**问题三**

**1. 概要**

本问题的目标是利用已知类别玻璃文物的化学成分数据，结合未分类文物的化学成分数据，通过聚类分析方法对未分类玻璃文物进行科学归类。

**2. 数据描述**

本问题涉及两组数据：

* **已分类玻璃文物化学成分数据**：包含了已知类别的玻璃文物的化学成分比例。这些文物的类别包括高钾玻璃和铅钡玻璃。
* **未分类玻璃文物化学成分数据**：包含了未分类的玻璃文物的化学成分比例，需通过模型预测其类别。

**3. 数据预处理**

为了确保模型能够正确分析和处理数据，首先需要对数据进行预处理：

1. **缺失值处理**：将所有缺失值填充为0，以防止在后续的分析中出现错误。
2. **特征一致性**：由于已分类和未分类玻璃文物的化学成分列可能不一致，需确保两者的特征列完全一致。这是通过取两组数据的共同列来实现的。

**4. 方法与模型**

本研究采用了**K-Means聚类**方法对未分类玻璃文物进行分类。具体步骤如下：

1. **数据标准化**：对化学成分数据进行标准化处理，使得每种成分在分析时具有同等重要性，避免由于量纲不同导致的偏差。
2. **主成分分析（PCA）降维**：为了便于可视化和提高计算效率，采用PCA方法将高维化学成分数据降至二维。
3. **K-Means聚类**：使用K-Means算法将文物分为两类（假设有两种主要类别：高钾玻璃和铅钡玻璃）。在已知数据上训练模型后，对未分类文物进行预测。

**5. 结果与分析**

通过PCA降维后的数据可视化显示，已知类别的玻璃文物在二维空间内形成了两个较为明显的聚类。通过K-Means聚类算法，对未分类文物的数据进行了聚类分析，并对其类别进行了预测。预测（结果见 *分类结果.xlsx*）预测结果显示，未分类文物被成功地划分到了与已知文物相似的类别中。图表, 散点图

描述已自动生成

这种分类方法具有以下优点：

1. **无需监督**：K-Means算法无需预先提供标签，能够自主发现数据的内在结构。
2. **解释性强**：PCA降维后的数据能够直观地展示不同类别文物的分布情况，为分类结果提供了直观的支持。

**6. 结论与展望**

通过化学成分分析与K-Means聚类算法的结合，成功地对未分类玻璃文物进行了科学归类。

未来的研究可以进一步优化分类算法，尝试引入更多的机器学习方法，如层次聚类或监督学习算法（如支持向量机、随机森林等），以进一步提高分类的准确性。