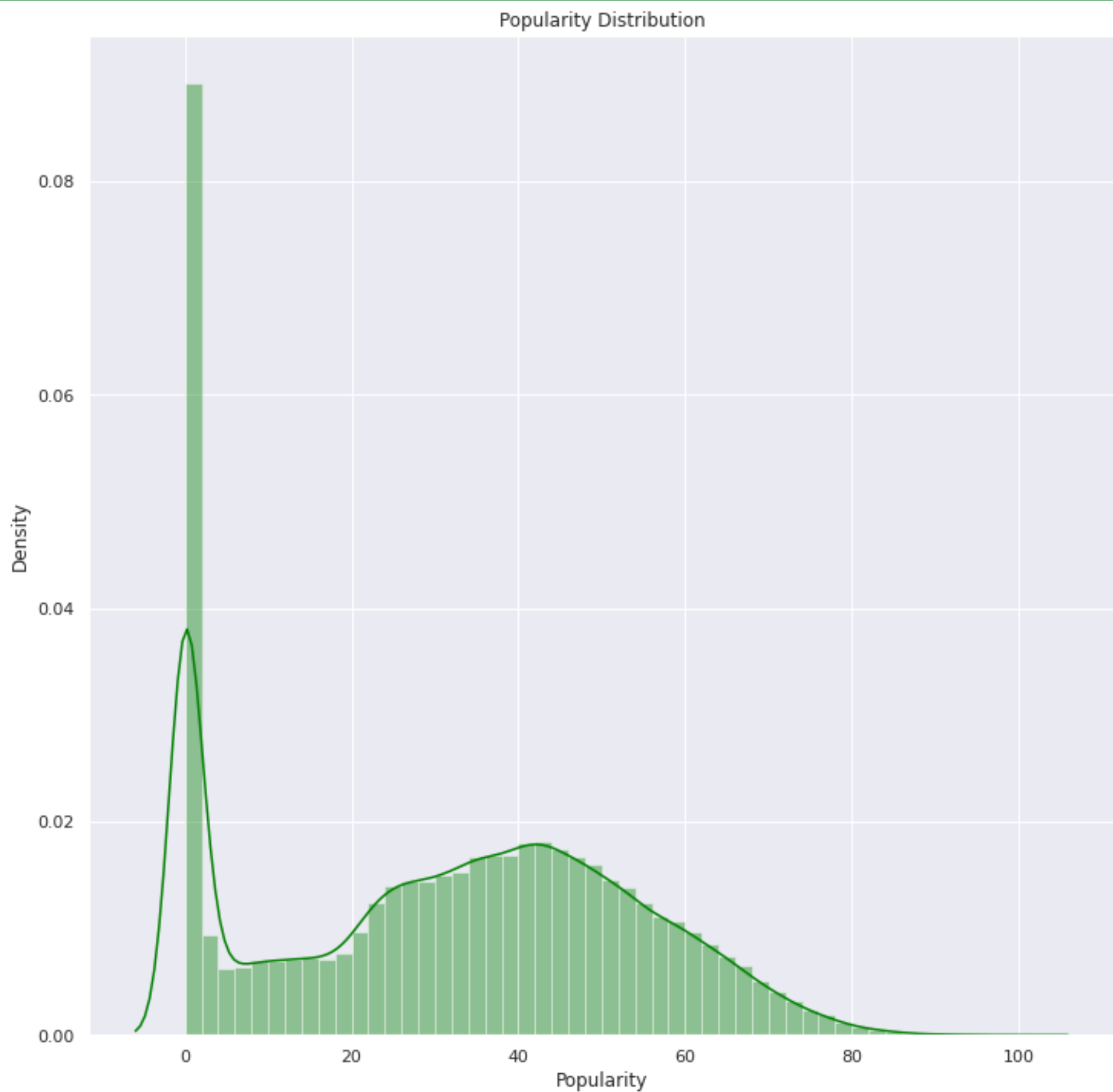
# Correlation and Popularity



# How is Popularity Measured?

- Total number of plays the track has had and how recent those plays are so natural year popularity is heavily influenced by year.
- The newer years have higher count of more recent plays



