Cluster for "**Prescribing of Opioids among Medical Professionals**"

By: Yonaton Heit

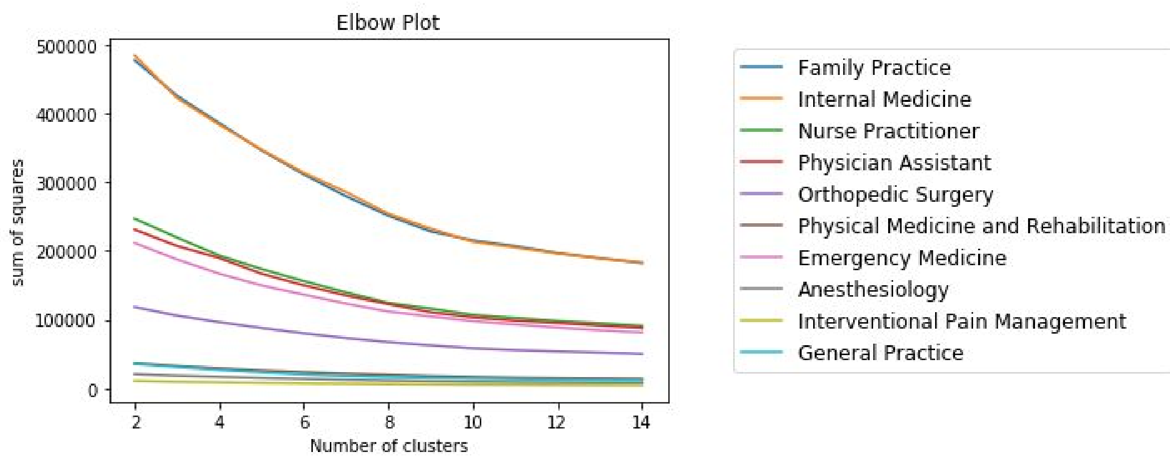
This is a continuation of the anomaly detection for opioid claims amount medical professionals. Previously, each health care provider was separated into separate specialty. Then using a boxplot, we were able to determine what normal ranges of the percent of drug claims that are opioids for healthcare providers which in a field. This method does not consider anything other factors other than specialty. For examine different geographic areas by different levels of opioid prescription. Here, clusters are created for each specialty based on multiple attributes and anomalies are determined based on the outliers of the cluster rather than the field alone. Attributes considered are location(longitude and latitude), city median income, temperature (min,max and mean), population, number of opioid claims, and percent of opioid claims. The number of total claims is left out because they can be completely dependent on number of opioid claims and percent of opioid claims. Having attribute that is completely co-dependant from other attributes is a poor practice for clustering. As of right now, we are only going to examine the clusters of the 10 specialties that prescribe the most opioids.

1. Creating Cluster for each Specialty

Each attribute has different ranges and scales of values. For example, mean temperatures ranges from 1-24 degrees C and median income ranges from \$0 to \$300000. In order to prevent attributes with larger variance from unduly weighting destination of clusters, each attribute is standardized (mean=0, standard deviation = 1). In addition, the attributes were decomposed with PCA in order to lower the number of attributes. It was found that of the 9 dimensions of the features, over 90% of the variance could be explained by the 6 dimensions. The decomposition revealed that the 4 of the mean temperature, min temperature, and latitude were highly correlated which could explain why 3 dimensions could be dropped.

Both elbow plots and silhouette scores were used to determine the optimal number of clusters to use for the K-means method. The elbow plot on Figure 1 was used to ballpark the number of clusters. Based on it, it was found that number of clusters for each specialty should be larger than 5. From this starting point (k>5), the silhouette score refined the value with the assumption that a larger score was a better fit.

Figure 1) The elbow plot of the top 10 specialties that prescribe opioids. This embow plot indicates that larger number of clusters would be optimal. We restricted the number of clusters to be greater than 5 and refined it using the silhouette score.



