EDA Project – Spotify Dataset

Consider you are Music Director/Mixing Engineer aiming to optimize new songs for popularity, leveraging insights from this dataset is key. Here are some questions categorized by analysis type:

1. Univariate Analysis:

1. What is the overall distribution of popularity scores across all tracks in the dataset? (Are most songs moderately popular, or is it skewed towards very high/low popularity?)
2. What is the average and typical range for duration_ms (song length)?
3. What are the most frequently occurring keys in the dataset, and what is their individual distribution?
4. How are tempo values distributed across all tracks? (Are songs generally fast, slow, or is there a wide spread?)
5. What is the distribution of acousticness scores? (Does the dataset lean towards acoustic or electronic sounds?)
6. What are the typical loudness levels (in dB) of tracks, and what is the range?
7. How is danceability distributed? (Are most songs highly danceable, or is there a mix?)

8. What is the distribution of energy levels in the dataset? (Are songs generally high or low energy?)
9. What are the most common time_signatures found in the music?
10. What is the distribution of speechiness? (Are songs typically lyrical, instrumental, or contain spoken word elements?)

2. Bivariate Analysis :

1. Is there a correlation between a song's duration_ms and its popularity? (Are shorter or longer songs more popular?)
2. How does danceability relate to popularity? (Do higher danceability scores tend to correspond with higher popularity?)
3. What is the relationship between energy and popularity? (Are high-energy tracks generally more popular than low-energy ones?)
4. Does loudness have a noticeable impact on popularity? (Are louder mixes preferred by listeners?)
5. Is there a relationship between acousticness and popularity? (Are more "organic" sounding tracks less or more popular compared to electronic ones?)

3. Multivariate Analysis

1. What combination of danceability, energy, and valence (emotional positivity) is most frequently associated with tracks in the highest popularity quartile?
2. Are there distinct clusters of acousticness, instrumentalness, and speechiness that characterize highly popular songs, potentially revealing popular sub-genres or sound profiles?
3. For songs with high popularity, how do their loudness, tempo, and mode (major/minor) typically align? (Can we identify a "popular mix recipe"?)

4. Time Series Analysis :

1. How has the average popularity of songs evolved over years? (Are songs becoming generally more or less popular over time?)
2. Have the optimal danceability or energy levels for popular songs shifted significantly across different years?
3. Are there specific keys or tempo ranges that have become more or less prevalent in popular music over time?
4. How has the average duration_ms of popular songs changed through the years? (Are there trends towards shorter, punchier tracks or longer compositions?)
5. Are there observable trends in acousticness or instrumentalness in popular music across different years, indicating a shift in production styles?