

# analysis

February 18, 2025

	provider_number	fy_start	fy_end	date_processed	date_created	beds	\
0	10001	1997-10-01	1998-09-30	2002-08-01	2001-06-27	425.0	
1	10001	2010-10-01	2011-09-30	2012-11-09	2012-11-09	408.0	
2	10001	2011-10-01	2012-09-30	2013-11-15	2013-11-15	408.0	
3	10001	2012-10-01	2013-09-30	2014-09-26	2014-03-17	408.0	
4	10001	2013-10-01	2014-09-30	2015-03-13	2015-03-11	398.0	

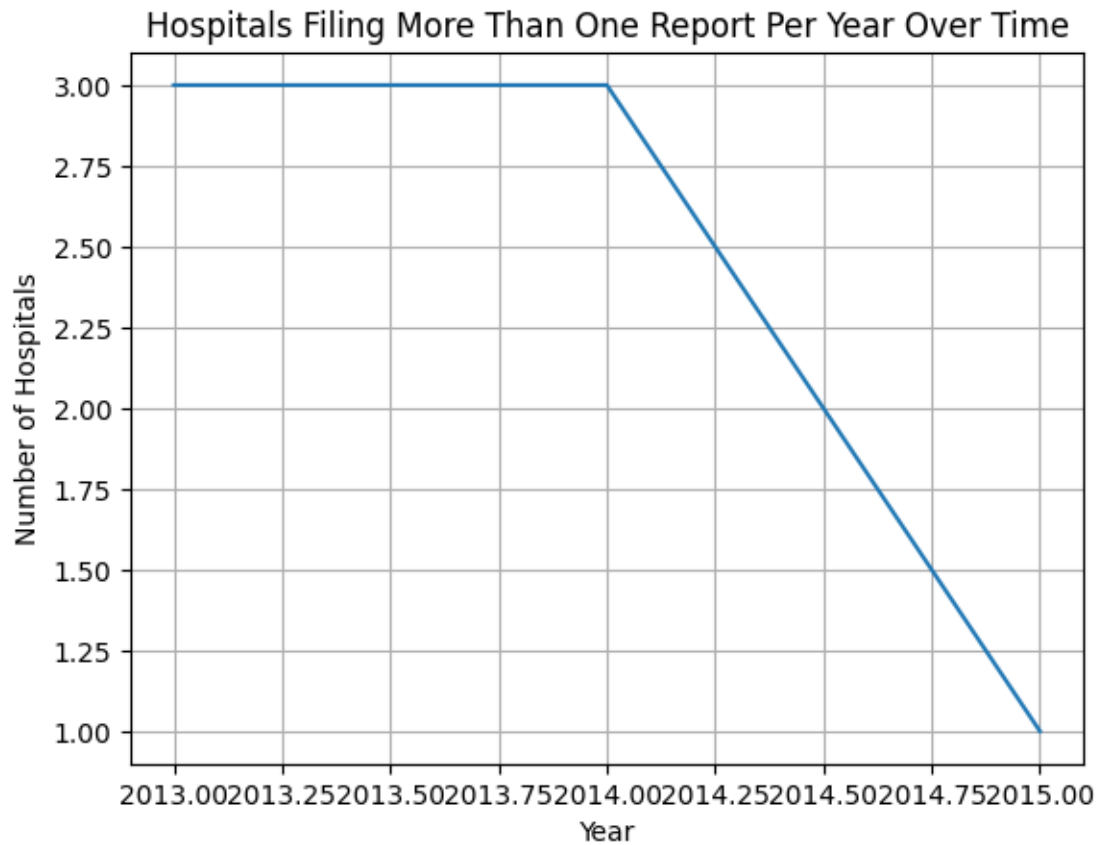
	tot_charges	tot_discounts	tot_operating_exp	ip_charges	...	\
0	3.048881e+08	1.489623e+08	133530676.0	37559906.0	...	
1	1.208332e+09	9.577269e+08	267292928.0	110484675.0	...	
2	1.263056e+09	1.008222e+09	274011907.0	116037225.0	...	
3	1.305720e+09	1.043723e+09	284850485.0	76820374.0	...	
4	1.451186e+09	1.160976e+09	313050295.0	129439329.0	...	

