**Group 1: Regular Data Science Questions**

1. Calculate the average 'Release Year' for each 'Genre'.

Disco: 1977; Folk: 1984; Grunge: 1991; Pop: 1979.33; Progressive Rock: 1979; R&B: 1975.14; Rock: 1972.35; Rock and Roll: 1957; Soft Rock: 1992; Soul: 1971; Synth-pop: 1985

1. Calculate the median of the ‘Number on Charts for the Year’ column

1.0

1. Find the longest and shortest 'Length in Minutes' across all songs. Convert 'Length in Minutes' to a numerical format for this calculation.

Longest: 482 minutes; Shortest: 124 minutes

**Group 2: Multistep Hard Data Science Questions**

1. For each 'Genre', calculate the average 'Number on Charts for the Year' and rank the genres based on this average.

Rock, Grunge, Pop, Soul, Soft Rock, R&B, Disco, Progressive Rock, Rock and Roll, Synth-pop, Folk

1. Identify the 'Artist' with the highest average 'Number on Charts for the Year' for songs released after 1980.

Nirvana

1. Create a new column 'Decade' based on 'Release Year' and calculate the median 'Length in Minutes' for each decade.

1950s: 2.41 minutes; 1960s: 3.14 minutes; 1970s: 3.97 minutes; 1980s: 4.22 minutes; 1990s: 4.52 minutes; 2000s: 3.93 minutes

1. For songs longer than the overall median length, calculate the standard deviation of their 'Release Year'.

9.38 years

1. Find the most frequent 'Album' for the top 10 songs with the longest 'Length in Minutes'.

A Day at the Races

1. For each artist, determine the range of years they have been active (based on the Release Year of their songs). Which artist has the widest range?

Elvis Presley

**Group 3: Multistep Data Analysis and Machine Learning Questions (Using Different ML Techniques)**

1. Use a Support Vector Machine (SVM) to classify songs as either 'Pop' or 'Rock' based on 'Release Year' and 'Number on Charts for the Year'. Provide the model's F1-score.

0.875

1. Implement a Neural Network to predict the 'Number on Charts for the Year' based on 'Release Year' and 'Length in Minutes'. Report the Mean Squared Error (MSE) of the model.

294.29

1. Apply Principal Component Analysis (PCA) to reduce the dimensions of the numerical features and then use K-Nearest Neighbors (KNN) to classify songs into 'Pop', 'Rock', or 'Other' genres. State the accuracy of the model.

33.33%

1. Use Linear Regression to predict the 'Release Year' based on 'Number on Charts for the Year'. Report the R-squared value of the model.

-0.261

1. Apply a Random Forest Regressor to predict 'Length in Minutes' from 'Release Year' and 'Number on Charts for the Year'. Provide the model's Root Mean Square Error (RMSE).

95.69

1. Implement a Naive Bayes classifier to predict whether a song is from the '70s, '80s, '90s, or '2000s based on 'Genre' and 'Number on Charts for the Year'. Report the Precision of the model.

0.167