Healthcare Fraud Detection using Clustering

Takeaway Summary

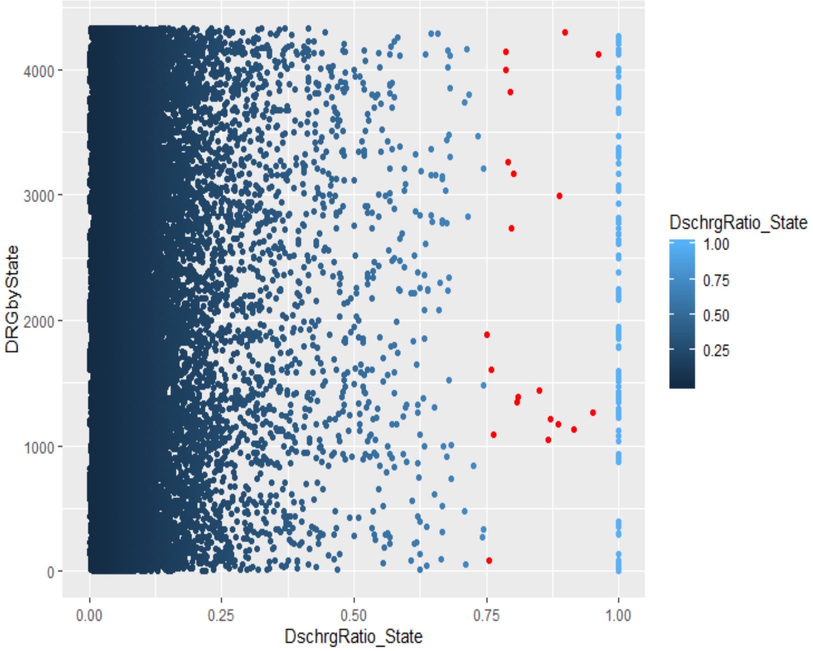
Ram Subramanian

MS Applied Analytics

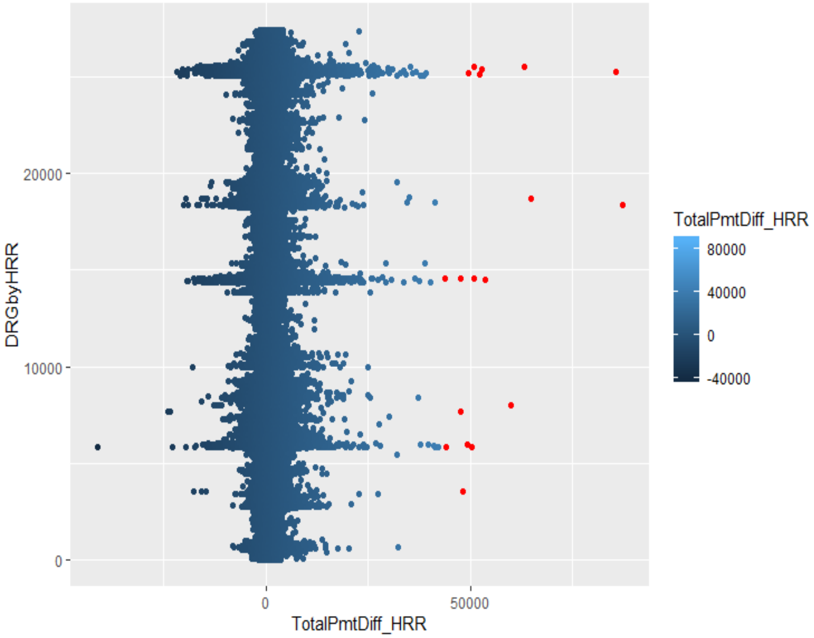
Columbia University, New York, NY

Healthcare Fraud Detection – Key Takeaways

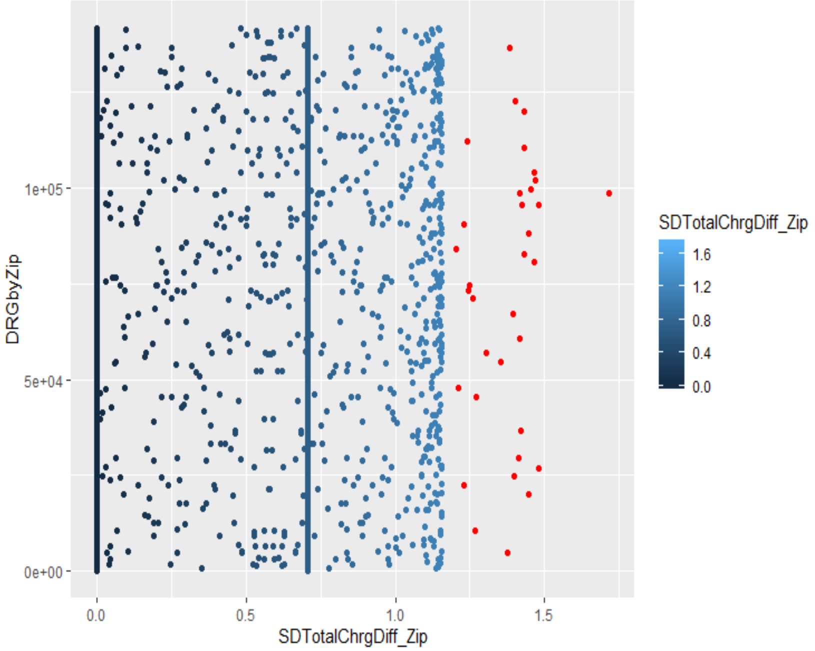
- ***Key Feature 1:*** Ratio of discharges between DRG-Provider combination and Total discharges for that DRG in the State



- ***Key Feature 2:*** Difference between Total payment to the provider for a DRG and the mean of the Total payment to the Provider for that DRG at a Hospital Referral Region (HRR) level



- ***Key Feature 3:*** Standard deviation between Average Covered Charges for a DRG and Mean Medicare Payments for every unique DRG-Zip combination



