



上图展示了各音频特征之间的相关性。根据常用标准，当预测变量之间的相关系数绝对值大于0.7时，通常认为存在较强的多重共线性问题。

分析结果

通过检查相关性矩阵，我们发现：

- 没有发现高度相关的特征对：**所有特征对之间的相关系数绝对值均小于0.7，表明模型中不存在严重的多重共线性问题。
- 最高相关性：**
 - energy与loudness之间的相关性相对较高（约0.69），但仍低于0.7的阈值
 - energy与acousticness之间存在较强的负相关（约-0.66）
 - 其他特征对之间的相关性普遍较低
- 独立性验证：**大多数预测变量之间的相关性较低，表明它们提供了相对独立的信息。

结论

基于相关性分析，我们的回归模型**不存在显著的多重共线性问题**。这意味着：

- 系数估计可靠：**回归系数的估计相对稳定，不会因为预测变量间的高度相关性而失真。
- 独立贡献明确：**每个特征对歌曲受欢迎程度的影响可以被视为相对独立的贡献。
- 解释性良好：**模型的解释性不会因为多重共线性而受到严重影响，我们可以相对准确地解释各个特征的影响。
- 无需变量选择或降维：**不需要通过剔除变量或主成分分析等方法来解决多重共线性问题。

虽然模型的整体拟合度较低（ $R^2 = 0.0905$ ），但这并非由多重共线性引起，而是因为歌曲受欢迎程度受到许多我们未能在模型中捕捉的因素影响，如艺术家知名度、营销推广、文化背景等非音频特征因素。

总之，我们的回归模型在多重共线性方面是健康的，各个预测变量提供了相对独立的信息，使得我们能够可靠地识别影响歌曲受欢迎程度的关键特征。