Analyzing Spotify Data

James Brenner, Hannah Summers, Sidhdhi Patel, Chloe Petitpas

The Business of Music: What Drives a Hit Song?

What components of a song contribute to its popularity and success?

- Music industry → records labels, artists, streaming services
- Strategically creating songs that will have an increased likelihood of success
- Allocating resources to market songs that have a higher potential for popularity
- Enhance recommendations to users on streaming platforms

Data Background

Data Collection:

- Spotify Web API
- Music from 1921 to 2020
- Last updated April 2021

Dependent Variable: Popularity

Independent Variables:

- Danceability
- Speechiness
- Acousticness
- Explicit Content
- Valence
- Instrumentalness
- Tempo
- Key
- Liveness
- Duration
- Energy
- Year

Variable	Mean	Standard Deviation	Min	Max
Popularity	26.60	22.10	0.00	100.00
Tempo (BPM)	117.09	30.54	0.00	243.51
Loudness (dB)	-11.67	5.67	-60.00	3.86
Duration (ms)	203,137	63,796	4,937	360,000
Year	1976.30	26.82	1920	2021
Danceability	0.54	0.17	0.00	0.99
Energy	0.48	0.27	0.00	1.00
Acousticness	0.51	0.38	0.00	1.00
Explicit (0/1)	0.07	0.26	0.00	1.00
Instrumentalness	0.18	0.32	0.00	1.00
Key (0-11)	5.21	3.51	0.00	11.00
Liveness	0.20	0.15	0.00	0.80
Speechiness	0.11	0.19	0.00	0.97
Valence	0.54	0.26	0.00	1.00

Research Model

Which research model?

- Ordinary Least Squares Regression
- Examines the relationship between the independent and dependent variables

Why?

- This allows for multiple independent variables
- Can quantify results through coefficients
- Can analyze predictive capabilities of the model

In the end this allows artists and producers to see general trends in what increases and decreases popularity

Results Table

Variable	Coefficient	Std Error	P-Value
Intercept	-668.7932	4.575	0.0
tempo	-0.0048	0.001	0.001
duration_ms	2.656e-05	7.34e-07	0.0
year	0.3539	0.002	0.0
danceability	0.7018	0.336	0.037
energy	-3.635	0.276	0.0
acousticness	-6.8637	0.198	0.0
explicit	8.8526	0.188	0.0
instrumentalness	-13.0836	0.142	0.0
key	-0.0469	0.012	0.0
liveness	-9.0378	0.308	0.0
speechiness	-19.2318	0.289	0.0
valence	2.9826	0.23	0.0

R-squared (0.415)

Interpretation: The R-squared value indicates that 41.5% of the variance in the popularity variable can be explained by the predictors in the model. This means that while the model explains a significant portion of the variability in popularity, there is still a large portion of variability (58.5%) that is unexplained by the current set of predictors.

Implication: The model is somewhat effective, but there may be additional factors influencing popularity that are not included in this analysis.

Results

Individual Variable Results:

- **Explicit**: Explicit content is a strong positive predictor of popularity.

- **Year**: Newer songs are more likely to be popular.

- **Energy**: Lower energy (slower, more mellow tracks) seems to decrease popularity.

- **Danceability**: More danceable songs are associated with higher popularity.

- **Instrumentalness**: More instrumental songs tend to be less popular.

- **Accousticness:** Songs that are more acoustic are generally less popular.

Results

Individual Variable Results (continued):

