# Healthcare Provider Data Analysis: Insights, Business Problem, and Optimization Strategies

## Advance Excel Project

POOJIT BHASKAR KASINA

DATA ANALYST TRAINEE & INTERN AT OESON
BATCH NO: OGTIP-BA/DA-AUS-06

OGTIP: OGTIPDAPA230

Mentor: Ms. Smruthi Mehta

# Contents

## Introduction

This report analyses healthcare data, focusing on information about Diagnosis-Related Groups (DRG) across the different states in the United States.

The dataset includes the following fields:

- DRG Definition: This tells the information about the diagnosis-related information of the aliment
- Provider Id: Authorized govt. id of the hospital
- Provider Name: Name of the hospital including details such as
- Provider Street Address, Provider City, Provider State, Provider Region and Provider Zip Code,
- Total Discharges: Number of patients discharged by specific hospitals and Cost fields such as Average Covered Charges, Average Total Payments, Average Medicare Payments
- Each record (or row) is a unique combination of ailment & and provider.

## Data Over View

- The Hospital Charges dataset has more than 150,000 records of treatment of various ailments in several US hospitals.
- The dataset has a record of 100 ailments
- It has a list of 3000+ healthcare providers (hospitals, medical centres, etc.)
- It has data from 50 states of the USA and approximately 2000 cities across the USA
- It describes the cost of treatment for approximately 7 million patients

| | A | B | C | D | E | F | G | H | I | J |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Column1 | Column2 | Column3 | Column4 | Column5 | Column6 | Column7 | Column8 | Column9 | Column10 |
| 2 | DRG Definition | Provider Id | Provider Name | Provider Street Address | Provider City | Provider State | Provider Zip Code | Hospital Referral Reg | Total Discharges | Average Covered Charges |
| 3 | 039 - EXTRACRANIAL PRO( | 10001 | SOUTHEAST ALABAMA M | 1108 ROSS CLARK CIRCLE | DOTHAN | AL | 36301 | AL - Dothan | 91 | $32963.07 |
| 4 | 039 - EXTRACRANIAL PRO( | 10005 | MARSHALL MEDICAL CEN | 2505 U S HIGHWAY 431 N | BOAZ | AL | 35957 | AL - Birmingham | 14 | $15131.85 |
| 5 | 039 - EXTRACRANIAL PRO( | 10006 | ELIZA COFFEE MEMORIAL | 205 MARENGO STREET | FLORENCE | AL | 35631 | AL - Birmingham | 24 | $37560.37 |
| 6 | 039 - EXTRACRANIAL PRO( | 10011 | ST VINCENT'S EAST | 50 MEDICAL PARK EAST D | BIRMINGHAM | AL | 35235 | AL - Birmingham | 25 | $13998.28 |
| 7 | 039 - EXTRACRANIAL PRO( | 10016 | SHELBY BAPTIST MEDICA | 1000 FIRST STREET NORTH | ALABASTER | AL | 35007 | AL - Birmingham | 18 | $31633.27 |
| 8 | 039 - EXTRACRANIAL PRO( | 10023 | BAPTIST MEDICAL CENTE | 2105 EAST SOUTH BOULE | MONTGOMERY | AL | 36116 | AL - Montgomery | 67 | $16920.79 |
| 9 | 039 - EXTRACRANIAL PRO( | 10029 | EAST ALABAMA MEDICAL | 2000 PEPPERELL PARKWA | OPELIKA | AL | 36801 | AL - Birmingham | 51 | $11977.13 |
| 10 | 039 - EXTRACRANIAL PRO( | 10033 | UNIVERSITY OF ALABAMA | 619 SOUTH 19TH STREET | BIRMINGHAM | AL | 35233 | AL - Birmingham | 32 | $35841.09 |
| 11 | 039 - EXTRACRANIAL PRO( | 10039 | HUNTSVILLE HOSPITAL | 101 SIVLEY RD | HUNTSVILLE | AL | 35801 | AL - Huntsville | 135 | $28523.39 |
| 12 | 039 - EXTRACRANIAL PRO( | 10049 | GADSDEN REGIONAL MEI | 1007 GOODYEAR AVENUE | GADSDEN | AL | 35903 | AL - Birmingham | 34 | $75233.38 |

*Figure 1 Raw Sample Health Care Data*

## Data Cleaning

- Firstly, I utilized Power Query, a data cleansing tool in Excel 2016 to eliminate any outliers in my dataset.

*Figure 2 Raw Data in Power Query*

- I removed duplicate records, blank/empty rows/null values, and made sure that the column data types were correctly defined to enhance the accuracy of my analysis.

- Secondly, I removed the negative values in the discharge field and I made the first row of the dataset my header which made it easier to sort and filter the data based on my analysis criteria.

