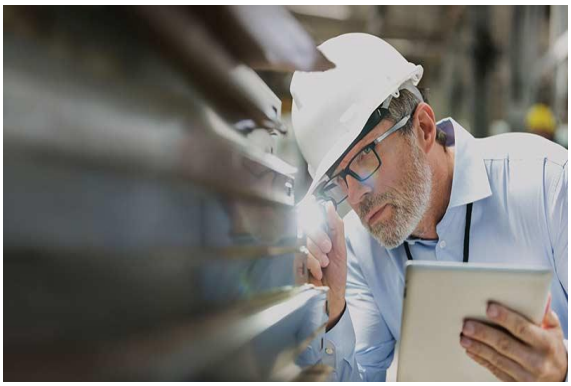
Exploratory Data Analysis on Workplace Fatalities

Imama Mansoor
imama.mansoor@gmail.com

Data Science Project 1

Introduction

Workplace fatality means a fatality that occurs to an employee (working for pay, compensation, or profit) or volunteer (exposed to the same work hazards and performing the same duties or functions as paid employees) while engaged in a legal work activity, or present at the site of the incident as a requirement of his or her job. A work relationship exists if an event or exposure results in a fatal injury to a person on or off the employer's premises and the person was there to work; or if the event or exposure was related to the person's work or status as an employee.



This project is all about exploratory data analysis on Workplace Fatalities and Injuries for FY 2012 in all the states of USA. A Power BI report is also attached for the visualization of the data to slice and dice, including descriptive analysis and giving lots of key insights and interesting analysis.

Dataset

The dataset includes data points on fatalities, injuries, inspections, and penalties for each state in the US for the year 2012. However, some fields in the dataset are missing and have been replaced with zero, I'll update those fields as soon I'll get the complete data. After data cleaning and removing some unnecessary fields, following are the important columns remained in the dataset for the analysis.

- State
- Number of Fatalities, 2012
- Rate of Fatalities, 2012
- State Rank, Fatalities 2012
- Number of Injuries/Illnesses 2012
- Injuries/Illnesses 2012 Rate
- Penalties FY 2013 (Average $)
- Penalties FY 2013 (Rank)
- Inspectors
- Years to Inspect Each Workplace Once
- State or Federal Program
- Location

Tools for Visualization

I have used Power BI which is a business analytics service by Microsoft.

The version used is 2.90.1081.0 64-bit (February 2021).

Key points for Analytical Storytelling

Some important key points which are taken care off while creating this report are as follows:

- Conventional wisdom says you have seven seconds during which your audience is making the decision about whether they are really engaging with you or not. Keeping this fact in mind I have used state wise analysis on the first page which gives the audience the impact that they have the control over all the data on a single click.
- Ben Shneiderman, an American Computer Scientist, who made well-formed dashboards, gave an idea for bringing data to life through the combination of visuals and narrative.
- Visual Processing stages like Preattentive Processing (Edges, Orientation, Color etc) and Serial Processing (Object Recognition). For example, color theme on each dashboard is soothing enough for the audience to keep them engaging in data wrangling. Secondly USA flag is there on the State wise analysis to create an environment for the states and dive deep to analyze each and every state. Thirdly descriptive analysis will not leave any question unanswered, as this dashboard cover all the statistical aspect of the particular field under consideration.
- Analytics map is also in mind while creating interesting insights.

