CS2316 Phase 3 Presentation - Nursing Home Penalties & Community Economics

By: Aditya Gutha and Anish Sannareddy

Point of Interest and Purpose

Why we chose this topic:

- Nursing homes play a major role in community health, yet many receive fines for failing to meet care and safety standards.
- We wanted to understand whether facilities in lower-income or economically vulnerable areas tend to receive more penalties.

What we expected to find:

- We predicted that lower-income areas and lower wages might be associated with more frequent or more severe penalties.
- We also expected penalties to cluster geographically rather than being evenly distributed.

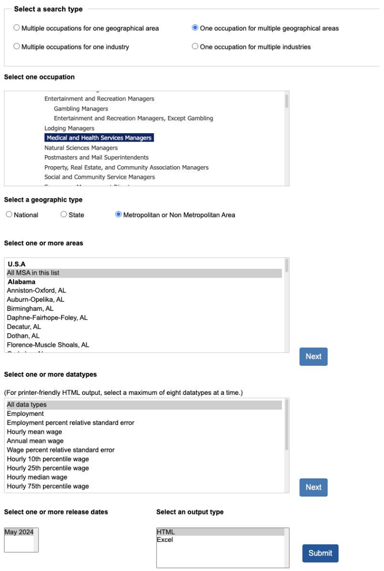
Datasets and Data Collection

Datasets used:

1. CMS Nursing Home Penalties (fines + payment denials)
2. BLS OEWS Wage Data for Medical & Health Services Managers
3. U.S. Census ACS API: poverty rate, education, housing indicators

How we collected them:

- CMS dataset downloaded directly from data.cms.gov.
- Wage data scraped using Selenium because BLS loads dynamically.
- Census indicators pulled through an API request with specified keys.



Select a search type

○ Multiple occupations for one geographical area ● One occupation for multiple geographical areas

○ Multiple occupations for one industry ○ One occupation for multiple industries

Select one occupation

Entertainment and Recreation Managers
Gambling Managers
Entertainment and Recreation Managers, Except Gambling
Lodging Managers
Medical and Health Services Managers
Natural Sciences Managers
Postmasters and Mail Superintendents
Property, Real Estate, and Community Association Managers
Social and Community Service Managers

Select a geographic type

○ National ○ State ● Metropolitan or Non Metropolitan Area

Select one or more areas

U.S.A
All MSA in this list
Alabama
Anniston-Oxford, AL
Auburn-Opelika, AL
Birmingham, AL
Daphne-Fairhope-Foley, AL
Decatur, AL
Dothan, AL
Florence-Muscle Shoals, AL

Select one or more datatypes

(For printer-friendly HTML output, select a maximum of eight datatypes at a time.)

All data types
Employment
Employment percent relative standard error
Hourly mean wage
Annual mean wage
Wage percent relative standard error
Hourly 10th percentile wage
Hourly 25th percentile wage
Hourly median wage
Hourly 75th percentile wage

