《数据科学导论》课程作业

本文由《数据科学理论与实践》（清华大学出版社）的作者朝乐门提供，是数据科学课程中的一些必做题目。

1.调查分析题

1. 结合自己的专业领域，调研数据科学及大数据在自己所属领域中的应用现状。
2. 结合自己的专业领域和研究兴趣，调查分析数据科学的典型和最佳实践。
3. 调查分析数据科学领域典型案例中常用的方法、技术和工具。
4. 调查分析数据科学与大数据技术相关的竞赛类项目，并给出自己的参与方案。
5. 调查分析数据科学与大数据技术相关的开源项目，并给出自己的参与方案。
6. 调查分析近3年在数据科学领域出版的专著。
7. 调查分析近3年的The Data Science Journal等数据科学领域的学术期刊上发表论文的主题。
8. 调查分析近3年的IEEE DSAA等数据科学领域国际会议的主要主题。
9. 用数据可视化方法（或故事化描述方法）展现数据科学的研究现状与发展趋势。
10. 结合数据调研，给出数据科学在自己的专业领域中可能的应用点和应用方式、

3.编程开发题

1. 提出一个有意思的研究假设或洞见，并用数据分析方法证明是否成立，并用可视化方法进行成果展示。
   * 数据集：数据集美国交通部提供的大数据（https://www.mot.gov.sg/）的航班信息；
   * 分析工具：R/Python
   * 分析算法：不限
2. 提出另一个有意思的研究假设或洞见，并用数据分析方法证明是否成立，并用 Markdown的形式给出分析过程和结果。
   * 数据集：数据集美国交通部提供的大数据（https://www.mot.gov.sg/）的航班信息；
   * 分析工具：Spark+R/Python
   * 分析算法：不限
   * 成果形式：Markdown

|  |
| --- |
| **特别声明**  1.本文件为中国人民大学朝乐门老师的课程《数据科学导论》（对应教材为《数据科学理论与实践》（清华大学出版社，2017））的配套教学资源。  2.非版权所有人的使用范围仅限于教师的课堂教学。版权归朝乐门所有，未经许可，不得以任何形式用于商业目的。  3.不要擅自放在包括百度网盘/文库、CSDN、豆丁网等第三方平台，否则版权所有人有权追究上传者和第三方平台的法律责任。  4.本文件将以【开源模式】更新和维护，欢迎您的补充与修改。通过chaolemen@ruc.edu.cn联系总负责人朝乐门老师，总负责人将根据您的贡献动态更新作者列表及排名。  朝乐门  2017-11-18 |