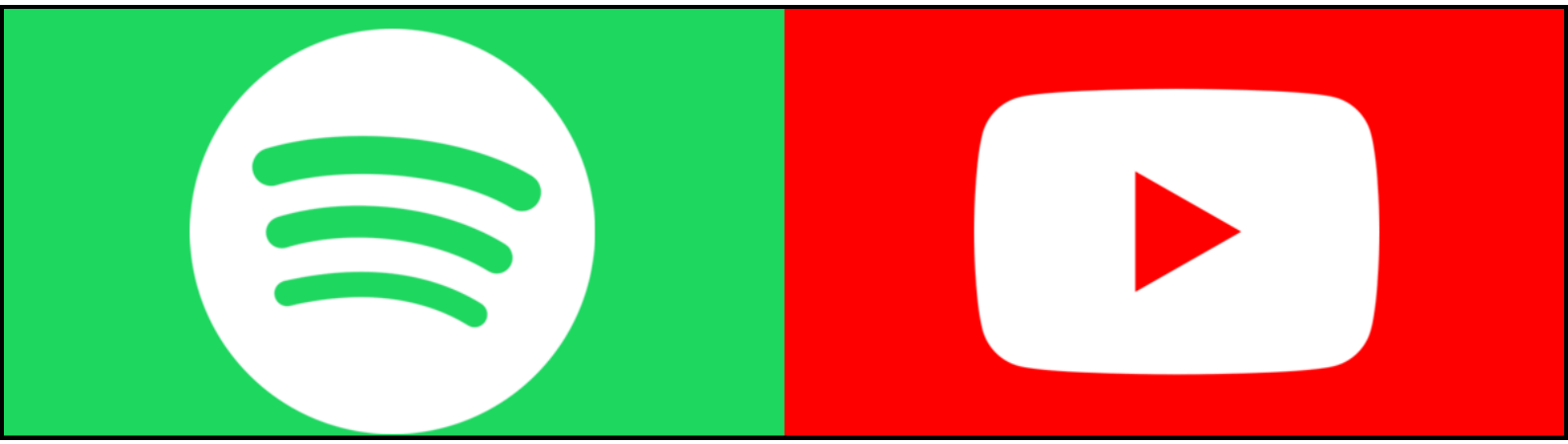

Spotify & YouTube Music – Power BI Project

Introduction to Spotify and YouTube

Spotify and YouTube are leading digital media platforms. Spotify is known for its music streaming and personalized playlists, while YouTube is a dominant video-sharing site offering music and user-generated content. This project outlines the **data cleaning process in Power BI**, detailing how missing values were handled, data types corrected, and consistency ensured.

Data Cleaning

1. Reorder and Rename Columns for Clarity

Objective: Improve readability and usability of the dataset.

Key Actions:

- Group columns logically: Spotify-related columns first, followed by YouTube-related.
- Rename columns to use title case (e.g., "spotify_info" → "Spotify Info").
- Retain key columns such as Track, Album, Channel, Views, etc.

Summary: Columns were reordered and renamed to enhance clarity and usability.

2. Identify and Handle Missing Values

Objective: Maintain data integrity by managing null values.

Approach:

- Remove rows with missing quantitative values (e.g., Views, Likes, Stream).
- Retain missing textual values (e.g., Description, Comments) as they do not affect analysis significantly.

Summary: Quantitative gaps removed; textual gaps retained for minimal impact.

3. Fix Irregularities in Merged Columns

Objective: Split combined fields for accurate data analysis.

Approach:

- Spotify_Info split into Spotify Link and Track ID using the | delimiter.
- YouTube_Info split into YouTube Link and Video Title based on character structure.
- Validation ensured links are functional and titles correct.

Summary: Merged fields were separated for improved data integrity.

4. Correct Case Sensitivity and Naming Conventions

Objective: Standardize formatting across dataset.

Approach:

- Column names converted to title case.
- Text entries (e.g., artist names, titles) standardized for consistent presentation.

Summary: Formatting improvements enhanced dataset consistency and readability.

5. Remove or Handle Irrelevant Columns

Objective: Streamline dataset by removing unnecessary columns.

Actions:

- Removed irrelevant columns like "Random Column 1" and "Random Column 2".
- Cleaned any random data from relevant fields.

Summary: Only meaningful data retained, reducing noise in analysis.

6. Handle Inconsistent Data Types

Objective: Ensure numerical columns are properly typed.

Actions:

- Columns like Views, Danceability, Energy were converted to decimal type.
- Power BI tools used to fix anomalies and validate formatting.

Summary: Numeric fields cleaned and correctly formatted.

7. Address Invalid Data and Ensure Correct Labeling

Objective: Remove incorrect entries and ensure accurate labels.

Steps:

- Replaced invalid entries (e.g., "invalid_data") in Views with nulls.
- Cleaned Album column to remove irrelevant or numeric entries.

Summary: Invalid entries corrected; labeling refined.

8. Check for and Remove Duplicate Rows

Objective: Maintain uniqueness in dataset.

Method:

- Used index columns and full dataset review to detect duplicates.
- Duplicates removed with final verification for accuracy.

Summary: Duplicate rows eliminated, preserving data quality.