Part 1: CMS Data Exercise

Recommendations:

1. Target states with **highest total contractor usage**. Highest 5 are: NY, PA, NJ, OH, NC.
2. Prioritize **CNA, LPN, RN** roles. They have the highest total hours.
3. Use metrics like **Contractor Dependency Ratio** and **Census to Staffing** ratio to make more targeted sales pitches to providers with high urgency of contractors.
4. Target '**For Profit'** ownership providers over **'non profit and/or government'** providers. For Profits have higher overall volume of total hours.

Since Clipboard health is a 2-sided marketplace, I will break down my analysis into multiple components and provide analysis that can help the sales team. First, I will focus on healthcare professionals and then the providers.

Healthcare Professionals

Looking at state level for the healthcare professionals, based on the provided data, **LPN, CAN, and RN** have the greatest number of total hours so we will focus on them.

Hrs LPN ctr	8,613,376
Hrs CNA ctr	2,694,586
Hrs RN ctr	846,487
Hrs LPNadmin ctr	37,397
Hrs RNDON ctr	82,337
Hrs RNadmin ctr	180,550
Hrs MedAide ctr	56,164

Figure 1 – Total Number of Hours Ctr

Visually, the choropleth below tells us a narrative that we already know. Bi-coastal states are denser and also have more total number of hours for healthcare professionals. I also have a tabular view of this data below. Looking at this per state, we can see NY, PA, NJ, OH, NC are the top 5 states. Sales teams should double down to gain market share in those states, if already present.

Darkest (Higher Number)

