**Hypothesis**: Higher Track Scores are associated with higher Spotify Popularity

**GitHub**: <https://github.com/taylorduncan/DSC530.git>

Outcome of the EDA

In our exploratory data analysis (EDA), we examined the relationship between Track Scores and Spotify Popularity among other variables. Our findings indicated a positive association between Track Scores and Spotify Popularity. The scatter plot of Spotify Popularity versus Track Score demonstrated a general upward trend, suggesting that tracks with higher scores tend to be more popular on Spotify. This relationship was further supported by the correlation coefficient, which showed a moderate positive correlation between the two variables.

Histograms of Track Scores and Spotify Popularity revealed that while both distributions were somewhat skewed, they were broadly centered around moderate values. The descriptive statistics, including mean and standard deviation, further underscored that higher Track Scores are generally associated with higher Spotify Popularity. The cumulative distribution function (CDF) of Spotify Popularity illustrated how popularity accumulates across different track scores, reinforcing the observed positive trend.

Missed Aspects and Additional Variables

While the analysis provided valuable insights, it may have missed the impact of additional factors such as genre, artist popularity, or track release date, which could influence both Track Scores and Spotify Popularity. Incorporating these variables might offer a more comprehensive understanding of the dynamics driving Spotify Popularity.

Assumptions and Potential Issues

One assumption made was that the relationship between Track Scores and Spotify Popularity is linear. While the scatter plot suggested a positive trend, this does not rule out potential non-linear relationships that could be explored further. Additionally, we assumed that the data quality was consistent across platforms and that outliers had been appropriately handled. However, further analysis could reveal whether outliers significantly affect the relationship between Track Scores and Spotify Popularity.

Challenges and Understanding

One challenge faced was determining the exact nature of the outliers. Although histograms and descriptive statistics pointed out extreme values, understanding their impact on the overall analysis required more detailed investigation. Additionally, while fitting distributions and conducting regression analysis, ensuring that the assumptions of the models (e.g., normality, linearity) were met proved to be complex.

In summary, the EDA indicated a positive association between higher Track Scores and higher Spotify Popularity, but the analysis could be enriched by including additional variables and exploring non-linear relationships. Addressing these aspects might lead to a more nuanced understanding of how Track Scores influence Spotify Popularity.