# Danceability & Valence analysis of the playlist

How is danceability and valence represented in the 100 most streamed songs? Both of these variables are created using Spotify’s own algorithm, defined as follows:

**Danceability** - describes how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity. A value of 0.0 is least danceable and 1.0 is most danceable.

**Valence** - A measure from 0.0 to 1.0 describing the musical positiveness conveyed by a track. Tracks with high valence sound more positive (e.g. happy, cheerful, euphoric), while tracks with low valence sound more negative (e.g. sad, depressed, angry).

In particular, I wanted to explore the relationship between the danceability and valence of a song, does a happy or more positive song mean that it more danceable?

# Summary Statistics

A summary statistics table was created using the df.describe() function to get a feel of the shape of the data:

Table

Description automatically generated

Overall, it appears that the danceability value of a song is higher than its valence. Looking at the mean for example, danceability is 0.67 compared to 0.50 for valence. For both variables, mean and median are quite similar (**danceability:** mean: 0.67, median: 0.69, **valence:** mean: 0.50, median: 0.48), suggesting that the distribution of data is quite symmetrical.

# Visualisation

A histogram can also explain the distribution of the two variables:

Chart, histogram

Description automatically generated

As mentioned above, the similar mean and median shows a symmetrical distribution for each variable. Valence peaks around 0.4, whereas danceability peaks around 0.8. Valence appears to be normally distributed, albeit with slightly more data in the right tail than the left. Danceability has a one-tailed poisson distribution, skewed towards the higher end of the scale. This shows that the songs in the top 100 tend to be more danceable.

A boxplot can describe the data even further:

Chart, box and whisker chart

Description automatically generated

The boxplots enforce the histogram’s findings that songs need to be at least slightly danceable to reach the top 100 in streams. Valence, on the other hand, is spread out across the board, suggesting that a popular song does not necessarily need to be positive.

# Correlation and Regression

Finally, to explain the relationship between the danceability and valence further a boxplot was created:

Chart, scatter chart

Description automatically generated

As expected, as the danceability level rises, as too does the valence. A correlation of 0.49 is moderately positive, but not as strong as expected. In addition, te r-square value of 0.24 suggests that the regression line is, in fact, a weak fit against the distribution of the data. There are songs that appear to go against the earlier assumptions.

### Examples of Songs that fit the model poorly

* High Valence, Low Danceability: Shawn Mendes - Treat You Better (Danceability: 0.44, Valence: 0.75)
* High Danceability, Low Valence: Kendrick Lamar - HUMBLE. (Danceability: 0.91, Valence: 0.42)

### Examples of songs that fit the model well

* Low Valence, Low Danceability: Billie Eilish - when the party's over (Danceability: 0.37, Valence: 0.20)
* High Valence, High Danceability: XXXTENTACION - Moonlight (Danceability: 0.90, Valence: 0.71)