Melissa Melnick
Clustering Analysis Project
IEMS 308
1.29.21

Executive Summary

Medicare is a system that provides health insurance to most U.S. citizens over the age of 65. In a super generalized description, Medicare works like this: the provider of the healthcare service submits the charge of the service to the insurance company. The insurance company pays for a certain percentage of the charge based on various policies, and the patient is responsible for the rest. However, there are different factors that can contribute what percent of the charge Medicare actually covers. This analysis aims to detect any factors that might lead to Medicare paying a larger percentage of one person's bill than another.

Luckily, the Center for Medicare and Medicaid Services (CMS) releases large amounts of its data. In 2018 it released its annual "Provider Utilization and Payment Data Physician and Other Supplier Public Use" file, which contains an enormous amount of data regarding healthcare services and, most importantly, how much Medicare covered for each of these different services.

The data was pared down into both a more streamlined demographic as well as relevant variables. This analysis only looks at data from the state of Missouri. Missouri provides interesting data due to having two large, but separate metropolitan areas (St. Louis and Kansas City) as well as a plethora of rural areas.

The biggest finding of this analysis is the fact that two regions of Missouri, the Southeastern region and the St.Louis Metropolitan area had nearly identical values for the average charges submitted by the medical provider, but Medicare covered on average 5% more of the charges in the St. Louis Metro region compared to the Southeastern region. This analysis may reveal an enormous moral breach in the United States Medicare system.

Methodology

The first step in any clustering analysis is to pare down the data set to only the necessary values. The data was pared down into both a more streamlined demographic as well as relevant variables. This analysis only looks at data

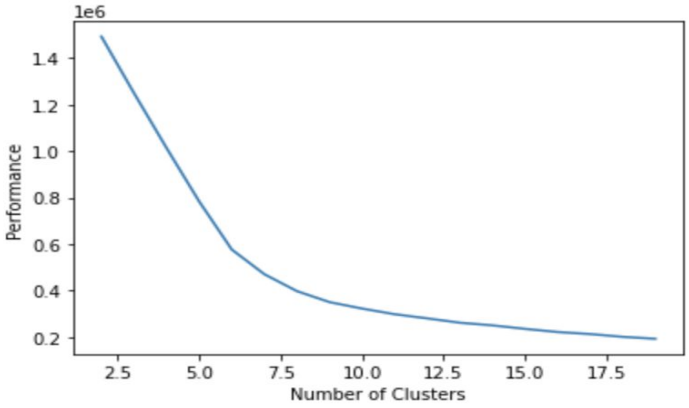
from the state of Missouri. Missouri provides interesting data due to having two large, but separate metropolitan areas (St. Louis and Kansas City) as well as a plethora of rural areas.

Via some basic exploratory analysis, the variables for this project were chosen. Included are:

- bene_unique_cnt (the amount of unique individuals that received a specific services
- average_submitted_charge_amt (the total amount of money submitted to Medicare by the provider)
- average_Medicare_payment_amt (the amount Medicare actually paid for),
- Zone (region of Missouri each of the services took place. These zones were calculated by the first 3 digits of the zipcode from the service)
    - The zones were then divided into 6 different zones: STL Metro, Northeast, Northwest, Southeast, Southwest, and KC Metro, and one-hot encoded.

Following the inclusion/exclusion of various variables, all outliers were removed from the sample. In this case, an outlier for a given variable was defined as being more than 2 standard deviations beyond the mean.

Next, the data was standardized for a more accurate fitting, and the optimal number of clusters was calculated using [AHH IDK WHAT THIS IS CALLED INSERT LATER]. To determine this, we look at where the 'elbow' or kink in the graph is. Based on this graph, it was determined that 6 clusters is the optimal amount. Finally, the data was ready for clustering.



Key Insights

The goal of this analysis is to determine if any specific variables have an impact on what percentage of a bill Medicare covers. To calculate that figure, the following equation is used:

Percent Paid(PP) = average_Medicare_payment amount / average_submitted_chrg_amt

Via examination of the clustering, it is fairly obvious that the biggest factor in determining which cluster a data point falls into is its location within the state. The table below shows approximately which region is associated with each cluster.

Cluster	Associated Region
0	NORTHWEST
1	SOUTHWEST
2	STL METRO
3	NORTHEAST
4	KC METRO
5	SOUTHEAST

After calculating the PP for all 6 clusters, it is clear to see that all of the clusters have a fairly similar percentage of the bills covered by Medicare with the exception of cluster five. Medicare covers on average **4% less** of the charges from those in cluster 0 relative to the next lowest, cluster three.

Cluster	Percent Paid by Cluster
0	26.45%
1	24.31%
2	**23.69%**
3	22.93%
4	26.41%
5	**19.91%**

It is extremely interesting to note that cluster two and cluster five had almost identical average submitted charges amounts, but Medicare covered on average 5% more of the bill, as shown by the table below.

Cluster	average_submitted_chrg_amt	average_Medicare_payment_amt
0	209.294697	55.352187
1	199.032762	48.383723
2	237.157214	56.195835
3	216.395654	49.634866
4	210.956379	55.724893
5	237.524044	47.287634

Referring back to the table with the locations of each cluster specified, we can see that cluster five refers to the Southeast region of Missouri, while cluster two refers to the St. Louis Metropolitan Area.

Conclusions

Medicare is funded entirely by taxpaying Americans. While this program is government sponsored, and intended to be equitable, the results of this analysis imply that the system gives a great advantage to those living in larger metropolitan areas.

St. Louis is a large city, with the entire metropolitan area having a population of more than 2 million people. On the other hand, the Southeastern region of Missouri is extremely rural, with few, if any, large cities or towns.

Those living in more rural regions often have less access to high quality health services, less quality education, and overall less general information. It is extremely worrisome if Medicare purposefully pays for a smaller proportion of the medical service charges for this simply due to the location of these citizens. There is absolutely no moral explanation for why these citizens are having less of their bills covered, compared to their fellow Missourians living in the city.

Next steps for this analysis would certainly be to dive more into this disparity and see if there are any other factors that contribute to this difference in coverage. Although those living in rural areas may not have access to all of the resources or information their city-living counterparts may have, that is no reason for them to receive less coverage from a government sponsored program.