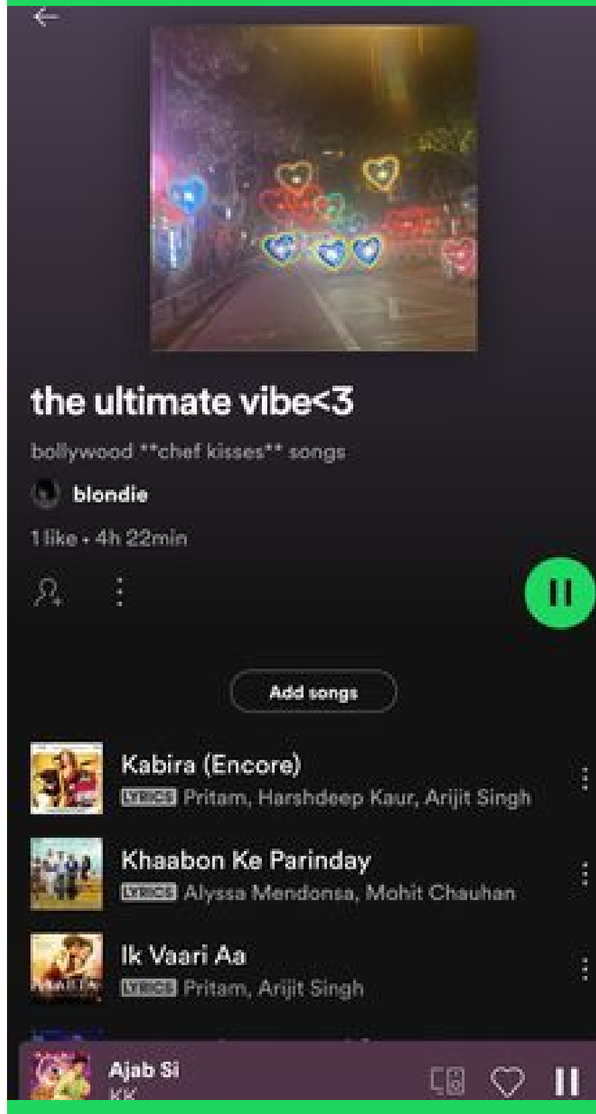
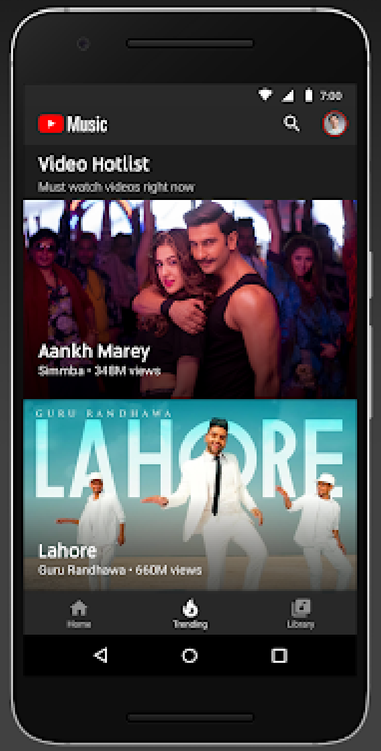
Music & Listen on Spotify

“Tune Trends: A Data-Driven Exploration of Music Streaming on YouTube and Spotify”

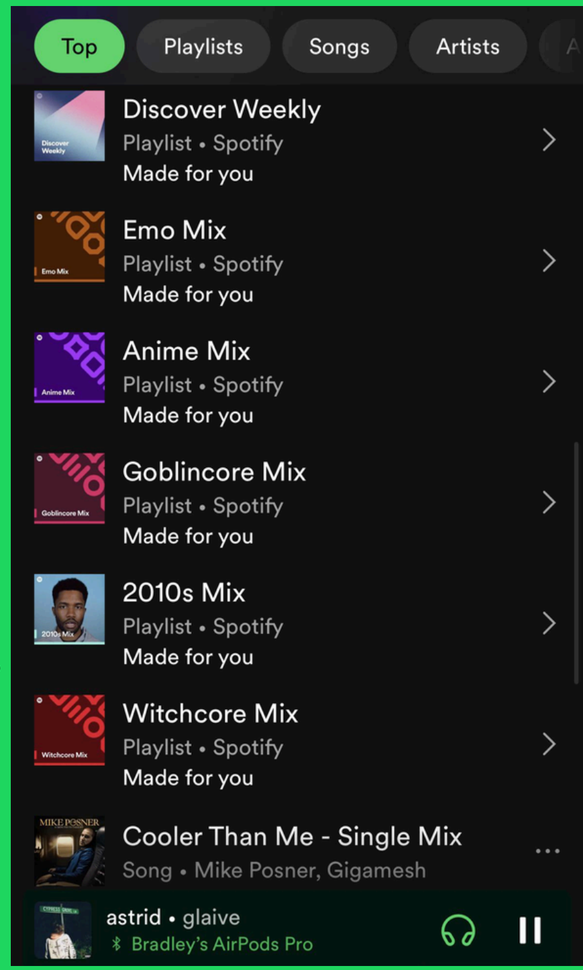
Presented by Jatin Limaye



INTRODUCTION



In the digital age, platforms like YouTube and Spotify have become integral to our entertainment and content consumption habits. Analyzing data from these platforms provides valuable insights into user behavior, content performance, and engagement trends. This project aims to harness the power of Power BI to conduct a comprehensive data analysis of YouTube and Spotify, enabling stakeholders to make data-driven decisions and understand trends in content consumption.



