Song Analysis using Power BI

- Import Data

1. Open Power BI Desktop.
2. Click on Home > Get Data> Text/CSV.
3. Select your dataset and load it.

1. Data Cleaning and Preparation

Using Power Query for Data Cleaning

1. Open Power Query Editor:
   - Click on Transform Data.

2.Replace Missing Values:

```
  // Replace null values in 'viewCount' column with 0
  let
     Source = <your source step>,
     ReplacedNulls = Table.ReplaceValue(Source, null, 0, Replacer.ReplaceValue,
{"viewCount"})
  in
     ReplacedNulls
```

3. Convert Data Types:

```
  // Convert 'publishedAt' to DateTime and 'viewCount', 'likeCount', 'commentCount' to number
type
  let
     Source = <your source step>,
     ChangedTypes = Table.TransformColumnTypes(Source, {
        {"publishedAt", type datetime},
        {"viewCount", Int64.Type},
        {"likeCount", Int64.Type},
        {"commentCount", Int64.Type}
     })
  in
     ChangedTypes
```

4. Add Custom Columns if Needed:

```
  // Add a custom column to extract the year from 'publishedAt'
  let
      Source = <your source step>,
      AddYearColumn = Table.AddColumn(Source, "Year", each Date.Year([publishedAt]),
Int64.Type)
  in
      AddYearColumn
```

Step 3: Exploratory Data Analysis (EDA)

Create Measures Using DAX

1. Total Engagement:

```
  TotalEngagement = SUM('YouTubeSongs'[likeCount]) +
SUM('YouTubeSongs'[commentCount])
```

2. Average Views Per Video:

```
  AvgViewsPerVideo = AVERAGE('YouTubeSongs'[viewCount])
```

3. Total Views by Channel:

```
  TotalViewsByChannel = CALCULATE(SUM('YouTubeSongs'[viewCount]),
ALLEXCEPT('YouTubeSongs', 'YouTubeSongs'[channelTitle]))
```

Create Visualizations

1. Distribution of Views:
   - Create a histogram by dragging the `viewCount` field to the `Values` section of a bar chart.

2. Trend Over Time:
   - Create a line chart with `publishedAt` on the x-axis and `viewCount` on the y-axis.

Step 4: Content and Channel Analysis

Create More DAX Measures

1. Popular Tags Analysis:

// Assuming tags are stored in a single column as comma-separated values, you might need to split and analyze them differently

Create Visualizations

1. Bar Chart for Channel Distribution:
   - Use a bar chart to show the number of videos per channel.

2. Word Cloud for Tags:
   - Use a custom visual (Word Cloud) to display popular tags.

Step 5: Temporal Trends Analysis

1. Views Over Time:

```
ViewsOverTime = CALCULATE(SUM('YouTubeSongs'[viewCount]), DATESINPERIOD('Calendar'[Date], MAX('Calendar'[Date]), -1, MONTH))
```

2. Heatmap for Engagement by Hour and Day:
   - Use a matrix visual with `Hour` and `Day` on axes and `viewCount` as values.

Step 6: User Engagement Insights

1. Correlation Between Likes and Comments:
   - Use a scatter plot to explore the relationship between `likeCount` and `commentCount`.

Sample Dashboard Setup

1. Main Dashboard:
   - Overview with key metrics: Total Views, Total Likes, Total Comments.
   - Time series line chart for views over time.
   - Bar chart for top channels by view count.

2. Content Analysis Dashboard:
   - Word cloud for popular tags.
   - Bar chart for video distribution by channel.

3. Engagement Dashboard:
   - Scatter plot for likes vs. comments.
   - Heatmap for engagement by hour and day.

Putting It All Together

Below are snippets that you can directly use within Power BI for various tasks:

Power Query M Code for Initial Transformation

```
let
    Source = Csv.Document(File.Contents("path_to_your_csv_file"), [Delimiter=",", Columns=12,
Encoding=1252, QuoteStyle=QuoteStyle.None]),
    PromotedHeaders = Table.PromoteHeaders(Source, [PromoteAllScalars=true]),
    ReplacedNulls = Table.ReplaceValue(PromotedHeaders, null, 0, Replacer.ReplaceValue,
{"viewCount", "likeCount", "commentCount"}),
    ChangedTypes = Table.TransformColumnTypes(ReplacedNulls, {
        {"publishedAt", type datetime},
        {"viewCount", Int64.Type},
        {"likeCount", Int64.Type},
        {"commentCount", Int64.Type}
    }),
    AddYearColumn = Table.AddColumn(ChangedTypes, "Year", each Date.Year([publishedAt]),
Int64.Type)
in
    AddYearColumn
```

DAX Measures for Analysis

```
TotalEngagement = SUM('YouTubeSongs'[likeCount]) + SUM('YouTubeSongs'[commentCount])
AvgViewsPerVideo = AVERAGE('YouTubeSongs'[viewCount])
TotalViewsByChannel = CALCULATE(SUM('YouTubeSongs'[viewCount]),
ALLEXCEPT('YouTubeSongs', 'YouTubeSongs'[channelTitle]))
ViewsOverTime = CALCULATE(SUM('YouTubeSongs'[viewCount]),
DATESINPERIOD('Calendar'[Date], MAX('Calendar'[Date]), -1, MONTH))
```