- This process allowed me to perform a thorough analysis of the entire dataset with a high level of certainty and accuracy.
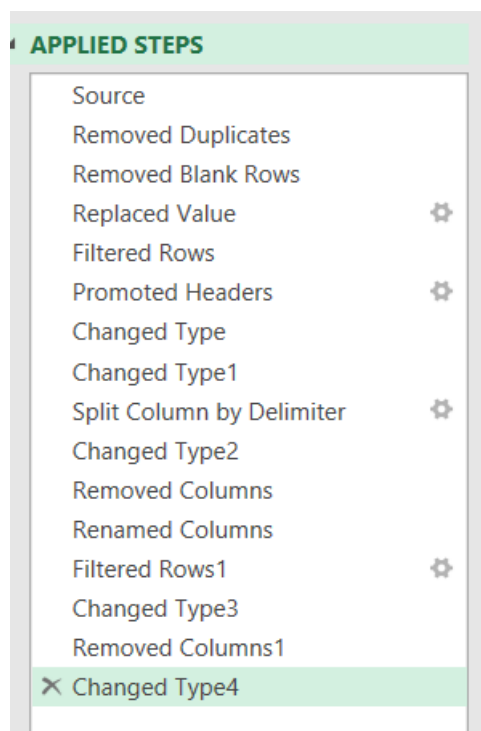


*Figure 3 Data Cleaning in Powery Query*

## Insights:

➢ After completing the process of data cleaning, I analysed the data and extracted valuable insights. These insights are beneficial for both healthcare providers and patients. By identifying patterns and trends within the data, I was able to identify potential solutions for healthcare providers to improve their services and patient outcomes. Additionally, patients can benefit from these insights by gaining a deeper understanding of their health and making more informed decisions regarding their care.

1. **Average Charges by Provider State:**

   Analysing the average covered charges by provider state to identify which states have the highest and lowest charges for the given dataset. This can help patients make informed decisions about where to seek medical care.

2. **Relationship Between Discharges and Payments:**

   To find out the relationship between the number of discharges and the average total payments. Do providers with more discharges tend to receive higher payments?

3. **Regional Payment Differences:**

   To Investigate payment costs across different provider regions. Are there significant variations in average payments, and find out the differences?

4. **Cost Provider Performance Comparison:**

   Identify top-performing and underperforming providers and compare the performance of individual providers based on metrics like "Average Covered Charges," "Average Total Payments," and "Average Medicare Payments."

5. **Medicare Payment Analysis:**

   To examine the data to understand the distribution of Medicare payments. Are there providers receiving significantly higher or lower Medicare payments?

# Business Objective, Problem, Solution and Data Visualization

1. Objective:  Relationship between DRG Definition and Costs

- Problem: Understand how different Diagnosis-Related Groups (DRG) definitions relate to healthcare costs.

- Objective: Analyse the data to determine if there are variations in costs based on DRG definitions. Identify which DRGs are associated with the highest and lowest average costs.

- Solution: Group the data by DRG Definition and calculate the average costs for each group. Create a bar chart or box plot to visualize the distribution of costs across different DRGs. This analysis can inform decisions related to resource allocation and pricing strategies for specific medical procedures or conditions.
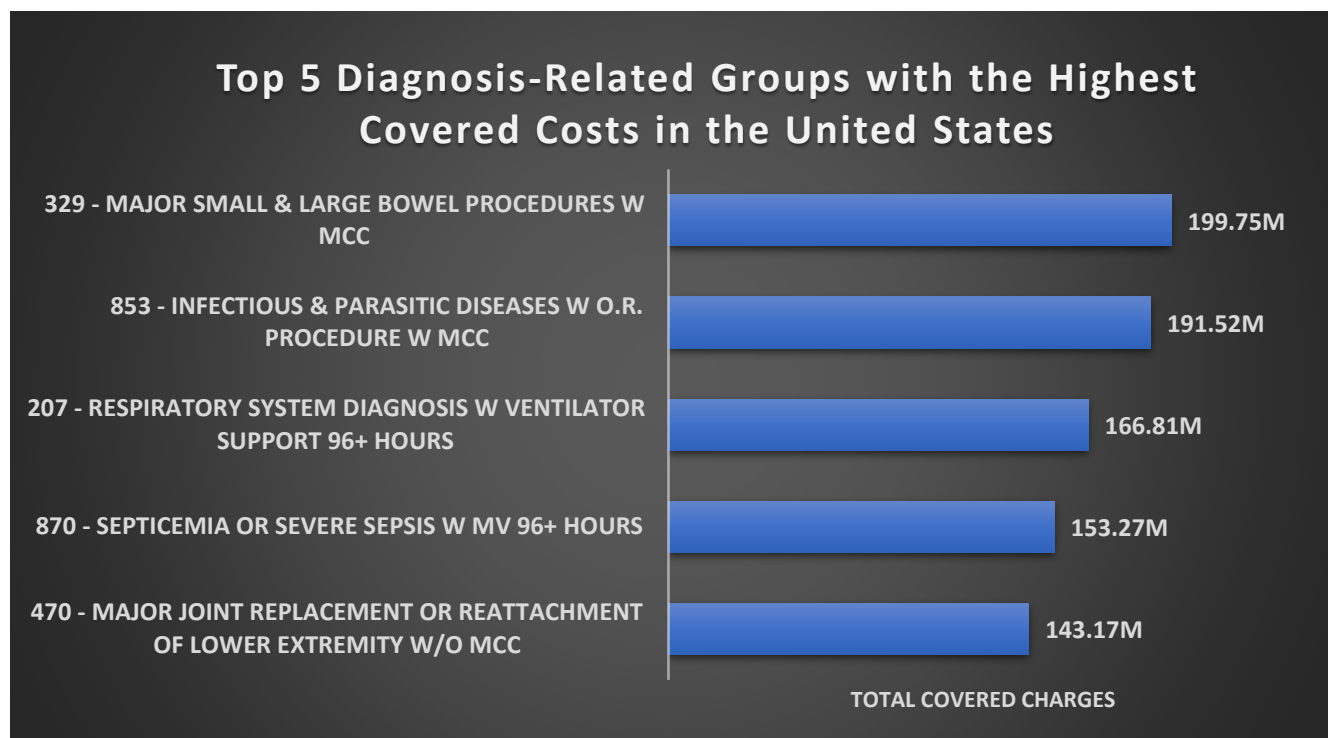


