Candace Warr

Project #1: Exploratory Data Analysis Report

REPORT GOALS

1. *Highlight at least one interesting aspect of the data.*
2. *Allows a reader to interactively explore other aspects of the data.*

REQUIREMENTS

1. *Which program, state or federal, has the highest rate of fatalities?*
    - The Federal program has the highest rate of fatalities at 128 per 100,000 full-time equivalent workers (State is at 88).
2. *Which state with a state program has the highest number of injuries/illnesses?*
    - California has the highest number of injuries/illnesses in the state program at 345,400.
3. *What is the relationship, if any, between "Average of Years to Inspect Each Workplace Once" and "Rate of Fatalities"?*
    - There appears to be no correlation between years to inspect each workplace once and the rate of fatalities as the correlation coefficient is 0.1223 (calculated in Excel) and the trend line is nearly horizontal. There are a few outliers that are influencing the data, so for future improvements to the project I would look further into the outliers - can they be removed?

SUBMISSION

1. *An Excel workbook and/or Power BI report (attached to the email). You must send an actual file, not publish a Power BI report to the web and send a link.*
2. *A document (Word file, PDF, or plain text) containing*:
    a. *Details on the questions explored, insights highlighted, and functionality provided by your project:*
        - It is very eye opening to see California and Texas in the top two for both Number of Injuries/Illnesses and Number of Fatalities - is this due to the population of the states, state size (area), or types of occupations?
        - Although Key Influencers indicate that injuries/illnesses increase when the number of inspectors decreases, it is not fully reflected as such in the line and clustered column chart. This may be due to a few outliers as well as some missing data.
            - When separating the Number of Inspectors by Program, state or federal, there are more state inspectors and a smaller rate of Fatalities by the state program.
        - There is a lot more that could be explored with this data (refer to "ideas for future improvement") to fill in the gaps and gain a better understanding of causes of workplace injuries/illnesses and fatalities.
    b. *Information about which Excel/Power BI features were used:*
        - Clustered Column Charts, Tables, Scatter Plot with Trend Line, Key Influencers
        - Correlation - Data Analysis in Excel
        - *Interactive Elements:* Slicer, Buttons

c. *What information from the Analytics Storytelling course you applied:*
    - I took a backwards approach in exploring relationships and information about the data that I had questions about myself before attacking the questions at hand. Because there was little to no correlation among the data, I chose to create visuals that would directly answer the questions posed about the data.
d. *Ideas for future improvement of your project (minimum of 3):*
    - Determine how injuries and illnesses are defined across both State and Federal programs
        - Are there particular types of injuries that occur more than others? Do these injuries lead to fatalities more than other injuries/illnesses?
    - Explore the outliers as well as the missing fields; clean the data
    - What is the distribution of injuries/illnesses among professions and how many of those injuries lead to fatalities in that profession? (is there a correlation among certain professions and number of injuries/illnesses and fatalities)
    - Does the number of inspectors relate to the size of the state: both area and population?
    - Review and include data from 2011 and 2013 to show any improvements or increases in injuries/illnesses and fatalities.