**Final Summary Report Project 1:**

**December 12, 2022**

**Team Members:**

**Dylan Kurbel, Raelle Nalos, Nicolette Keplinger, and Iqra Imam**

**Intro:**

We investigated a dataset published by the U.S. Department of Health & Human Services (Healthdata.gov). Our dataset is available via API, but since it is still actively updated, we curated a .csv to keep analyses consistent at one time point. This file can be found on our repo (<https://github.com/nkeplinger/healthcare_module7/tree/main/Resources>/[COVID\_staff\_shortages.csv](https://github.com/nkeplinger/healthcare_module7/blob/main/Resources/COVID_staff_shortages.csv))

This dataset has published reported staff shortages in hospitals across the US since Jan 1 2020 and is actively publishing data daily. Data include daily accounts of hospitals in every US state that report shortages as “yes”, “no”, or “not reported”. This dataset includes a total of 153 columns containing additional hospital stats each day such as geography, covid deaths, patient occupancy and availability. These data led us to further investigate any correlations, relationships, and trends that these other statistics may have with critical staff shortages. We broke this research into 4 topic groups (one for each group member) for which we all investigated a series of hypotheses via data exploration and analyses. Our findings are summed up below. Note: All supporting and referenced figures can be found in the git repo “outputs” directory (<https://github.com/nkeplinger/healthcare_module7/tree/main/output>).

**Research Sub-Category 1: Chronological Analysis of Staff Shortages via Month, Season, and Weekday.**

1. Hypothesis 1-1:
   * Null 1.1: There was no relationship between seasons and staffing shortage.
   * Alternative 1.1: There was a relationship between the seasons and staffing shortages.
   * Result Summary: the Pearson coefficient and p-value indicate we can reject the null hypothesis. Shortages increase on average with progression of seasons from Spring to Winter, most heavily in Fall and Winter.
2. Hypothesis 1-2:
   * Null 1.2: There was no relationship between day of the year and staffing shortage.
   * Alternative 1.2: There was a relationship between the day of the year and staffing shortages.
   * Result Summary: the linear regression r-squared value indicates we can reject the null hypothesis. On average, the first and last 30 days of each year see the greatest impact of staffing shortages, with a significant increase in the last 5 months of the year.
3. Hypothesis 1-3:
   * Null 1.3: There was no relationship between day of the week and staffing shortage.
   * Alternative 1.3: There was a relationship between the day of the week and staffing shortages.
   * Result Summary: the Pearson coefficient and p-value indicate we can reject the null hypothesis. Wednesday through Friday see the greatest levels of reported staffing shortages, dwindling as the weekdays approach Tuesday which saw the lowest average shortages.

**Research Sub-Category 2: Staff Shortages Vs. Covid Deaths from 2020 to 2022**

1. Hypothesis 2-1:
   * Null Hypothesis 2-1: There is no relationship between staff shortages and covid deaths from 2020 to 2022
   * Alternative Hypothesis 2-1: Covid deaths having a significant impact on hospital staff or shortage or vice versa.
   * Data Analysis 2.1: Comparing figures 2.1 and 2.6 . There is no significant correlation in the number of covid deaths and staff shortages across hospitals in NY and TX .

Results Summary 2-1: Data figures Figure 2.3 shows a negative correlation between Covid Deaths vs Covid Deaths.

Figure 2.1. There is an outlier in the NY data for Q2 of 2022.

Figure 2.6. The TX data is spread more evenly.

