A Project Report on Comprehensive Analysis of Workplace Incidents Data

MD ASHIQUR RAHMAN

February 24, 2023

Abstract

This technical report describes an exploratory data analysis (EDA) and data simulation project relating to workplace injury incidents. This project involves the generation of simulated data representing workplace incidents over the course of one year. The information included in the simulated data includes incident dates, incident types (e.g., Falls, Slips, Equipment Malfunctions), and severity levels (Low, Medium, High). This project is primarily intended to provide an understanding of the application of data generation and data exploration techniques utilizing Python programming. This report begins with the creation of synthetic data, which is used to simulate workplace incidents and store them in a structured manner. In the following section, the data generated will be explored in more detail. There are several aspects of the data exploration phase, including data summary statistics, frequency analyses of incident types and severity levels, identification of patterns and trends, and visualization of the data. In addition, the report examines the relationship between incident severity levels and incident types. The key findings from the data exploration are highlighted, including the most common types of incidents and insights into incidents of high severity. Furthermore, the report discusses data cleaning, and addressing any inconsistencies or missing values. This analysis can be utilized to inform decision-making and safety measures in an occupational environment, highlighting the potential benefits of data-driven approaches based on Python programming.

Contents

List of Figures

1 INTRODUCTION

This project involves a comprehensive and in-depth investigation of workplace incidents, including data collection, analysis, and visualization. Our overall objective is to gain profound insights into workplace incidents, shedding light on their frequency, underlying patterns, and levels of severity.

Understanding and mitigating workplace incidents is of paramount importance in Occupational Health and Safety. As a result of such incidents, productivity, safety, and overall operational efficiency may be adversely affected. Using data-driven methodologies, this project seeks to address these concerns head-on, providing a framework for analyzing and improving workplace safety.

The project begins with the generation of synthetic data, which is meticulously crafted to simulate real-world workplace incidents. A synthetic dataset provides a controlled environment for experimentation and serves as the basis for subsequent analyses. This type of data generation technique is particularly valuable in scenarios in which access to extensive real-world data is limited or restricted.

The next step in the project is the exploration of data generated by the data generation phase. The purpose of this phase is to examine in detail the synthetic dataset, delving into various aspects of the incident data. Among the key components of this analysis are summary statistics, a breakdown of incident types and their frequencies, severity distributions, and patterns that may appear over time. The meticulous examination of the dataset enables the project to reveal latent insights that may have been overlooked in the past.

Furthermore, the visualization of data plays a crucial role in elucidating complex patterns and trends. Using visualization techniques, the project seeks to provide stakeholders with a visual representation of incident data that is intuitive and accessible to them. With the aid of meticulously crafted graphs and charts, it becomes possible to identify trends, anomalies, and relationships within the data set.

Crucially, this project recognizes the critical intersection of incident type and severity. It embarks on an exploration of how different incident types correlate with varying levels of severity. This deeper understanding can guide the formulation of targeted strategies for preventing and mitigating high-severity incidents.

1.1 ROLE AND CONTRIBUTIONS

Throughout the execution of this project, I assumed the sole and comprehensive responsibility as its principal contributor. My involvement in this endeavor extended across all critical phases, and I meticulously designed and implemented the entire workflow. I was responsible for data generation, data exploration, data cleaning, and the creation of enlightening visualizations. By combining these tasks, the project's direction evolved from conceptualization to a tangible and insightful data-driven analysis.

2 SYSTEM REQUIREMENTS

a) **Python**
The code is written in Python, so you must have Python installed on your system

b) **Python Libraries**
The script utilizes several Python libraries, which are commonly used in data analysis such as numpy, pandas, matplotlib, and seaborn.

c) **Operating System**
The code should work on various operating systems, including Windows, macOS, and Linux.

d) **Integrated Development Environment (IDE)**
While not strictly required, using Jupyter Notebook, Visual Studio Code, or PyCharm can make it easier to work with the code.

3 PROJECT WORKFLOW

a) **Generating Simulated Data**
Initially, I used Python's NumPy and Pandas libraries to initiate the project. As part of my research, I generated synthetic data that mimicked real workplace incidents with the aim of ensuring precision and statistical accuracy. The generated dataset contained vital attributes, such as incident dates, incident types (including "Falls," "Slips," and "Equipment Malfunctions"), and severity levels (ranging from "Low" to "High"). We meticulously crafted this synthetic dataset to incorporate the nuances of real-world situations. Upon successful generation, the dataset was saved as "accident.csv."