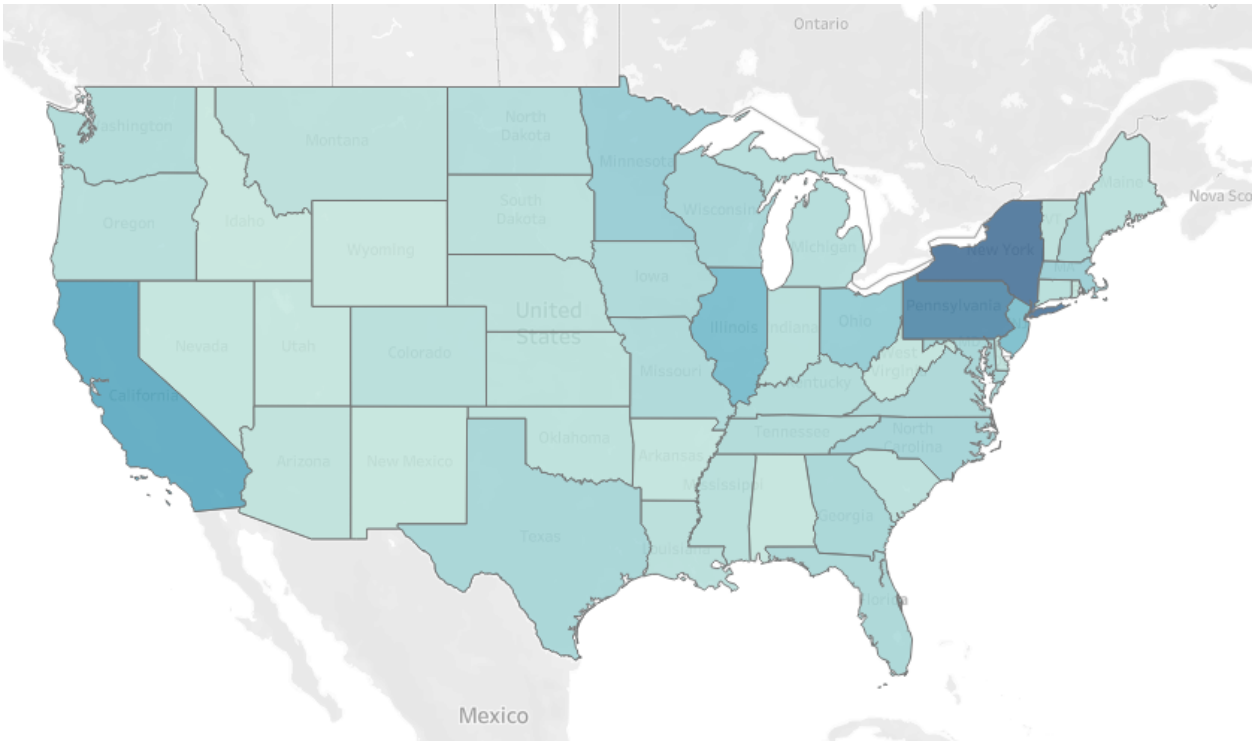

Figure 2 - Choropleth of Total Number of Hours for LPN, CNA, RN in Q1 2024 (Contractor)

Highest

State	Hrs LPN ..	Hrs CNA ctr	Hrs RN ctr
NY	1,216,465	343,309	156,247
PA	1,107,096	288,949	82,519
NJ	502,040	138,123	38,533
OH	372,453	98,916	33,950
NC	371,391	69,838	13,874
MA	363,958	81,165	40,476
IL	318,700	147,370	66,313
TX	314,968	62,895	14,474
CA	289,389	198,050	26,534
FL	280,548	52,962	22,187

Lowest

RI	27,172	9,093	4,542
MT	25,273	25,276	5,748
SD	23,830	27,675	4,514
ID	16,729	1,982	1,232
NV	15,283	8,417	3,565
UT	8,816	8,047	5,761
WY	8,797	3,981	1,822
AK	6,388	2,600	1,508
DC	5,152	2,229	2,828
PR	70	0	258

Figure 3 - Highest Cumulative Hours for LPN, CNA, RN in Q1 2024

Now let's break down the healthcare professionals into different segments. Based on Q1 '24 data, we can forecast the number of hours CNA, LPN, and RNs would take up in the next months. It is insightful to know that there isn't much seasonality from a month over month basis. While this is a quick analysis, we can make this forecast more accurate with more historical data that we can extract.

