

Process Book

Group Members: Meena Balamurugan and Everett Lopez

Project Repository: <https://github.com/dataviscourse2025/final-project-treblemakers.git>

1. Overview and Motivation

Our project explores the evolving landscape of music through data-driven visualizations using Spotify's Web API. While our original focus was on analyzing audio features (such as danceability, energy, tempo, and valence) across genres and decades, we encountered an important technical limitation early in the process: several key Spotify endpoints related to track-level audio features have been deprecated for new applications.

Rather than allowing this limitation to halt progress, we adapted our approach to study marketing-oriented metrics that still reflect important aspects of musical trends. Specifically, we shifted our focus to examining artist popularity, market concentration, and artist efficiency across major genres. This revised approach continues to answer questions about how music evolves and competes over time. The difference is that it is now from a marketing and cultural influence perspective rather than an auditory one.

Our motivation remains rooted in understanding how genres and artists shape, and are shaped by, global listener behavior. By visualizing these relationships interactively, we aim to make industry-level insights intuitive for both casual audiences and marketing strategists.

2. Related Work

A major source of inspiration for this project was Spotify Wrapped, which demonstrates how large-scale music data can be transformed into clear, engaging, and highly narrative visual summaries. Wrapped's use of bold color, simple metrics, and intuitive storytelling influenced our focus on designing visualizations that feel accessible while still revealing deeper industry patterns. Its emphasis on personalized insights also guided our decision to structure our visuals around artist influence, genre concentration, and long-term popularity trends. It helps frame complex data in a way that feels interpretable and meaningful.

3. Questions

Based on the metrics available through Spotify's API and the artist-level dataset produced by our collection script, our project focuses on answering the following questions:

- Which genres are dominated by a small group of extremely popular artists, and which have a more balanced distribution of influence?
- How efficiently do artists convert general popularity into an engaged follower base? (i.e., does high visibility always translate to strong followership?)
- How do genre popularity trends change over time?

- Which genres show rising momentum, and which appear to be declining in cultural relevance?
- How do legacy genres (e.g., Rock, Jazz) compare with globally emerging ones (e.g., K-Pop, Latin) in long-term listener engagement?

These questions directly shaped the visualizations we created and reflect the shift from audio-feature analysis to marketing-oriented insights.

4. Data and Methodology

We collected all data using Spotify's Web API, relying primarily on the `/v1/search` endpoint to retrieve artists and tracks across ten major genres. Because Spotify has deprecated several key endpoints, including genre seeds and related artists, we could not rely on Spotify's built-in genre hierarchy. As a result, we developed our own genre-filtering and expansion logic to build consistent datasets across genres.

For each genre, our script performed large-batch searches (up to 1,000 results) and filtered artists using strict matching to the root genre and a minimum follower threshold of 500,000 to ensure industry relevance. When a genre returned too few qualified artists, our code automatically expanded the search using the most frequently co-occurring subgenres detected in the initial sample. This allowed us to construct stable 50-artist datasets per genre despite the lack of official Spotify genre-seed support.

After deduplication, the top 10 artists were extracted for the leaderboard view, while the full artist sets powered the Bubble Map, Genre Concentration Index, and Efficiency Map. To compute long-term trends, we retrieved up to 50 tracks per year (2010–2025) using genre- and subgenre-based queries, averaging popularity to generate annual trend curves. This pipeline produced structured, comparable datasets across genres while working within Spotify's modern API limitations.

5. Exploratory Data Analysis

Before building our final visualizations, we conducted a series of lightweight exploratory tests to understand the structure and limitations of the dataset produced by our API pipeline. Our earliest visual checks included simple bar charts of artist follower counts, scatterplots comparing popularity and followers, and small tables summarizing genre distributions. These quick sketches helped confirm that some genres, such as Pop and K-Pop, were heavily dominated by a few extremely influential artists, while others, like Jazz or Country, showed a more balanced distribution. This initial imbalance highlighted the need for a visualization that could express market concentration, which later became the Genre Concentration Index.

We also plotted preliminary scatterplots of popularity versus followers, which revealed that these two metrics often diverged more than expected: some artists had high popularity but relatively modest follower counts, while others showed the opposite pattern. This insight led directly to the

creation of the Efficiency Map, designed to visualize how effectively artists convert general visibility into an engaged audience.

Finally, while testing our yearly popularity retrieval, we created rough line plots of average popularity from 2010–2025. These early trends showed clear rises in certain global genres (e.g., K-Pop, Latin) and plateauing patterns in legacy genres. This motivated the inclusion of the Popularity Curve, allowing users to compare long-term genre trajectories.

