**Final Report**

PATIENTS’ EXPERIENCE IN PUBLIC HEALTHCARE

By

Khoa Nguyen

Lisha Choudhary

Martin Vivas

Esther Okolie

Neha Pathak

# TABLES OF CONTENT

[TABLES OF CONTENT](#_um39bn809j4d)

[INTRODUCTION](#_ybv77mj54vs7)

[DATA DESCRIPTION & PRE-PROCESSING](#_2a3i7cr61ajn)

[Source of data](#_9gytltxvzisc)

[Dataset Description](#_eii3ypp8ish1)

[Pre-Processing](#_39tw64s4v4xe)

[DESCRIPTIVE ANALYSIS](#_paveuv5w5hel)

[Separate Analysis](#_erjk9e1vktbd)

[Payment Analysis](#_d4rrtfd5hvee)

[Payment by Regions](#_p4j46xke0lwa)

[Payment by States](#_qdwiknq6rq93)

[Payment by Selected States](#_cxyq4ru0325k)

[Payment: Disease-wise](#_43o2re7c77wk)

[Payment: Disease-wise by 5 Selected States](#_46lpgh2dq498)

[Hospital Consumer Assessment Analysis](#_vb45x24pfuxs)

[Analysis by Category](#_7graw3qb3sl7)

[Comparison of Categories of Regions](#_tbppi9evjo8i)

[Analysis by State](#_rxxxdazbtht5)

[Analysis of Overall Rating Score](#_p4v1x8vae3ir)

[Overall Rating Comparison by State](#_64yxoa37o2yw)

[Cleanliness Score Comparison by Region](#_d5ip2u7nu8id)

[Timely and Effective Care Analysis](#_t69kswj66zlt)

[Subtitle if you need](#_hxlf1sfts3h)

[Subtitle if you need](#_x9u2h0fh7jr7)

[Joint Analysis](#_juqbalq6kqpi)

[Results and Evaluation Plan](#_fwvi459gby9a)

[Completed Tasks](#_ysiqb256s93)

[Challenges & Future Work](#_dka2ie6iy2cf)

[Learning Outcome & Recommendation](#_ev610zp1838g)

[Citations](#_n281wiq2qog1)

# INTRODUCTION

Public healthcare stands as a cornerstone of societal well-being as it represents the principles of fairness and inclusivity by ensuring that all individuals, regardless of their financial circumstances, can access essential healthcare services. Keisler-Starkey, Bunch, and Lindstrom's (2023) findings underscore the significance of public healthcare in the United States, revealing that among the 92.1% of the population who had insurance coverage in 2022, a notable portion—36.1%—relied on public plans, with the remaining 65.6% covered by private plans. However, the contrast between these two sectors extends beyond the coverage statistics. Private plans offer a plethora of choices for providers and benefits. Public plans, on the other hand, only consist of two main programs: Medicare and Medicaid services. Due to the many limitations that a public healthcare consumer might face, our team wants to generalize their experience in terms of payment made by the government, hospital assessment from surveys and effectiveness of treatment. These analysis will be then broken down into 3 levels: nation, regions, and states.

# DATA DESCRIPTION & PRE-PROCESSING

## Source of data

The health data is provided by Medicare and Medicaid, which is available on the website of the Center of Medicare and Medicaid Services (CMS).

The region and state was provided by Kaggle (S. Omer, 2018)

## Dataset Description

Data includes 3 different tables:

|  | CSV Name | Rows | Columns |
| --- | --- | --- | --- |
| HCAHPS-Hospital | HCAHPS-Hospital.csv | 449,097 | 22 |
| Payment and Value of Care | Payment\_and\_Value\_of\_Care-Hospital.csv | 18,560 | 22 |
| Timely and Effective Care | Timely\_and\_Effective\_Care-Hospital.csv | 115,498 | 16 |
| Region and State | state\_regions.csv | 4 | 54 |

|  | Description |
| --- | --- |
| HCAHPS-Hospital | The data records responses from The HCAHPS Patient Survey, which is a survey instrument and data collection methodology for measuring patients’ perceptions of their hospital experience. The HCAHPS survey contains patient perspectives on care and patient rating items that encompass key topics: communication with hospital staff, responsiveness of hospital staff, communication about medicines, discharge information, cleanliness of hospital environment, quietness of hospital environment, and transition of care. |
| Payment and Value of Care | Data is collected from Medicare patients, Medicare managed care patients and non-Medicare patients. Timely and effective care measures include severe sepsis and septic shock, COVID-19 Vaccination, cataract care follow-up, colonoscopy follow-up, heart attack care, preventive care, cancer care measures, stroke, and venous thromboembolism. |
| Timely and Effective Care | The payments included in this measure are price-standardized and risk-adjusted for heart attack, heart failure, pneumonia, and hip/knee replacement. Hospitals’ rates are compared to the national mean payment to categorize whether a hospital’s payment rate is less than the national mean payment, no different than the national mean payment, or greater than the national mean payment. |
| Region and State | The states are matched with the appropriate regions and divisions. |

## Pre-Processing

Each dataset was performed separately so there are different pre-process steps. But they follow the same principles:

* Checking and converting data type
* Checking for missing values or “Not Available” values
* Removing columns with unnecessary information, such as footnotes
* Renaming values or columns for better clarity or joining purposes
* Summarizing information by different methods:
  + Create columns and generate a pivot table for measurements
  + Break measurements into different data frames based on measurements and data types (categorical or numerical)
* Joining datasets with the region table to add a new column for the region.

# DESCRIPTIVE ANALYSIS

Data is analyzed separately and then as a whole

## Separate Analysis

### Payment Analysis

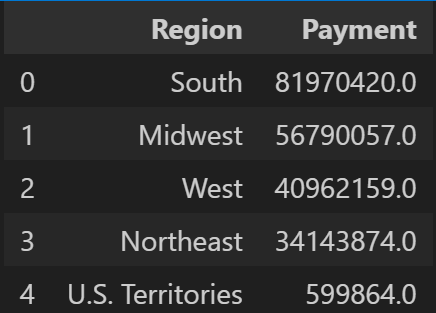
The Payment and Value of Care analysis gives the amount the government funds health facilities through Medicare and Medicaid.

The payment was based on 4 diseases, which are:

* Heart attack
* Heart failure
* Hip/Knee replacement
* Pneumonia

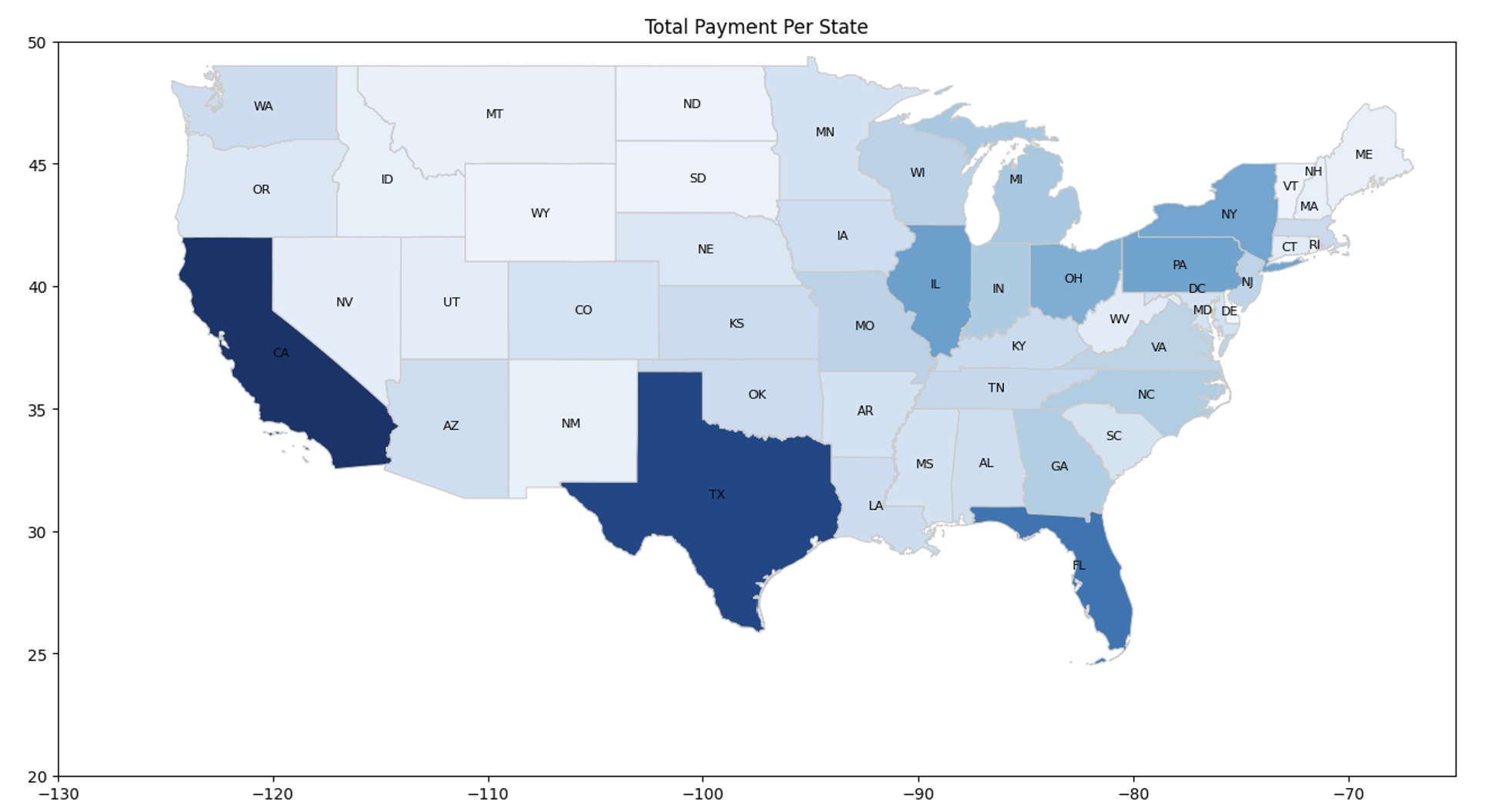
The first analysis was made to determine the total amount paid in the entire country which was over 214 million. This amount was paid to 4510 different facilities during the reporting year.

#### Payment Analysis by Regions

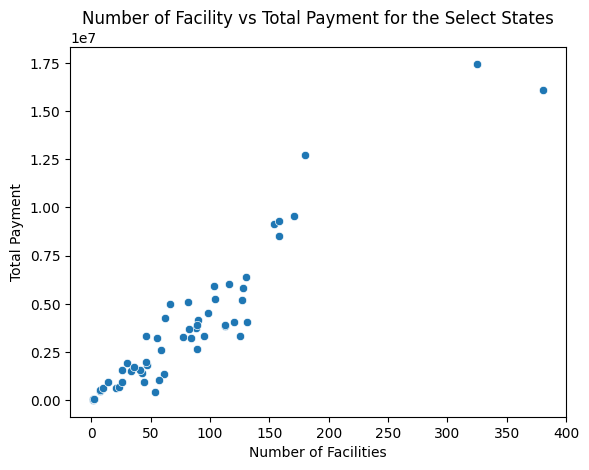


Looking at the payment distribution for the regions, we see that the South amassed the most payments with almost 82 million, followed by the Midwest. The region with the least payment is the US Territories with a total amount of less than 600k dollars.

