Yijing Barry Zhang
IEMS 308

Investigation into Medicare Emergency Medicine Providers

Executive Summary
Medicare beneficiaries are one of the most vulnerable subpopulations to the current pandemic due to the qualifying age of Medicare. We have seen a surge in the need for emergency medical attention for these elderly patients and it seems imminent to investigate the capability and quality of care that patients can receive at these emergency departments so that we can discover vulnerabilities and allocate national resources more efficiently. This study specifically looks into the regional differences in emergency care providers who participate in Medicare through the percentage of medical cost that Medicare pays as well as the average number of services that each provider needs to perform.

Clustering is an analytics technique that explores and classifies the data by grouping a dataset into smaller groups based on similarities of certain features. In order to effectively analyze the regional differences, clustering is used in this study to identify the geographical regions of interest. The resulting clusters divide up the US into smaller regions based on the proximity of nearby emergency care providers.

A review of the characteristics in these clusters demonstrate noticeable differences between large urban areas and rural areas, with urban areas generally having higher percentage of Medicare payment as well as lower average number of procedures per emergency room. While the variance among the regions are small, likely due to certain national standards, these noticeable trends could suggest inequalities in access to less costly and more timely emergency care.

Problem Statement:
The current COVID-19 pandemic has put significant stress on our healthcare system. According to the CDC, older adults, who constitute the vast majority of Medicare users, seem to be at higher risk for more serious COVID-19 illness and are therefore more likely to require emergency medical care. This warrants investigation into the current state of Emergency Departments with Medicare participation to uncover regional trends that pose vulnerabilities to a surge of elderly COVID patients. Specifically, I want to look at the ratio of medical cost that Medicare pays (indicator of how good Medicare coverage is) and average number of service provided (proxy indicator of how occupied the ERs would be)

Methodology:
Data:
The data used is the Medicare Provider Utilization & Payment Data provided by The Centers for Medicare and Medicaid Services. For this specific analysis, I used US mainland-only medical providers who practice emergency medicine, participate in Medicare, and practice in a medical facility. The selection of US-only providers is to better capture US-only regional trends, and the selection of medical professionals in medical facilities is to make sure that our insights are specific and relevant to the current pandemic (since only emergency departments have the capability to treat critical COVID patients). In order to better interpret the geographical locations of these medical providers, I also utilized an external data source to convert zip codes in the original data to longitude and latitude.

EDA and Cleaning:

While the data is well formatted, there are certain anomalies that were brought to my attention during the exploratory phase. Namely, there exists very large values in certain fields that have significantly skewed the data. In order to maintain the integrity of the data while still obtaining meaningful results, I decided to remove any entries that are more than 4 standard deviations away from the mean.

Features:

I chose a set of features and created an additional feature for our analysis. For clustering, I clustered only based on the latitude and longitude of the medical providers. These GCS coordinates, in the span of one country, would not produce much error and are directly used as linear coordinates without transformation or normalization. In order to calculate the percentage of medical cost that Medicare covers, an additional "coverage rate" feature was created as the ratio between average_Medicare_payment_amt and average_Medicare_allowed_amt.

Analysis:

The clustering was done on the GCS coordinates to find concentrated clusters that divide up the US. Ideally these clusters will have clearly defined boundaries such that each cluster can be seen as a distinct geographical region and represent the characteristics of the emergency rooms in the region. I initially used the elbow method to select the optimal number of clusters, but no clear "elbow point" was found under 30. This led to a hypothesis that the optimal number of clusters would be fairly large. I continued to observe improvements in silhouette score with additional clusters at larger increments of clusters and in order to have sufficient number of items in each cluster for explainability, I stopped increasing the number of clusters at 200.

The geospatial clustering reached a mean silhouette score of .61 based on a sample size of 10000. The score ranges from -1 to 1, with 1 signaling a very well-suited clustering. This means that this clustering had few overlaps and fairly effectively separated the data into meaningful clusters. There are clusters that clearly center around cities, especially large metropolitan areas. The clusters in less populated areas are more scattered and have longer intrapoint distances.

After the clusters were created, quantitative and visual analyses were done on the cluster properties. Since there are national standards in place, we did not expect nor observe significant variation in the average coverage rate among the clusters (less than 10% of the clusters were 1 standard deviation away from the mean), however, there is a noticeable difference in the coverage rate along the coasts (especially in Urban areas) and central/southern regions of America (mostly rural areas.) This difference is especially apparent in the geographical visualization in the appendix.

The regional differences in the total number of services provided is not as strong but we can once again see that ERs in larger metropolitan areas (such as NYC, Chicago, and San Francisco) have relatively lower average number of services per provider than rural areas (clusters in Wyoming, North Carolina, Tennessee, etc.) Clusters in Southeast and Midwest regions have relatively higher average number of services per provider and those clusters are also shown in the plot in the appendix below.

