RN Ratio

## **1) Quantify the 1% Outliers in Terms of Facilities**

**Goal**: Understand *how many facilities* that top 1% represents and what their characteristics are.

* **Method**:
  1. **Identify Outlier Rows**
     + Already identified: ~13,000 rows that exceed the 99th percentile for the ratio (or hours).
  2. **Group by provnum**
     + Count how many outlier rows each facility has.
     + This will reveal which facilities are frequent outliers (e.g., a handful of facilities each with many outlier rows) vs. many facilities each with a few outlier rows.
  3. **Facility Count**
     + Summarize how many unique facilities appear in the top 1% (i.e., have at least one outlier row).

**Deliverables**:

* A **table** listing outlier facility count, total contract hours, and ratio ranges among these outliers.
* A **top 10** or **top 50** facility list by total outlier rows or total outlier hours.

## **2) Characterize the Extreme Outliers (High & Low Ratios)**

**Goal**: Determine the nature of facilities/days with extremely high vs. extremely low (near-zero) contract reliance.

* **Method**:
  1. **High-Ratio Subset**: Rows where rn\_temp\_ratio > 0.99 (or the top 1%).
     + Summarize facility type, ownership, region/state, size (mdscensus), etc.
  2. **Low-Ratio Subset**: Rows where rn\_temp\_ratio < 0.01 (or near 0).
     + Same set of facility attributes for contrast.
  3. **Compare & Contrast**:
     + Are high-ratio facilities typically in certain states, certain ownership structures, or do they have large/small resident census?
     + Do low-ratio facilities differ systematically?

**Deliverables**:

* **Descriptive stats** (mean or median mdscensus, state distributions, ownership breakdown) for high-ratio and low-ratio groups.
* **Visualization**: Possibly a small multiples plot or separate boxplots showcasing how these two extreme groups differ in size, location, or staff usage.

## **3) Investigate Exclusive Contract Facilities**

**Goal**: Identify the subset of facilities that rely **entirely** (or near 100%) on contract RN hours.

* **Method**:
  1. **Compute Mean RN Ratio per Facility**
     + If mean(rn\_temp\_ratio) is above 0.99 (for example), categorize the facility as “exclusive contract.”
  2. **Cross-check** with total hours
     + Confirm whether these facilities are large vs. small.
  3. **Time-Series for Exclusives**
     + Focus on these “exclusive” contract facilities over time: do they remain fully contract-based every single day?

**Deliverables**:

* A **list of “exclusive contract” facilities**, their average census, total RN hours, and region.
* **Trend plots** for a few sample facilities to show their daily ratio is indeed near 1.0 consistently.

## **4) Deep-Dive on Outlier Facilities & Hours**

**Goal**: Understand why outlier facilities might have both high contract and high employee hours simultaneously—are they simply very large?

* **Method**:
  1. **Select Facilities** with the highest total contract hours (e.g., top 50) and see if they also rank high on total employee hours.
  2. **Aggregate** their total or average hours to see if these are indeed large facilities (compare to mdscensus).
  3. **Possibility**: Some large facilities might run both many employee and contract hours daily to handle specialized shifts or staff shortages.

**Deliverables**:

* A **scatter plot** of total contract hours vs. total employee hours at the facility level, colored by average census.
* **Conclusions** on whether big user facilities are truly “huge census” sites or if they follow unique staffing models.

## **5) State-Level Variation**

**Goal**: Explore whether certain states or regions strongly favor contract usage or impose different legislative constraints.

* **Method**:
  1. **Group** the data (or just the outliers) by state.
  2. **Compute** average or median RN ratio and total contract hours.
  3. **Compare** states with the highest ratio vs. those with the lowest ratio. Possibly link to known labor laws or union presence.

**Deliverables**:

* A **ranking of states** by mean or median ratio, possibly a bar chart sorted by contract usage.
* **Interpretation** of how legislation or local labor markets might drive differences.

## **6) Facility Staffing Models & the Role of Clipboard**

**Goal**: Tie these extremes back to potential solutions or marketing angles (Clipboard’s offering).

* **Method**:
  1. **Segment** facilities by staffing model: high-employee, high-contract, balanced, etc.
  2. **Analyze** potential cost, quality, or operational benefits of shifting from near-100% contract to a more balanced approach—or from near-0% to some flexible contract usage.
  3. **Clipboard Value Proposition**: If heavy contract facilities face high costs, how can a “cheaper contractor” solution help? If near-zero contract facilities struggle with coverage, how can they be persuaded to try contract staff?

**Deliverables**:

* A **segmentation** of facilities into different staffing “archetypes,” describing how your contract solution might target each group’s pain points or benefits.

## **7) Weekday/Weekend or Shift-Specific Patterns**

**Goal**: Confirm if no clear usage patterns exist or if certain times, like weekends, show a spike.

* **Method**:
  1. **Day-of-Week** grouping: Already tested for CNA, but replicate for RN ratio specifically.
  2. **Shift-level** if data is available: Aggregating hours by morning/evening/night.

**Deliverables**:

* **Trend lines** for each day-of-week or shift, either at an aggregated or facility-level viewpoint.
* Clarification of whether patterns remain “flat” or if the weekend truly stands out.

**8) Merge External Datasets (Cost, Ownership, Penalties)  
Goal: Determine whether outlier facilities incur higher costs, face more regulatory actions, or have different ownership structures.**

* **Incorporate Cost Data**
  + **Agency fees, overtime expenses, or total labor costs.**
  + **Check whether heavy contract usage aligns with higher overall costs or penalty amounts.**
* **Ownership & Penalties**
  + **Are these outlier facilities for-profit vs. nonprofit?**
  + **Do they have more citations or fines?**
* **Analysis**
  + **Join these external fields by facility ID (provnum), then create correlation/regression or group comparisons (e.g., outlier vs. non-outlier).**
* **Deliverables:**
  + **A table or chart showing cost metrics (e.g., average daily cost, penalty counts) for outlier vs. typical facilities.**
  + **Additional layer of insight on whether high ratio usage translates to higher penalties or lower quality scores.**

## **9) Next Analysis Steps**

1. **Consolidate All Findings**
   * Summarize outlier facility counts, their total hours, states, and potential reasons for extreme usage.
2. **Refine**: Decide if we keep top 1% outliers in the main dataset or handle them separately for modeling and forecasting.
3. **Link to Cost and Regulatory Data**
   * Merge with external penalty/cost data to see if high contract usage correlates with higher spending or more citations.
4. **Investigate LPN & CNA** roles similarly
   * We already know RN usage correlates with LPN/CNA usage. A deeper, multi-role approach could yield more robust staffing “profiles.”

## **Summary**

By following these structured next steps, you’ll be able to:

* **Pinpoint** exactly which facilities dominate the outlier space and whether they are truly large or just extremely contract-heavy.
* **Explain** high and low ratio differences (and potential reasons behind them).
* **Integrate** state-level or ownership-level insights to understand legislative or operational nuances.
* **Propose** solutions or marketing angles for high- vs. low-ratio facilities, highlighting where cost savings or coverage gaps might be addressed with your product (Clipboard).

Ultimately, this roadmap will help refine your data-driven narrative: which facilities or groups are prime candidates for intervention, what operational patterns cause high contract usage, and how can your team tailor solutions (e.g., cheaper contract staff or partial contract coverage) to fit distinct staffing models.