Nurse staffing strategies for enhanced patient care

Matthew Bain

2024-03-22

abstract I analyze a medical staffing dataset and identify avenues to improve work satisfaction among nurses and the quality of care provided at United States medical institutions.

Table of contents

Imports ... 1
The dataset ... 2
  Load the data ... 2
  Inspect the data ... 2
  Group the features ... 4
  Clean the data ... 6
Explore the dataset ... 6
  Visualize distributions ... 6
  Visualize relationships ... 6
  Compare groups ... 7
Feature engineer ... 7
  Join geographical data ... 7
  Join seasonal data ... 7
Analyze geography ... 7
Analyze seasonality ... 7
Model ... 8
Extra visualizations ... 8
  Sparklines ... 8
Archive ... 8
Bibliography ... 8

[…]

Imports

```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

from great_tables import GT
from pandas.plotting import scatter_matrix

from src.stylesheet import customize_plots
from src.inspection import make_df, display, display2
```

The dataset

Load the data

We begin by exploring the data to get to know the features and patterns on which we will base our analysis.

```python
if 'data' not in locals():
    data = pd.read_csv(
        "../data/raw/PBJ_Daily_Nurse_Staffing_Q1_2024.zip",
        encoding='ISO-8859-1',
        low_memory=False
    )
else:
    print("data loaded.")
```

Inspect the data

```python
GT(data.sample(10))
```

PROVNUM	PROVNAME	CITY	STATE	COUNTY_NAME				
475003	BIRCHWOOD TERRACE REHAB & HEALTHCARE	BURLINGTON	VT	Chittenden	2024Q1	[illegible]	157.50.00.00.00.065.57.08.00.00.0146.76.7065.17578.750.00.00.00.00.00.0	
675748	HERITAGE TRAILS NURSING AND REHABILITATION CENTER	CLEBURNE	TX	Johnson	2024Q1	[illegible]	100.00.08.08.0025.45.4820.47.4770.065.8131.05.61.540.00.0029.89.890.0	
6750…	DIVERSICARE OF LULING	LULING	TX	Caldwell	2024Q1	[illegible]	108.00.08.08.009.43.430.00.00.0022.91.9141.41.2411.68.6811.42.420.0	
555…	[illegible]	[illegible]	CA	Tulare	2024Q1	[illegible]	108.00.08.08.00.00.00.08.82.8242.43.1395.95.170.00.00.00.00.00.00.0	
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	

```python
df = data.describe().round(1)
# display(Markdown(df.to_markdown()))
GT(df.reset_index())
```

index	COUNTY_FIPS	WorkDate	MDScensus	Hrs_RNDON	Hrs_RNDON_emp	Hrs_RNDON_ctr	Hrs_RNadmin	Hrs_RNadmin_emp	Hrs_RNadmin_ctr	Hrs_RN	Hrs_RN_emp	Hrs_RN_ctr	Hrs_LPNadmin	Hrs_LPNadmin_emp	Hrs_LPNadmin_ctr	Hrs_LPN	Hrs_LPN_emp	Hrs_LPN_ctr	Hrs_CNA	Hrs_CNA_emp	Hrs_CNA_ctr	Hrs_NAtrn	Hrs_NAtrn_emp	Hrs_NAtrn_ctr	Hrs_MedAide	Hrs_MedAide_emp	Hrs_MedAide_ctr
count	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1330966.0
mean	[illegible]	[illegible]	83.4	5.2	5.1	0.1	10.3	10.0	0.2	34.4	31.5	3.0	6.6	6.6	0.1	66.5	59.8	[illegible]	[illegible]	[illegible]	13.0	4.2	4.2	0.1	8.5	8.3	0.2
std	[illegible]		49.1	4.5	4.5	0.9	14.9	14.6	1.8	34.7	31.4	10.7	10.7	10.6	1.3	48.4	44.8	[illegible]	[illegible]	[illegible]	32.6	13.1	12.7	2.1	17.6	17.2	2.2
min	[illegible]	[illegible]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
25%	[illegible]	[illegible]	51.0	0.0	0.0	0.0	0.0	0.0	0.0	13.0	12.0	0.0	0.0	0.0	0.0	32.8	28.2	0.0	97.0	88.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
50%	[illegible]	[illegible]	76.0	8.0	8.0	0.0	7.5	7.4	0.0	25.6	24.2	0.0	0.0	0.0	0.0	56.9	50.8	[illegible]	[illegible]	136.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0
75%	[illegible]	[illegible]	104.0	8.0	8.0	0.0	16.0	16.0	0.0	44.8	41.5	0.0	9.0	8.8	0.0	88.6	81.2	[illegible]	[illegible]	203.1	11.0	0.0	0.0	0.0	11.2	10.8	0.0
max	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Group the features