Ideas for Future Improvement

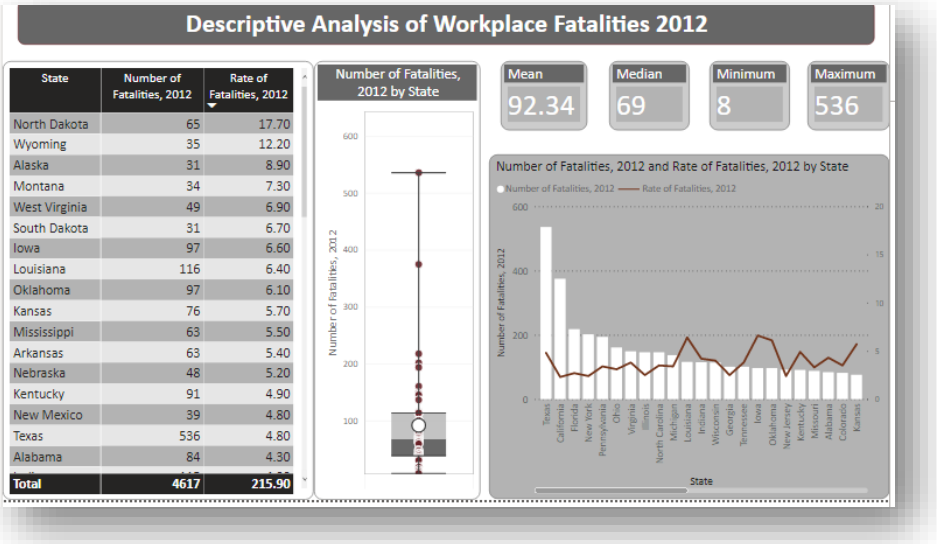
- A complete dataset containing all the values will result to create good insights as compared to an incomplete data.
- Irrelevant data is of no use, so there is no point of using it.
- Data cleaning is the most important step which should be done before analyzing or even start wrangling it.
- Small datasets have very limited visualization so try to work on Big data to create visualization on a broader spectrum.
- Storytelling concepts will be of more usage, while giving presentations instead of creating narrations.
- Practice on more datasets and explore Power BI for hands on experience of newer visualization techniques.

Details for Visualization

State wise Analysis

You can slice and dice the data state wise as you want to analyze the Workplace fatalities data, just by clicking on the State name, you will get all the related data as Number of Fatalities, Number of Injuries, Inspectors, the particular state rank, including the geographic location of the state by map. Not only this, you will also get to know the "Years to Inspect Each Workplace Once" and "Program". **Isn't this great to get the data on a single click!**

Descriptive Analysis



State	Number of Fatalities, 2012	Rate of Fatalities, 2012
North Dakota	65	17.70
Wyoming	35	12.20
Alaska	31	8.90
Montana	34	7.30
West Virginia	49	6.90
South Dakota	31	6.70
Iowa	97	6.60
Louisiana	116	6.40
Oklahoma	97	6.10
Kansas	76	5.70
Mississippi	63	5.50
Arkansas	63	5.40
Nebraska	48	5.20
Kentucky	91	4.90
New Mexico	39	4.80
Texas	536	4.80
Alabama	84	4.30
Total	4617	215.90

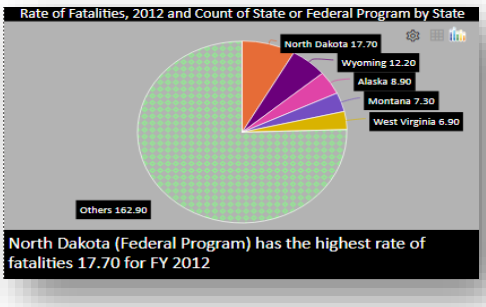
I have selected **Workplace Fatalities** and **perform descriptive analysis** on it, as a result you will get the knowledge about Mean, Median, Minimum and Maximum. I have also shown the particular fields like State, Number of Fatalities and rate of Fatalities for the fiscal year 2012 in tabular form for your ease. There is a Box and Whisker

plot for graphically depicting the numerical data through quartiles. The mean is shown by the white circle in between. The small circles in deep red color shows the Number of Fatalities relative to state. You can also analyze the relationship between Number of Fatalities and Rate of Fatalities by State through this Line and Stacked Column chart. You can explore this visualization further by clicking on any state and all the visualization will filter accordingly. Go ahead and explore…

Question and Answer

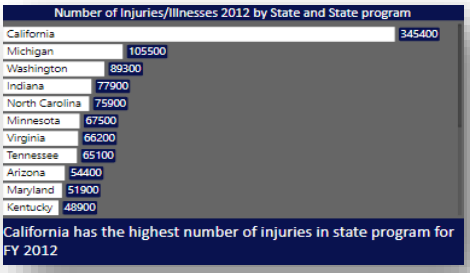
❖ Which program, state or federal, has the highest rate of fatalities?

