Sales Data Analyst Case Study - Clipboard Health

Christopher Legarda

February 4, 2025

Contents

1 Introduction

This report presents an analysis of the CMS Payroll Based Journal (PBJ) Daily Nurse Staffing dataset for Q2 2024. The goal is to derive insights for the Clipboard Health sales leadership team and provide strategic recommendations. Additionally, this document includes SQL queries addressing key sales data questions.

1.1 Objectives

- Analyze nursing home staffing trends, including employee vs. contractor utilization.
- Identify staffing shortages & evaluate compliance with CMS' minimum staffing standards.
- Provide data-driven insights to optimize workforce planning and improve care quality.
- Demonstrate SQL proficiency in sales data analysis and reporting.

2 Data Analysis

2.1 Dataset Overview

The PBJ Daily Nurse Staffing dataset from CMS tracks daily staffing hours for long-term care facilities. Each record represents a facility's staffing per workday, distinguishing between employees and contractors across key nursing roles: Registered Nurses (RNs), Licensed Practical Nurses (LPNs), and Certified Nursing Assistants (CNAs).

2.2 Exploratory Data Analysis (EDA)

This analysis focuses on frontline staffing patterns & compliance with CMS' minimum staffing requirement of 3 hours & 29 minutes per resident per day. Key areas include:

- Employee vs. Contractor Distribution: Identifying facilities heavily reliant on contract nurses, which may indicate workforce instability.
- Regional Staffing Trends: Highlighting states & counties with the lowest staffing hours per resident.
- CMS Compliance: Evaluating facilities failing to meet the 3.48-hour per resident threshold.
- Impact on Facility Quality: Examining how staffing shortages correlate with CMS ratings & patient care outcomes.

2.3 Recommendations

1. Identifying Staffing shortages Findings:

- Top 5 facilities have staffing time below 5 minutes per resident per day, indicating critical understaffing.
- The lowest-staffed facility (Bear Mountain at Worcester) provides only 12 seconds per resident per day.
- Even at the higher end, these staffing times fall significantly below federal recommendations.

Recommendation:

- Target facilities with low staffing per resident.
- Offer rapid hiring and placement solutions to facilities with the most severe shortages.

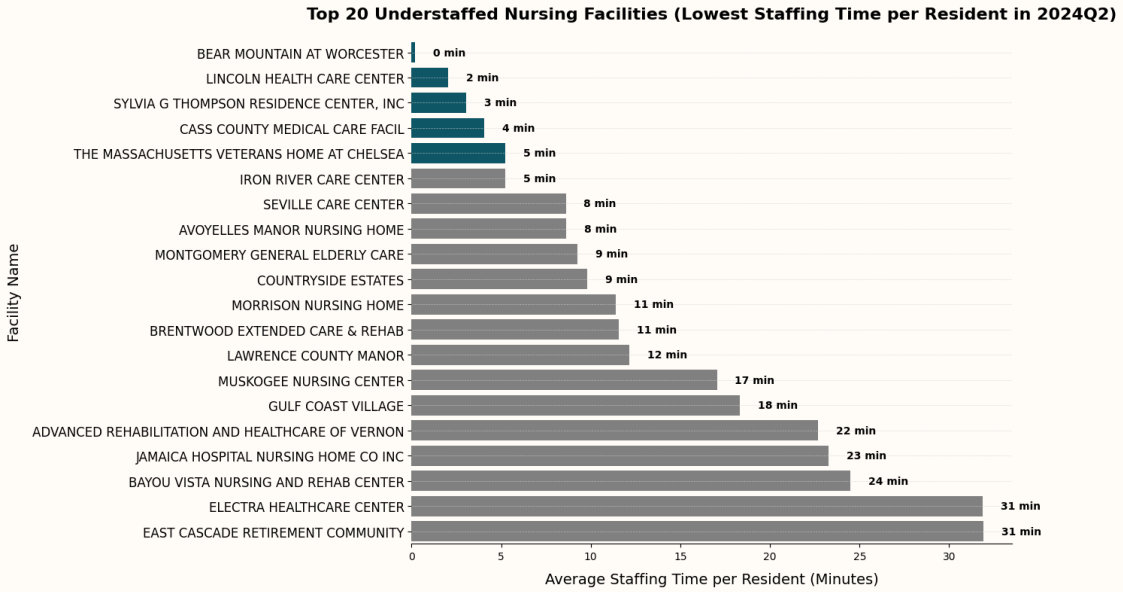

Figure 1: Top 20 Understaffed Nursing Facilities – Critical Staffing Shortages (2024 Q2).

2. Identify High-Demand States & Counties Findings:

- Missouri, Texas, Oklahoma, Indiana & New Mexico fall below CMS-recommended staffing levels.
- Some counties are experiencing extreme staffing shortages, with nurses & caregivers spending less than an hour per resident daily.

