

Data analysis / case report

Dataset name:

Injuries CSV File

Data source:

<https://think.cs.vt.edu/corgis/csv/injuries/>

Data analysis project location:

<https://github.com/Hypnootis/data-analytics/tree/main/probability-statistics>

Data format:

Csv-file.

Data description and limitations:

The Occupational Safety and Health Administration (OSHA) collected work-related injury and illness data from employers within specific industry and employment size specifications from 2002 through 2011. Roughly 25 000 data points.

The date information is only limited to the year the case was reported in, so we can't i.e see if the winter seasons have more cases in certain industries than others.

Even though the data is of injuries in the US, not all states participated in the 2010 ODI (OSHA Data Initiative; non-participating states: Alaska, Oregon, Puerto Rico, South Carolina, Washington and Wyoming), so data collected in 2009 is not as complete.

(Population source: U.S. Census Bureau QuickFacts)

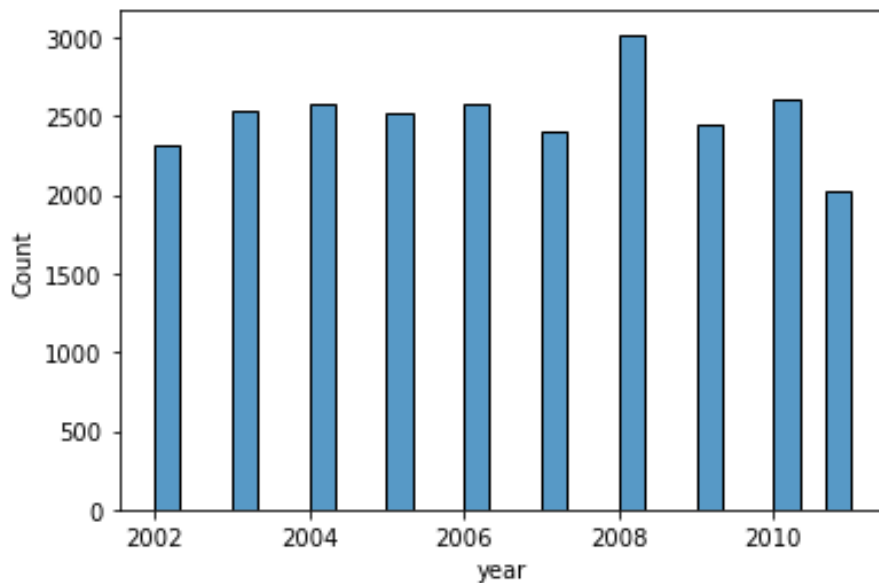
Data cleanup procedure:

The businesses have a "second name", which seems to be a more recognizable name for the companies behind them, but not nearly all of them have it. These also have no impact on the data so I took them out.

Also location details like street name and zip code don't seem important for analysis purposes. I will however keep state and city, as those are more broad and might result in interesting findings.

Data analysis:

Mean of cases per year is 2501, and the distribution hovers at around that mark in the plot:



Average amount of days away is 2,4 days

Looking at the top 10 companies with the most reports:

Business Name	Reports
Home Depot USA Inc	266
Lowe's Home Centers Inc	205
Lowe's Companies, Inc	83
THE HOME DEPOT USA INC	72
FEDEX EXPRESS	58
United Parcel Service Inc	54
Wal-Mart Stores Inc	51
United Parcel Service Inc OH	33
Kindred Healthcare Inc	27
Menard Inc	24

Home Depot and Lowe's are both home improvement store chains, dealing in manufacturing and logistics. FEDEX and United Parcel Service are both logistics companies, so these industries seem to see the most reports.

You can also see in the data the very similar company names, and Lowe's Home Centers seems to be a subsidiary of Lowe's Companies, but I couldn't find any information on the Home Depot situation. Could just be a quirk in the data, since it has been gathered over a decade.

