数据预处理是数据分析和机器学习过程中一个非常关键的步骤。它主要包括以下几个步骤：

数据清洗：处理缺失值、重复数据和异常值。常见的方法有：

填补缺失值（如使用均值、中位数、众数等）。

删除含有缺失值的记录。

使用插值法或模型预测来填补缺失值。

处理重复数据，通常是删除重复记录。

处理异常值，可以使用统计方法（如三倍标准差法）或模型检测异常值。

数据集成：将来自不同来源的数据进行整合。常见的方法有：

合并数据表（如使用 SQL 的 JOIN 操作）。

数据库中的数据聚合。

数据转换：

归一化/标准化：将数据缩放到相同的范围内。常见的方法有 Min-Max 归一化、Z-score 标准化。

编码：将分类变量转换为数值变量。常见的方法有独热编码（One-Hot Encoding）、标签编码（Label Encoding）。

特征工程：从现有数据中提取或创造新的特征。可以包括特征组合、特征分解等。

数据降维：减少特征数量，常见的方法有：

主成分分析（PCA）。

线性判别分析（LDA）。

特征选择：如基于方差、基于模型的重要性等方法。

数据分割：将数据集划分为训练集和测试集（有时还包括验证集）。常见的方法有：

随机划分。

按时间序列划分（对时间序列数据）。

交叉验证（如 k 折交叉验证）。

我们的代码功能：

读取数据：读取包含谷歌指数数据的 CSV 文件，假设文件中有两列数据：date（日期）和 search\_index（搜索指数）。

日期处理：将日期列转换为日期格式，并将其设为索引，以便进行时间序列分析。

归一化处理：使用 MinMaxScaler 将搜索指数归一化到 [0, 1] 的范围。

可视化：使用 Matplotlib 进行可视化，比较归一化前后的搜索指数。