**问题一：数据清洗与预处理分析思路与实现方案**

### 1. ****问题目标****：

根据提供的电站数据（包括发电量、环境监测数据、天气数据等），我们需要对数据进行清洗与预处理，为后续分析（如清洗时机预测、积灰检测等）做好准备。任务主要包括：

数据缺失值的处理

异常值的处理

时间格式统一与数据按小时汇总

### 2. ****数据清洗与预处理步骤****：

#### 步骤 1：数据导入与时间格式统一

读取所有电站的发电数据、环境监测数据和天气数据，并确保时间字段统一格式为 datetime 类型。

将时间戳统一为小时级别（即将分钟和秒钟的部分统一为整点）。

#### 步骤 2：处理缺失值

对于发电量和环境数据中的负值，视为异常值（例如，发电量为负数），应将其视为缺失值（NaN），并通过线性插值法填充缺失值。

#### 步骤 3：异常值处理

检查数据中是否存在不合理的值，如异常的发电量（过高或过低的值）或其他环境数据（例如辐照强度过高或过低）。

#### 步骤 4：按小时聚合数据

按小时对所有电站的数据进行汇总，发电量使用累加方法，其他环境数据（如温度、辐照度等）则使用平均值进行聚合。

#### 步骤 5：合并所有数据

将每个电站的发电数据、环境数据和天气数据按时间进行合并，确保每个时间点都有对应的发电量、环境数据和天气信息。

#### 步骤 6：添加元信息

为每条记录添加电站编号（如电站1、2、3、4）和装机容量等信息，便于后续分析。