TRAVNUR
Improving Healthcare Through Data Mining

Verab Chitchyan

Gabriella de Asis

Emmanoel Dermkrdichyan

Carson Logston

Anh (Steven) Nguyen

TRAVNUR

DESCRIPTION OF ORIGINAL DATASET

```
sys:1: DtypeWarning: Columns (18,20,26,28,30,36,40,56,57,60,61) have mixed types.
```

```
state                                    object
date                                     object
total number of hospitals                float64
number of hospitals participated         float64
```

Data Type (df.dtypes)

```
(21748, 65)
```

Shape (df.shape)
for 01/01/2020 - 04/10/2021

```
RangeIndex: 21748 entries, 0 to 21747
Data columns (total 65 columns):
 #   Column                                      Non-Null Count   Dtype
---  ------                                      --------------   -----
 0   state                                       21746 non-null   object
 1   date                                        21746 non-null   object
 2   total number of hospitals                   0 non-null       float64
 3   number of hospitals participated            21746 non-null   float64
 4   critical_staffing_shortage_today_yes        21746 non-null   float64
 5   critical_staffing_shortage_today_no         21746 non-null   float64
```

```
28  staffed_icu_adult_patients_confirmed_and_suspected_covid                14630 non-null   object
29  staffed_icu_adult_patients_confirmed_and_suspected_covid_coverage       21746 non-null   float64
30  staffed_icu_adult_patients_confirmed_covid                              14780 non-null   object
31  staffed_icu_adult_patients_confirmed_covid_coverage                     21746 non-null   float64
32  total_adult_patients_hospitalized_confirmed_and_suspected_covid         15135 non-null   object
33  total_adult_patients_hospitalized_confirmed_and_suspected_covid_coverage 21746 non-null   float64
```

Information (df.info())

DESCRIPTION OF NEW DATASET

```
state                                                              object
date                                                               object
total number of hospitals                                         float64
number of hospitals participated                                  float64
```

Data Type (df.dtypes)

```
(14310, 62)
```

Shape (df.shape)
for 07/15/2020 - 04/10/2021

```
RangeIndex: 14310 entries, 0 to 14309
Data columns (total 62 columns):
 #   Column                                        Non-Null Count   Dtype
---  ------                                        --------------   -----
 0   state                                         14310 non-null   object
 1   date                                          14310 non-null   object
 2   number of hospitals participated              14310 non-null   int64
 3   critical_staffing_shortage_today_yes          14310 non-null   int64
 4   critical_staffing_shortage_today_no           14310 non-null   int64
 5   critical_staffing_shortage_today_not_reported 14310 non-null   int64
```

Information (df.info())

STATISTICAL DESCRIPTION

```
        previous_day_admission_pediatric_covid_suspected_coverage
count                                          14310.000000
mean                                             102.069881
std                                               94.748319
min                                                0.000000
25%                                               45.000000
50%                                               85.000000
75%                                              129.000000
max                                              592.000000
```

```
critical_staffing_shortage_today_yes
                         14310.000000
                            16.822292
                            23.367231
                             0.000000
                             3.000000
                            10.000000
                            22.000000
                           188.000000
```

Description (df.describe())

