­

**D214 Capstone Project:**

**PA2 – Healthcare Providers and Their**

**Impact on Overall Hospital Ratings**

Jason Willis

College of Information Technology,

Western Governors University

Dr. Daniel Smith

September 08, 2022

Table of Contents for Each Rubric

Part 1: Research Question

[Describe Purpose, Summarize Research Question and Define Objectives:](#RubricA) 3

***Part 2: Data Collection***

[[Summarize Data Preparation](#RubricA), Perform Exploratory Data Analysis:](#RubricA) 3

***Part 3: Data Extraction and Preparation***

[Describe Data Extraction and Preparation:](#RubricA) 3

***Part 4: Analysis***

[Describe Analysis](#RubricA) Performed: 3

***Part 5: Data Summary and Implications***

[Outcomes, Implications and Recommendations:](#RubricA) 3

Part F: Sources

List Sources and Bibliography: 8

# Hospital Ratings

The Patient Survey – Hospital Consumer Assessment of Healthcare Providers and Systems is a dataset provided by the Centers for Medicare and Medicaid Services. This survey poses questions asked of patients and their ratings over a few different clinical perspectives. What can a hospital learn from this survey? Can they affect the outcome, and if so, what services could they focus on? According to Schmocker (2015) “Readiness for discharge appears to be a clinically useful patient-reported metric, as those RFD have higher satisfaction with the hospital and physicians.” Is this the only or best metric to use or can a hospital focus on provider care and strengthen their overall service ratings?

## A – Research Question

Is communication from a doctor more statistically significant to a patient’s overall hospital rating than a nurse?

**Null hypothesis** – Doctor communication does not have a more statistically significant impact on the overall hospital rating when compared to a nurse.

**Alternate Hypothesis** - Doctor communication has a more statistically significant impact on the overall hospital rating when compared to a nurse.

**B – Data Collection**

The Patient Survey – Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) dataset (2022) was selected to provide questions posed to patients about their care. The questions focused on were: “Nurse Communication”, “Doctor Communication” and “Overall Hospital Rating”. A 5-star rating system was utilized.

This survey dataset captures all three measures compared in this analysis for over 4,000 clinics providing over 450,000 rows of data. Most of the time spent was to prepare the data for analytical techniques applied in this study. This effort will be discussed in section C – Data Extraction and Preparation. After the dataset was identified, it was downloaded from the Centers for Medicare & Medicaid Services (CMS, 2022) and then loaded into a data frame (Figure 1).

Table

Description automatically generated

Figure - Load Dataset from \*.csv File

**C – Data Extraction and Preparation**

Once the data is loaded into a data frame, unnecessary columns were dropped and renamed for easier processing. A data type warning was provided for columns with multiple data types since they would cause errors during the analysis process. Some columns had “Not Applicable” and “Not Available” mixed in the same column as the ratings provided by patients. These non-numerical data points were removed, and the data series was converted to an integer data type. (Figure 2) While processing the data isn’t as intuitive as using a graphical user interface, this approach is very efficient as data scales.

Graphical user interface, text, table

Description automatically generated with medium confidence

Figure - Remove Unnecessary Columns, Clean up Mixed Data Types and Rename Columns

**D – Analysis**

Exploratory data analysis was performed on the refined data frame. Here are a few explanations of what was performed:

* Info() method to verify column names, null-value counts, and data types; see Figure 3.
* Shape and describe() methods were used to understand the data frame’s shape, count, unique categorical entries, most frequent with count, mean, standard deviation, minimal, maximum and quantiles 25%, 50% and 75% of the rating values. See Figure 4.
* Head() method was used to understand the layout of the data frame. An attribute of -5 showed the first and last 5 rows. See Figure 5.
* A Seaborn heatmap was used to show any null values graphically. Additionally, Pandas dropna() and .isnull() methods were used to help verify. See Figure 6.
* Questions and ratings were counted. Then the questions were grouped using the .groupby() method to show each question’s mean rating value. See Figure 7.
* Ratings histogram was created, providing visual distribution. See Figure 8.
* Boxplots were created to display the minimum, first quartile, median, third quartile, and maximum values of each grouped question. See Figure 9.
* One-way Analysis of Variance (ANOVA) was calculated. See Figure 10. According to Norman, 2010 “Parametric statistics can be used with Likert data, with small sample sizes, with unequal variances, and with non-normal distributions, with no fear of ‘‘coming to the wrong conclusion’’. These findings are consistent with empirical literature dating back nearly 80 years.” One disadvantage of choosing ANOVA to analyze Likert scale data seemed to be within the limitation of the survey interpretations themselves. The questions to be rated are still able to be interpreted by the individual which may differ when compared to the research objectives.