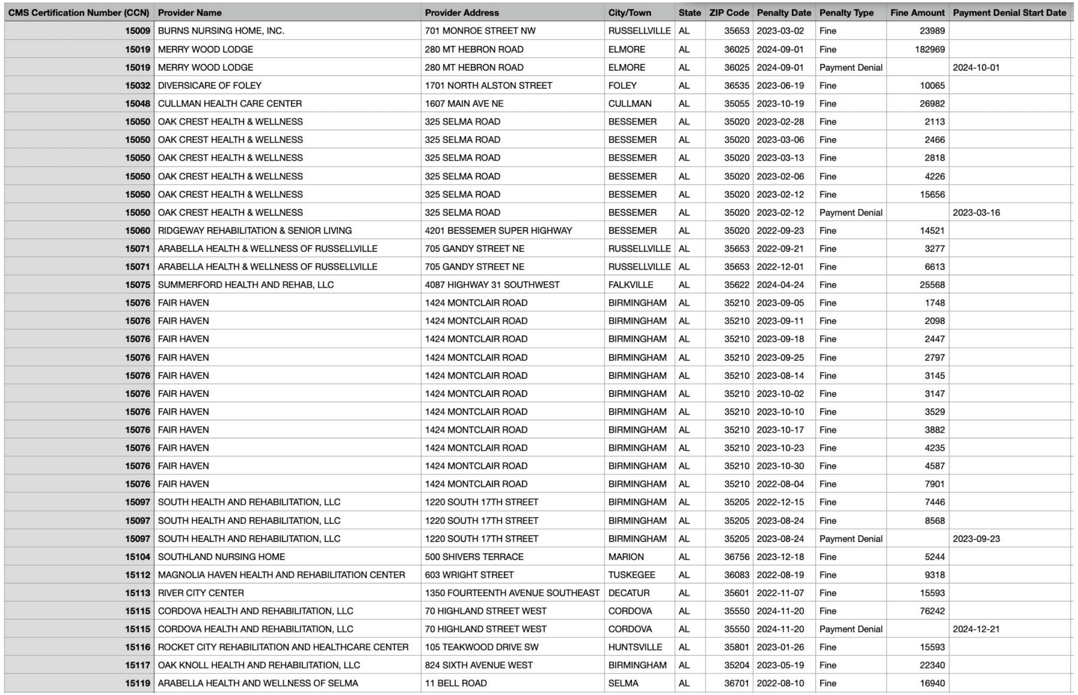
Select one or more release dates

May 2024

Select an output type

HTML
Excel

Next
Next
Submit



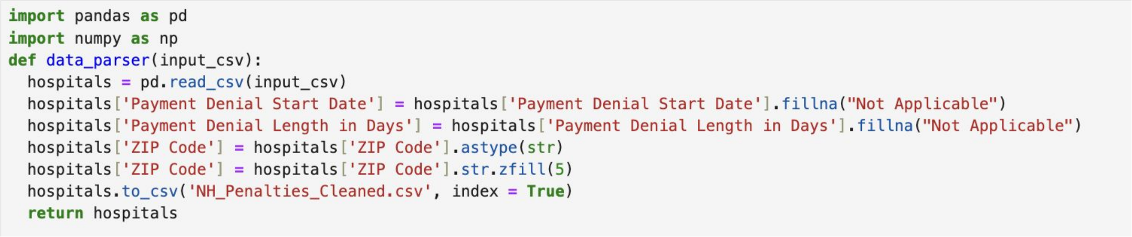
CMS Certification Number (CCN)	Provider Name	Provider Address	City/Town	State	ZIP Code	Penalty Date	Penalty Type	Fine Amount	Payment Denial Start Date
15009	BURNS NURSING HOME, INC.	701 MONROE STREET NW	RUSSELLVILLE	AL	35653	2023-03-02	Fine	23989	
15019	MERRY WOOD LODGE	280 MT HEBRON ROAD	ELMORE	AL	36025	2024-09-01	Fine	182969	
15019	MERRY WOOD LODGE	280 MT HEBRON ROAD	ELMORE	AL	36025	2024-09-01	Payment Denial		2024-10-01
15032	DIVERSICARE OF FOLEY	1701 NORTH ALSTON STREET	FOLEY	AL	36535	2023-06-19	Fine	10065	
15048	CULLMAN HEALTH CARE CENTER	1607 MAIN AVE NE	CULLMAN	AL	35055	2023-10-19	Fine	26982	
15050	OAK CREST HEALTH & WELLNESS	325 SELMA ROAD	BESSEMER	AL	35020	2023-02-28	Fine	2113	
15050	OAK CREST HEALTH & WELLNESS	325 SELMA ROAD	BESSEMER	AL	35020	2023-03-06	Fine	2466	
15050	OAK CREST HEALTH & WELLNESS	325 SELMA ROAD	BESSEMER	AL	35020	2023-03-13	Fine	2818	
15050	OAK CREST HEALTH & WELLNESS	325 SELMA ROAD	BESSEMER	AL	35020	2023-02-06	Fine	4226	
15050	OAK CREST HEALTH & WELLNESS	325 SELMA ROAD	BESSEMER	AL	35020	2023-02-12	Fine	15656	
15050	OAK CREST HEALTH & WELLNESS	325 SELMA ROAD	BESSEMER	AL	35020	2023-02-12	Payment Denial		2023-03-16
15060	RIDGEWAY REHABILITATION & SENIOR LIVING	4201 BESSEMER SUPER HIGHWAY	BESSEMER	AL	35020	2022-09-23	Fine	14521	
15071	ARABELLA HEALTH & WELLNESS OF RUSSELLVILLE	705 GANDY STREET NE	RUSSELLVILLE	AL	35653	2022-09-21	Fine	3277	
15071	ARABELLA HEALTH & WELLNESS OF RUSSELLVILLE	705 GANDY STREET NE	RUSSELLVILLE	AL	35653	2022-12-01	Fine	6613	
15075	SUMMERFORD HEALTH AND REHAB, LLC	4087 HIGHWAY 31 SOUTHWEST	FALKVILLE	AL	35622	2024-04-24	Fine	25568	
15076	FAIR HAVEN	1424 MONTCLAIR ROAD	BIRMINGHAM	AL	35210	2023-09-05	Fine	1748	
15076	FAIR HAVEN	1424 MONTCLAIR ROAD	BIRMINGHAM	AL	35210	2023-09-11	Fine	2098	
15076	FAIR HAVEN	1424 MONTCLAIR ROAD	BIRMINGHAM	AL	35210	2023-09-18	Fine	2447	
15076	FAIR HAVEN	1424 MONTCLAIR ROAD	BIRMINGHAM	AL	35210	2023-09-25	Fine	2797	
15076	FAIR HAVEN	1424 MONTCLAIR ROAD	BIRMINGHAM	AL	35210	2023-08-14	Fine	3145	
15076	FAIR HAVEN	1424 MONTCLAIR ROAD	BIRMINGHAM	AL	35210	2023-10-02	Fine	3147	
15076	FAIR HAVEN	1424 MONTCLAIR ROAD	BIRMINGHAM	AL	35210	2023-10-10	Fine	3529	
15076	FAIR HAVEN	1424 MONTCLAIR ROAD	BIRMINGHAM	AL	35210	2023-10-17	Fine	3882	
15076	FAIR HAVEN	1424 MONTCLAIR ROAD	BIRMINGHAM	AL	35210	2023-10-23	Fine	4235	
15076	FAIR HAVEN	1424 MONTCLAIR ROAD	BIRMINGHAM	AL	35210	2023-10-30	Fine	4587	
15076	FAIR HAVEN	1424 MONTCLAIR ROAD	BIRMINGHAM	AL	35210	2022-08-04	Fine	7901	
15097	SOUTH HEALTH AND REHABILITATION, LLC	1220 SOUTH 17TH STREET	BIRMINGHAM	AL	35205	2022-12-15	Fine	7446	
15097	SOUTH HEALTH AND REHABILITATION, LLC	1220 SOUTH 17TH STREET	BIRMINGHAM	AL	35205	2023-08-24	Fine	8568	
15097	SOUTH HEALTH AND REHABILITATION, LLC	1220 SOUTH 17TH STREET	BIRMINGHAM	AL	35205	2023-08-24	Payment Denial		2023-09-23
15104	SOUTHLAND NURSING HOME	500 SHIVERS TERRACE	MARION	AL	36756	2023-12-18	Fine	5244	
15112	MAGNOLIA HAVEN HEALTH AND REHABILITATION CENTER	603 WRIGHT STREET	TUSKEGEE	AL	36083	2022-08-19	Fine	9318	
15113	RIVER CITY CENTER	1350 FOURTEENTH AVENUE SOUTHEAST	DECATUR	AL	35601	2022-11-07	Fine	15593	
15115	CORDOVA HEALTH AND REHABILITATION, LLC	70 HIGHLAND STREET WEST	CORDOVA	AL	35550	2024-11-20	Fine	76242	
15115	CORDOVA HEALTH AND REHABILITATION, LLC	70 HIGHLAND STREET WEST	CORDOVA	AL	35550	2024-11-20	Payment Denial		2024-12-21
15116	ROCKET CITY REHABILITATION AND HEALTHCARE CENTER	105 TEAKWOOD DRIVE SW	HUNTSVILLE	AL	35801	2023-01-26	Fine	15593	
15117	OAK KNOLL HEALTH AND REHABILITATION, LLC	824 SIXTH AVENUE WEST	BIRMINGHAM	AL	35204	2023-05-19	Fine	22340	
15119	ARABELLA HEALTH AND WELLNESS OF SELMA	11 BELL ROAD	SELMA	AL	36701	2022-08-10	Fine	16940	

