Hip Hop Popularity Analysis Proposal

Research Question
How do specific song characteristics—such as tempo, danceability, energy, and loudness—influence the popularity of hip hop tracks on streaming platforms like Spotify?

Variables and Visualizations

1. Tempo
- **Variable**: Tempo (beats per minute)
- **Visualization**: Scatter plot with tempo on the x-axis and popularity on the y-axis.
- **Justification**: This scatter plot will help reveal any correlation between the tempo of a song and its popularity, indicating whether faster or slower songs tend to perform better.

2. Danceability
- **Variable**: Danceability (a measure of how suitable a track is for dancing)
- **Visualization**: Histogram of danceability scores segmented by popularity quartiles.
- **Justification**: A histogram allows us to examine how danceability varies across popularity levels and to assess whether more danceable tracks are generally more popular.

3. Energy
- **Variable**: Energy (intensity and activity level of the track)
- **Visualization**: Box plot of energy levels by popularity tier.
- **Justification**: A box plot will help us observe the spread and median energy levels for different popularity levels, which could indicate whether high-energy tracks are more appealing.

4. Loudness
- **Variable**: Loudness (average decibel level of the track)
- **Visualization**: Scatter plot with loudness on the x-axis and popularity on the y-axis.
- **Justification**: This scatter plot and the correlation analysis will help determine if louder songs are generally more popular.

5. Lyrical Sentiment (Optional)
- **Variable**: Lyrical Sentiment (categorized as positive, neutral, or negative using sentiment analysis)
- **Visualization**: Pie chart showing sentiment distribution for popular vs. less popular songs.
- **Justification**: A pie chart will allow us to see if there's a dominant sentiment among popular tracks, giving insight into how lyrics might influence popularity.

Analysis Plan

1. Tempo
- **Method**: Calculate the correlation between tempo and popularity to assess any statistically significant relationship.

- **Assumptions**: We assume that tempo directly influences listener engagement and could impact popularity.

2. Danceability
- **Method**: Compare the median danceability scores for popular and less popular tracks using a histogram.
- **Assumptions**: Danceability scores are assumed to correlate with listener enjoyment, particularly in social or energetic contexts.

3. Energy
- **Method**: Use a box plot to examine energy level distributions across popularity tiers.
- **Assumptions**: Higher energy levels are assumed to attract more listeners due to their engaging nature.

4. Loudness
- **Method**: Calculate the correlation between loudness and popularity.
- **Assumptions**: Loudness contributes to a track's appeal, potentially making louder songs more popular.

5. Lyrical Sentiment (Optional)
- **Method**: Use sentiment analysis to categorize lyrics and assess their distribution across popularity levels.
- **Assumptions**: Sentiments in lyrics could emotionally engage listeners, influencing popularity.

Hypotheses and Expected Outcomes

1. Tempo
- **Hypothesis**: Faster tempos are likely associated with higher popularity.
- **Expected Outcome**: A positive correlation between tempo and popularity, where popular songs tend to have a moderate to high tempo.

2. Danceability
- **Hypothesis**: Higher danceability scores will correlate with higher popularity.
- **Expected Outcome**: Popular songs will likely have higher median danceability scores, indicating a preference for danceable tracks.

3. Energy
- **Hypothesis**: Higher energy levels are associated with popularity.
- **Expected Outcome**: Popularity tiers with the highest songs are expected to have higher median energy levels.

4. Loudness
- **Hypothesis**: Louder tracks may be more popular.

- **Expected Outcome**: A positive correlation between loudness and popularity.

5. Lyrical Sentiment (Optional)
- **Hypothesis**: Tracks with positive or intense (negative) sentiment might be more popular.
- **Expected Outcome**: Popular songs are expected to have a higher frequency of positive or intense sentiments.