describe
argmin, argmax
idxmin, idxmax
sum
mad
prod
var
kurt
cumsum
cummin, cummax
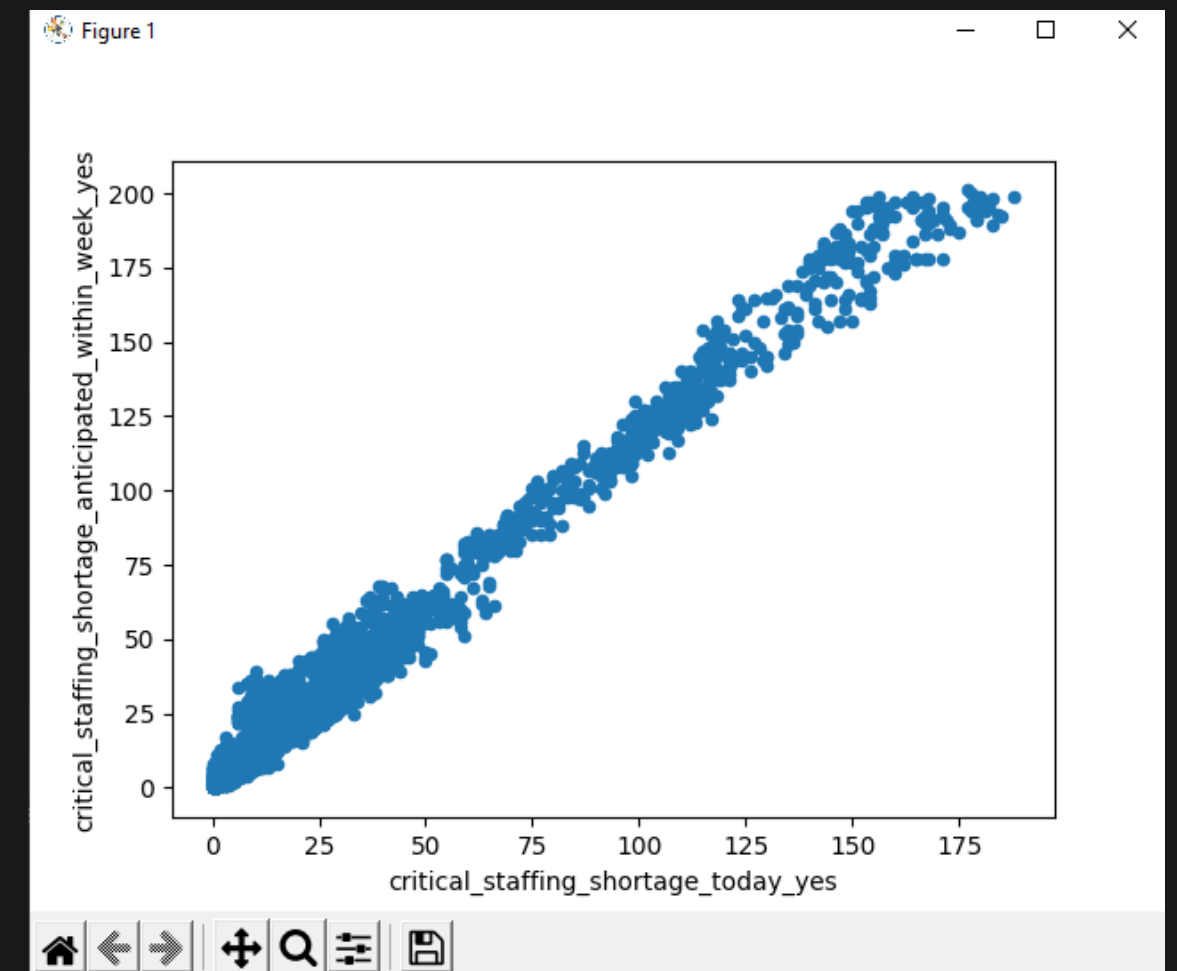cumprod
diff
pct_change

CORRELATIONS

- critical_staffing_shortage_today_yes
- critical_staffing_shortage_anticipated_within_week_yes
- total_adult_patients_hospitalized_confirmed_and_suspected_covid
- total_adult_patients_hospitalized_confirmed_covid
- total_pediatric_patients_hospitalized_confirmed_and_suspected_covid
- total_pediatric_patients_hospitalized_confirmed_covid

total_adult_patients_hospitalized_confirmed_and...	0.813178
total_pediatric_patients_hospitalized_confirmed...	0.813773
total_pediatric_patients_hospitalized_confirmed...	0.813354

Scatterplots

Using variables:
- critical_staffing_shortage_anticipated_within_week_yes (y)
- criticial staffing shortage shortage today yes (x)

- Positive correlation is shown in the graph
- Can conclude that when there is an increase in staff shortages at the hospital, the amount of anticipated staffing required increases as well.