*Figure 4 Top 5 Diagnosis*

## Data insights from Visualization

- ✓ The most expensive medical procedure in the data is "329 - MAJOR SMALL & LARGE BOWEL PROCEDURES W MCC," costing $199.75 M.

- ✓ The data includes the top three diagnoses related to respiratory, infections and specific procedures with or without complications.

- ✓ Patients can easily compare the prices of these diagnoses across different states to make informed healthcare decisions.
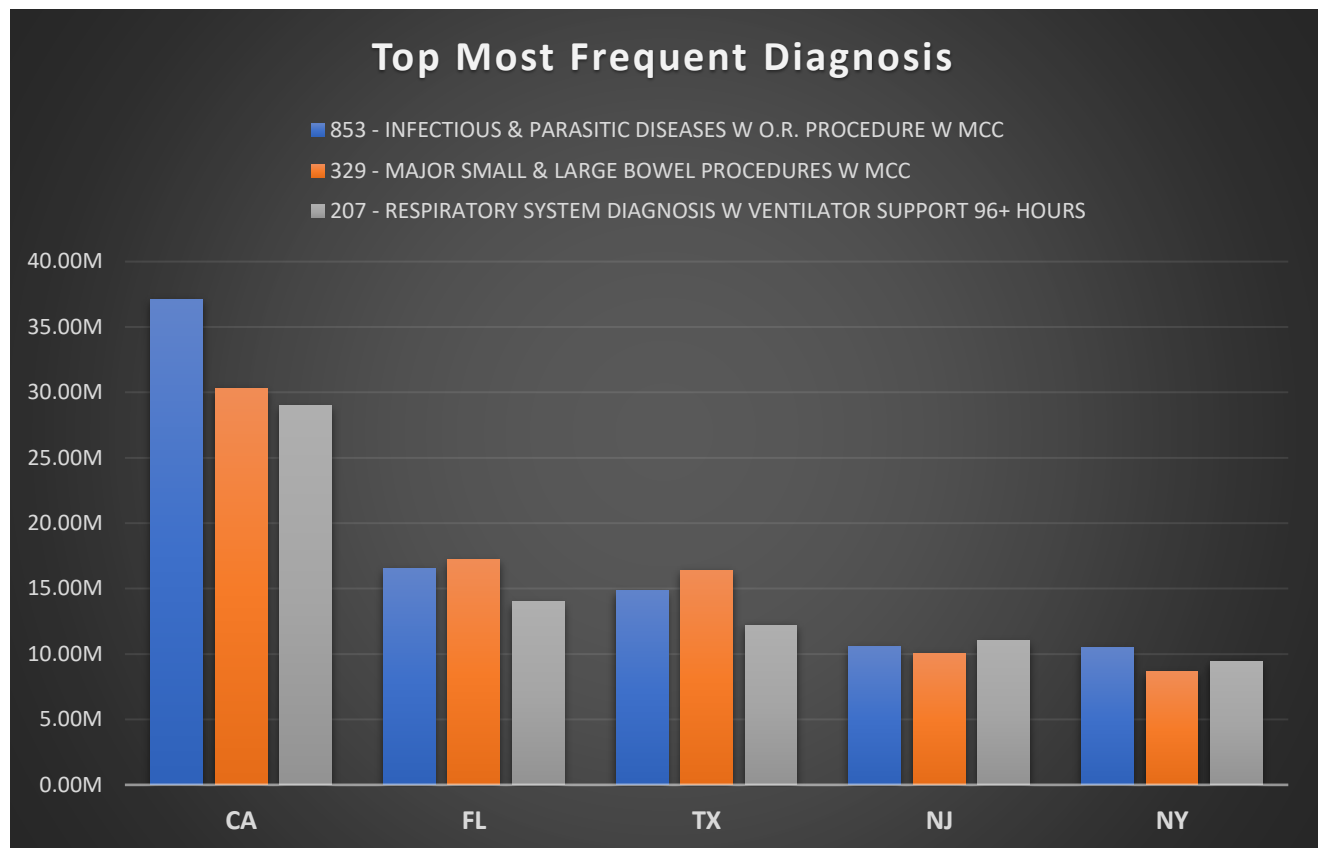
*Figure 5 Most Frequent Diagnosis*

### Key Points from Data Visualization:

✓ Majority of Diagnoses Occur in California: Looking at the data visualization, we can see that most diagnoses are happening in the state of California. This means that a significant number of people are receiving medical diagnoses in California compared to other states.

