Spotify V/S Youtube

Spotify vs YouTube – Cross-Platform Popularity

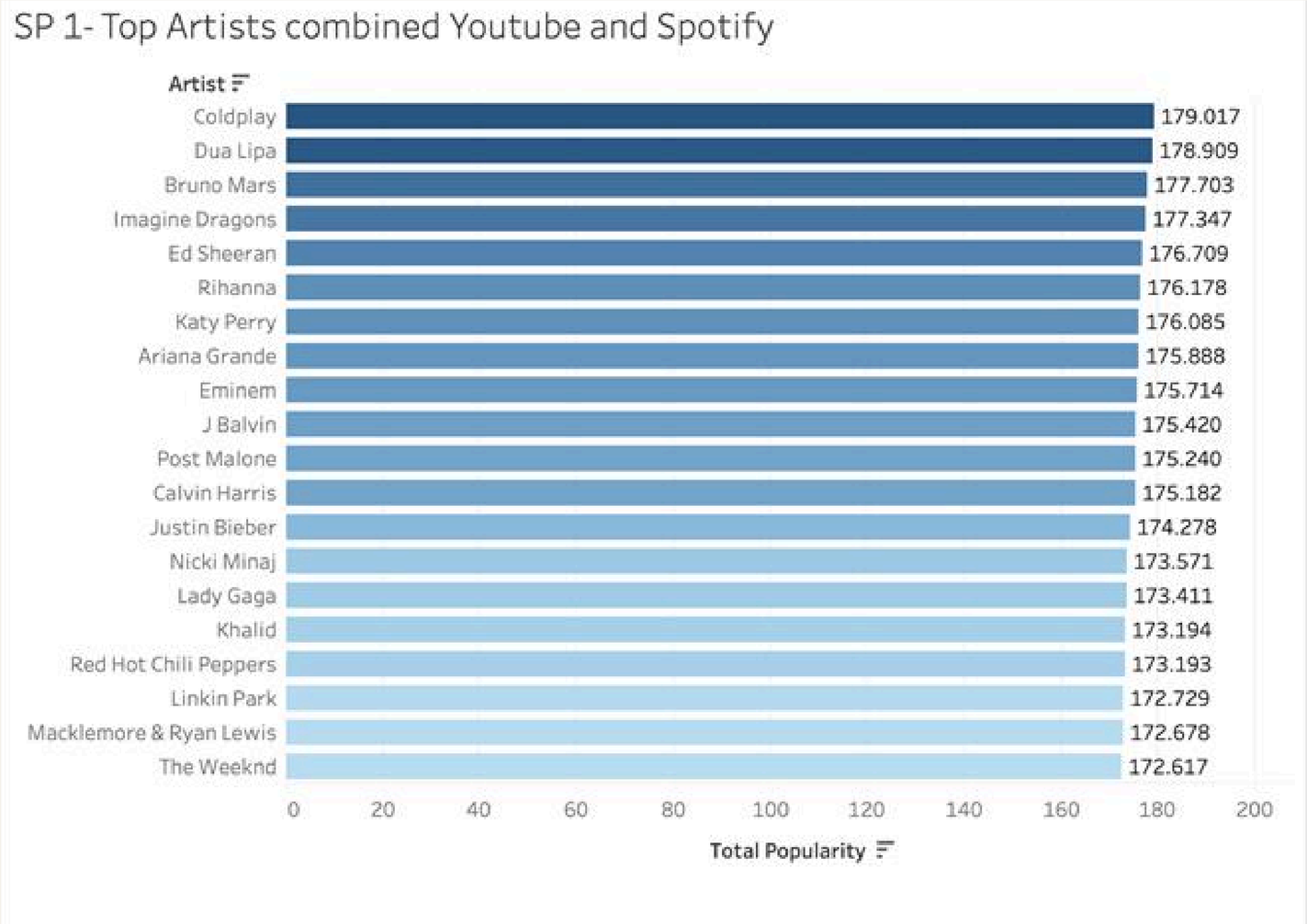
Compare how songs perform on Spotify vs YouTube using the spotify_youtube dataset.