2. Examination of the Clusters.

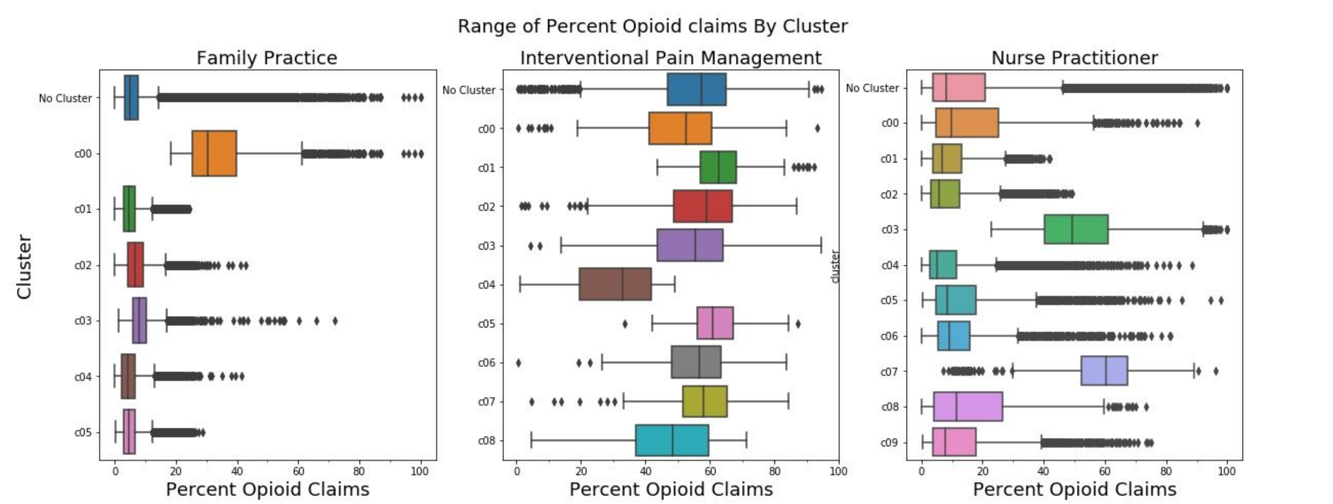
Using boxplots we can visualize the normal range of percent opioid claims of the clusters within a specialty. There were 3 distinct patterns of clustering for the specialties which is shown on Figure 2. Figure 2 also shows boxplot of the normal range of opioid prescription without clustering.

The first major pattern is that within a specialty, the clusters have similar medians and distribution of percent of opioids (though not identical). This is a similar range to that of the total data within the specialty without clustering. The exception is a single cluster that has a completely different median and range. The specialties that follow this pattern are family practice, internal medicine, emergency medicine and general practice. It is of interest that all these specialties prescribed a broad set of drugs, not just opioids. A significant number of healthcare providers that were considered outliers without clustering are no longer outliers when placed into a cluster with a different range from the unclustered data. This has the effect of lowering the number of outliers with the except of emergency medicine. The single cluster with different range for emergency medicine range of values was lower than the other clusters. Since only higher opioid claim outliers are considered, clustering did not lower the number of considered outliers for emergency medicine. This clustering pattern would suggest that a lower number of clusters would suffice, this pattern was only seen for large cluster numbers (k>5).

The second cluster patterns for a specialty seen is clusters having broad ranges in their boxplot. The range of the distribute of these unclustered specialties can covers close to 0\% to 100\% percent opioid claims which makes use of percent opioid claims useless for outliers detection. The boxplot of the clusters cover a smaller range but they is still wide. Outlier detection for opioid fraud is difficult with a wide range since it is more likely the case data points within the expected range. Specialties that have this pattern are orthopedic surgery, physical medicine and rehabilitation, anesthesiology, and interventional pain management. All specialties that treat pain and would be expected heavily prescribe opioids.

The last cluster patterns seen is there two distinct sets of similar clusters. One with a similar percent opioid boxplot to the uncluster and another set of clusters with a different range. Both nurse practitioner and physician assistant fall into this pattern. Nurse practitioners have 10 clusters, 8 with similar boxplots to the unclustered and 2 similar to each other. Physician assistant have 11 clusters, 8 with similar boxplots to the unclustered and 3 similar to each other.