#### Payment by States



From the data, the total number of states that received payment are 56 states including US territories. An analysis was made to show the distribution of the payments in each of these states. From the map, we see that the states that stood out and received the most payments are California, Texas, and Florida. One possible reason is that these states with higher payments also have a larger number of facilities. The Scatter plot below also shows a positive relationship or correlation between the number of facilities and the payments received.



We would narrow our analysis to look at these states and possible reasons they received more funds in comparison to other states.

##### Payment Analysis by Selected States

Further narrowing the analysis to a comparative study of highly funded states with the state of current residence Indiana. Distribution of payment could be directly affected by population and number of facilities in all respective states. The approximate amount of population is compared below.

**California (CA):**

**Payment:** $17,450,375.00

**Population (Large):** 39.5 million (as of 2021).

**Extensive Healthcare Facilities:** Boasts numerous hospitals, though the exact count may vary.

**Robust Economy:** California has one of the largest economies globally, contributing significantly to tax revenue.

**Texas (TX):**

**Payment:** $16,080,207.00

**Population (Large):** 29.1 million (as of 2021).

**Diverse Healthcare Landscape:** Offers a wide range of medical services across its urban and rural areas.

**Economic Strength**: Strong economy, attracting federal support for various sectors, including healthcare.

**Florida (FL):**

**Payment**: $12,702,206.00

**Substantial Population:** 21.7 million (as of 2021).

**Varied Healthcare Services:** Florida's healthcare system caters to diverse needs, including those of its large elderly population.

**Tourism Boost:** Florida's healthcare infrastructure supports millions of annual tourists, in addition to its resident population.

**Indiana (IN):**

**Payment**: $6,041,647.00

**Population**: 6.7 million (as of 2021).

**Healthcare Landscape**: The Healthcare system emphasizes collaboration between urban medical centers and rural healthcare providers.

**Economic Factors**: Federal funding supports efforts to improve healthcare access and affordability in Indiana.

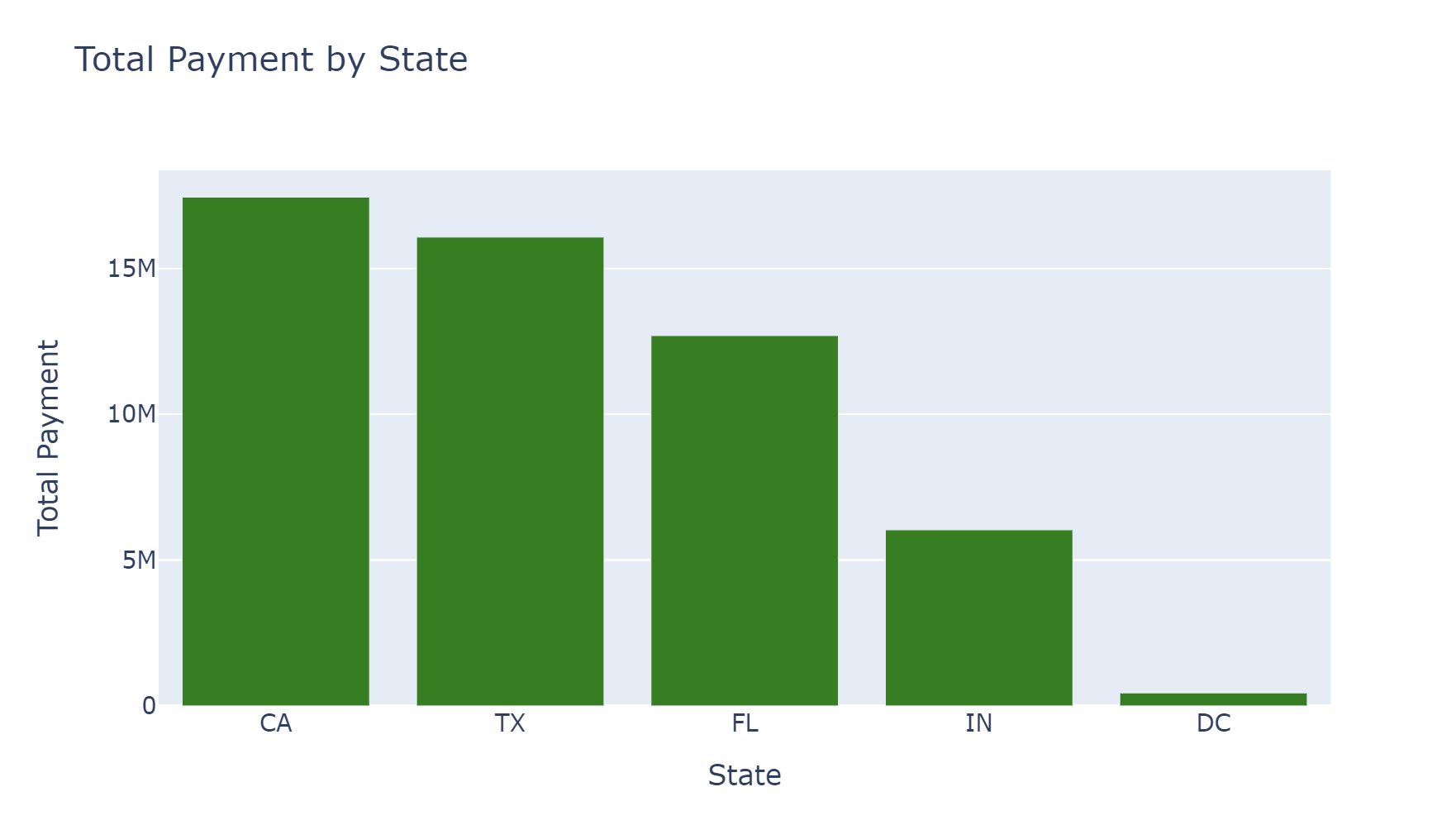
**District of Columbia (DC):**

**Payment**: $448,774.00

**Population**: 700,000 residents (as of 2021).

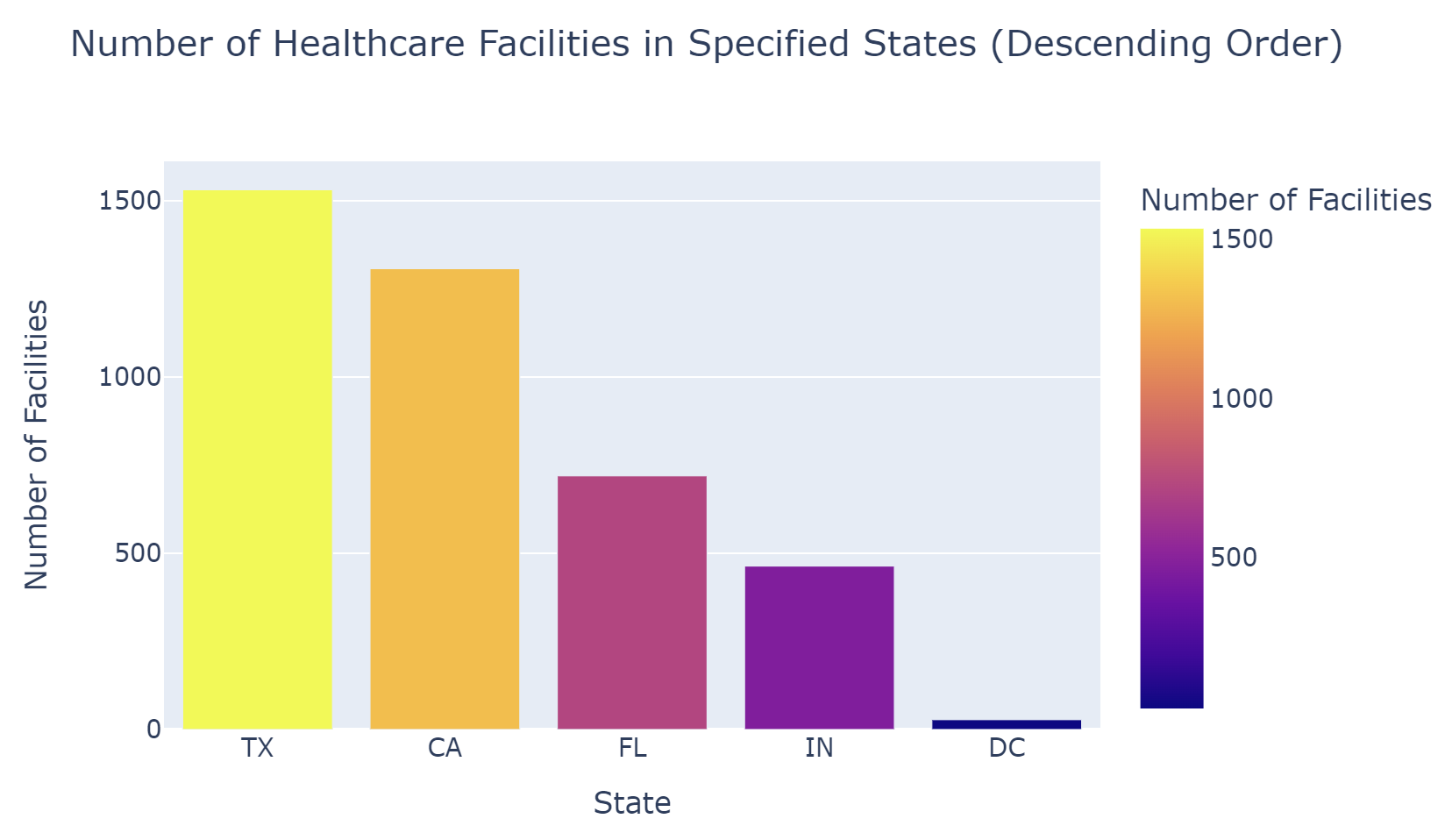
**Healthcare Challenges**: D.C.'s hospitals face unique challenges due to the transient nature of its population and high rates of chronic diseases.

**Importance of Funding:** Federal support is crucial for maintaining critical healthcare services in the nation's capital.

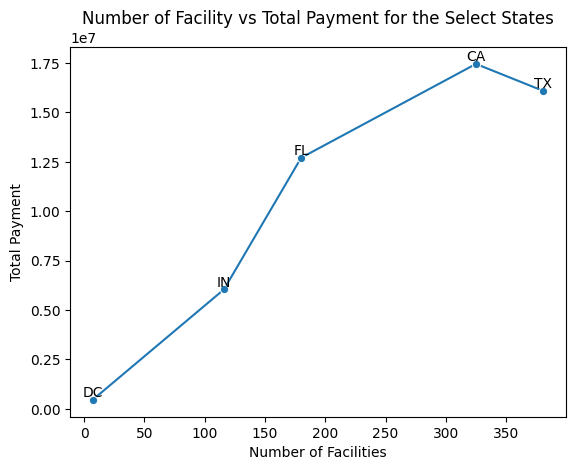


The above analysis leads to a graph plot as shown with California leading the states in funding, while Indiana lands 4th in this comparative study.