Scatterplots

Using variables:
- critical_staffing_shortage_anticipated_within_week_yes (y)
- total_adult_patients_hospitalized_confirmed_covid (x)

- Positive correlation is shown in the graph
- Can conclude that when there is an increase in patients being hospitalized, the amount of anticipated staffing required increases as well.
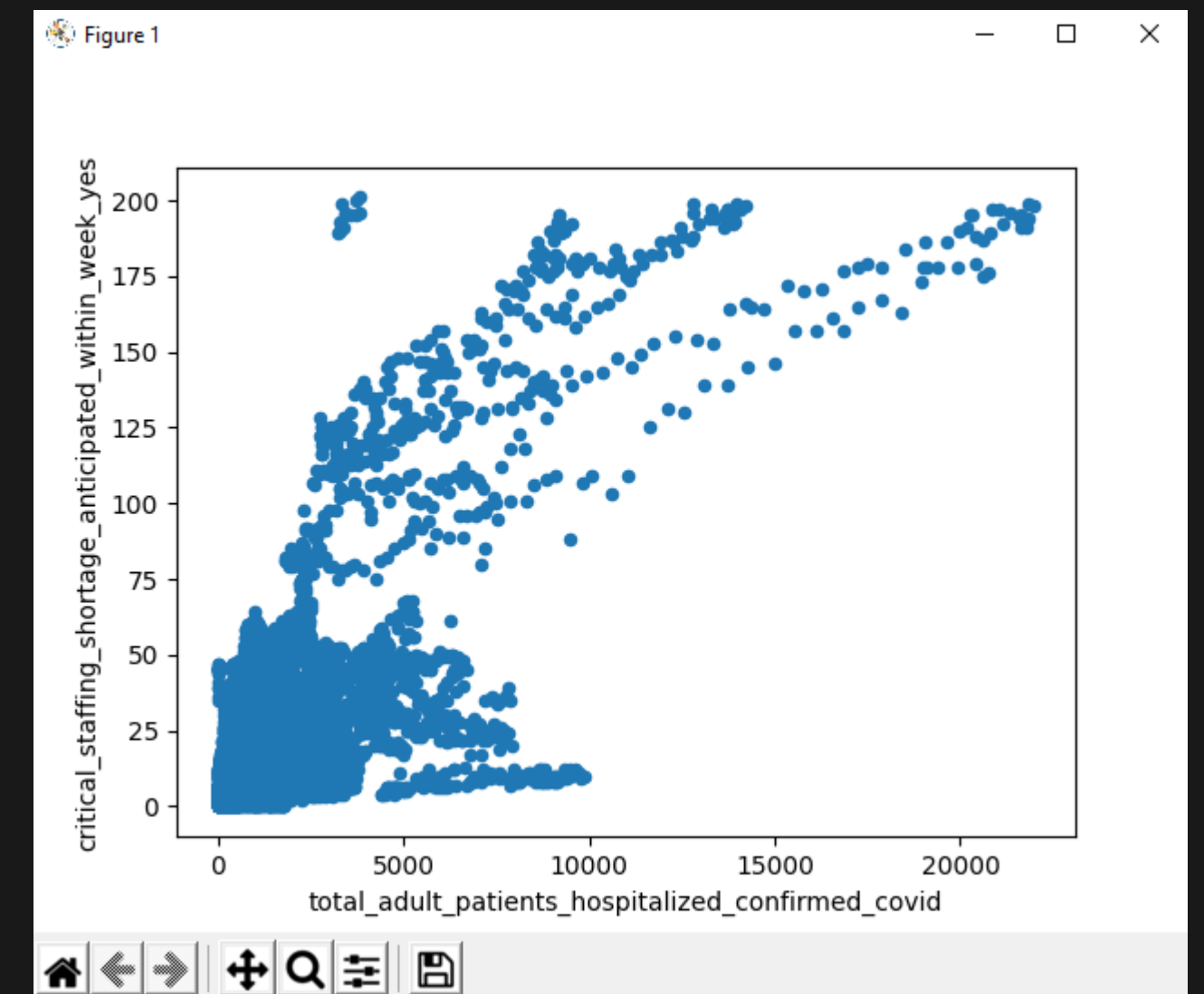
Scatterplots

Using variables:

- total_adult_patients_hospitalized_confirmed_and_suspected_covid (y)
- total_adult_patients_hospitalized_confirmed_covid (x)

