Modeling Private Insurance Payment Amounts Using Medicare

Columbia University Data Science Institute

Authors: Ryan Rogers, Sarthak Arora, Parv Joshi, Shruti Kaushal, and Tyler Marshall

Industry Mentors: Katherine Etter, Cindy Tong, and Ziyu Tan

Faculty Mentors: Adam Kelleher and Cathy Li

Data Science Capstone Project with Johnson & Johnson

Project Context

For the Johnson & Johnson (J&J) Medical Devices Health Economics team, it is important to understand the value they provide to potential consumers. One key factor when determining hospital profitability is private insurance payments, so it is advantageous for the team to have predictions on these values. For our project, we used IBM MarketScan® Commercial Claims & Encounters (CCAE) database and Medicare payment information to model private insurance payment amounts for specific procedures across Metropolitan Statistical Areas (MSAs) in the continental United States.

Feature Selection and Design

The CCAE database contained information about procedures, geography (MSA & State level) and insurance reimbursement amounts (commercial and medicare) from years 2018 to 2020. Since these features didn't capture market conditions that have a direct impact on healthcare costs, we added more features relating to Economic, Demographic, and Population Profile:

Economic	Income per capita; Employment Rate
Demographic	Total Population; Median Age; Sex Ratio; Number of Races; Median Household Income, size; Fertility Rate
Population Profile	Percentage of people in poverty, married, educated, veterans, disabled, and immigrants; Percentage of people with medicare, private insurance, and no insurance

Figure 1: Names and types of the features added

Threshold, Missing Value, and Cluster Analysis

The J&J team originally used observations with 50 or more private claims for modeling. We revised this threshold to 35 after performing statistical analysis under business assumptions that resulted in additional training data. We also observed that certain features interacted differently with each procedure. For example, breast reconstruction was correlated to sex ratio while ankle fix wasn't. To incorporate surgery specific information we segmented the data by clustering procedures based on costs using optimized KMeans. We subsequently constructed models for each segment.

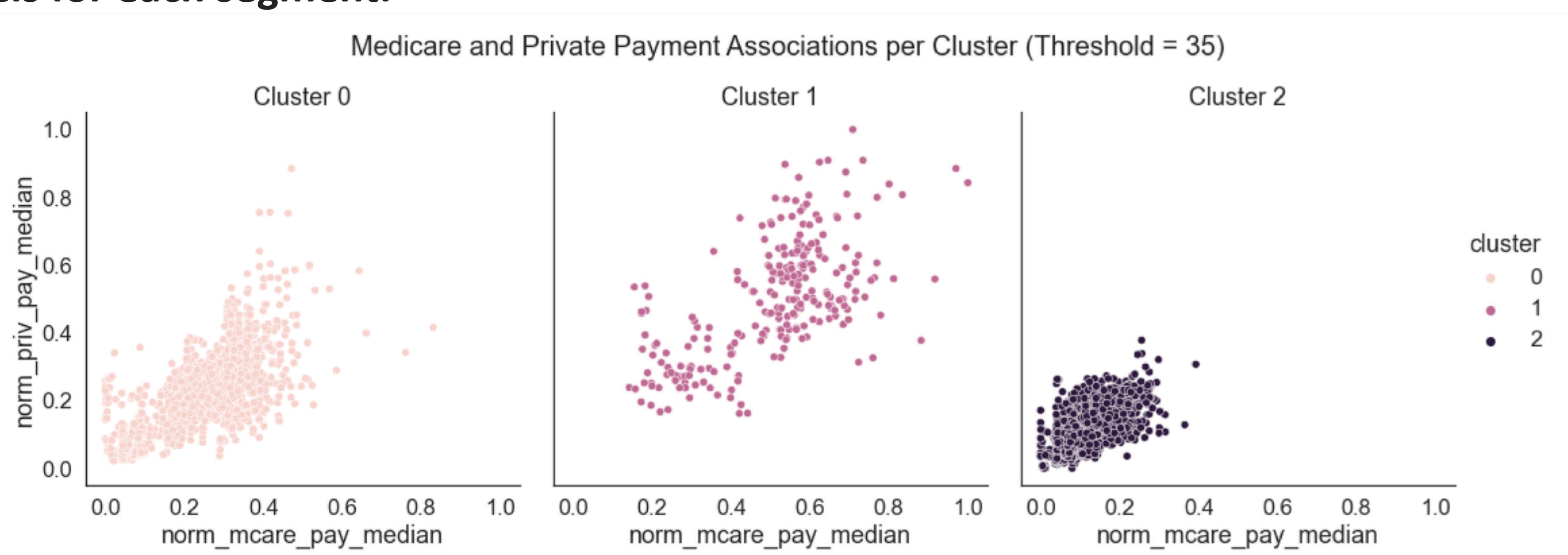

Figure 2: Associations of normalized private and medicare payments by cluster

Cluster	Procedures
0	breast reconstruction, mastectomy, navigation, rtc slap bank, partial shoulder arthroplasty, clavicle fixation, radius/ulna internal fixation, bunionectomy, pnn, fess, septoplasty, bsp, bariatric, lap appendectomy, liver ablation, kidney ablation, hysterect, orthovisc monovisc, robotic assisted surgery, pka, prostatectomy
1	ant tls fusion, post cerv fusion, post tls fusion, hepat, intracranial thromb
2	ant cerv fusion, tsa, proximal humerus, tha, revision tha, hip fracture fixation, tka, revision tka, femoral shaft fixation, prox tibia fixation, ankle fix, thoracic, lung ablation, laac, colorect, hernia, cardiac ablation, cardiac ablation additional discrete, cardiac ablation linear focal, cardiac ablaton ice, cardiac ablaton anesthesia, tpa

Figure 3: Procedure group clusters formed using KMeans clustering

Monotonicity and Business Rules

Interpretability was a major issue with the original J&J model. In particular, it could predict payments related to Ambulatory Surgical Centers (ASCs) as more expensive than Inpatient care. This was considered illogical given the nature of care. To remedy this, we incorporated monotonicity constraints into our model building to prevent this occurrence.

Results

Clustering helped us use different models for different procedure groups, resulting in a lower test MAPE for each procedure. This caused a reduction in the average test MAPE value across all procedures from 16.11% to 14.72%.

Model	Test MAPE
Original DataRobot Model	16.11%
RF Model Without Clustering	16.16%
RF Model With Clustering	14.72%

Table 1: Resulting test MAPEs for new and original model. Cluster scores were calculated using only major procedure groups for original model.

Appendix for Details

Conclusions and Recommendations

Through experimentation, we determined that breaking procedure groups into clusters, then using this clustering pattern to create several models yielded the best results. We also enforced business rules ignored by the original model by adding monotonicity constraints to our models.

We recommend incorporating surgical complexity to better capture determinants of surgery costs while clustering procedures. Models can be made more robust by switching to a different evaluation metric since MAPE inherently introduces a bias against over-forecasts.

Acknowledgments

We would like to thank the Johnson & Johnson team for their mentorship throughout this process, especially Cindy Tong, Ziyu Tan, and Kade Etter. We also want to acknowledge Professor Adam Kelleher and Cathy Li for their guidance during this project.

References

1. FRED | St. Louis Fed. (n.d.). https://fred.stlouisfed.org/searchresults/?st=metropolitan+statistical+area
2. US Census Bureau. (2022, November 16). Data. Census Website. https://www.census.gov/data.html