Analyzing the number of facilities for each state for supporting the direct relation number of facilities and Payment

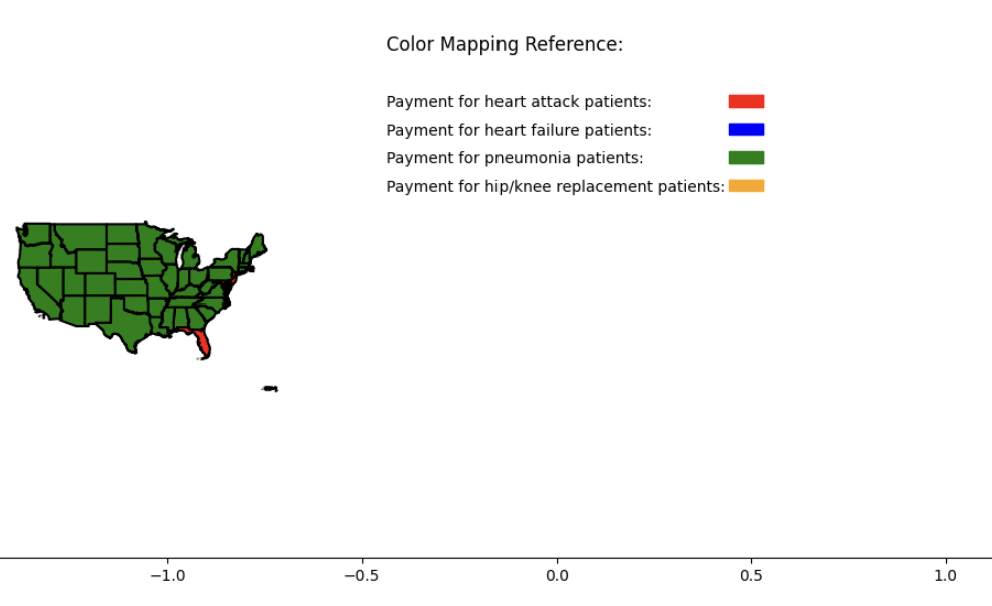


A comparative analysis of both Payment and Number of Facilities among selected states is shown below.

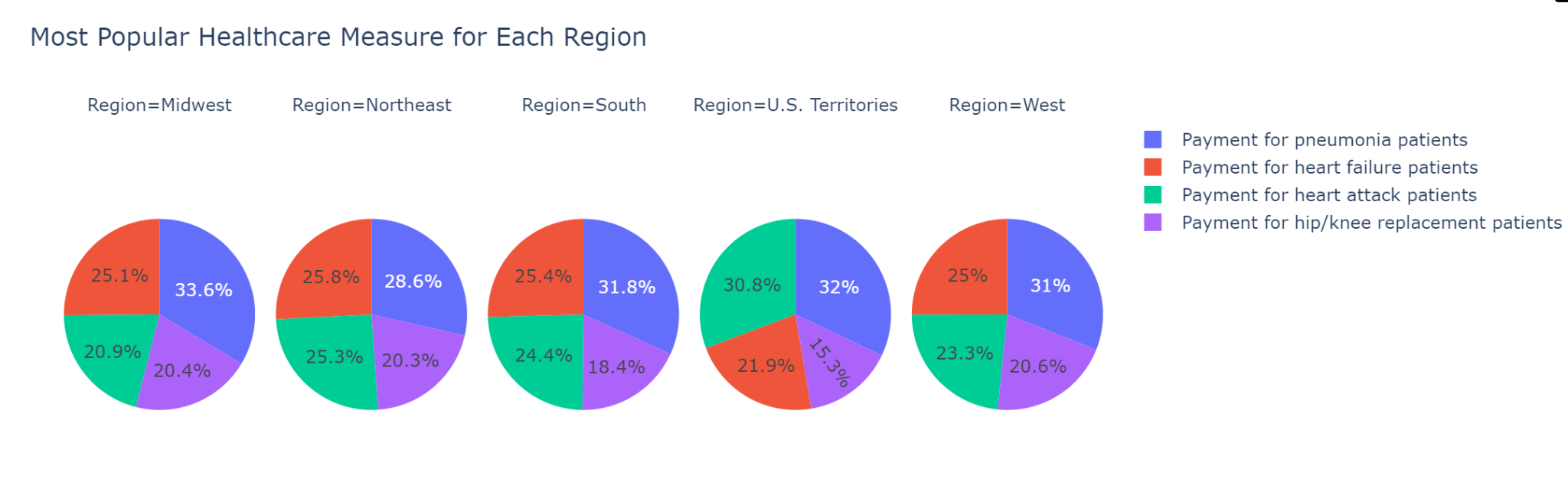


#### Payment Analysis: Disease-wise

Based on observations, it was found that pneumonia is the most funded disease in all states. This means that large number of the population in the US have suffered from this disease.The graph below shows this.

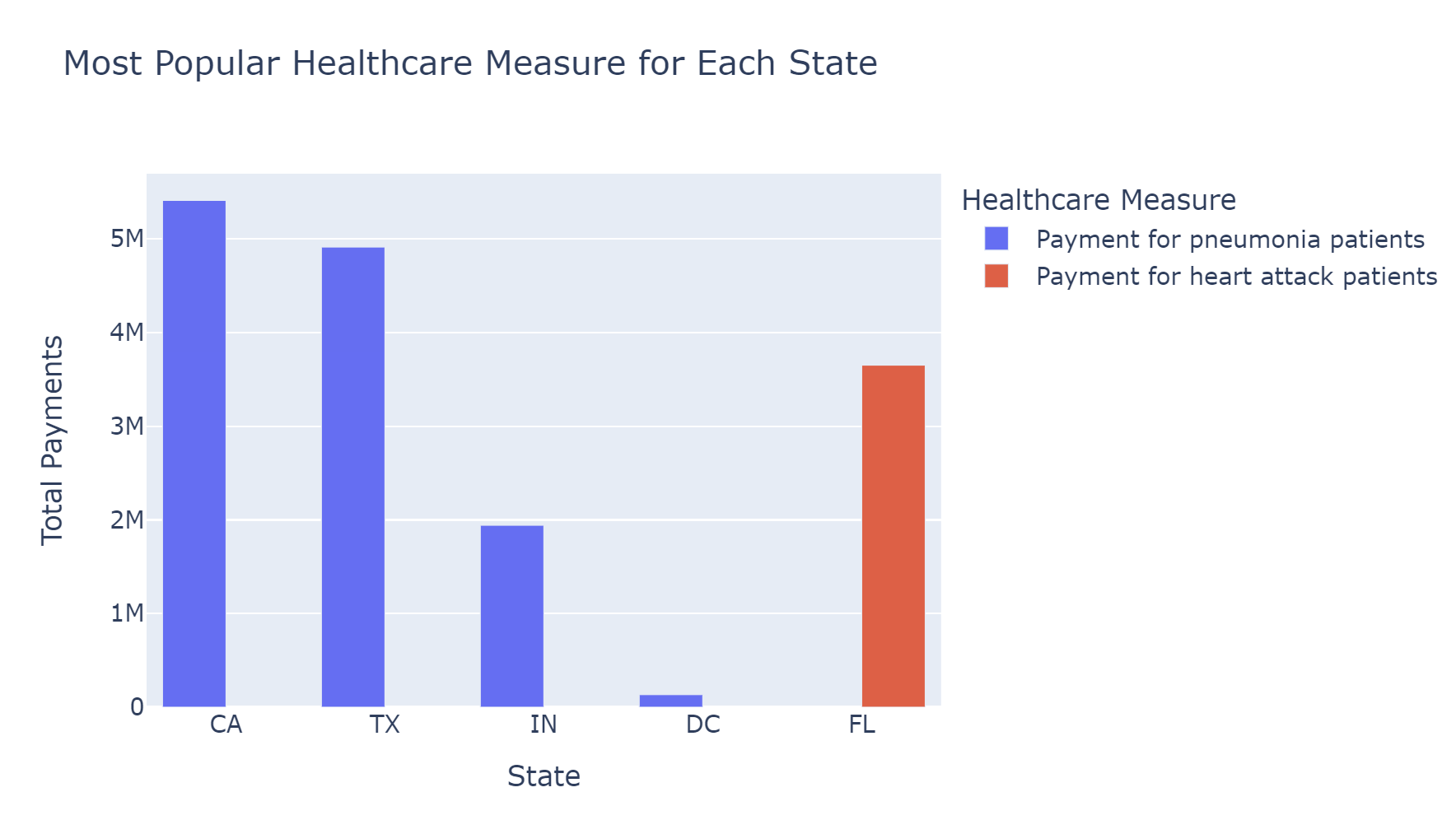


##### Payment Analysis: Disease-wise by Region



From the graph, it shows that in each region, pneumonia is the highest funded disease and hip/knee replacement is the lowest. Heart failure came second in the Midwest, Northeast, South and the West but came third in US Territories. Heart attack came thirdcin all regions except in US Territories.

##### Payment: Disease-wise by 5 Selected States



Further analysis carried out on the 5 selected states shows that pneumonia is the disease with the highest funding for all 4 states except in Florida which had Heart Attack as the most funded. It can be assumed that either there are better hospitals that cater to Heart attack patients or there are more patients with heart attack in the state than patients with other diseases. Another reason could be that it is very expensive to treat heart attack cases in that state.

### Hospital Consumer Assessment Analysis

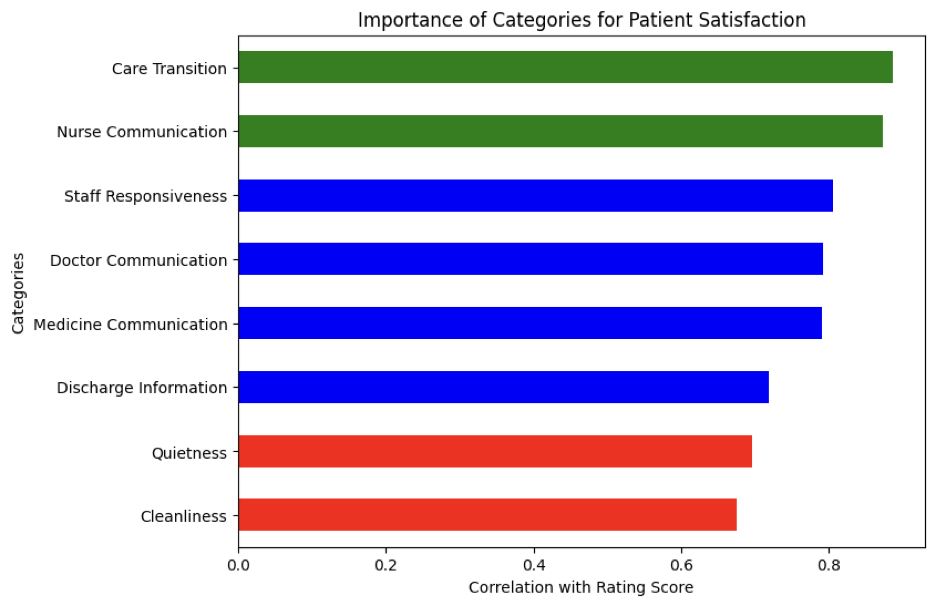
The hospital consumer assessment analysis was based on a 22-question patient survey. Those questions fall into one of the following categories:

* Nurse communication: How well nurses treated patients, listened to them, and explained things during their care.
* Doctor communication: How well doctors treated patients, listened to them, and explained things during their care.
* Staff responsiveness: How promptly patients received help when they requested it.
* Communication about Medicines: How well patients were informed about the purpose and potential side effects of medications.
* Discharge Information: How well patients were informed about symptoms to watch for after leaving the hospital.
* Care Transition: How well patients were informed about managing their health after leaving the hospital, including when to take medications.
* Hospital Cleanliness: How often were the room and areas clean?
* Hospital Quietness: How often were the room and areas quiet?