✓ Top Diagnoses Are Expensive: The data also shows that the most common diagnoses come with substantial covered charges. Specifically, these top diagnoses cost about $37.1 million, $30.31 million, and $28.99 million. This suggests that treating these particular health conditions can be quite expensive.

✓ Gathering information for Government Agencies: This can be valuable for government agencies responsible for healthcare and public health. It draws attention to the need for better treatment and care for patients in hospitals and indicates that the government should focus on improving the quality and affordability of healthcare services for people with these common diagnoses to ensure better overall health in the population.

✓ In simpler terms, the data shows that a lot of people in California are getting diagnosed with certain health issues, and treating these issues can be costly. This should prompt government agencies to pay extra attention to making healthcare more effective and affordable for patients with these diagnoses to promote better health in the community.

2.  Objective: Relationship between Geographic Location and Average Total Payments

-   Problem: To Determine how geographic location (state) affects average total payments for healthcare providers.

-   Objective: By Analysing the data to identify trends and patterns in average total payments across different states, and explore whether certain states have significantly higher or lower average payments than others.

-   Solution: Create a bar chart or map to show the variations in payments by state. This analysis can help identify regions with cost and inform strategies for cost optimization and treatment that will be better with the help of information provided by the government.
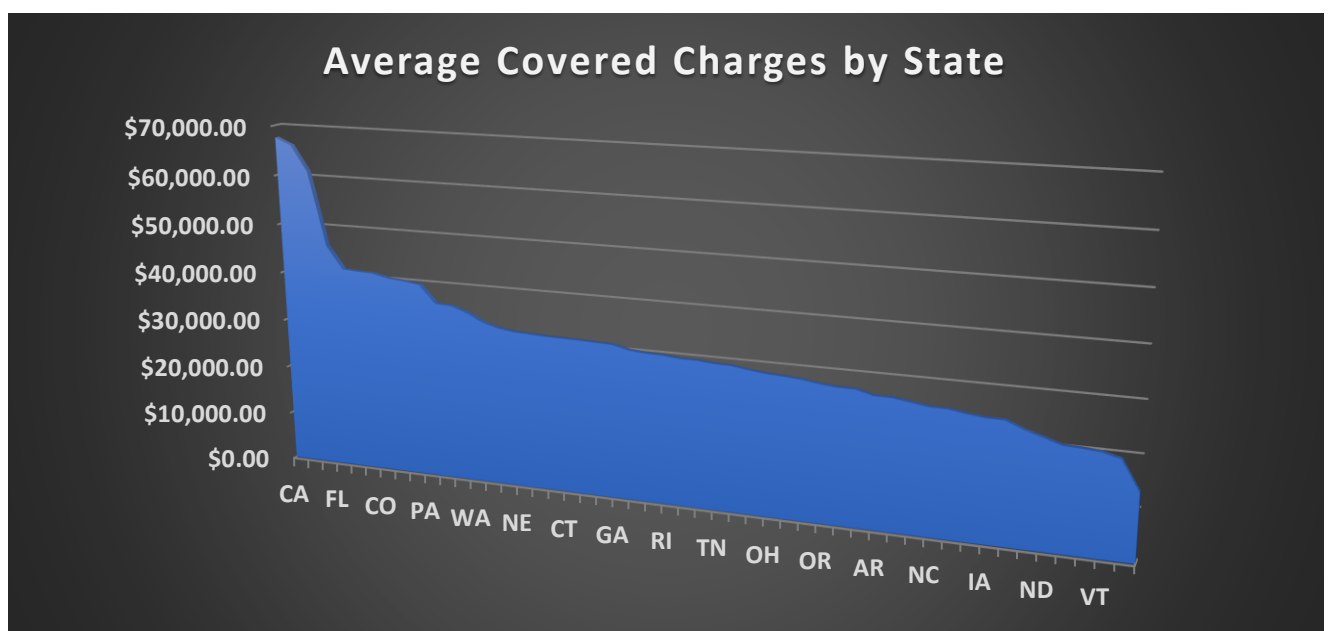


*Figure 6 Avg Covered Charges*

## Key points from Data Visualization

✓   California's High Average Cost: In the area chart, it's evident that California (CA) had the highest average covered cost, with an average of approximately $67,567.45 per case. This means that, on average, medical expenses were relatively high for patients in California.