- **Tempo:** For each 1-unit increase in tempo, popularity decreases by 0.0048, holding other variables constant
- **Duration (milliseconds):** For every additional millisecond in a song popularity score increases by 0.00002656
- **Key:** More instrumental songs tend to have lower popularity decreasing by about 13 per unit increase
- **Liveness:** Songs with higher liveness scores tend to be less popular
- **Speechiness:** Songs with more spoken content tend to be significantly less popular
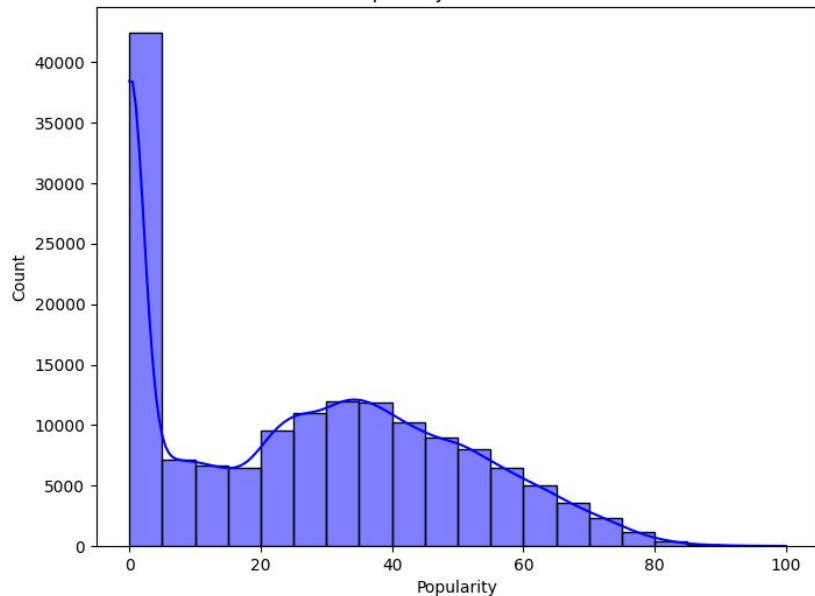- **Valance:** Songs that are more positive (i.e. higher valance) tend to be more popular.