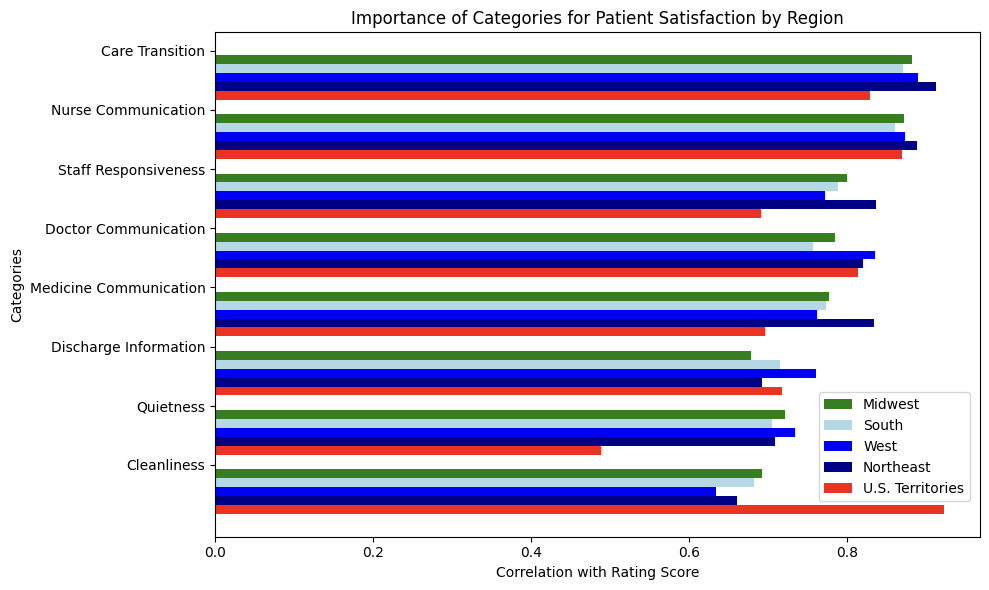
#### Analysis by Category

The first analysis was made to identify the most important categories determining the overall score rating. Analyzing each category's impact on determining a high overall score is crucial for directing efforts towards improving patient satisfaction.

By doing a general analysis of all the eight factors for Patient Satisfaction, we could tell that Care Transition and Nurse Communication are the two factors that ranked the highest in all the regions. Whereas, Quietness and Cleanliness are ranked the least.



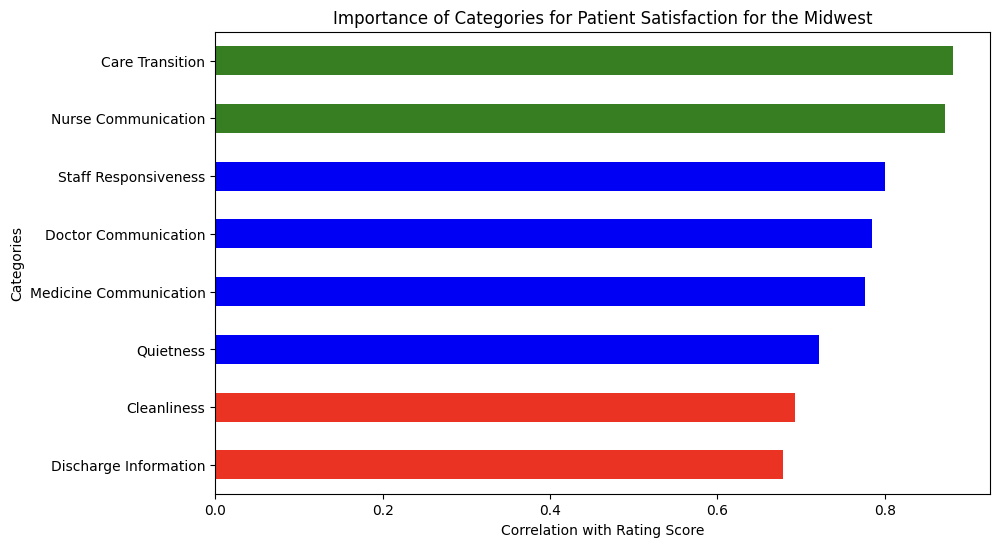
After finding this, an analysis was made to identify if the same factors are important in each of the U.S. regions.



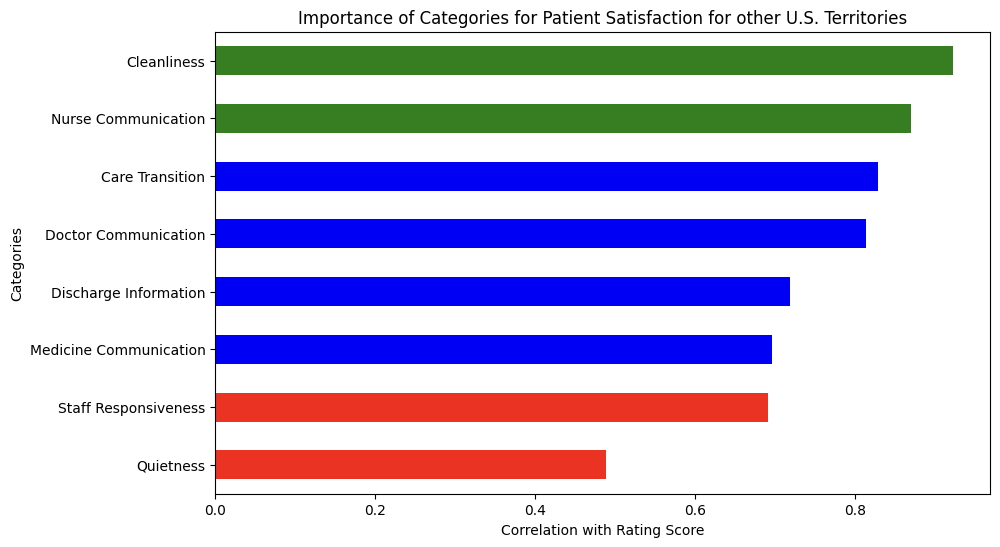
This visualization gives a lot of information, although it might seem like a lot to take in at once.It helps to identify if there's anything unique about each region in the U.S. There are some identifiable trends, for example, the Northeast seems to prioritize good communication about medicines and prompt staff responsiveness more than other regions. One interesting thing is that U.S. territories seem to place a higher emphasis on cleanliness compared to the states.

##### Comparison of Categories of Regions

In the Midwest, discharge information doesn't weigh as heavily on the overall Rating Score compared to other regions nationwide, as we can see that it is the least important category. All the other categories are very similar, maintaining Care Transition and Nurse communication as the top two important categories.



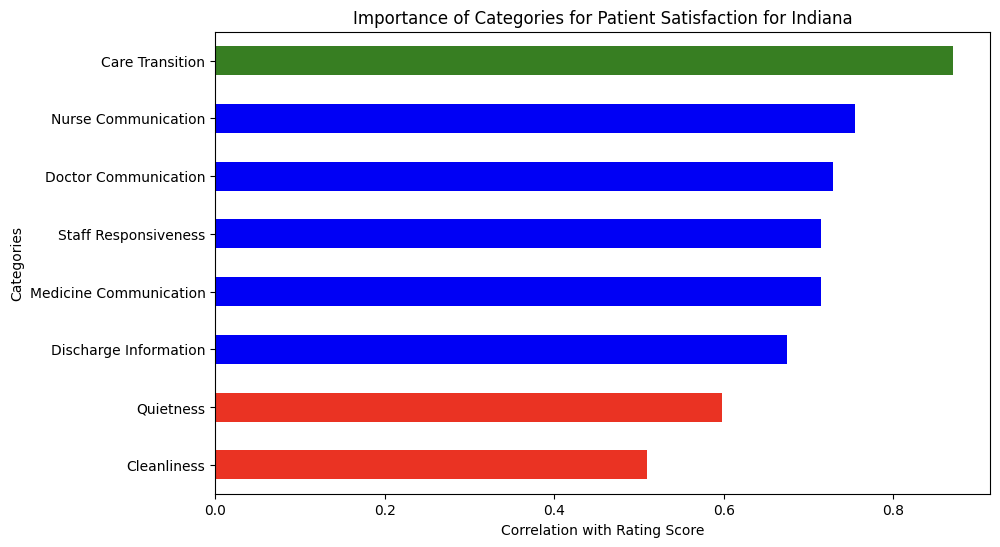
In contrast to other regions where cleanliness ranks lower in importance, other U.S. territories highlight its significance in patient satisfaction, standing out as a significant factor. However, quietness seems to have less impact in these territories compared to other regions, with a correlation of less than 0.5 with the Overall Rating Score.



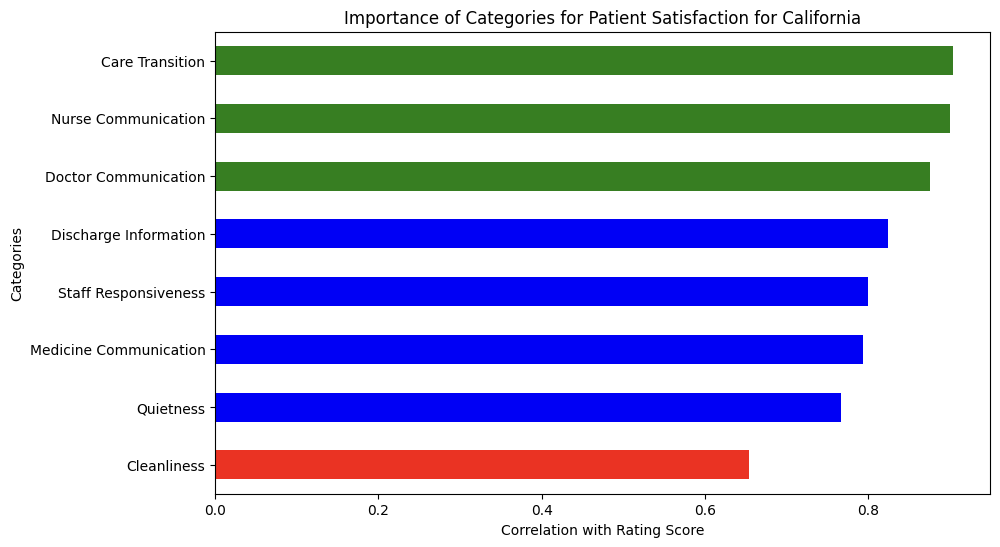
##### Analysis by State

A closer examination of category importance was conducted, focusing on specific states.

It was found that Indiana places considerably higher importance on Care Transition compared to other categories. Moreover, the remaining categories align with national trends, with cleanliness ranking as the least important factor.