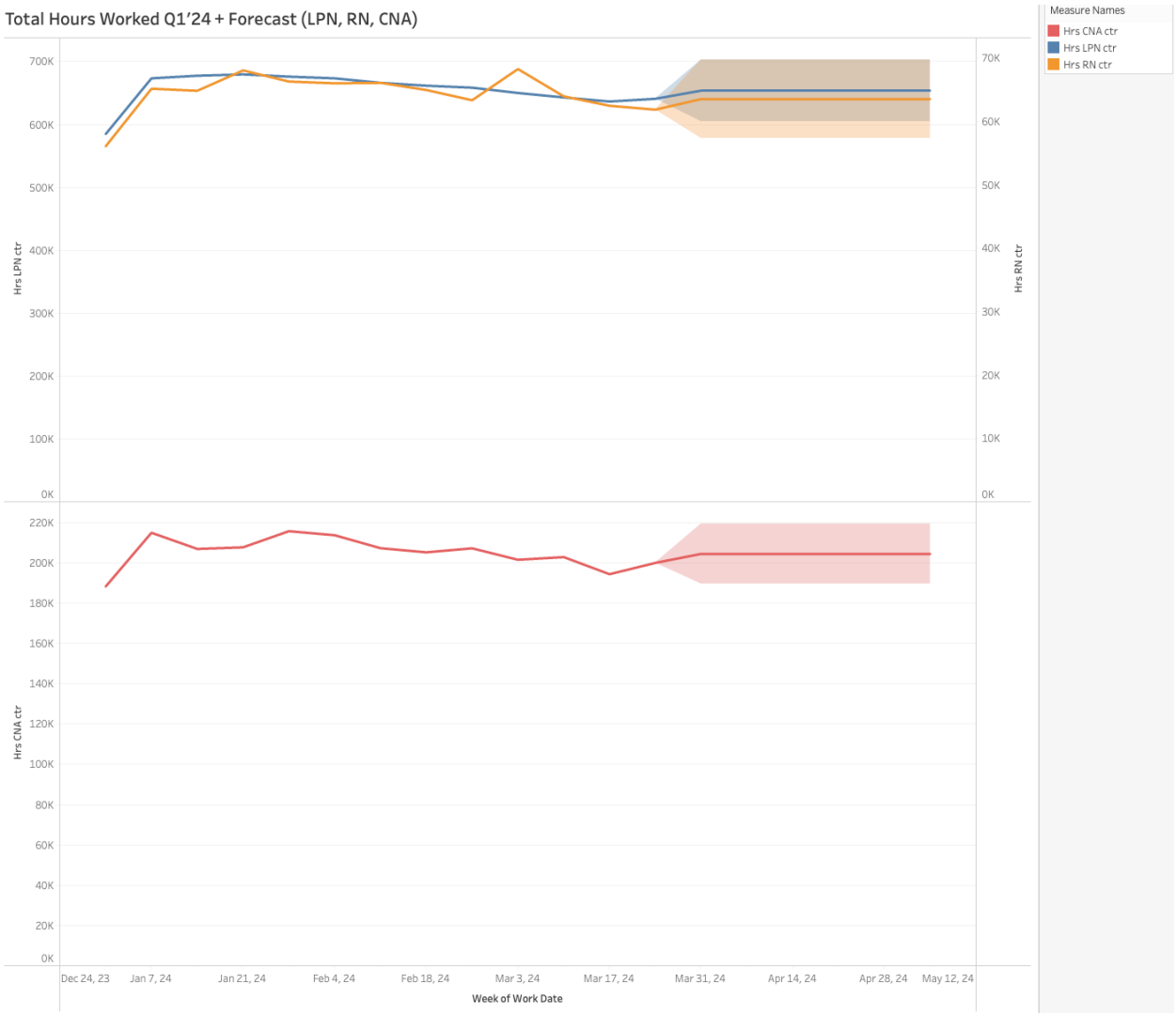

Figure 4 – Total Hours Worked Q1'24 + Forecasts

Figure 5 below shows the Total Addressable Market, calculated by number of total hours * average range of hourly pay. There is information asymmetry and I am making assumptions that are 'best guesses'. We can use the same method to extrapolate for other roles too (RNDON, RNadmin, etc). These average hourly rates are according to Salary.com. Note that these do not reflect differentiated pay for overtime and also may vary per state. You can tell that LPN is the most lucrative job category, followed by CNA and RN.

Job Category	**Hourly Rate**	**TAM**	**Clipboard Profit (assuming 10% cut)**

RN	$45	$ 38,091,915.00	$ 3,809,191.50
LPN	$26	$ 223,947,776.00	$ 22,394,777.60
CNA	$17	$ 51,197,134.00	$ 5,119,713.40

Figure 5 – TAM and Profit Margins per Quarter

Another metric we can leverage is high contractor dependency. From the 8 different job categories, **LPN, CNA, and RN** have the highest cumulative number of hours. Based on this observation, I want to figure out what is the contractor dependency for those roles. For LPN, CNA, and RN, I defined the 'contractor_dependency_ratio (or cd_ratio)' as:

Contractor Dependency Ratio (cd_ratio) = contractor / (employee + contractor)

I then listed the provider name as 'PROVNAME' for values that have a higher contractor dependency than the baseline value (mean value), sorted descending. Cd_ratio is useful since all healthcare professionals that work for Clipboard are considered a contractor. Those providers with high ratios will be good targets to use Clipboard's services. In short, the highest ranked providers will likely depend heavily on contractor labor.

[85]:

	PROVNAME	LPN_cont_ratio
10746	ROSEMONT AT STONE MOUNTAIN	3448.117984
9778	PLEASANT ACRES REHABILITATION AND NURSING CENTER	1558.426948
6084	HUNT NURSING & REHAB CENTER	1402.215156
13555	WALNUT HILLS NURSING HOME	1245.829338
4710	GARDENS AT STEVENS, THE	1023.94279
...	...	...
2957	COLUMBINE MANOR CARE CENTER	8.875765
6465	KINDRED HOSPITAL - LOUISVILLE	8.873292
7919	MEDFORD MULTICARE CENTER FOR LIVING	8.86979
6099	HUNTINGTON HILLS CTR FOR HEALTH AND REHABILITA...	8.862287
7298	LONG ISLAND STATE VETERANS HOME	8.855005