Data Cleaning

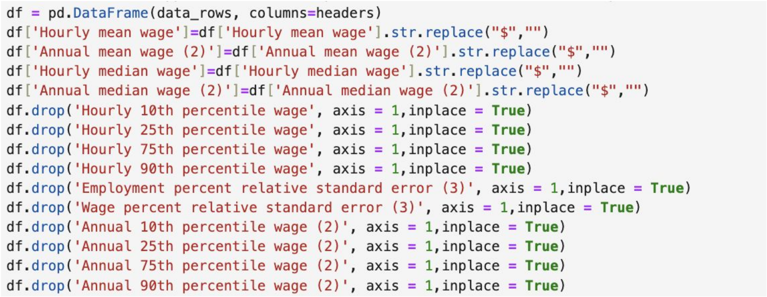
Cleaning steps:

- Removed missing values and filled non-applicable entries with “N/A” or 0.
- Stripped “$”, commas, and special characters from numeric fields.
- Correct zip codes since Python sometimes interprets those as numbers and strips the leading digit.
- Aggregated penalties by state and city so all three datasets could merge cleanly.

Zip Code Fix



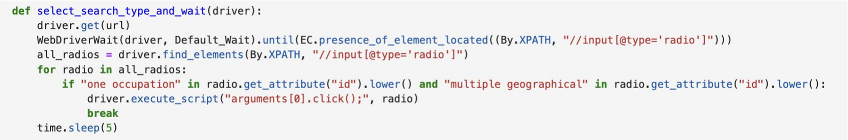
```
import pandas as pd
import numpy as np
def data_parser(input_csv):
  hospitals = pd.read_csv(input_csv)
  hospitals['Payment Denial Start Date'] = hospitals['Payment Denial Start Date'].fillna("Not Applicable")
  hospitals['Payment Denial Length in Days'] = hospitals['Payment Denial Length in Days'].fillna("Not Applicable")
  hospitals['ZIP Code'] = hospitals['ZIP Code'].astype(str)
  hospitals['ZIP Code'] = hospitals['ZIP Code'].str.zfill(5)
  hospitals.to_csv('NH_Penalties_Cleaned.csv', index = True)
  return hospitals
```



```
df = pd.DataFrame(data_rows, columns=headers)
df['Hourly mean wage']=df['Hourly mean wage'].str.replace("$","")
df['Annual mean wage (2)']=df['Annual mean wage (2)'].str.replace("$","")
df['Hourly median wage']=df['Hourly median wage'].str.replace("$","")
df['Annual median wage (2)']=df['Annual median wage (2)'].str.replace("$","")
df.drop('Hourly 10th percentile wage', axis = 1,inplace = True)
df.drop('Hourly 25th percentile wage', axis = 1,inplace = True)
df.drop('Hourly 75th percentile wage', axis = 1,inplace = True)
df.drop('Hourly 90th percentile wage', axis = 1,inplace = True)
df.drop('Employment percent relative standard error (3)', axis = 1,inplace = True)
df.drop('Wage percent relative standard error (3)', axis = 1,inplace = True)
df.drop('Annual 10th percentile wage (2)', axis = 1,inplace = True)
df.drop('Annual 25th percentile wage (2)', axis = 1,inplace = True)
df.drop('Annual 75th percentile wage (2)', axis = 1,inplace = True)
df.drop('Annual 90th percentile wage (2)', axis = 1,inplace = True)
```