Healthcare Fraud Detection – Key Takeaways

	Features	Insights
1.	Ratio of discharges between DRG-Provider combination and total discharges at a State, Zip code and HRR level	How patients from the same location are potentially targeted by provider groups as per Level 6 of Healthcare Fraud Control (Sparrow, 2000)
2.	Differences and deviations of charges and payments related to every DRG at a State, Zip and HRR level relative to the mean values	a. Upcoding – Billing for service with higher reimbursement rate b. Excessive or Unnecessary Services

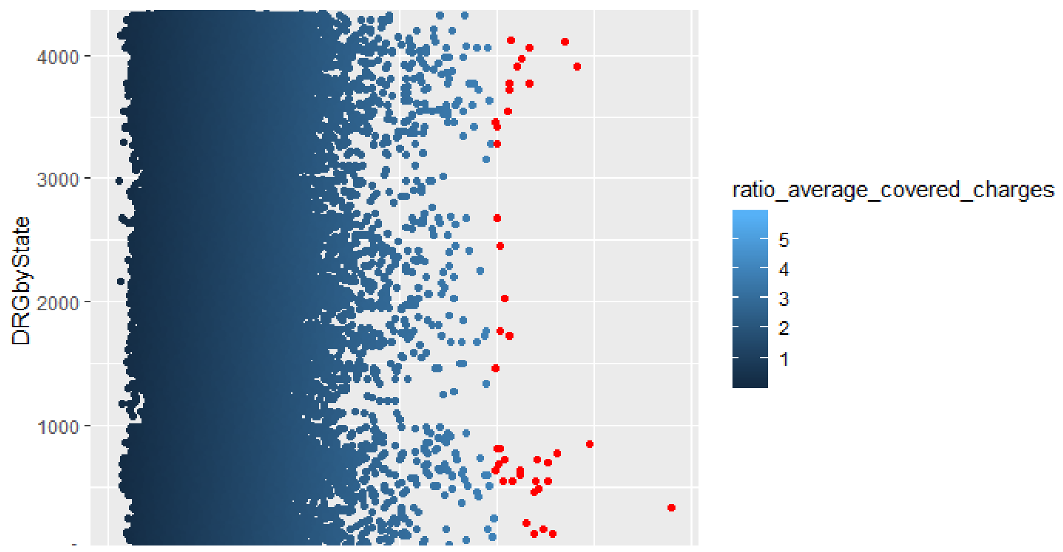
Note: Cut-off for the outliers will need to be set for every feature with its own criteria based on expert interviews

