**数据说明**

本次比赛构建了一个模拟现实招聘市场的虚拟数据集，旨在为参赛者提供一个深入分析和探索就业趋势的丰富资源。该数据集包含了430,664条详尽的招聘通知，覆盖了169,540种多样化的职位、267,296家不同企业、158个行业类别，并涵盖了371个行政区划。数据集中设置了缺失值、异常值或数据不一致等噪声信息（例如无效的薪酬、缺失的行业类别等），需要参赛者注意并处理。数据以xlsx格式提供。

|  |  |  |
| --- | --- | --- |
| JobWanted.xlsx | | |
| **字段名称** | **字段含义** | **相关说明** |
| job\_title | 职位名称 | 职位名称是指该职位在公司中的正式称谓，它通常描述该职位的主要工作内容和职责，比如软件工程师、产品经理、人力资源专员等。本数据为每种职位分配唯一的虚拟代码，示例为：51b08e609b8cd069a426306d4deec371eK |
| city | 行政区划 | 企业所在的行政区划 |
| salary | 薪酬 | 薪酬是指该企业为员工工作所支付的金钱或福利的总额。本数据构建的虚拟薪酬示例为：5-10K、5-10K·13薪 |
| experience | 经验要求 | 经验要求是指该企业对申请该职位的候选人所需具备的工作经验的年限或水平的说明，比如1-3年、应届等。本数据为每种经验要求分配唯一的虚拟代码，示例为：EdD |
| education | 学历要求 | 学历要求是指该企业对申请特定职位的候选人所需具备的教育背景的说明，比如学历不限、本科等。本数据为每种学历要求分配唯一的虚拟代码，示例为：GI |
| company | 企业 | 发布招聘通知的企业 |
| company\_type | 行业类别 | 行业类别是指该企业主要经营活动的市场领域或经济部门，比如医疗保健行业、金融服务行业等。本数据为每种行业类别分配唯一的虚拟代码，示例为：type\_BLfSmG |