Below is an updated version of our comprehensive report on the three key RN-related variables—**hrs\_rn\_emp**, **hrs\_rn\_ctr**, and **rn\_temp\_ratio**—enhanced with recent findings from advanced outlier detection, facility-level analyses, correlation checks, and time-series explorations.

## **1. Overview & Key Observations**

### **Distribution and Skewness**

* **hrs\_rn\_emp (Employee Hours)**
  + Ranges from 0 to ~888 hours, with a median of ~24 hours.
  + ~1% of rows exceed ~148 hours.
  + Histogram and boxplot reveal a **long right tail**, signifying a small fraction of extremely high employee-hour entries.
* **hrs\_rn\_ctr (Contract Hours)**
  + Ranges from 0 to ~512 hours, with a median of 0 hours.
  + ~1% of rows exceed ~44 hours.
  + The majority of records report **no** contract hours (0), reflecting limited usage in most facilities on most days; however, a notable minority of records show high contract hour totals.
* **rn\_temp\_ratio (Temporary Staffing Ratio)**
  + Defined as \frac{\text{hrs\_rn\_ctr}}{\text{hrs\_rn\_emp} + \text{hrs\_rn\_ctr}}.
  + Most values (75%+) are 0, indicating no contract RN usage on many days; about 1% are near 1, signifying full reliance on contract RNs.

### **Visualizations**

* **Histograms**
  + Confirm heavily skewed distributions.
  + Large concentration at 0 for **hrs\_rn\_ctr** and **rn\_temp\_ratio**, and a smaller but important tail of extremely large values for **hrs\_rn\_emp** and **hrs\_rn\_ctr**.
* **Boxplots**
  + Show that extreme values are heavily **right-tailed** for both employee and contract hours.
  + The ratio boxplot highlights the 0 vs. near-1 phenomenon (bimodal pattern).
* **Scatter Plot (hrs\_rn\_emp vs. rn\_temp\_ratio)**
  + Cluster near (x>0,y=0)(x>0, y=0) => Some employee hours but **no** contract hours.
  + Small cluster near y=1y=1 => Entirely contract-based RNs for those facilities/days.

### **Outliers**

* **Top 1%** of rows exhibit **very high** employee or contract hours (above 148 or 44 hours, respectively) or an **rn\_temp\_ratio** near 1.
* Domain context suggests some facilities do legitimately log large aggregated hours or exclusively contract-based RN shifts.

## **2. Advanced Findings**

### **2.1 Extreme Outlier Detection**

* The **99th percentile** stands at ~148 hours (employee), ~44 hours (contract), and a ratio of ~1.0.
* Each metric shows around 13,000 outliers (1% of total rows).
* **Interpretation**: While 99% of facility-days remain within typical ranges, a small portion demonstrates **exceptionally high** reliance on employees or contracts.

### **2.2 Facility-Level Analysis**

* Facilities with the **highest average RN temp ratio** frequently have near‐100% contract usage (zero employee RN hours).
* Facilities with the **highest total RN contract hours** sometimes have substantial employee hours as well, but rely more heavily on contracts overall.
* **Conclusion**: A **handful of facilities** exhibit disproportionate contract usage, whether in total volume or ratio.

### **2.3 Correlation with Other Nursing Variables**

* **hrs\_rn\_emp** strongly correlates with **hrs\_cna\_emp** (~0.59) and more modestly with **hrs\_lpn\_emp** (~0.28).
  + Suggests that facilities with high RN employee hours also tend to staff LPN and CNA roles with employees.
* **hrs\_rn\_ctr** moderately correlates with **hrs\_cna\_ctr** (~0.42).
* **rn\_temp\_ratio** correlates with **hrs\_rn\_ctr** (~0.62), confirming that the ratio naturally rises as contract hours increase.
* **Conclusion**: There's a **tendency for consistent staffing approaches** across roles—facilities that rely heavily on contract RNs often do so for LPNs or CNAs, while those using robust employee staffing do so across multiple roles.