2064 rows × 2 columns

Figure 6 - Facilities with Highest LPN Contractor Dependency Ratio

Furthermore, another metric that is useful is “Census to Staffing Ratio (CSR)”, relative to average. CSR can be interpreted as the number of people listed in the census relative to the number of healthcare professionals. I created a new value called ‘Total_ctr’, which adds all 8 job categories. I then compared it to the census. The higher the CSR value is, the more likely the provider is understaffed and needs more staffing.

[133]:

	PROVNAME	CSR
9994	PROMEDICA SKILLED NURSING & REHAB (WASHINGTON ...	1122.76936
9995	PROMEDICA SKILLED NURSING & REHAB (WEST DEPTFORD)	870.111839
11554	SOUTH COAST POST ACUTE	833.839513
3344	CREEKSIDE POST-ACUTE	599.7068
4643	FUTURE CARE CAPITAL REGION	464.939378
...	...	...
1430	BEECHWOOD HOME FOR INCURABLES	5.289862
7764	MASONIC CARE COMMUNITY OF NEW YORK	5.288548
4308	FALL CREEK REHABILITATION AND HEALTHCARE CENTER	5.286826
1049	AVALON HEALTH CARE CENTER AT STONERIDGE	5.283843
5040	GOOD SHEPHERD VILLAGE	5.283119

3262 rows × 2 columns

Figure 7 - Facilities with Highest CSR Census to Staffing Ratio

Providers

Now, let’s look at the providers. I merged the original dataset with the ‘CMS Certification Number’ with various datasets: QualityMsr_Claims, ProviderInfo, and SurveySummary. The columns of interest are: **overall rating, Staffing Rating, Adjusted Score, observed score, Total Number of Penalties**. The premise is to see whether these new variables are correlated to our Total hours for LPN, CNA, and RN. If so, we can use these as features for machine learning modeling and other predictive work as well. The correlation is not obvious but there seems to be a ‘low positive correlation’.

After some data wrangling, I merged on the CMS number to be the same as the PROVNUMBER (015009 != 15009), we have the table below. Comparing to the total hours for LPN, CNA, and RN. From here, we can also filter it on a provider level to see how it would change.

