**数据**

本案例所采用数据为“financejob.csv”文件，主要字段描述如下：

|  |  |  |
| --- | --- | --- |
| 变量  英文名称 | 变量  中文名称 | 变量描述 |
| company | 公司名称 | 字符型变量。例：联合资信评估有限公司。 |
| nature | 公司属性 | 定性变量。共7个水平，分别为：国企，民营，股份制企业，上市公司，合资，外商独资，其他。 |
| size | 公司规模 | 定性变量。共6个水平，分别为：10000人以上，1000-999人，500-999人，100-499人，20-99人，20人以下。 |
| sector | 公司行业大类 | 定性变量。共4个水平，分别为：基金/证券/期货/投资，信托/担保/拍卖/典当，银行，保险。 |
| subsector | 公司子行业 | 定性变量。共9个水平，分别为：基金，证券，期货，信托，担保，银行，保险，租赁，其他。 |
| position | 招聘岗位 | 具体的招聘岗位名称。 |
| category | 岗位类别 | 字符型变量。共200个类别的职位，例：证券分析/金融研究，股票/期货操盘手。 |
| impposition | 重点岗位 | 定性变量。共7个水平，分别为：风控岗，投行岗，项目管理岗，销售岗，研究岗，交易岗，其他。 |
| bond | 债券岗位 | 二值变量。1为债券类岗位。 |
| descrip | 岗位职责 | 字符型变量。例：岗位职责：1）寻找项目投资机会，包括债权和股权项目；2）项目分析，包括对接融资方、企业尽调、项目分析、上会讨论等；3）项目签约并落实风控；4）贷后管理。 |
| lowsalary | 最低薪资 | 数值型变量。取值范围：1000-100000元/月。 |
| highsalary | 最高薪资 | 数值型变量。取值范围：2000-150000元/月。 |
| avesalary | 平均薪资 | （最低薪资+最高薪资）/2。取值范围：1750-115000元/月。 |
| education | 学历 | 定性变量。共4个水平，分别为：不限，大专，本科，硕士及以上。 |
| experience | 经验 | 定性变量。共5个水平，分别为：不限，1-2年，3-4年，5-6年，7年及以上。 |
| coding | 证书编程 | 定性变量。证书：从业资格证、司考、CPA。编程：R、Python、SQL等。 |
| charact | 特质 | 定性变量。4个水平，分别为：责任心、学习能力、知识、团队意识。 |

**代码**

代码运行需知：

1. 请在每份code中自行设置更改自己的工作路径，并将上述所提到的文件放到该工作路径下。
2. 文件financejob.csv为清洗后可供分析的数据。
3. 本案例的分析软件为R语言，代码共由3部分组成，三个代码可单独运行。
4. 若按code1 – code3的顺序运行代码，则代码顺序和PPT呈现顺序一致。
5. 词云分析（文本挖掘）对应代码code1-wordcloud.R用到了三份停词文件——“stopwords1.txt”，“stopwords2.txt”以及“stopwords3.txt”来去掉不相关词汇。
6. code中所涉及画图部分区分mac和windows系统，mac系统请运行mac代码段，windows系统请运行win代码段。在画图时，如遇到提示系统未安装该字体可以忽略或者自行安装相应字体。
7. 请在UTF-8编码形式下运行code。