## **Comprehensive Project Status Report: Nursing Home Staffing Analysis (February 12, 2025)**

**1. Executive Summary**

This report provides a comprehensive overview of the Nursing Home Staffing Analysis project, designed to investigate the impact of temporary staffing in nursing homes using data from Q2 2024. The project is currently progressing through the **Prepare** and initial **Process** phases. Significant progress has been made in data ingestion, cleaning, standardization, and initial metric development for Registered Nurses (RNs) and Certified Nursing Assistants (CNAs). Key initial findings include the identification of skewed staffing data, bimodal contract usage patterns, facility-level outliers in contract staffing, and weekend spikes in CNA contract ratios. Actionable next steps are clearly defined, focusing on deeper facility-level and shift-level analysis, integration of cost and quality metrics, and the generation of data-driven recommendations for Clipboard Health sales leadership. The primary open issue is resolving a data type conversion problem with the "workdate" column, which is currently being addressed. Overall, the project is on track to deliver valuable insights into nursing home staffing dynamics and demonstrate the value proposition of Clipboard Health's on-demand staffing solutions.

**2. Project Background and Objectives**

* **Project Title:** Nursing Home Staffing Analysis - Q2 2024
* **Objective:** To analyze nursing home staffing data from the second quarter of 2024 (Q2 2024) to understand:
  + The prevalence and patterns of temporary (contract) staffing in nursing homes.
  + The impact of temporary staffing on quality of resident care, operational efficiency, and staffing costs.
  + The predictability of staffing gaps within nursing homes.
* **Context:** Nursing homes are facing significant staffing shortages and increasingly rely on temporary staff. This project leverages data to understand the implications of this trend and to position Clipboard Health as a valuable solution.
* **Target Audience:** Clipboard Health Sales Leadership Team, to inform sales strategies and demonstrate the platform's value to potential nursing home clients.
* **Data Sources:**
  + **Clipboard Health Internal Staffing Data:** (Conceptual - not directly used in this exercise but represents the context)
  + **Publicly Available CMS Datasets:**
    - Payroll Based Journal (PBJ) Daily Nurse Staffing (Q2 2024)
    - PBJ Daily Non-Nurse Staffing (Q2 2024)
    - Skilled Nursing Facility Quality Reporting Program Provider Data (Jan 2025)
    - Nursing Home Survey Data (Jan 2025)
    - MDS Quality Measures (Jan 2025)
    - Nursing Home Ownership (Jan 2025)
    - Health Citations (Jan 2025)
* **Guiding Questions:** The analysis is structured around five SMART questions, focusing on RN and CNA staffing ratios, intra-quarter variations, cost correlations, and staffing pattern predictability (detailed in Appendix A).

**3. Progress to Date**

**3.1. Ask Phase (Completed)**

* **Problem Definition:** Clearly defined the business problem of increasing reliance on temporary staffing in nursing homes and its potential negative impacts.
* **Stakeholder Understanding:** Addressed the needs of Clipboard Health sales leadership, focusing on actionable insights for sales strategy.
* **SMART Question Formulation:** Developed five specific, measurable, actionable, relevant, and time-bound (SMART) questions to guide the analysis (Appendix A).
* **Domain Research Integration:** Incorporated domain research to refine questions and ensure relevance to the nursing home industry.

**3.2. Prepare Phase (Largely Completed)**

* **Data Ingestion:** Successfully ingested data from all planned CMS datasets (PBJ Nurse & Non-Nurse, Quality Reporting, Survey, MDS, Ownership, Citations) initially in CSV and converted to Parquet format.
* **Data Cleaning & Standardization:**
  + Resolved encoding inconsistencies in raw data files.
  + Converted datasets to efficient Parquet format.
  + Addressed missing values and duplicate rows.
  + Conducted initial outlier investigation.
  + Standardized column names to lowercase with underscores.
  + Converted date columns to datetime format.
* **Data Validation:** Performed data quality checks including missing value analysis, duplicate checks, descriptive statistics, and verification against raw CSV files to ensure data integrity.
* **Prepared Datasets:** Cleaned and standardized datasets are saved in the ./data/prepared/ folder.