Figure 2) Each of the boxplots are represent the three clustering patterning of each specialty described in this section 2. The top boxplot for each set of clusters shows the unclustered data from that specialty. The boxplot below shows the data distribution for each cluster.

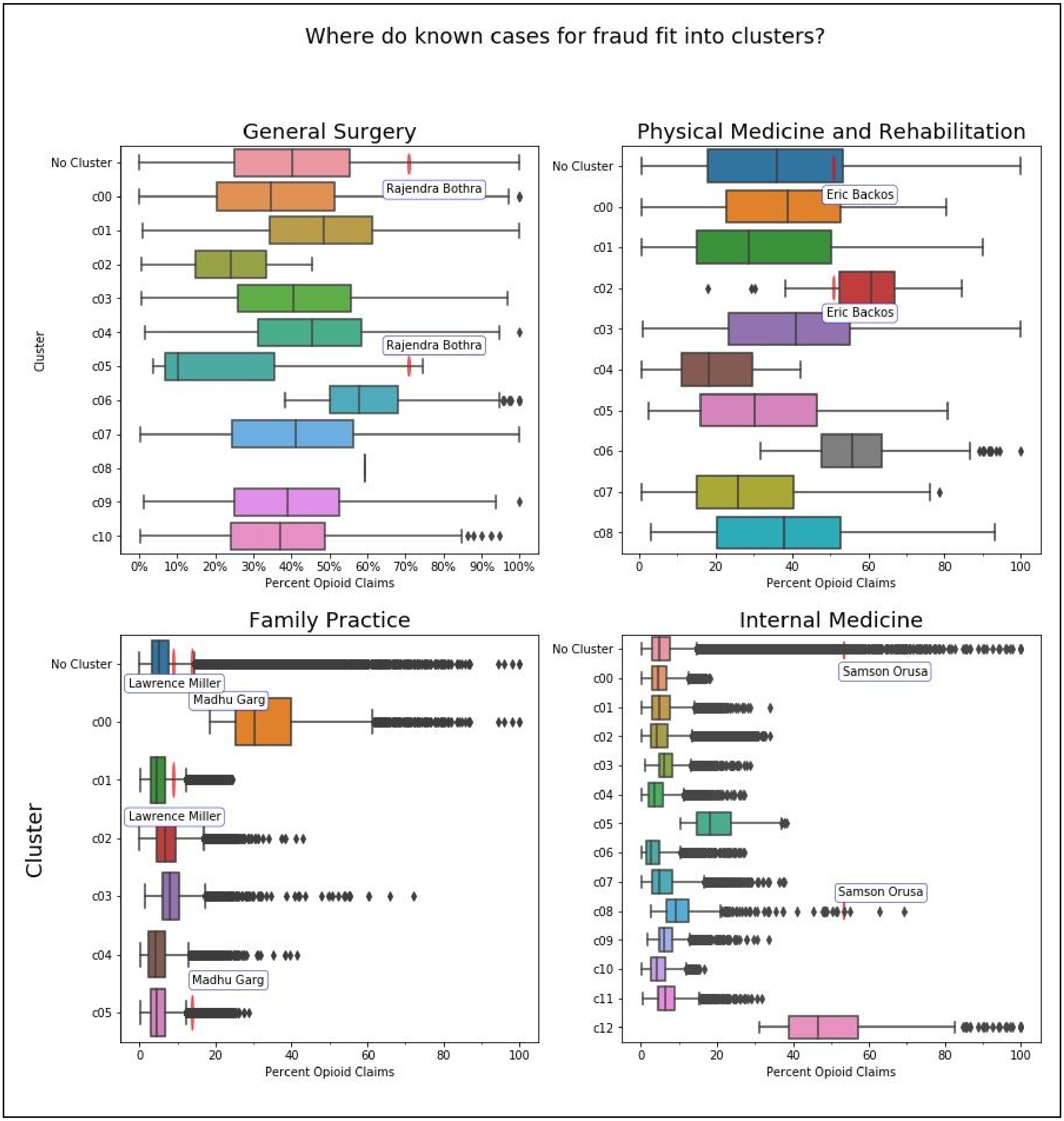


3. Check Clusters against Known Cases of Fraud

In 2018, the United States Department of Justice had a massive crackdown on the healthcare fraud [1]. 601 individuals were prosecuted, including 165 doctors, nurses and other licensed medical professionals. Of these charged, 162 were prosecuted for prescribing and distributing opioid and other narcotics. There is complete list of health care provider that have been changed. From various online articles on specific cases from this round of arrests, we were able to find 5 healthcare providers who were both persecuted for opioid fraud and were on the 2013 Medicare Part D database.