Removing “$” symbol

Selenium to select button



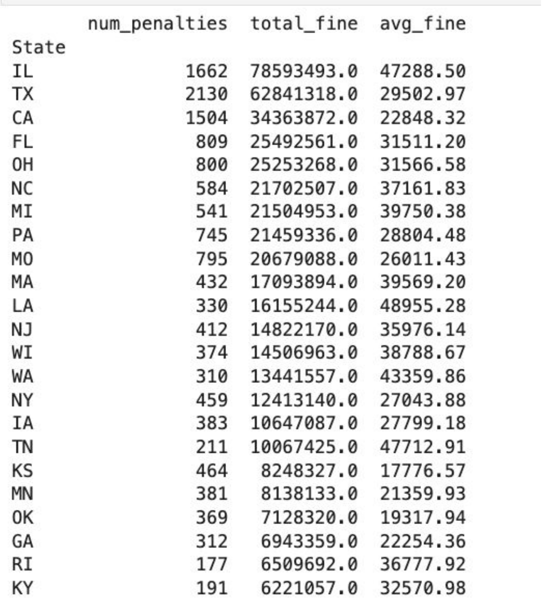
```
def select_search_type_and_wait(driver):
    driver.get(url)
    WebDriverWait(driver, Default_Wait).until(EC.presence_of_element_located((By.XPATH, "//input[@type='radio']")))
    all_radios = driver.find_elements(By.XPATH, "//input[@type='radio']")
    for radio in all_radios:
        if "one occupation" in radio.get_attribute("id").lower() and "multiple geographical" in radio.get_attribute("id").lower():
            driver.execute_script("arguments[0].click();", radio)
            break
    time.sleep(5)
```

Data Analysis Overview

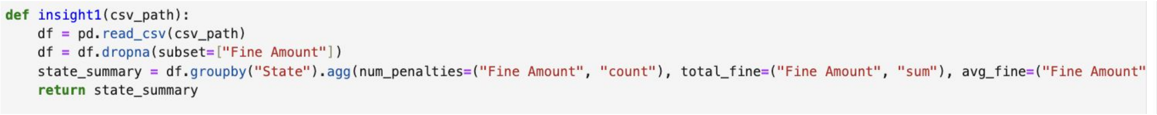
Methods used:

- Group-by summaries (penalties per state, per provider, per city).
- State-level clustering using K-Means to identify risk groups.
- Linear regression using poverty rate to predict city-level penalty counts.
- Comparison of Illinois city penalties with wage data from BLS.