	secondary_mcare_payment	street	city	state	zip	\
0	60111.0	1108 ROSS CLARK CIRCLE	DOTHAN	AL	36301-	
1	35367.0	NaN	NaN	NaN	NaN	
2	55155.0	NaN	NaN	NaN	NaN	
3	58522.0	NaN	NaN	NaN	NaN	
4	56453.0	NaN	NaN	NaN	NaN	

	county	hvpb_payment	hrrp_payment	year	source
0	HOUSTON	NaN	NaN	1998	unique reports
1	NaN	NaN	NaN	2011	unique reports
2	NaN	NaN	NaN	2012	unique reports
3	NaN	-101118.0	NaN	2013	unique reports
4	NaN	68134.0	NaN	2014	unique reports

[5 rows x 26 columns]

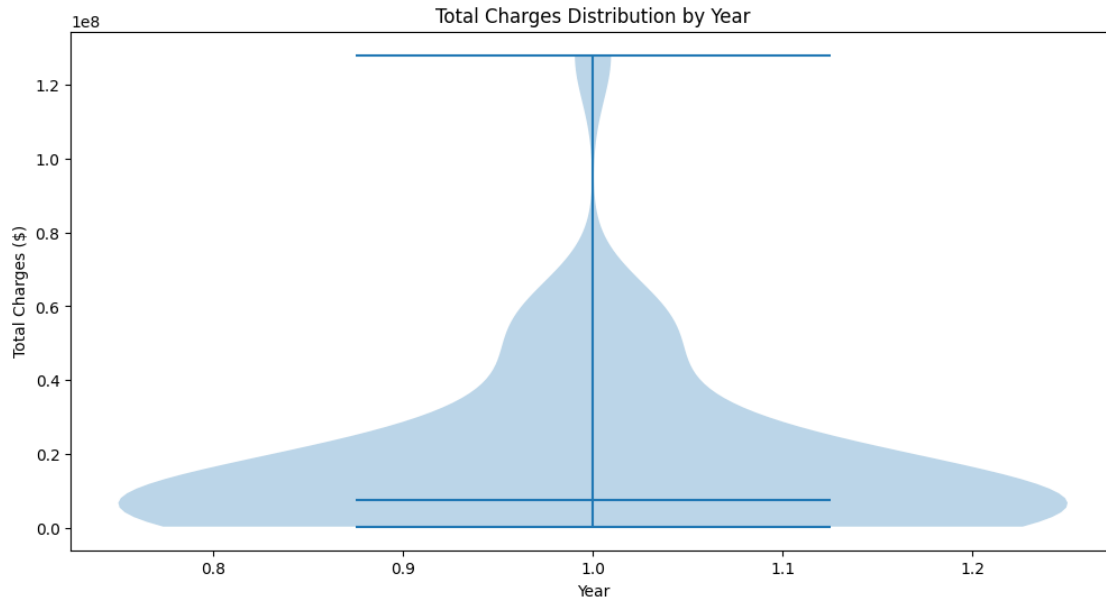
1. How many hospitals filed more than one report in the same year? Show your answer as a line graph of the number of hospitals over time.



2. After removing/combining multiple reports, how many unique hospital IDs (Medicare provider numbers) exist in the data?

Total Unique Hospitals: 8812

3. What is the distribution of total charges (tot\_charges in the data) in each year? Show your results with a “violin” plot, with charges on the y-axis and years on the x-axis.