Listing 1: Python code for generating simulated data

```
import numpy as np
import numpy as np
import pandas as pd
np.random.seed(113)

timeframe = pd.date_range(start="2023-01-01", end="2023-12-31",
    freq="D")
typeofincident = np.random.choice(["Fall", "Slips", "Equipment␣
    Malfunction"], size=len(timeframe))
severity = np.random.choice(["Low", "Medium", "High"], size=len(
    timeframe))
data = {
    "Date": timeframe,
    "Incident␣Type": typeofincident,
    "Severity": severity
}
randomdataforworkplaceinjury = pd.DataFrame(data)
randomdataforworkplaceinjury.to_csv("accident.csv", index=False)
print("Data␣has␣been␣saved␣successfully")
```

b) **Exploring the Generated Data**

This project included a critical phase involving a comprehensive analysis of the newly generated dataset. Through the use of Python's Pandas library, I ingested and meticulously examined the data by assessing its fundamental structure and basic statistics. In addition, key statistics were calculated and presented, such as data distribution and summary statistics, which provided valuable insight into the characteristics of the dataset.

Listing 2: Python code for Exploring the Generated Data

```
readingthedata = pd.read_csv("accident.csv")
print(readingthedata)

print(readingthedata.head())
print(readingthedata.tail())
print(readingthedata.info())
print(readingthedata.describe())
```

c) **Data Cleaning**

Any data analysis endeavor must be conducted with the highest level of data quality. Fortunately, the generated dataset was found to be relatively clean, with no missing values to be addressed. It was this fortuitous circumstance that allowed a simplified data analysis process to take place, as the dataset was deemed ready for further analysis.

d) **Data Visualization**

Through the use of Python's Matplotlib library, I have been able to create highly compelling data visualizations. I sought to depict the distribution of incident types in a visually engaging and informative manner. A bar chart was created to illustrate the frequency of each type of incident. The visualization was instrumental in providing a concise overview of the occurrences of incidents within the dataset.

Listing 3: Python code for Data Visualization

```
import matplotlib.pyplot as plt
incidentcounting = readingthedata['Incident␣Type'].value_counts()
plt.figure(figsize=(15, 8))
plt.bar(incidentcounting.index, incidentcounting.values)
plt.xlabel('Incident␣Type')
plt.ylabel('Count')
plt.title('Distribution␣of␣Incident␣Types')
plt.xticks(rotation=45)  # Rotate x-axis labels for better
    readability
plt.show()
```

e) **Identifying Patterns and Trends**

A consideration of the temporal aspect of the dataset was necessary. The "Date" column was meticulously transformed into a datetime format to facilitate the analysis of temporal data. Through this strategic approach, I was able to discern patterns and trends related to incident types over time and identify any noteworthy variations or trends within the dataset.

Listing 4: Python code for Identifying Patterns and Trends

```
readingthedata['Date'] = pd.to_datetime(readingthedata['Date'])
print(readingthedata['Date'])
```

f) **Investigating High Severity Incidents**

An understanding of the underlying factors of high severity incidents was of particular interest. I conducted a dedicated examination of incidents categorized as "High" severity in order to gain a deeper understanding of this matter. This subset of data was carefully preserved as "highseverity.csv." The goal was to identify which incident types were most frequently associated with high-severity outcomes, therefore providing insight into potential areas of concern.

Listing 5: Investigating High Severity Incidents

```
highseverityincidents = readingthedata[readingthedata['Severity']
   == 'High']
print(highseverityincidents)
highseverityincidents.to_csv("highseverity.csv", index=False)
print("Data␣has␣been␣saved␣successfully␣as␣highseverity.csv")
```

g) **Finding out which incident type causes high severity**

This critical question was addressed by strategizing the severity levels and analyzing the mean severity. As a result of this analytical endeavor, valuable insights were gained regarding the factors contributing to incidents of high severity.

