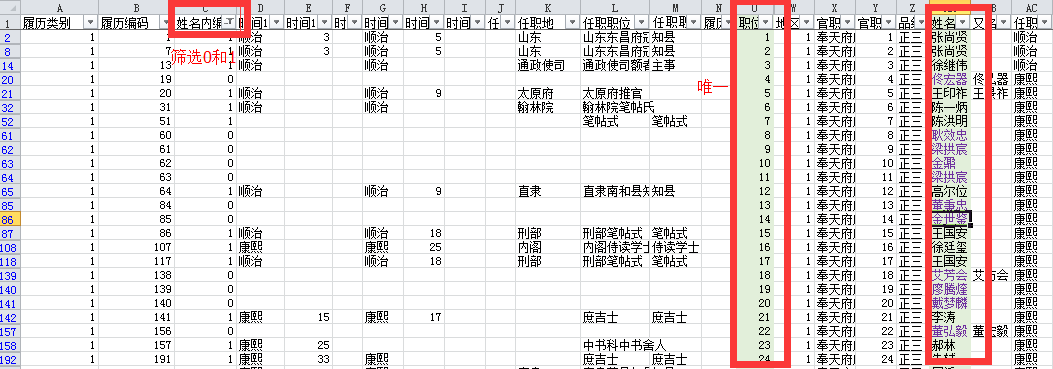
1. **数据处理前的加工**
2. 核对各项数据，是否符合数据库分析处理的格式，然后集中修改。
3. 将其中的纪年方式改为公元纪年，D列“时间1朝代”+E列“时间1年” 换算后填入列 F列“时间1公历年”。同理，“时间2朝代”和“时间2年”同样转换后填入“时间2公历年”如下图：

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 列 | C | D | E | C | H | I |
| 表头 | 时间1朝代 | 时间1年 | 时间1公历年 | 时间2朝代 | 时间2年 | 时间2公历年 |
| 内容 | 乾隆 | 40 | 1775 | 乾隆 | 40 | 1775 |

3、列O前的履历信息数据说明：



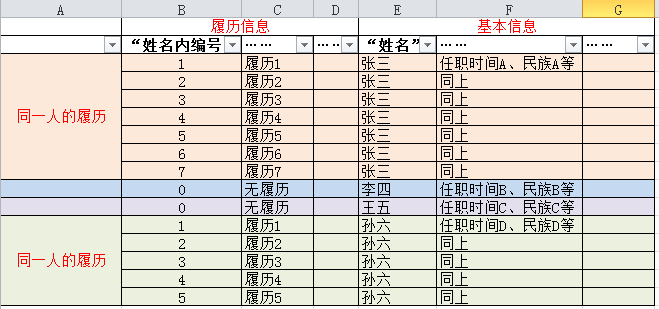
列C“姓名内编号”：如果官员无履历信息则编号为0，如果有履历信息为1至n,如上图中1、2……6则表示某人共有6条履历信息，但是列O “职位总序号”都为1，说明这是同一人的履历信息。所以只要筛选列C“姓名内编号”为0和1的数据就能保留下一个官员唯一的一条基本信息。（因为每条履历信息后面对应的基本信息都是一样的。例如上图中某人有1至6条履历信息，其列O后面的基本信息都是一样的，也就是有重复的6行，筛选后只保留姓名内编号为1的这条，也就只剩下此官员一条信息了。）筛选后如下图：



可见，上图中“姓名”一列的官员姓名是按照任职时间先后排列的。

注：为了直观了解数据库的数据构成，将简化后的数据库如下图展示出来。

原始数据库如下：



筛选“姓名内编号”为“0”和“1”之后的数据库，是按照任职时间A到D进行排列的，与上图相比省略了一部分履历信息。如下图。