Q1.

- **Examine the dataset for any missing values. Which columns contain null values?**
- **How should missing values in the Views and Likes columns be handled? Should they be filled with a default value, removed, or handled in another way? Justify your approach.**

Solution

STREAM
null
null
null
null
null
null
null
null
null
null
null
null
null
null
null
null
null
null
null
null

COMMENTS	LIKES	views	licensed	OFFICIAL_VIDEO
null	null	null	null	null
null	null	null	null	null
null	null	null	null	null
null	null	null	null	null
null	null	null	null	null
null	null	null	null	null
null	null	null	null	null
null	null	null	null	null
null	null	null	null	null
null	null	null	null	null
null	null	null	null	null
null	null	null	null	null
null	null	null	null	null
null	null	null	null	null
null	null	null	null	null
null	null	null	null	null
null	null	null	null	null
null	null	null	null	null
null	null	null	null	null
null	null	null	null	null

After Examine the data above column contains null values

- In my point of view removing rows is better approach because the null values not in much quantity.

Justification

- Simplicity: Removing rows can be straightforward and avoids potential complications from imputing values.
- Data Quality: Ensures that the remaining data is complete and potentially of higher quality for analysis.

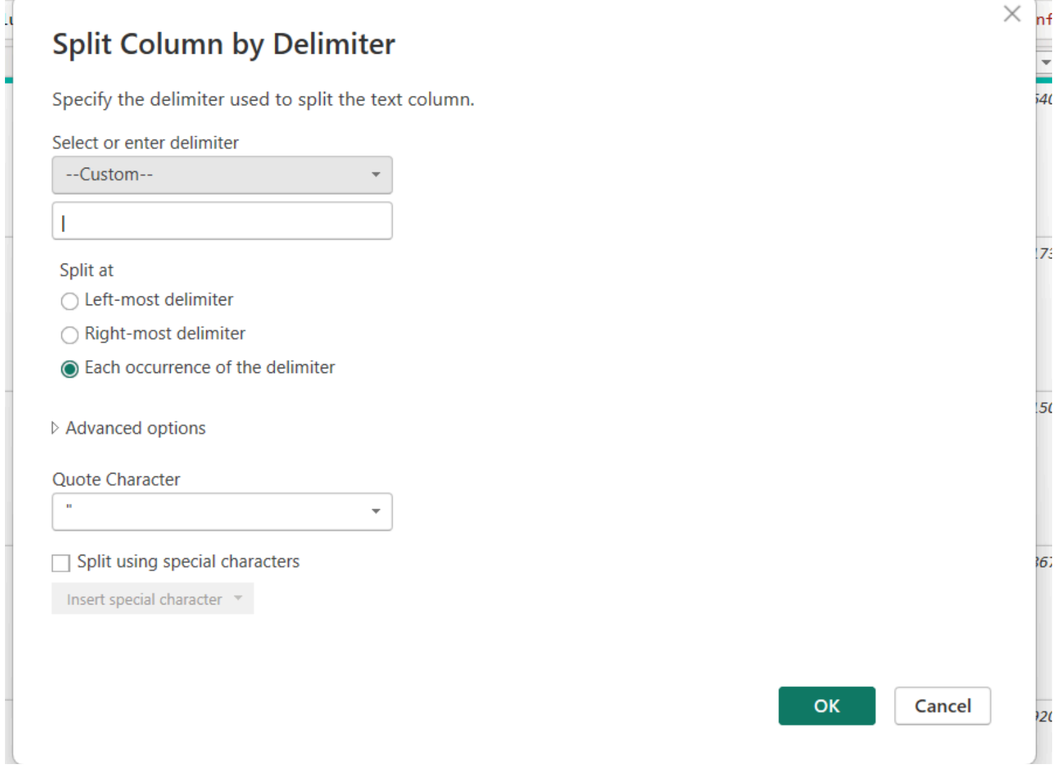
Q2.