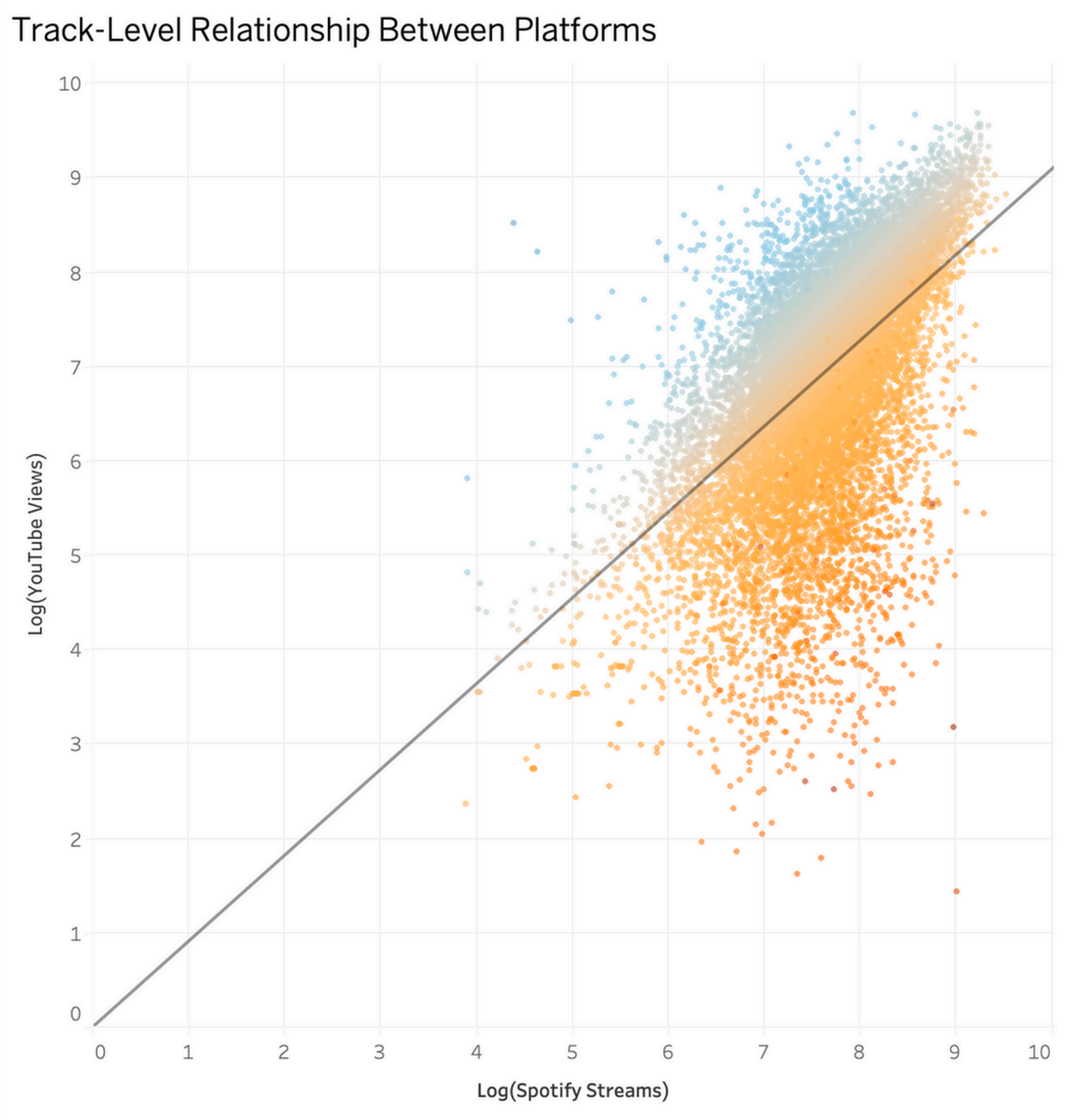
- **Main questions:**
  - How closely are streams and views related across platforms?
  - Which audio features (danceability, energy, etc.) are associated with higher popularity?
  - Which tracks or artists are platform-specific outliers (YouTube-first vs Spotify-first)?

