

2024年第三届钉钉杯大学生大数据挑战赛复赛题目

复赛B：医疗门诊患者及用药数据案例分析

一、问题背景:

智慧医疗的出现，主要是因为传统医疗存在管理系统的不完善、医疗成本高、渠道少、覆盖面低等问题，因此需要建立一套智慧的医疗信息网络平台体系，使患者用较短的等疗时间、支付基本的医疗费用，就可以享受安全、便利、优质的诊疗服务。新质生产力在医疗领域的体现，在于创新驱动发展、数字化和智能化、产业深度融合、优化资源配置、政策支持与环境建设、高质量发展、患者体验改善以及跨界合作等方面。其不仅推动了医疗服务质量的提升，也为医疗行业的未来发展提供了新的方向和动力。习近平总书记曾指出，信息化可以使医疗卫生服务“如虎添翼”。从“互联网+医疗”、“AI+大健康”到“5G+智慧医疗”，互联网时代的数字红利让老百姓的获得感、幸福感、安全感不断得到提升。

门急诊收费系统是用于处理医院门急诊划价和收费的计算机应用程序，包括门急诊划价、收费、退费、打印报销凭证、结帐、统计等功能。医院门诊划价、收费系统是直接为门急诊病人服务的，主要目的是减少病人排队时间，提高划价、收费工作的效率和服务质量，减轻工作强度，优化执行财务监督制度的流程。住院收费系统是用于住院病人费用管理的计算机应用程序，住院收费管理系统的设计应能够及时准确地为患者和临床医护人员提供费用信息，及时准确地为患者办理出院手续，支持医院经济核算、提供信息共享和减轻工作人员的劳动强度。总体来说，智慧医疗建设是新质生产力在卫生健康领域的着力点，新质生产力是助力智慧医疗建设的重要手段。

该数据来自某医院近些年门诊医疗数据，已经做了脱敏和数据变换处理。共1个excel文件，文件内共6列。数据具体说明见表1。

表1

变量编号	变量名	释义
1	日期	数据对应时间标签
2	门诊患者人次数	门诊当日患者就诊人次数
3	门诊收入OBS_T01_MZSR68	门诊当日住院、药品、器材、检查等总收入
4	药品总收入	当日就诊患者用药总收入
5	当日病房收入	当日病房住院、药品、器械等总收入
6	当日病房收入对应科室	上述数据对应的病房科室

二. 解决问题

1. 聚类分析

(1) 通过数据分析选取合适的指标，这些指标的目的在于能够描述不同科室、不同日期下数据的画像特征，说明选择这些指标的理由和合理性。

(2) 结合选取的特征指标，使用2种聚类算法，对不同科室、不同日期的数据进行聚类分析，找出具有相似特征的数据。自行选择聚类算法并设计模型的结构和参数。自行选择或设计美观、合理的可视化图表展示聚类结果，并对聚类结果的机理、原因、特点进行详细的分析。

2. 特征分析

根据上一问的聚类结果，对不同类的数据特性、特征进行定性和定量的分析，分析内容包括但不限于时间、科室类型、门诊人数对药品收入、门诊收入的影响，请提供模型的设计思路、实现过程、完整代码和分析结果。