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

[…]

0.1 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
```

0.2 The dataset

0.2.a 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.")
```

0.2.b Inspect the data

```
data.sample(5)
```

	PROVNUM	PROVNAME	CITY	STATE	COUNTY_NAME	COUNTY_FIPS	CY_Qtr	WorkDate	MDScensus	Hrs_RNDON	...	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
481867	205086	PINNACLE HEALTH & REHAB AT N BERWICK	NORTH BERWICK	ME	York	31	2024Q1	20240125	38	8.00	...	0.0	159.75	159.75	0.0	0.00	0.00	0.0	0.00	0.00	0.0
333320	155263	SYCAMORE CARE STRATEGIES	LOOGOOTEE	IN	Martin	101	2024Q1	20240319	93	8.95	...	0.0	62.54	62.54	0.0	0.00	0.00	0.0	0.00	0.00	0.0
386166	165386	COLONIAL MANOR OF ELMA	ELMA	IA	Howard	89	2024Q1	20240223	35	7.50	...	0.0	47.25	47.25	0.0	0.00	0.00	0.0	9.25	9.25	0.0
980747	395456	GARDENS AT WYOMING VALLEY, THE	WILKES-BARRE	PA	Luzerne	79	2024Q1	20240218	87	0.00	...	0.0	175.00	175.00	0.0	0.00	0.00	0.0	0.00	0.00	0.0
304194	145984	ALDEN NORTH SHORE REHAB & HCC	SKOKIE	IL	Cook	31	2024Q1	20240317	32	7.25	...	7.5	113.25	113.25	0.0	15.25	15.25	0.0	0.00	0.00	0.0

```
data.info(memory_usage=False)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1330966 entries, 0 to 1330965
Data columns (total 33 columns):
 #   Column            Non-Null Count    Dtype
---  ------            --------------    -----
 0   PROVNUM           1330966 non-null  object
 1   PROVNAME          1330966 non-null  object
 2   CITY              1330966 non-null  object
 3   STATE             1330966 non-null  object
 4   COUNTY_NAME       1330966 non-null  object
 5   COUNTY_FIPS       1330966 non-null  int64
 6   CY_Qtr            1330966 non-null  object
 7   WorkDate          1330966 non-null  int64
 8   MDScensus         1330966 non-null  int64
 9   Hrs_RNDON         1330966 non-null  float64
 10  Hrs_RNDON_emp     1330966 non-null  float64
 11  Hrs_RNDON_ctr     1330966 non-null  float64
 12  Hrs_RNadmin       1330966 non-null  float64
 13  Hrs_RNadmin_emp   1330966 non-null  float64
 14  Hrs_RNadmin_ctr   1330966 non-null  float64
 15  Hrs_RN            1330966 non-null  float64
 16  Hrs_RN_emp        1330966 non-null  float64
 17  Hrs_RN_ctr        1330966 non-null  float64
 18  Hrs_LPNadmin      1330966 non-null  float64
 19  Hrs_LPNadmin_emp  1330966 non-null  float64
 20  Hrs_LPNadmin_ctr  1330966 non-null  float64
 21  Hrs_LPN           1330966 non-null  float64
 22  Hrs_LPN_emp       1330966 non-null  float64
 23  Hrs_LPN_ctr       1330966 non-null  float64
 24  Hrs_CNA           1330966 non-null  float64
 25  Hrs_CNA_emp       1330966 non-null  float64
 26  Hrs_CNA_ctr       1330966 non-null  float64
 27  Hrs_NAtrn         1330966 non-null  float64
 28  Hrs_NAtrn_emp     1330966 non-null  float64
 29  Hrs_NAtrn_ctr     1330966 non-null  float64
 30  Hrs_MedAide       1330966 non-null  float64
 31  Hrs_MedAide_emp   1330966 non-null  float64
 32  Hrs_MedAide_ctr   1330966 non-null  float64
dtypes: float64(24), int64(3), object(6)
```

```
data.describe().round(1)
# display(Markdown(data.describe().to_markdown()))
```