✓   Vermont's Low Average Cost: On the other hand, Vermont (VT) had the lowest average covered cost among the areas, with an average of about $13,377.80 per case. This indicates that medical expenses for patients in Vermont were comparatively lower.

✓   Visualization of Average Costs: The area graph effectively illustrates and compares the average costs across different areas. It provides a clear visual representation of how much, on average, patients were charged for healthcare services in each location.

3.  Objective: Relationship between Geographic Location (City) and the Number of Healthcare Providers

- Problem: how the geographic location (City) is related to the number of healthcare providers?

- Objective: Analyse the data to determine if there are regional differences in the number of healthcare providers and if certain states have more or fewer providers.

- Solution: Count the number of healthcare providers for each state and create a bar chart. This analysis can help identify areas with potential provider shortages or areas with a surplus of healthcare facilities.
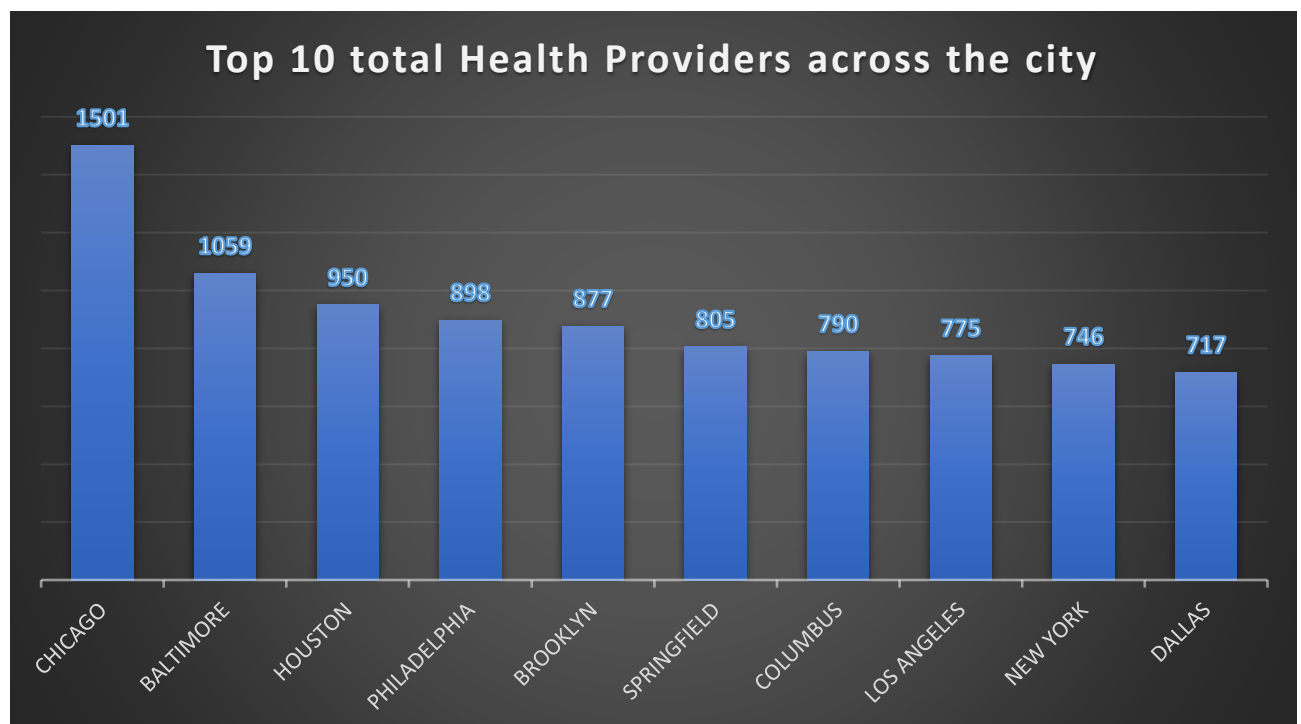


*Figure 7 Total health Providers*

## Key points from Data Visualization:

✓ Chicago Leads in Healthcare Facilities: Among the top 10 cities across different states, Chicago has the most healthcare facilities, with approximately 1501 health providers. This indicates that Chicago offers a wide range of healthcare services and providers.

✓ Cities with Fewest Facilities: Conversely, when examining the data, it's clear that Woodland Hills, Trophy Club, and Zuni have the fewest healthcare facilities. These cities offer relatively fewer options for healthcare services compared to others.

✓ Using Visualization can help people make informed choices about where to seek healthcare. When people see that Chicago has the highest number of facilities, they may be more inclined to consider it as a suitable location for their healthcare needs.