State (NH Pr..	Hrs LPN ctr	Hrs CNA ctr	Hrs RN ctr	Overall Rating	Total Number of ..	Staffing Rating	Adjusted Score	Total nursing sta..
NY	1,216,465	343,309	156,178	1,765	765	1,579	17,425	24,418
PA	1,096,176	283,812	80,934	1,951	1,032	2,117	19,759	28,836
NJ	502,040	138,123	38,533	1,079	693	959	11,073	14,187
OH	372,453	98,916	33,950	2,679	1,621	2,028	25,494	43,000
NC	371,391	69,838	13,874	1,063	773	999	13,760	19,296
MA	363,958	81,165	40,476	977	764	999	11,888	14,303
IL	316,896	147,181	65,671	1,646	2,499	1,337	23,556	30,061
TX	313,779	62,154	14,426	3,001	2,747	2,117	34,365	56,034
CA	285,967	197,348	26,245	3,577	2,616	3,180	34,469	43,236
FL	280,548	52,962	22,187	2,142	1,243	2,029	24,065	27,548
MD	214,133	63,487	22,352	634	289	628	6,772	9,405
VA	199,080	50,432	12,657	766	303	605	9,160	13,425
TN	191,429	44,021	11,783	769	382	665	9,883	13,388
GA	186,627	59,980	9,700	852	567	786	10,660	16,792
MO	172,269	48,898	9,969	1,185	1,097	989	13,633	24,669
WI	163,979	64,124	20,121	951	652	1,018	9,343	14,711
MI	148,141	36,278	[illegible]	[illegible]	[illegible]	[illegible]	13,183	18,943
KY	143,428	43,827	[illegible]	[illegible]	[illegible]	[illegible]	9,378	12,416
IN	130,595	31,942	[illegible]	[illegible]	[illegible]	[illegible]	14,615	25,301
LA	123,222	28,777	2,103	570	514	553	10,792	11,153
SC	121,879	23,847	7,241	496	299	473	6,454	8,358
CO	114,772	31,415	10,110	608	564	598	4,550	9,768
MN	110,249	94,295	39,011	1,107	733	1,346	7,157	14,100
CT	106,469	30,045	10,410	565	290	596	6,655	7,139
WA	102,283	44,281	11,564	548	452	567	5,230	8,713
NH	94,492	49,929	6,680	199	60	200	2,221	3,299
Null	87,649	47,600	17,933	788	481	777		8,114
OR	86,793	32,325	10,548	320	208	385	3,244	4,764
NE	86,388	23,790	7,260	520	304	538	3,645	8,629
MS	81,781	19,991	2,282	443	218	615	7,431	8,236
VT	79,395	25,349	3,632	100	60	105	1,111	1,960
IA	72,825	49,496	16,066	1,144	847	1,277	7,502	18,039
ME	71,063	23,696	7,538	222	69	286	1,999	3,547
AZ	65,318	19,014	7,536	452	92	382	3,934	6,610
DE	61,702	17,054	6,297	126	57	139	1,201	1,926
WV	61,340	6,914	2,245	290	125	298	3,441	4,543
KS	47,484	22,677	7,556	743	494	836	6,179	11,998
AL	47,042	7,749	2,824	629	184	685	6,928	9,751
NM	42,976	4,709	2,421	169	235	151	2,277	3,300
OK	42,387	21,668	3,829	745	618	680	7,967	13,932
AR	38,515	10,248	2,062	636	180	600	7,350	10,123
ND	33,904	41,323	4,860	225	64	307	1,138	3,125
HI	32,103	14,485	7,378	147	72	157	844	1,253
RI	27,172	9,093	4,542	228	233	259	2,335	3,304
SD	22,268	25,909	4,173	228	128	261	1,333	3,789
MT	21,411	23,860	5,657	155	208	161	1,082	2,788
ID	16,729	1,982	1,232	266	109	248	1,626	3,396
NV	15,202	8,417	3,565	168	59	169	1,574	2,442
UT	8,804	8,041	5,736	283	188	265	2,167	4,656
WY	8,797	3,981	532	85	69	91	720	1,391
AK	6,388	2,600	1,508	48	41	66	239	644
DC	5,152	2,229	2,828	52	62	62	357	475
PR	70	0	258	15	32	21		148

State (NH ProviderInfo Sep2024.csv1): **WI**
Hrs CNA ctr: **64,124**

Figure 8 – Total Number of Hours CNA, RN, LPN ctr vs Other Metrics per State

Figure 9 is broken out by **ownership** type. Government and non profit ownership types have a lower volume of total hours whereas for profit has a higher volume of hours, thus making for profit providers may be the bigger opportunity for the sales team. When prioritizing sales pitches to providers, Sales can prioritize 'For profit' organizations, just based on the volume of total number of hours they can provide, which means more business.

Ownership Type (NH ProviderInfo Sep202..	Hrs LPN ctr	Hrs CNA ctr	Hrs RN ctr	Overall Rating	Total Number of ..	Staffing Rating	Adjusted Score
For profit - Corporation	3,058,740	828,405	269,512	15,060	10,612	13,586	173,223
For profit - Limited Liabilit..	2,284,486	612,445	209,182	9,399	8,822	8,550	122,622
Non profit - Corporation	1,446,622	547,113	165,107	8,302	3,506	8,401	64,555
For profit - Partnership	431,395	122,126	38,868	1,317	993	1,216	15,100
For profit - Individual	349,139	97,344	38,426	1,897	1,588	1,699	22,113
Government - County	325,044	135,701	28,324	1,208	504	1,166	9,388
Non profit - Church related	174,485	87,538	22,459	1,080	484	1,124	7,473
Government - State	148,779	87,450	23,377	401	140	460	1,170
Non profit - Other	134,072	55,404	13,908	1,097	560	1,071	8,297
Null	87,649	47,600	17,933	788	481	777	
Government - Hospital dis..	80,226	19,474	3,999	733	492	543	6,774
Government - City/county	59,381	31,579	5,933	163	75	173	1,292
Government - Federal	22,343	11,633	4,065	43	33	51	499
Government - City	11,014	10,774	5,394	209	125	223	655