Advanced Pie and Donut Chart – xViz is used for this visualization which is depicting State by Highest Rate of Fatalities by State or Federal program. As shown North Dakota which is under the Federal program has the highest rate of Fatalities for the fiscal year 2012 having value 17.70.

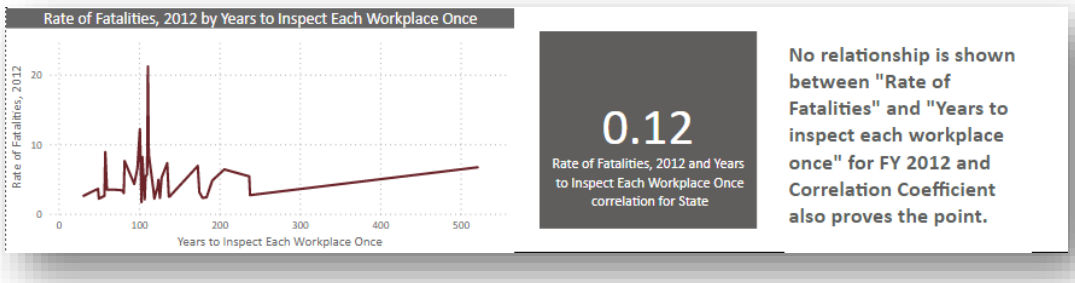


❖ Which state with a state program has the highest number of Injuries/Illnesses?

In this visualization I have used Horizontal Bar chart to illustrate about the state with the highest number of injuries under state program. California has the highest number of injuries/illnesses for fiscal year 2012 equivalent to 345400.

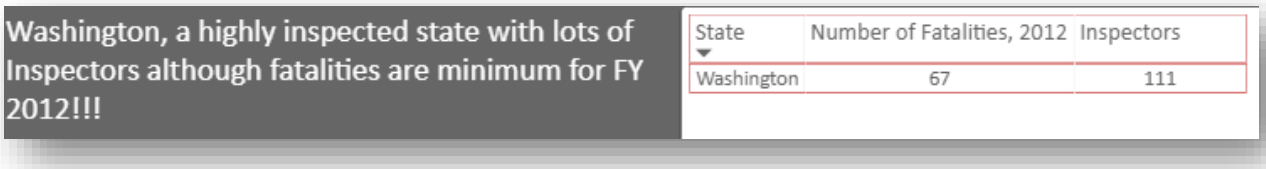


❖ What is the relationship, if any, between "Average of Years to Inspect Each Workplace Once" and "Rate of Fatalities"?



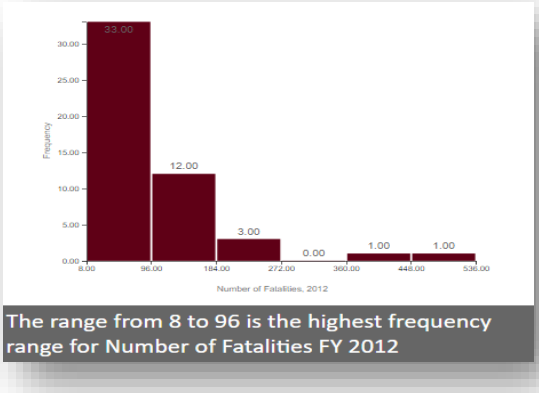
In this visualization I have used Line chart to understand the relationship between "Years to Inspect Each Workplace Once" and "Rate of Fatalities" for the fiscal year 2012. However, I don't find any relationship between these two terms, as shown in the chart. For further confirmation I have also checked the correlation between these two fields and the result is 0.12 which also proves my analysis.