	COUNTY_FIPS	WorkDate	MDScensus	Hrs_RNDON	Hrs_RNDON_emp	Hrs_RNDON_ctr	Hrs_RNadmin	Hrs_RNadmin_emp	Hrs_RNadmin_ctr	Hrs_RN	…	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	1330966.0	1330966.0	1330966.0	1330966.0	1330966.0	1330966.0	1330966.0	1330966.0	1330966.0	1330966.0	…	1330966.0	1330966.0	1330966.0	1330966.0	1330966.0	1330966.0	1330966.0	1330966.0	1330966.0	1330966.0
mean	92.0	20240218.7	83.4	5.2	5.1	0.1	10.3	10.0	0.2	34.4	…	6.5	171.2	158.2	13.0	4.2	4.2	0.1	8.5	8.3	0.2
std	99.2	83.0	49.1	4.5	4.5	0.9	14.9	14.6	1.8	34.7	…	16.2	113.7	106.3	32.6	13.1	12.7	2.1	17.6	17.2	2.2
min	1.0	20240101.0	1.0	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%	31.0	20240125.0	51.0	0.0	0.0	0.0	0.0	0.0	0.0	13.0	…	0.0	97.0	88.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
50%	69.0	20240215.0	76.0	8.0	8.0	0.0	7.5	7.4	0.0	25.6	…	0.0	148.1	136.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0
75%	127.0	20240309.0	104.0	8.0	8.0	0.0	16.0	16.0	0.0	44.8	…	5.8	217.0	203.1	11.0	0.0	0.0	0.0	11.2	10.8	0.0
max	840.0	20240331.0	743.0	327.8	327.8	42.0	266.2	266.2	92.5	908.6	…	454.0	1857.0	1573.0	694.3	452.0	279.0	280.5	395.6	395.6	128.9

0.2.c Group the features

```
df = data.loc[:, [
    "STATE",
    "COUNTY_NAME", "COUNTY_FIPS",
    "CITY",
    "PROVNAME", "PROVNUM",
    # "MDScensus"
]].value_counts()
df.to_frame()
# GT(df.reset_index().head(n=5))
```

STATE	COUNTY_NAME	COUNTY_FIPS	CITY	PROVNAME	PROVNUM	count
AK	Anchorage	20	ANCHORAGE	PRESTIGE CARE & REHAB CENTER OF ANCHORAGE	025025	91
OH	Allen	3	LIMA	LIMA CONVALESCENT HOME	366297	91
				SHAWNEE MANOR	365361	91
				SPRINGS OF LIMA THE	366464	91
				SPRINGVIEW MANOR	366221	91
…	…	…	…	…	…	…
IN	Tippecanoe	157	WEST LAFAYETTE	HERITAGE HEALTHCARE	155402	91
				INDIANA VETERANS HOME	155787	91
				UNIVERSITY PLACE HEALTH CENTER AND ASSISTED LIVING	155725	91
				WESTMINSTER VILLAGE - WEST LAFAYETTE	155177	91
WY	Weston	45	NEWCASTLE	WESTON COUNTY HEALTH SERVICES	535023	91

```
display2(
    "data['STATE'].value_counts()",
    "data['COUNTY_NAME'].value_counts()",
    "data['CITY'].value_counts()",
    "data['PROVNAME'].value_counts()",
    "data['MDScensus'].value_counts()",
    width="340px",
    globs=globals()
)
```

```
<IPython.core.display.HTML object>
```

```
data[["CY_Qtr", "WorkDate", "MDScensus"]]
```

	CY_Qtr	WorkDate	MDScensus
0	2024Q1	20240101	50
1	2024Q1	20240102	49
2	2024Q1	20240103	49
3	2024Q1	20240104	50
4	2024Q1	20240105	51
…	…	…	…
1330961	2024Q1	20240327	81
1330962	2024Q1	20240328	83
1330963	2024Q1	20240329	85
1330964	2024Q1	20240330	82
1330965	2024Q1	20240331	82