Figure 9 – Total Number of Hours CNA, RN, LPN ctr vs Other Ratios per Ownership

Next Steps

- If internal data were available, I'd like to know what Clipboard's market penetration and breakdown is relative to competitors. I would like to further enhance the Choropleth with Clipboard's presence, and color-code it in granularity of providers. It could be possible that some providers have both Clipboard and competitor.
- Are we having staffing issues in some areas? Is network effects a big deal for staffing? For example, how important is it for a Nurse to have a friend or known colleague already working in the facility?
- What is the business model when Clipboard works with providers? Do they pay a flat fee, or a subscription? Knowing this information, I can extrapolate future sales/potential market and capture both sides of the two-sided marketplace.

Sources:

https://www.registerednursing.org/rn-salary/

https://www1.salary.com/Healthcare-Nursing-Salaries.html

https://nurse.org/education/CNA-salary/

https://data.cms.gov/provider-data/search?page-size=50&theme=Nursing%20homes%20including%20rehab%20services

Part 2: SQL Test

1.

```
SELECT
        c.customer_name,
        p.product_name,
        s.total_amount
FROM sales as 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  >= CURRENT_DATE - INTERVAL 30 DAY;
```

2.

```
SELECT
        p.category,
        SUM(s.quantity * p.price) as total_revenue
FROM products p
JOIN Sales s on p.product_id = s.product_id
WHERE s.sale_date >= CURRENT_DATE - INTERVAL 1 YEAR
group by p.category
```

3.

```
SELECT c.customer_name
FROM Customers c
```

```
JOIN Sales s on c.customer_id = s.customer_id
where sales_region = 'West'
and YEAR(sale_date) = 2023
```

4.

```
SELECT
        c.customer_id,
        c.customer_name,
        COUNT(s.sales_id) as total_sales,
        SUM(s.quantity) as total_quantity,
        SUM(s.quantity * p.price) as total_revenue
FROM Customers c
JOIN Sales s on c.customer_id = s.customer_id
JOIN Products p on c.product_id = p. product_id
GROUP BY customer_name
```

5.

```
WITH cte (
        SELECT
                s.customer_name,
                SUM(s.quantity * p.price) as total_revenue
        FROM Sales s
)
SELECT
        customer_name,
        total_revenue
FROM cte
GROUP BY customer_name
ORDER BY total_revenue DESC
LIMIT 3
```

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
```

7.

```
SELECT
        c.sales_region as region,
        c.customer_name,
        CASE WHEN c.sign_up_date >= CURRENT_DATE - INTERVAL 6 MONTH THEN
New' Else 'Existing' END as customer_status
FROM Customer c
```

8.

```
SELECT
        EXTRACT MONTH(sale_date) as month,
        EXTRACT YEAR(sale_date) as year,
        SUM(s.total_amount) as total_sales
FROM Sales s
WHERE  s.sale_date between  CURRENT_DATE -  INTERVAL 12 MONTH AND
CURRENT_DATE
GROUP BY year,month
ORDER BY year, month
```

9.

```
SELECT p.category
FROM Products p
JOIN Sales s ON p.product_id = s.product_id
GROUP BY  p.category
HAVING
        SUM(s.quantity * p.price) > 50000 AND
        MAX(s.sale_date)  >= CURRENT_DATE - INTERVAL 6 MONTH;
```

10.

```
WITH CTE as(
        SELECT
                s.sales_id,
                s.total_amount,
                (s.quantity * p.price) as expected_sales
        FROM Sales s
        JOIN Products p on s.product_id = p.product_id
)
SELECT sales_id
FROM CTE
WHERE total_amount != expected_sales
```