We note that there are 91 records per provider (`len(data["WorkDate"].unique())`) and 1,330,966 records in the table overall. The following table, which collapses the raw data across providers, thus has 14,626 $\left(\frac{1330966}{91}\right)$ entries.

```python
df = data.loc[:, [
    "STATE",
    "COUNTY_NAME", "COUNTY_FIPS",
    "CITY",
    "PROVNAME", "PROVNUM",
    # "MDScensus"
]].value_counts()
# df.to_frame()
GT(df.reset_index().head(n=20))
```

STATE	COUNTY_-NAME	COUNTY_-FIPS	CITY	PROV-NAME	PROVNUM	count
AK	Anchorage	20	ANCHOR-AGE	PRESTIGE CARE & REHAB CENTER OF AN-CHORAGE	025025	91
OH	Allen	3	LIMA	LIMA CONVA-LESCENT HOME	366297	91
OH	Allen	3	LIMA	SHAWNEE MANOR	365361	91
OH	Allen	3	LIMA	SPRINGS OF LIMA THE	366464	91
OH	Allen	3	LIMA	SPRINGVIEW MANOR	366221	91
OH	Allen	3	SPENCERVILLE	ROSELAWN MANOR	365744	91
OH	Ashland	5	ASHLAND	BRETHREN CARE VILLAGE HEALTH CARE CENTER	366166	91
OH	Ashland	5	ASHLAND	CRYS-TAL CARE CENTER OF ASH-LAND	366239	91
OH	Ashland	5	ASHLAND	GOOD SAM-[illegible]	365093	91
OH	Ashland	5	ASHLAND	[illegible]	365646	91
OH	**Ashtabula**	**7**	[illegible]	[illegible]	**365[illegible]**	**91**

```
GT(data[["CY_Qtr", "WorkDate", "MDScensus"]].head())
```

CY_Qtr	WorkDate	MDScensus
2024Q1	20240101	50
2024Q1	20240102	49
2024Q1	20240103	49
2024Q1	20240104	50
2024Q1	20240105	51

Clean the data

Explore the dataset

Visualize distributions

Visualize relationships

```
attributes = ["Hrs_RN", "Hrs_LPN_ctr", "Hrs_CNA", "Hrs_NAtrn", "Hrs_MedAide"]
n = len(attributes)

fig, axs = plt.subplots(n, n, figsize=(8, 8))
scatter_matrix(
    data[attributes].sample(200),
    ax=axs, alpha=.7,
    hist_kwds=dict(bins=15, linewidth=0)
)
fig.align_ylabels(axs[:, 0])
fig.align_xlabels(axs[-1, :])
for ax in axs.flatten():
    ax.tick_params(axis='both', which='both', length=3.5)

# save_fig("scatter_matrix_plot")

plt.show()
```

