

Data Scientist Capstone

项目定义

标准	符合要求
项目概述	学员提供了该项目高度概括的介绍。提供了足够的背景信息，如问题的领域、项目来源和所用数据集或输入数据。
问题描述	清晰地定义了需要解决的问题。提出了解决问题的思路，包括讨论了预期解决方案。
评价指标	清晰定义了能衡量模型或结果的性能的评价指标。根据问题的特点调整了评价指标。

分析

标准	符合要求
探索性数据分析	报告了有关问题的特征和计算统计量，讨论了有关数据集的特征和计算统计量，并且完整描述了输入空间或输入数据。明确了所用数据或输入的异常或特点。
数据可视化	构建了数据可视化来更好地展示你在数据探索性分析时的发现。保证可视化方案适用于你要绘制的数据图表。

方法步骤

标准	符合要求
数据预处理	清晰记录了所有预处理步骤。已经修正了所用数据中的异常和特点。如果不必做数据预处理，务必清晰解释其合理性。
代码实现	完整记录了代码实现的过程，包括将给定数据集或输入数据的评价指标、算法和技术。讨论了代码实现过程遇到的复杂情况。
模型改进	清晰地记录了对所用的算法和技术改进的过程。报告了初始和最终的解决方案，如果中间结果有意义也应报告。

结果

标准	符合要求
模型的评价和验证	如果使用了模型，应该满足如下条件：详细评估了最终模型的质量——比如模型参数。用了某种分析方法来验证模型的鲁棒性。当然，如果你的选题更适合用数据可视化或者其他不涉及机器学习的方法，也可以不用模型。
讨论结果	详细讨论了最终的结果。探索为何某些技术要比其他技术表现更优或如何改进，并对上述过程进行记录。

结论

标准	符合要求

标准	符合要求
反思	学员充分总结了解决方案的完整流程，讨论了该项目中他认为有趣或有挑战的一两个方面。
改进	讨论了如何改进代码的实现，至少一个方面即可。考虑了这些改进对应的可能的解决方案，并将它们和目前的方案进行对比。

可交付成果

标准	符合要求
博客文章或应用程序	<p>如果你选择提交一篇博客文章，需要满足如下条件：项目报告内容组织合理，容易被技术人员读懂。每个章节清晰、简洁、明确。没有语法和拼写错误。列出所有引用的完成项目时用到的材料。</p> <p>如果你选择提交一个网络应用程序，需要满足如下条件：网络应用程序要用数据告诉用户该网络应用程序的表如何。该应用程序不需要支持网络访问，但是需要提供一份指南，指导如何在本地运行该应用程序。</p>
Github 代码库	<p>必须为该项目创建一个 Github 代码库。该代码库必须包括一个 README.md 文件。该文件需要告诉读者用到的依赖库，做该项目的动机，简述仓库中的每个文件总结分析的结果以及必要的致谢。如果学员提交的是网络应用程序而不是博客文章，那么还需要在README文件或 Jupyter notebook 里写明项目定义、分析和结论。学员不应该使用其他人的代码来完成该项目。但是完成该项目的过程可以参考网上的其他资料，如StackOverflow 和 Kaggle。</p>

标准	符合要求
最佳实践原则	代码排版整齐，且使用 DRY 原则。提供了 README 文件。使用了 PEP8 规范代码的风格。使用软件工程和沟通课程中介绍的良好实践原则。