Using the cluster boxplot on Figure 3, we examined whether these medical professionals were marked as outliers.These five healthcare providers are Bothra Rajendra (general surgery) [2], Eric Backo (physical medicine and rehabilitation) [2], Madhu Garg(family practice) [3], Lawrence Miller (family practice) [4], and Samson Orusa (internal medicine) [5]. General surgery, specialty practiced by Bothra Rajendra, was not one of the fields previously examined so the clustering process that has been described was performed. The field of general surgery's boxplots has a broad range of values in the distribution which makes outlier detection more difficult. Without clustering Bothra Rajendra percent opioid claims fall clearly inside the expected range and is not marked as an outlier (without clustering the boxplot distribution covers 0% to 100%). With clustering, Bothra Rajendra in a borderline datapoint (right below the top whisker.) Both Madhu Garg and Samson Orusa were clear outliers within their cluster. Without clustering, Madhu Garg is below the edge of the distribution. Lawence Miller is just below the edge of the distribution within his cluster. Finally Eric Backos failed to be marked as an outlier; the median percent opioid claims of this cluster is higher the value found for Eric Backos. The final results are of the known cases of fraud, two where detected as an outlier, two were borderline, and failed to be detected.

Figure 3) The five known cases of fraud where placed into the boxplot of their specialty. Similar to Figure 2, the top boxplot for each specialty is the unclustered and the rest are clustered.



4. Conclusion

The purpose of this project is to demonstrates the ability of clustering of healthcare providers in order to detect anomaly opioid prescription in hopes to determine possible cases of fraud. It is important to note that analysis is limited by the lack of known causes of fraud. It would be more desirable to have a large cases of fraud. Since two of the five cases of known of fraud were borderline outliers, it may be warranted to examine the boundaries to determine better conditions for outlier. Considering the time scale of this project, its will have to be done in the future. In two of the cases, clustering did help identify outliers (or borderline outliers) where the unclustered model did not.

Based on results, these clusters can help identify outliers opioid prescribed by healthcare providers that may be fraudulent. It is recommended that outliers within their cluster (and borderline outliers) should be investigated for possible cases for fraud. This methodology was found to be more useful for specialties that have a narrow range of opioid claims such as family practice and internal medicine. For future consideration of anomaly detection, it may be beneficial to focus on the opioids that are commonly abused rather than all opioids. An addition factor that should be consider is the dollar amount of claims which may be a strong indicator of fraud. The dollar amount of claims was considered by Source 6 and successfully identified fraudulent claims within the Interventional Pain Management which our model should struggle with due to the broad range of percent of opioid claims.

Sources: (Not in the format of the final document. The correct format will be use in the final document)

[1]:https://www.justice.gov/opa/pr/national-health-care-fraud-takedown-results-charges-against-601-individuals-responsible-over

[2]:https://www.detroitnews.com/story/news/local/macomb-county/2018/12/06/feds-allege-six-detroit-area-doctors-fueled-opioid-crisis-health-care-fraud-conspiracy/2225239002/

[3]https://patch.com/california/glendora/ex-glendora-doctor-sentenced-prison-selling-pain-pills-addicts-la

[4]:http://www.fox29.com/news/local-news/prosecutors-4-area-doctors-arrested-in-pill-mill-investigation

[5]:https://www.theleafchronicle.com/story/news/crime/2018/12/14/who-samson-orusa-clarksville-doctor-charged-opioid-fraud-scheme/2312021002/

[6]:https://www.splunk.com/blog/2017/09/28/building-a-60-billion-data-model-to-stop-us-healthcare-fraud-with-splunk-and-machine-learning.html