Recommendation:

- Develop marketing campaigns for high-need regions, emphasizing emergency staffing solutions.
- Partner with state agencies and local nursing facilities to offer immediate staffing support.

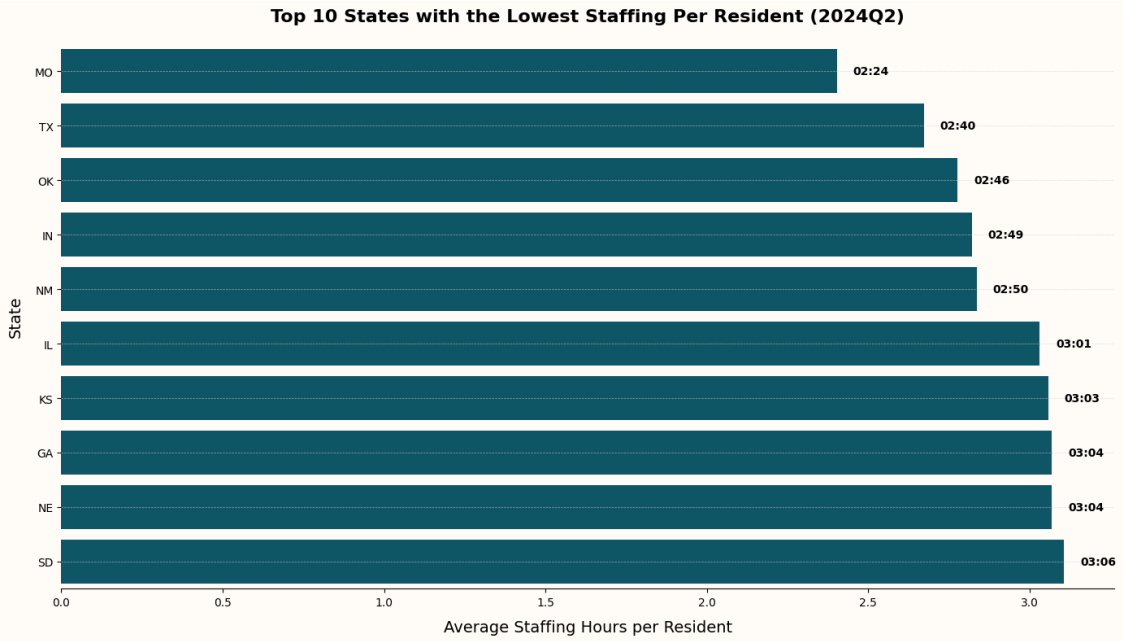

Figure 2: Top 10 States with the Lowest Staffing Hours Per Resident (2024 Q2)

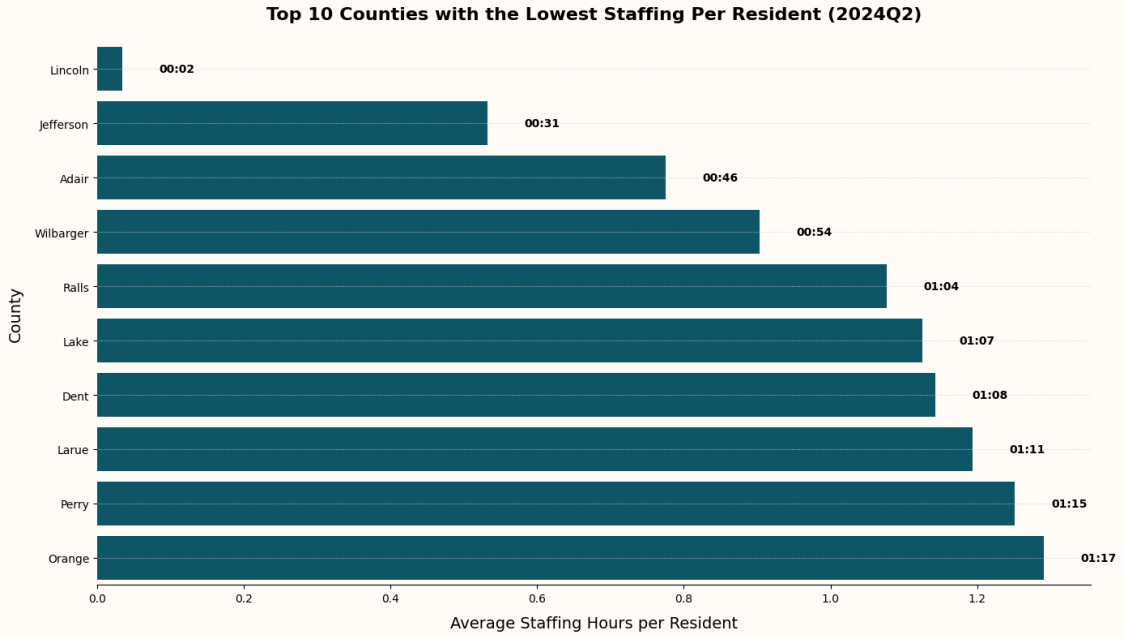

Figure 3: Top 10 Counties with the Lowest Staffing Hours Per Resident (2024 Q2)

3. Compare RN, LPN, and CNA Staffing Findings:

- CNAs are the backbone of long-term care facilities, performing direct patient care.
- LPNs contribute less than half the hours of CNAs, indicating they play a supporting.
- RNs work the least hours, suggesting their role is focused on supervision, medication administration, and clinical decision-making rather than direct care.