- Positive correlation is shown in the graph
- The amount of patients that are suspected and confirmed versus only confirmed shows that the majority of both variables is confirmed. They almost identically line up on the chart.
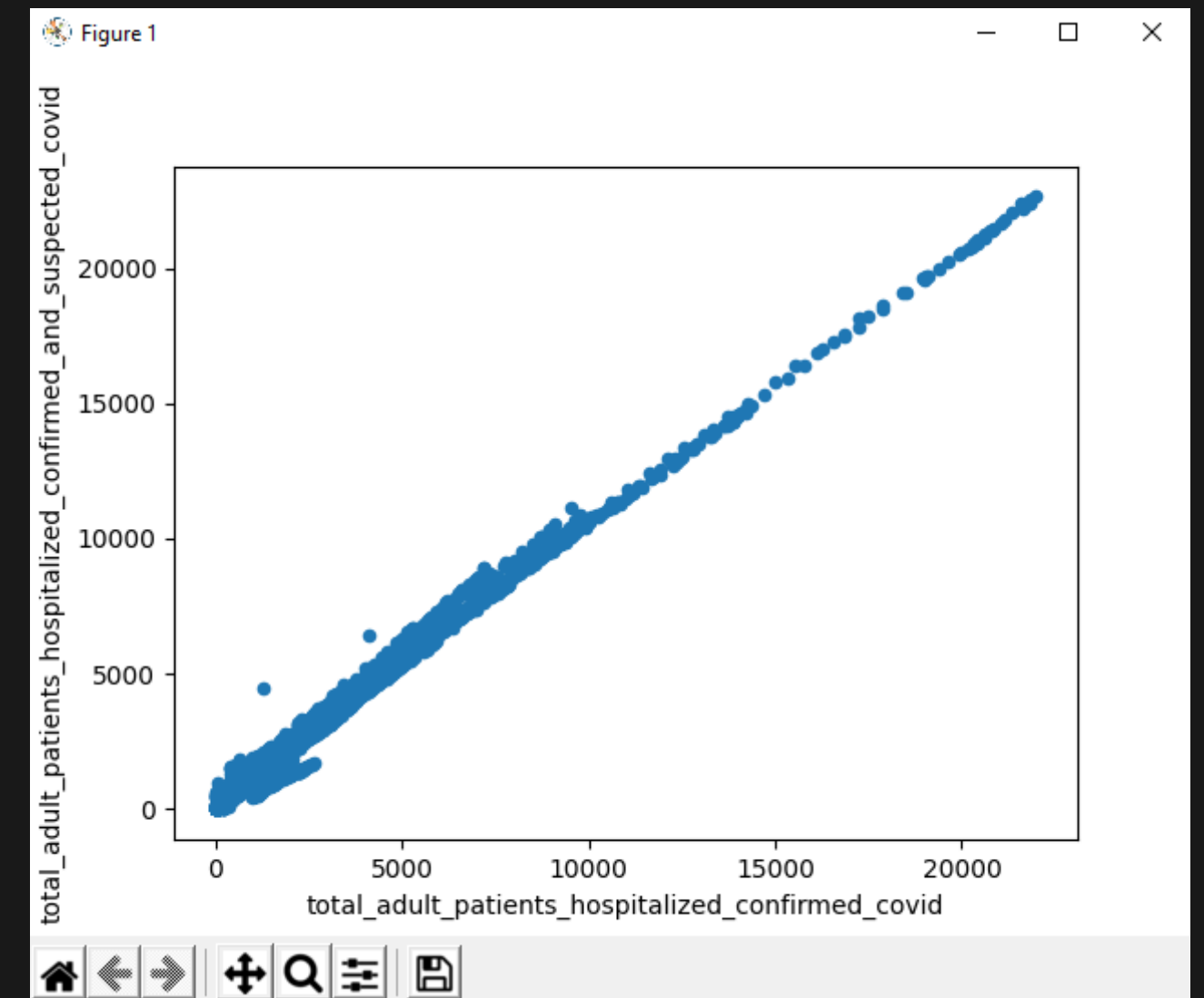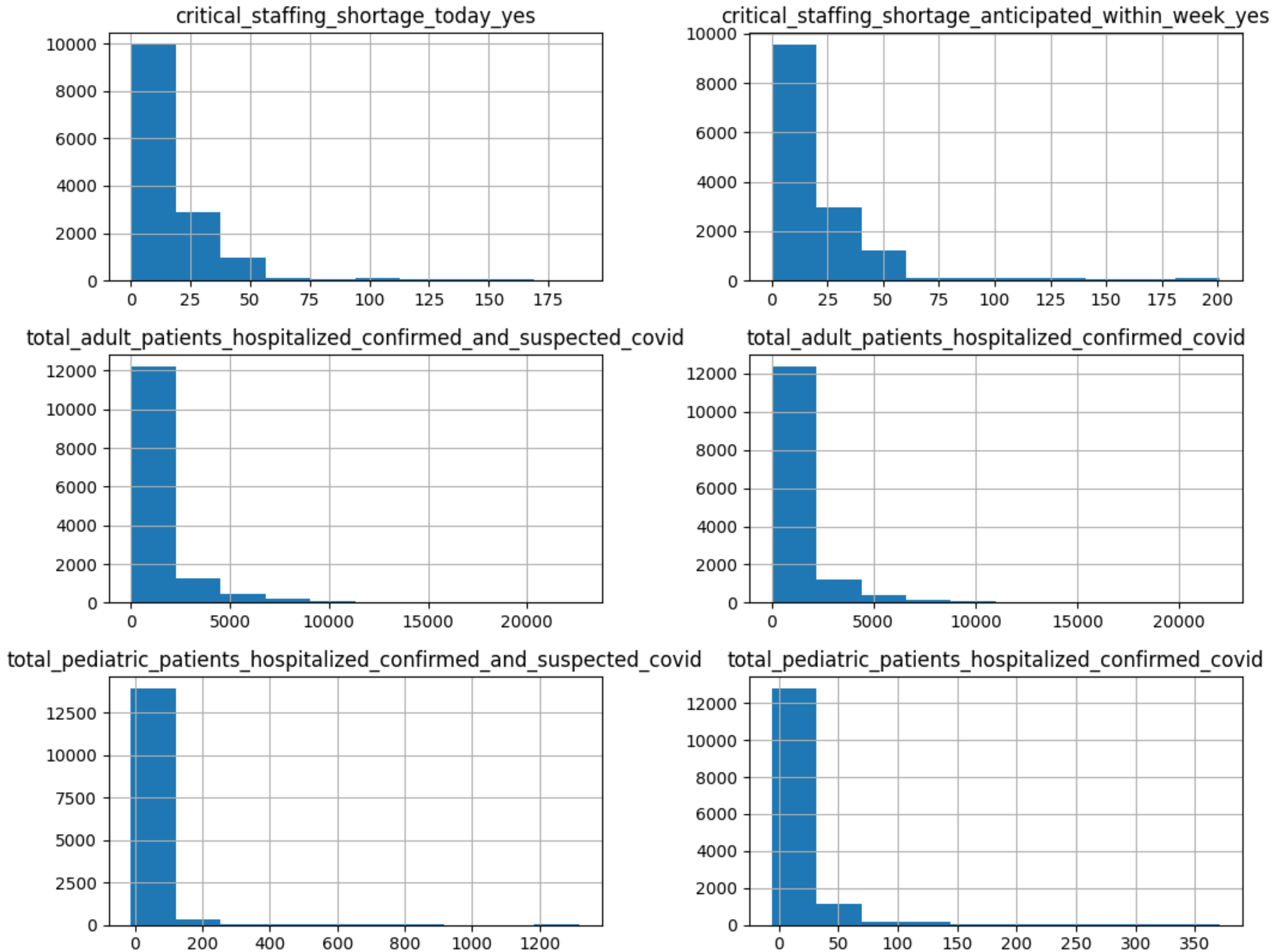
FEATURE SKEWNESS

