Florida Medicare Cluster Analysis
Yahia El Bsat
IEMS 308

Executive Summary

Medicare expenditure on surgical procedures in the state of Florida amounted to more than $4 Bn in 2016 alone. A K-Mean clustering analysis of surgery providers and their submitted charges to Medicare shows that the cardiovascular surgery types are being underfunded or are showing a much higher submission amount than the standard amount meaning that Medicare needs to investigate further into those type of surgeries and either revise its allowed amount or impose more stringent allowed amount and further auditing on providers who tend to submit amounts higher than the allowed amount in an effort to combat fraud.

This report summaries the problem statement then outlines the methodology used in analyzing the 2012-2016 Medicare Service Provider dataset for the state of Florida using K-means clustering technique. It summarizes the results and provides insights into how Medicare can improve its financial operation and outlines next steps for the analysis.

Problem Statement

In 2016 the total Medicare spend in the state of Florida alone amounted to over $21 Bn 20% of which was spent on surgery alone. The state of Florida is a big retirement state with a 17% of its population aged 65 years or older. Surgical practices vary among practices and the charges allowed by Medicare as a standard for each surgical type do not match with the service provider claim amount. There could be some fraud involved and the following analysis tries to identify what are the different clusters or groups of healthcare providers in the state of Florida based on the practice they provide, the type of service they did and the claim amount they requested during that time period. Identifying such groups will help Medicare combat fraud and identify target groups of providers where more auditing needs to be performed.

Assumptions

- The data provided by Medicare on its public records is accurate and aggregates the average spend amount by service provider and CPT code correctly
- Service codes (CPT and HCPCS) are correctly reported by service providers. That is service providers do not charge for a service different from what they performed. However such variability should be detected in the clusters especially when taking into account the type of surgery performed and the provider type.
- The MEDICARE_PARTICIPATION_INDICATOR variable accurately describes those practitioners who filed for Medicare and when the variable is false, no claim has been made
- String data like the provider credentials are selected from a list of option meaning no typing differences in the same credential exist. If that is the case, then the algorithm will consider the same credential typed differently as different credential type impacting the final classification

Methodology

To identify the groups of service providers, K-means clustering was used.
First, the most up to date data was downloaded from the Medicare site as tab delimited text format reflecting claims between 2012-2016 for all service providers among all states. The data header was edited manually to take out the copyright signature on the second row. Then an analysis of the different

features provided was performed to understand the different variables, their composition, and their distribution. This was done in coordination with a documentation file downloaded from the Medicare site to explain what each variable means. There was a total of 26 features included in the dataset and around 10 million rows.

Given the specificity of our problem, that is Medicare spend on surgical practices in the state of Florida, the data was then filtered to only include records where Medicare Participation is True, and where the State is Florida. Further, one of the features included HCPCS codes for categories I to III. Category I is the American Medical Association classification for physician services, other categories reflect non physician services and products which are not of out interest. Thus the data was filtered only to include the data from category 1 only. Further, by evaluating and understanding the most up to date CPT codes from the AMA website, the data was filtered to include only codes relevant to surgical practices.

After the data was filtered, the features to include in the K-mean clustering was performed. Those were as follow:

- Provider credentials
- Provider gender
- Provider type
- Place of service
- CPT surgical practice code (aggregated to 2 digit level to reduce number of features)
- Number of services provided
- Number of distinct beneficiary
- Number of distinct beneficiary per day of service
- Average Medicare allowed amount

The choice of features to include was made relevant to the problem at hand. Given that we are trying to identify different groups of providers with varying characteristics and surgical procedure types those feature seemed the most relevant to include in the analysis. Geographical data for example was deemed less relevant and thus was not included and the other variables were used for filtering so they all have the same value.

It is important to note that the CPT variable was created by taking the first 2 digits of the HCPCS level 1 code as an analysis of the CPT codes shows that it is an appropriate level of analysis without going into much details. Categories included at this level are for example 'Surgical Procedure on the Digestive System" versus more specific like "Stomach", "Pancreas", or "Liver" for example. A full list can be found here: https://coder.aapc.com/cpt-codes-range/79 .
Note that average submitted charge amount and average Medicare payment amount were excluded from the clustering as those will be used in the analysis part and aggregated by cluster to answer the problem of potential fraud.

Following the selection of features, one hot encoding for the categorical features was performed. This resulted in a total of 594 numerical features over 103k unique records which went directly into the clustering algorithm after being normalized using Sklearn standard scalar function. K-mean clustering was then performed on the above normalized data for a range of clusters between 2 and 12 to identify the optimal number of clusters for the analysis. The following graph shows the k-mean score for different number of clusters.

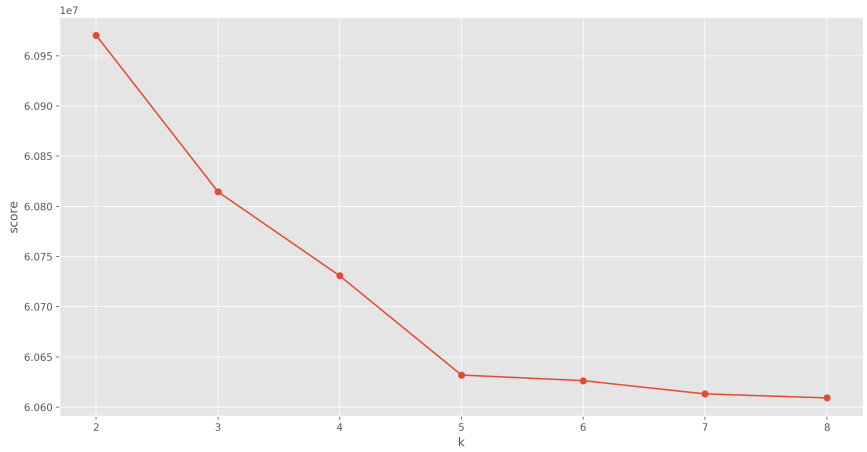

Figure 1: K Mean results for different number of clusters

Given figure 1, the optimal number of clusters was identified as 5 and thus 5 was the chosen number of clusters for running the algorithm.
Then the k-mean algorithm was run for 10 times with k = 5 and a random centroid generation for a maximum number of iterations of 3000 to get a most optimal result.
Using PCA transformation, the result of the clustering was as follow:

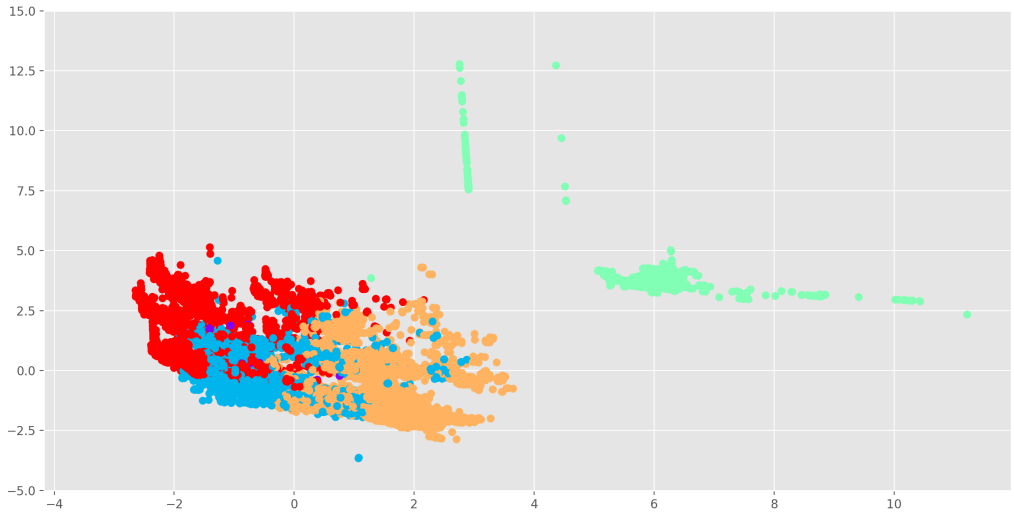

Figure 2: Clusters of Florida Medicare Service providers with K-mean using k = 5

As figure 2 shows, there is a clear distinction between the green, red, blue and orange clusters. The purple ones are less dense and mixed among the others. That can be due to the PCA technique used to transform 594 features to 2 to be able to visualize on a 2D graph.

In preparation for the analysis, different features were created by aggregating for each cluster, those are:

- Average Payment: weighted average by line service count of the Average Medicare payment by provider and CPT code

- Average Submission Amount: weighted average by line of service count of the average submitted amount by provider and CPT code
- Difference Submit Pay: difference between the average submitted amount and the weighted average payment amount
- Difference Allowed Submit: difference between the average allowed amount and the weighted average submitted amount

Analysis

Resulting from the previous preparation of the results, we get the following characteristics for the 5 clusters:

Cluster	Providers	% of Payment	Avg Payment	Avg Submission	Avg Allowance	Submission – Payment	Allowed – Submission	Highest CPT
1	268	0.1	$87	$230	$109	$143	$22	67 (Eye)
2	35428	21.5	$58	$241	$155	$183	-$86	51 (Urinary)
3	5398	20.2	$93	$616	$831	$523	$215	45 (Digestive)
4	29711	30.8	$321	$1504	$491	$1183	-$1013	36 (Cardiovascular)
5	32405	27.4	$53	$133	$142	$80	$9	13 (Integumentary)

Table 1: Summary Results of the Clusters

Table 1 above shows the results of the different clusters. Each cluster had a characteristic type or range type of surgeries performed (column of highest CPT) which describes what kind of surgeries are performed. Striking clusters are clusters 2, 4 which together account for 52.3% of Medicare total payment and where the average submitted amount to Medicare is higher than the allowed charge amount. It is more acute for cardiovascular surgeries (cluster 4) where that difference is $1013 which is much higher than in other clusters. This group of physicians thus will require more special handling and more auditing to make sure that charges are submitted are accurate or at least the allowed amount must be revised to take into account higher surgery costs for example. We can see however that the actual payment amount by Medicare is much lower than the requested/submitted amount.

Conclusion

The key insight in this analysis is that cardiovascular surgeries and similar surgeries in the state of Florida are leading to much higher charges and charge submission to Medicare then the allowed amount by Medicare and payment of Medicare only covers about 21% of the cost. Thus Medicare will have to spend time understanding what the charges for those type of surgeries are and maybe revaluate its allowed charges to cover such surgeries especially that cardiovascular surgeries are life critical and can save lives.

Next Steps

Given the conclusion that cardiovascular surgeries are being underfunded, a follow up step would be to filter the data of providers only among cardiovascular surgeries and do some analysis, potentially further clustering within the group, to understand if the charges are uniform among providers for the same service or we see differences between providers. If it is the former then funding needs to be revaluated for this particular surgery type and if it is the latter than this is a potential source of fraud and Medicare needs to audit such submission amount requests further to understand why certain providers are trying to request higher payment amounts.