**OUTLIERS**

To detect outliers for every column except the first two (`track\_name` and `artist(s)\_name`), we'll use statistical methods. A common approach is to look at the distribution of each column and identify values that are far from the central tendency. We'll use the Interquartile Range (IQR) method, where values outside 1.5 times the IQR below the first quartile or above the third quartile are considered outliers.

Let's perform outlier detection on the relevant columns:

It appears that we encountered an error due to non-numeric data in some columns while trying to calculate the interquartile range. This issue likely arises in columns that are not suitable for numerical outlier analysis, such as categorical data or text fields.

To resolve this, we need to ensure that we only apply outlier detection to numeric columns. Let's adjust our approach to detect outliers only in columns with numeric data types. We'll reattempt the outlier detection with this consideration.

We have identified outliers in several numeric columns. Here are the columns with outliers and the count of outliers in each:

- `artist\_count`: 27 outliers

- `released\_year`: 151 outliers

- `in\_spotify\_playlists`: 109 outliers

- `in\_spotify\_charts`: 78 outliers

- `in\_apple\_playlists`: 78 outliers

- `in\_apple\_charts`: 9 outliers

- `in\_deezer\_charts`: 143 outliers

- `in\_shazam\_charts`: 108 outliers

- `bpm` (Beats Per Minute): 5 outliers

- `danceability\_%`: 3 outliers

- `energy\_%`: 4 outliers

- `instrumentalness\_%`: 87 outliers

- `liveness\_%`: 44 outliers

- `speechiness\_%`: 136 outliers

### Handling Outliers:

Handling outliers depends on the context and importance of each feature. Common strategies include:

1. \*\*Trimming\*\*: Removing the outlier rows.

2. \*\*Capping\*\*: Replacing outliers with a maximum or minimum value (such as the 1st or 99th percentile).

3. \*\*Transformation\*\*: Applying a transformation (e.g., log transformation) to reduce the effect of outliers.

### Next Step:

We need to decide how to handle these outliers. Would you like to remove them, cap them, or apply a transformation? Also, after handling outliers, we will proceed with feature scaling. Let me know your preferences for handling these outliers.