Overall, our exploratory checks highlighted key structural patterns in the data, such as artist dominance, differences between popularity and follower engagement, and clear long-term shifts in genre momentum. These findings directly shaped the visualizations we chose to build and helped us focus on the metrics that best captured how genres and artists compete and evolve over time.

6. Design Evolution

Our visual design evolved as we evaluated how effectively each prototype communicated the insights we wanted to highlight. One significant change occurred in the Genre Concentration Index (GCI). In the milestone phase, the GCI was implemented as a donut chart, but user testing and our own observations revealed that it was difficult to judge segment proportions in a circular layout. Because linear encodings support more accurate comparison, we redesigned the GCI as a horizontal bar chart (reflected in the final dashboard), which makes the top, mid-tier, and small-artist shares easier to interpret and aligns better with perceptual principles.

The Efficiency Map preserved its scatterplot design but grew in functionality. The milestone version only displayed popularity versus follower count. In the final version, users can click on any quadrant to reveal additional artist details, allowing deeper exploration without adding visual clutter.

The Popularity Curve underwent the largest conceptual pivot. Our initial design envisioned an artist-level time series showing album releases, popularity trends, and audio feature patterns. However, several Spotify endpoints required for this analysis, especially those returning audio features and detailed artist history, were deprecated for new applications. To adapt, we redesigned the visualization to operate at the genre level, aggregating up to 50 tracks per year (2010–2025) and plotting the average yearly popularity. Users can also compare two genres at once, preserving the ability to tell long-term evolutionary stories even without access to the original artist-level data.

Overall, these changes demonstrate a consistent design philosophy: when faced with ambiguity, complexity, or missing data, we chose clarity, interpretability, and robustness. Each redesign strengthened the narrative power of the dashboard while staying grounded in accessible Spotify data.

7. Design and Implementation

After pivoting from feature-based audio analysis to marketing-focused insights, we designed five complementary visualizations implemented using D3.js, JavaScript, and CSS, orchestrated through a main controller script. Each visualization highlights a distinct dimension of the modern music industry:

Bubble Map

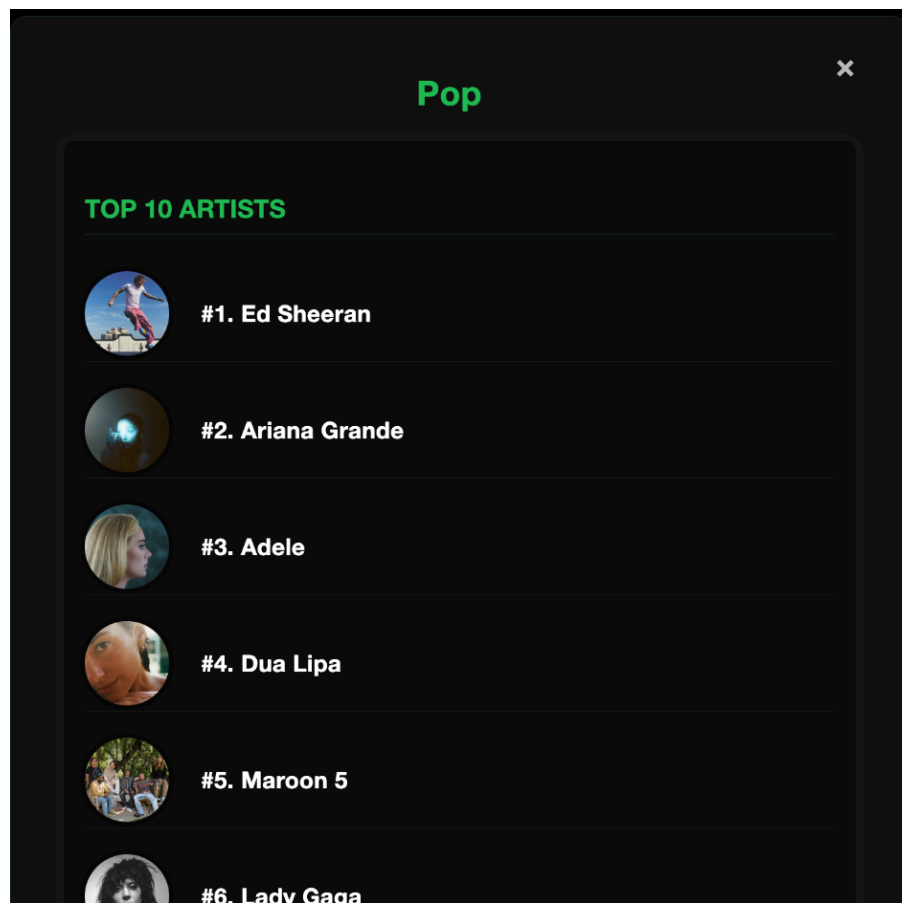
The Bubble Map serves as the entry point to the entire interface, offering a high-level visual overview of the ten selected genres. The 10 genres are represented as large circular bubbles arranged in a pleasant 3–4–3 layout, balancing symmetry with visual interest. Artist image thumbnails are applied to each bubble using SVG patterns; by default they appear in grayscale to reduce visual clutter, but on hover they animate into full color with a subtle scale-up effect to indicate interactivity. Clicking a bubble opens a modal panel that drives the rest of the experience, automatically loading the Top-10 Artist List, Genre Concentration Index, Efficiency Map, and Popularity Curve for the selected genre. This design keeps the landing view clean and uncluttered while still acting as a navigational hub that ties the entire system together.