Recommendation:

- CNAs perform the most hours, so Clipboard Health should focus on expanding its CNA talent pool.
- Target facilities experiencing CNA shortages to prevent patient care disruptions.

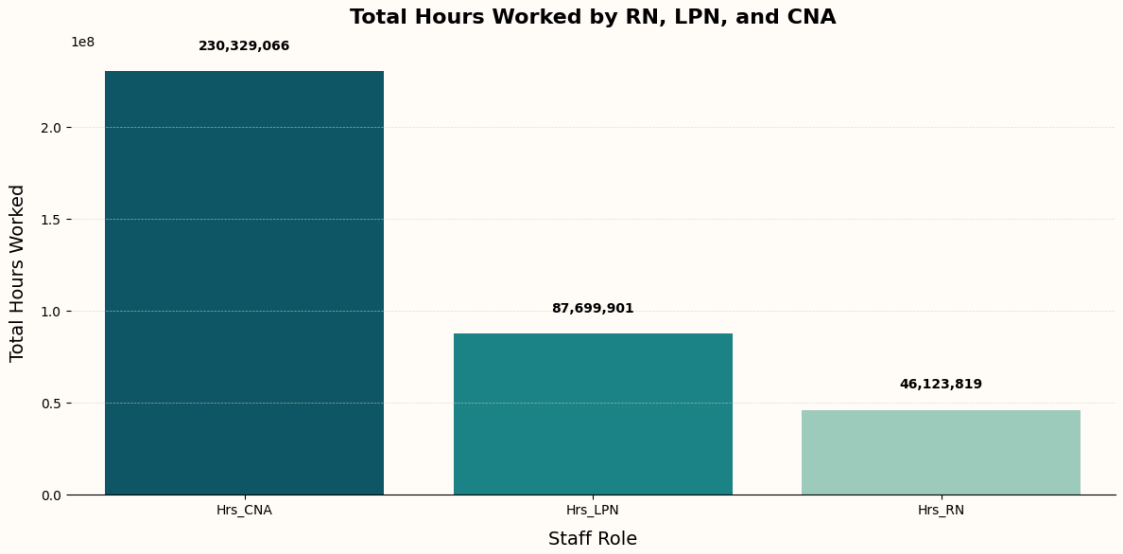

Figure 4: Total Hours Worked by CNAs, LPNs, and RNs in Long-Term Care Facilities (2024 Q2)

4. Contractors vs. Employees in High-Volatility Facilities Findings:

- Some facilities rely heavily on contractors (above 35%) for their staffing needs.
- Facilities like Marigold Rehabilitation HCC (59.3% contractors) show a high dependency on external staffing solutions, which aligns with Clipboard Health's service model.

Recommendation:

- Targeting nursing homes where contractors make up 30%+ of staff.
- Focusing on states/counties with a high % of contractor use.
- Offering flexible staffing solutions to address high variability in these facilities.