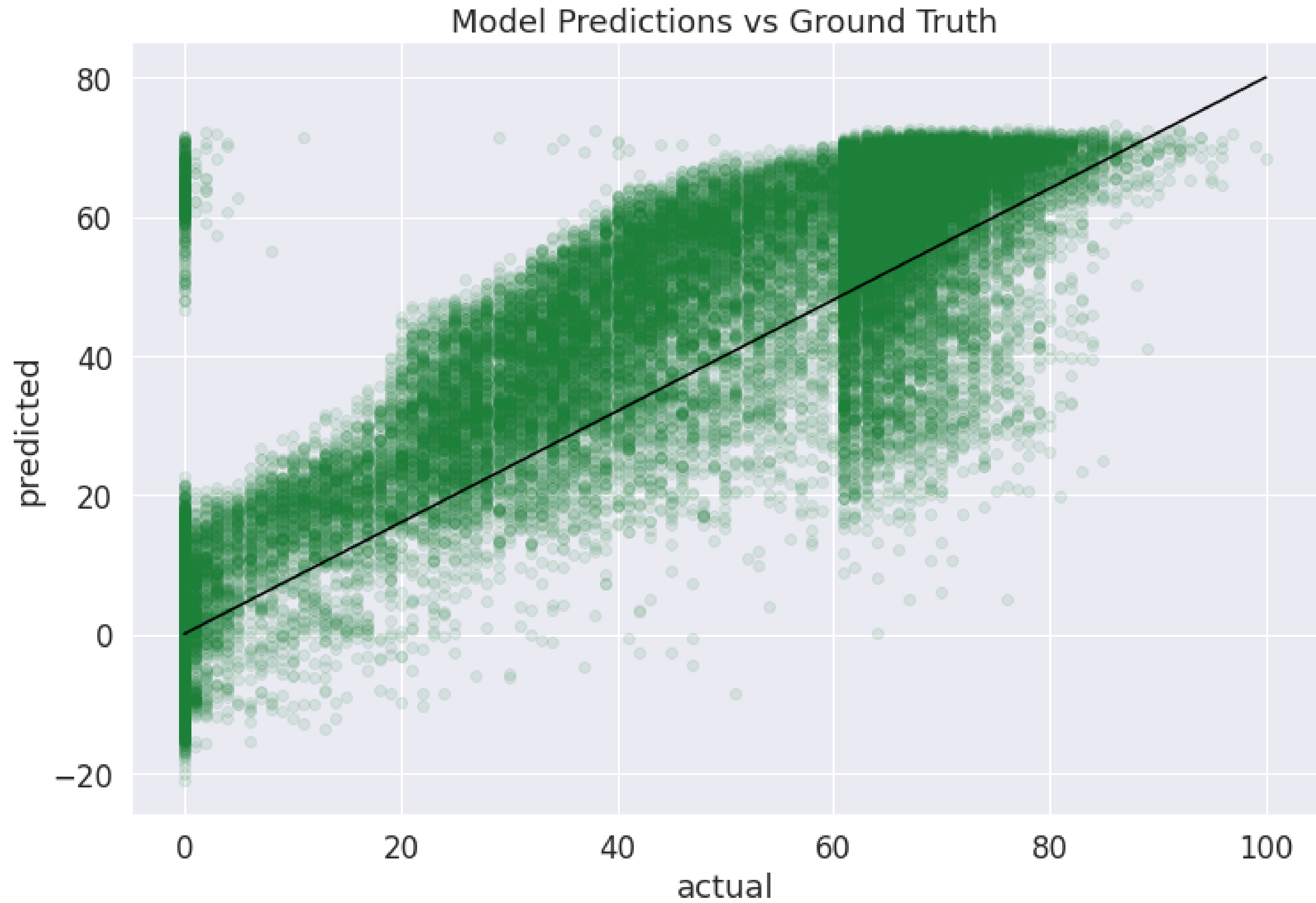


# POPULARITY DISTRIBUTION

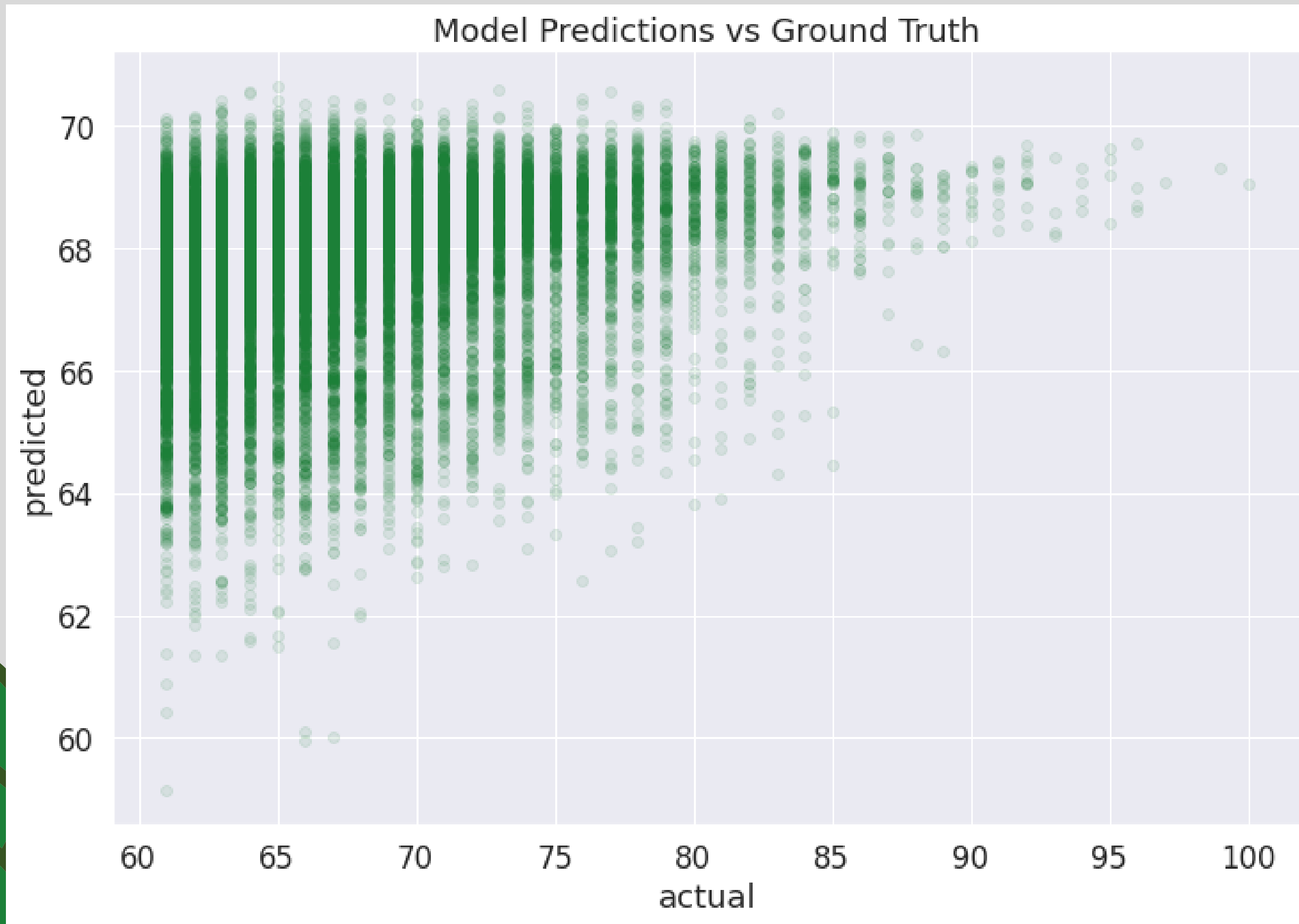
Overwhelming frequency of 0's, and greater than 75 are considered outliers

# LINEAR REGRESSION MODEL

MAE = 8.7449  
 $R^2 = 0.8223$



WHY NOT HIGHER?





# RECOMMENDATION MODEL

# RECOMMENDATIONS

```
X=df[['valence','year','acousticness' , 'danceability', 'energy', 'explicit','liveness','loudness','mode','speechiness']]
y = df['name']
neigh = NearestNeighbors(n_neighbors=10)
neigh.fit(X,y)
ind=neigh.kneighbors(X.iloc[[567]])[1]
for i in ind:
    print(df.iloc[i][['name', 'artists']])
```

## Cardigan by Don Toliver



	name	artists
57830	When You Love Me	['Memphis Minnie']
39126	New Orleans Stop Time	['Memphis Minnie']
1591	Pálida Noche - Instrumental (Remasterizado)	['Francisco Canaro']
39299	No Seas Malita - Instrumental (Remasterizado)	['Francisco Canaro']
20628	No Seas Amlita	['Francisco Canaro']
75774	Quand Il Joue De L'accordéon	['Fréhel']
39184	Farolito de papel - Instrumental (Remasterizado)	['Francisco Canaro']
57871	En la Trampa - Remasterizado	['Francisco Canaro', 'Charlo']
20610	Flores Secas - Remasterizado	['Francisco Canaro', 'Charlo']
75666	La Guardia Vieja - Instrumental (Remasterizado)	['Francisco Canaro']

# Suggestions

- Add more songs above 70 popularity
- Look for songs with more energy and loudness
- Less acoustic songs