广州商学院

信息技术与工程学院（部）2021—2022学年第二学期

《数据分析与处理实验》考查内容及评分标准

**题目：自拟**

任选一个商业网站，使用网络爬虫技术采集该网站的数据，利用数据分析技术对采集的数据进行数据分析作为期末的论文。

具体要求：

1、编写爬虫程序，将采集到的商业网站数据存储到CSV文件或者Excel文件中，作为数据分析的数据来源。

2、编写Python程序对采集的数据进行数据处理。

3、将处理好的数据，完成5-6个模块的数据分析。

4、编写的数据处理程序、数据分析程序要运行正常，在技术层面要吻合教学过程中涉及到的知识，如：数据分析所需要的模型或者算法的选取、应用窗体的搭建（UI）、图表的绘制、有效价值的结论等。

《数据分析与处理实验》论文评分标准

**一、论文成绩各部分的依据和量化如下：**

1、 论文（50%）其内容要符合《广州商学院课程论文撰写基本规范》。

2、 答辩（50%）其内容包含如大作业评分标准所述。

**二、课程论文评分标准：**

**课程论文成绩评定表**

|  |  |
| --- | --- |
| 评价项目 | 评价内容 |
| 1.创新性、合理性（5分） | 5分：表明设计具有创新性、合理性；  3-4分：表明设计具有新意、合理性；  1-2分：表明设计具有合理性。 |
| 2.选题难度、实现复杂性（10分） | 9-10分： 难度系数高，设计实现功能复杂；  7-8分：难度系数低，设计实现功能较复杂；  6分及以下：难度系数很低，设计实现功能一般。 |
| 3.完成情况包括数据分析的应用、工作量（60分） | 52-60分：有完整的符合标准的文档，有正确的数据可视化技术设计方案，设计较全面，优化得当；  44-51分：有完整的符合标准的文档，有正确的数据可视化技术方案，设置无明显错误；  36-43分：有完整的符合标准的文档，有基本的数据可视化技术，设计方案正确，但有少数失误；  36分以下：有相对完整的符合标准的文档，有基本的数据可视化技术，设计方案基本正确，但缺少部分必要功能，存在问题。 |
| 4.设计报告（25分） | 23-25分：优秀。表明报告清楚详细地说明了设计的过程，格式规范、内容完整，阐述清晰，层次分明；  19-22分：良好。表明报告说明了设计的过程，格式较规范、内容比较完整，阐述清晰，有层次；  15-18分：中。表明报告说明了设计的过程说明不太清晰，格式不规范、内容比较简单；  15分以下：差。表明报告过于简单。 |
| 总分 | （100分） |

**三、答辩环节检查评分标准：**

|  |  |
| --- | --- |
| 一级指标 | 二级指标 |
| 爬虫内容（20分） | 1.爬取的内容积极健康，信息数据充实、准确；（10分） |
| 2. 爬取的数据主题突出，能围绕一个主题及相关内容进行爬取；（5分） |
| 3. 数据存储到CSV文件或者Excel文件，设计合理，无乱码或其他非正常的字符；（5分） |
|  |  |
| 数据处理（20分） | 1. 编写Python程序对采集的数据进行数据预处理；（10分） |
| 2.预处理能完成数据格式标准化，异常数据清除，错误纠正，重复数据的清除等；（10分） |
|  |  |
|  |  |
| 技术应用（60分） | 1. 能正确应用数据分析所需要的模型或者算法（10分）； |
| 2. 能正确完成窗体的搭建（UI），UI能正常转为Python程序（20分） |
| 3. 能正确进行数据分析，进行图表的绘制，并提取出有价值的信息，形成有效的结论。（30分） |
| 总分 | （100分） |
|  |  |