- **Purpose**: Provide actionable insight for labels/artists on where to focus promotion (video vs audio, which types of songs travel best across both platforms).

Top Artist Combined - YouTube and Spotify

- Aggregated data to the artist level using total popularity (log streams + log views).
- Top 20 artists by combined popularity include Coldplay, Dua Lipa, Bruno Mars, Imagine Dragons, and Ed Sheeran.
- All Top 20 have very similar scores (~172–179), showing a tight, concentrated market.
- Artist-level correlation between avg log views and avg log streams is ~0.64, indicating strong cross-platform alignment.



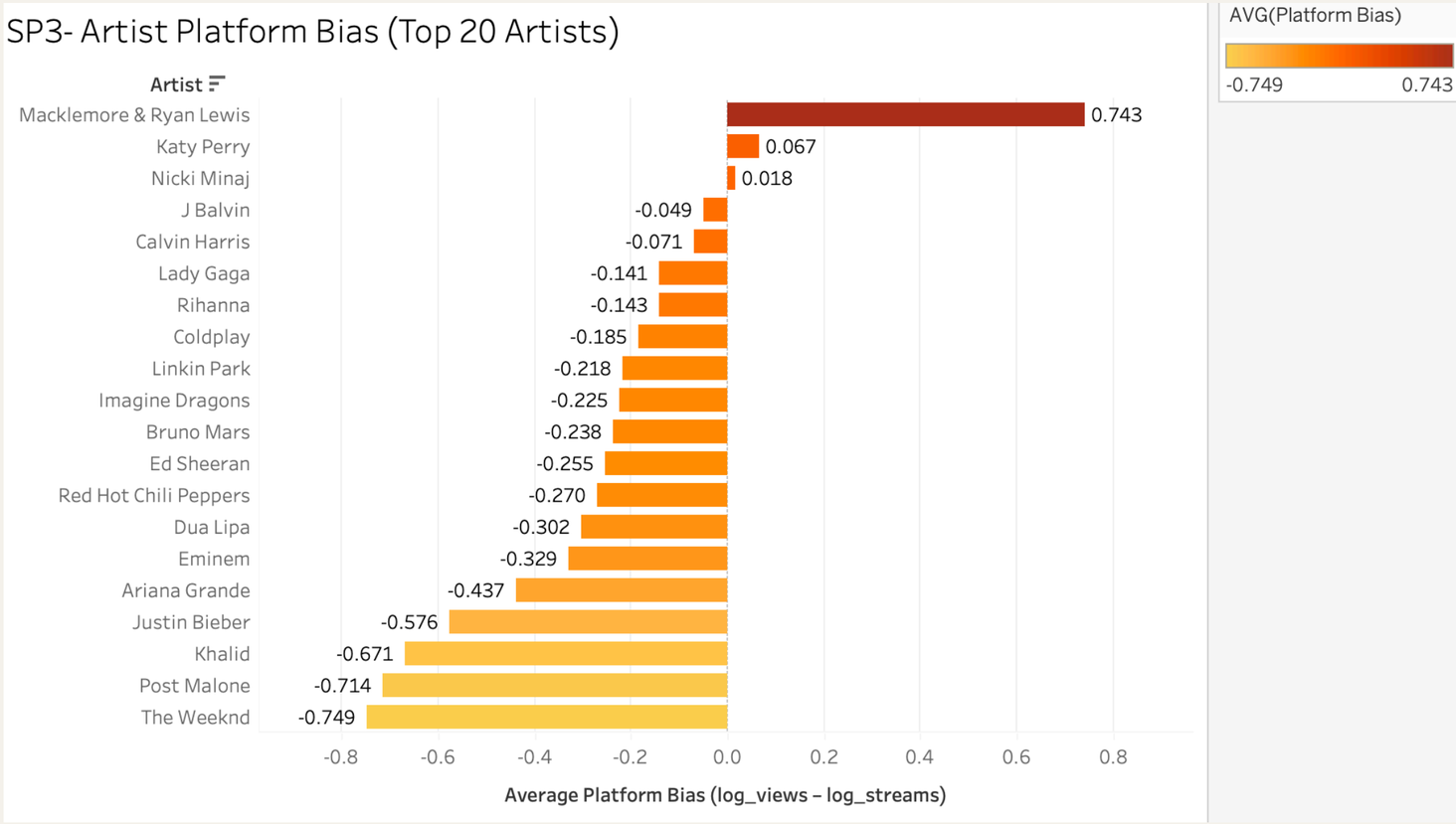


Track-Level Relationship Between Platforms

- Each point = one track, plotted as log_streams (Spotify) vs log_views (YouTube).
- Overall trend is clearly upward – more streams are associated with more views.
- Points colored by platform_bias (blue = Spotify-biased, orange/red = YouTube-biased)
- Points above the diagonal line are YouTube-heavy tracks.
- Points below are Spotify-heavy tracks.

Platform Bias for Top Artists

- **Platform bias** = log_views – log_streams
- **Positive** → YouTube-heavy, **Negative** → Spotify-heavy
- Most of the Top 20 artists are slightly Spotify-heavy (small negative bias)
- YouTube-heavy standouts: Macklemore & Ryan Lewis, Katy Perry, Nicki Minaj
- Spotify-heavy standouts: The Weeknd, Post Malone, Khalid
- **Takeaway**: Even top artists don't split evenly across platforms



Audio Features & Popularity

- Features that help popularity (**positive coefficients**):
  1. Danceability – biggest driver, especially for YouTube (+0.56 vs +0.11 on Spotify).
  2. Loudness – small positive effect on both platforms.

- Features that hurt popularity (**negative coefficients** on both):
  1. Speechiness (≈ −1.0 YT, −0.9 SP), Instrumentalness, Acousticness, Liveness, Valence → spoken-word, Instrumental, acoustic, or "live" tracks under-perform.

- Upbeat, loud, danceable vocal songs travel best across both platforms; acoustic / instrumental / speech-heavy tracks are less scalable.

