**数据库数据数据处理工作方案**

**一、研究计划：**

目前有三个职官的数据库，研究计划如下:

数据库1 数据量小

数据库2 数据量大

其一需要对数据库1和数据库2的数据进行全面的比较，包括任职时间、个人身份信息、以及出身情况等，输出统计结果与数据。并且对数据库1、2、3的官职、姓名和科举进行比较。

其二，以数据库2为基础，与相关数据库连接，对官员任职流动情况进行分析。

**二、具体分工：**

数据搜集整理-李荣倩（lirongqian@sjtu.edu.cn 电话 18217293515）

数据比对、分析-王洋 （flybird\_wy@foxmail.com 电话 18616730721）

**三、工作进程与预算：**

1，第一阶段：（2500元）两个数据库的比较：数据库1和数据库2的比较；数据库1、2、3的比较。

2，第二阶段：（2500元）以数据库2为基础，与其他数据库连接，进行相关分析。（数据库2需要补充数据）

**四、工作内容**

**（一）第一阶段**

**1、对象**

两个数据库的比较： 数据库1和数据库2，数据库1、2、3的“科举”比较。

（1）数据库2大约800多条数据，其中姓名有重复，因不同时间任多个官职，容量大

（2）数据库1大约800多条数据，姓名也有重复（因不同时间任多个官职），摘取的某几个年份的职位所对应的官员，相对于数据库2的官员较少，容量小。

（3）数据库3，包含数据库1/2的部分信息。

**2、**计算有姓名、官职、民族、旗分、科举信息记录占总记录的百分比。

**3、**以时间上的“同一季”、“连续两季”、“连续三季”、“一年”分别比较：

（1）比较**①**，官职、姓名的匹配情况。

要求：输出的结果——匹配的百分比，未匹配的百分比，输出匹配与未匹配的数据（两个excel表格）

（2）比较**②**，在①的基础上，“民族”的匹配情况。

要求：输出的结果——匹配的百分比，未匹配的百分比，输出匹配与未匹配的数据（两个excel表格）

（3）比较**③**，在①②的基础上，“旗分”的匹配情况。

要求：输出的结果——匹配的百分比，未匹配的百分比，输出匹配与未匹配的数据（两个excel表格）

（4）比较**④**，在①②③基础上，“科举”的匹配情况。

要求：输出的结果——匹配的百分比，未匹配的百分比，输出匹配与未匹配的数据（两个excel表格）

（5）比较**⑤**，通过①②③④，数据完全匹配的情况

要求：输出的结果——匹配的百分比，输出匹配的数据（一个excel表）

（6）比较**⑥**，在同一“官职”、“姓名”基础上，单独比较“科举”的匹配情况。

要求：输出的结果——匹配的百分比，未匹配的百分比，输出匹配与未匹配的数据（两个excel表格）

**4、输出的结果和形式：**

4个比较的匹配与未匹配的百分比、匹配与未匹配数据的表格。

**5、图形示意**

**⑤**

**④**科举

**③**旗分

**②**民族

**①**官职、姓名

**⑥**科举

官职、姓名

**6、结合数据库3**

* 计算数据库1与数据库3人名和科举相同的百分比，并列出数据库3与数据库1相同的记录（即人名和科举相同的记录）
* 计算数据库2与数据库3人名和科举相同的百分比，并列出数据库3与数据库1相同的记录（即人名和科举相同的记录）
* 计算数据库1、数据库2和数据库3人名和科举相同的百分比，并列出数据库3与数据库1相同的记录（即人名和科举相同的记录）

**（二）第二阶段**

**1、对象：**数据库2 与其他数据库连接

**2、数据库：**需在第一阶段 数据库2的基础上补充官职——“副都统”的数据形成新的数据库2，目前尚未完成此数据库。

**说明：**

**1、基本概念**

|  |  |  |
| --- | --- | --- |
|  | **所用概念** | **表中数据** |
| 民族 | 民人 | 汉 |
| 旗人 | 满洲、汉军、蒙古 |
| 季 | 春季 | 1 |
| 夏季 | 2 |
| 秋季 | 3 |
| 冬季 | 4 |
| 同一季 | 表中“季”的数据相同 |  |
| 连续两季 | 表中“季”的数据差1， |  |
| 连续三季 | 表中“季”的数据相差2， |  |
| 一年 | 表中“季”的数据相差3， |  |

**2、工作进度**

* 首先完成第一阶段的工作，验收后支付2500元酬金
* 当数据库2所有数据补充完善后，进行第二阶段的工作，验收后支付2500元酬金。

**3、要求**

* 要求数据分析后提交相关结果。
* 将每项工作进程的程序语言一同打包提交，方便我们发现不理解的分析过程，按照程序语言对照理解。