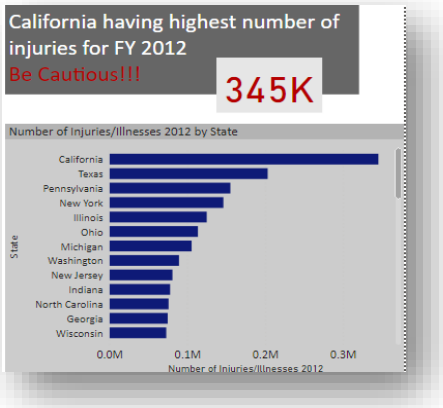
Interesting Insights



Washington DC, is that particular state which has quite minimum number of fatalities i.e 67 but a lot of inspectors i.e 111 are allotted for this state. This might be because of the reason as Washington is the capital state of America. I have used table visualization and a text box for the illustration of this insight.

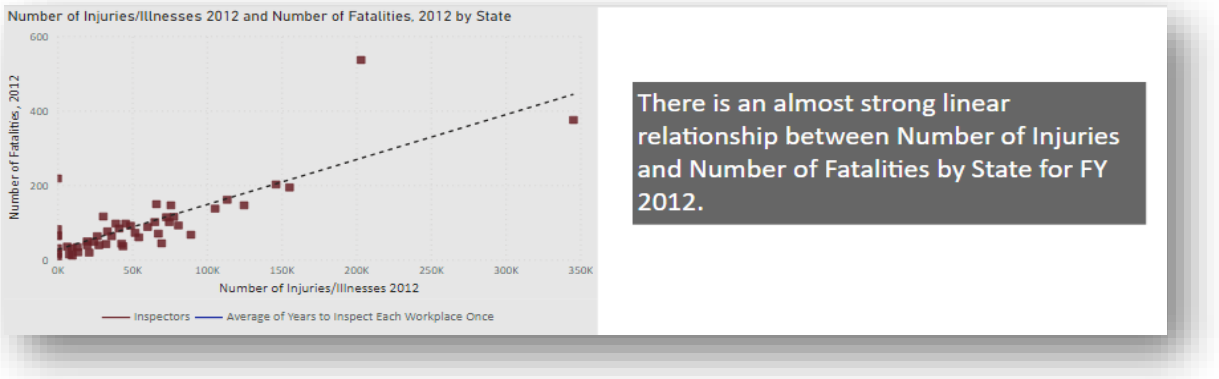
Consider Number of Fatalities for the Fiscal year 2012. I have used Histogram and a text box for the illustration of this insight. It is clearly shown that the range from 8 to 96 is the highest frequency range having 33 value of frequency. This depicts common number of Fatalities for most states. Histograms are one of the easiest ways to understand any visualization.





California having highest number of Injuries resulted in a huge number of Fatalities. Nearly 3.0 million nonfatal workplace injuries and illnesses were reported by private industry employers in 2012, resulting in an incidence rate of 3.5.

"An increase in workplace fatalities is a serious concern for Cal/OSHA. We are analyzing the data to bolster and direct our enforcement and education efforts" said Cal/OSHA Chief Douglas Parker in a press release.



For further analysis of the given dataset, I tried to find some relationship among the fields. I have used scatter chart and found a strong linear relationship between Number of Injuries and Number of Fatalities for the fiscal year 2012 by state. These small squares represent each state. For more ease and further analysis, I have drawn a trend line too from the analytics pane.

If you want to explore the data of “Inspectors” and “Years to Inspect Each Workplace Once”, this visualization will help to get the date for these fields relative to a particular state.

References

https://www.onlyinyourstate.com/north-dakota/reasons-to-never-move-to-nd/
https://www.safewise.com/blog/safest-cities-north-dakota/
https://www.bls.gov/news.release/archives/osh_11072013.pdf
https://hrwatchdog.calchamber.com/2020/01/california-fatal-workplace-injuries-increased-in-2018/#:~:text=After%20several%20years%20of%20declines,up%20from%20376%20in%202017)