Assignment: Spotify Songs Analysis

Download the "spotify_songs.csv" file and import the dataset into RStudio. The dataset contains information about songs on Spotify. Use R programming to answer the following questions. **Note**: Since we do not cover all the methodologies required for the analysis (particularly for question 5) in the classroom, you are encouraged to use large language models to help you work on your assignment.

1. Dataset exploration:
   a. How many songs are there in total in the dataset?
   b. How many distinct playlists are there in the dataset?
   c. How many distinct artists are there in the dataset?

2. Track popularity visualization:
   a. Plot a histogram of the overall track popularity.
   b. Plot a histogram of the average track popularity by playlist genre.

3. Identify and discuss the features that make a song more "danceable." Use both regression and correlation analyses.

4. Perform a multiple linear regression to predict track popularity using numeric features and the genre of the playlist and answer the following questions
   a. As an agent of a label company looking for young talents and their songs, identify the top three important features in a song that make it more likely to be popular.
   b. Assess the goodness of fit of your model. If the model fits the data well, explain why you think it's possible to "quantify art." If not, discuss what component(s) might be missing.
   c. Evaluate whether your regression suffers from multicollinearity.

5. A record label company is interested in studying the evolution of "trendy" within the music industry. For instance, which brands or genres are popular each year? Are song titles getting shorter or longer over time? Does featuring a collaboration (e.g., 'feat. XXXX') make a song more popular?

Your report should be an executive summary with a maximum length of five pages. Please explain the key numbers in your analysis in plain language and provide proper visualizations to support your findings.