- **The Spotify_Info and Youtube_Info columns contain merged data separated by delimiters. Split these columns back into their original components. What are the original components, and how can you ensure that the split data is clean and accurate?**

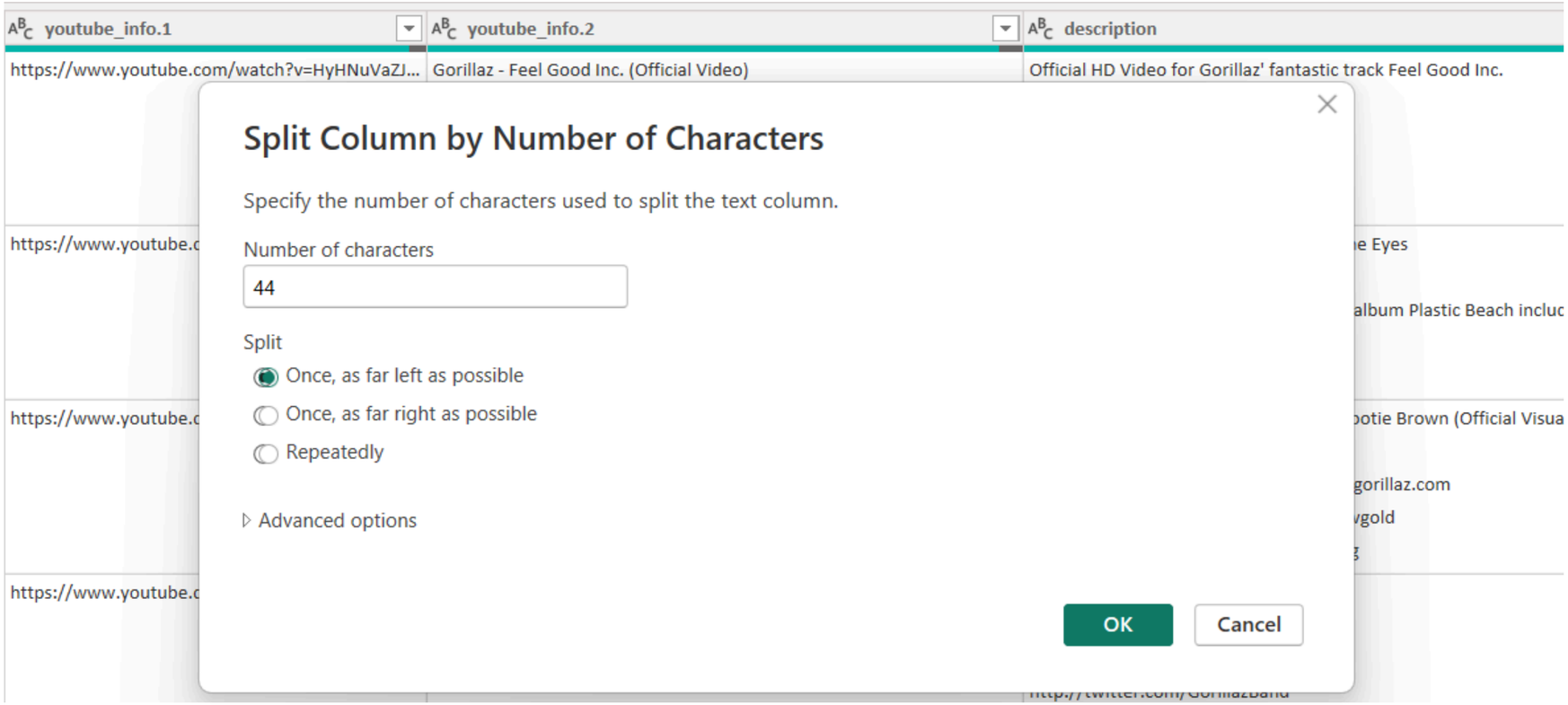
- **After splitting, remove any unnecessary delimiters or prefixes/suffixes that do not belong.**

Solution

After splitting the merged column of spotify info it split into two columns which name link of the album and track id that is the original component and to ensure their accuracy and clean we check the both column data on the spotify app



We also split the youtube column into the name of Link of the song and track id of the song



- After splitting we also extract the hyphen from the Track column with the extraction function

TRACK	ALBUM	ALBU
FEEL GOOD INC.	Demon Days	album
RHINEST		
new gol		
ON MEL		

Text Before Delimiter

Enter the delimiter that marks the end of what you would like to extract.

Delimiter

_

Advanced options

Scan for the delimiter

From the end of the input

Number of delimiters to skip

0

OK Cancel

TRACK
FEEL GOOD INC.
RHINESTONE EYES
new gold (feat. tame impala and bootie brown)
ON MELANCHOLY HILL