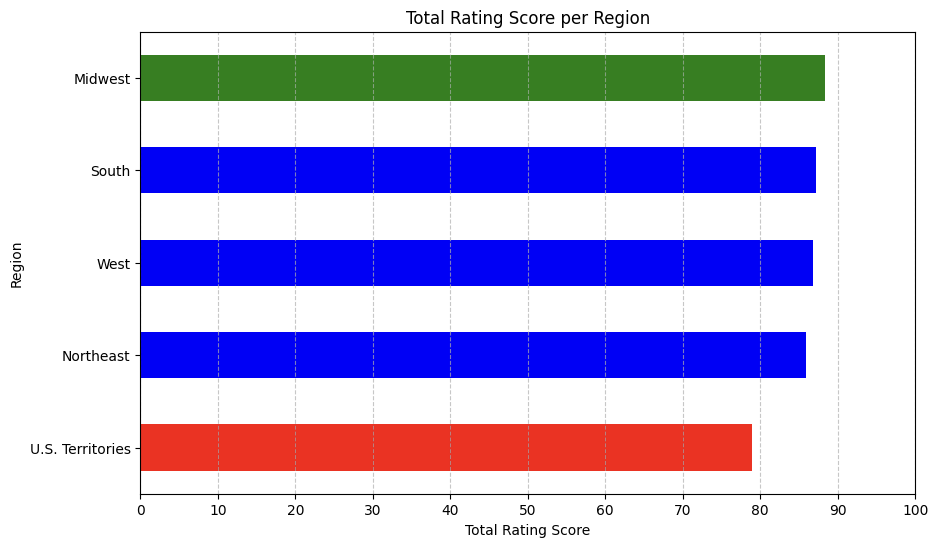


Upon analyzing California, it's evident that Care Transition and Nurse Communication remain the most critical categories. Interestingly, Doctor Communication also emerges as a significant factor. Once more, Cleanliness appears at the bottom of the list in terms of importance.



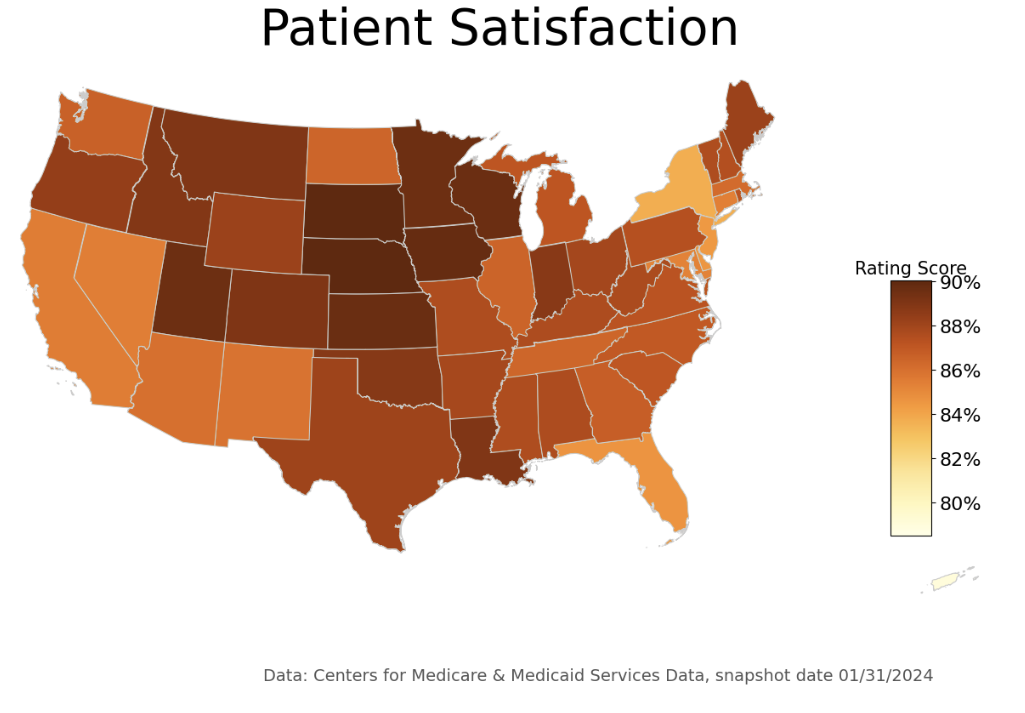
#### Analysis of Overall Rating Score

Analyzing the Overall Rating Score provides valuable insights into regional trends. Generally, the four main regions in the U.S. exhibit similar scores, with the Midwest showing slightly higher ratings for overall patient satisfaction. Conversely, Other U.S. Territories notably record the lowest overall rating scores.



##### Overall Rating Comparison by State

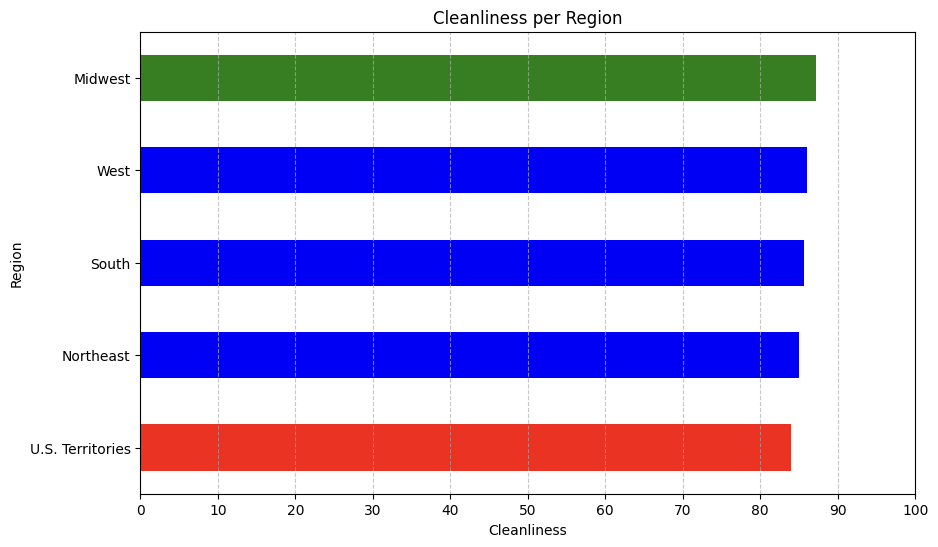
To gain deeper insights into Overall Rating Scores and Patient Satisfaction, a state-level analysis was conducted and visualized using a U.S. heat map. Several states in the Midwest region stand out with the highest Rating Scores. States such as North Dakota, South Dakota, Iowa, Minnesota, and Wisconsin, all situated on the western side of the Midwest, demonstrate particularly high scores. Additionally, within the Midwest but on its eastern side, Indiana shows an above-average overall rating score.



In the Southern region, states like Texas, Louisiana, and Oklahoma show slightly higher-than-average Rating Scores. However, compared to other significant states, California and Florida show lower-than-average Rating Scores. Notably, many states in the Northeast region display some of the lowest Rating Scores across the U.S.

##### Cleanliness Score Comparison by Region

Given that cleanliness was one of the attributes receiving lower ratings, a thorough examination of this category was conducted by region to extract insights. The Midwest region ranked highest compared to others. Interestingly, while it was anticipated that other U.S. territories might have a lower score on cleanliness, the actual scores were not as low as anticipated.

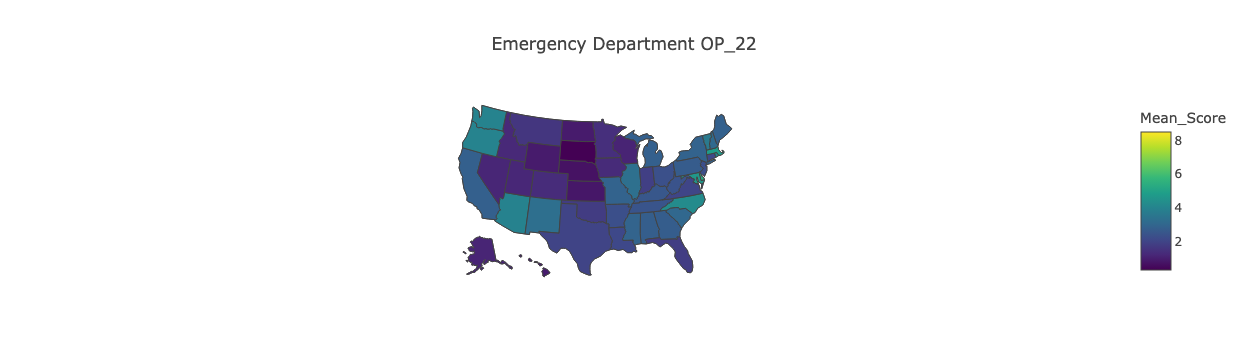


### Timely and Effective Care Analysis

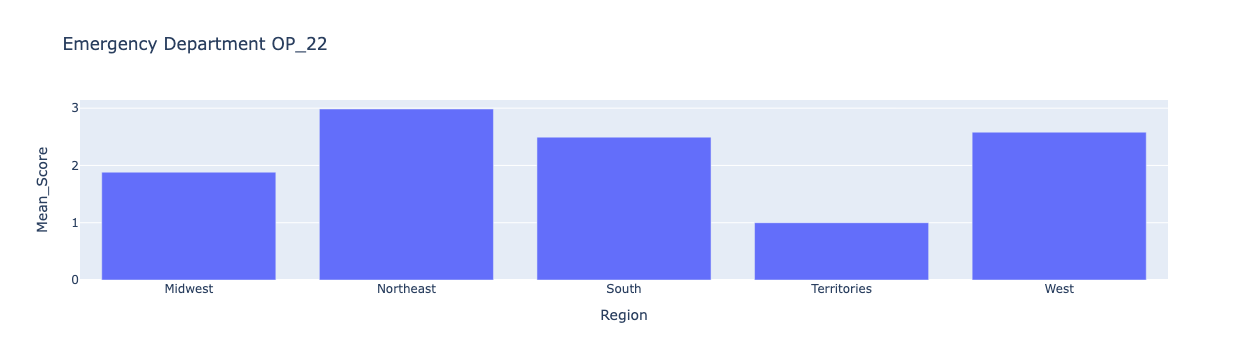
In order to measure the timeliness and effectiveness of healthcare treatment, the data reports a multitude of factors, such as Emergency Department, Healthcare Personnel Vaccination, Colonoscopy care, 'Electronic Clinical Quality Measure, 'Sepsis Care, etc. Due to the sheer amount of data, our team will focus on a few measures:

* Emergency - 1: Percentage of Leaving Emergence Without Being Seen
* Emergency - 2: Number of Minutes Patients Spent in the ER before being sent home.
* Personnel Vaccination: Vaccination Coverage among Personnel
* Stroke: Percentage of Patients Receiving brain scan results within 45 minutes of arrival
* Heart Attack: Number of Minutes Before outpatients with chest pain or possible heart attack who needed specialized care were transferred to another hospital

#### Emergency - 1: Percentage of Leaving Emergence Without Being Seen

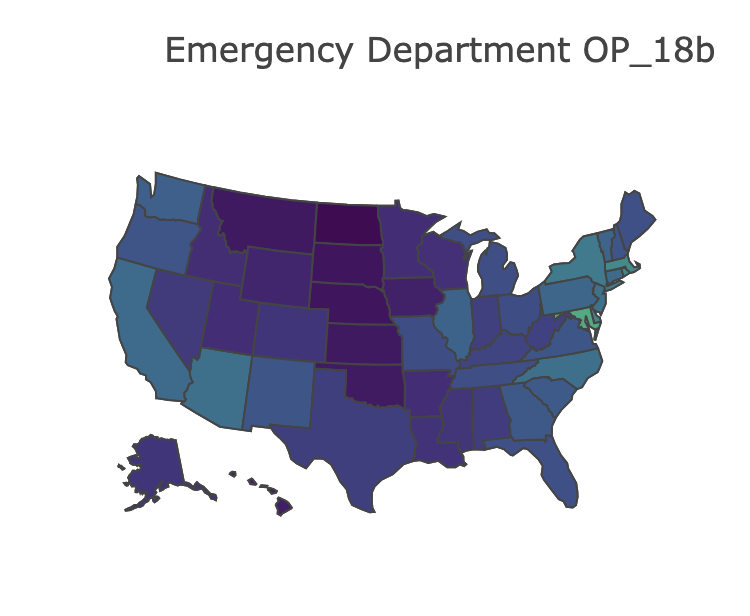
The measure ID is OP-22. As it records the percentage of patients not seen in the emergency department, a low value would be desirable. Based on the following graph, we saw that the darkest area is around the midwest region. 