# COST ANALYSIS

4.  Objective: Relationship between Total Payments and Costs across Health Providers

- Problem: Understand the relationship between the number of total paymentss and healthcare costs.
- Objective: Analyse the data to assess whether providers with more discharges tend to have higher or lower costs.
- Solution: Analyse the correlation between total discharges and costs. Create a stacked bar chart to visualize the relationship to finding affordable health providers



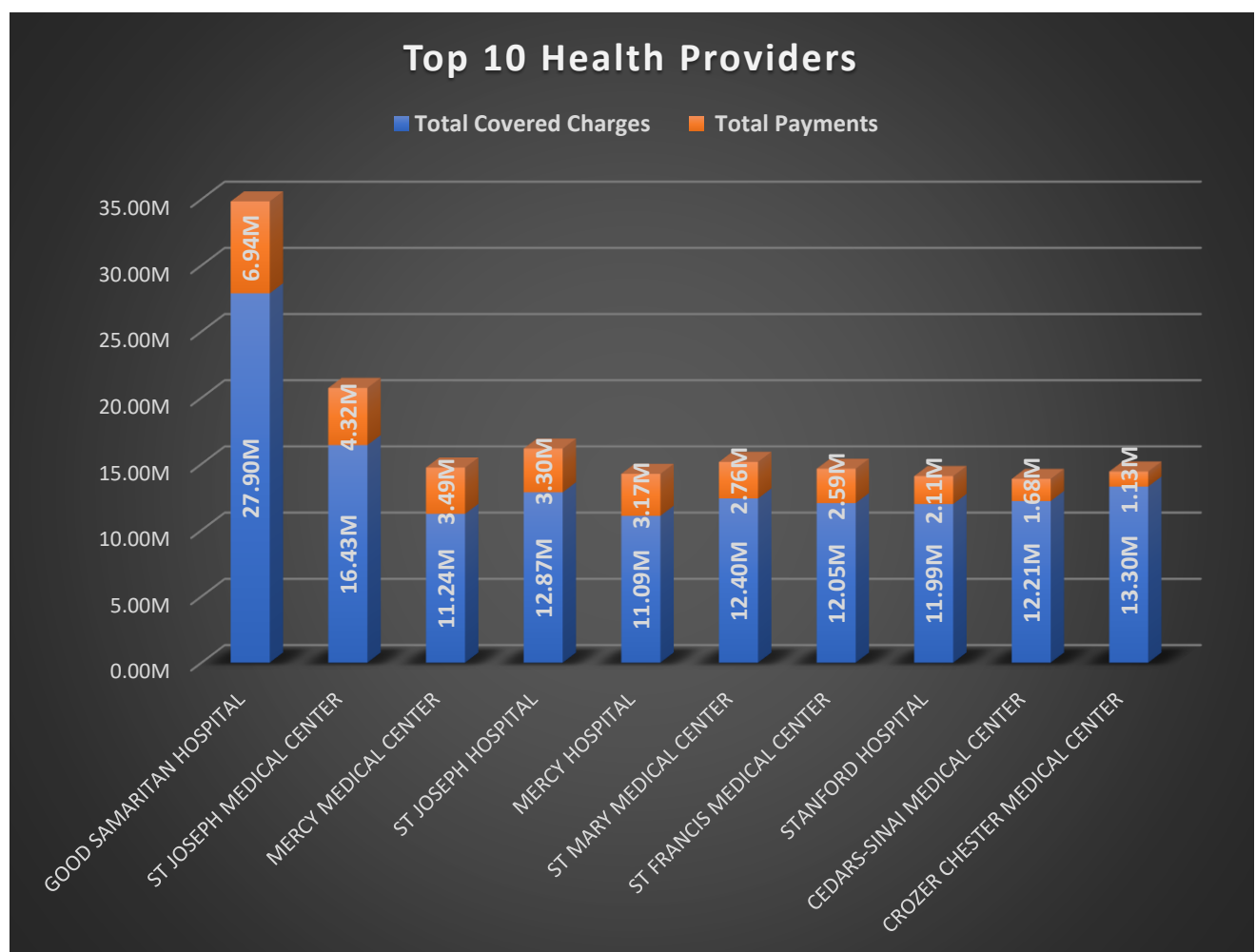*Figure 8 Health Provider Analysis*

## Insights from Data Visualization:

✓  Among the top 10 health providers, Good Samaritan Hospital has the highest total covered charges and total payments. They receive the most money for their healthcare services.

✓ Community General Hospital Has Lower Costs: In contrast<mark>, Community General Hospital has the lowest total covered charges,</mark> amounting to $2,995.61, and the lowest total payments, which are $3,623.44. This indicates that their healthcare services are relatively more affordable compared to others.