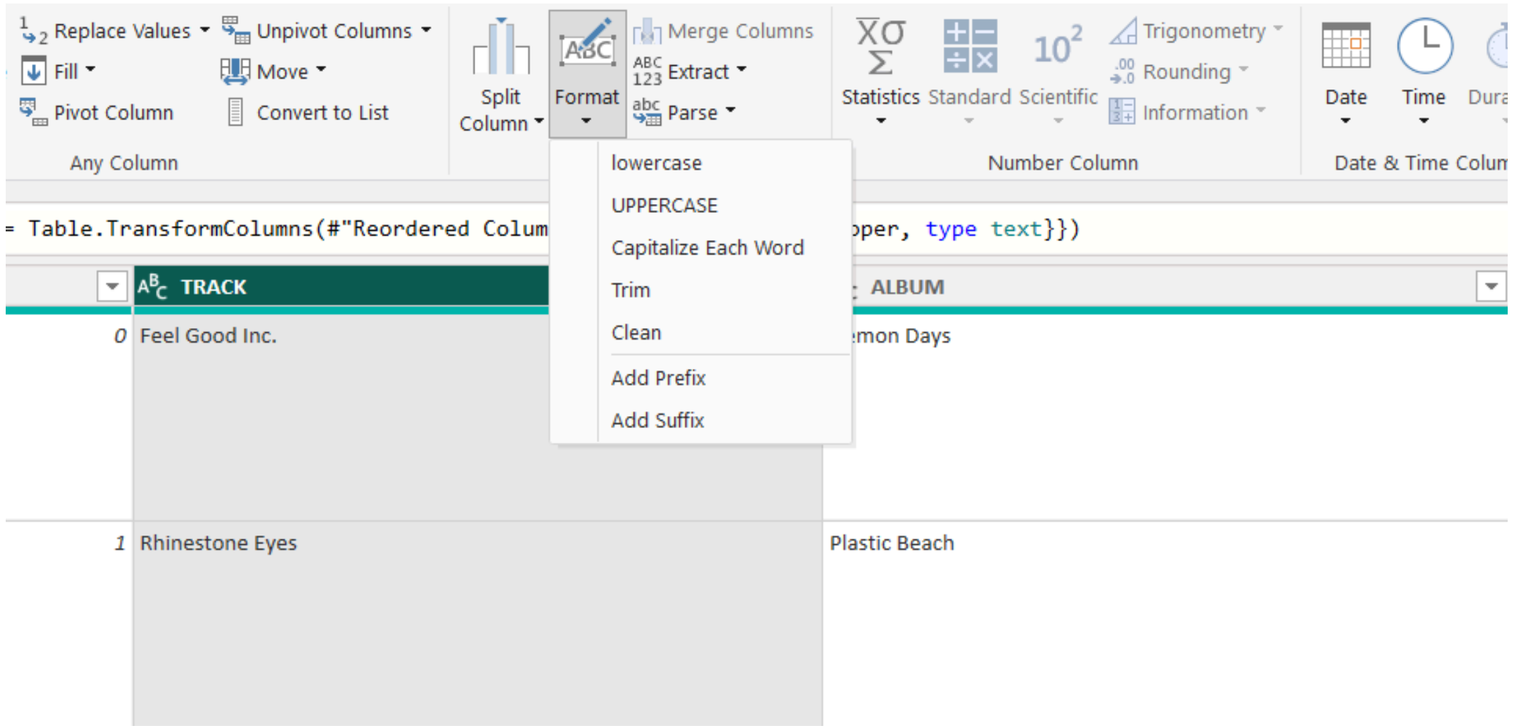
Q3.

- **The column names have inconsistent case sensitivity (some are uppercase, others lowercase). Standardize all column names to follow a consistent format (e.g., all lowercase with underscores).**

- **Fix any data entries where case sensitivity might affect consistency (e.g., artist names or track titles). Ensure that the Artist and Track columns are formatted consistently**

Solution

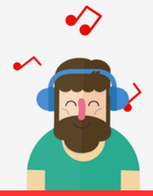
- All columns in the data is in proper format and standard form. This consistency is crucial for maintaining accuracy in analysis, seamless integration of data from various sources, and the reliability of data relationships within Power BI.
- The data in the table where artist names or title track also in the proper or standard format.



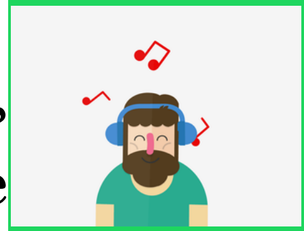
We can standardized the data through available option and we can also choose the data types according to the requirement of the dataset. We remove any error with the option remove error and replace error.

Q4.

- **Identify and remove any irrelevant or randomly generated columns that do not provide useful information for analysis. Which columns should be removed, and why?**
- **If any random data exists in relevant columns, clean or remove those entries.**



Solution



- After Examine data In the table we found random_column_1 , RANDOM_COLUMN_2 can not make any sense so we delete the column.

- In my case i not found any data but if it found then we replace the data with the replace function

Q5.