4. What is the distribution of estimated prices in each year? Again present your results with a violin plot, and recall our formula for estimating prices from class. Be sure to do something about outliers and/or negative prices in the data.

```
-----
ValueError                                Traceback (most recent call last)
Cell In[15], line 26
    24 # Plot
    25 plt.figure()
--> 26 plt.violinplot(estimated_price_df, showmedians=True)
    27 plt.xticks(range(1, len(ep_years) + 1), ep_years, rotation=45)
    28 plt.xlabel('Year')

File ~/Library/Frameworks/Python.framework/Versions/3.11/lib/python3.11/
site-packages/matplotlib/pyplot.py:2981, in violinplot(dataset, positions,
vert, widths, showmeans, showextrema, showmedians, quantiles, points,
bw_method, data)
    2976 @_copy_docstring_and_deprecators(Axes.violinplot)
    2977 def violinplot(
    2978     dataset, positions=None, vert=True, widths=0.5,
    2979     showmeans=False, showextrema=True, showmedians=False,
    2980     quantiles=None, points=100, bw_method=None, *, data=None):
-> 2981     return gca().violinplot(
    2982         dataset, positions=positions, vert=vert, widths=widths,
    2983         showmeans=showmeans, showextrema=showextrema,
    2984         showmedians=showmedians, quantiles=quantiles, points=points,
    2985         bw_method=bw_method,
    2986         **({"data": data} if data is not None else {}))
```

```
File /Library/Frameworks/Python.framework/Versions/3.11/lib/python3.11/
↳site-packages/matplotlib/__init__.py:1433, in _preprocess_data.<locals>.
```

```
↳inner(ax, data, *args, **kwargs)
    1430 @functools.wraps(func)
    1431 def inner(ax, *args, data=None, **kwargs):
    1432     if data is None:
-> 1433         return func(ax, *map(sanitize_sequence, args), **kwargs)
    1435     bound = new_sig.bind(ax, *args, **kwargs)
    1436     auto_label = (bound.arguments.get(label_namer)
    1437                  or bound.kwargs.get(label_namer))
```

```
File /Library/Frameworks/Python.framework/Versions/3.11/lib/python3.11/
↳site-packages/matplotlib/axes/_axes.py:8043, in Axes.violinplot(self, dataset,
↳positions, vert, widths, showmeans, showextrema, showmedians, quantiles,
↳points, bw_method)
```

```
    8040     kde = mlab.GaussianKDE(X, bw_method)
    8041     return kde.evaluate(coords)
-> 8043 vpstats = cbook.violin_stats(dataset, _kde_method, points=points,
    8044                             quantiles=quantiles)
    8045 return self.violin(vpstats, positions=positions, vert=vert,
    8046                  widths=widths, showmeans=showmeans,
    8047                  showextrema=showextrema, showmedians=showmedians)
```

```
File /Library/Frameworks/Python.framework/Versions/3.11/lib/python3.11/
↳site-packages/matplotlib/cbook/__init__.py:1523, in violin_stats(X, method,
↳points, quantiles)
```

```
    1520 stats = {}
    1522 # Calculate basic stats for the distribution
-> 1523 min_val = np.min(x)
    1524 max_val = np.max(x)
    1525 quantile_val = np.percentile(x, 100 * q)
```

```
File <__array_function__ internals>:200, in amin(*args, **kwargs)
```

```
File /Library/Frameworks/Python.framework/Versions/3.11/lib/python3.11/
↳site-packages/numpy/core/fromnumeric.py:2946, in amin(a, axis, out, keepdims,
↳initial, where)
```

```
    2829 @array_function_dispatch(_amin_dispatcher)
    2830 def amin(a, axis=None, out=None, keepdims=np._NoValue, initial=np.
↳_NoValue,
    2831         where=np._NoValue):
    2832     """
    2833     Return the minimum of an array or minimum along an axis.
    2834     (...)
    2944     6
    2945     """
-> 2946     return _wrapreduction(a, np.minimum, 'min', axis, None, out,
```

