First Data Science Project: Exploratory Data Analysis
## By Salsabil Chowdhury Tory

I have created both Excel and Power BI reports to analyze the given data in two different platforms. I believe Power BI can create a better visual representation altogether.

The Excel report:

In this report, I will present an analysis of the provided data using Excel pivot tables and slicers. I used these features to identify patterns and relationships in the data.

Workplace Fatalities and Injuries: Using the Excel pivot table with slicers program, I found that the Federal program has the highest rate of workplace fatalities, while California has the highest number of injuries/illnesses. This information provides valuable insights into the safety and health of workplaces in these areas.

Years to Inspect and Rate of Fatalities: To explore the relationship between "Average of years to inspect Each Workplace Once" and "Rate of Fatalities," I evaluated the correlation coefficient in Excel. The correlation coefficient is 0.122, indicating a weak positive correlation between the two variables. I used both the correlation matrix and the data analysis feature to arrive at this conclusion.

	Rate of Fatalities, 2012	Years to Inspect Each Workplace Once
Rate of Fatalities, 2012	1	
Years to Inspect Each Workplace Once	0.122329	1

'Rate of Fatalities, 2012' and 'Years to Inspect Each Workplace Once' appear to...

so the correlation coefficient is very low amost close to zero.
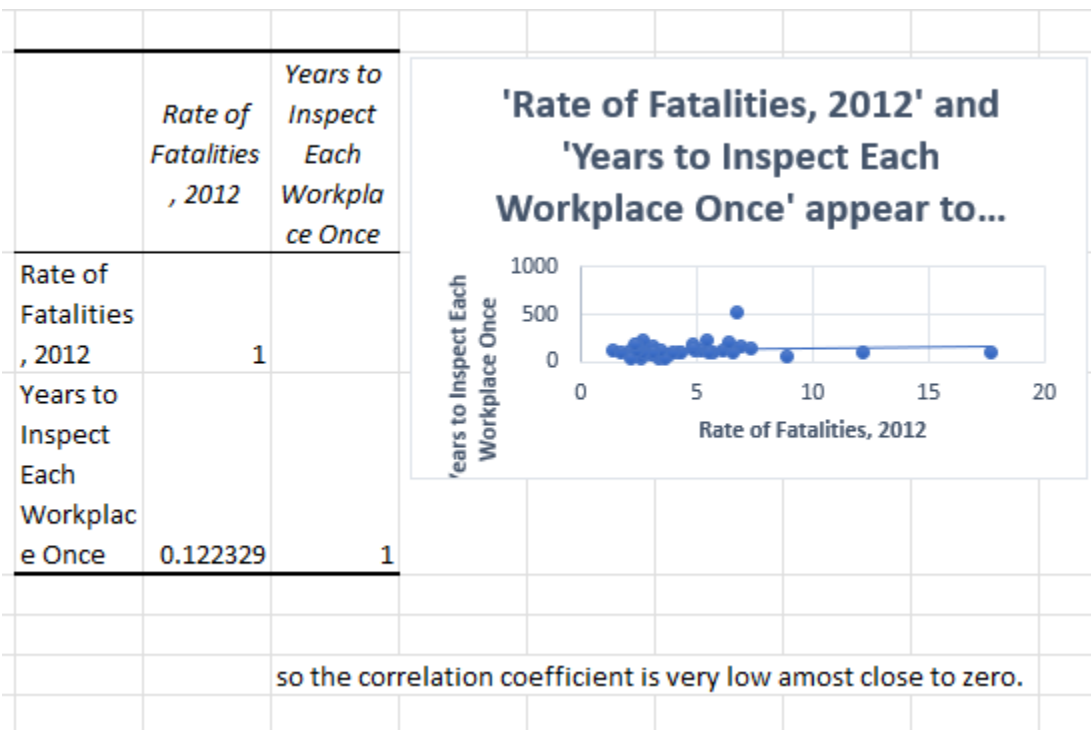
Overall, this Excel analysis provides valuable insights into the safety and health of workplaces in the Federal program and California. The correlation analysis between "Average of years to inspect Each Workplace Once" and "Rate of Fatalities" shows a weak positive correlation, suggesting that longer inspection intervals may be associated with a

slightly higher rate of fatalities. This information can be used to inform workplace safety policies and practices.