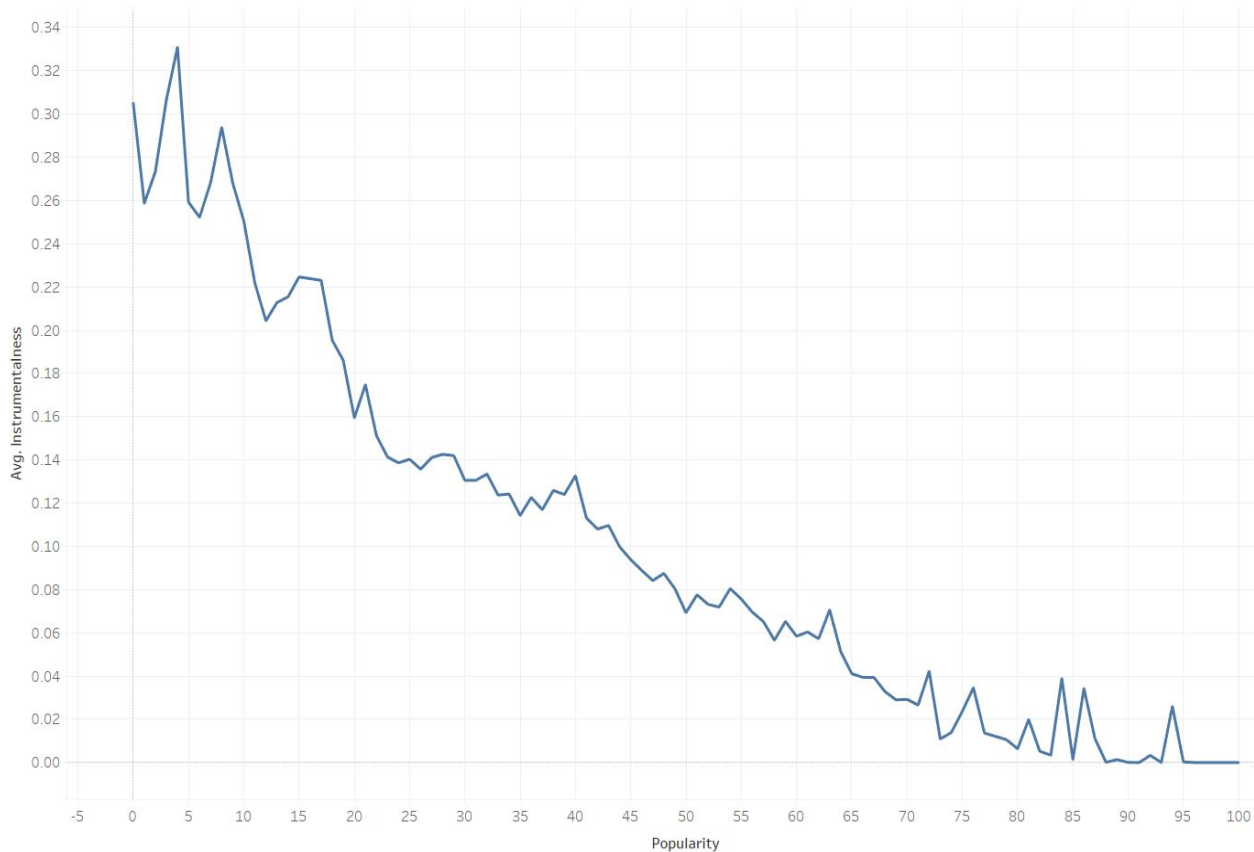# Visualization Before Analysis

Feature Correlation Heatmap
