

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



```
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.

```
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

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

[illegible]

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

	COUN-	MD-	Hrs_H-	Hrs_R-	Hrs_H-	Hrs_P-	Hrs_LP-	Hrs_H-	Hrs_-	
	TW-	H-	HNRN-	Nad-	Nad-	Nad-	Hrs_H-	Hrs_LP-	Hrs_H-	CNANA-
	RPS-	State-	CD-	MD-	ON-	Prctim-	impr-	impr-	impr-	impr-
1330966	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
20240121	15.73	4.52	5.10	1.10	3.00	0.24	4.15	3.06	6.66	0.16
9012	3.49	1.45	4.50	1.44	1.46	1.84	7.31	4.07	0.70	6.13
20240110	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
20240112	3.51	0.00	0.00	0.00	0.00	0.00	0.13	0.12	0.00	0.00
20240021	5.76	0.80	0.80	0.07	7.57	4.02	5.64	2.00	0.00	0.00
20240130	9.10	4.08	0.80	0.16	1.60	0.44	8.15	0.09	0.88	0.08
20240023	17.43	32.73	27.82	26.62	26.29	25.89	44.20	24.62	8.61	5.46

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 ($\frac{1330966}{91}$) entries.

```
df = data.loc[:, [
    "STATE",
    "COUNTY_NAME", "COUNTY_FIPS",
    "CITY",
    "PROVNAME", "PROVNUM",
    # "MDScensus"
]]
df.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 SAKE AND OVER THE RIVER NURSING CENTER	365093	91
OH	Ashland	5	ASHLAND	ASHLAND BROOKH- AMPTON CAMP AND HOSPITAL NURSING CENTER	365646	91
OH	Ashtabula	7	ASHTABULA BLVD 1	ASHTABULA BLVD 1	366000	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()
```

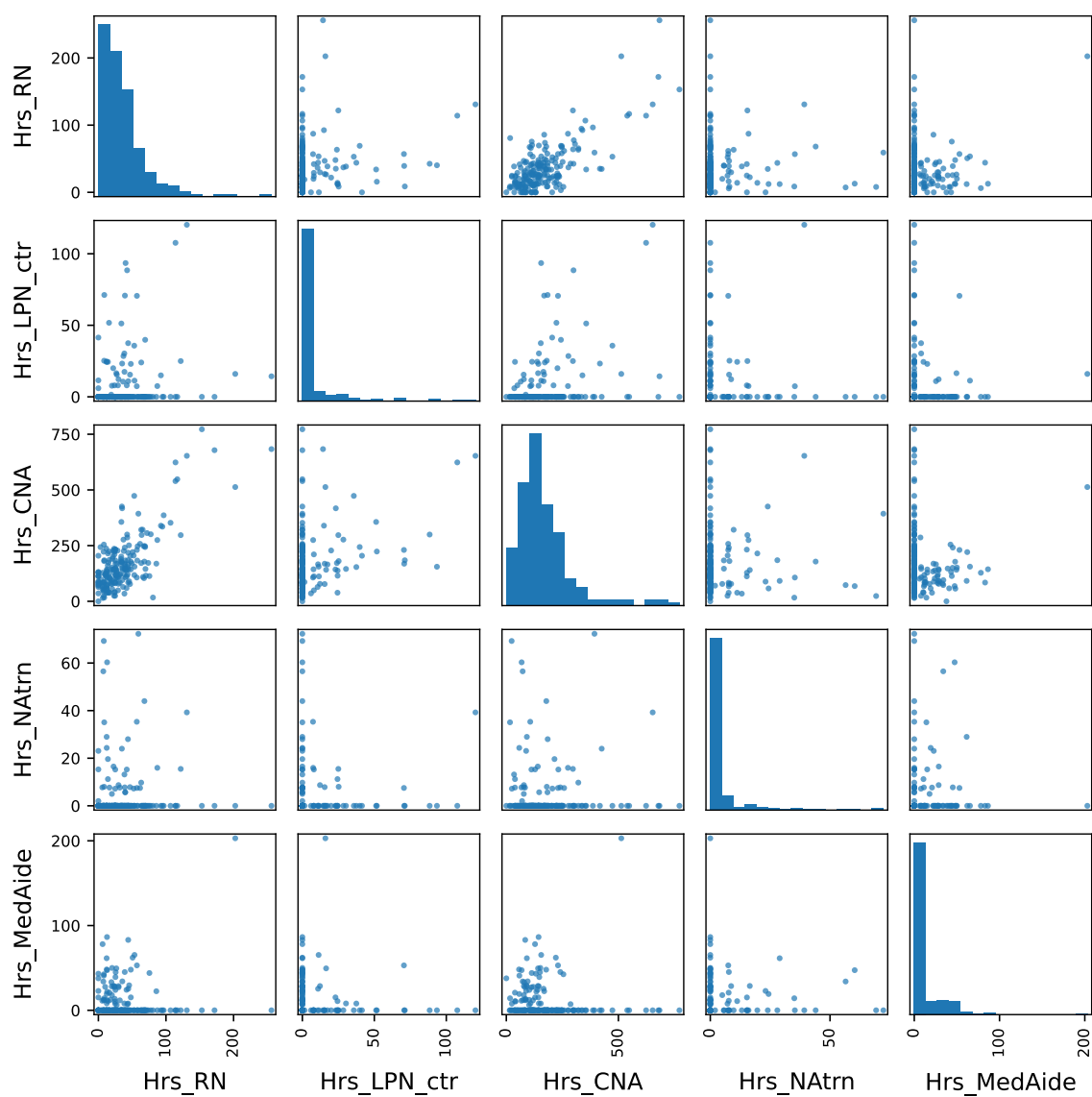


Figure 1: Scatter matrix of nursing worker working hours

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	
279K285K286K286K282K245K242K277K285K285K285K281K246K244K278K284K286K286K280K247K237K	

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

Archive

Bibliography