**Consolidated Dataset Overview Report**

This report provides an overview of the seven CMS datasets used in the Nursing Home Staffing Analysis project. It describes each dataset, its columns, data types, the meaning of each column, and its potential applications within the project.

**1. pbj\_nurse**

* **Overview:** This is the primary dataset for analyzing nursing staff hours (RNs, LPNs, CNAs, and others) at nursing facilities, broken down by employee and contract staff. It contains daily records for each facility.
* **Columns:**
  + provnum: object - CMS Certification Number (CCN) - Unique identifier for the facility. *Application: Joining with other datasets.*
  + provname: object - Provider Name - Name of the nursing facility. *Application: Facility identification, reporting.*
  + city: object - City where the facility is located. *Application: Geographic analysis.*
  + state: object - State where the facility is located. *Application: Geographic analysis, state-level comparisons.*
  + county\_name: object - County where the facility is located. *Application: Geographic analysis.*
  + county\_fips: int64 - County FIPS code - Numerical code for the county. *Application: Geographic analysis, joining.*
  + cy\_qtr: object - Calendar Year and Quarter (2024Q2). *Application: Data validation.*
  + workdate: datetime64[ns] - Date of the work record. *Application: Time-series analysis, daily/weekly aggregations.*
  + mdscensus: int64 - MDS Census - Number of residents assessed using the Minimum Data Set (MDS). *Application: Analyzing staffing relative to resident population.*
  + hrs\_rndon: float64 - Total hours worked by RN Director of Nursing (DON).
  + hrs\_rndon\_emp: float64 - Hours worked by RN DON (Employee).
  + hrs\_rndon\_ctr: float64 - Hours worked by RN DON (Contract).
  + hrs\_rnadmin: float64 - Total hours worked by RNs in administrative roles.
  + hrs\_rnadmin\_emp: float64 - Hours worked by RNs in administrative roles (Employee).
  + hrs\_rnadmin\_ctr: float64 - Hours worked by RNs in administrative roles (Contract).
  + hrs\_rn: float64 - Total hours worked by Registered Nurses (direct care + DON + admin). *Application: Key variable for RN staffing analysis.*
  + hrs\_rn\_emp: float64 - Hours worked by Registered Nurses (Employee). *Application: Key variable for RN staffing analysis.*
  + hrs\_rn\_ctr: float64 - Hours worked by Registered Nurses (Contract). *Application: Key variable for RN staffing analysis.*
  + hrs\_lpnadmin: float64 - Total hours worked by LPNs in administrative roles.
  + hrs\_lpnadmin\_emp: float64 - Hours worked by LPNs in administrative roles (Employee).
  + hrs\_lpnadmin\_ctr: float64 - Hours worked by LPNs in administrative roles (Contract).
  + hrs\_lpn: float64 - Total hours worked by Licensed Practical Nurses (LPNs) / Licensed Vocational Nurses (LVNs).
  + hrs\_lpn\_emp: float64 - Hours worked by LPNs/LVNs (Employee).
  + hrs\_lpn\_ctr: float64 - Hours worked by LPNs/LVNs (Contract).
  + hrs\_cna: float64 - Total hours worked by Certified Nursing Assistants (CNAs). *Application: Key variable for CNA staffing analysis.*
  + hrs\_cna\_emp: float64 - Hours worked by CNAs (Employee). *Application: Key variable for CNA staffing analysis.*
  + hrs\_cna\_ctr: float64 - Hours worked by CNAs (Contract). *Application: Key variable for CNA staffing analysis.*
  + hrs\_natrn: float64 - Total hours worked by Nurse Aides in Training.
  + hrs\_natrn\_emp: float64 - Hours worked by Nurse Aides in Training (Employee).
  + hrs\_natrn\_ctr: float64 - Hours worked by Nurse Aides in Training (Contract).
  + hrs\_medaide: float64 - Total hours worked by Medication Aides.
  + hrs\_medaide\_emp: float64 - Hours worked by Medication Aides (Employee).
  + hrs\_medaide\_ctr: float64 - Hours worked by Medication Aides (Contract).
* **Key Uses:** Calculating RN and CNA temporary staffing ratios, analyzing relationships with resident census, identifying trends, and identifying outlier facilities.