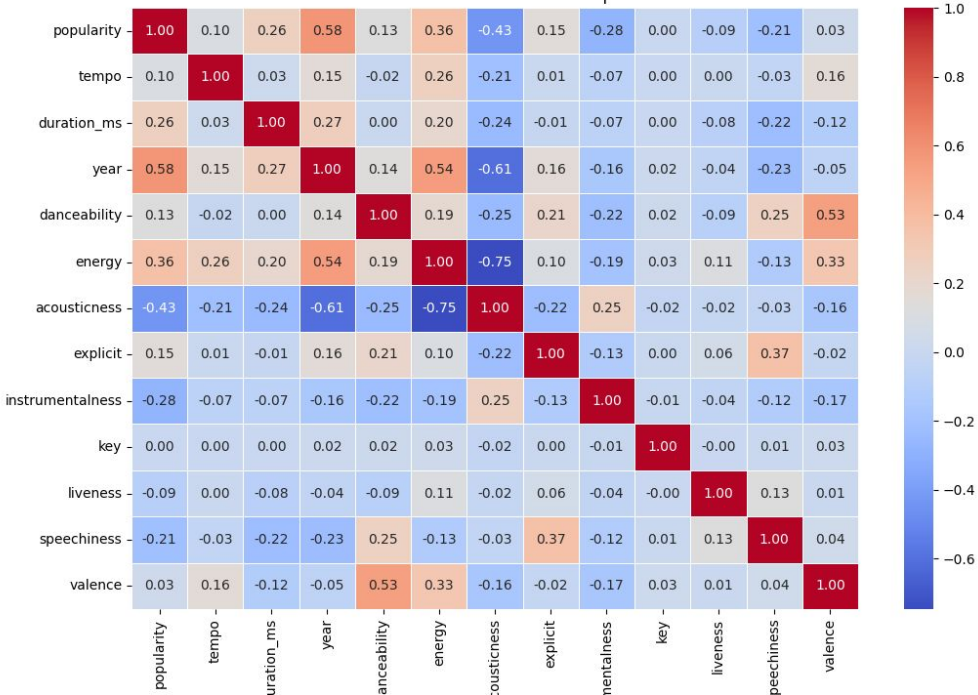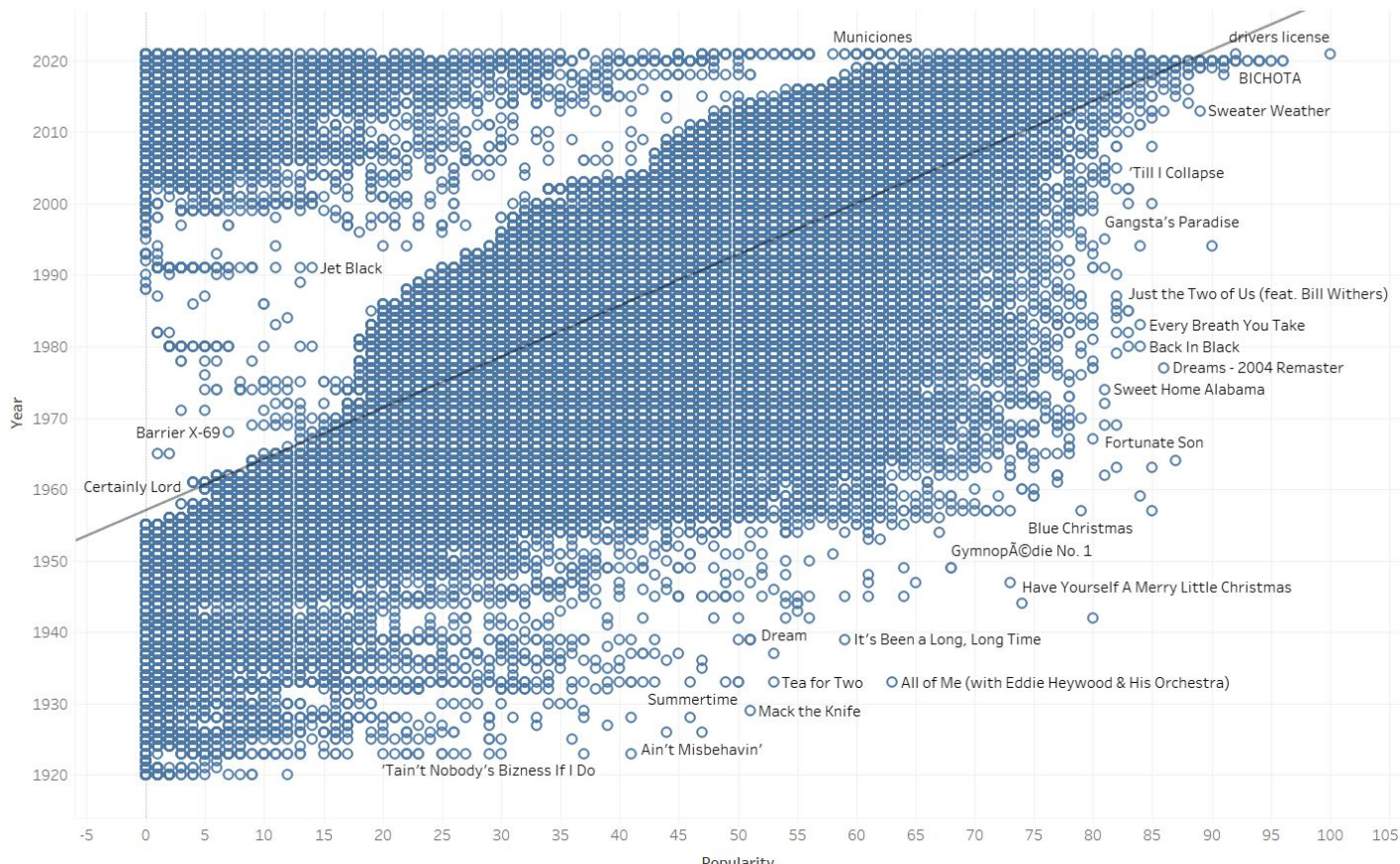

