Anastasia Erofeeva

Graduate Project

Brief description of problem:

In a perfect world, every hospital would provide the best quality of care, with the best possible outcomes, for the lowest price. However, this is not always the case. The "Payment and value of care – Hospital" dataset from Data.Medicare.org shows that, for the same treatments, there is variation in quality of care and hospital charges between hospitals in the country. There is even variation between hospitals in the same state or the same city. This dataset is publicly available, but it has 19K rows and 23 columns, so, just by looking at the raw data, it is difficult to see the whole picture and make recommendations. A summary of the data, including data visualizations, would help people better understand the variation in hospitals around the country, as well as within their states, in order to make informed decisions.

Libraries:

- **Pandas:** I used the Pandas library to convert the csv file from Data.Medicare.org into a DataFrame, which allowed me to split the data by type of illness and value of care category. I was then able to iterate over these subsets of the data to identify the best hospitals in the nation, the best hospitals in each state, the states with one or more top hospitals (better outcomes and lower prices), and the price distributions for treatment in different hospitals, for each of the four categories of illness in the dataset (heart attack, heart failure, hip/knee replacement, pneumonia).
- **Matplotlib:** I used the Matplotlib library to make box plots, which provided a visualization of Outcome vs Price for each category of illness.
- **Plotly:** I used the Plotly library to make choropleth maps, which showed the states with one or more top hospitals (better outcomes and lower prices) for each category of illness.

Something interesting I learned:

I am interested in health care, data science, and finance, so I really enjoyed working on this project. It was great to get some experience working with a relatively big dataset, as well as using common data science libraries. In addition, this project allowed me to practice going through the data science process of obtaining the data, cleaning the data, exploring the data, making models/visualizations of the data, and interpreting the results. Something interesting I learned from visualizing the data, is that higher treatment prices do not necessarily mean better outcomes. I supposed that would be the case, but seeing the box plots of the price distributions for each outcome type (better, average, worse) was still somewhat surprising. The box plots for hip/knee replacement were especially interesting, because they showed that most hospitals with better than average outcomes also had lower than average prices.

Limitations:

- Unfortunately, many values in the dataset were 'Not Available', so the hospitals with those values could not be included in the results.
- Some states/territories did not have any hospitals with better outcomes and lower payment. In those cases, I compared the outcome category (better, average, or worse) and then the payment (lower, average, or higher) to determine the 'best' hospital for that state, using the

ranking order shown below. A limitation of this method, however, is that with the data provided in the dataset, there is no way to tell which 'better' is the best 'better', which 'average' is the best 'average', etc. It would be very helpful to have more specific information on the mortality and complications, other than just general categories. Here is the order I used to rank hospitals:

1. Better complications and lower payment (BEST)
2. Better complications and average payment
3. Better complications and higher payment
4. Average complications and lower payment
5. Average complications and average payment
6. Average complications and higher payment
7. Worse complications and lower payment
8. Worse complications and average payment
9. Worse complications and higher payment (WORST)

- If I had more time, I would also make the following improvements to the plots:
    - Show details about the best hospitals, not just the number of hospitals, when hovering over states in the choropleth map.
    - Change the colorbar tick marks for the choropleth maps to integers instead of floats, because it does not make sense to have 1.5 hospitals.
    - Color significant outliers in the box plots green or red, depending on if they represent hospitals that provide great value of care for a low price, or poor value of care for an unusually high price.

Examples of output:

This page shows screenshots of the output for one of the four categories of illness - hip/knee replacement.

Note: In order to fit everything on one page, the screenshots of the table and list only show a few rows of the entire output.