Power BI report:

In this report, I will present a visual analysis of the provided data using Power BI. The data was first imported, cleaned and transformed in data query to obtain a clean structure for
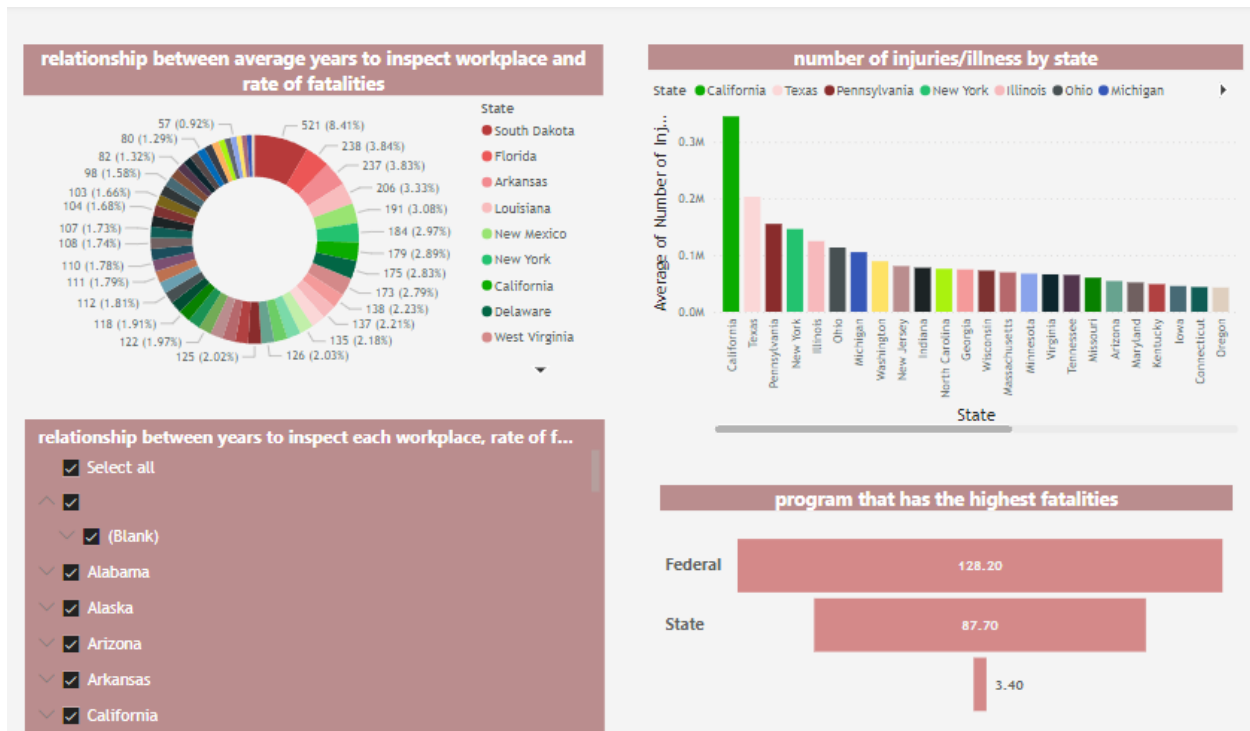
the visual report. Based on the questions given in the project, I created different charts that fit the scenario the best.

Data Transformation and Cleaning: To create a clear and effective visual report, I first imported the data and transformed it in data query. I cleaned the data and removed all unnecessary information to ensure a clean structure.

Visual Report Creation: I have created different charts to analyze the data and present the results visually.

Donut Chart: To show the relationship between "average years to inspect workplace" and "rate of fatalities", I created a donut chart. This chart allows us to quickly see the correlation between the two variables.

Bar Chart: To show the number of injuries/illness by state, I created a bar chart. This chart is useful to compare the number of injuries/illnesses across different states.

Funnel Chart: To show which program "state" or "federal" has the highest fatalities, I created a funnel chart. This chart provides an overview of the fatalities in each program and allows us to see which program has the highest fatalities.

Slicer: Finally, I created a slicer to show the visual relationship between "years to inspect each workplace", "rate of fatalities", "and number of injuries/illnesses by state". This slicer allows us to select either all the states or one or more states to see how the other parameters relate. The slicer also interconnects with the donut, bar and funnel charts in the report and shows the interactive relation between the parameters.

Overall, this Power BI visual report is an effective way to analyze the provided data. The charts and slicer provide a clear and interactive representation of the data, making it easier to understand the relationships between different parameters.

3 Ideas for future improvement of this project

- Integrating predictive analytics: Incorporating predictive analytics along with the historical data can provide insights into potential workplace hazards or risks. For example, using machine learning algorithms to predict the likelihood of workplace accidents or injuries based on past incidents, demographic information, and other relevant data can help organizations reduce the future workplace hazards.

- Expanding data sources: If more than one dataset can be analyzed from additional sources, we can get even better insights into workplace safety. For example, employee surveys or safety audits can be beneficial for better analysis.

- Developing a real time dashboard: Real-time dashboard can display the most updated data on workplace safety enabling organizations to make quantitative decisions more promptly.