**# Medicare Nursing Home Quality Analysis**

**📊 Project Overview**

***# Project Title: Analyzing Safety and Compliance Features in Nursing Home Facilities***

This project analyzes data from Medicare and Medicaid-certified nursing homes in the United States. The focus is on identifying trends in deficiencies, penalties, staffing, ownership, and quality measures using a comprehensive dataset provided by Medicare.gov.

*15,000* # *Medicare and Medicaid-certified nursing homes in the country.*

**📌 Objectives**

*#To* ***analyze the safety and compliance features*** *in nursing home facilities and their potential impact on facility operations,*

*# focusing on identifying key features related to safety and compliance.*

**📁 Dataset**

The dataset was obtained from the official [Medicare.gov] (https://data.medicare.gov) website. It includes:

- Deficiencies

- Penalties

- Provider information

- Survey summaries

- Ownership types

- Quality metrics (MDS, MSR)

**## 🧰 Tools & Technologies**

- \*\*Python\*\*: Jupiter Notebooks

- \*\*Libraries\*\*: pandas, matplotlib, seaborn, plotly

**- \*\*Visualization\*\*: Bar plots, heatmaps, scatter plots**

**- \*\*Source Control\*\*: Git & GitHub**

**🧠 Key Insights**

- \*\*For-profit nursing homes\*\* had significantly more survey deficiencies than non-profit facilities.

- \*\*Nursing homes with more staffing hours per resident per day\*\* tended to have fewer penalties.

- \*\*States with larger elderly populations\*\* showed varied levels of quality care and deficiencies.

- \*\*Ownership type\*\* plays a major role in penalty occurrences and survey results.

**📈 Visualizations**

A set of charts are used to illustrate:

- Deficiencies by state

- Ownership type vs quality scores

- Penalty distributions

- Staffing comparisons

**## ✅ Project Features**

- Exploratory Data Analysis (EDA)

- Data visualization and storytelling

- Feature engineering (ownership groupings, staffing bins, etc.)

- Insights for healthcare stakeholders, patients, and policy analysts

**## 📌 Future Work**

- Predictive modeling of nursing home deficiencies

- Sentiment analysis on nursing home reviews (if available)

- Integration with CMS hospital-level data for cross-facility insights

**🚀 Author**

Michael Lam  
📍 Application Analyst | Healthcare IT | Data Analytics  
<https://github.com/Michaellam001>

**Understanding the Code:**

**ATTEN: please see GitHub for code:** [**https://github.com/Michaellam001/Nursing-Home-Medicare-Medicaid-Dataset-comparables/blob/main/Nursinghomecompare.ipynb**](https://github.com/Michaellam001/Nursing-Home-Medicare-Medicaid-Dataset-comparables/blob/main/Nursinghomecompare.ipynb)

A screenshot of a computer program

AI-generated content may be incorrect.

***Here's a summary of what the code accomplishes:***

*#Loading Data in Chunks: The data from each CSV file is read in chunks to prevent memory overflow.*

*#Renaming Columns: Standardizes column names to ensure consistent merging.*

*#Optimizing Memory Usage: Converts object columns to categorical types and downcasts numeric columns to save memory.*

*#Merging in Stages: Joins DataFrames in a staged manner to handle large datasets without exceeding memory limits.*

*#Memory Cleanup: Deletes intermediate DataFrames and uses garbage collection to free up memory.*

*#Next Steps*

*#Verify Results: Check the merged DataFrame (df\_combined) to ensure the joins have been performed correctly and the data is as expected.*

*#Perform Additional Analysis: With the combined DataFrame, you can now proceed to any further analysis, visualizations, or modeling.*

A screenshot of a computer

AI-generated content may be incorrect.

*#Inspect the Data: Look at the structure, types, and summary statistics of your dataset to get a feel for the data.*

*#****Data Types:*** *Ensure the data types are appropriate for analysis (e.g., dates are in datetime format).*

**print(df\_combined.info())**

**print(df\_combined.describe(include='all'))**

A white background with black and white text

AI-generated content may be incorrect. *# Data Cleaning and Preparation*

*# Handle Missing Values: Check for and address missing values. You can choose to fill them in, drop rows/columns, or use imputation methods.*

print(df\_combined**.**isnull()**.**sum())

*#Convert Data Types: Ensure that date columns are in the correct format and categorical columns are appropriately encoded.*

df\_combined['filedate\_survey\_summary'] **=** pd**.**to\_datetime(df\_combined['filedate\_survey\_summary'])

**import** matplotlib.pyplot **as** plt

**import** seaborn **as** sns

*# Example: Histogram of a numerical column*

df\_combined['ZIP']**.**hist(bins**=**50)

plt**.**title('Distribution of ZIP Codes')

plt**.**show()

*# Example: Value counts for a categorical column*

print(df\_combined['OWNERSHIP']**.**value\_counts())

A graph with numbers and text

AI-generated content may be incorrect.

***#here is some operational fire safety side of things lets say we want to locate sprinklers and smoke detectors***

A screenshot of a computer code

AI-generated content may be incorrect.

A graph with blue dots

AI-generated content may be incorrect.

A black square with white text

AI-generated content may be incorrect.*#Correlation Analysis: Check for correlations between numerical variables.*

correlation\_matrix **=** df\_combined**.**corr()

sns**.**heatmap(correlation\_matrix, annot**=True**, cmap**=**'coolwarm')

plt**.**title('Correlation Matrix')

plt**.**show()

A white background with black text

AI-generated content may be incorrect.

A screenshot of a computer screen

AI-generated content may be incorrect.

A computer screen shot of a computer code

AI-generated content may be incorrect.

A graph of a number of facilities

AI-generated content may be incorrect.

A screenshot of a computer program

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.



A screenshot of a computer code

AI-generated content may be incorrect.