

Grammy Career Trajectories - App Summary

What It Is

A data analysis project that compares Grammy winners and nominees with modern streaming metrics to test whether awards predict long-term success. It targets major Grammy categories from 1959-2024 using public web and API data.

Who It's For

Primary persona: a music industry or data storytelling analyst evaluating awards vs. commercial impact.

What It Does

- Scrapes Grammy category tables from Wikipedia into CSV data files.
- Includes a Spotify fetcher script scaffold to collect monthly listeners and follower counts.
- Plans fuzzy name matching to normalize artist identities across datasets.
- Merges Grammy and Spotify data into a final analysis-ready dataset.
- Provides Billboard chart scraping utilities for longitudinal chart presence.
- Supports exploratory analysis and visualization in a Jupyter notebook.

How It Works (Repo Evidence)

- Inputs: Wikipedia pages for Grammy categories, Spotify API credentials, and optional Billboard charts.
- Scraping: scripts/01_scrape_grammys.py -> data/raw/grammy_data_raw.csv.
- Spotify metrics: scripts/02_fetch_spotify.py (stub) -> data/raw/spotify_data_raw.json.
- Normalization/merge: scripts/03_normalize_data.py (stub) -> data/final/grammy_spotify_merged.csv.
- Analysis: notebooks/exploratory_analysis.ipynb consumes processed/final data.
- Static site build/deploy pipeline: Not found in repo.

How To Run (Minimal)

- python -m venv venv, then activate the venv.
- pip install -r requirements.txt
- Copy .env.example to .env and set SPOTIFY_CLIENT_ID and SPOTIFY_CLIENT_SECRET.
- python scripts/01_scrape_grammys.py
- Optional next steps: scripts/02_fetch_spotify.py and scripts/03_normalize_data.py (currently stubs).
- Local site preview/build commands: Not found in repo.