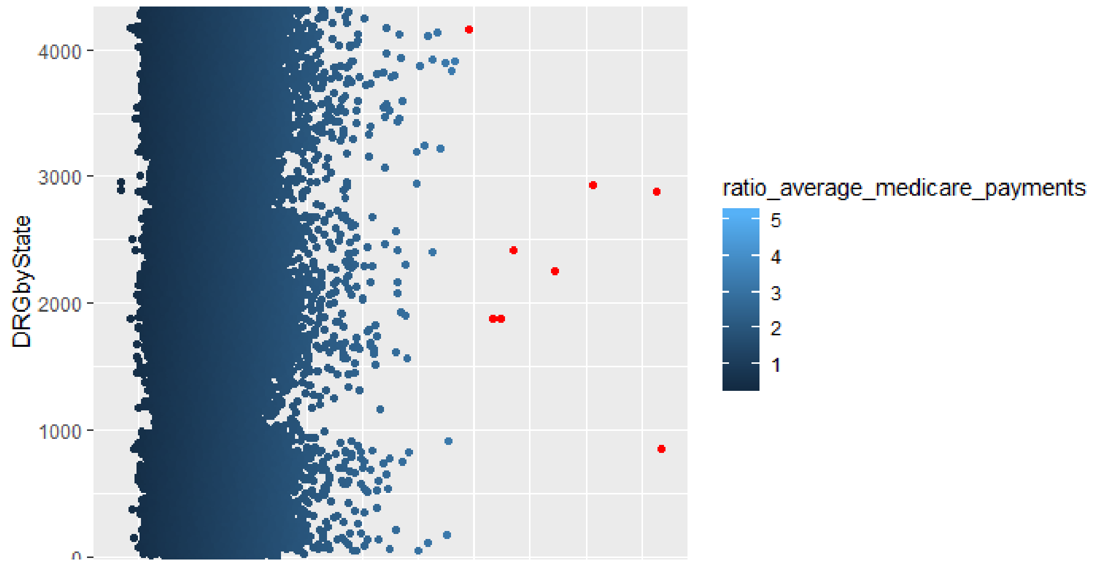
Additional Feature engineering & Principal Component Analysis

- New features added to perform peer comparisons on payments and charges among providers of the same state

