# **Comprehensive Report on the “Ask” Phase**

This document provides a detailed record of the **Ask** phase for our nursing home staffing data analysis project. The goal is to capture all relevant details so that any stakeholder—technical or non-technical—can pick up where we left off, understand the rationale behind the project, and confidently proceed with the next phases. While this project is quite technical, we’ve also included plain-language explanations for clarity.

## **1. Introduction & Background**

### **1.1 Project Context**

Healthcare facilities frequently face staffing challenges, particularly when deciding how to balance **permanent (employee) staff** and **temporary (contract) staff** to achieve both **cost-effectiveness** and **quality of care**. Our team is analyzing data from the **Centers for Medicare & Medicaid Services (CMS)**—specifically, the **Payroll-Based Journal (PBJ) data** for Q2 2024. These datasets provide **daily staffing hours** by role (e.g., RNs, LPNs, CNAs, administrators, therapists), split into employee vs. contract hours, alongside daily resident census counts.

In addition, we have supplemental data files such as:

* **Skilled Nursing Facility Quality Reporting Program Provider Data (Jan2025)**
* **NH Survey Summary (Jan2025)**
* **NH Quality Measures from MDS (Jan2025)**
* **NH Ownership (Jan2025)**
* **NH Health Citations (Jan2025)**

When combined, these sources allow us to investigate how contract vs. employee staffing correlates with **clinical outcomes, facility characteristics, ownership details,** and **regulatory citations**. Our findings will inform an on-demand staffing platform’s strategy—ultimately helping the sales team position its solutions effectively to healthcare facilities.

## **2. Problem Definition (What We Are Trying to Solve)**

### **2.1 Business Need**

**Clipboard Health** (or any similar staffing solution) seeks to **demonstrate the value** of its on-demand marketplace in filling critical staffing gaps. Traditional staffing agencies can be expensive and slow to respond, whereas a technology-driven, on-demand platform can theoretically mitigate workforce shortages more **cost-effectively** and **efficiently**.

However, to make this case compelling, we must answer fundamental questions:

1. **Where** (which facilities, states, ownership types) is contract staffing usage highest?
2. **When** (time of day, month, or quarter) do facilities most rely on contract staff?
3. **Why** might some facilities have higher reliance on contract labor? (E.g., staff shortages, location constraints, facility size, ownership structure.)
4. **What** is the potential effect on quality or regulatory outcomes when using a high percentage of contract staff?

### **2.2 Scope of Analysis**

* **Data Coverage**: Q2 2024 PBJ Nurse and Non-Nurse datasets form our core. Supplementary Jan2025 files (Quality Reporting, MDS, Ownership, Surveys, Citations) extend the context.
* **Timeframe**: Though PBJ data are daily, we focus primarily on aggregated or comparative analyses across Q2 2024. (We may do finer time-series breakdowns if needed.)
* **Stakeholders**:
  1. **Sales Leadership** – Interested in data-backed evidence to convince healthcare facilities of on-demand staffing benefits.
  2. **Operations / Analytics Teams** – Want to explore patterns in staffing usage to improve the platform’s matching algorithms or marketing campaigns.
  3. **Potential Facility Clients** – May review these insights to see how on-demand staffing aligns with compliance, resident quality, or cost goals.

## **3. Stakeholders & Their Expectations**

1. **Sales Leadership / Business Development**
   * **Expectation**: Actionable metrics and recommendations to identify high-impact markets or facility segments.
   * **Layman’s Terms**: They want to see how big the “opportunity” is and how they can pitch a modern staffing model.
2. **Technical / Analytical Staff**
   * **Expectation**: Full details of data cleaning, transformations, outlier handling, methodology, and code samples for replicability.
   * **Layman’s Terms**: They will be looking under the hood—so we’ll provide robust documentation, code snippets, and definitions for all key metrics.
3. **Executives / Decision-Makers**
   * **Expectation**: High-level insights about business implications, cost, and risk.
   * **Layman’s Terms**: Want to see at-a-glance how contract vs. permanent staffing impacts bottom-line or compliance metrics.

## **4. Desired Outcomes & Objectives**

### **4.1 Primary Objectives**

1. **Quantify Contract vs. Employee Staffing**
   * Identify daily ratios of contract staff hours to total staff hours (nursing & non-nursing).
   * Highlight trends across facility sizes, ownership types, or geographies.
2. **Link Staffing to Quality & Citations**
   * Examine if higher contract usage correlates with certain quality metrics (e.g., readmission rates, pressure ulcer rates) or with more frequent health citations.
3. **Identify Cost/Operational Indicators**
   * If feasible, incorporate cost or penalty data (where available) to see the financial ramifications of contract-heavy staffing.
4. **Deliver Strategic Recommendations**
   * Provide up to five targeted actions for the sales team, such as focusing on certain facility segments or marketing a particular ROI advantage.

### **4.2 Secondary Objectives**

* **Assess Seasonal/Weekly Variations**
  + Does contract usage spike on weekends or holidays?
* **Pinpoint Facilities That May Benefit Most**
  + For instance, small rural facilities with high turnover might rely heavily on short-term contract staff.