Data Summary for Hip/Knee Replacement

Best Hospitals in the Nation

	Facility Name	City	State	Lower estimate	Higher estimate	Value of care category
0	WASHINGTON REGIONAL MEDICAL CENTER	FAYETTEVILLE	AR	$18,527	$19,486	Better Complications and Lower Payment
1	ST LUKE'S MEDICAL CENTER	PHOENIX	AZ	$17,968	$18,762	Better Complications and Lower Payment
2	JOHN MUIR MEDICAL CENTER - CONCORD CAMPUS	CONCORD	CA	$18,747	$19,767	Better Complications and Lower Payment
3	SEQUOIA HOSPITAL	REDWOOD CITY	CA	$20,083	$21,122	Better Complications and Lower Payment
4	PROVIDENCE SAINT JOHN'S HEALTH CENTER	SANTA MONICA	CA	$18,339	$18,892	Better Complications and Lower Payment
5	HOAG ORTHOPEDIC INSTITUTE	IRVINE	CA	$19,275	$19,821	Better Complications and Lower Payment
6	ADVENTIST HEALTH ST HELENA	SAINT HELENA	CA	$16,334	$16,923	Better Complications and Lower Payment
7	MERCY MEDICAL CENTER REDDING	REDDING	CA	$18,902	$19,864	Better Complications and Lower Payment
8	CEDARS-SINAI MEDICAL CENTER	LOS ANGELES	CA	$19,977	$20,691	Better Complications and Lower Payment
9	WASHINGTON HOSPITAL	FREMONT	CA	$16,752	$17,246	Better Complications and Lower Payment
10	EISENHOWER MEDICAL CENTER	RANCHO MIRAGE	CA	$19,276	$20,007	Better Complications and Lower Payment
11	ST FRANCIS HOSPITAL & MEDICAL CENTER	HARTFORD	CT	$19,569	$20,213	Better Complications and Lower Payment
12	CHRISTIANA CARE HEALTH SERVICES, INC.	NEWARK	DE	$18,091	$18,623	Better Complications and Lower Payment
13	BLAKE MEDICAL CENTER	BRADENTON	FL	$17,803	$18,723	Better Complications and Lower Payment
14	PHYSICIANS REGIONAL MEDICAL CENTER - PINE RIDGE	NAPLES	FL	$18,390	$19,101	Better Complications and Lower Payment

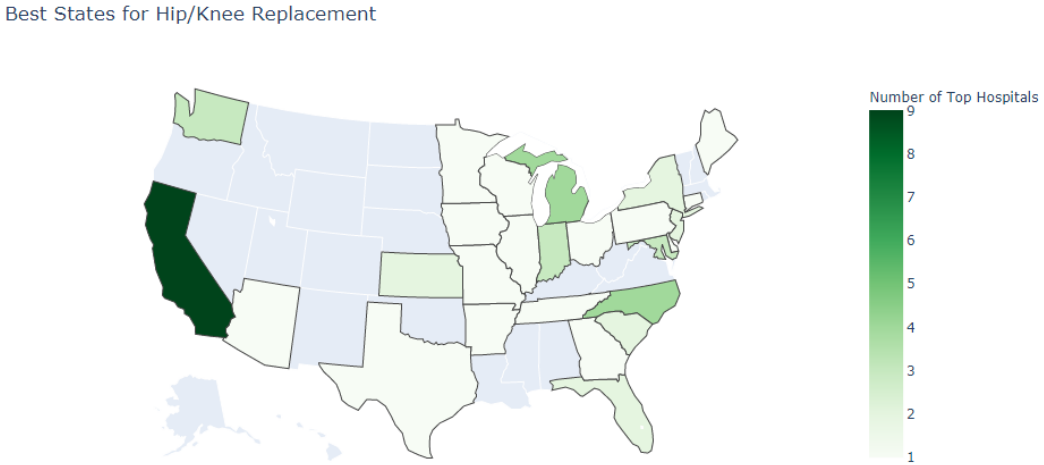
...

Best Hospitals in Each State/Territory

```
AK: ALASKA NATIVE MEDICAL CENTER - Anchorage
    Average Complications and Lower Payment ($16274)

AL: THOMAS HOSPITAL - Fairhope
    Average Complications and Lower Payment ($19330)

AR: WASHINGTON REGIONAL MEDICAL CENTER - Fayetteville
    Better Complications and Lower Payment ($19009)

AZ: ST LUKE'S MEDICAL CENTER - Phoenix
    Better Complications and Lower Payment ($18357)

CA: ADVENTIST HEALTH ST HELENA - Saint Helena
    Better Complications and Lower Payment ($16622)

CO: POUDRE VALLEY HOSPITAL - Fort Collins
    Average Complications and Lower Payment ($17274)

CT: ST FRANCIS HOSPITAL & MEDICAL CENTER - Hartford
    Better Complications and Lower Payment ($19890)

DC: SIBLEY MEMORIAL HOSPITAL - Washington
    Average Complications and Lower Payment ($20318)

DE: CHRISTIANA CARE HEALTH SERVICES, INC. - Newark
    Better Complications and Lower Payment ($18360)
```

...

States with the Highest Number of Above Average Hospitals



Value of Care (Complications vs Price)