Exploring Spotify Songs: Popularity and Danceability

C4: Dat Nguyen, Hsin-Pan Chen, Zhengyuan Pei, Xiaoyu Ma, Victoria Nguyen, Chia-Chun Hung, Yang-Hsuan Lin
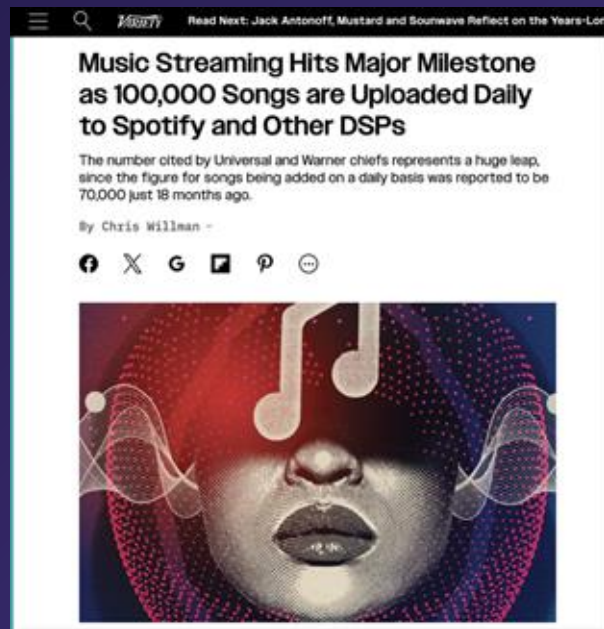
Spotify Song   ♥ |◀ ◀◀ ❚❚ ▶▶ ▶| ●────────────   👤

Why a song takes off?

"100,000+" (Daily Uploads)

Songs that can get more Attention
Will stands out

But how?

Danceability? Valence? Speechiness?
Our Guess: **Danceability** !

Music Streaming Hits Major Milestone as 100,000 Songs are Uploaded Daily to Spotify and Other DSPs

The number cited by Universal and Warner chiefs represents a huge leap, since the figure for songs being added on a daily basis was reported to be 70,000 just 18 months ago.

By Chris Willman –

Spotify Song

Then How to answer this?

- **Testable hypothesis:**

  Songs with higher danceability will have higher popularity

- **Two metrics from Spotify:**

  Danceability - Independent Variable

  Popularity - Dependent Variable

Ideal Experiment: Randomized Playlist Exposure

User **A**	Random Assignment **B**	Danceability Playlists **C**	Compare Popularity **D**

- **Random Assignment:** Users are randomly assigned to playlists made of high/low and mixed danceability.

- **Standardize exposure:** Artist popularity, release date, exposure, algorithmic visibility

- **Control for confounder:** Hold key factors (energy, valence, tempo, artist popularity) constant so only danceability varies.

Limitations

Observational data limits causal inference:

Because songs aren't randomly assigned and Spotify's algorithm shapes exposure, our regression shows correlations, not causal effects.

Danceability's Causal Effect Is Unknown:

Missing factors like marketing and playlist placement mean the coefficient can't reflect danceability's true causal impact.

The results have limited generalizability:

Spotify's opaque scoring and recommendation systems mean our patterns may not generalize across genres, users, or contexts.

Spotify Song   ♥   ⏮ ⏪ ⏸ ⏩ ⏭

Data Description & Structure

Dataset

- **Source:** Spotify Features.csv (Website: Kaggle, Author: Somu Mourya, 2023)
- **Observation unit:** Each row = one Spotify track
- **Size:** 232,726 songs, across multiple genres and artists
- **Goal:** Investigate how a song's danceability relates to its popularity on Spotify

Aggregation Level

- No aggregation by artist or album
- Each observation represents one unique song (track-level)