**Text

Description automatically generated with medium confidence**

Figure - Pandas .info() Method

Table

Description automatically generated

Figure - Pandas .describe() Method

Table

Description automatically generated

Figure - Pandas head() Method

Text

Description automatically generated with low confidence

Figure - Check for Missing or Null Values

Text

Description automatically generated

Figure - Group Question and Rating Data to Aggregate

Chart, bar chart

Description automatically generated

Figure - Ratings Distribution

Chart, box and whisker chart

Description automatically generated

Figure - Boxplot of Questions

Table

Description automatically generated

Figure - One-Way ANOVA

**E – Data Summary and Implications**

Review of hypothesis: Is communication from a doctor more statistically significant to a patient’s overall hospital rating than a nurse?

**Null hypothesis** – Doctor communication does not have a more statistically significant impact on the overall hospital rating when compared to a nurse.

**Alternate Hypothesis** - Doctor communication has a more statistically significant impact on the overall hospital rating when compared to a nurse.

Analysis of Variance, ANOVA for convenience, was performed on the data set to ascertain if “…a significant difference among the groups tested” as stated by Dr. Sewell during hist Lecture: D207 T2 – Welcome to D207 EDA Webinar. (n.d.) ANOVA uses an F-statistic which measures mean equality of a group and a p-value to measure probability under the assumed hypotheses. The F-statistic of the data was 0.6141 and the p-value was 0.541; thus, we fail to reject the null hypotheses. To express this in another way, the result states the means are at least as far apart as observed; given there are no underlying differences between said means. Analysis shows a tight range between doctor, nurse and overall ratings. Both independent variables seem to be important to a hospitals overall rating.

**Recommendations:** Since a patients experience with their doctor and nurse are both important to the hospital’s overall rating, continued training and improved provider/patient relations should be strived for. Additionally, more specific questions could be added to the patient survey to dig deeper into understanding what key behaviors could be championed to improve. A limitation within the current survey point’s to how providers communicate, but this really isn’t the whole story. Trying to understand why a patient provided a certain rating will help illuminate where focus is needed.

**F – Sources**

* Help using Markdown: <https://www.markdownguide.org/basic-syntax/>
* MacTeX: <https://tug.org/mactex/mactex-download.html>
* Matplotlib Help: <https://matplotlib.org/2.1.2/api/_as_gen/matplotlib.pyplot.plot.html>
* Numpy Help: <https://numpy.org/doc/stable/>
* Pandas Help: <https://pandas.pydata.org/docs/user_guide/index.html#user-guide>
* Python Help: <https://docs.python.org/3.9/library/index.html>
* Scipy.stats Help: <https://docs.scipy.org/doc/scipy/reference/tutorial/stats.html>
* Matplotlib: <https://matplotlib.org/stable/index.html>
* Seaborn: <https://seaborn.pydata.org/api.html>
* References: See the references section.

## References

Sewell, W. (n.d.). Lecture: D207 T2 – Welcome to D207 EDA Webinar. Western Governors University. Found Here: <https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=fcf752f1-6ff7-4286-9100-ad1f016a98d6>

Patient Survey – Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS). (2022). Centers for Medicare & Medicaid Services (CMS). Found Here: <https://data.cms.gov/provider-data/dataset/dgck-syfz>

Norman G. Likert scales, levels of measurement and the "laws" of statistics. Adv Health Sci Educ Theory Pract. 2010 Dec;15(5):625-32. doi: 10.1007/s10459-010-9222-y. Epub 2010 Feb 10. PMID: 20146096.

Schmocker R.K., Holden S.E., Vang X, Leverson G.E., et. al., Association of Patient-Reported Readiness for Discharge and Hospital Consumer Assessment of Health Care Providers and Systems Patient Satisfaction Scores: A Retrospective Analysis. J Am Coll Surg. 2015 Dec;221(6):1073-82.e1-3. doi: 10.1016/j.jamcollsurg.2015.09.009. Epub 2015 Sep 25. PMID: 26474513; PMCID: PMC4662900.