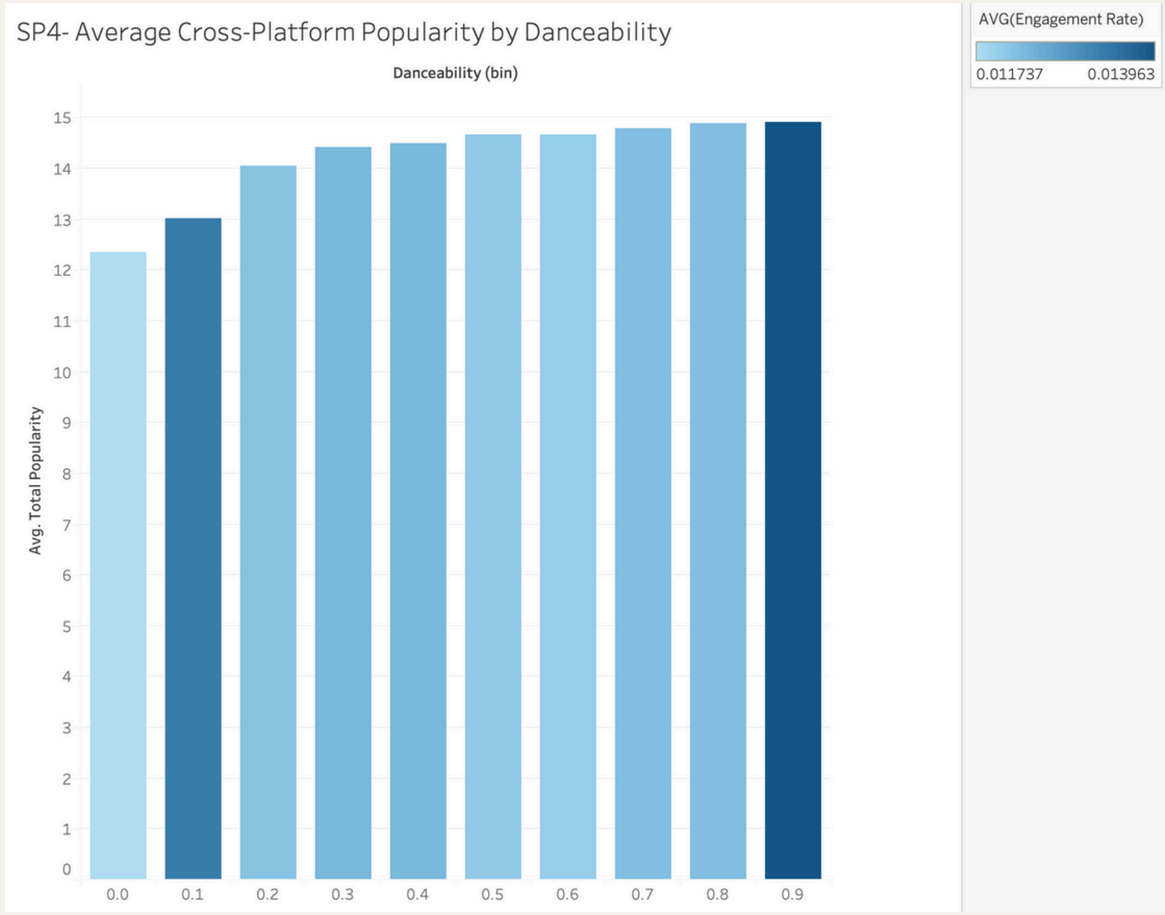
- Being extremely high-energy or happy on top of already danceable/loud doesn't give extra lift.

```
CORRELATION OF AUDIO FEATURES WITH log_views AND log_streams:
                  log_views  log_streams
log_views          1.000000     0.591648
log_streams        0.591648     1.000000
Danceability       0.157783     0.057218
Energy             0.191339     0.061898
Loudness           0.306468     0.162259
Speechiness       -0.051750    -0.117811
Acousticness      -0.174027    -0.115895
Instrumentalness  -0.238646    -0.129046
Liveness          -0.004712    -0.050411
Valence            0.095699    -0.001190
Tempo              0.050138     0.030694
Duration_ms        0.067780     0.019553

REGRESSION COEFFICIENTS FOR log_views (YouTube):
Danceability: 0.5600
Energy: -0.3391
Loudness: 0.0699
Speechiness: -1.0292
Acousticness: -0.1171
Instrumentalness: -0.5680
Liveness: -0.0820
Valence: -0.1204
Tempo: 0.0009
Duration_ms: 0.0000

REGRESSION COEFFICIENTS FOR log_streams (Spotify):
Danceability: 0.1142
Energy: -0.4662
Loudness: 0.0302
Speechiness: -0.9163
Acousticness: -0.2475
Instrumentalness: -0.2130
Liveness: -0.1694
Valence: -0.1401
Tempo: 0.0005
Duration_ms: 0.0000
```