Type	Variable	Description
Main IV (X)	Danceability	Describes how suitable a track is for dancing (0.0 = least, 1.0 = most danceable), reflecting rhythm stability, beat strength, and overall regularity.
Main DV (Y)	Popularity	Indicates a song's popularity (0–100), with higher values representing more frequently played tracks.
Controls	Energy, Valence, Loudness, Tempo, Acousticness, Instrumentalness, Genre	Capture other aspects of sound and style that may also affect popularity.

| Variable	Mean	Median	SD	Min	Max
Popularity	41.13	43	18.19	0	100
Danceability	0.55	0.57	0.19	0.06	0.99

- The average popularity score is around 41, showing most songs fall in the midrange of Spotify popularity
- On average, songs have a moderate danceability
- The relatively high standard deviation of popularity (18.2) indicates strong variation across tracks

Spotify Song

♥   ⏮ ⏪ ⏸ ⏩ ⏭   👤

Popularity

Right-skewed — many songs have mid-level popularity (30–60), while only a small number reach very high scores


Distribution of Popularity
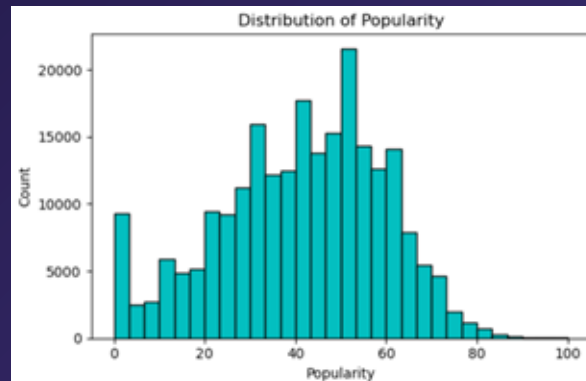
Danceability

Roughly bell-shaped and centered around 0.5 – 0.6. Suggests Spotify's catalog contains mostly moderately danceable songs rather than extremes