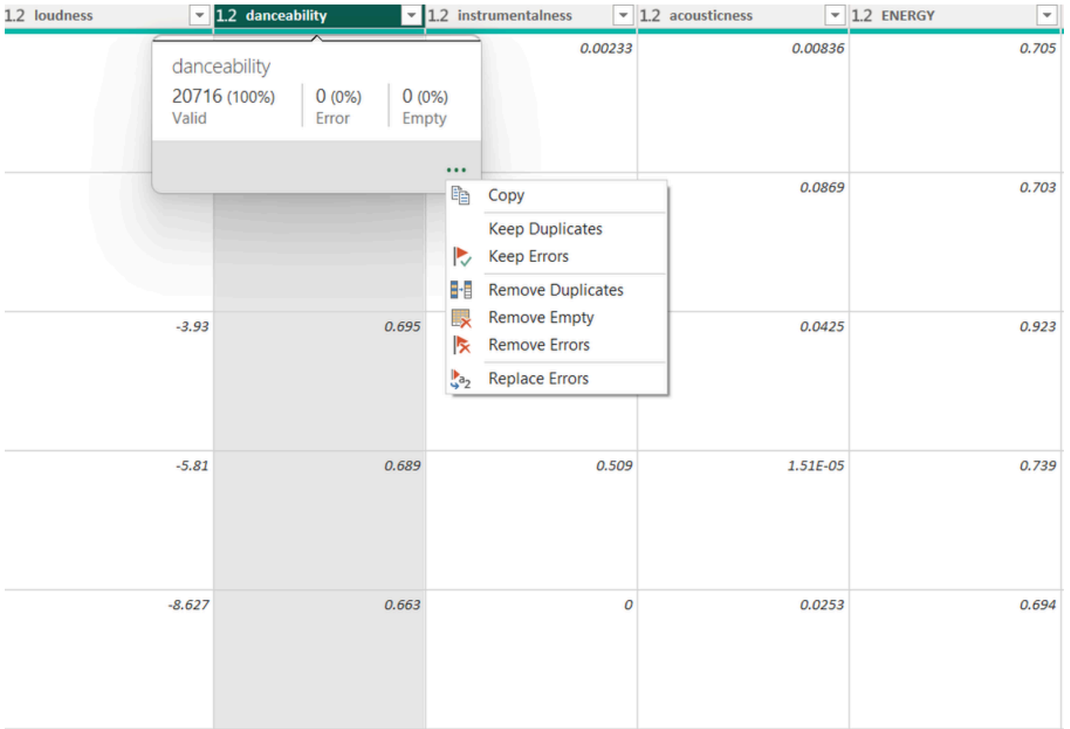
- **Some columns that should be numeric (e.g., Danceability, Energy) are stored as text. Convert these columns back to numeric format. What steps would you take to identify and fix any issues that arise during this conversion?**
- **Ensure that all numeric columns are in the correct format and handle any non-numeric values or anomalies.**

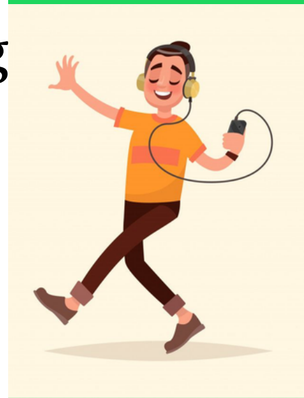
Solution

- The Both column Danceability, Energy not contain text they are in the numeric format and if the error occur like nan or unknown then replace with null which arranged all in the format. if the error still occur then corrected with option or feature of remove error.

- All the numeric columns in the table are arranged in the proper form and not occur any non numeric error in that column

Some occur which are occur after replace then we use remove error and the remove error feature remove all error from the columns because some error are not detected easily so with this feature we correct other errors also which create problem in the data cleaning process.





Q6.

- **Check the Views column for any entries labeled as "invalid_data" or any other incorrect values. Replace these entries and justify your method.**
- **Ensure that all values in the Album column are correctly labeled and that there are no numeric entries or irrelevant data.**

Solution

1. Replace the invalid_data with the null.By leaving the invalid or missing field as null, you can clearly indicate to the user that something is wrong with their input. This way, you can display an appropriate error message like:

- "The data you entered is invalid."
- "This field cannot be left blank."
- "Please enter valid information."