**2. pbj\_non\_nurse**

* **Overview:** Mirrors pbj\_nurse, but for *non-nursing* staff (administrative, therapy, social work, etc.), providing daily hour data.
* **Columns:**
  + provnum, provname, city, state, county\_name, county\_fips, cy\_qtr, workdate, mdscensus: Same meaning and applications as in pbj\_nurse.
  + **Hours Columns (General Pattern):** hrs\_[role], hrs\_[role]\_emp, hrs\_[role]\_ctr for various non-nursing roles, where:
    - hrs\_[role]: Total hours worked by the role.
    - hrs\_[role]\_emp: Hours worked by the role (Employee).
    - hrs\_[role]\_ctr: Hours worked by the role (Contract).
  + **Key Roles (Examples):** hrs\_admin, hrs\_admin\_fn, hrs\_meddir, hrs\_othmd, hrs\_pa, hrs\_np, hrs\_clinnrsspec, hrs\_pharmacist, hrs\_dietician, hrs\_feedasst, hrs\_ot, hrs\_otasst, hrs\_otaide, hrs\_pt, hrs\_ptasst, hrs\_ptaide, hrs\_respther, hrs\_resptech, hrs\_spclangpath, hrs\_therrecspec, hrs\_qualactvprof, hrs\_othactv, hrs\_qualsocwrk, hrs\_othsocwrk, hrs\_mhsvc.
* **Key Uses:** Analyzing non-nursing staffing patterns, comparing contract vs. employee usage across support functions, and relating to operational costs.

**3. qrp\_provider**

* **Overview:** Skilled Nursing Facility Quality Reporting Program (SNF QRP) data, providing quality performance measures.
* **Columns:**
  + cms\_certification\_number\_(ccn): object - Unique CMS Certification Number (provider number). *Application: Linking datasets.*
  + provider\_name: object - Facility name.
  + address\_line\_1: object - Street address.
  + city/town: object - City.
  + state: object - State.
  + zip\_code: int64 - ZIP code.
  + county/parish: object - County/parish.
  + telephone\_number: object - Phone number.
  + cms\_region: int64 - CMS region.
  + measure\_code: object - Specific quality measure code. *Application: Identifying different quality metrics.*
  + score: object - Facility's score on the measure. *Application: Analyzing quality outcomes.*
  + footnote: object - Footnote about the score.
  + start\_date: datetime64[ns] - Reporting period start date.
  + end\_date: datetime64[ns] - Reporting period end date.
  + measure\_date\_range: object
  + location1: object - Facility address.
* **Key Uses:** Linking staffing data to quality outcomes to assess the impact of temporary staffing.

**4. nh\_survey**

* **Overview:** Results from nursing home surveys (inspections), including deficiencies.
* **Columns:**
  + cms\_certification\_number\_(ccn): object - Unique CMS Certification Number. *Application: Linking datasets.*
  + provider\_name: object - Facility name.
  + provider\_address: object - Address.
  + city/town: object - City.
  + state: object - State.
  + zip\_code: int64 - ZIP code.
  + inspection\_cycle: int64 - Inspection cycle (1, 2, or 3).
  + health\_survey\_date: datetime64[ns] - Date of health survey.
  + fire\_safety\_survey\_date: datetime64[ns] - Date of fire safety survey.
  + total\_number\_of\_health\_deficiencies: int64 - Total health deficiencies.
  + total\_number\_of\_fire\_safety\_deficiencies: float64 - Total fire safety deficiencies.
  + count\_of\_[deficiency\_type]\_deficiencies: int64 or float64 - Counts of specific deficiency types (e.g., abuse, quality of life, resident assessment, nursing services, resident rights, nutrition, pharmacy, environmental, administration, infection control, etc.). *Application: Detailed analysis of compliance issues.*
    - Freedom from abuse and neglect.
    - Quality of Life.
    - Resident assesment.
    - Nursing and Physician Services.
    - Resident Rights Deficiencies.
    - Nutrition and dietary.
    - Pharmacy service.
    - Enviromental.
    - Administration Deficiencies.
    - Infection Control.
    - Emergency Preparedness.
    - Automatic sprinklers.
    - Construction.
    - Services.
    - Corridor Walls and Doors
    - Egress.
    - Electrical.
    - Emergency plans.
    - Fire alarm systems.
    - Smoke.
    - Interior.
    - Gas, Vaccum, Electrical Systems.
    - Hazardous area.
    - Illumination and Emergency Power.
    - Labratories.
    - Medical Gases and Anaesthetizing Areas.
    - Smoking Regulations.
    - Miscellaneous.
  + location: object
  + processing\_date: datetime[ns]
* **Key Uses:** Correlating staffing with survey outcomes (number and type of deficiencies).