Insight 1: Penalties by State

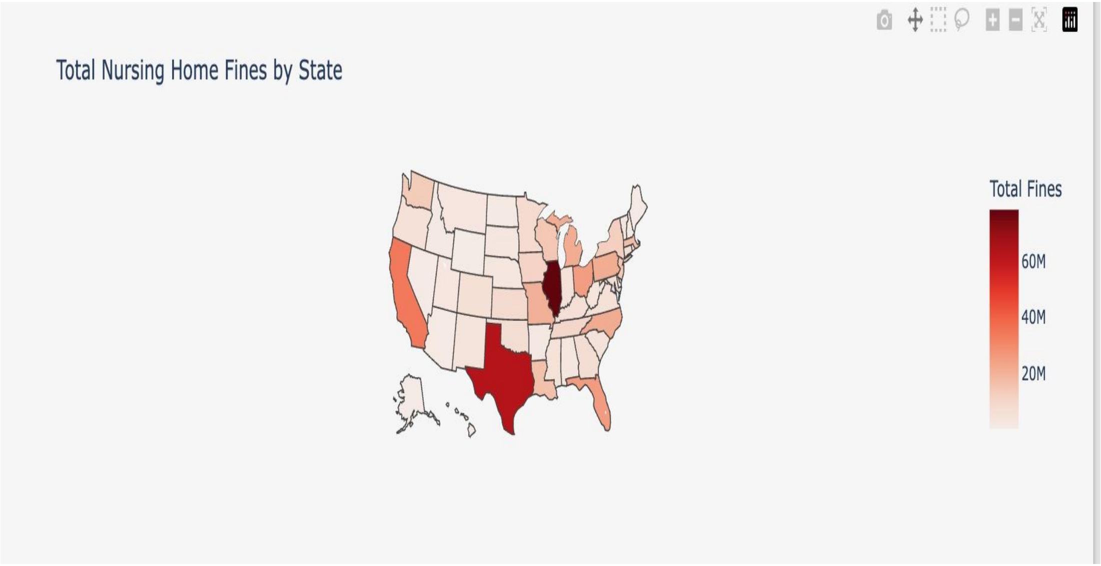


```
       num_penalties  total_fine  avg_fine
State
IL              1662  78593493.0  47288.50
TX              2130  62841318.0  29502.97
CA              1504  34363872.0  22848.32
FL               809  25492561.0  31511.20
OH               800  25253268.0  31566.58
NC               584  21702507.0  37161.83
MI               541  21504953.0  39750.38
PA               745  21459336.0  28804.48
MO               795  20679088.0  26011.43
MA               432  17093894.0  39569.20
LA               330  16155244.0  48955.28
NJ               412  14822170.0  35976.14
WI               374  14506963.0  38788.67
WA               310  13441557.0  43359.86
NY               459  12413140.0  27043.88
IA               383  10647087.0  27799.18
TN               211  10067425.0  47712.91
KS               464   8248327.0  17776.57
MN               381   8138133.0  21359.93
OK               369   7128320.0  19317.94
GA               312   6943359.0  22254.36
RI               177   6509692.0  36777.92
KY               191   6221057.0  32570.98
```



```python
def insight1(csv_path):
    df = pd.read_csv(csv_path)
    df = df.dropna(subset=["Fine Amount"])
    state_summary = df.groupby("State").agg(num_penalties=("Fine Amount", "count"), total_fine=("Fine Amount", "sum"), avg_fine=("Fine Amount"
    return state_summary
```

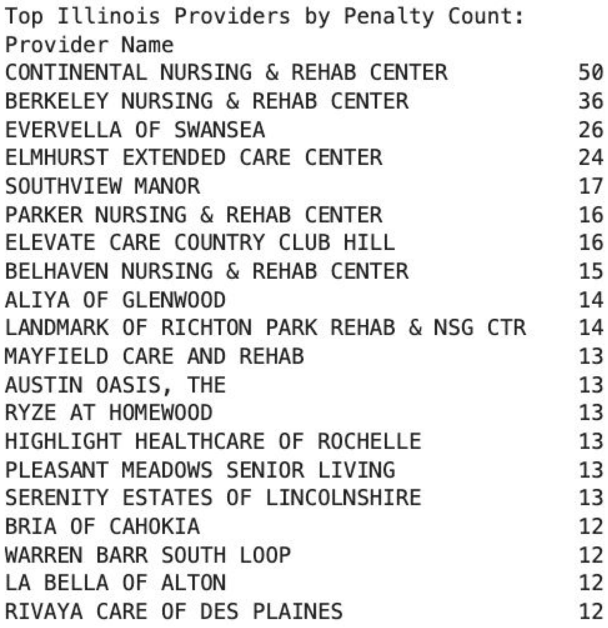
Visualization 1: State Penalty Summary



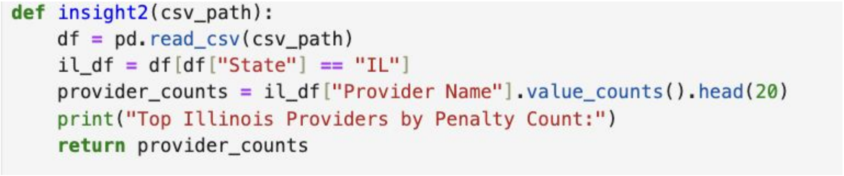
Key finding:

- States vary massively in their total fines—from over $4 million to less than $20k.
- Penalties are not evenly distributed, indicating major geographic differences in nursing home performance or regulatory enforcement.
- The state of Illinois has the highest aggregate fine amount.