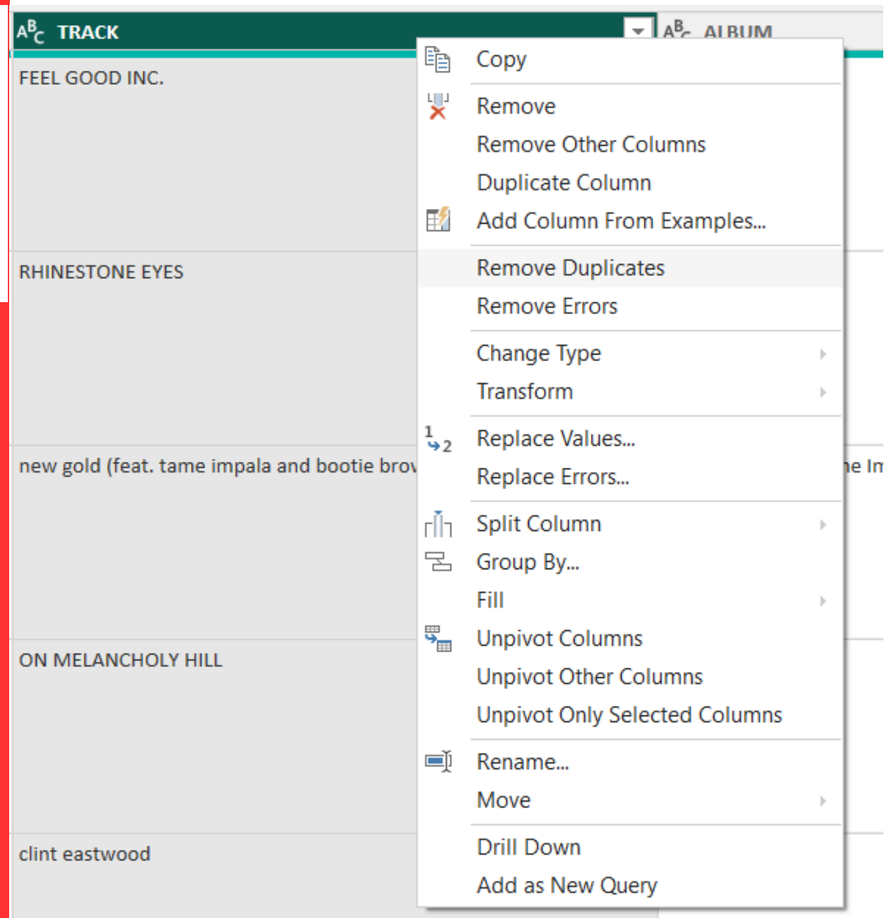
The system informs the user about the specific fields that are null or invalid, encouraging them to review and correct their input. This prevents invalid data from being processed or stored, ensuring the integrity of the data being handled.

- The data entered in the album column is clean and in the relevent form.

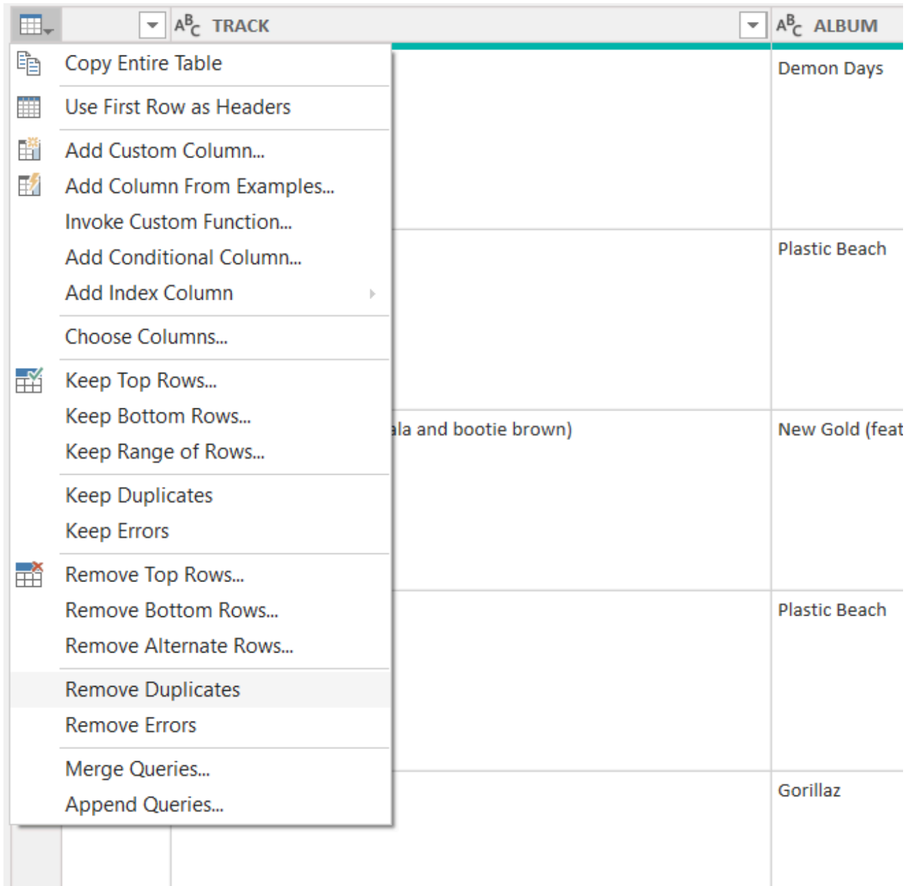
Q7. Identify and remove any duplicate rows in the dataset? How can you ensure that the remaining data is unique and accurate.