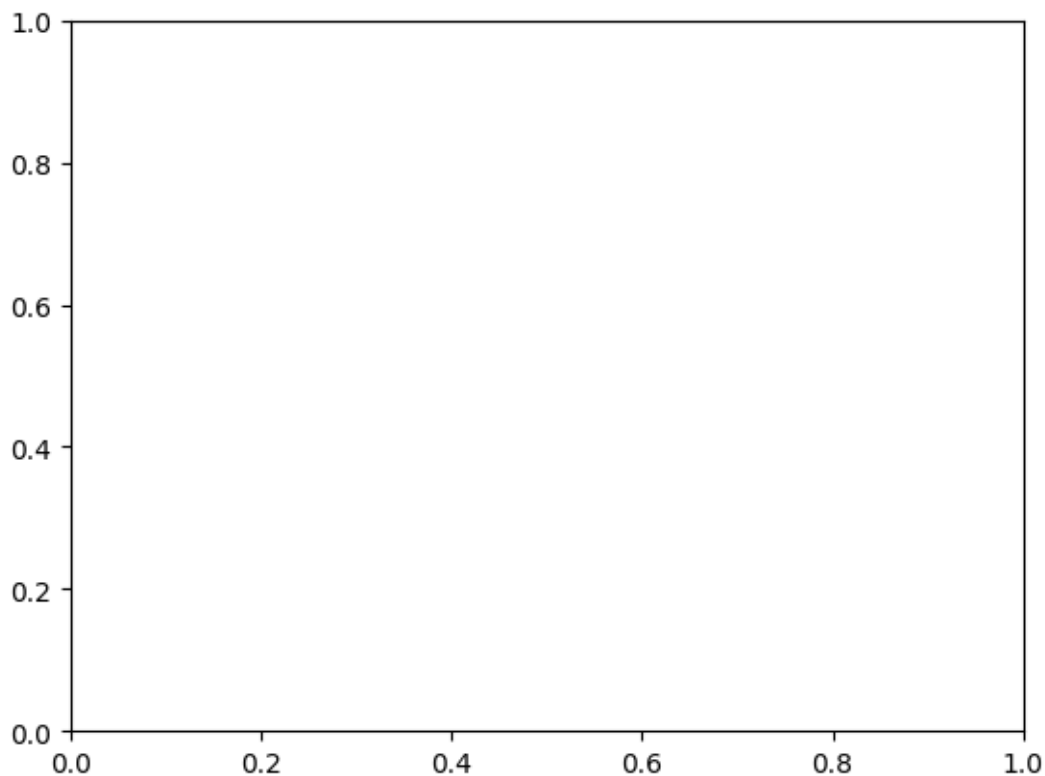
```

2947
↳ keepdims=keepdims, initial=initial, where=where)

File ~/Library/Frameworks/Python.framework/Versions/3.11/lib/python3.11/
↳ site-packages/numpy/core/fromnumeric.py:86, in _wrapreduction(obj, ufunc,
↳ method, axis, dtype, out, **kwargs)
    83         else:
    84             return reduction(axis=axis, out=out, **passkwargs)
---> 86 return ufunc.reduce(obj, axis, dtype, out, **passkwargs)

ValueError: zero-size array to reduction operation minimum which has no identity

```



5. Calculate the average price among penalized versus non-penalized hospitals.

```

Cell In[7], line 5
    hcris_2012 = final_hcris[(finally['year'] == 2012) & (hcris_data['beds'] >
↳ 30)].copy()

SyntaxError: invalid syntax

```

6. Split hospitals into quartiles based on bed size. To do this, create 4 new indicator variables,

where each variable is set to 1 if the hospital's bed size falls into the relevant quartile. Provide a table of the average price among treated/control groups for each quartile.

7. Find the average treatment effect using each of the following estimators, and present your results in a single table:

Nearest neighbor matching (1-to-1) with inverse variance distance based on quartiles of bed size  
Nearest neighbor matching (1-to-1) with Mahalanobis distance based on quartiles of bed size  
Inverse propensity weighting, where the propensity scores are based on quartiles of bed size  
Simple linear regression, adjusting for quartiles of bed size using dummy variables and appropriate interactions as discussed in class

8. With these different treatment effect estimators, are the results similar, identical, very different?
9. Do you think you've estimated a causal effect of the penalty? Why or why not? (just a couple of sentences)
10. Briefly describe your experience working with these data (just a few sentences). Tell me one thing you learned and one thing that really aggravated or surprised you.

```
Cell In[19], line 2
```

```
jupyter nbconvert --to pdf --no-input --output homework2.pdf "analysis.ipyn"
```

```
SyntaxError: invalid syntax
```