Insight 2: Illinois Nursing Home Penalties

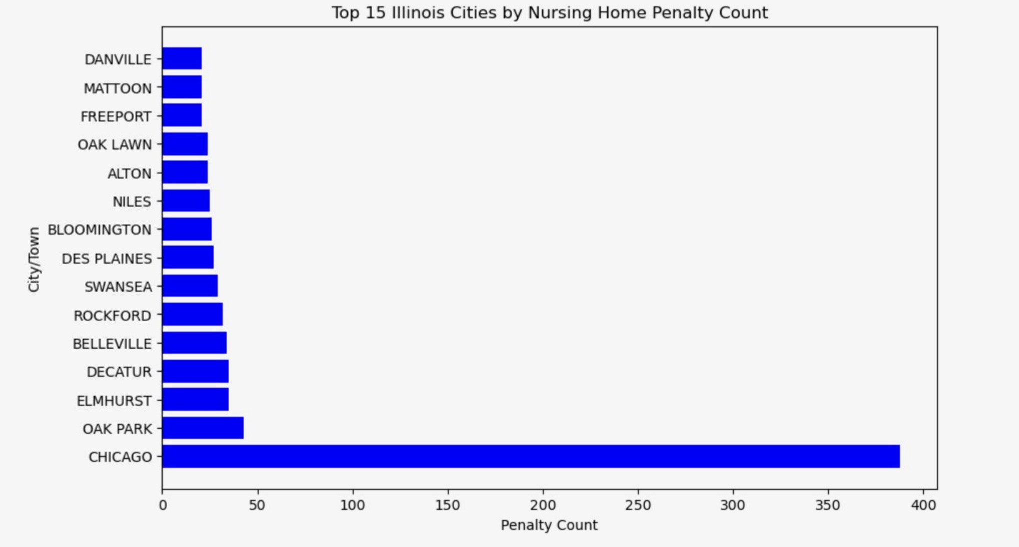


```
Top Illinois Providers by Penalty Count:
Provider Name
CONTINENTAL NURSING & REHAB CENTER          50
BERKELEY NURSING & REHAB CENTER             36
EVERVELLA OF SWANSEA                        26
ELMHURST EXTENDED CARE CENTER               24
SOUTHVIEW MANOR                             17
PARKER NURSING & REHAB CENTER               16
ELEVATE CARE COUNTRY CLUB HILL              16
BELHAVEN NURSING & REHAB CENTER             15
ALIYA OF GLENWOOD                           14
LANDMARK OF RICHTON PARK REHAB & NSG CTR    14
MAYFIELD CARE AND REHAB                     13
AUSTIN OASIS, THE                           13
RYZE AT HOMEWOOD                            13
HIGHLIGHT HEALTHCARE OF ROCHELLE            13
PLEASANT MEADOWS SENIOR LIVING              13
SERENITY ESTATES OF LINCOLNSHIRE            13
BRIA OF CAHOKIA                             12
WARREN BARR SOUTH LOOP                      12
LA BELLA OF ALTON                           12
RIVAYA CARE OF DES PLAINES                  12
```



```python
def insight2(csv_path):
    df = pd.read_csv(csv_path)
    il_df = df[df["State"] == "IL"]
    provider_counts = il_df["Provider Name"].value_counts().head(20)
    print("Top Illinois Providers by Penalty Count:")
    return provider_counts
```

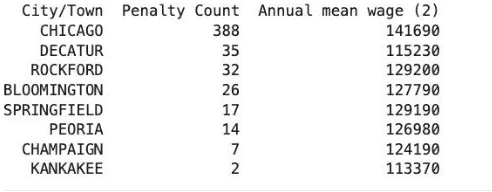
Visualization 2: Illinois Nursing Home Penalty Count



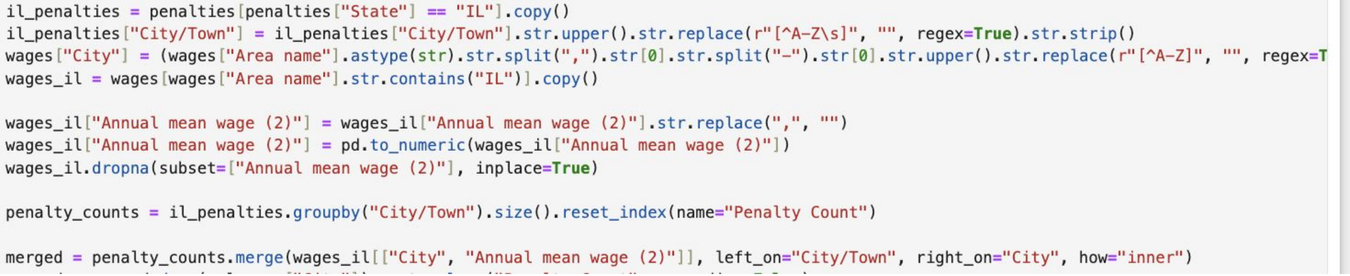
Key finding:

This visualization shows the top 15 Illinois cities ranked by total nursing home penalties. Chicago leads with nearly 400 penalties, far more than any other city. This suggests that Illinois' nursing home quality problems are not spread evenly across the state but are highly concentrated in Chicago. This could relate to a higher population count in Chicago but we will touch on this later.

Insight 3: Penalty Count and Mean Wage of IL Cities



```
   City/Town  Penalty Count  Annual mean wage (2)
     CHICAGO            388                141690
     DECATUR             35                115230
    ROCKFORD             32                129200
 BLOOMINGTON             26                127790
 SPRINGFIELD             17                129190
      PEORIA             14                126980
   CHAMPAIGN              7                124190
    KANKAKEE              2                113370
```