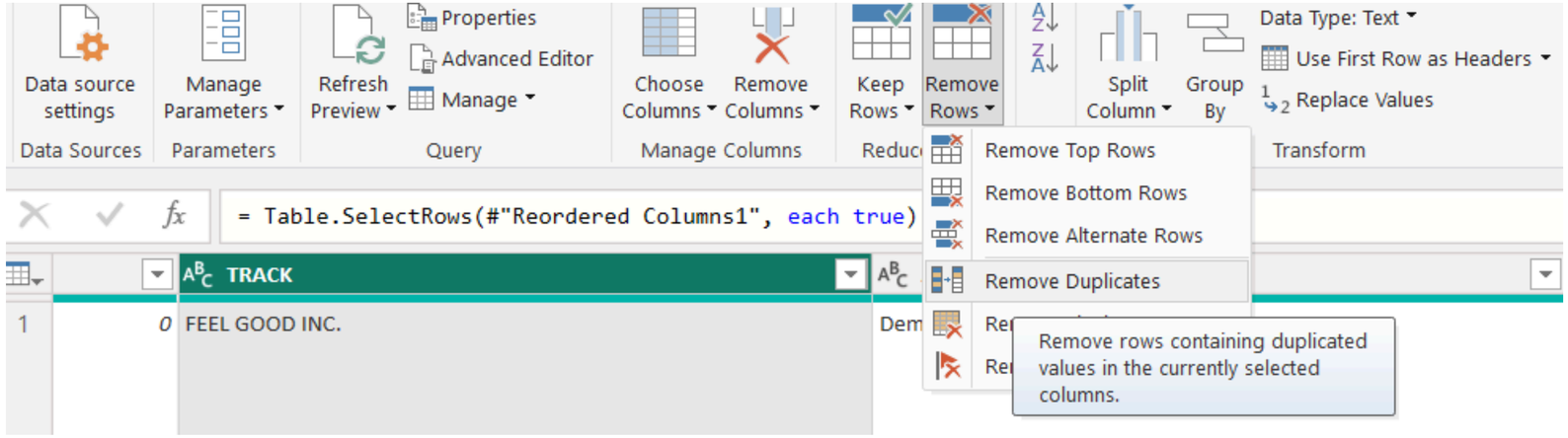
Solution

This method for remove duplicate from particular column



This method for remove duplicate from all rows and columns





Ensure Remaining Data is Unique and Accurate

- ## Check for Remaining Duplicates:

1. To ensure there are no remaining duplicates, you can use the Group By feature to aggregate data and check for any anomalies.
2. Go to the Home tab and select Group By.
3. Choose the columns you want to group by and summarize the data to check if there are any unexpected results or counts.

Removing duplicates from a dataset is necessary for several important reasons:

- Ensures Accuracy
- Improves Data Integrity
- Enhances Data Quality
- Facilitates Efficient Analysis
- Prevents Misleading Insights
- Reduces Data Storage and Processing Costs
- Improves Data Consistency

Q8.

1. **Reorder the columns in a logical sequence to improve the dataset's readability and usability. What order makes the most sense for this dataset?**
2. **Rename columns where necessary to ensure that their names clearly reflect the data they contain.**

Solution

1.2 speechiness	1.2 loudness	1.2 danceability	1.2 ENERGY	1.2 instrumentalness	1.2 acousticness	1.2 ter
0.613	0.177	-6.679	0.818	0.705	0.00233	0.00836
0.0463	0.0302	-5.815	0.676	0.703	0.000687	0.0869
0.116	0.0522	-3.93	0.695	0.923	0.0469	0.0425
0.064	0.026	-5.81	0.689	0.739	0.509	1.51E-05
0.0698	0.171	-8.627	0.663	0.694	0	0.0253

PROPERTIES

Name

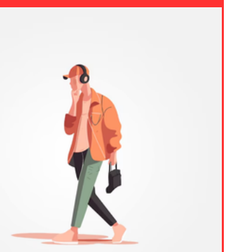
Spotify_Youtube_Unclean_v2

All Properties

APPLIED STEPS

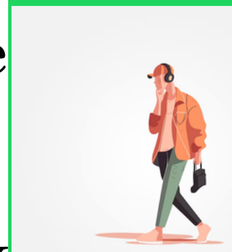
- Source
- Promoted Headers
- Changed Type
- Reordered Columns
- Removed Columns
- Renamed Columns
- Removed Duplicates
- Split Column by Delimiter
- Changed Type1
- Renamed Columns1
- Split Column by Position
- Changed Type2
- Extracted Text Before Delimiter
- Replaced Value
- Changed Type3
- Removed Errors
- Extracted Text Before Delimit...
- Extracted Text After Delimiter
- Reordered Columns1
- Filtered Rows

Reordering columns and renaming them are essential steps in making a dataset more readable and user-friendly.

