**数据分析工程师技术笔试题目**

# 答题要求说明

## 1.笔试题目说明

附件为具体的笔试题目，其中：

【一、业务和数据导读】部分是对电力一些基本概念的讲解以及数据内容的描述，作为笔试的行业知识背景，请在答题前仔细阅读并理解。

【二、三】为具体的考核内容，请根据题目要求针对性作答。可将相应结果内容直接补充到题目要求下方，需要提供佐证材料的需要单独提交附件的，请标注对应的附件名称。

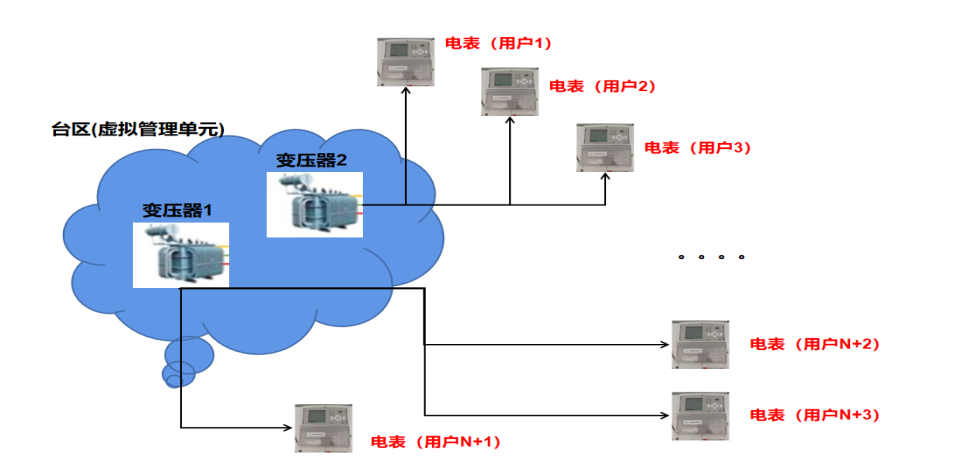
## 2.答题注意事项

分析题重点考核分析思维，技术考核题部分侧重考核数据处理和观测分析能力，总体要求90分钟内完成，请大家根据题目设置合理分配答题时间。部分题目灵活性高可自由发挥，面试官会根据作答情况适当予以考虑。

# 附件一：笔试题目

## 一、业务和数据导读

电网在对供电设备和用电用户进行管理时通常会通过划分成不同的层级的管理单元来开展，其中其中供电范围和台区是常见的2个管理概念。供电单位可以理解为省市区县不同的管理组织机构，台区是配电网管理的一个较小设备单元。以下是台区的拓扑关系的简易视图：



可以近似理解，“台区-变压器-电表”可以建立逐层对应的关系，其中台区与变压器一对多，变压器与电表一对多。

其他概念说明：行业分类、城乡分类、用电类别是从不同维度标记用户分类，通常用标准代码标识。

现有以下数据表：

（1）电表用电数据：包含电表编号，对应变压器编号，日期以及每6个小时的电量；

（2）台区与变压器对应关系：包含台区编号，变压器编号；

（3）台区统计信息：台区下对应的各类用户的数量及容量统计信息，共 2个表格；

根据提供的4张excel数据表，请选择自己最熟练掌握的技术工具（如sql、python、Excel其中一种或多种），完成以下技术考核内容，并请标注具体使用的技术工具类型。

## 二、分析题（20分）

用户用电并缴纳相应的电费是电力营销的基本业务。近年来，缴费业务发生了很大的变化，主要表现在以下方面：①缴费的渠道变多，除了上门收费、电力营业厅缴费、银行代扣、代收机构缴费等常规方式，支付宝、微信、掌上电力、网上国网APP等渠道的使用率在不断提高；②缴费方式后由付费模式（先用电后根据电力系统按期抄表结算发行的电费账单进行缴费）、充值卡（预购电量，用完后充值）等转变为可以预存电费（主动存一笔金额到电力账号）。通过不同渠道或不同付费方式进行缴费，用户缴纳电费的及时性、缴纳电费的次数等也体现出不同的特征。

根据以上描述，若要对某一地市用户的缴费行为特征开展画像分析，你认为可以按照什么样的逻辑开展，请简要阐述你的分析思路。

## 三、技术考核内容（80分）

技术工具选型：（请在此备注说明）

## 3.1异常处理（共20分）

对所提供的数据表进行观测，从数据中发现了什么问题（如空值、异常值等）？针对发现的问题是怎么处理的？请做简要说明，可以用图表佐证。

（*需提交最终处理后的完整Excel数据清单，现场笔试使用SQL的，可以直接备注数据库表名并提交代码*）

3.2数据加工（共30分）

（1）数据聚和：将两张台区信息表进行拼接并统计最终的记录数（4分）；

（2）数据去重：对三份数据分别进行删除重复项并统计每份数据最终的记录数（6分）；

（3）求和：对电表用电数据进行处理，得到各电表每日用电量，并计算所有电表日电量的最大值、最小值及发生时间。（8分）

（4）分窗计算：根据合并后台区信息表，计算每条记录用户数占对应台区用户总数的占比，并放入存入相应的统计结果表中（6分）；

（5）数据关联：将三份数据进行关联拼接，并统计最终得到的记录数（6分）；

（6）分组求和：对关联后的数据进行处理，得到①台区日电量②各一级单位每日用电量，并提交数据结果表格（10分）；

（*需提交最终处理后的完整Excel数据清单，现场笔试使用SQL的，可以直接备注数据库表名并提交代码*）

3.3可视化及算法预测（30分）

（1）根据前述数据处理与统计结果，你能得到什么结论？通过图表进行可视化展示来支撑你的结论。（图表可用Excel生成）

（2）请任意选择一种算法，构建预测模型对二级单位未来一天用电量进行预测，输出预测结果。（不要求准确性，需提交代码和预测结果。）