Questions:

Data Exploration:

Plot the distribution of different features present in the dataset using appropriate visualization techniques. Use libraries like matplotlib and seaborn to visualize various feature distributions. Histograms can be especially helpful.

I made hist plot for all numerical values
Analyze the correlation between different features using the .corr() function provided by pandas and visualize them using a heatmap. Explain your findings. Should there be any other variables added based on the entire dataset?

I find energy and loudness are very related, feels like I should remove it?
feels like tempo should be used for clustering

Model Enhancement:

Implement a method to handle outliers in the dataset before feeding it into the KMeans model. Use techniques like Z-score or the Interquartile Range (IQR) method. Explain your approach and how it helps in better model training.

IQR removes a lot more then zscore way, maybe use 3 for zscore threshold is just like that?

silhouette_score get a little higher after removing outliers, not sure why
Explore different clustering algorithms other than KMeans (such as DBSCAN and Agglomerative Clustering) and compare their performance.

kmeans is by far the best

dbscan is really bad, not sure why?

AgglomerativeClustering is super slow, only pick first 10000 samples for testing, don't really know how to improve , cluster number is smaller the better, but silhouette_score is always better than dbscan, not sure why?

Implement a method to determine the optimal number of clusters using methods like the Elbow Method or Silhouette Analysis. Discuss different techniques for finding the optimal number of clusters and explain your chosen method.

used elbow from teammate code
made code of silhouette_score, set sample_size=10000 to make it a lot faster

User Interaction and Feedback:
Improve the user interaction part of the code to accept multiple favorite songs and provide multiple recommendations by modifying the input gathering step.

not done yet!!, not sure what we need to do

Implement a method for collecting user feedback on the recommended songs by incorporating an interactive prompt after recommendations are made.

not done yet!!, not sure what to do with feedback

innovations:
made all operation as functions, so it's easy to adjust and test within one cell. just comment out stuff you don't need or tweak the varables
tried to add onehotencoder for genre, and tried to use pca to make calculation faster, not sure if helped on removing noise