

Nursing Home Quality, Staffing

US Data on Certified Beds, Health Deficiencies, Ratings and Staffing Hours

(<https://www.kaggle.com/datasets/thedevastator/nursing-home-quality-staffing>)

Data Source

The data is an external data sourced by [data.world](#); however, it can be found on Kaggle.

[data.world](#) is a public benefit corporation, hosting government datasets. This dataset is a government dataset.

Data collection

Data collection is administrative.

The data was collected via survey by State surveyors during on-site evaluations, and also by reports produced by the nursing homes. Nursing homes are required by Section 6106 of the Affordable Care Act to submit auditable payroll-based staffing and resident census data quarterly.

Data was collected and processed in September, 2016.

Data Content

This dataset provides comprehensive information on currently active nursing homes, such as the type of ownership, number of certified beds, legal business name, and continuing care retirement community status. It also includes ratings and scores associated with quality measures, staffing levels and other factors used in the Five-Star Rating System.

Data Relevance

With this data set, the nursing homes in the US can be accessed and compared based on the ratings and availability of staff and facilities in the nursing homes. It is also to gain insights into how different levels of certification, ownership type, and automatic sprinklers systems affect overall ratings, health inspection ratings, quality measures ratings, and fines payment denials in nursing homes.

The data set was chosen because I am interested in knowing how nursing homes in the US are rated, including the factors that influence this rating. Also, because it meets the requirements of Achievement 6 Project Brief.

Data Profile

Please see attached link.

Data Wrangling and Consistency Checks

The original data set had shape (15640, 82).

However, after first cleaning, resultant shape is (15050, 59).

Dropping Columns and missing values

- ❖ The following columns were dropped because they are not necessary for the analysis:
 - Provider Name
 - Provider Address
 - Provider Zip Code
 - Provider Phone Number
 - Provider SSA County Code
 - Cycle 1 Number of Standard Health Deficiencies
 - Cycle 1 Number of Complaint Health Deficiencies
 - Cycle 2 Number of Standard Health Deficiencies
 - Cycle 2 Number of Complaint Health Deficiencies
 - Cycle 3 Number of Standard Health Deficiencies
 - Cycle 3 Number of Complaint Health Deficiencies

- ❖ The columns for footnote were dropped after dropping all missing values, because they were in the data set to give reasons for missing values.
- ❖ The missing values in the rating columns had reason of either “Too new to be rated” or “Not available”. Therefore, these missing values were dropped as the relative number of missing values is not significant and will not have any impact on the analysis (590 rows).
- ❖ The rows of nursing homes with value “True” in the column “Most Recent Health Inspection More Than 2 Years Ago” were dropped (14 rows) because of the old age of the inspection of these homes.

Mixed Data Type

The following variables had mixed data type:

- With a Resident and Family Council
- Cycle 1 Standard Survey Health Date
- Cycle 2 Standard Health Survey Date
- Cycle 3 Standard Health Survey Date

The mixed data type has been corrected.

Descriptive Statistics

Please see Notebook

Data Limitation

Data was processed in 2016 (7 years ago). There must have been updates to the values in the data set.

Exploratory Questions

1. What is the impact of the different variables on the assessment and ratings of the nursing homes?
2. Which states, cities or counties have nursing homes with good ratings? (and vice versa).
3. What is the relationship between the quality of the nursing homes and the staffing/staffing hours of the nursing homes?
4. Are there relationships between the ratings of the nursing homes?
5. Why do the nursing homes have varying staffing hours per resident per day? Are there nursing homes with similar staffing hours? Does this factor affect the rating or not?