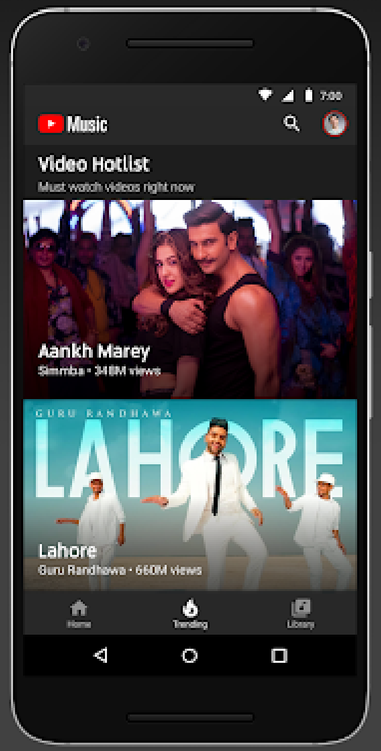


- "Ordering the data is helpful for anyone analyzing it to make decisions. It puts the right information in the right place, so even someone with no prior knowledge of the data can offer suggestions and develop an understanding of it. Data is arranged systematically and in an organized manner to enhance clarity and support accurate analysis."

- Renaming columns in a dataset improves clarity and usability by making it easier for users to understand and find the data they need. Descriptive names eliminate confusion and reduce the learning curve for new users, streamline data analysis by facilitating easier querying and filtering, and enhance communication among team members. Consistent naming also helps prevent errors and ensures data visualizations are clear and interpretable. Column names make the dataset more accessible and valuable for accurate analysis and decision-making.



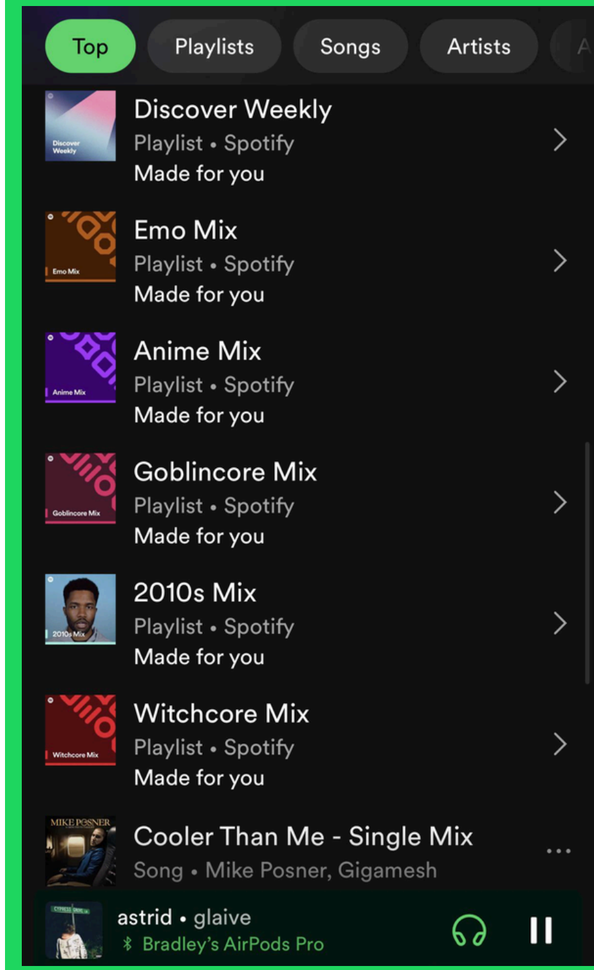
CONCLUSION

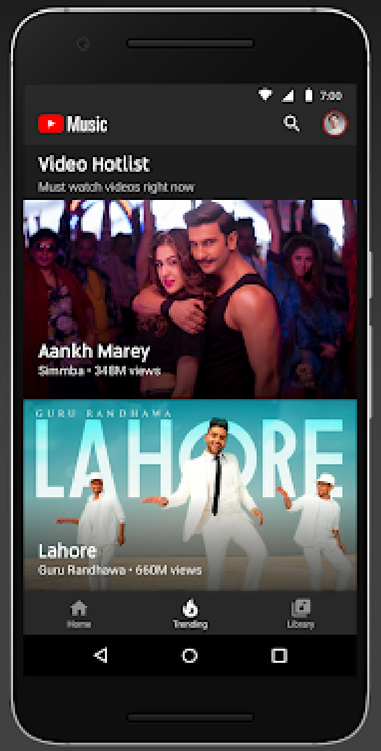


In short, proper data cleaning ensures accurate and actionable insights. Key steps include:

- Fixing Null Values: Ensures data completeness.
- Consistent Structuring: Simplifies analysis.
- Removing Unnecessary Columns: Reduces clutter.
- Reordering and Renaming Columns: Improves readability.
- Eliminating Duplicates: Prevents redundancy.
- Removing Invalid Data: Maintains accuracy.
- Standardizing Text: Enhances consistency.

These steps create a reliable dataset, leading to clearer insights and better decision-making.





Thank You

If you like my content then share it