Phase 3 CS 2316 One Pager

By: Aditya Gutha and Anish Sannareddy

Link To Presentation:
https://www.youtube.com/watch?v=LMQZRh2KXkU

Nursing homes are central to community health, yet many receive fines or payment denials for failing to meet quality and safety standards. Our project investigates how these penalties are distributed across the United States and whether facilities in lower-income or more vulnerable areas tend to receive more penalties. By combining federal data on nursing home penalties, wages for health-care managers, and community-level socioeconomic indicators, we explore how local conditions may shape nursing home quality.

1 Objective

Our objective is to use three combined datasets to answer questions such as:
- Are nursing homes that receive frequent penalties concentrated in specific states or regions?
- Within a state (focusing on Illinois), which cities experience the highest concentration of nursing home penalties?
- How do local wages and poverty levels relate to the number and severity of penalties?

More broadly, we aim to generate data-driven insights about how local socioeconomic conditions relate to nursing home quality.

2 Methodology

Data sources:
- CMS "Nursing Home Penalties" dataset (fines and payment denials).
- BLS OEWS wage data for Medical and Health Services Managers (scraped with Selenium).
- U.S. Census community indicators
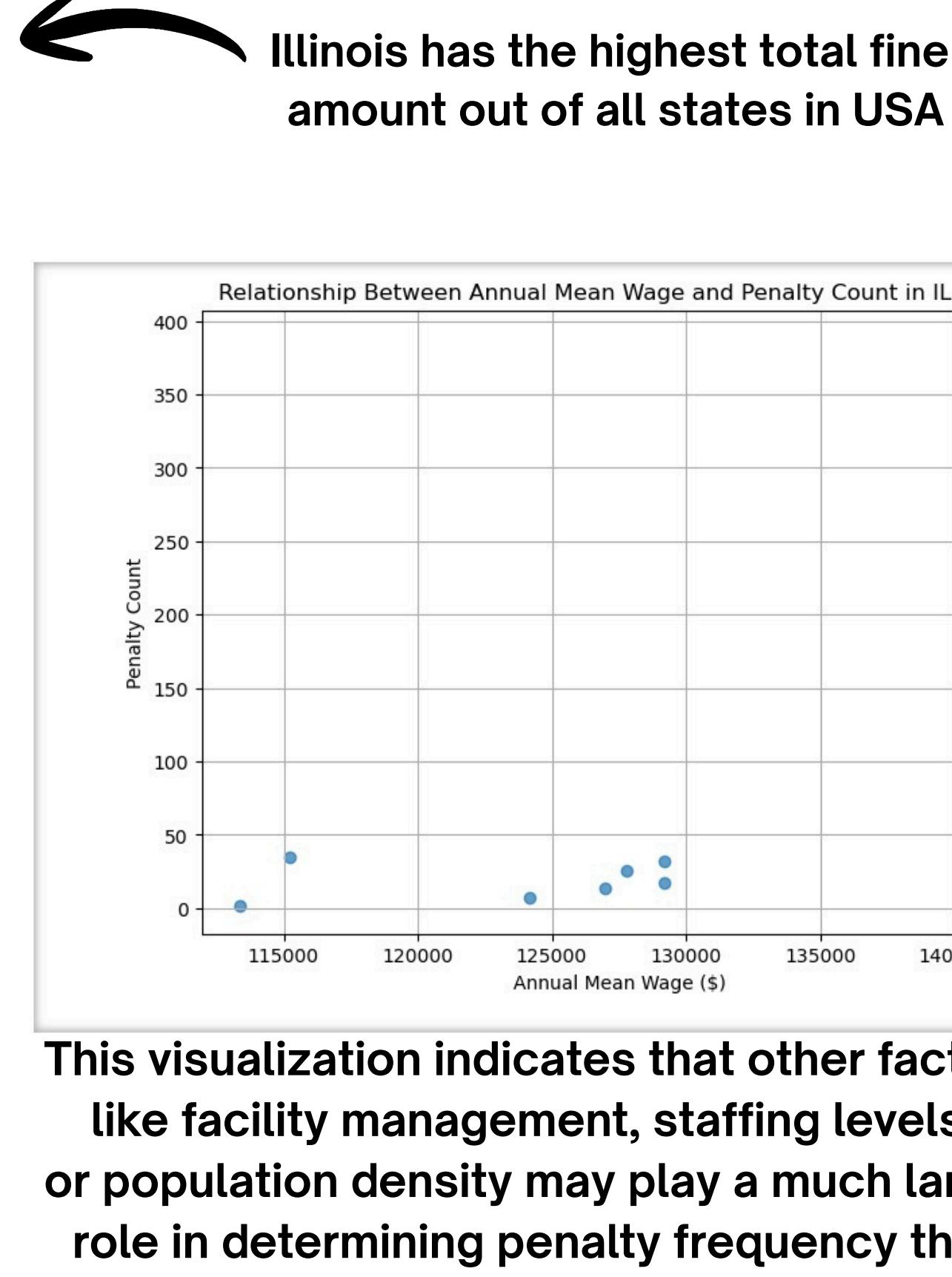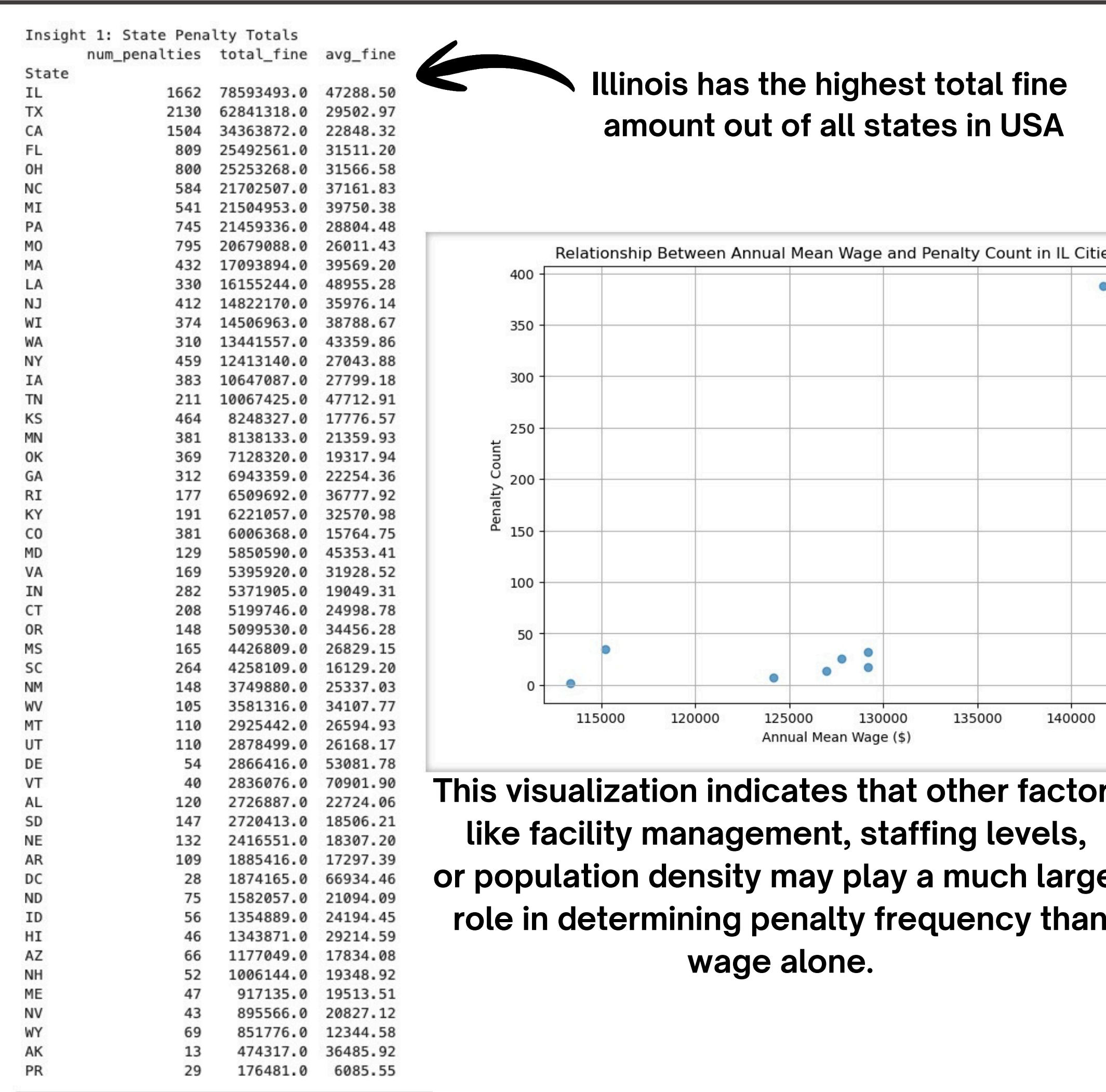
Analysis methods:
- Descriptive group-bys (penalties per state, penalties per provider/city).
- Linear regression using percent below poverty level to predict city-level penalty counts.
- K-Means clustering of states based on total fines and number of denials to identify "high-risk" vs. "low-risk" clusters.

Cleaning & integration:
- Cleaned missing values, removed symbols (e.g., "$", commas) from numeric columns, and standardized location fields (uppercasing city names, stripping punctuation, aligning state abbreviations).
- Aggregated penalties by state and by city, then merged with wage data (by Illinois city) and ACS poverty data (by place/city).