Distribution of Danceability

Correlation Matrix

Popularity vs. Danceability
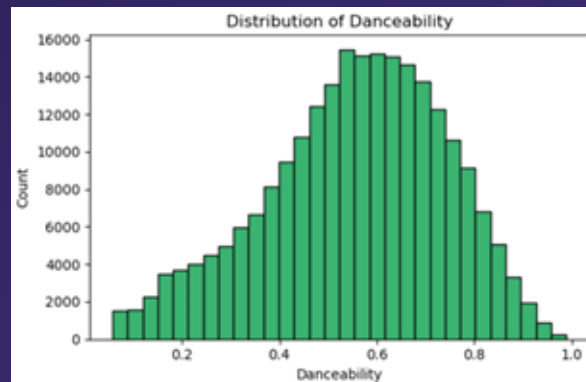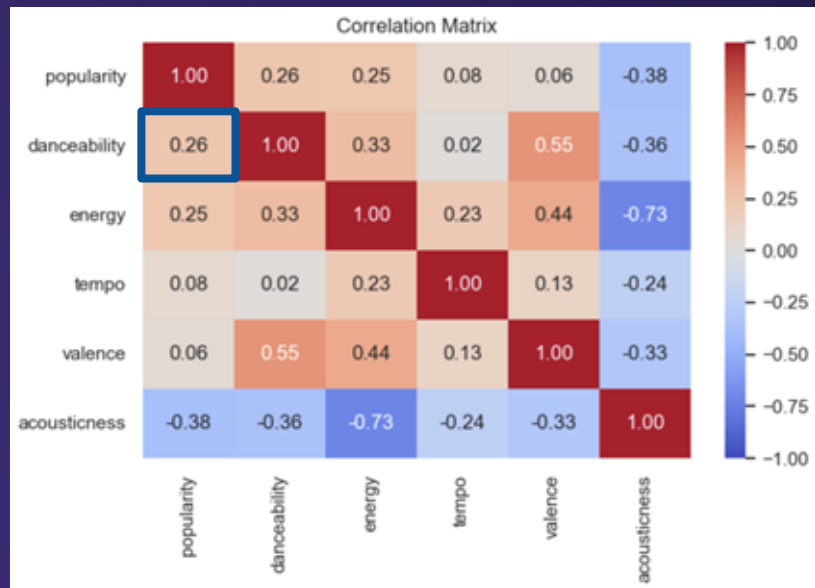
- Modest positive relationship (r = 0.26)
- Popularity is influenced by multiple song features including energy and valence
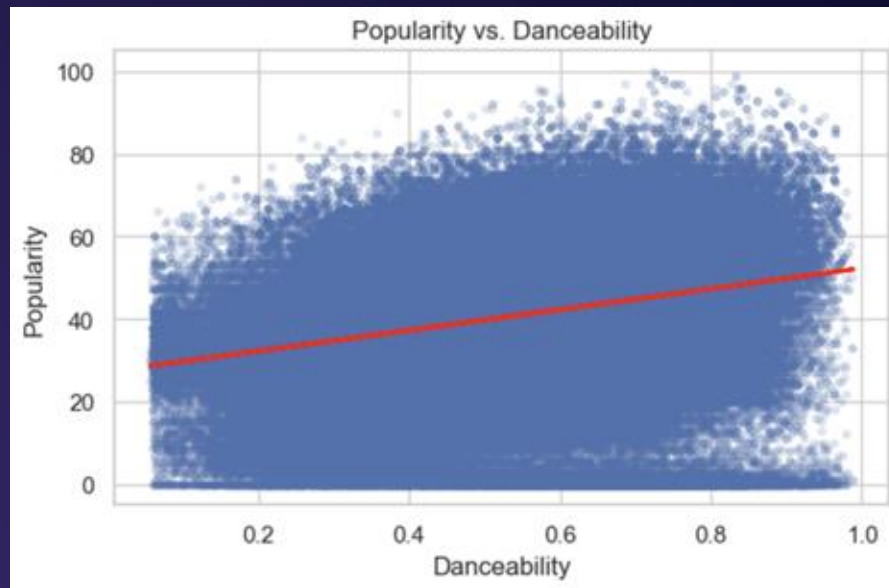
- Upward-sloping trend line
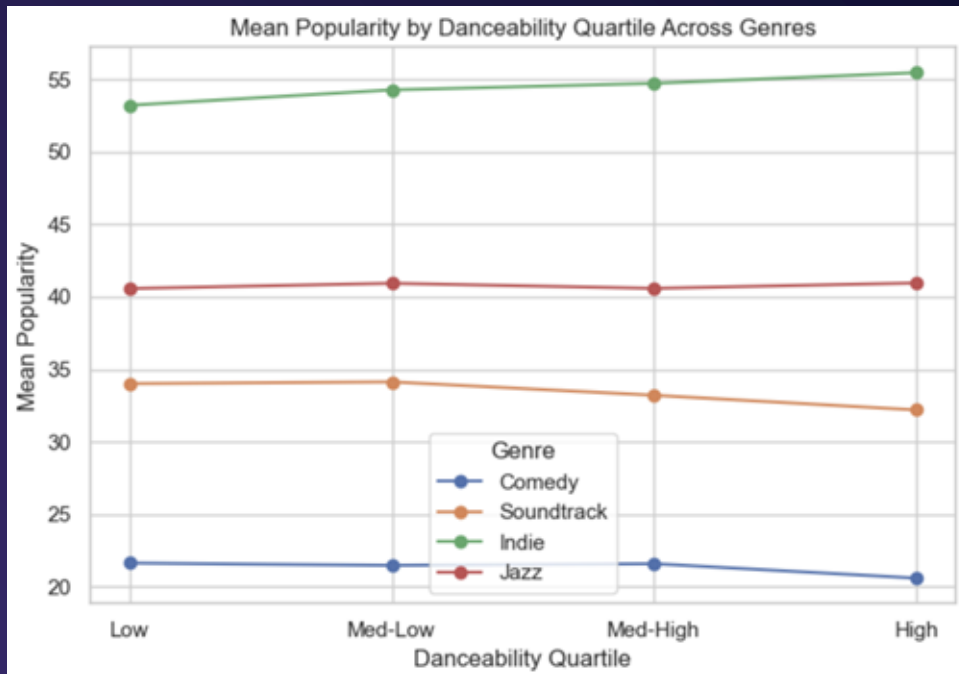- Wide vertical spread shows many other factors also influence popularity