## **5. Constraints & Considerations**

1. **Data Limitations**
   * PBJ data only covers **paid hours**—unpaid overtime or volunteer hours aren’t included.
   * Some facilities may be excluded due to “aberrant” data or incomplete submissions.
   * Supplemental datasets (e.g., Quality Reporting, Ownership) might have mismatched coverage periods.
2. **Time & Resource Constraints**
   * This is a **case study** for an interview context, so time is limited. We aim for a robust but concise analysis.
   * The volume of data can be large, so **Python** or **SQL** is preferred over Excel for heavy lifting.
3. **Methodological Constraints**
   * We might observe correlations (e.g., high contract usage with certain citations) but can’t automatically claim causation without deeper investigation.
4. **Data Security & Compliance**
   * While this is a test environment (case study), in production we’d need to secure facility-level data and comply with HIPAA or other relevant regulations if real patient data were involved.

## **6. Key Questions (SMART Questions)**

Here are the main queries guiding our analysis—each framed to be **Specific, Measurable, Action-Oriented, Relevant, and Time-Bound**:

1. **Facility Contract Reliance**
   * “Which facilities in Q2 2024 have a contract-staff ratio exceeding 40%, and how does this ratio vary by size and ownership?”
     + **Layman’s Terms**: We want to find out who uses a lot of temp staff and see if that usage is tied to how big they are or who owns them.
2. **Quality Correlation**
   * “How do daily contract hours per resident correlate with NH Survey Summary or MDS-based quality measures in Q2 2024?”
     + **Layman’s Terms**: Does a facility’s reliance on outsiders (contractors) affect the quality of care?
3. **Citation & Survey Outcomes**
   * “Do facilities with consistently high contract staffing see higher or lower frequencies of health citations in Q2 2024?”
     + **Layman’s Terms**: If a place has a lot of temp nurses or staff, does it get in more trouble with health inspectors?
4. **Ownership Type & Patterns**
   * “Are for-profit or chain-owned facilities in Q2 2024 more likely to have higher contract usage than nonprofits?”
     + **Layman’s Terms**: Check if business model (e.g., big chain or private) influences how often they bring in outside staff.
5. **Recommendations for Sales Leadership**
   * “Based on these insights, which facility categories (size, location, ownership) present the best opportunities for pitching an on-demand staffing solution?”
     + **Layman’s Terms**: Figure out who we should talk to first when selling the flexible staffing model.

## **7. Success Criteria & Deliverables**

### **7.1 Success Criteria**

* **Data Quality**: We successfully merge multiple datasets (PBJ, MDS, Survey Summaries, Ownership, etc.) with minimal mismatch or missing fields.
* **Analytical Rigor**: We produce robust descriptive and exploratory analyses that clearly highlight patterns or correlations relevant to staffing usage.
* **Clarity & Actionability**: Our final outputs (reports, charts, presentations) must be understandable for a broad range of stakeholders, from analysts to sales directors.

### **7.2 Deliverables**

1. **Comprehensive Written Report**
   * Detailing the data cleaning, methodology, findings, and recommended actions.
   * Including technical appendices for code snippets (Python, SQL).
2. **Visual Summaries**
   * Graphs or dashboards showing contract usage trends, facility distribution by usage ratio, possible correlation heatmaps with quality measures.
3. **Executive Summary / Sales Handout**
   * A short, high-level briefing that highlights top takeaways for immediate decision-making.

## **8. Layman’s Explanation of the Project Rationale**

Even though we’re dealing with multiple, somewhat technical datasets, the core idea is straightforward:

* **Nursing homes need enough staff to care for residents**—that’s essential for quality outcomes and compliance with regulations.
* **Sometimes they run short of permanent staff**, so they hire temporary (contract) staff, which can be flexible but often more expensive or harder to integrate.
* We want to **find out exactly how big this “temp staffing” usage is**, how it changes day-to-day, and **whether it impacts facility quality scores** or health citations.
* Armed with that information, we can **tell prospective clients** how an on-demand staffing platform might solve their biggest staffing headaches **more efficiently** than traditional staffing agencies.

## **9. Next Steps After the Ask Phase**

1. **Data Inventory & Access**
   * Confirm credentials for all relevant datasets (PBJ nurse, PBJ non-nurse, QRP, MDS, ownership, etc.).
   * Assess completeness and any potential data gaps.
2. **Preliminary Data Profiling**
   * Load the datasets into Python or a SQL environment to check row counts, missingness, and consistent date formats.
3. **Finalize Scope**
   * Determine if we need further segmentation (e.g., focusing only on certain states or facility sizes).
   * Decide which quality metrics from MDS or QRP are most relevant.
4. **Draft a Project Timeline**
   * Outline how long ingestion, cleaning, analysis, and final reporting might take.

With these steps, we’ll be ready to proceed into the **Prepare** (data gathering), **Process** (cleaning), and **Analyze** (exploration/modeling) phases.

# **Conclusion**

In summary, the **Ask** phase for this CMS-based staffing analysis project is now well-defined. We have:

* **Clarified the business problem**: Assessing contract vs. employee staffing across nursing homes, exploring impact on quality and citations.
* **Identified key stakeholder needs**: From sales leadership seeking marketing angles, to executives wanting big-picture ROI, to technical staff requiring data details.
* **Laid out SMART questions**: Each focusing on how staffing levels correlate with facility characteristics and performance.
* **Established constraints & success criteria**: Recognizing data limitations, confidentiality requirements, and a project scope that emphasizes Q2 2024 but may expand.

This level of clarity ensures a smooth handoff, should another analyst or team take over. It also positions us to maintain a robust audit trail of all decisions. By following these guidelines, the next phases (data prep, cleaning, analysis) can be approached confidently and systematically, keeping us aligned with both technical and business objectives.