0.2.d Clean the data

0.3 Explore the dataset

0.3.a Visualize distributions

0.3.b 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()
```

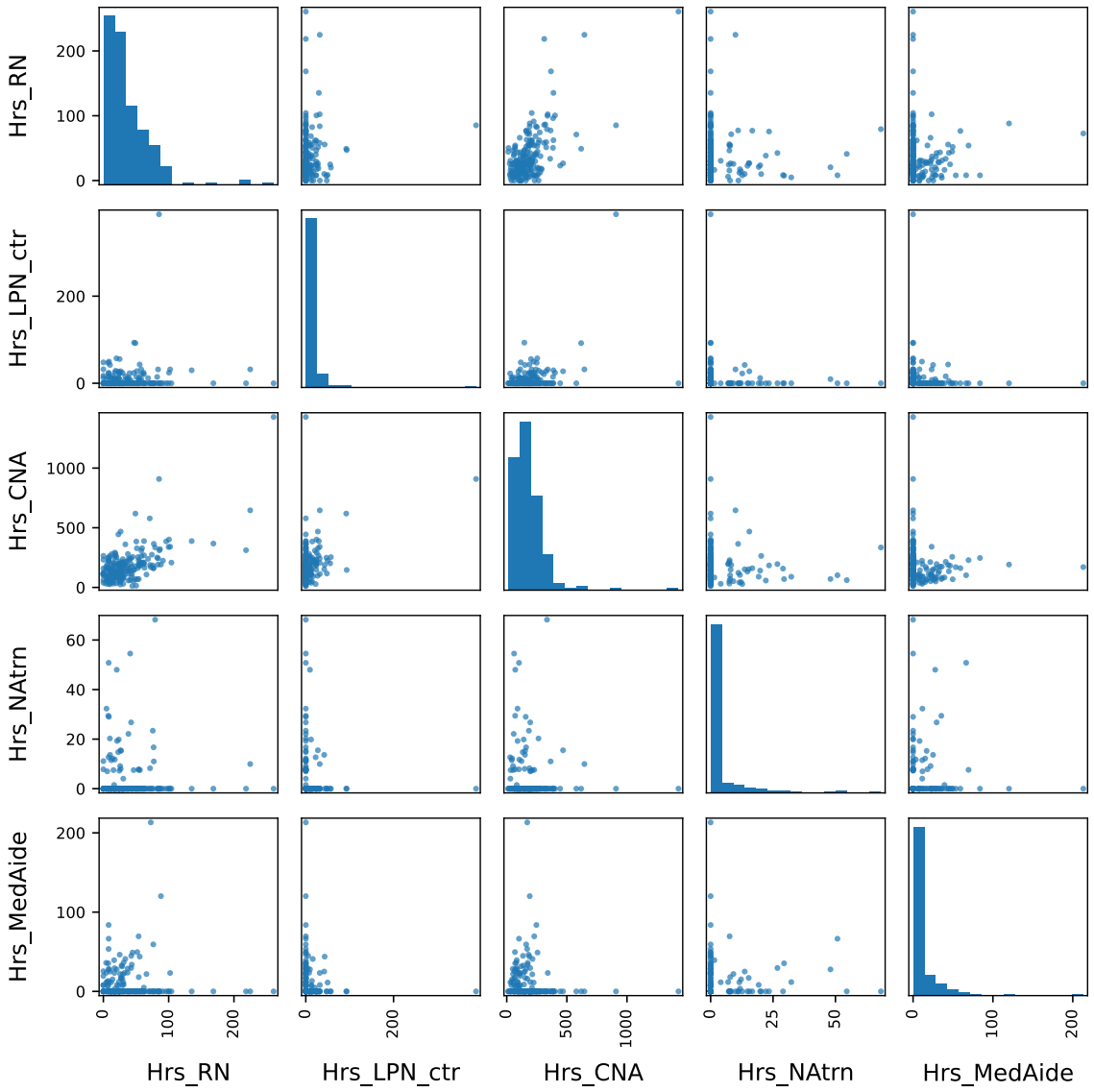

Figure 1: Scatter matrix of nursing worker working hours

0.3.c Compare groups

0.4 Feature engineer

0.4.a Join geographical data

0.4.b Join seasonal data

0.5 Analyze geography

0.6 Analyze seasonality

0.7 Model

0.8 Extra visualizations

0.8.a Sparklines

```python
# TODO: pivot on day

data_pivoted = data.pivot_table(
    index="STATE",
    columns="WorkDate",
    values="Hrs_RN",
    aggfunc='mean'
)

# Resetting the index for easier column access
# data_pivoted.reset_index(inplace=True)
data_pivoted.head()
```

WorkDate / STATE	[illegible]
AK	[illegible]
AL	[illegible]
AR	[illegible]
AZ	[illegible]
CA	[illegible]

```python
# (
#     GT(data_pivoted, rowname_col="STATE")
#     .fmt_nanoplot(
#         columns=data_pivoted.columns[1:],
#         reference_line="mean",
#         reference_area=["min", "q1"]
#     )
#     .fmt_nanoplot(
#         columns=data_pivoted.columns[1:],
#         plot_type="bar",
#         reference_line="max",
#         reference_area=["max", "median"]
#     )
# )
```

0.9 SQL

0.10 Archive

Bibliography