



Analyzing the Relationship Between Genre Count and Artist Success

Introduction

Music consumption is shaped by trends, culture, and the diversity of sounds artists create. This project explores whether the number of genres an artist covers has any impact on their popularity on Spotify. By analyzing real Spotify charts and artist metadata, we uncover patterns in genre diversity and its relationship with audience appeal.

Objectives

our hypothesis main question is: Do artists who span multiple genres have higher popularity?

1. Does an artist who performs in multiple genres become more popular?
2. Is there a linear or noticeable relationship between genre count and popularity?
3. Can we predict popularity using genre count?
4. How do different datasets (primary Spotify API data vs secondary dataset) compare?

Data Collection

Dataset Source: Spotify Charts, Spotify API, Kaggle

Primary Data:

- Weekly Top-200 Spotify charts from 20 regions
- Artist information collected through the Spotify API

Secondary Data:

- Clean Spotify dataset from Kaggle for comparison

Preparation:

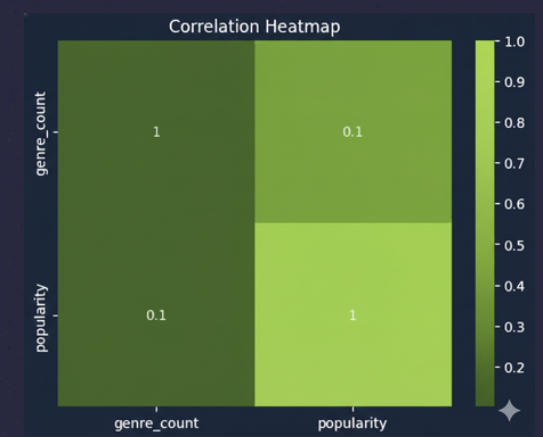
- Downloaded and merged all chart files
- Extracted unique artists
- Retrieved additional details via the API

Data Analysis

- Collected and cleaned global Spotify artist data from 20 regions
- Measured genre diversity using genre counts
- Compared popularity between single-genre and multi-genre artists using visualizations
- Used correlation, scatter plots, and regression to analyze relationships
- Identified dataset limitations (regional imbalance, missing genre data, popularity bias)
- Highlighted how genre diversity may influence artist popularity



Scatter plot - Popularity vs Number of Genres



Boxplot - Popularity by Genre Category

Models and Findings

- **Linear Regression** served as the interpretable baseline, showing how each genre affects popularity when controlling for others.
- **Random Forest** captured nonlinear relationships and complex interactions, improving predictive performance and highlighting influential genres.
- **XGBoost** further modeled subtle dependencies within the top 50 genre features, achieving competitive performance on structured data.

Dataset included 702 artists with numeric features (genre_count, followers) and one-hot encoded vectors for the top 50 genres

Popularity depends more on which genres an artist has than how many. Some genres boost popularity while others reduce it, and the difference between single- and multi-genre artists isn't statistically significant

Conclusions

- Specific genres—not genre diversity—drive popularity; popularity is more sensitive to the type of genre than the number of genres an artist has.
- Nonlinear models outperform linear ones, indicating popularity results from complex interactions across genre combinations rather than isolated features.
- Genre_count is useful as a supporting feature in ensemble models but not as a standalone predictor.
- Multi-genre artists show higher average popularity, but the difference is not statistically significant, suggesting genre diversity may help but is not decisive.
- Regional genre effects matter: some genres (e.g., Khaleeji) reduce predicted popularity, while others (e.g., Egyptian pop) consistently increase it.