# **Data Loading and Preparation**

· Use pandas to load our core CMS datasets (pbj\_nurse, pbj\_non\_nurse, qrp\_provider, nh\_survey, nh\_quality\_mds, nh\_ownership, nh\_citations).

· Validate data quality (checking shapes, missing values, duplicates, and proper date formatting with pd.to\_datetime).

· Standardize column names and merge key fields (e.g., provnum/CCN) for cross-dataset analysis.

# **Aggregated Analysis & Time Series**

· Compute daily aggregated staffing ratios (for nursing, non-nursing, CNA, and RN) using pandas groupby.

· Use resample and rolling window functions to generate time series plots (matplotlib/seaborn) that reveal daily and day-of-week patterns (e.g., weekend upticks in nursing contracts).

# **Deep-Dive into Specific Queries**

## **A. Facility Segmentation & Outlier Characteristics**

· Group by facility (provnum) to calculate contract ratios and identify the top 1% outliers.

· Characterize these facilities by resident census, ownership, and location (using nh\_ownership and qrp\_provider data).

· Compute total contract hours aggregated at the facility level.

## **B. Exclusive Contract Facilities & Exclusive RN Days**

· Filter records to isolate facilities/days with near-100% contract usage.

· Conduct time series analysis on these subsets to uncover scheduling anomalies or operational drivers.

## **C. Employee vs. Contract Tails Relationship**

· Plot distributions (histograms, scatter plots) of employee and contract hours.

· Compute correlations (using pandas corr()) to assess if extreme employee hours counterbalance contract usage.

## **D. Organizational Models & Contractor Economics**

· Infer cost pressures by using proxy measures from staffing patterns (frequency of high contract usage and outlier aggregates).

· Develop a conceptual framework outlining potential cost-savings and operational benefits of Clipboard Health’s on-demand staffing solution.

## **E. High Staffing Facilities**

· Identify facilities with both high employee and contract hours by merging with resident census data.

· Use descriptive statistics (describe()) and visualizations to compare these facilities with lower-staffed counterparts.

## **F. State Variations & Legislative Implications**

· Group by state to calculate and visualize state-level staffing ratios.

· Annotate these visuals with external state regulatory information and labor market indicators.

## **G. Further Analysis on Usage Patterns**

· Implement clustering (e.g., KMeans from scikit-learn) on features like total staffing hours, contract ratios, and census to identify distinct facility segments.

· Explore advanced time series methods to detect subtle usage patterns within clusters.

## **H. CNA Oscillation Patterns**

· Reassess CNA contract ratios versus resident census using scatter plots and correlation analysis, with a focus on weekday versus weekend differences.

## **I. Nursing vs. Non-Nursing Weekly Dynamics**

· Create pivot tables (using pandas pivot\_table) and visualize day-of-week trends to understand anomalies (e.g., Monday peaks in non-nursing contracts versus weekend rises in nursing contracts).

· Explore potential reasons by linking to operational policies if available.

## **J. Intra-Quarter Inconsistencies & Customer Segmentation**

· Calculate variability measures (standard deviation, rolling windows) of temporary staffing ratios over the quarter.

· Map these patterns to known organizational models and segment potential customers for targeted on-demand staffing solutions.

# **Integration of Domain Knowledge & Additional Datasets**

· In addition to the CMS datasets, we should consider integrating external datasets to enrich our analysis:

· External Cost Data: If available, Medicare cost reports or financial datasets that capture overtime expenses, agency fees, and penalty amounts would enable direct cost correlation analyses.

· Labor Market & Economic Data: Datasets from sources like the Bureau of Labor Statistics (BLS) can offer insights into regional labor market conditions that might influence staffing patterns.

· Regulatory Data: State-level regulatory or legislative datasets can help explain variations in staffing ratios due to differing minimum staffing standards or labor laws.

These additional data sources can be imported into Python (using pandas or APIs) and merged with our CMS datasets using common identifiers (e.g., state, county, facility) for a richer, multi-dimensional analysis.

# **Reporting and Recommendations**

· Consolidate all outputs (aggregated metrics, visualizations, clustering results) into a Jupyter Notebook.

· Document insights and actionable recommendations in markdown cells to build a compelling business case for Clipboard Health’s on-demand staffing solution.

By executing this plan entirely in Python, we ensure reproducibility and flexibility. Integrating additional cost, economic, and regulatory datasets will provide a more holistic view of the factors driving temporary staffing patterns in nursing homes, ultimately strengthening our analysis and recommendations.