**本科生课程设计报告**

**《XXXX数据可视化分析》**

**课程名称《大数据可视化技术》**

**姓 名**

**学 号**

**学 院**

**专 业**

**课程性质 专业必修课**

**任课教师 黄诚惕**

**注意事项：**

1. 课程设计报告要有题目、摘要、关键词、正文及参考文献。摘要100-200字左右，关键词3～5个；参考文献不少于5篇；论文字数不少于1000字；要有自己的见解与心得；附查重报告，且查重率不高于20%；
2. 课程论文应符合一般学术规范，严禁抄袭或应付了事；
3. 课程论文均须用A4纸双面打印(查重报告附后)，加装本封面，左侧装订。

|  |  |  |  |
| --- | --- | --- | --- |
| 教师评阅意见： | | | |
| 论文成绩 |  | 评阅日期 |  |

**课程设计内容描述**

以下是为《大数据可视化技术》课程设计的5个参考题目，请任选一个完成：

### ****课题1：城市交通流量时空可视化分析****

****目标****：基于时空数据探索城市交通模式与拥堵规律  
****要求****：

1. ****数据预处理****：

* 清洗出租车轨迹数据（如缺失坐标、异常速度），提取时间-空间特征（如高峰时段、区域热力）。
* 融合外部数据（如天气、节假日）分析拥堵诱因。

1. ****可视化设计****：

* 动态热力图展示不同时段区域热度变化。
* 轨迹线聚类分析（如机场→市区热门路线）。
* 交互功能：时间轴滑动对比工作日/周末差异。

1. ****技术栈****：自选可视化工具  
   ****数据集****：

* 纽约出租车公开数据集（NYC TLC Trip Data）：含经纬度、时间戳、费用等字段（[链接](https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page" \t "https://yiyan.baidu.com/chat/_blank)：）。

### ****课题2：社交媒体用户关系网络可视化****

****目标****：分析社交网络中的用户影响力与社群结构  
****要求****：

1. ****数据预处理****：
   * + 从Twitter/微博API抓取用户关注关系，构建加权网络图（节点为用户，边为互动频率）。
     + 提取用户属性（如粉丝数、发帖频率）作为节点元数据。
2. ****可视化设计****：
   * + 力导向图展示核心用户节点及传播路径。
     + 社群检测算法（如Louvain）高亮子群，并用颜色区分。
     + 交互功能：点击节点展开用户属性及关联话题。
3. ****技术栈****：自选可视化工具。  
   ****数据集****：

* 斯坦福网络分析项目（SNAP）的Twitter数据集，SNAP: Network datasets: Social circles（ https://snap.stanford.edu/data/egonets-Twitter.html）。

### ****课题3：电商用户行为路径分析****

****目标****：通过用户点击流数据优化网站转化率  
****要求****：

1. ****数据预处理****：
   * + 解析用户会话日志（如页面跳转、停留时间），构建用户旅程图。
     + 标记关键事件（如加入购物车、支付）。
2. ****可视化设计****：
   * + 桑基图展示用户从浏览到购买的路径流失。
     + 漏斗图对比不同营销活动的转化效果。
     + 交互功能：筛选特定用户群体（如新客/老客）或设备类型。
3. ****技术栈****：自选可视化工具  
   ****数据集****：

* 阿里云天池电商用户行为数据集（含用户ID、商品ID、行为类型等字段，需注册下载）。

### ****课题4：新闻文本主题演化可视化****

****目标****：追踪新闻热点随时间的变化趋势  
****要求****：

1. ****数据预处理****：
   * + 对新闻标题/正文进行分词、LDA主题建模，提取主题关键词。
     + 过滤低频词，合并同义词。
2. ****可视化设计****：
   * + 词云时间轴动态展示主题词变化。
     + 堆叠面积图对比不同主题热度周期。
     + 交互功能：点击主题词跳转原始新闻列表或关联事件。
3. ****技术栈****：自选可视化工具。  
   ****数据集****：

* GDELT全球新闻事件数据库（按日期/主题分类，[链接](https://www.gdeltproject.org/" \t "https://yiyan.baidu.com/chat/_blank)）。

### ****课题5：环境传感器数据多维分析****

****目标****：探索空气质量与气象因素的关联关系  
****要求****：

1. ****数据预处理****：
   * + 融合PM2.5、温湿度、风速等多源数据，处理时间对齐问题。
     + 标记异常值（如传感器故障导致的数据跳变）。
2. ****可视化设计****：
   * + 平行坐标图展示多变量关联（如PM2.5与风速的反比关系）。
     + 地理散点图叠加气象站位置与污染等级。
     + 交互功能：刷选特定天气条件（如降雨/高温）或时间段。
3. ****技术栈****：自选可视化工具。  
   ****数据集****：

* 中国空气质量在线监测分析平台（含全国站点实时数据，[链接](https://www.aqistudy.cn/" \t "https://yiyan.baidu.com/chat/_blank)）。

### ****通用课设要求****

1. ****报告规范****：
   * + 行文有条理、结构合理、格式统一规范（字体、锻炼、章节等，可参考工学院设论文格式规范）
     + 需包含问题定义、数据预处理流程、可视化设计思路、技术实现细节及结论。
     + 附代码注释和可视化截图。
2. ****代码提交****：
   * + Jupyter Notebook或GitHub仓库（含数据清洗脚本+可视化代码）或可以执行的代码压缩包（需附上开发环境介绍及可运行的完整代码及相应注释说明）。
3. ****演示要求****：
   * + 5分钟视频展示交互功能，重点说明设计选择背后的可视化原则（如避免误导性图表）。

****评分重点****：

* 数据处理的合理性（20%）
* 可视化编码的有效性（40%，如颜色/尺寸映射是否直观）
* 交互功能的创新性（30%）
* 报告与演示的清晰度（10%）

**注：请独立完成，同一题目不得有内容雷同！**

除要求的可视化设计内容，可自行添加个性化设计部分，包括交互功能，这两部分都又比较大的分值比重，有创新的可以加分。

可根据同学们的兴趣调整题目（但难度应不低于参考的5个题目），鼓励使用真实业务场景数据以增强实践价值。

提交截止时间：2025年6月20日