**Research Sub-Category 3: Staff shortages Vs. Adult and Pediatric Patient Numbers**

1. Hypothesis 3-1:
   * Null Hypothesis 3-1: There is no relationship between staff shortages and inpatient occupancy (Adult and/or Pediatric) from 2020 to 2022 in the US
   * Alternative Hypothesis 3-1: There is a relationship between critical staff shortages and adult and/or pediatric inpatient occupancy from 2020 to 2022
   * Data Analysis 3-1: When viewing all three variables together (percent staff shortage, percent adult inpatient occupancy and percent pediatric inpatient occupancy), there were no obvious trends/relationships that stood out across the calendar months from Jan 2020 to Dec 2022 (Figure 3-1). However, when breaking them down by year and variable, there were some linear relationships among these variables. The percent inpatient occupancy of adults versus the percent staff shortages all had a positive linear relationship in years 2020 (Figure 3-1a), 2021 (Figure 3-1b), and 2022 (Figure 3-1c). In contrast, the percent occupancy of pediatric patients had no relationship in 2020 (Figure 3-1d), a slightly positive relationship in 2021 (Figure 3-1e), and a stronger negative relationship in 2022 (Figure 3-1f). This Data supports the alternative hypothesis that there is a relationship between critical staff shortages and adults and pediatric inpatient occupancy from 2020 to 2022, with the exception of pediatric patients in 2020.
   * Result Summary 3-1: These data (Figures 3-1, 3-1a-f) supports the alternative hypothesis that there is a relationship between critical staff shortages and adults and pediatric inpatient occupancy from 2020 to 2022, with the exception of pediatric patients in 2020.
2. Hypothesis 3-2:
   * Null Hypothesis 3-2: There is no relationship between staff shortages and confirmed covid cases (Adult and/or Pediatric) from 2020 to 2022 in the US
   * Alternative Hypothesis 3-2: There is a relationship between critical staff shortages and confirmed covid cases in adult and/or pediatrics from 2020 to 2022
   * Data Analysis 3-2: Similar to 3-1, in 3-2 percent staff shortages, total adult covid cases, and total pediatric covid cases were examined for the calendar year 2020-2022. However, because of the drastic variability of the y-axis variables, they were plotted on independent graphs (Figure 3-2a, 3-2b, and 3-2c, respectively). There appears to potentially be similar trends of all variables increasing around months 5-15 and again around months 25. These trends were confirmed with scatterplots examining the linear regression between covid cases and percent staff shortages. Specifically, there is a strong positive linear relationship between confirmed adult covid cases across the country and percent staff shortages from 2020-2022 (Figure 3-2d). There were also reduced, but still positive relationships between confirmed pediatric cases across the country and percent staff shortages from 2020-2022 (Figure 3-2e). This data supports the alternative hypothesis that there is a relationship between critical staff shortages and BOTH pediatric and adult confirmed covid cases in US hospitals from 2020 to 2022.
   * Result Summary 3-2: These data (Figures 3-2a-e) supports the alternative hypothesis that there is a relationship between critical staff shortages and BOTH pediatric and adult confirmed covid cases in US hospitals from 2020 to 2022.

**Research Sub-Category 4: Hospitals per State vs. Shortages and Occupancy**

While looking at the data for critical staffing shortage, we wanted to see how the geography, or the size of the states would impact the critical staffing shortages, as well as the percentage of inpatient beds used in comparison to the inpatient beds. Because the initial data set was not enough, an additional data set was merged to to include the number of hospitals in a state. To do this, we had to make sure the states were matching, and there were not any additional territories which would skew the data set.

1. Hypothesis 4-1:
   * Null Hypothesis 4-1: There is no relationship between the number of states in a hospital and average critical staffing shortage per state.
   * Alternative Hypothesis 4-1: There is a relationship seen in the number of in a hospital, and the average critical staffing per state.
   * Data Analysis 4-1:First we took a look at the relationship between the number of hospitals in a state and the average critical staffing shortage per state. This was done by taking the average of critical staffing shortages per day in a state, and comparing it to the number of states in a hospital to see if the number of hospitals would impact the staffing shortage. The null hypothesis states that there is no relationship between the number of states in a hospital, and the average critical staffing shortage. When first looking at the number of hospitals per state, we tried to look at which states have the most hospitals. The three states that were the outliers were California with 339 hospitals, Florida with 221 Hospitals, and Texas with 369 hospitals. Then a scatter plot was created to see the relationship, and we saw a high positive correlation, showing that the more hospitals there are in a state, the higher average staffing shortage there was in a day for a state. The ones with the highest values were California and Texas with the most shortages.
   * Result Summary 4-1: These data (Figure 4-1 and 4-2) support that there is a positive correlation, where the more hospitals there are, the more average critical staffing shortages there are per state
2. Hypothesis 4-1:
   * Null Hypothesis 4-2: There is no relationship between the number of hospitals in a state and the % occupancy of inpatient beds
   * Alternative Hypothesis 4-2: There is a relationship between the number of hospitals in a state and the % occupancy of inpatient beds
   * Data Analysis 4-2: Afterwards, while looking at geography, we wanted to take a look at how the number of hospitals in a state would impact the percent occupancy of inpatient beds. The null hypothesis stated that there is no relationship between the number of hospitals in a state, and percent occupancy of inpatient beds. Rhode Island had the highest percentage of beds occupied, and Wyoming and American Samoa had the lowest percentage of beds occupied. This was compared in order to see what if the number of hospitals in a state played a role in the percent occupied inpatient beds, especially since previously it was seen that there was a relationship between the number of hospitals in a state and average shortage.
   * Results Summary 4-2: From the analysis, it is seen that there is no relationship seen between the hospitals in a state and average percent occupancy for inpatient beds, and no values and states that are outliers. This means that the null hypothesis is correct.

**Presentation:**

<https://docs.google.com/presentation/d/16ZlD2ZP7MW3f2VNf2Mx0vtbWuTpn_NLg3J3VEBjYXsY/edit#slide=id.gbdde9ca734_0_2667>

**Sources:**

https://healthdata.gov/Hospital/COVID-19-Reported-Patient-Impact-and-Hospital-Capa/g62h-syeh

https://www.ahd.com/state\_statistics.html