

Introduction

Medicare is a federal government-sponsored health insurance program that provides coverage for people who are 65 or older, select younger people with disabilities, and people with End-Stage Renal Disease (permanent kidney failure requiring dialysis or a transplant.)¹ There are 44 million beneficiaries, 15% of the U.S. population, currently enrolled in the Medicare program. Enrollment is projected to rise to 79 million by 2030.²

There is ongoing political and economic debate around Medicare's contribution to the federal deficit. Though the specific data and arguments for this debate fall outside of the scope of this report, we will explore a patient dataset that provides insights into the frequencies and financial costs of 100 disease related groups classified and covered by Medicare. This dataset is made available by the Centers for Medicare and Medicaid Services, a federal agency that administers the Medicare program.³ The data include Medicare Inpatient Prospective Payment System (IPPS) payments for the top 100 most frequently billed discharges at the national level, paid under Medicare based on a rate per discharge using the Medicare Severity Diagnosis Related Group (MS-DRG) for Fiscal Year 2011. Using this data directly, patients insured through Medicare can make informed financial decisions by comparing the amount charged by individual hospitals within their proximities and nationwide for services classified under the relevant DRG.⁴

The dataset is available here:

<https://data.cms.gov/Medicare/Inpatient-Prospective-Payment-System-IPPS-Provider/97k6-zzx3>

The dataset fields are as follows. Clarifications are made for select terms that are relevant to this report.

DRG Definition: code and description used as a classification system that groups similar clinical conditions and the procedures furnished by the hospital during the stay, *Provider Id:* Id used by Medicare, *Provider Name:* Hospital or treatment center, *Provider Street Address, Provider City, Provider State, Provider Zip Code, Hospital Referral Region Description, Total Discharges:* number of discharges billed by the provider for inpatient hospital services *Average Covered Charges:* Provider's average charge for services covered by Medicare for all discharges in the DRG. Vary among hospitals due to differences in charge structures. *Average Total Payments:* Average of Medicare payments to the provider for the DRG including the DRG amount, co-payments, deductibles, and other payments. *Average Medicare Payments:* Average of only the Medicare payments to the provider for the DRG

This report evaluates two variables, the 1. total discharges and 2. total average medicare payments for each of the 100 disease-related-groups in the dataset.

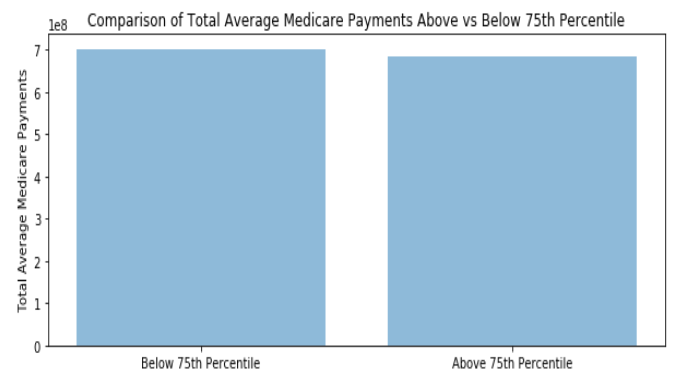
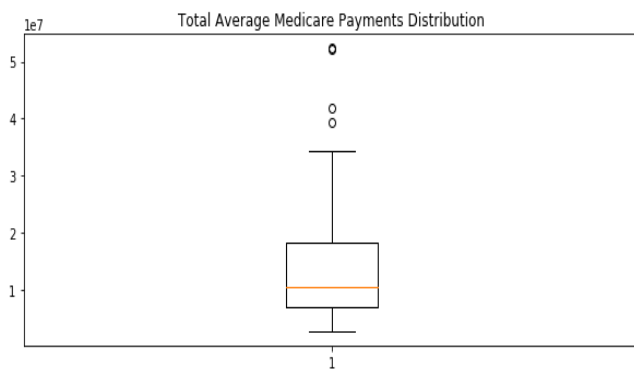
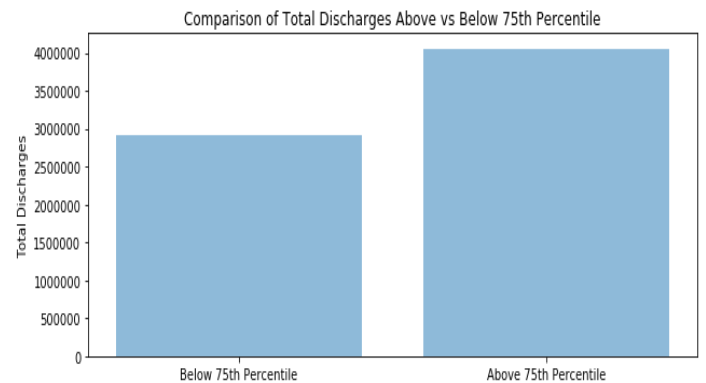
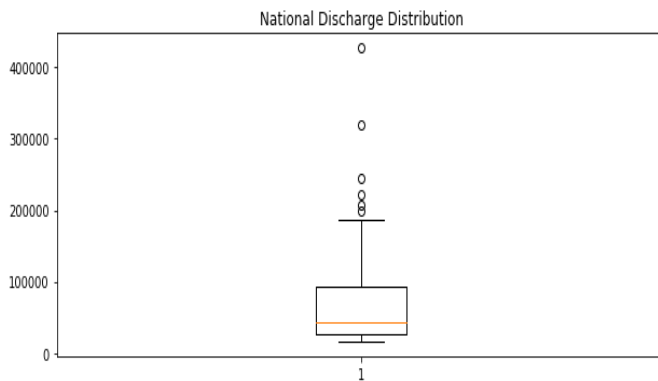
Click [here](#) to access all code relevant to this report on my Github:

Copy/paste if link broken:

<https://github.com/Gulab88/National-Medicare-Inpatient-Dataset/blob/master/Thinkful%20Prep%20Capstone%201%20-%20National%20Medicare%20Inpatient%20Dataset.ipynb>

Summary Statistics and Visualizations: Statistical analysis of a sample was not necessary since the population set was small enough for direct statistical analysis.

Summary Statistics	Total Discharges (Associated with each DRG)	Total Average Medicare Payments (Associated with Each DRG)
Minimum Value	16,157	\$2,824,156.80
Maximum Value	185,599	\$34,277,714.96
Interquartile Range	67,007.75	\$11,357,202.45
Range Excluding Outliers	169,442	\$31,453,558.16
Range Including Outliers	411,050	\$49,512,227.44
Mean	69,753	\$13,851,541.69
Mean Excluding Outliers	56,990	\$12,265,847.91
Median	42,186	\$10,593,718.55
Mode	100 equally common values	100 equally common values
Standard Deviation	67,482.91	\$9,876,772.58



Analytic Questions

1. Which disease-related groups (DRGs) have the highest discharge totals? Are there trends for the total discharges associated with each DRG?

Outliers on the Total Discharges associated with each DRG Boxplot

Disease-Related Group (DRG)	National Total Discharges	Percent of All Discharges (6,975,318)
194 - SIMPLE PNEUMONIA & PLEURISY W CC	198,390	2.84%
690 - KIDNEY & URINARY TRACT INFECTIONS W/O MCC'	206,695	2.96%
292 - HEART FAILURE & SHOCK W CC	222,038	3.18%
392 - ESOPHAGITIS, GASTROENT & MISC DIGEST DISORDERS W/O MCC	244,854	3.51%
871 - SEPTICEMIA OR SEVERE SEPSIS W/O MV 96+ HOURS W MCC	319,072	4.57%
470 - MAJOR JOINT REPLACEMENT OR REATTACHMENT OF LOWER EXTREMITY W/O MCC	427,207	6.12%

The presence of 6 outliers, all representing discharge frequencies greater than the maximum*, is immediately apparent upon examination of the boxplot. These outliers account for 23.18% of total national discharges, meaning that the outliers will significantly affect measures of central tendency for total discharges. The mean excluding outliers is 12,763 discharges smaller. The outliers can serve as special points of interest for reducing Medicare costs since they correspond with DRGs that disproportionately account for such a high number of discharges. Upon crude examination, the boxplot shows that the total discharge sums below the 75th percentile and above the 75th percentile are close in value. This prompted the analyst to create two groups, one representing each sum, and visualize them using a bar graph (included above). Total discharges above the 75th percentile account for about 52% of all discharges, and the total discharges below the 75th percentile account for the remaining 48%. The 25 DRGs represented above the 75th percentile have a much wider distribution of discharge totals than the ones below. The median is also skewed toward the smaller discharge totals, illustrated that more DRGs have smaller discharge totals than larger ones within the context of this dataset. These observations could be the starting point for further analysis.

*maximum refers to the upper whisker of the boxplot

2. Which DRGs are the most expensive for Medicare? Are there trends for the total average Medicare payments associated with each DRG?