Danceability & Cross-Platform Popularity



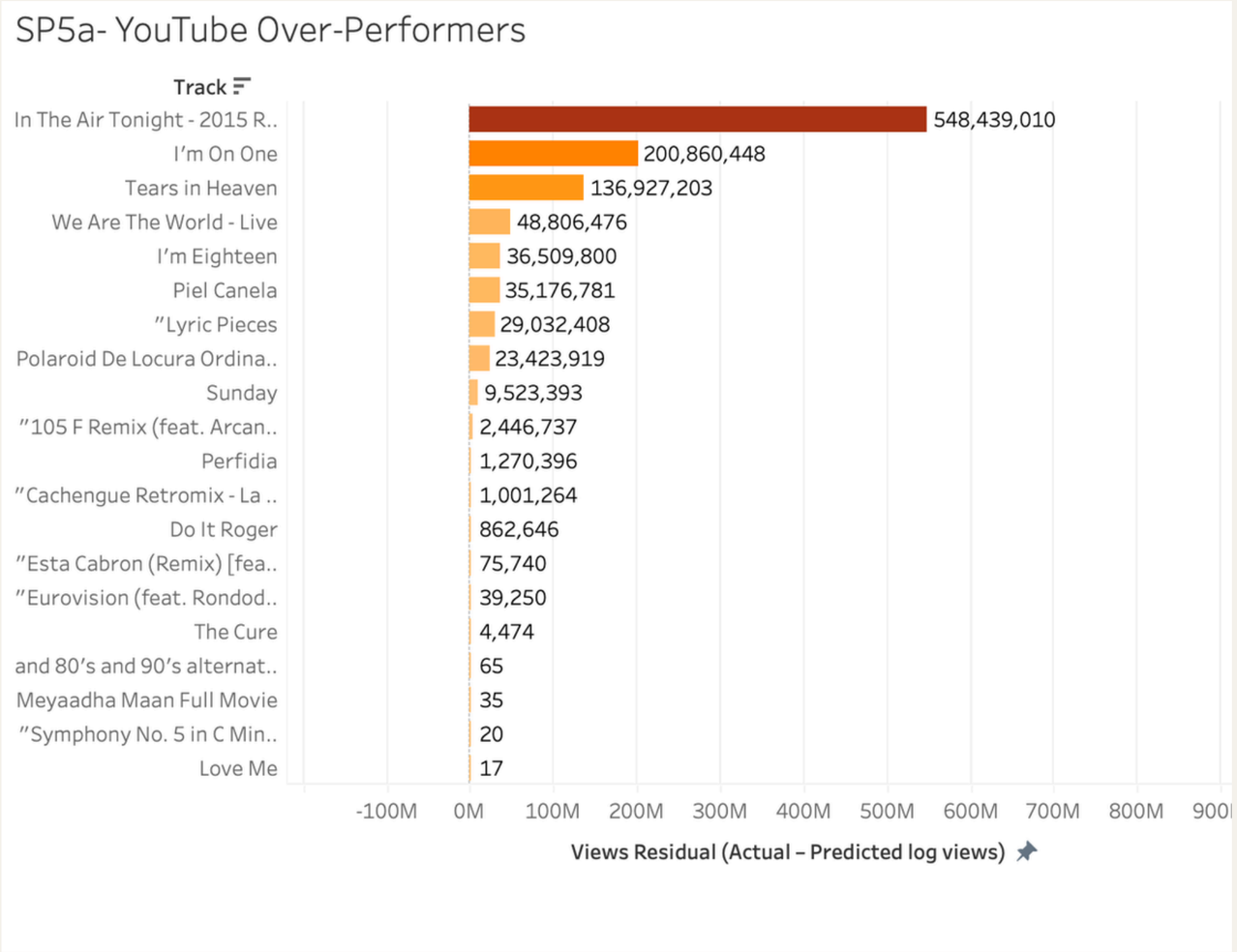
- Popularity increases as danceability rises.
- Engagement rate (color) is reasonably stable but slightly higher in some mid-bins.
- Danceability correlates positively with both log_views (0.16) and log_streams (0.06).
- In multivariate regression, Danceability has strong positive coefficient for both Youtube(~0.56) and Spotify (~0.11).
- Some features like Instrumentalness, Acousticness and Speechiness are negatively associated with popularity.

Python Model: log_views ~ log_streams

- Slope close to 1 ⇒ a 1% increase in Spotify streams is associated with roughly a 1% increase in YouTube views on a log scale.
- We use this model to compute views_residual = actual log_views − predicted log_views, which measures over- or under-performance on YouTube given Spotify performance.
- This residual is the basis for identifying YouTube over-performers (SP5a).