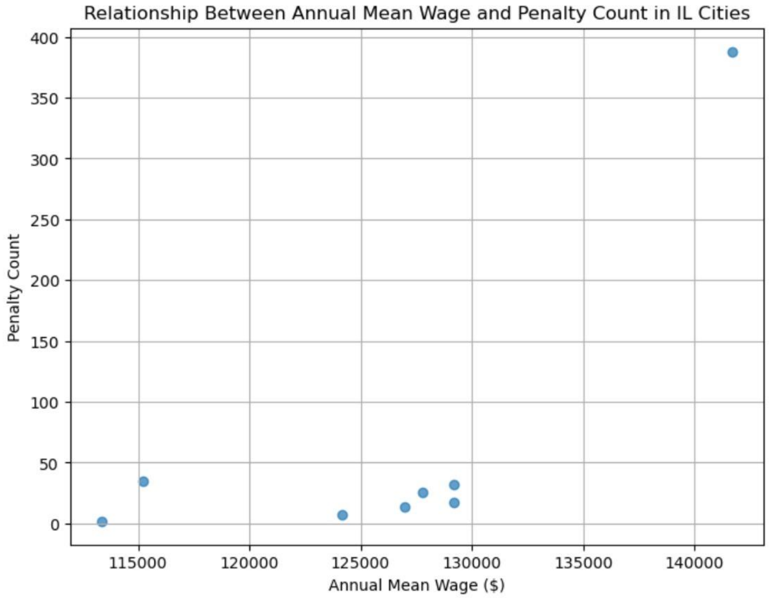
```python
il_penalties = penalties[penalties["State"] == "IL"].copy()
il_penalties["City/Town"] = il_penalties["City/Town"].str.upper().str.replace(r"[^A-Z\s]", "", regex=True).str.strip()
wages["City"] = (wages["Area name"].astype(str).str.split(",").str[0].str.split("-").str[0].str.upper().str.replace(r"[^A-Z]", "", regex=T
wages_il = wages[wages["Area name"].str.contains("IL")].copy()

wages_il["Annual mean wage (2)"] = wages_il["Annual mean wage (2)"].str.replace(",", "")
wages_il["Annual mean wage (2)"] = pd.to_numeric(wages_il["Annual mean wage (2)"])
wages_il.dropna(subset=["Annual mean wage (2)"], inplace=True)

penalty_counts = il_penalties.groupby("City/Town").size().reset_index(name="Penalty Count")

merged = penalty_counts.merge(wages_il[["City", "Annual mean wage (2)"]], left_on="City/Town", right_on="City", how="inner")
```

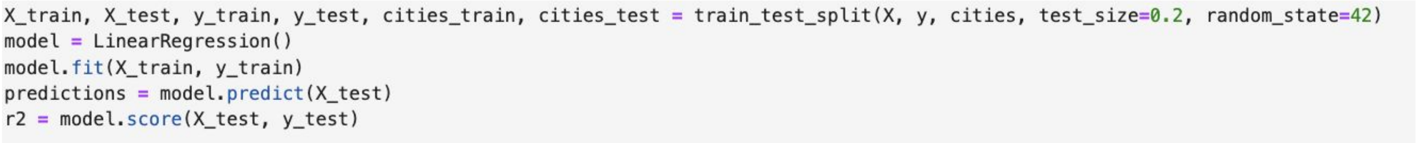
Visualization 3: Mean Wage Vs. Penalty Count in IL Cities



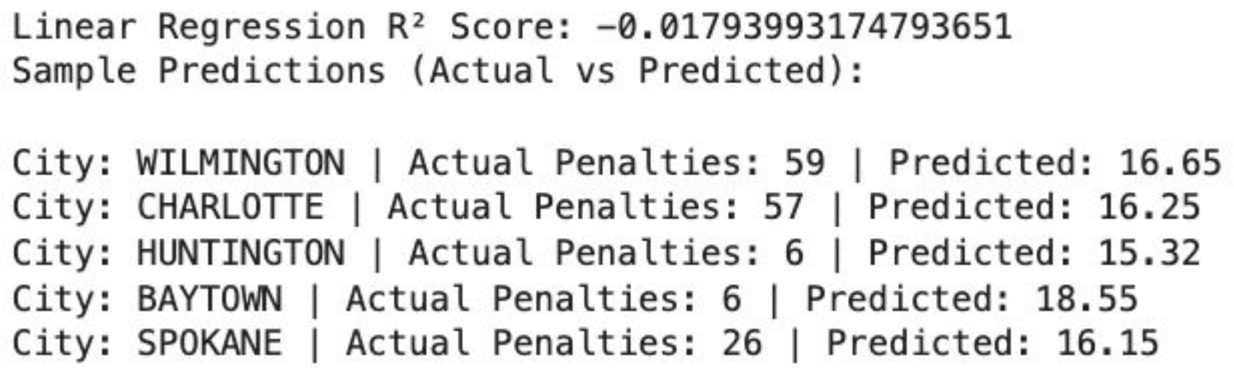
Key Findings:

- The scatterplot shows weak or no correlation between annual mean wage for healthcare managers and nursing home penalty counts in Illinois.
- Most cities fall in a tight range of wages ($110k–$130k) but have vastly different penalty numbers.
- Chicago stands out as a major outlier with far higher penalties, suggesting other factors like facility size, staffing pressure, or population density can play a larger role than wage levels alone.