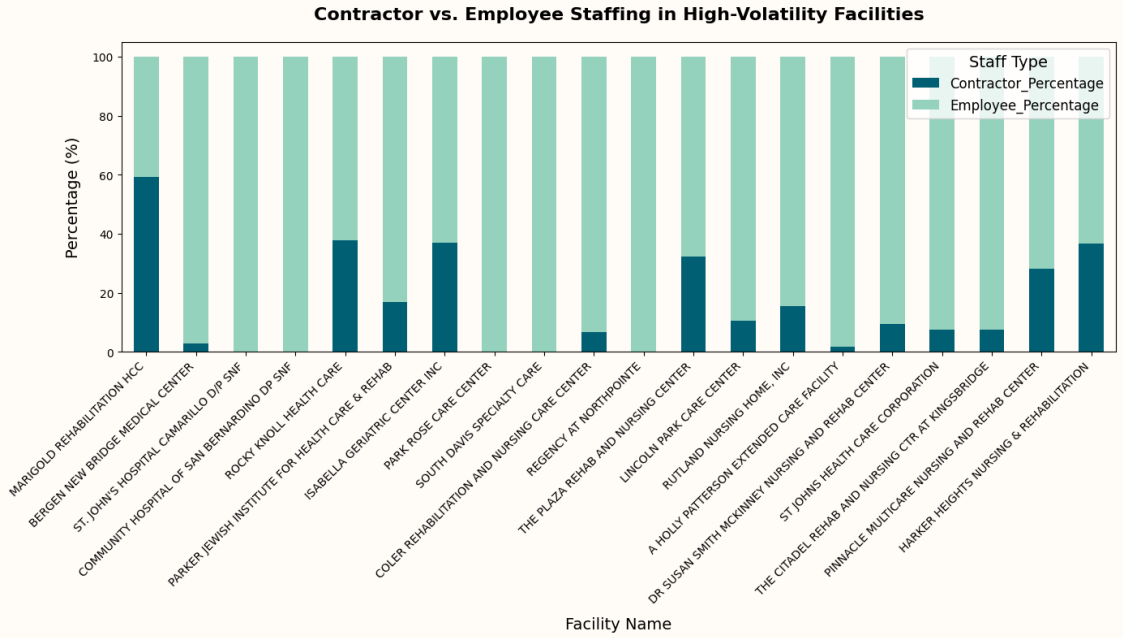

Figure 5: Contractor vs. Employee Staffing in High-Volatility Nursing Facilities (2024 Q2)

3 SQL Test Solutions

Here are SQL queries for the given sales data analysis questions.

3.1 Query 1: Sales in Last 30 Days

```
SELECT
    c.customer_name,
    p.product_name,
    s.total_amount
FROM Sales s
JOIN Customers c ON s.customer_id = c.customer_id
JOIN Products p ON s.product_id = p.product_id
WHERE s.sale_date >= CURDATE() - INTERVAL 30 DAY;
```

3.2 Query 2: Total Revenue by Product Category

```
SELECT
    p.category,
    SUM(s.total_amount) AS total_revenue
FROM Sales s
JOIN Products p ON s.product_id = p.product_id
WHERE s.sale_date >= CURDATE() - INTERVAL 1 YEAR
GROUP BY p.category
ORDER BY total_revenue DESC;
```

3.3 Query 3

```
SELECT DISTINCT c.customer_id, c.customer_name, c.sales_region
FROM Sales s
JOIN Customers c ON s.customer_id = c.customer_id
WHERE YEAR(s.sale_date) = 2023
AND c.sales_region = 'West';
```

3.4 Query 4

```
SELECT
    c.customer_name,
    COUNT(s.sales_id) AS total_sales,
    SUM(s.quantity) AS total_quantity,
    SUM(s.total_amount) AS total_revenue
FROM Sales s
JOIN Customers c ON s.customer_id = c.customer_id
GROUP BY c.customer_name
ORDER BY total_revenue DESC;
```

3.5 Query 5

```
SELECT
    c.customer_name,
    SUM(s.total_amount) AS total_revenue
FROM Sales s
JOIN Customers c ON s.customer_id = c.customer_id
WHERE YEAR(s.sale_date) = 2023
GROUP BY c.customer_name
ORDER BY total_revenue DESC
LIMIT 3;
```

3.6 Query 6

```
SELECT
    p.product_name,
    SUM(s.quantity) AS total_quantity_sold,
    RANK() OVER (ORDER BY SUM(s.quantity) DESC) AS rank
FROM Sales s
JOIN Products p ON s.product_id = p.product_id
WHERE YEAR(s.sale_date) = 2023
GROUP BY p.product_name
ORDER BY rank;
```

3.7 Query 7

```
SELECT
    customer_name,
    sales_region,
    CASE
        WHEN sign_up_date >= CURDATE() - INTERVAL 6 MONTH THEN 'New'
        ELSE 'Existing'
    END AS customer_category
FROM Customers;
```

3.8 Query 8

```
SELECT
    DATE_FORMAT(s.sale_date, '%Y-%m') AS month_year,
    SUM(s.total_amount) AS total_sales
FROM Sales s
WHERE s.sale_date >= CURDATE() - INTERVAL 12 MONTH
GROUP BY month_year
ORDER BY month_year ASC;
```

3.9 Query 9

```
SELECT
    p.category,
    SUM(s.total_amount) AS total_revenue
FROM Sales s
JOIN Products p ON s.product_id = p.product_id
WHERE s.sale_date >= CURDATE() - INTERVAL 6 MONTH
GROUP BY p.category
HAVING total_revenue > 50000
ORDER BY total_revenue DESC;
```

3.10 Query 10

```
SELECT
    s.sales_id,
    c.customer_name,
    p.product_name,
    s.quantity,
    p.price,
    s.total_amount,
    (s.quantity * p.price) AS expected_total
FROM Sales s
JOIN Products p ON s.product_id = p.product_id
JOIN Customers c ON s.customer_id = c.customer_id
WHERE s.total_amount != (s.quantity * p.price);
```