5. Objective: Relationship between health provider and Medicare payment

- Problem: There may be disparities in Medicare payments among different healthcare providers, which could indicate issues with reimbursement rates and billing practice

- Solution: Conduct a comprehensive analysis of Medicare payments to various healthcare providers, examining payment rates, frequency, and any discrepancies.
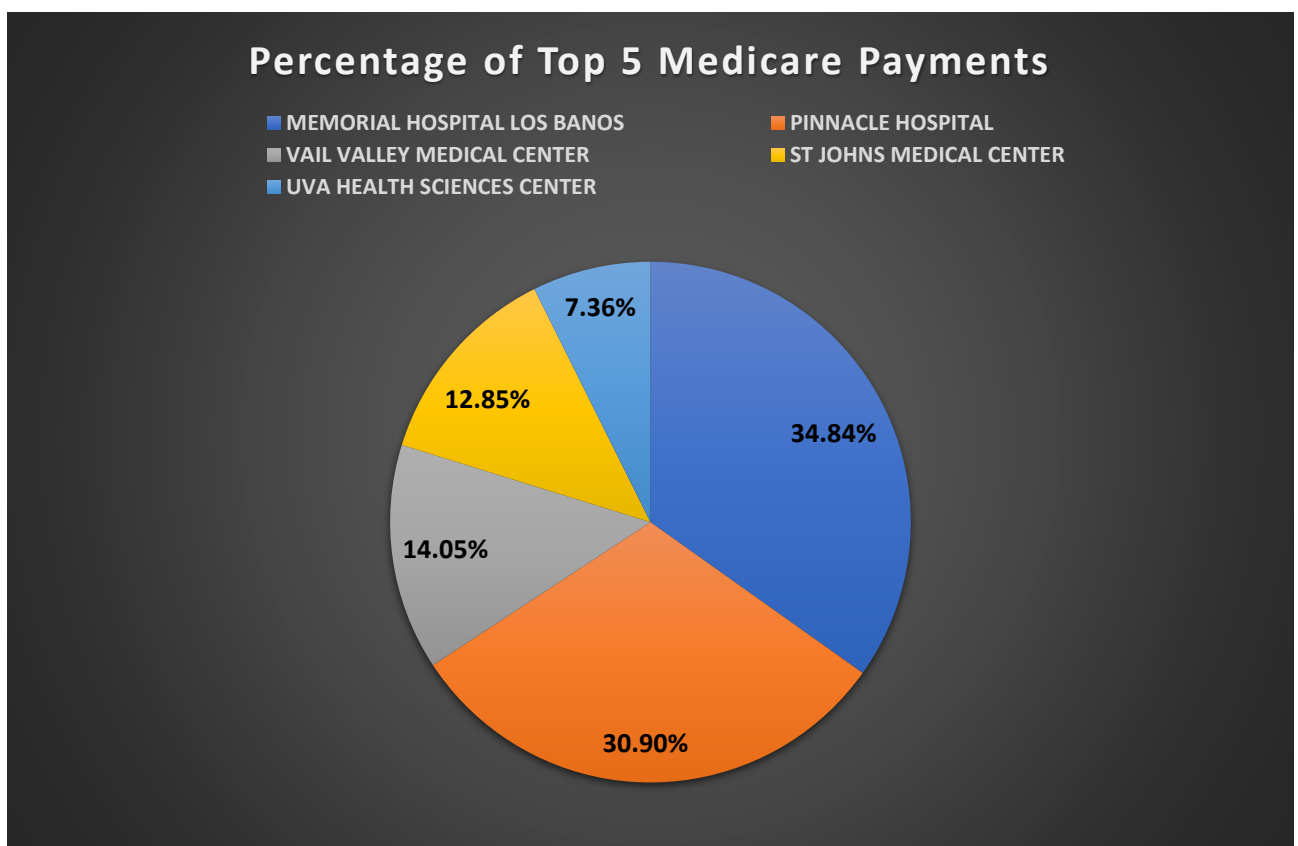


*Figure 9 Medicare Payment Analysis*

## Data Insights from Visualization:

✓ Looking at the pie chart of the top 5 Medicare payments, it's evident that the majority of Medicare payments, approximately <mark>34.84%</mark>, go to <mark>MEMORIAL HOSPITAL LOS BANOS</mark>. This information can help patients understand where significant financial support is available.

✓ On the other hand, <mark>SONOMA DEVELOPMENTAL CENTER</mark> receives the lowest payments, totalling about $6,641.81. This indicates that they receive relatively less financial support from Medicare compared to others.

6. Objective: Examine how healthcare services and payments differ in various regions to help make better healthcare and more efficient for everyone.

- Problem: Some areas have higher healthcare costs and limited access to quality care, causing unfairness in healthcare.