### **2.4 Advanced Outlier Profiling**

* Looking at rows that exceed the 99th percentile in **any** of the three RN variables yields ~37K outlier rows.
* **State-level** grouping of outliers indicates some states have higher absolute counts, potentially due to larger facility densities or local labor practices. Average RN temp ratios among these outliers range from ~0.14 to >0.80, showing variability in how different regions rely on contract staff.

## **3. Time‐Series Analysis**

Despite investigating monthly and weekly aggregations, the dataset appears **remarkably stable** from April to June 2024:

* **Monthly Means**: ~31–32 hrs for employee, ~2.8–2.9 hrs for contract, ratio ~0.065.
* **Weekly Means**: Slightly more granular, but still ~31–32 hrs for employee, ~3 for contract, ratio ~0.06–0.07.
* Even a **two‐step facility‐level** approach (summing daily hours per facility then averaging across facilities) reveals limited variation: ~200 hrs weekly employee vs. ~15–16 contract hours on average, ratio ~0.07.

### **Interpretation of Flat Lines**

* **Aggregate Steadiness**: On average, facilities show consistent RN staffing hours, with no major spikes during Q2 2024.
* **Possible Local Fluctuations**: True changes might exist at the **daily** or **facility** level but are overshadowed by large sample sizes and consistent day‐to‐day usage patterns.

## **4. Key Insights & Recommendations**

1. **Heavily Skewed Data**
   * Most rows show modest employee hours and zero contract hours; a small subset exhibits extremely high values.
   * Consider **winsorizing** or separate analysis of outliers if extreme values hamper certain modeling steps.
2. **Bimodal Contract Usage**
   * Many entries have **no** contract usage (0) or nearly **all** contract usage (~1 ratio), indicating distinct staffing models across different facilities or days.
3. **Facility Outliers**
   * A small number of facilities drive a large portion of extreme contract usage.
   * Further domain analysis might investigate the reasons—**ownership structure**, **financial constraints**, or **acute staff shortages**.
4. **Correlations Across Roles**
   * Facilities either invest in employees across RNs/LPNs/CNAs or they rely heavily on contract staff.
   * Interventions to reduce contract usage in RNs may need to address broader staffing patterns, not just RNs alone.
5. **Stable Aggregated Time Series**
   * At the **monthly or weekly** level, staffing patterns appear **flat** in Q2 2024.
   * This stability can be a valid finding, but if you suspect hidden peaks, **zoom in** on a smaller subset or look at **daily facility-level** data.

### **Next Steps**

* **Validate Outliers**: Cross-check whether extremely high hours or near‐100% contract usage is consistent with facility capacity or known special circumstances.
* **Merge External Datasets**: Incorporate cost, ownership, and penalty data to see if these outlier facilities also stand out for higher costs, lower quality scores, or regulatory actions.
* **Drill Down**: If deeper variability is suspected, evaluate day‐to‐day shifts at a single facility or small groups of interest.
* **Consider Exclusion or Segmentation**: Decide whether to exclude top 1% outliers or keep them in a separate “high‐outlier” segment for specialized modeling or analysis.

## **5. Conclusion**

Our refined analysis of **hrs\_rn\_emp**, **hrs\_rn\_ctr**, and **rn\_temp\_ratio** reconfirms that:

* The data exhibits a **highly skewed** distribution, with most rows at low or zero contract usage and a small set of outliers.
* **Bimodal** patterns are evident in the RN ratio (0 vs. 1).
* **Time-series** trends over Q2 2024 remain flat at an aggregate level, suggesting stable average usage across facilities.
* **Correlations** point to an interdependent staffing structure: those who rely heavily on contract RNs often do likewise for LPNs/CNAs.

Overall, these results spotlight a relatively **stable** but **skewed** RN staffing landscape. To gain further clarity, future steps might focus on deeper facility-level or daily analyses, verifying outlier validity, and examining financial or quality impacts associated with heavy reliance on contract staff.

If you need additional detail on any specific subset analysis, or if you’re ready to integrate external measures (like cost/penalty data) or advanced predictive models, we can proceed accordingly.