Top 10 Artist List

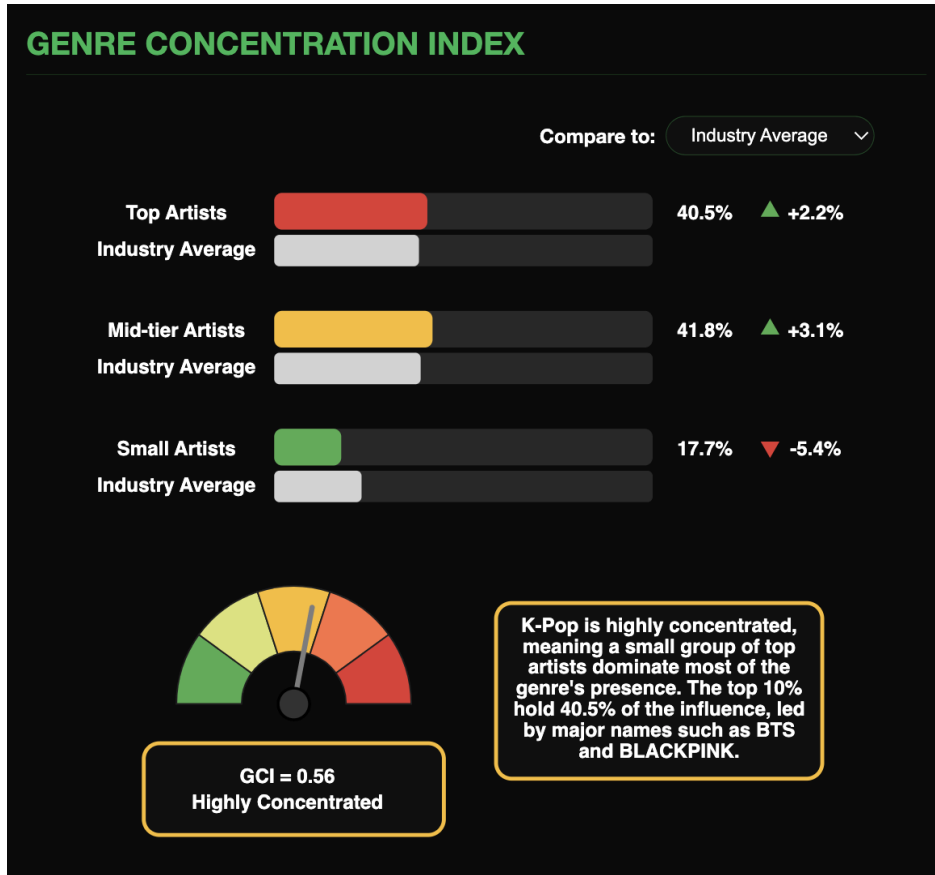
Inside each genre's modal, a ranked list shows the top ten artists in the sample by followers and popularity. Each entry includes the artist's image and name, presented in a simple,

leaderboard-style format. This view provides an immediate sense of who defines the genre and how dominance is distributed among its leading figures.