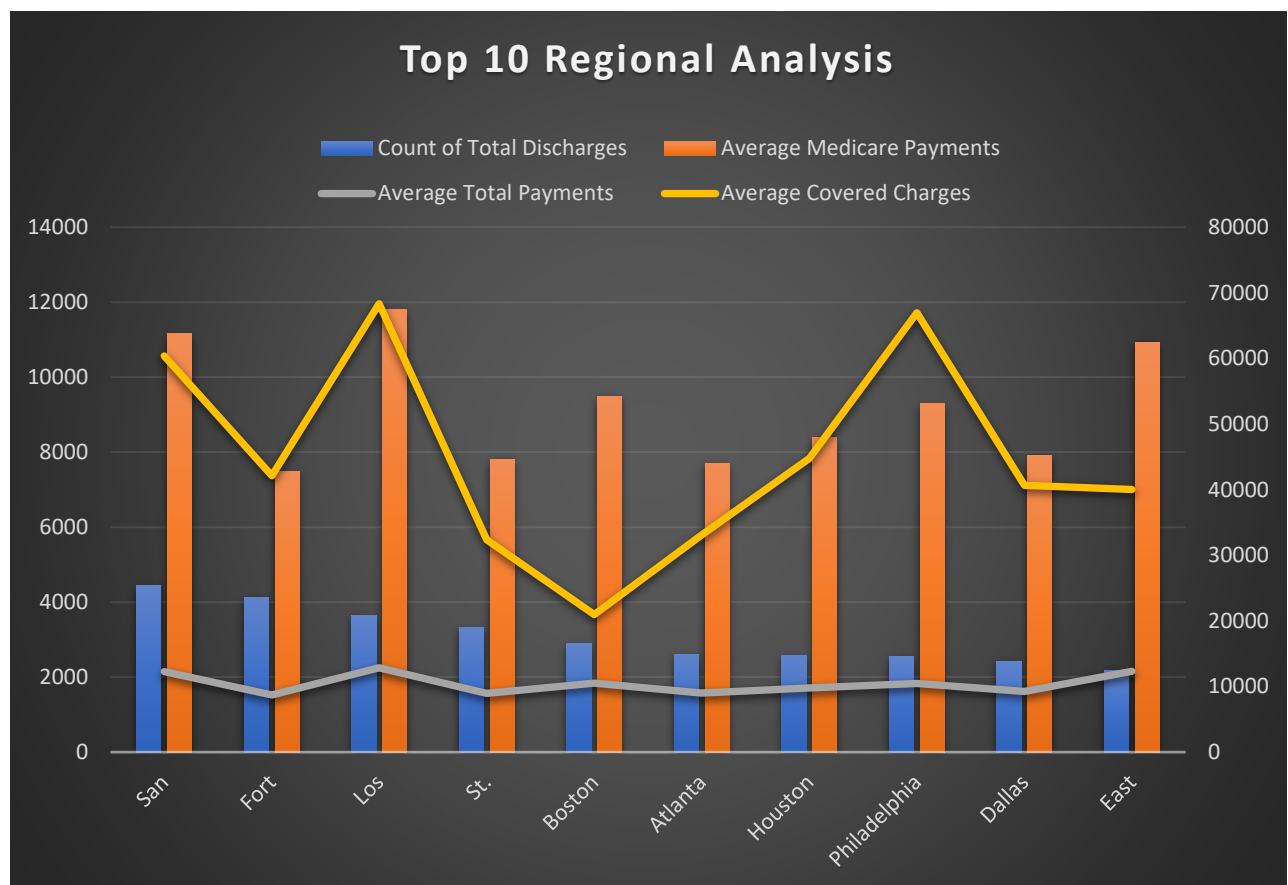- Solution: Create policies and actions to make healthcare fairer and more affordable in all regions.



*Figure 10 Regional Cost Performance*

Insights for Data Visualization:

✓ Regional Discharge Disparities: The count of total discharges varies significantly across regions. "San" has the highest count of total discharges at 4,434, indicating a potentially higher patient volume, while "East" has the lowest count at 2,187.

✓ Average Medicare Payments: The "San" region stands out with the highest average Medicare payments per discharge, reaching $11,154.07. This suggests that healthcare services in this region may be relatively more expensive or involve specialized treatments.

- ✓ Total Payments and Covered Charges: Regions like <mark>"Los"</mark> and <mark>"Philadelphia"</mark> have relatively higher average total payments and average covered charges per discharge, indicating potentially higher overall healthcare costs and billing rates.
- ✓ Cost-Efficient Regions: <mark>"Boston"</mark> and <mark>"Dallas"</mark> regions exhibit relatively lower average Medicare payments, total payments, and covered charges, suggesting that they may offer more cost-efficient healthcare options for patients.
- ✓ Policymakers and healthcare providers can use this data to target areas for cost analysis and explore opportunities for improved healthcare access and affordability.

These objectives, problems, and solutions can provide valuable insights into various aspects of healthcare management, cost optimization, and resource allocation.

## Recommendations

- ➢ Discharge information will be useful for the government to go further into the reasons where patients get to know about the treatment provided by health provider
- ➢ Regular Updates on cost predictions will be good for patients with budget
- ➢ Knowing the cost of the treatment with or without Medicare might encourage patients awareness
- ➢ Hospitals should invest in better systems to ensure they bill correctly, submit claims on time, and collect payments efficiently, reducing the gap between charges and payments.
- ➢ Medicare should provide clear guidelines and training to healthcare providers on how to correctly bill and communicate with Medicare to optimize payments.
- ➢ Regions with lower payments and disparities should invest in healthcare infrastructure and workforce development to improve access to quality care.
- ➢ To Encourage regions to adopt cost-efficient healthcare practices and share best practices among them