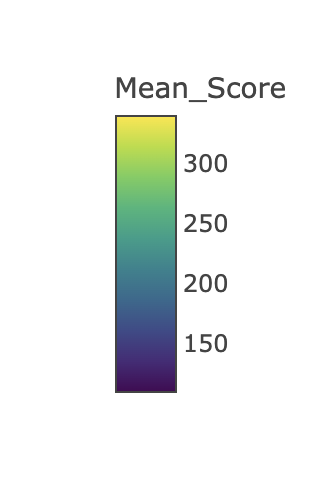
The following graph offers the same information in more detail. One notable point is that territories have much lower value compared to the Midwest region.

Then, it is broken down into the state level. It is seen that while most states are consistent at around 2-3%, Washington D.C has more than 8% of patients unseen during their emergency visits. This is an alarming note. 

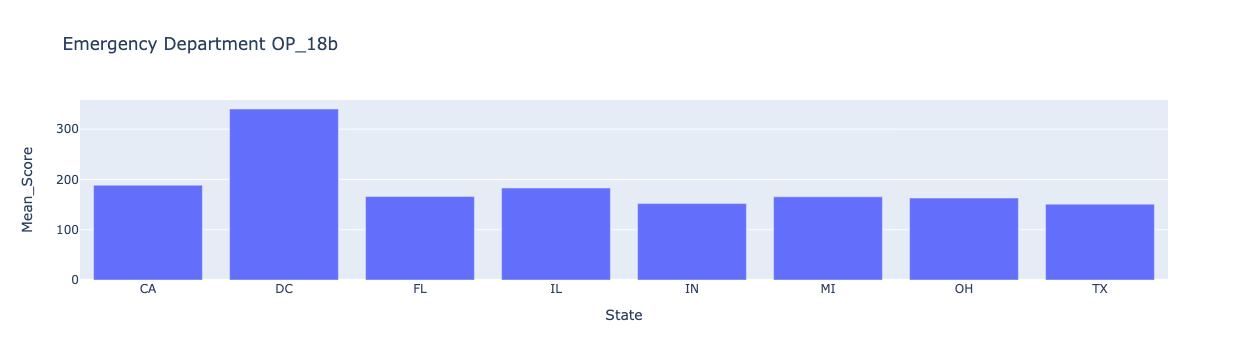
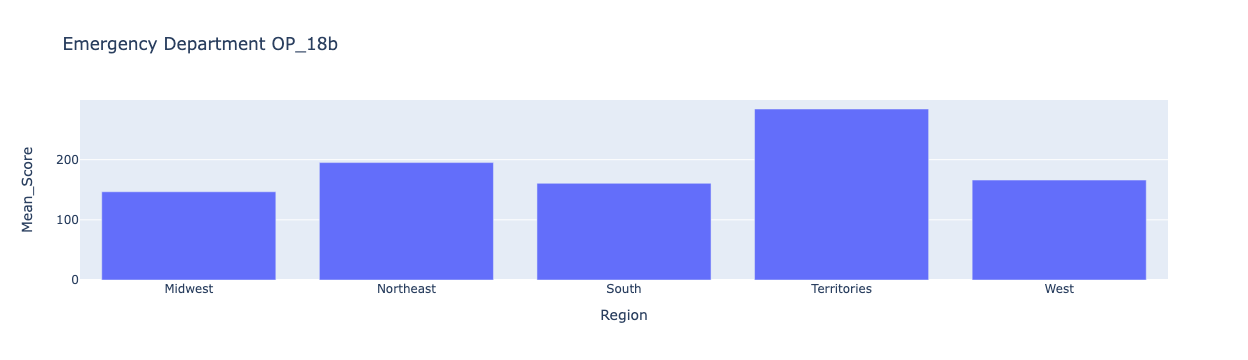
#### 

#### Emergency - 2: Number of Minutes Patients Spent in the ER before being sent home.





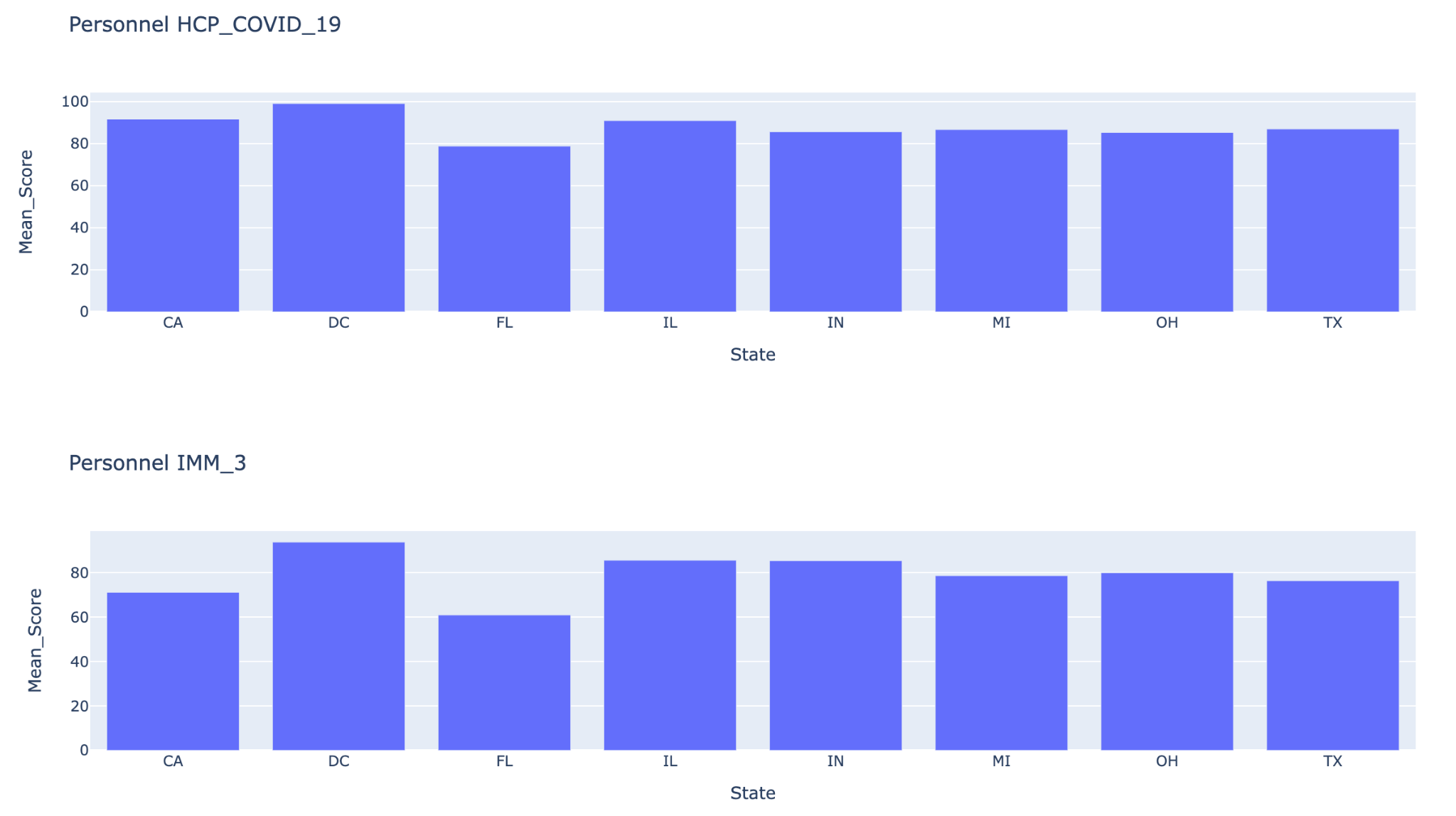
The measure ID is OP-18b or OP-18. It measures the average (median) time patients spent in the emergency department before leaving from the visit. Unlike the previous categories, patients on average spent much more time in the emergency department in U.S territories according to the following two graphs. Moreover, it is seen that D.C not only has the most patients unseen rate, but also a long time spent in the emergency department.



#### Personnel Vaccination: Vaccination Coverage among Personnel

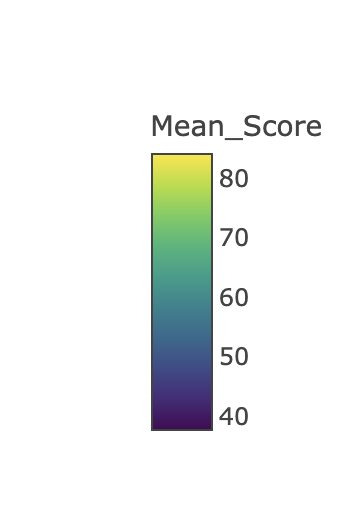
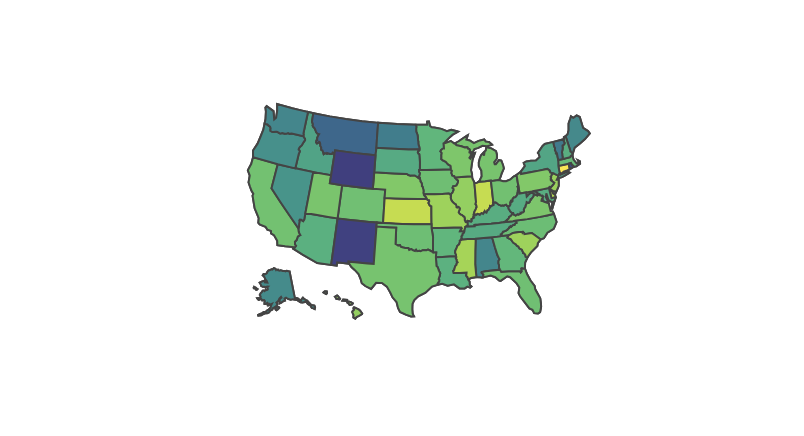
The measure ID is HCP\_COVID\_19 and IMM\_3, which detail the vaccination rate in facilities for COVID-19 and Influenza. Overall, most regions share similar rates, but it should be noted that Covid-19 has much higher coverage than Influenza in general.