**5. nh\_quality\_mds**

* **Overview:** Quality measures from the Minimum Data Set (MDS), reflecting resident care.
* **Columns:**
  + cms\_certification\_number\_(ccn): object - Unique CMS Certification Number. *Application: Linking datasets.*
  + provider\_name: object - Facility name.
  + provider\_address: object - Address.
  + city/town: object - City.
  + state: object - State.
  + zip\_code: int64 - ZIP code.
  + measure\_code: int64 - Specific quality measure code.
  + measure\_description: object - Description of the quality measure.
  + resident\_type: object - "Long Stay" or "Short Stay" residents.
  + q1\_measure\_score, q2\_measure\_score, q3\_measure\_score, q4\_measure\_score: float64 - Quarterly scores.
  + footnote\_for\_q1\_measure\_score, footnote\_for\_q2\_measure\_score, footnote\_for\_q3\_measure\_score, footnote\_for\_q4\_measure\_score: float64 - Footnotes for quarterly scores.
  + four\_quarter\_average\_score: float64 - Average score across four quarters.
  + footnote\_for\_four\_quarter\_average\_score: float64 - Footnote for average score.
  + used\_in\_quality\_measure\_five\_star\_rating: object - "Y" or "N" for Five-Star inclusion.
  + measure\_period: datetime64[ns] - **Currently all missing (NaT). This needs investigation.**
  + location: object - Facility location.
  + processing\_date: datetime64[ns] - Data processing date.
* **Key Uses:** Linking staffing to resident care outcomes, using quarterly scores. The missing measure\_period is a significant issue.

**6. nh\_ownership**

* **Overview:** Information on nursing home ownership.
* **Columns:**
  + cms\_certification\_number\_(ccn): object - Unique CMS Certification Number. *Application: Linking datasets.*
  + provider\_name: object - Facility name.
  + provider\_address: object - Address.
  + city/town: object - City.
  + state: object - State.
  + zip\_code: int64 - ZIP code.
  + role\_played\_by\_owner\_or\_manager\_in\_facility: object - Owner/manager role (e.g., "CORPORATE DIRECTOR", "5% OR GREATER DIRECT OWNERSHIP INTEREST"). *Application: Analyzing ownership structure.*
  + owner\_type: object - "Individual" or "Organization". *Application: Categorizing ownership.*
  + owner\_name: object - Name of the owner.
  + ownership\_percentage: object - Percentage of ownership.
  + association\_date: datetime64[ns] - **Currently all missing (NaT). This needs investigation.**
  + location: object - Facility location.
  + processing\_date: datetime64[ns] - Data processing date.
* **Key Uses:** Investigating whether ownership type (for-profit, non-profit, etc.) correlates with staffing or quality.

**7. nh\_citations**

* **Overview:** Data on citations and penalties related to health and safety issues.
* **Columns:**
  + cms\_certification\_number\_(ccn): object - Unique CMS Certification Number. *Application: Linking datasets.*
  + provider\_name: object - Facility name.
  + provider\_address: object - Address.
  + city/town: object - City.
  + state: object - State.
  + zip\_code: int64 - ZIP code.
  + survey\_date: datetime64[ns] - Date of the survey.
  + survey\_type: object - Type of survey ("Health").
  + deficiency\_prefix: object - Prefix for the deficiency code ("F").
  + deficiency\_category: object - Category of the deficiency (e.g., "Quality of Life and Care Deficiencies"). *Application: Analyzing types of violations.*
  + deficiency\_tag\_number: int64 - Specific tag number for the deficiency.
  + deficiency\_description: object - Description of the deficiency.
  + scope\_severity\_code: object - Code indicating the scope and severity of the deficiency (e.g., "D", "E", "F"). *Application: Assessing severity of violations.*
  + deficiency\_corrected: object - Whether the deficiency was corrected.
  + correction\_date: datetime64[ns] - Date the deficiency was corrected.
  + inspection\_cycle: int64 - Inspection cycle (1, 2, or 3).
  + standard\_deficiency: object - "Y" or "N".
  + complaint\_deficiency: object - "Y" or "N".
  + infection\_control\_inspection\_deficiency: object - "Y" or "N".
  + citation\_under\_idr: object - "Y" or "N".
  + citation\_under\_iidr: object - "Y" or "N".
  + location: object - Facility location.
  + processing\_date: datetime64[ns] - Data processing date.
* **Key Uses:** Linking staffing patterns to the frequency and severity of citations, providing a proxy for compliance and potential financial penalties.