Outliers on the Total Average Medicare Payments associated with each DRG Boxplot

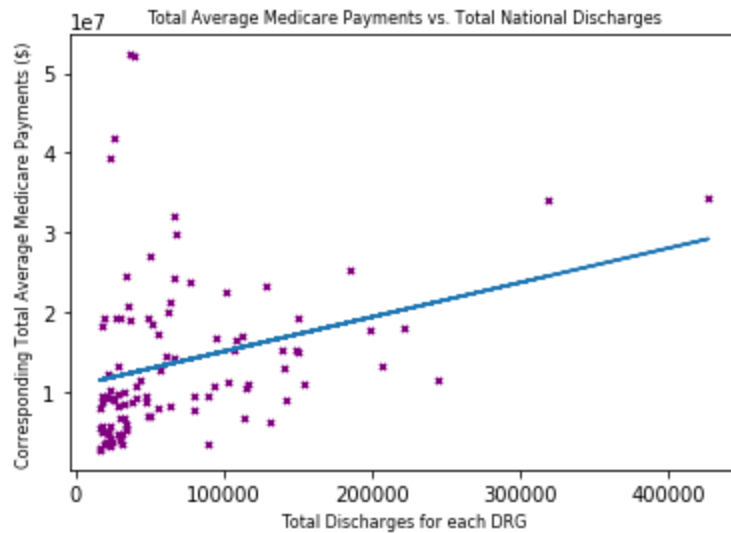
Disease-Related Group (DRG)	Total Average Medicare Payout	Percent of Sum Total Average Medicare Payout (\$1,385,154,169.05)
870 - SEPTICEMIA OR SEVERE SEPSIS W MV 96+ HOURS	\$39,343,567.52	2.84%
207 - RESPIRATORY SYSTEM DIAGNOSIS W VENTILATOR SUPPORT 96+ HOURS	\$41,902,364.87	3.03%
853 - INFECTIOUS & PARASITIC DISEASES W O.R. PROCEDURE W MCC	\$52,038,586.16	3.76%
329 - MAJOR SMALL & LARGE BOWEL PROCEDURES W MCC	\$52,336,384.24	3.78%

The presence of 4 outliers, all representing total average Medicare payments greater than the maximum*, is immediately apparent upon examination of the boxplot. These outliers account for 13.41% of total national discharges, meaning that the outliers will significantly affect measures of central tendency for total average Medicare payments, though they may affect these measures less than the outliers for DRG discharge totals do. The mean excluding outliers is about 1.6 million dollars less. The outliers can serve as special points of interest for reducing Medicare costs since they correspond with DRGs that disproportionately account for higher totals for average Medicare payments. Upon crude examination, the boxplot shows that the sums for total average Medicare payment below the 75th percentile and above the 75th percentile are close in value. This prompted the analyst to create two groups, one representing each sum, and visualize them using a bar graph (included above). Total average Medicare payments above the 75th percentile account for about 49% of all values, and the total average Medicare payments below the 75th percentile account for the remaining 51%. The 25 DRGs represented above the 75th percentile have a wider distribution of associated total average Medicare payment amounts than the ones below. The median is also skewed toward the smaller average Medicare payments, illustrating that more DRGs have smaller total average Medicare payment amounts than larger amounts within the context of this dataset. These observations could be the starting point for further analysis.

*maximum refers to the upper whisker of the boxplot

3. Is there a relationship between the total discharges for each DRG and the associated average Medicare payments? Do DRGs with higher total discharges also receive a higher average Medicare payout?

In order to answer these questions, I plotted total average Medicare payments for each DRG against the total discharges for the same DRG using a scatterplot. I then found a line of best fit and calculated the p-value to determine if the null hypothesis, namely that no relationship exists between these two variables, could be rejected. The p-value for this regression line is 0.003, which means the results are highly statistically significant. The p-value, combined with the line of best fit's positive slope, supports the claim that DRGs with higher total discharges are also the same ones with higher average Medicare payments. This may seem obvious, because Medicare payments are made based on discharges, but this correlation indicates that total average Medicare payments are relatively standard across the DRG range even if the provider requests dramatically different compensations across the DRG range.



Proposal for Further Research

This dataset and the analytic report could be further utilized to evaluate if preventive health measures for the most costly DRGs are effective in reducing the number of discharges associated with the DRG. Of course, the desired reason behind the reduced discharge total would be a reduced total of patients presenting with the DRG-associated symptoms. This research project would entail finding a dataset showing types of preventive health measures and outcomes in subjects who qualify for Medicare, namely the elderly and disabled. In order to accomplish this, I would need to learn how to most efficiently select, organize, and combine information from two or more large datasets. I would also need to learn how best to use python and its packages to conduct descriptive, predictive, and prescriptive analytics in order to present the findings as actionable data.

Sources

1. Medicare.gov
2. assets.aarp.org/rgcenter/health/fs149_medicare.pdf
3. https://en.wikipedia.org/wiki/Centers_for_Medicare_and_Medicaid_Services
4. <https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/Inpatient2011.html>