The story remains the same when data is broken down into state levels. States have good coverage of COVID-19, with Florida at the lowest of 80%. The influenza vaccination rate, on the other hand, mainly stays around 80% if not less. 

#### 

#### Stroke: Percentage of Patients Receiving brain scan results within 45 minutes of arrival



#### 

For this category, a light-colored area would suggest a higher rate of patients receiving CT brain scans within 45 minutes of arrival given stroke symptoms. As depicted, most states have a high rate, with Kansas achieving over 80% percentage. In contrast, New Mexico and Wyoming rank last around 40%.



#### Heart Attack: Number of Minutes Before outpatients with chest pain or possible heart attack who needed specialized care were transferred to another hospital

Based on the following graphs, there are no significant differences between regions, except for U.S. territories. In terms of individual states, California stands out over 100 minutes to get outpatients to their specialized facility.

#### 

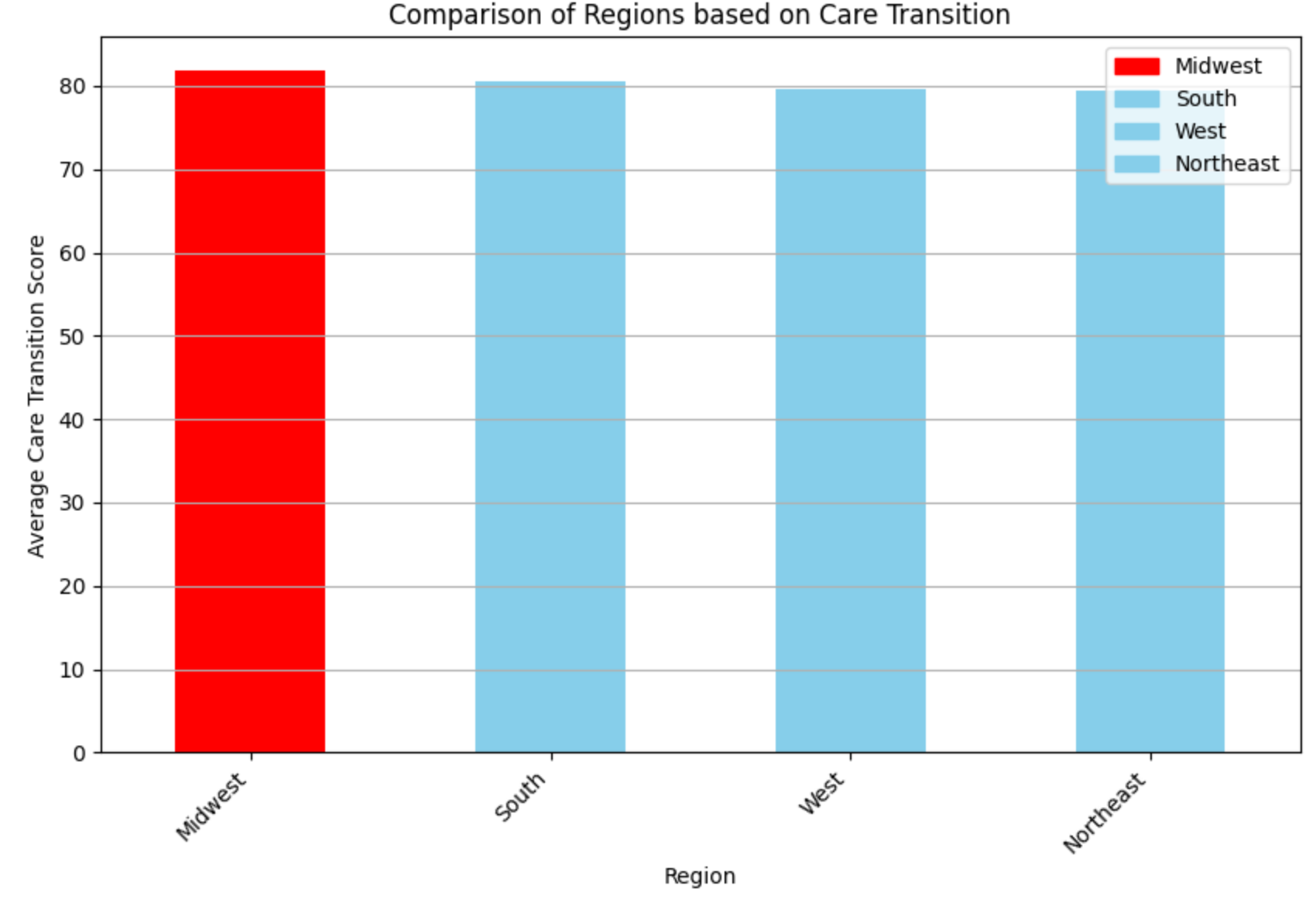
#### Conclusion

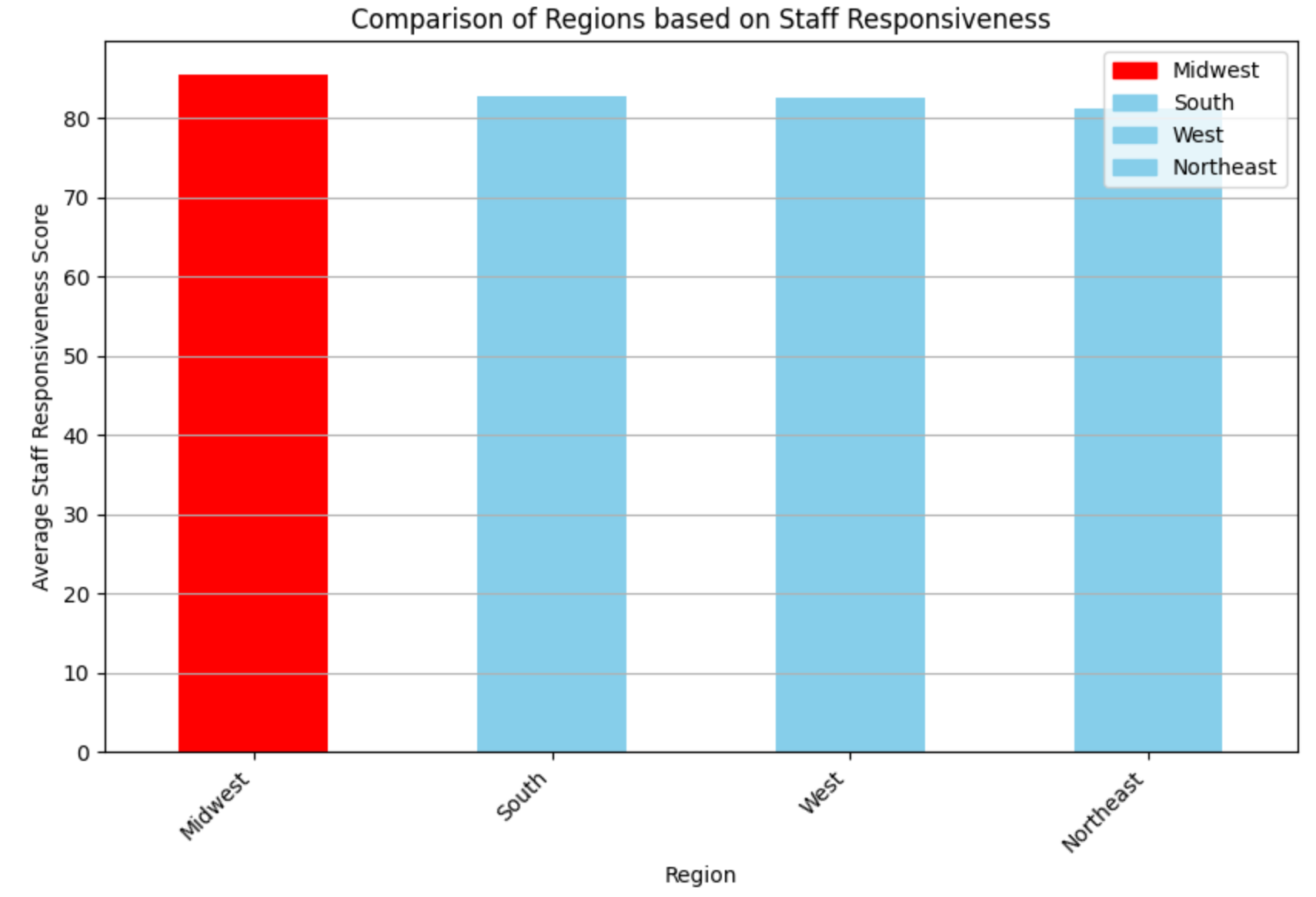
It is safely concluded that Midwest is the best in terms of timeliness and effectiveness as it has the lowest number of patients left unseen and time spent in the department, good vaccination coverage among personnel, and brain scan rate. Another notable note is that U.S. territories perform worse than U.S. states. Moreover, D.C has one of the worst emergency services for public healthcare.

## Joint Analysis

|  | PAYMENT | NO. OF FACILITIES | HOSPITAL RATING | STAFF RESPONSIVENESS | CARE  TRANSITION | EMERGENCY | VACCINATION |
| --- | --- | --- | --- | --- | --- | --- | --- |
| MIDWEST | 56,790,057 | 1,342 | 88.4 | 1st | 1st | 1.88% | 87% |
| SOUTH | 81,970,420 | 1,697 | 87.16 | 2nd | 2nd | 2.49% | 85% |
| WEST | 40,962,159 | 925 | 86.78 | 3rd | 3rd | 2.58% | 90% |
| NORTHEAST | 34,143,874 | 552 | 85.85 | 4th | 4th | 2.98% | 94% |

Through the joint analysis, we can observe that the Midwest ranks the highest in hospital rating, Staff Responsiveness, Care Transition and Emergency. South ranks first in Payment and No. of Facilities and ranks second overall. Whereas, Northeast ranks the last overall.