Skewness (df.skew())

```
critical_staffing_shortage_today_yes                                    3.584149
```

```
critical_staffing_shortage_anticipated_within_week_yes                  3.510674
```

```
total_pediatric_patients_hospitalized_confirmed_and_suspected_covid_coverage   2.600400
total_pediatric_patients_hospitalized_confirmed_covid                   4.543498
total_pediatric_patients_hospitalized_confirmed_covid_coverage          2.612634
total_staffed_adult_icu_beds_coverage                                   2.630415
inpatient_beds_utilization                                             -0.104708
```

Asymmetrical and highly skewed, with the exception of patient bed utilization which is approximately symmetric
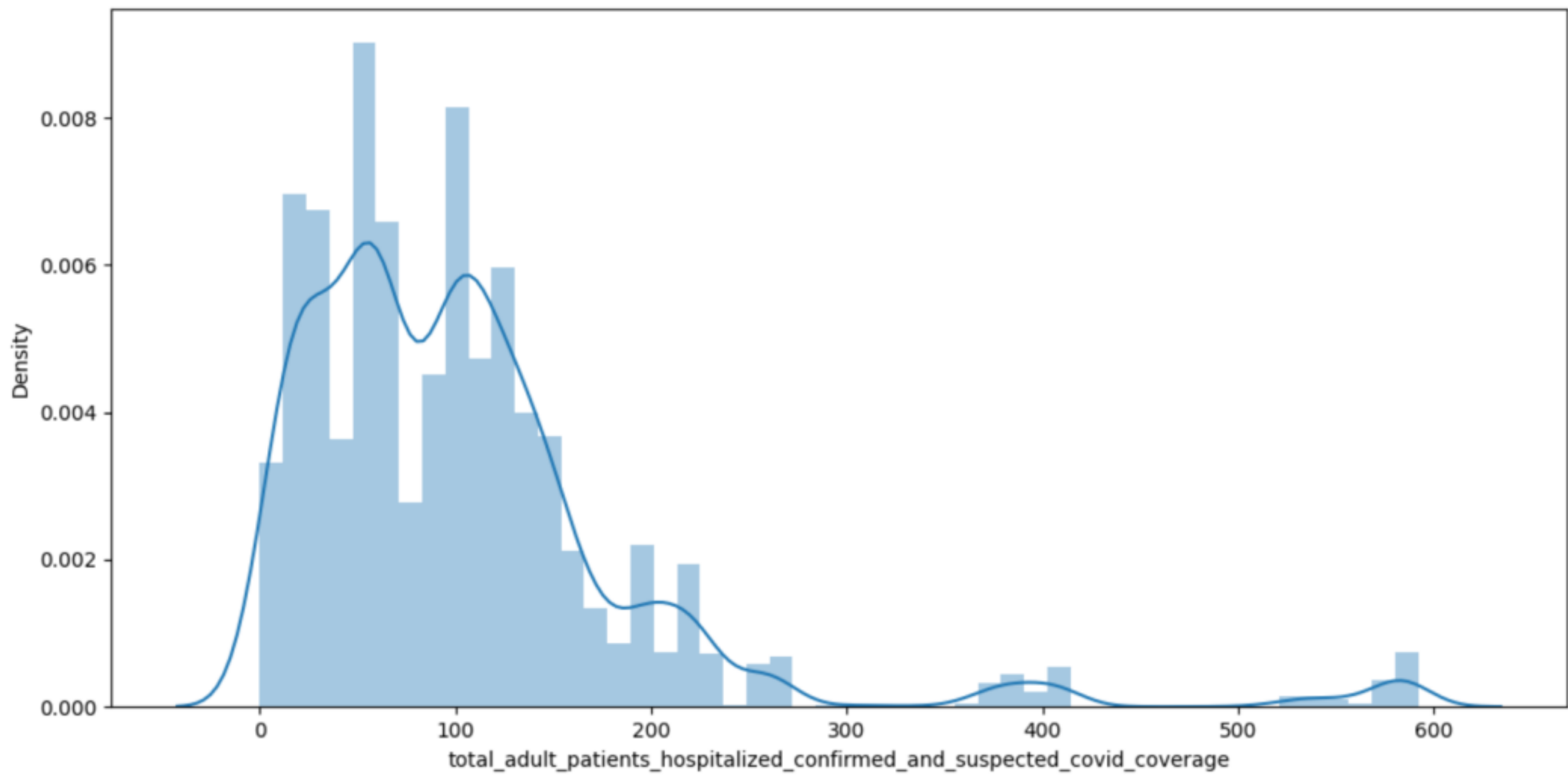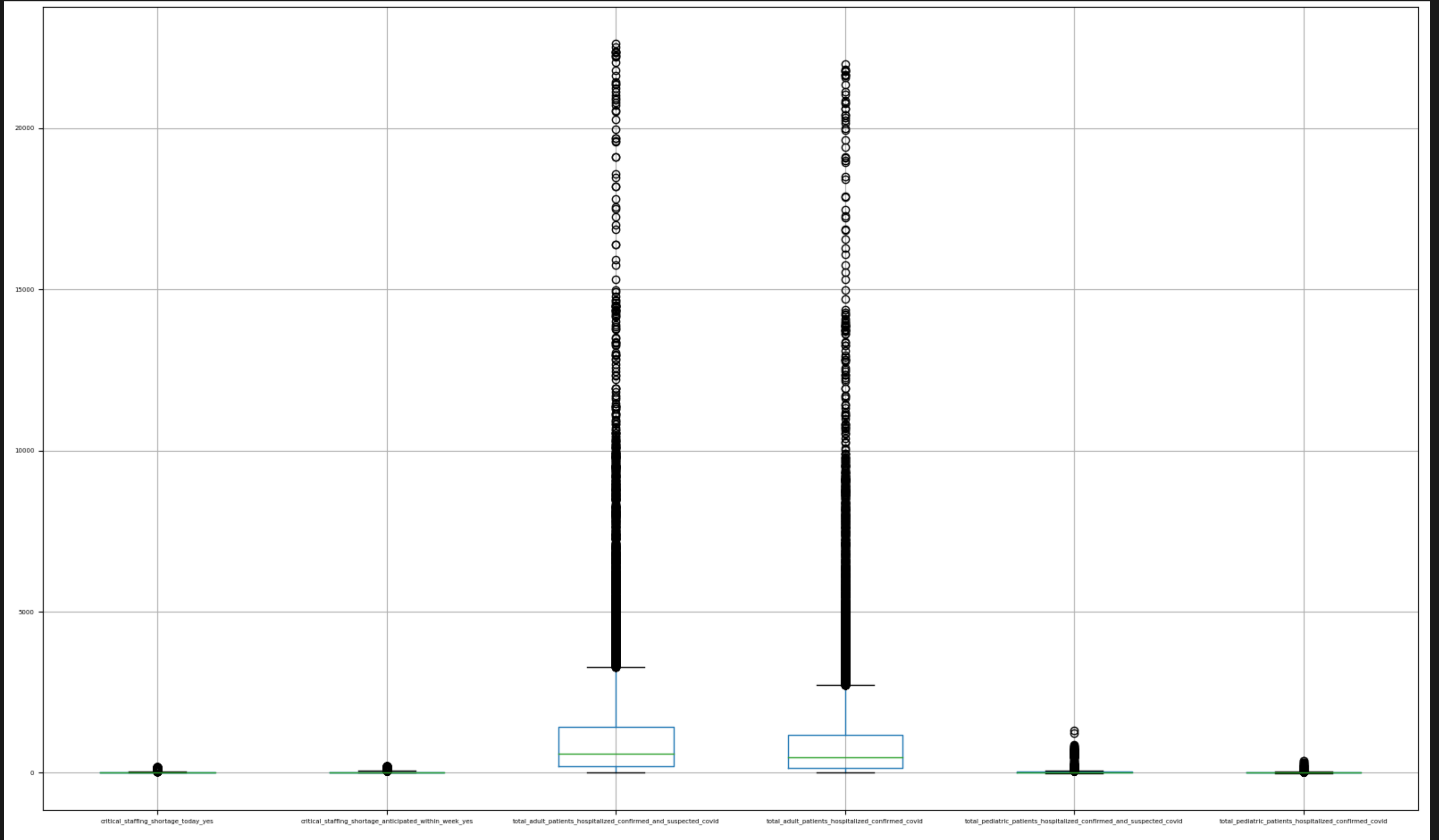
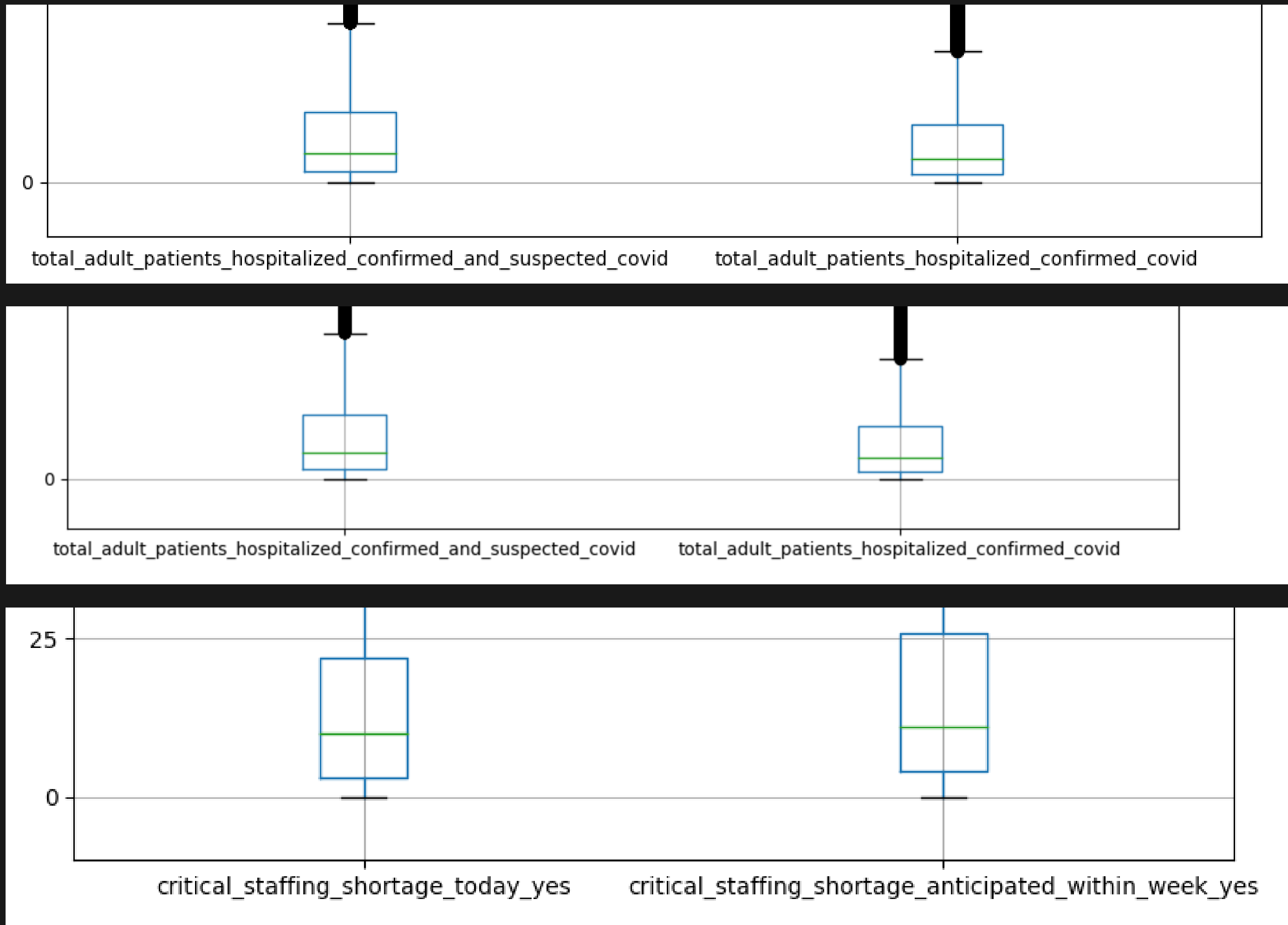**CLASS DISTRIBUTION**

df.boxplot()

CLASS DISTRIBUTION

df.boxplot()

total_adult_patients_hospitalized_confirmed_and_suspected_covid

total_adult_patients_hospitalized_confirmed_covid

total_adult_patients_hospitalized_confirmed_and_suspected_covid

total_adult_patients_hospitalized_confirmed_covid

critical_staffing_shortage_today_yes

critical_staffing_shortage_anticipated_within_week_yes

CONCLUSION

- From the dataset we have found variables with strong correlations that provide us with ample support for our company's mission

QUESTIONS?
ANSWERS?
CRIES OF OUTRAGE?