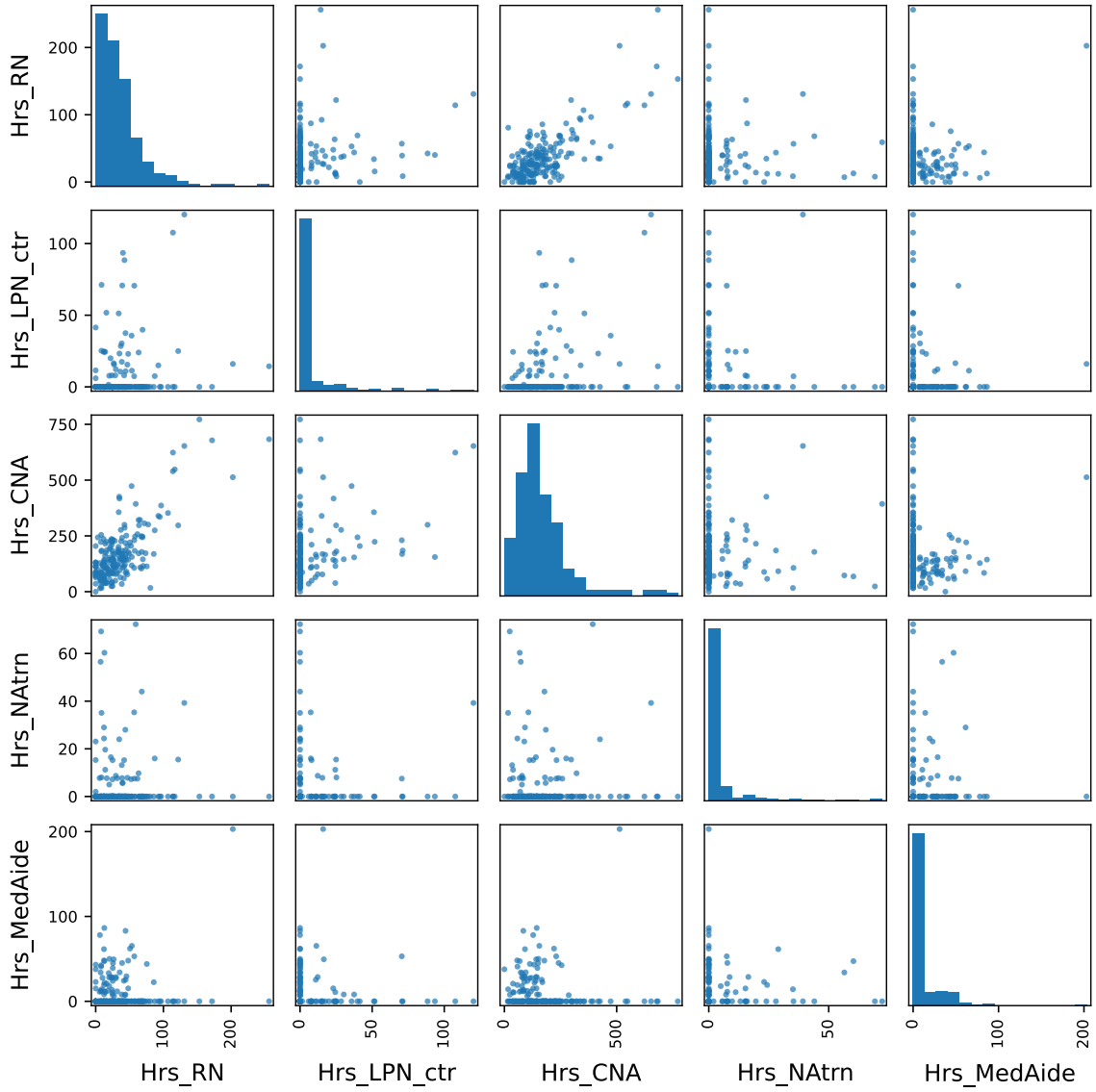

Figure 1: Scatter matrix of nursing worker working hours

Compare groups

Feature engineer

Join geographical data

Join seasonal data

Analyze geography

Analyze seasonality

Model

Extra visualizations

Sparklines

```
# Plot sparklines of average work hours across 91 days by state
(
    GT(gt_df.head(), rowname_col="STATE")
    .fmt_nanoplot(
        columns="lines",
        reference_line="mean",
        reference_area=["min", "q1"]
    )
    .tab_header(
        title="Nurse hours worked in the United States",
        subtitle="The top 5 busiest states",
    )
    .tab_stubhead(label="State")
    .cols_label(
        lines="Total hours worked over 91 days",
    )
)
```

Nurse hours worked in the United States

The top 5 busiest states

State	Total hours worked over 91 days
2.80K2.87K2.84K2.46K2.38K2.91K2.91K2.98K2.97K2.89K2.47K2.37K2.88K2.88K2.89K2.90K2.80K2.36K2.24K	
63.1K62.4K60.0K46.8K46.3K59.9K62.7K63.5K62.8K59.9K48.0K47.0K59.6K61.9K63.4K62.0K58.3K46.4K44.3K	
53.5K52.4K50.8K40.0K39.0K50.4K52.6K52.7K52.0K49.6K39.6K38.7K50.3K53.1K54.0K53.5K51.5K39.9K37.7K	
31.1K31.0K30.5K25.0K24.2K29.8K30.6K30.9K31.0K30.7K25.3K24.3K30.4K30.8K31.1K31.0K30.3K24.7K23.1K	
K279K285K286K286K282K245K242K277K285K285K285K281K246K244K278K284K286K286K280K247K237K	

Figure 2: Sparklines of average work hours across 91 days by state.

Archive

Bibliography