Advanced Topics in Health Care Data Analytics and Data Mining

Analyzing Inpatient PCCR Costs Using k-Means Cluster Analysis

*Yash Puri, Doreen Accorley, Peter Mukiibi,*

*Biniyam Abebe, Puja Yadav*

March 25, 2025

GROUP 1

**INTRODUCTION**

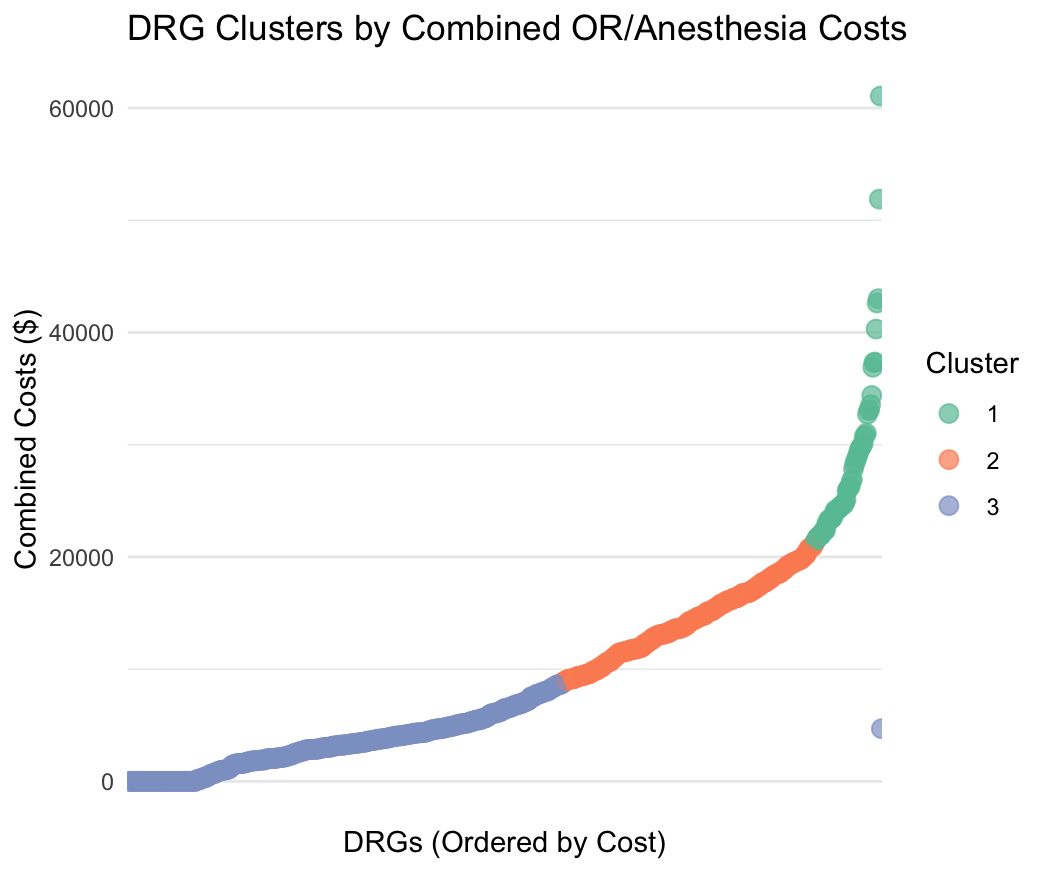
In this case study, we use diagnostic-related groups (DRGs) from 2016 Vermont Inpatient Claims data and linked Revenue Code data to examine cost trends from inpatient hospitalizations. We are able to extract information about two specific codes of interest related to operating room and anesthesiology expenses by filtering DRG data and combining DRGs with PCCR average prices. To find discrete cost groups within the total costs, we aggregate these expenses from the two departments and apply k-means clustering. Our goal in conducting this clustering analysis is to shed light on Vermont Hospitals' income trends and resource allocation.

**TOOLS USED**

R

EXCEL

**Filtering and organizing data**   
Cluster analysis will be used to examine and interpret cost categories from inpatient hospital diagnostic-related groups (DRGs) in the 2016 Vermont Inpatient Claims and Revenue Codes datasets. DRGs are assigned to each inpatient hospitalization, and the Revenue Code dataset details their expenditures.   
Before understanding cost patterns, we restructured our data. We first limited our DRG range to DRG 20 and DRG 977, which represent “Intracranial vascular procedures with PDX hemorrhage with MCC” and “HIV with or without other related condition”. We next used unique patient identifiers to integrate the inpatient and revenue records using an inner join.   
The Primary Cost Centers totaled all charges. Hospitals can organize bills and charges by department using PCCR coding. This approach avoids double charging for DRGs from various departments by grouping charges by department. All charges under $100 were deleted to avoid needlessly modest expenses that might obscure cost groups in our cluster analysis.   
After aggregating and labeling charges by PCCR, we calculated our filtered DRGs.



This chart provides a refined visualization of DRG (Diagnosis-Related Group) clusters based on combined operating room and anesthesia costs, with the DRGs ordered from the lowest to the highest cost along the x-axis. The y-axis represents the total cost in dollars, and each point is color-coded to indicate its cluster: blue for low-cost (Cluster 3), orange for medium-cost (Cluster 2), and green for high-cost (Cluster 1) procedures. The smooth upward trajectory of the data points reveals a clear cost gradient and distinct clustering. Low-cost procedures, which dominate the lower range of the cost spectrum, are consistently represented by blue dots and typically fall below $10,000. Medium-cost procedures in the orange cluster span approximately $10,000 to $25,000, while the green cluster represents high-cost procedures, generally exceeding $25,000, with a few outliers approaching or surpassing $60,000.

This clustering helps identify which DRGs contribute most significantly to hospital expenditures. The blue cluster, being low-cost and uniform, suggests predictable and efficient procedures ideal for outpatient care or high-volume scheduling. The orange cluster presents opportunities for process improvement and resource optimization, while the green cluster requires careful management due to its high resource demand and potential for cost variability. These insights can guide strategic decision-making in operational planning, budgeting, clinical staffing, and reimbursement strategy. By visually segmenting DRGs by cost, this chart offers a powerful tool for hospitals and health administrators to prioritize cost-containment efforts and align resources with clinical and financial goals.