Popularity Across Danceability Quartiles

- Median popularity rises consistently from low to high danceability groups
- Large variation within each quartile indicates that danceability alone does not fully determine popularity

Popularity Across Danceability Quartiles
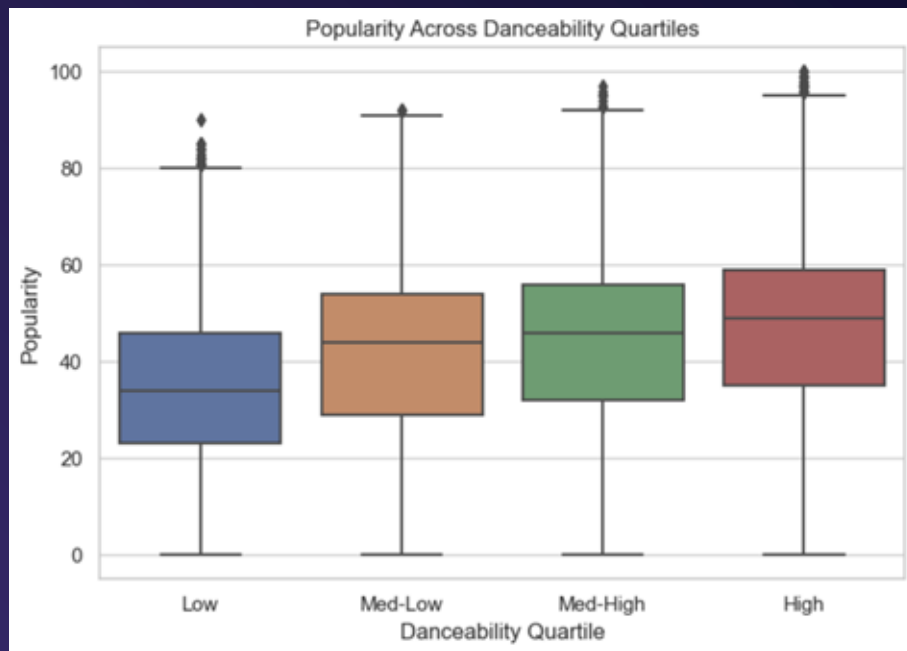
Genre Differences

- Patterns differ by genre — Indie and Jazz show a positive trend, while Soundtrack and Comedy show negative patterns.
- Different genres different in energy, valence, and tempo
- Danceability does not influence all genres same