Genre Concentration Index

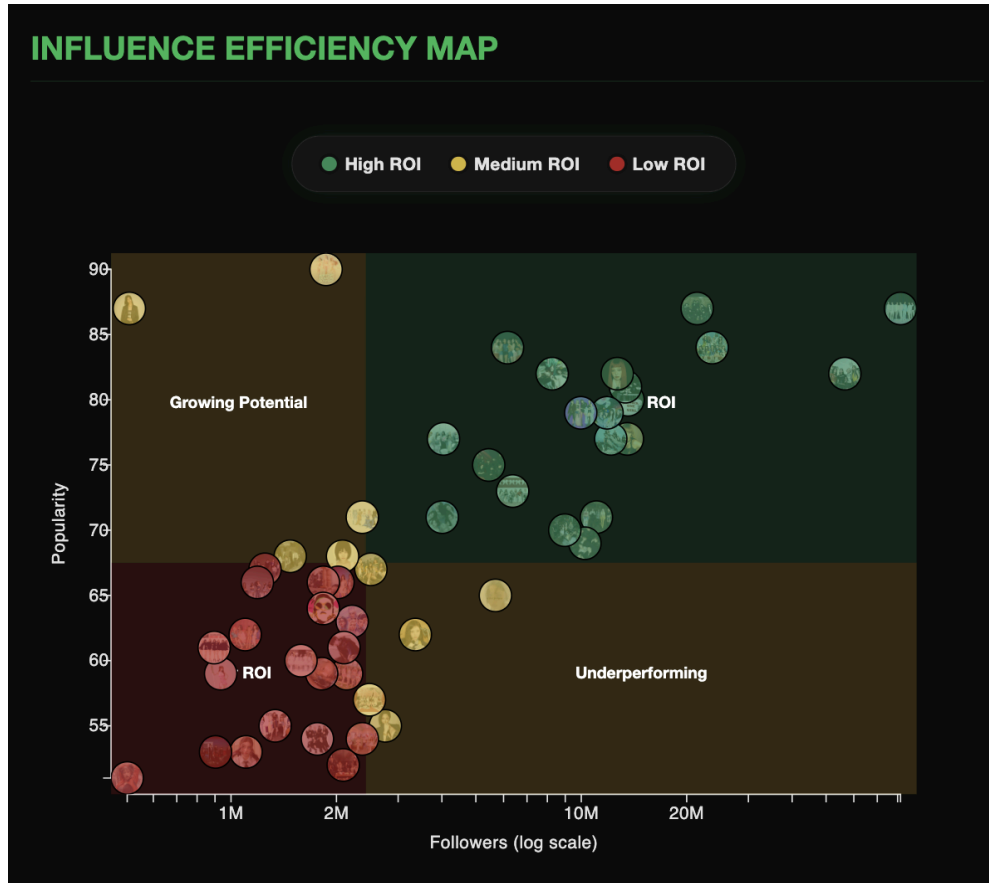
The Genre Concentration Index visualizes how influence is distributed within a genre by weighting each artist using a custom formula based on popularity and follower count. The implementation computes tier shares (top 10%, mid 30%, bottom 60%) and displays them as animated horizontal bars, allowing users to quickly see whether a genre is dominated by a few artists or more evenly distributed. A comparison dropdown lets users compare the selected genre to either any other genre or to an “industry average,” which is defined as the mean tier distribution across all genres included in this project. Hovering over each bar reveals example artists from that tier, while a color-coded gauge meter provides an overall concentration score (from evenly distributed to highly concentrated). An accompanying text panel summarizes the results in plain language. Together, these elements make the GCI an intuitive and interpretable tool for understanding how influence is distributed within a genre.



Influence Efficiency Map

The Influence Efficiency Map visualizes how artists within a genre balance two key marketing metrics: popularity and follower base. Each artist is represented as a circular image positioned using a log-scaled x-axis (followers) and a linear y-axis (popularity). We used a log scale for follower counts because artist audiences vary by several orders of magnitude. In our exploratory samples, follower distributions ranged from emerging artists to global superstars, creating extreme right-skew. A log scale prevents superstar outliers from flattening the visual space and ensures that all artist tiers remain perceptually distinguishable.