3 Analysis

- The clustering results show that nursing home penalties are geographically concentrated, with a small group of states forming a clear high-risk tier characterized by very large fines and multiple payment denials. This can suggest systematic differences in oversight or staffing rather than surrounding economic conditions.
- Illinois stands out with the highest total fine amount nationwide, but the penalties within the state are concentrated in only a few cities. This can point to localized issues rather than statewide issues.
- The wage-versus-penalty visualization for Illinois indicates little correlation between annual mean wages and penalty counts, meaning wage levels alone do not explain facility performance. Other factors, such as training quality, facility resources, or population density, likely play a larger role.

```
Insight 1: State Penalty Totals
      num_penalties   total_fine   avg_fine
State
IL    1662   78593493.0   47288.50
TX    2130   62841318.0   29502.97
CA    1584   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    12413340.0   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
CO    381    6006368.0    15764.75
MD    129    5058590.0    45353.41
VA    169    5395920.0    31928.52
IN    282    5371985.0    19049.31
CT    208    5199746.0    24998.78
OR    148    5099530.0    34456.28
MS    165    4426809.0    26829.15
SC    264    4258109.0    16129.20
NM    148    3749800.0    25337.03
WV    105    3581316.0    34107.77
MT    110    2925442.0    26594.93
UT    110    2878499.0    26168.17
DE    54     2866416.0    53081.78
VT    40     2036076.0    70901.90
AL    120    2726887.0    22724.06
SD    147    2720413.0    18506.21
NE    132    2416551.0    18307.20
AR    109    1885416.0    17297.39
DC    28     1874165.0    66934.46
ND    75     1582057.0    21094.09
ID    56     1354889.0    24194.45
HI    46     1343871.0    29214.59
AZ    66     1177049.0    17834.08
NH    52     1006144.0    19348.92
ME    47     917135.0     19513.51
NV    69     851776.0     12344.58
WY    43     895566.0     20827.12
AK    13     474317.0     36485.92
PR    29     176481.0     6085.55
```

Figure: Relationship Between Annual Mean Wage and Penalty Count in IL Cities (x-axis: Annual Mean Wage ($); y-axis: Penalty Count)

Illinois has the highest total fine amount out of all states in USA

This visualization indicates that other factors like facility management, staffing levels, or population density may play a much larger role in determining penalty frequency than wage alone.

```
    State   Fines        Denials   Cluster
14  IL      78593493.0   676       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
25  MO      20679088.0   185       1
19  MA      17093894.0   32        0
18  LA      16155244.0   47        0
31  NJ      14822170.0   17        0
48  WA      14506963.0   39        0
34  NY      12413340.0   22        0
12  IA      10647087.0   112       0
43  TN      10067425.0   59        0
16  KS      8248327.0    57        0
23  MN      8138133.0    81        0
36  OK      7128320.0    71        0
10  GA      6943359.0    42        0
40  RI      6509692.0    24        0
17  KY      6221057.0    32        0
6   CO      6006368.0    42        0
20  MD      5058590.0    0         0
46  VA      5395920.0    0         0
5   CT      5199746.0    63        0
37  OR      5099530.0    0         0
24  MS      4426809.0    23        0
41  SC      4258109.0    16        0
32  NM      3749800.0    28        0
50  WV      3581316.0    0         0
26  MT      2925442.0    8         0
45  UT      2878499.0    15        0
8   DE      2866416.0    7         0
47  VT      2830076.0    12        0
1   AL      2726887.0    16        0
42  SD      2720413.0    0         0
29  NE      2416551.0    10        0
2   AR      1885416.0    15        0
7   DC      1874165.0    0         0
28  ND      1582057.0    0         0
13  ID      1354889.0    2         0
11  HI      1343871.0    0         0
3   AZ      1177049.0    0         0
30  NH      1006144.0    3         0
51  WY      917135.0     7         0
33  NV      895566.0     2         0
51  WY      851776.0     0         0
0   AK      474317.0     0         0
39  PR      176481.0     0         0
```

This data clusters states based on their "risk".
Cluster 0 = Highest total fines & highest number of denials (high-risk states)
Cluster 1 = Medium fines and/or moderate denials (mid-risk states)
Cluster 2 = Low fines and few to no denials (low-risk states)
This clustering helps reveal clear geographic patterns in nursing home performance and highlights where quality problems are most concentrated.

4 Conclusion

Our analysis shows that nursing home penalties are not evenly distributed but instead cluster in specific states and cities, indicating clear geographic disparities in care quality. While socioeconomic factors like wages and poverty levels show some relationship to penalty frequency, they do not fully explain why certain regions consistently perform worse.