Mean Popularity by Danceability Quartile Across Genres

Spotify Song

Multivariable Regression Model

```
                              OLS Regression Results
===============================================================================
Dep. Variable:           popularity   R-squared:                      0.213
Model:                          OLS   Adj. R-squared:                 0.213
Method:               Least Squares   F-statistic:                 1.052e+04
Date:             Tue, 02 Dec 2025   Prob (F-statistic):              0.00
Time:                     14:24:51   Log-Likelihood:             -9.7740e+05
No. Observations:           232725   AIC:                         1.955e+06
Df Residuals:               232718   BIC:                         1.955e+06
Df Model:                        6
Covariance Type:          nonrobust
```

```
===============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------
const         61.5806      0.321    191.674      0.000      60.951      62.210
danceability  15.8952      0.233     68.214      0.000      15.439      16.352
energy       -15.6161      0.249    -62.654      0.000     -16.105     -15.128
valence      -11.3418      0.164    -69.031      0.000     -11.664     -11.020
tempo          0.0013      0.001      1.111      0.267      -0.001       0.003
acousticness -16.1897      0.144   -112.507      0.000     -16.472     -15.908
loudness       0.9791      0.010     93.869      0.000       0.959       1.000
===============================================================================
Omnibus:                   3210.158   Durbin-Watson:                  0.443
Prob(Omnibus):                0.000   Jarque-Bera (JB):            3300.652
Skew:                        -0.284   Prob(JB):                        0.00
Kurtosis:                     2.864   Cond. No.                    1.49e+03
===============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.49e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Spotify Song

Prediction Interval Analysis: Row 12345
Track: "Submerge" by Movements

Key Features:
- Danceability: 0.342
- Energy: 0.45
- Valence: 0.115
- Acousticness: 0.0095
- Loudness: -8.665 dB
- Musical Structure
- Tempo: 152.308 BPM
- Key: C#
- Mode: Major
- Time Signature: 3/4

Predicted: 50.2

18.6   Error: -8.24   81.9

Actual: 42

Popularity Score

Spotify Song
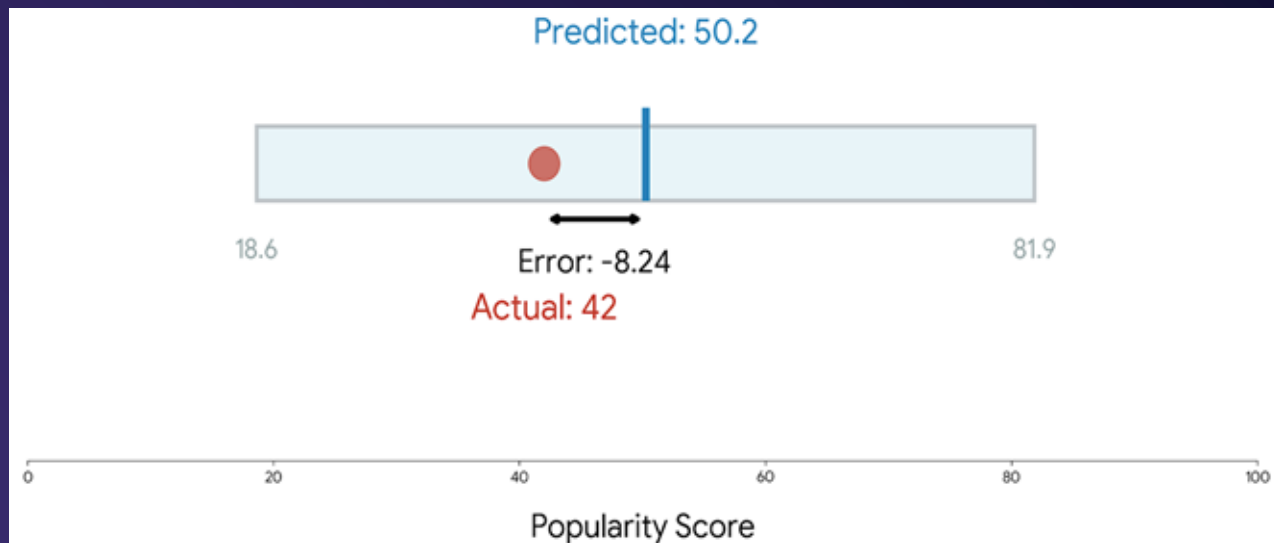
Conclusion

- Higher danceability, higher popularity.
- Danceability alone does not explain popularity.
- Popularity is multi-factor: energy, tempo, valence, and genre also influence outcomes.
- Danceability boosts a song's chances but does not guarantee a hit. A song still needs the right emotional and musical context to truly break through.

Spotify Song   ❤   ⏮ ⏪ ⏸ ⏩ ⏭