**FIRST DATA SCIENCE PROJECT –**

**WORKPLACE FATALITIES IN THE VARIOUS STATES OF UNITED STATES IN THE YEAR 2012**

**PROJECT SUMMARY**

**Rate of Fatalities (RoF)**

From the **lowest RoF** to the **highest** **RoF**, the Bar chart (RoF -Federal State Program) shows the variation of adequacy of Federal/State program and gaps in implementation along with the States. The Adequacy of Federal/State program can be measured in terms of achieving the goal of maximum possible reduction of RoF with less gaps in implementation of the program.

**Variable Factors**

1. Rate of Fatalities (RoF)
2. Workers’ population in the State
3. Adequacy of the Federal / State program meeting the goal of reducing maximum the RoF
4. Federal / State program implementation gaps
5. Number of workplaces where Federal/State programs implemented

Per Federal Program, the **Highest** **RoF** in 2012 is in the state of **North Dakota** where the program was inefficient to meet the goal of reducing the RoF with more gaps in implementation and the **lowest** **RoF** is in **Massachusetts**, where the program was adequate and implemented strongly with only few gaps. However, the average Rate of Fatalities (RoF) is estimated to **4.43,**

**= 4.43**

**\*\*\*\***Per state program, the **highest** **RoF** is in **Wyoming**, where program implemented was inefficient to handle further reduction of RoF effectively with more gaps, however the **lowest RoF** is in **Washington** where program implemented was highly adequate with few implementation gaps.

**= 4.03**

**So, Federal Program has the highest Rate of Fatalities (RoF) in the year 2012**

**BOX AND WHISKER PLOT** (Plotted on Log Scale)

|  |  |  |
| --- | --- | --- |
| **BOX AND WHISKER PLOT Statistical Observations** | **Federal** | **State** |
| Minimum RoF | **1.40** | **1.40** |
| Second Quartile (The RoF below which the lower 25% of the States are contained) | **2.00 – 3.00** | **1.70 – 2.70** |
| Median RoF (The middle in the range of RoF Values) | **4.0 in the range in between 3.10 – 5.10** | **3.2 in the range in between 2.80 – 3.70** |
| Third quartile (The RoF above which the upper 25% of the data are contained) | **5.30 – 7.30** | **3.80 – 6.60** |
| Maximum RoF (The largest RoF in the data set) | **17.70** (outlier) | **12.20** (outlier) |
| The average RoF | **4.43** | **4.03** |

**Based on the above observations, Federal program has the highest Rate of Fatalities (RoF)**

**Highest Number of Injuries / Illness**

From the Bar Chart, the State with a State program has the highest number of injuries / illnesses is California State with number of injuries/illnesses around **345K.** This is an outlier in the data set as the next highest number of injuries/illnesses in the State of Michigan with **106K.**

The number of injuries / illnesses varies with States depend on the following variables

**Variables**

1. Workers’ population in the State
2. Adequacy of the Federal/ State program meeting the goal of reducing maximum number of injuries / illnesses
3. Federal / State program implementation gaps
4. Number of workplaces where Federal/State programs implemented

**Average of Years to Inspect Each Workplace Once**

**Vs**

**Rate of Fatalities (RoF) in Scatter Plot**

The average of Years to inspect each workplace once can be assessed in terms of Inspection on demand **(IOD)** in years on a workplace. The scatter plot shows the correlation between inspection on demand **(IOD)** and RoF. Each observation in the plot has two coordinates. The inspection on demand **(IOD)** represents X-coordinate and RoF represents the Y-coordinate.

The line of best fit is the trend line that fits the data the most closely. The trend line on the plot shows an uphill move from left to right. This indicates a positive correlation between the X and Y coordinates ie a positive relationship between the inspection on demand **(IOD)** and Rate of Fatalities (RoF).

The Positive correlation shows when the inspection on demand **(IOD)**is high in a State then the Rate of Fatalities (RoF) also high. The trend line can also be used to predict the inspection on demand (IOD) and RoF if not available for a State in 2012. This will help us to remove the outliers in the data set and predict the most accurate one.

**Variables affecting IOD –**

a) No of Workplaces,

b) No. of dedicated Federal / State Inspectors,

c) Time to complete inspection in a workplace

d) Future Improvements in the Workplace

**Fatal Occupational Injuries by Major events in 2012 – Pie Chart**

A pie chart is also given showing Fatal Occupational injuries by major events in 2012. These are categorized mainly as: -

Fatal Injuries due to

1. Contact with objects and equipment - **39.13%**
2. Falls, Slips and Trips – **37.02 %**
3. Exposure to harmful substances – **17.43%** and
4. Fires and Explosions – **6.42%**

**Power BI Features Used in the Project**

1. Transform Data – to clean and arrange data in the respective fields
2. Data Visualizations (Details, Format, and Analytics features used wherever required) such as
3. Stacked Column Chart – used for comparing Rate of Fatalities (RoF) in 2012 by States with Federal/State inspection programs and also used for Highest number of injuries / illnesses in 2012 by State
4. Slicer with Filter feature is used for selection of Federal or State programs
5. Box card – used for showing average RoF
6. Box – and – Whisker plot used for statistical observations
7. Scatter Chart showing Average RoF and Average of Years to inspect Each workplace once used with Trend line from Analytics feature of the visualizations.
8. Pie Chart used for showing Fatal Occupational Injuries in 2012 by major events.
9. Other Power BI Features such as Fields Data Set, Text Box etc.

**Use of Analytical Story Telling**

**Preattentive** processing technique was used in the visual processing stage. **Preattentive** processing steps – 3 steps – 1) Enclosure, 2) Color and 3) Size.

* Relay on the chart themselves to drive the story instead of rows and columns of data set.
* Legends can be integrated in the chart itself instead of existing as separate elements.
* Give the viewers an easy takeaway.

The chart uses a single color across all bars and orders the bars descending by RoF. Notice how the eyes more easily follow the length of the bars down and across the States from largest to smallest.

The vertical bar chart has a dark black border. It’s probably the first thing can be noticed about the chart. Once we remove the border, as shown in the chart in the middle, we notice that the chart category labels have a diagonal orientation. They stand out because nothing else in the chart is diagonal. It’s a bit distracting and difficult to read.

**Recommendations for Future Improvements**

Based on the above studies and / or charts, the following the recommendations for future improvement to reduce substantially the Rate of Fatalities in the workplaces

1. Updated and / or new codes and recommended practices to be included in the State / Federal programs in order to prevent fatalities in the workplaces.
2. Mandatory requirements of Occupational Safety and Process Safety controlled by laws and regulations
3. Training on the implementation new codes and recommended practices
4. Incident investigation and reporting corrective actions
5. Learn from the previous experience
6. More stringed inspection and control programs required