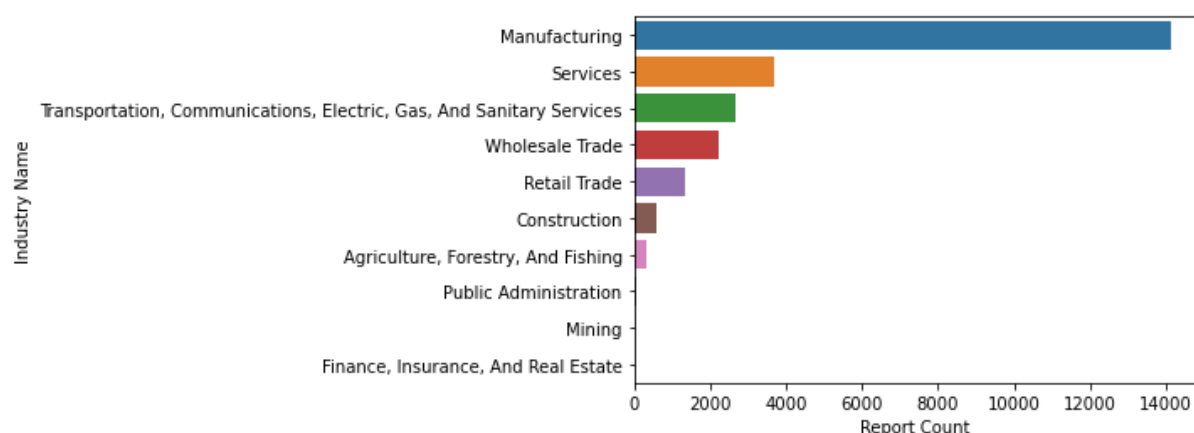
Here are the 5 most represented states in the dataset, California coming in with a crushing victory.

CA	2477
TX	1649
PA	1484
OH	1435
IL	1349

The most represented city is surprisingly Houston, Texas, and a Californian city doesn't come in till third place, Los Angeles. California's population in 2010 was around 37 million, whereas Texas' was 25 million. We can make an educated guess that California's overrepresentation could be due to a higher population.

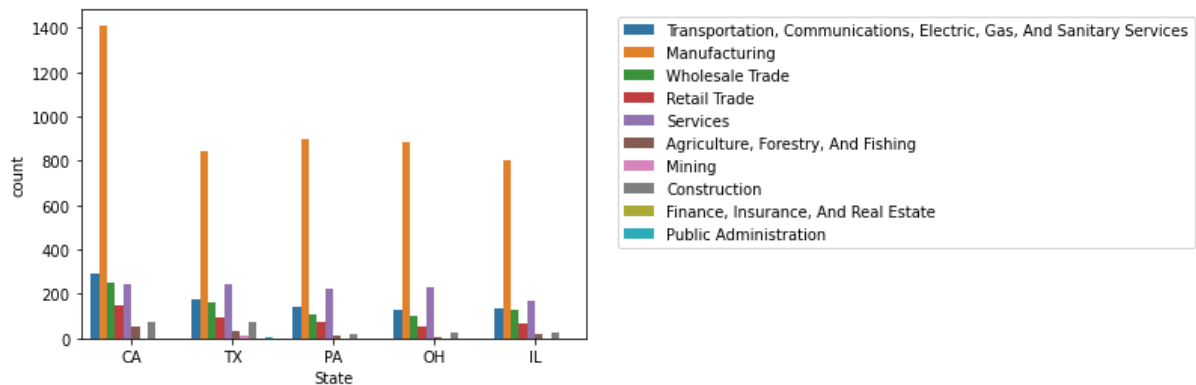
Houston	201
Chicago	125
Los Angeles	111
Cleveland	102
Cincinnati	94

Looking at the distribution by industry, manufacturing has the most cases reported, nearing 14 000. An interesting fact is the single (1) case reported in the industry of Finance, Insurance and Real Estate. Unfortunately our data set doesn't have information about the injury.



Looking at the top 5 states in reports, we can see that they do have an emphasis on manufacturing based on the reports. What is interesting here is that ignoring cases in

manufacturing, California has reported a similar amount of incidents compared to the other states, citing either a bigger amount of workers in manufacturing, or less safe conditions in manufacturing on average. Keeping the population difference in mind, you can make of that what you will.



Looking at population numbers, Pennsylvania's population in 2010 was around 12,8 million, Ohio at 11,5 million and Illinois at 12,8 population. This goes roughly along with the population distribution of the time, so looking at population does not give accurate results if wanting to look for correlations.

Results and summary:

Based on the dataset used, businesses dealing in manufacturing appear to have a higher rate of incidents, with a major representation by businesses dealing with home improvement. Location does not seem to play a factor in the number of incidents, apart from a higher population = more workers = more incidents reported.