Popularity Distribution

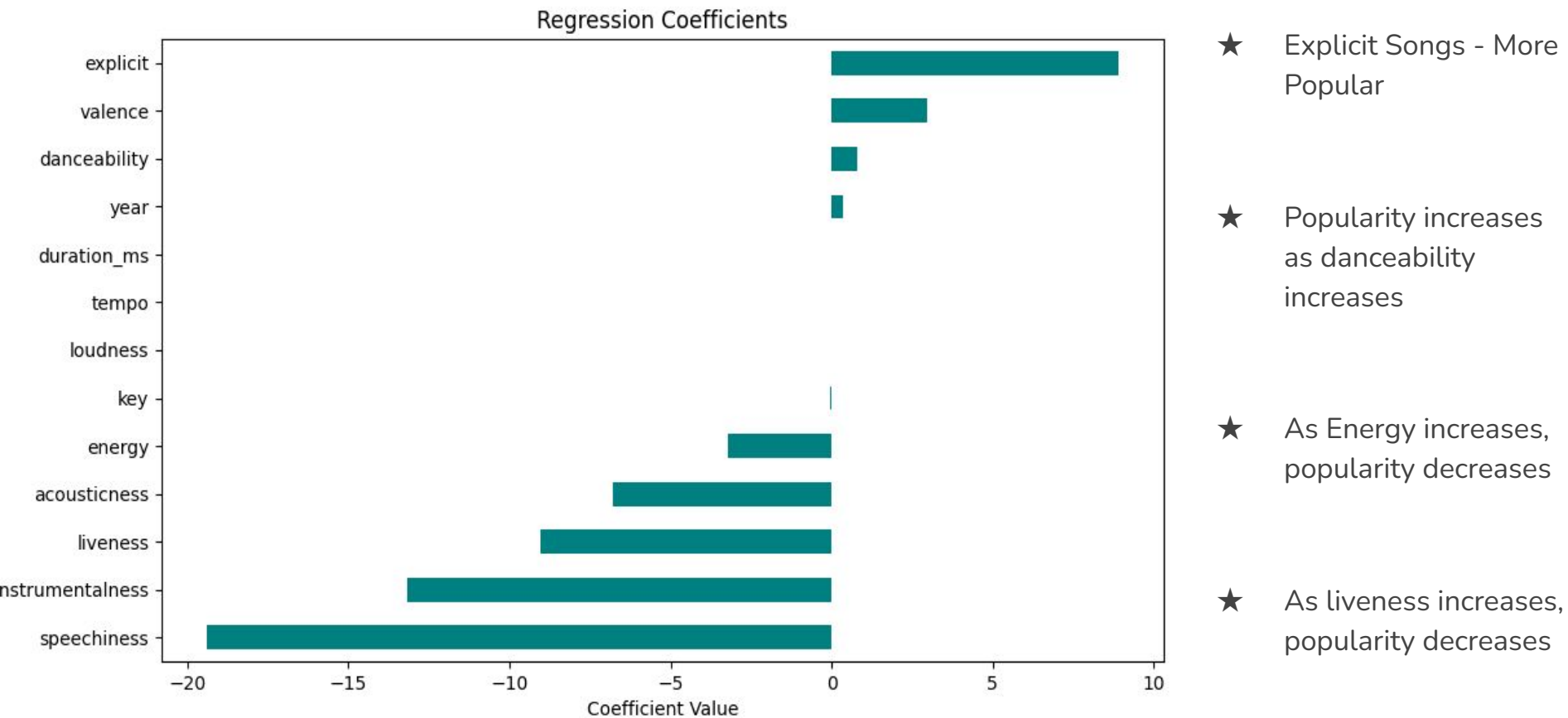# Visualization After the Analysis

Instrumentalness vs Popularity

Visualization After the Analysis

Year vs Popularity

Visualization After the Analysis

Regression Coefficients

★ Explicit Songs - More Popular

★ Popularity increases as danceability increases

★ As Energy increases, popularity decreases

★ As liveness increases, popularity decreases

How Our Results Can Shape the Music Industry

Increase in Popularity:

- Danceability
- Explicit Content
- Valence (Happier/Positive)

Decrease in Popularity:

- Instrumentalness
- Low Energy
- Speechiness
- Liveness

Record Labels and Artists:

- Produce songs that are danceable & positive
- Consider adding explicit content for engagement
- Avoid excessive instrumentals & speech-heavy songs
- Promote songs that align with popular characteristics

Streaming Services:

- Optimize algorithm to prioritize positive & danceable songs
- Curate playlists with danceable, explicit, and positive tracks

Thank You!