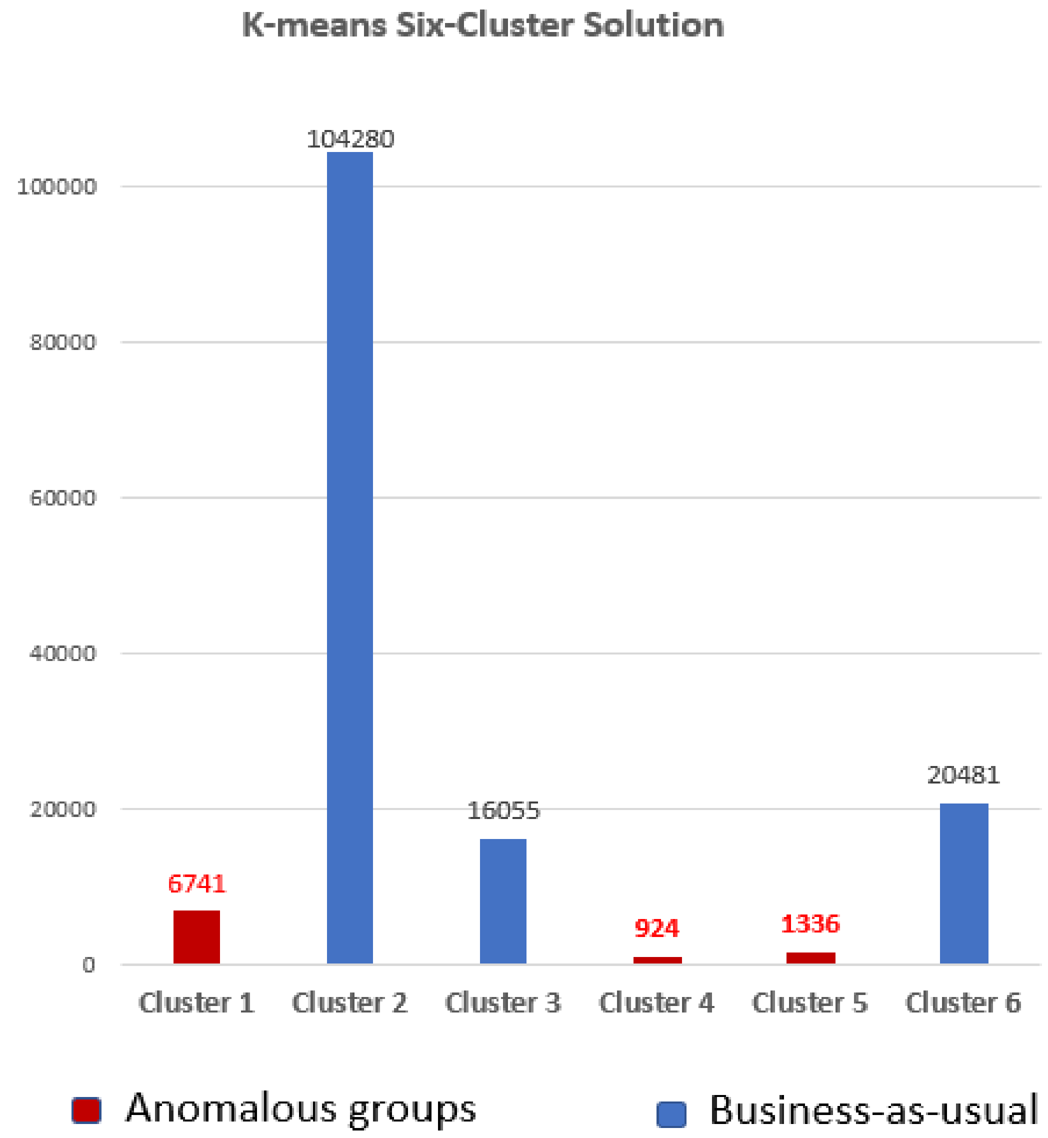
- Principal Component Analysis performed for dimension reduction to avoid higher weightage to a specific dimension measured by multiple variables

- Based on the variances explained by the components, we keep the top 8 Principal Components that capture up to 95% of the information





Anomalous Groups from Cluster Solutions



- A **six-cluster solution** arrived at using
  - Initial observations of cluster distribution
  - Data driven methods like Within sum of squares plot and Ratio plot

- Conclusions using the average statistics of the variables,
  - **Clusters 4 and 5** (**1.5%** of total cases) deserve a careful and thorough inspection
  - **Cluster 1** (**4.5%** of total cases) needs a preliminary level inspection
  - Remaining **bigger clusters (2,3,6)** should fall under the **business-as-usual** category

- The size of the six-cluster solution indicates that clusters 4,5 and 1 correspond to **6% of the total observations** that need inspection for healthcare waste and abuse

Additional Key Points:

- Difference between abusive behavior and fraudulent behavior in both nature and proportions
- Variable selection to involve logical inference validated by expert interviews for more efficient and effective fraud detection
- Suspect of abusive behavior could also be one with a lower quality prescribing behavior – need for collating multiple variables
- Choice of the unit of analysis (e.g. physician, hospital, DRG) is important in healthcare fraud detection as features chosen accordingly