1. **课程名称**

学位学生学业成绩数据集

**课程概述**

对于作业，您需要探索数据分析技术在提供的数据集上的应用。您必须研究与数据集相关的数据问题，特别考虑问题域的独特属性，并在其上测试一种或多种技术。

您的分析应该深入而详细，而且必须比本课程中已经涵盖的内容更深入。您必须采用数据探索、操作、转换和可视化概念来指导您完成解决方案过程。解释和证明所选择的技术是非常重要的。

您可能还需要预处理您的数据以使其成为适当的格式。作业应涉及多种技术，将其分类为不同的标准，并对每个标准中使用的命令进行详细探索。概述调查结果，对其进行分析并使用适当的图表正确证明其合理性。此外，还需要一份支持文档来反映使用 R 编程概念的图形和代码。

1. **本课程的目标**

该作业将帮助您探索和分析一组数据，并将其重建为有意义的决策表示。

1. **类型**

个人作业

1. **课程描述**

该数据集包含学位学生在学术课程结束时的三年期最终分数，其中包含可能会或可能不会影响这些学生成绩的几个特征。它包括他们的个人详细信息、学习成绩、家庭背景和日常生活。并且需要分析这些因素如何影响他们的学习成绩。该数据集包含 33 列，具有 16 个整数数据类型值、9 个字符串值和 8 个布尔值。

**技巧**

使用本课程中涵盖的各种数据探索、操作、转换和可视化技术来探索数据集的技术。并且作为附加特征必须探索可以提高检索效果的进一步概念。

为此作业提供的数据集与学位学生的学业成绩有关。它包含 32 列和 396 行。它包含与他们的学术、家庭背景和日常生活相关的数据。

1. **一般要求**

* R 程序应该可以无错误地编译和执行。
* 应该对用户的每个条目进行验证以避免逻辑错误。
* 数据集中不允许重复。
* 你应该;
  + 包括良好的编程习惯，例如注释、变量命名约定和缩进。
  + 在检查数据时，从互联网上进行额外的研究，以理解给定数据集的知识和信息。
* 分析应该是有意义和有效的，为决策提供信息。
* 实现的任何附加功能都必须提高检索效果。
* 在学生的情况下：
  + ***作业演示失败，作业总分若超过50%，将调整为现有总分的50%。***
  + ***发现涉及抄袭的，将按照APU和斯塔福德郡大学关于抄袭的规定进行处理。***

1. **可交付成果：**

* 完整的 **脚本** （源代码）和 **报告** 必须提交给 **APU学习管理系统（Moodle）**.
* RScript（程序代码）：
  + 以您的姓名和 TP 编号命名文件。
  + 通过键入您的姓名和 TP 编号开始程序中的前两行。例如：

＃ 名称

#TP123456

* 对于每个问题示例，给出一个 id 并解释您想知道的内容。例如：

# 问题 1：学生如何在 Level 3 中取得更好的成绩。

# 分析 1-1：找出额外教育支持与 L1 和 L2 分数之间的关系……

# 分析1-2：找出学习时间与……的关系。

# 分析1-3：找出……之间的关系

* 对于每个额外的功能示例，给出一个 id 并提供解释。

# 额外功能 1

# 关于额外功能的评论

**8.2 *文件：课程报告***

* 作为评估的一部分，您必须以印刷版和软拷贝形式提交项目报告，其格式如下：

1. 封面：

所有报告都必须有封面。可在报告前放置一块保护性透明塑料片以保护封面。封面应提供以下详细信息：

* 模块
* 课程名称
* 录取
* 学生姓名和身份证
* 分配日期（分发报告的日期）。
* 完成日期（提交报告的日期）。

1. 内容：

* 介绍和假设（如果有）
* 数据导入/清洗/预处理/转换
* 每个问题必须从一个单独的页面开始，并包含：
  + 分析技术——数据探索/操作/可视化
  + 带有说明的源代码屏幕截图。
  + 带有解释的输出/绘图的屏幕截图。
  + 根据获得的结果概述调查结果。
* 额外的功能说明必须在单独的页面中，并包含：
  + 带有说明的源代码屏幕截图。
  + 带有解释的输出/绘图的屏幕截图。
  + 解释添加此额外功能如何改善结果。

1. 结论
2. 参考

* 报告中使用的字体大小必须为 12pt，字体为 Times New Roman。报告中不允许包含完整的源代码。报告必须打字并清楚打印。
* 您可以从 Internet 或书籍中获取算法和信息。文件中应清楚地引用资源。
* 所有参考文献都必须使用哈佛命名约定，如下所示：

*该理论于 1970 年首次提出 (Larsen, AE 1971)，但此后一直被驳回。MK Larsen (1983) 是反对派中最有活力的人之一......*

*/\*\**

*\* 以下源代码来自 (Danang, SN 2002)*

*\*/*

*int noshape=2;*

*noshape=GetShape();*

* 文档或源代码末尾的参考列表必须按以下格式指定：

*Larsen, AE 1971，水生科学文献指南，McGraw-Hill，伦敦。*

*拉森，MK 1983， 英国医学杂志 [在线]，可从 http://libinfor.ume.maine.edu/acquatic.htm 获取（1995 年 11 月 19 日访问）*

*岘港，SN，2002，寻找类似图像 [在线]，代码项目，\*可从 http://www.codeproject.com/bitmap/cbir.asp 获得，[2006 年 9 月 14 日访问]*

* 有关其他引用类型的更多信息，请参见 Petrie, A., 2003, UWE Library Services Study Skills: How to reference [online], England, University of Western England，可从 http://www.uweac.uk/library 获得/resources/general/info\_study\_skills/harvard2.htm，[2003 年 9 月 4 日访问]。

1. **作业评估标准**

作业评估由三个部分组成：编码 (50%)、文档 (30%) 和演示 (20%)。各组成部分的详细划分如下：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **编码 (50%)** | | **文件 (30%)** | | |
| **标准** | **分配的标记** | **标准** | **分配的标记** | |
| 数据探索 | 10% | 报告结构和参考文献 内容：   * R 概念的描述和理由合并。 * 编写屏幕截图、图表   项目描述、限制和结论 | 30% | |
| 数据操作 | 10% |
| 数据转换 | 10% |
| 数据可视化 | 20% |
| **介绍 (20%)** | | | | |
| **标准** | | | | **分配的标记** |
| * 能够回答讲师针对已完成和呈现的工作提出的问题 | | | | 20% |

1. **开发工具**

为此作业编写的程序应在 R Studio 中编写

1. **学术诚信**

* 您应该在课程期间保持最高水平的学术诚信。
* 抄袭是一种严重的违法行为，将根据 APU 和斯塔福德郡大学关于抄袭的规定进行处理。

.