Insight 4: Linear Regression Prediction



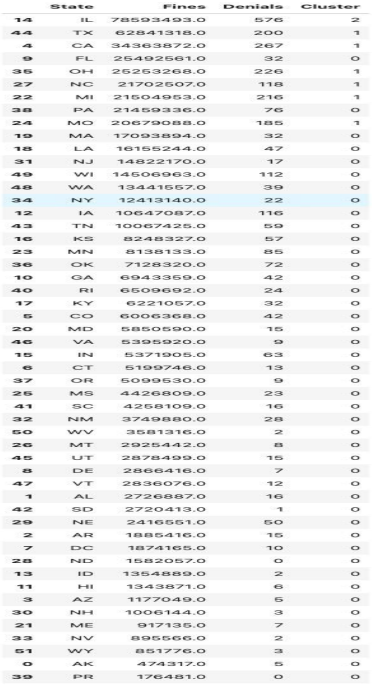
```
X_train, X_test, y_train, y_test, cities_train, cities_test = train_test_split(X, y, cities, test_size=0.2, random_state=42)
model = LinearRegression()
model.fit(X_train, y_train)
predictions = model.predict(X_test)
r2 = model.score(X_test, y_test)
```



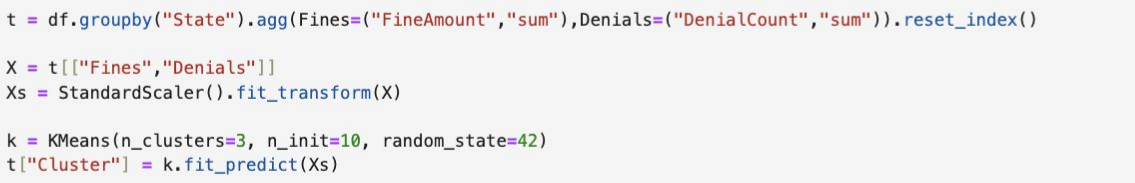
```
Linear Regression R² Score: -0.01793993174793651
Sample Predictions (Actual vs Predicted):

City: WILMINGTON | Actual Penalties: 59 | Predicted: 16.65
City: CHARLOTTE | Actual Penalties: 57 | Predicted: 16.25
City: HUNTINGTON | Actual Penalties: 6 | Predicted: 15.32
City: BAYTOWN | Actual Penalties: 6 | Predicted: 18.55
City: SPOKANE | Actual Penalties: 26 | Predicted: 16.15
```

Insight 5: K-Means Cluster



	State	Fines	Denials	Cluster
14	IL	78593493.0	576	2
44	TX	62841318.0	200	1
4	CA	34363872.0	267	1
9	FL	25492561.0	32	0
35	OH	25253268.0	226	1
27	NC	21702507.0	118	1
22	MI	21504953.0	216	1
38	PA	21459336.0	76	0
24	MO	20679088.0	185	1
19	MA	17093894.0	32	0
18	LA	16155244.0	47	0
31	NJ	14822170.0	17	0
49	WI	14506963.0	112	0
48	WA	13441557.0	39	0
34	NY	12413140.0	22	0
12	IA	10647087.0	116	0
43	TN	10067425.0	59	0
16	KS	8248327.0	57	0
23	MN	8138133.0	85	0
36	OK	7128320.0	72	0
10	GA	6943359.0	42	0
40	RI	6509692.0	24	0
17	KY	6221057.0	32	0
5	CO	6006368.0	42	0
20	MD	5850590.0	15	0
46	VA	5395920.0	9	0
15	IN	5371905.0	63	0
6	CT	5199746.0	13	0
37	OR	5099530.0	9	0
25	MS	4426809.0	23	0
41	SC	4258109.0	16	0
32	NM	3749880.0	28	0
50	WV	3581316.0	2	0
26	MT	2925442.0	8	0
45	UT	2878499.0	15	0
8	DE	2866416.0	7	0
47	VT	2836076.0	12	0
1	AL	2726887.0	16	0
42	SD	2720413.0	1	0
29	NE	2416551.0	50	0
2	AR	1885416.0	15	0
7	DC	1874165.0	10	0
28	ND	1582057.0	0	0
13	ID	1354889.0	2	0
11	HI	1343871.0	6	0
3	AZ	1177049.0	5	0
30	NH	1006144.0	3	0
21	ME	917135.0	7	0
33	NV	895566.0	2	0
51	WY	851776.0	3	0
0	AK	474317.0	5	0
39	PR	176481.0	0	0



```
t = df.groupby("State").agg(Fines=("FineAmount","sum"),Denials=("DenialCount","sum")).reset_index()

X = t[["Fines","Denials"]]
Xs = StandardScaler().fit_transform(X)

k = KMeans(n_clusters=3, n_init=10, random_state=42)
t["Cluster"] = k.fit_predict(Xs)
```

Conclusion and What We Learned

Final takeaway:
Our project shows that nursing home quality issues are uneven across the U.S. and are especially concentrated in specific regions. While community economic conditions influence these patterns, they do not provide the full explanation. Identifying these high-risk clusters can help guide oversight, resource allocation, and future improvements in long-term care quality.

New skills:

- Learned Selenium web scraping to extract data from a dynamically loaded website.
- Worked with external APIs (U.S. Census).
- Worked with matplotlib for some visualizations
- Cleaned and merged multiple datasets with inconsistent structure.
- Implemented clustering and regression models not heavily covered in class.