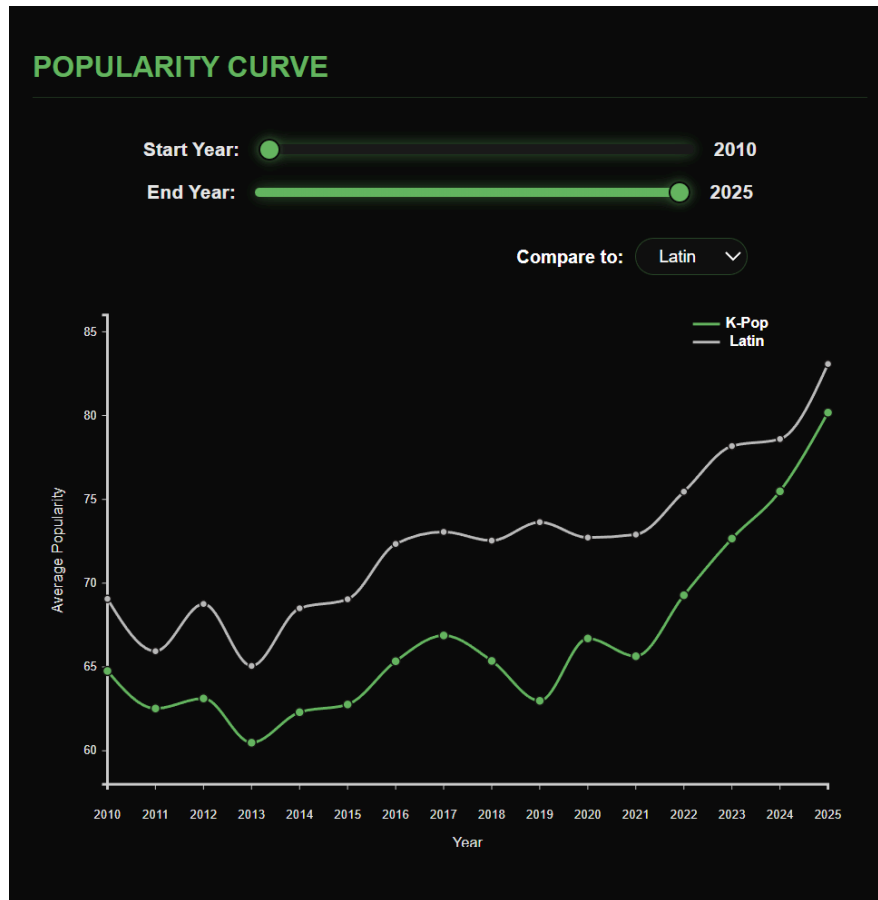
To help users interpret performance at a glance, the chart dynamically divides into four ROI quadrants based on the median follower count and median popularity. Each quadrant corresponds to a strategic tier: High ROI, Growing Potential, Underperforming, and Low ROI. Each tier is highlighted with translucent color overlays that update during zooming and panning. Clicking on any quadrant opens a mini-dashboard overlay summarizing the number of artists in that tier, average popularity, average followers, and a scrollable, sortable list of artist profiles. These design choices emphasize perceptual clarity: quadrant shading supports rapid categorization, and the detail-on-demand dashboard allows focused comparison without overcrowding the main view. The result is an interactive tool that communicates how efficiently artists convert audience size into cultural impact within a genre.



Popularity Curve

The Popularity Curve visualizes long term momentum within a genre by plotting the average popularity of its tracks from 2010 to 2025. The line chart is fully interactive, and users can adjust the displayed year range with two synchronized sliders that update the chart instantly. Each genre is drawn as a smooth trend line with clear markers at each yearly data point. A comparison menu allows users to overlay a second genre in a neutral color, which supports direct visual comparison across the same time period. When the user hovers over the chart, a vertical guide line appears along with two tooltip cards that display the exact popularity values for both the main genre and the comparison genre in the selected year.

The visualization emphasizes clarity and ease of interpretation. The animated line drawing appears only when the user switches genres or comparison targets, which makes changes easy to follow without creating visual distraction. A simple legend labels both lines and adjusts automatically to maintain readability. Together, these choices turn the Popularity Curve into an effective exploratory tool that helps users compare long term popularity trajectories and understand whether a genre is rising, declining, or staying stable over time.



8. Evaluation

Each visualization directly contributed to answering our core questions about genre structure, artist influence, and long-term popularity trends. The Genre Concentration Index addressed our question about whether genres are dominated by a small group of artists or more evenly distributed. It showed clear variation: some genres cluster heavily around a few stars, while others maintain broader mid-tier participation. The Influence and Efficiency Map answered our question about which artists convert visibility into meaningful engagement by revealing mismatches between popularity and follower count, and by highlighting which artists outperform expectations within each ROI quadrant. The Popularity Curve helped us understand long-term evolution by showing how a genre’s average track popularity rises, falls, or stabilizes across time.

Together, these insights directly answered our guiding questions about concentration, efficiency, and genre evolution. The visualizations work well overall, and user interactions such as quadrant detail views and comparison modes strengthened interpretability. Future improvements could include incorporating audio features if Spotify reopens those endpoints, adding finer subgenre filters, and expanding the temporal resolution of popularity data. Even with current limitations, the system provides a coherent and informative perspective on how genres and artists behave on Spotify.