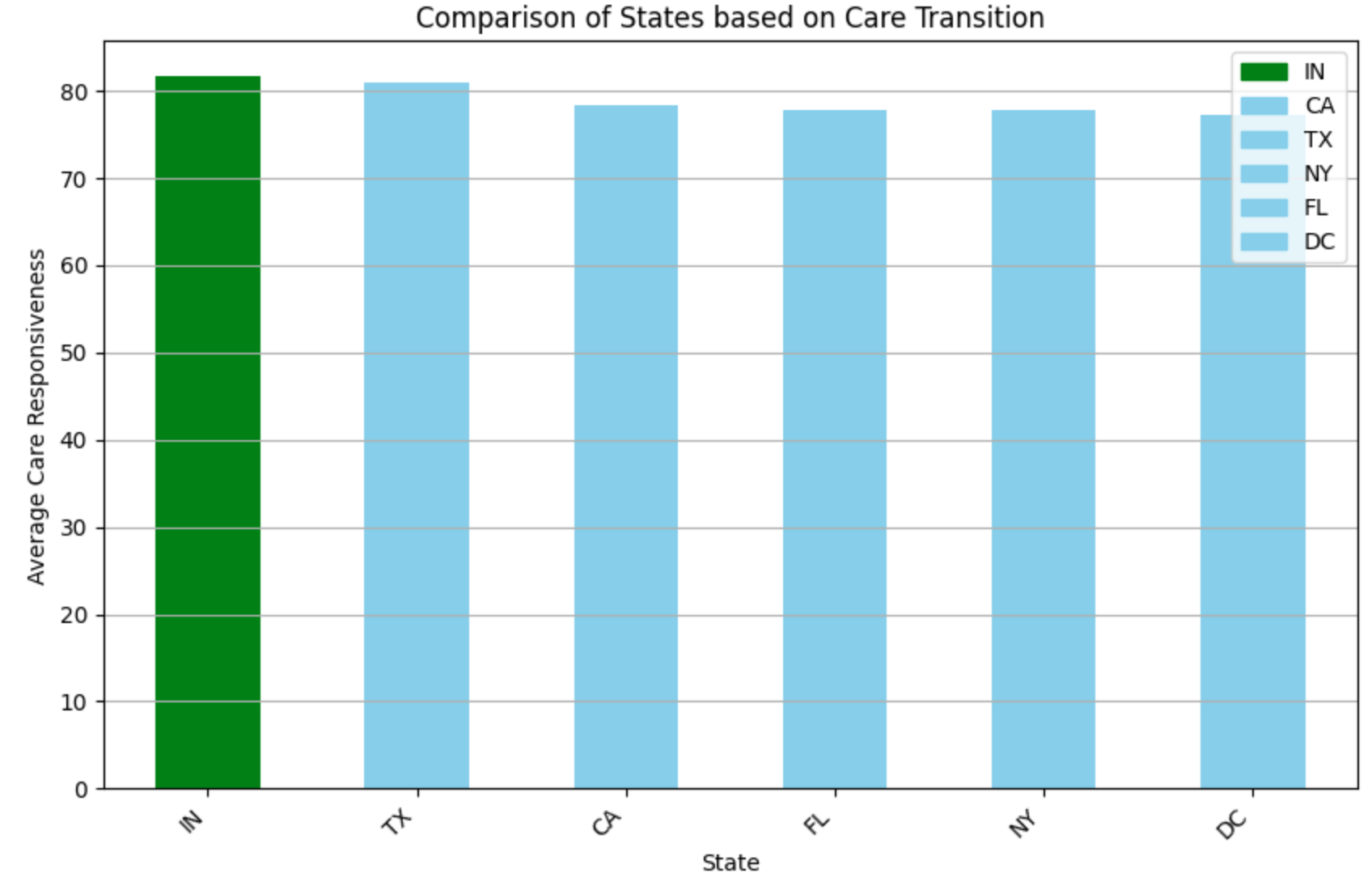
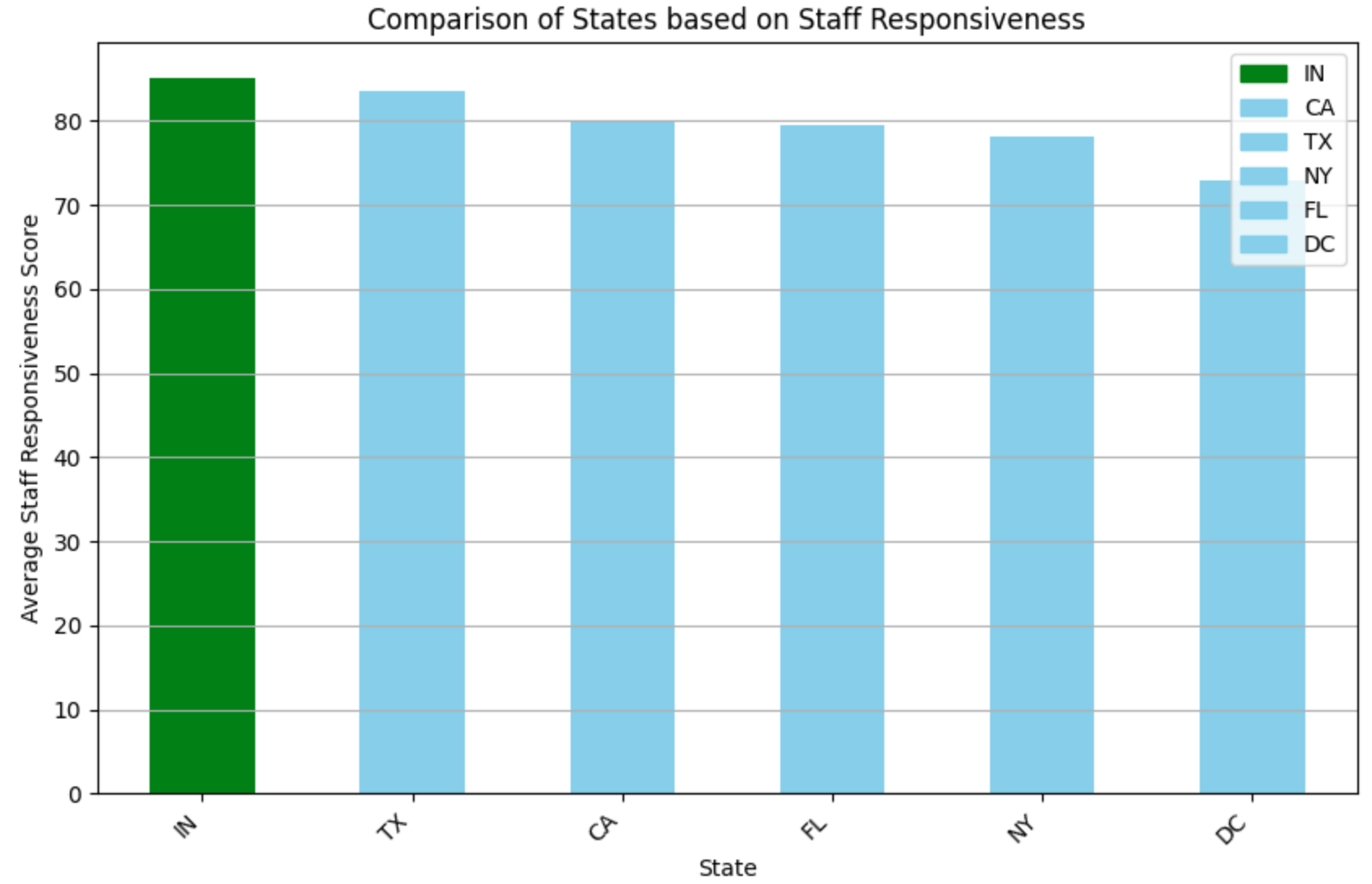




Through the above graphs, we can understand that Midwest ranks first In Staff Responsiveness and Care Transition. Further, the Northeast ranks the lowest among the other regions.

|  | PAYMENT | NO. OF FACILITIES | STAFF RESPONSIVENESS | CARE  TRANSITION | EMERGENCY | VACCINATION |
| --- | --- | --- | --- | --- | --- | --- |
| IN | 6,041,647 | 116 | 1st | 1st | 1.88% | 85.65% |
| CA | 17,450,375 | 319 | 3rd | 3rd | 2.83% | 91.64% |
| TX | 16,080,207 | 358 | 2nd | 2nd | 2.58% | 90% |
| DC | 448,774 | 6 | 5th | 5th | 2.98% | 94% |
| FL | 12,702,206 | 177 | 4th | 4th | 2.98% | 94% |

By doing the joint analysis state-wise we concluded that Indiana ranks the highest in Staff Responsiveness, Care Transition and Emergency. California ranks highest in Payments and DC has the overall lowest scores in all the categories.



By doing the state-wise analysis for Care Transition and Staff Responsiveness it is evident that Indiana ranks first in both of them along with Emergency which makes them the categories that are prioritized the most.

The project focuses on descriptive analysis, which makes it difficult to run an evaluation. However, in the project plan, it is stated that there would be clustering models, so our team can divide the original dataset into training and testing datasets to test the clustering model and its accuracy. This could be used to compare with another dataset that includes a systemic rating, which is called the Overall Star Ratings. They are designed to assist patients, consumers, and others in comparing hospitals side by side. The hospitals can receive between one and five stars, with five stars being the highest rating, and the more stars, the better the hospital performs on the quality measures. Most hospitals will display a three-star rating. For more information, go to the hospital

# Results and Evaluation Plan

Results from Joint Analysis

# Completed Tasks

* Data Cleaning and Transformation to have a more meaningful dataset to analyze and generate visualizations based on the Hospital Patient Surveys Dataset.
* Data Collection and Visualization for the state, region, and division dataset, to analyze datasets grouped by regions.
* Data Analysis and Visualization on features correlation among the categories and overall score rating.
* In-depth analysis of region and state overall score and by categories to show what stands out.

# Challenges & Future Work

There are many challenges our team encountered, which are as follows:

* Datasets were broad with many different tables and unfamiliar terminologies.
* In terms of discoveries, many of the findings can be similar or consistent, which makes it difficult to make a good story or distinguish them.
* Merging tables is difficult due to different features and formats. For example, table A has 2 columns: measurement and score, and a facility in the dataset can have different records of measurement. This is the main issue when trying to merge because each table will have different measurements, so table A can have 9 records of facility 001 but table B can have 4 records of facility 001. To avoid this problem, information has to be flattened to only 1 record. However, this would create tables with extremely high amounts of columns.
* There is data from different years on the CMS data source. However, due to the collection period, data can remain and be updated over time. This is an issue when merging tables of the same topic from different periods.
* The scope of work was changed from ranking hospitals to descriptive analysis to be more appropriate for the audience.

Based on the challenges, our team has a list of future works:

* To see trends, the team can attempt to:
  + For the same year, our team can merge tables from different topics. Then, analysis can be performed for each year.
  + Data for the same topic is merged over a collection of years. Then, analysis can be performed for each topic.
* In addition to the current analysis, other factors, such as the cost of living, and public education can be added to create a more comprehensive understanding of people under public health insurance.
* This work has not yet been evaluated properly. Thus, in the future, it can be compared to other research and findings.

# Learning Outcome & Recommendation

What we learn:

* Get a better understanding of how payments from the government through Medicare and Medicaid are distributed to various states and regions.
* Handling large volumes of data, even from reputable sources like the CMS, necessitated extensive data wrangling. This involved tasks such as data cleaning and transformation.
* Making graphs and charts easy to read for the audience is very important. It should be easy to understand without even having to explain much.

For future teams who decide to work on healthcare analysis, there are a few important notes:

* Healthcare datasets are not often available to the public with CMS being the only exception. They will require permission from institutions.
* When working with healthcare, it is expected that the audience has a basic level of understanding of health and disease terms.

# Citations

1. Keisler-Starkey, K., Bunch, L. N., & Lindstrom, R. A. (2023, September). Health insurance coverage in the United States: 2022. https://www.census.gov/content/dam/Census/library/publications/2023/demo/p60-281.pdf
2. Shacham, Omer. “USA States to Region.” *Kaggle*, 12 June 2018, www.kaggle.com/datasets/omer2040/usa-states-to-region.