**3.3. Process Phase (In Progress)**

* **Key Metric Development:** Defined and documented key metrics with clear formulas for each SMART question (Appendix B). Metrics include:
  + RN Temporary Staffing Ratio, CNA Temporary Staffing Ratio, Temporary Staffing Ratios for Nursing and Non-Nursing Roles, Facility Temporary Staffing Percentage, Daily Temporary Staffing Ratio.
* **Initial Metric Calculation and Analysis:**
  + **RN Temporary Staffing Ratio Analysis:** Explored distribution, outliers, facility-level variations, correlations with other nursing roles, and state-level variations. Identified skewed data, bimodal contract usage, and facility outliers.
  + **CNA Temporary Staffing Ratio Analysis:** Analyzed daily fluctuations, correlation with resident census (moderate negative), day-of-week patterns (weekend spikes), and facility-level insights.
* **Outlier Analysis:** Identified and started to characterize outlier facilities and days with extreme staffing ratios.
* **Time-Series Exploration (Initial):** Attempted time-series analysis but is currently blocked by the "workdate" issue.

**4. Key Initial Findings**

* **Skewed Staffing Data:** Data for employee and contract hours, and the temporary staffing ratio, are highly skewed, with most facilities showing modest employee hours and zero or minimal contract hours, while a small subset exhibits very high values.
* **Bimodal Contract Usage:** RN temporary staffing ratio shows a bimodal pattern – many instances of zero contract usage and a significant number of instances with near-total contract usage.
* **Facility-Level Outliers:** A small number of facilities disproportionately contribute to extreme contract staffing usage, both in total volume and as a ratio.
* **Correlation Across Roles:** Facilities tend to adopt consistent staffing strategies across different nursing roles (RNs, LPNs, CNAs) – either relying heavily on employees or on contract staff across the board.
* **Stable Aggregate Time Series (Q2 2024):** At a monthly and weekly aggregate level, RN staffing patterns in Q2 2024 appear relatively stable, with limited fluctuation.
* **Weekend Spikes in CNA Contract Ratio:** CNA temporary staffing ratio consistently increases on weekends, suggesting scheduling gaps and a greater reliance on contract CNAs during weekends.
* **Negative Correlation: CNA Ratio vs. Resident Census:** A moderate negative correlation suggests that on days with higher average resident census, facilities tend to slightly *decrease* their reliance on contract CNAs, potentially utilizing more employee staff or overtime instead.

**7. Overall Project Assessment and Next Steps**

* **Project Status:** The project is progressing well and is currently in the initial stages of the **Process Phase**, building upon a robust **Prepare Phase**.
* **Strengths:** Strong foundation in data preparation, clearly defined metrics, actionable next steps, initial findings providing valuable insights.
* **Challenges:** Need to resolve the "workdate" issue, handle data skewness and outliers effectively, and successfully integrate external cost and quality data.
* **Immediate Next Steps:**
  + **High Priority:** Resolve the "workdate" data type issue to enable time-based analysis.
  + **Continue Process Phase Activities:** Systematically execute the planned next steps outlined in Section 5, focusing on facility-level analysis, shift-level analysis (if feasible), and the integration of external datasets.
  + **Document Continuously:** Maintain detailed documentation of all steps, findings, and decisions.

**8. Appendix A: SMART Questions**

1. **Refined RN Staffing Analysis:** "What is the ratio of contract hours versus employee hours for Registered Nurses (RNs) in Q2 2024 in facilities with fewer than 120 residents?"
2. **Intra-Quarter Variation for CNAs:** "What are the short-term (within Q2 2024) changes in the ratio of temporary versus permanent staffing for Certified Nursing Assistants (CNAs), and how do these changes relate to fluctuations in the number of residents?"
3. **Comparative Trend and Cost Correlation Analysis:** "Over Q2 2024, how do temporary staffing trends differ between direct care (nursing) and support (non-nursing) roles, and what correlations exist between these trends and extra costs such as overtime expenses, temporary staffing fees, and penalties?"
4. **Cost Implication Forecasting:** "What are the cost differences in Q2 2024 between facilities with over 50% temporary staffing versus those with less than 50% for both nursing and non-nursing roles, and how would a 10% reduction in temporary staffing affect overall staffing costs?"
5. **Intra-Quarter Staffing Pattern Analysis:** "What are the short-term (within Q2 2024) variations in temporary staffing levels for both nursing and support roles, and do these patterns show predictable times when staffing gaps occur?"