

第一次选拔题：数据恢复与归类

处理数据是现代大数据时代的一项非常重要的工作，尤其是数据的恢复和归类整理技术，更是关键与核心。

本题目附带一份残缺数据表（见附件），表中所提供的数据，是自去年新冠疫情爆发以来，直到今年 2 月，某平台公布的部分国家（共 19 个国家，国家名已经隐去）的疫情数据。为了论文建模的需要，其中有些数据已经被特意删除，数据表的第一行是对应国家的大约人口数。由于人工输入数据的客观事实，不排除某些国家上报的数据滞后，使得平台公布的个别数据出现异常，需要各个队伍完成以下的任务：

- A. 检查表中是否有异常数据，有哪些是？如何修改？
- B. 对于这 19 个国家，大部分都有局部数据缺失，请对缺失的数据作恢复。
- C. 对于国家 04、国家 07、国家 12、国家 17，对它们的起源数据进行恢复。
- D. 根据已有数据，对所有国家在今年 3 月 1 日~5 日的疫情做出预测。
- E. 对这 19 个国家进行两种以上形式的分类，并给出分类的依据和结果。

本题目同时还训练各个队伍的写作能力，要求每支队伍按照比赛的要求与模式写作论文答题。对于任务 ABCD，请将你们计算得到的完整结果填入到 Excel 表（附件 1）相应位置，你们填入或修改的数据，请用**红色 12 号字体加粗**（见附件 D308 数据示例）。对于任务 E，只需要将结果写在论文中。

注意：请将你队伍的论文和结果文件，[以附件的形式于 7 月 30 日 21 时之前发到邮箱 shizi0906@126.com](mailto:shizi0906@126.com)。

为方便统一管理，请规范你们的论文，结果文件，邮件的命名：

论文命名为“队号+第一次论文”，比如说“Z17 第一次论文”。

结果文件命名为“队号+第一次结果”，比如说“Z17 第一次结果”。

邮件的主题是“队号+第一次训练题”，比如说“Z17 第一次训练题”。