An interesting theme that emerged is when the graph of coverage rate and numbers of services are superimposed (see appendix.) While the two variables cannot be directly compared, they are put on the same standardized scaling and we can see that there exist clear regional differences between clusters that

exist in major metropolitan areas and the rest of the US. Namely, urban areas seem to have higher Medicare coverage rate and lower numbers of services done per provider. These are both favorable traits since they suggest less timely and less costly emergency care attention. This could suggest vulnerabilities to the pandemic in terms of both timely healthcare and financial cost for elderly patients using Medicare in rural areas.

Limitations:

Geospatial clustering: Given that longitudes and latitudes do not accurately represent distance as we perceive them, the k-means clustering done here is likely an inexact but close proxy to the surface distance between the points. These differences should not overly influence the results since we are limited to a rather small range (mainland US only) but should ideally be accounted for in future studies.

Lack of Features: While this analysis provides a good initial exploration, there are other factors that it cannot account for because of the lack of data. Utilization, for example, is a more direct measurement in how effective an emergency department is at handling a surge of critical condition patients. Adding on those additional data would paint a more holistic picture of the current threats.

Conclusions:

From the analysis of our geographical clusters, we observe certain regional differences that are worth our attention. The Medicare providers in clusters along the coasts seem to cover a larger portion of the medical cost for emergency medical services. This trend is nearly reversed for the average number of services provided per provider. This trend suggests a potential divide between rural and urban American emergency health providers. The combination of more crowded emergency rooms and smaller coverage percentage by Medicare that is widespread in rural America can suggest some degree of health care inequality and could present significant problems with access to emergency care in the current pandemic.

Next Steps:

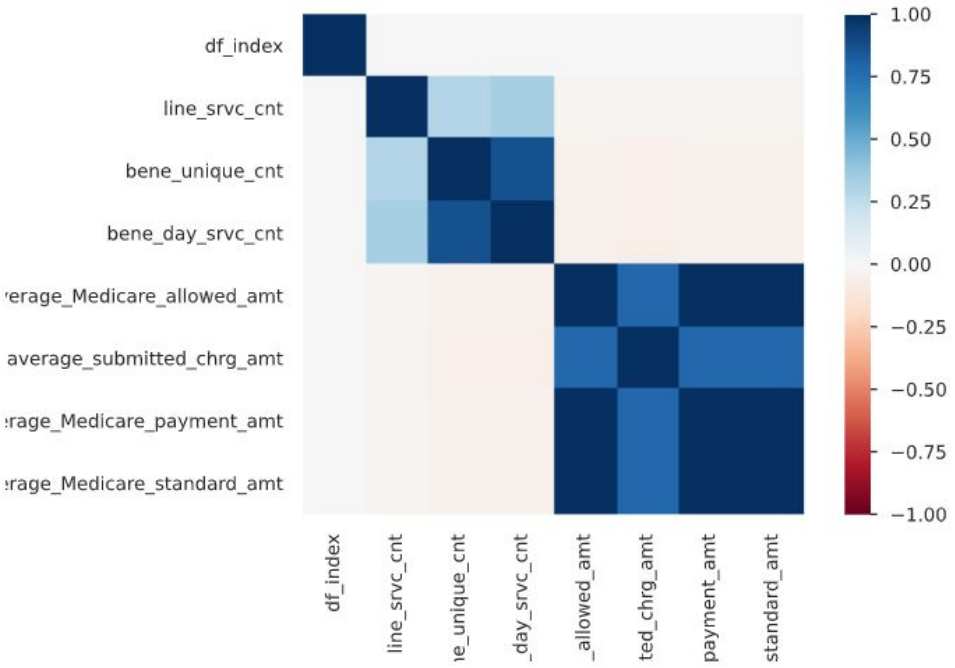
One of the common techniques done to further investigate clustering results is a further clustering within the clusters. In certain areas such as especially dense urban areas or rural areas with especially dense ER presence, it might be worth our attention to further cluster within them such that we have a better understanding of the phenomenons.

We might also benefit from incorporating demographics data as well as additional domain expertise in Medicare policies. We could then better understand the regional differences highlighted in this analysis and have more accurate directions of future targets for improvements.

Appendix:

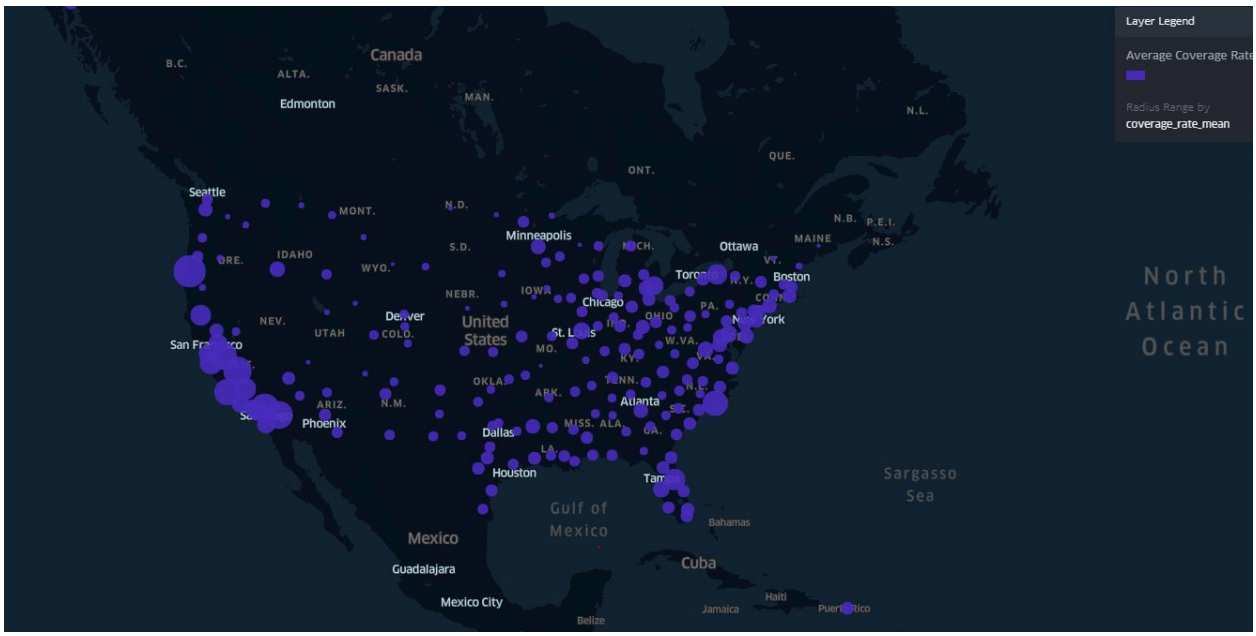
EDA Results:

- High Correlation among the Medicare charge and number of services (pairwise correlation among the numeric variables

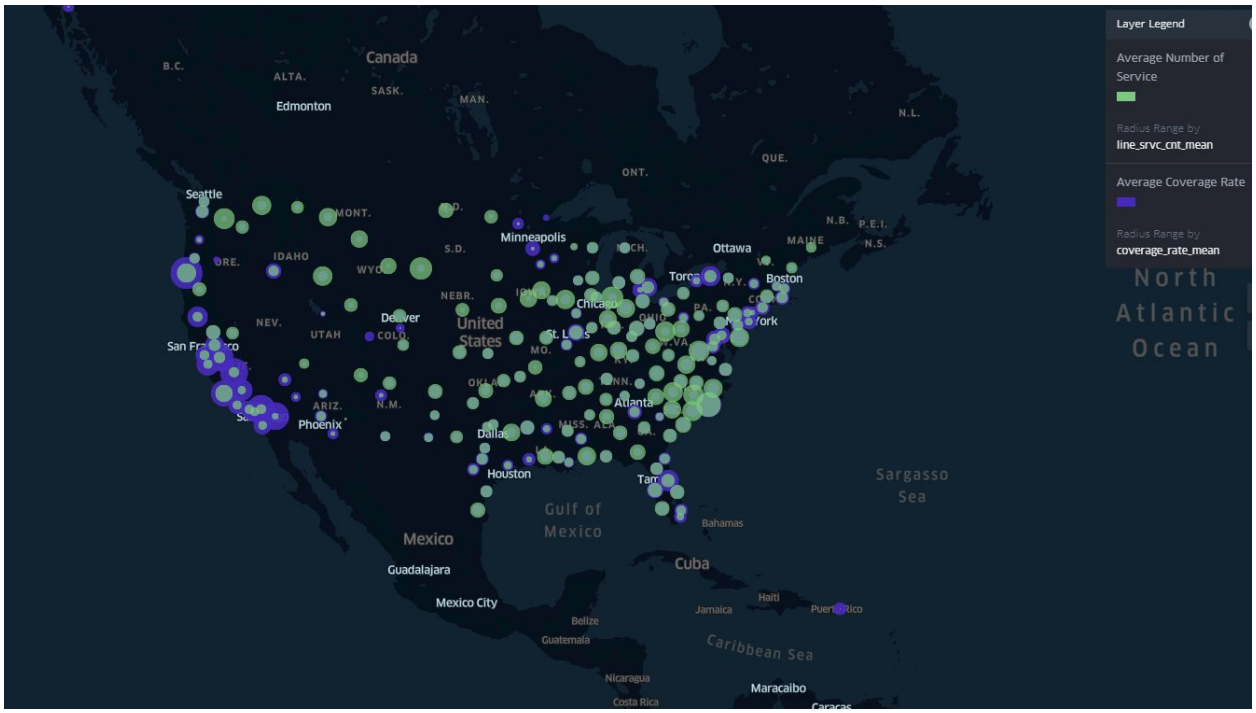


Visualization

Mean coverage amount by Medicare for each cluster (larger size indicating higher coverage)



Mean number of services done each cluster (large size indicating large numbers of service provided)



Superimposed graphs of coverage rate and number of services