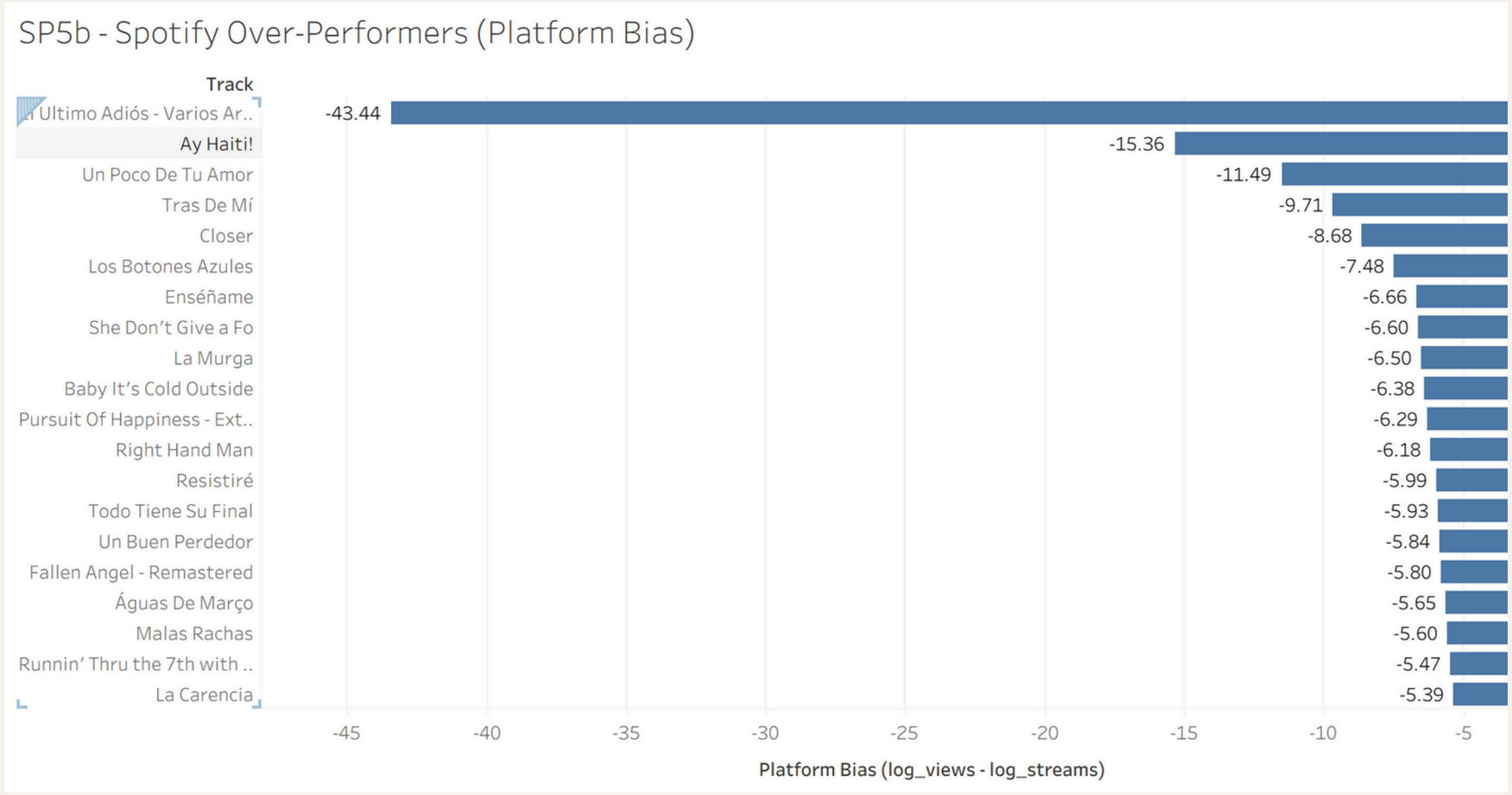
YouTube Over-Performers (residuals)

These are the songs that punch above their weight on YouTube. Their Spotify numbers are not bad, but the regression says their YouTube views are way higher than we'd expect. From a business view, these are video-first tracks where investing in YouTube promotion makes a lot of sense.



Spotify Over-Performers (residuals)

On the flip side, these songs are clearly Spotify-heavy. They stream really well, but their YouTube presence lags the model's expectation. They behave more like 'audio' experiences than video products



Final Takeaways

- Cross-platform alignment: Track-level and artist-level correlations between Spotify and YouTube popularity are around 0.6, so hits generally succeed on both platforms.
- Feature-driven success: Danceability, loudness and (to some extent) energy are positively associated with popularity, while instrumental or highly acoustic tracks under-perform.
- Platform-specific outliers: Using residuals and platform bias, we identify: YouTube over-performers – video-driven, viral, or nostalgic songs; Spotify-heavy tracks – audio-first listening tracks.
- Business implications: Use this analysis to tailor promotion by platform: Video-first content → prioritize YouTube / short-form video; Audio-first content → focus on Spotify playlists and audio-based marketing.

Thank You