Listing 6: Finding out which incident type causes high severity

```
severitymapping = {"Low": 1, "Medium": 2, "High": 3}
readingthedata['Severity'] = readingthedata['Severity'].map(
   severitymapping)
meanseverity = readingthedata['Severity'].mean()
print("Mean␣Severity:", meanseverity)
```

h) **Visualizing Incident Type vs. Severity**

As a culmination of the project's data-driven exploration, I developed a heatmap. By combining Seaborn and Matplotlib, this visualization masterpiece illustrates the intricate relationship between incident types and severity levels in an intuitive way. It provided a visual conclusion to the project's investigative journey, providing a powerful tool for identifying incident types that are most commonly associated with high-severity outcomes.

Listing 7: Visualizing Incident Type vs. Severity

```
import seaborn as sns
incident = readingthedata.pivot_table(index='Incident␣Type',
   columns='Severity', values='Date', aggfunc='count')
plt.figure(figsize=(10, 6))
sns.heatmap(incident, annot=True, cmap='coolwarm')
plt.xlabel('Severity')
plt.ylabel('Incident␣Type')
plt.title('Incident␣Type␣vs.␣Severity')
plt.show()
```

4 RESULT ANALYSIS

The project's comprehensive analysis of workplace incidents using data generation, exploration, and visualization techniques has resulted in valuable insights into incident occurrence, patterns, and severity within the context of industrial and systems engineering, especially focusing on human factors.

a) **Incident Occurrence**

This dataset provides a temporal perspective on workplace incidents for the entire year of 2023. The number of incidents occurred sporadically throughout the year, with some variations in time. It appears that external factors may influence incident occurrence, warranting further investigation.

b) **Incident Types**

There were three primary categories of incident types identified during an analysis: ”Falls,” ”Slips,” and ”Equipment Malfunction.” This insight provides a foundational understanding of the distribution of incident types within the workplace. ”Equipment Malfunction” was the most common incident type, followed by ”Fall” and ”Slips.” See Figure 1.

```
print(readingthedata[['Incident Type']]. value_counts())
```

```
Incident Type
Equipment Malfunction    127
Fall                     126
Slips                    112
Name: count, dtype: int64
```

Figure 1: Incident Type

c) **Severity Levels**

As shown in Figure 2, The severity levels of workplace incidents were classified as ”Low,” ”Medium,” and ”High.” Most incidents were classified as ”Low.” However, a substantial portion of the incidents were classified as ”High.” This distribution illustrates the importance of understanding and addressing high-severity incidents, as they may have more significant consequences for workplace safety and efficiency.

d) **Incident Type vs. Severity**

In addition to bar charts and a heatmap, as shown in Figure 3 and in Figure 4, visualizations were used to illustrate the relationships between incident types and severity levels.

```
print(readingthedata['Severity']. value_counts())
```

```
Severity
Low       139
Medium    120
High      106
Name: count, dtype: int64
```

Figure 2: Severity Levels

```
print(readingthedata.groupby('Incident Type').agg({'Severity': 'count'}))  # Count incidents by type
```

```
                       Severity
Incident Type
Equipment Malfunction       127
Fall                        126
Slips                       112
```

Figure 3: Incident Type vs. Severity

In particular, ”Equipment Malfunction” and ”Fall” incidents were more likely to be classified as ”High” severity incidents. Considering these findings, it may be necessary to implement targeted safety measures and interventions in order to prevent these types of incidents from occurring.

e) **Mean Severity Analysis**

The average severity of all incidents was determined by calculating the mean severity across all incidents. While fluctuations in mean severity over time may indicate changes in workplace safety conditions, they should be closely monitored. See figure 5.

f) **High Severity Incidents**

A subset analysis was conducted on incidents categorized as ”High” in severity. The purpose of this investigation was to identify the most common types of incidents that contribute to high severity. According to the results, ”Equipment Malfunction” was the most common incident type associated with high severity levels. See figure 6.

g) **Insights into Incident Causes**

Calculating the mean severity and mapping severity levels to numerical values provided insight into the factors contributing to high-severity incidents. It is possible to use this quantitative approach as a basis for further root cause analysis and prevention efforts.

According to the findings of this study, data-driven analysis plays a crucial role in Occupational Health and Safety, especially when it comes to addressing human factors related to

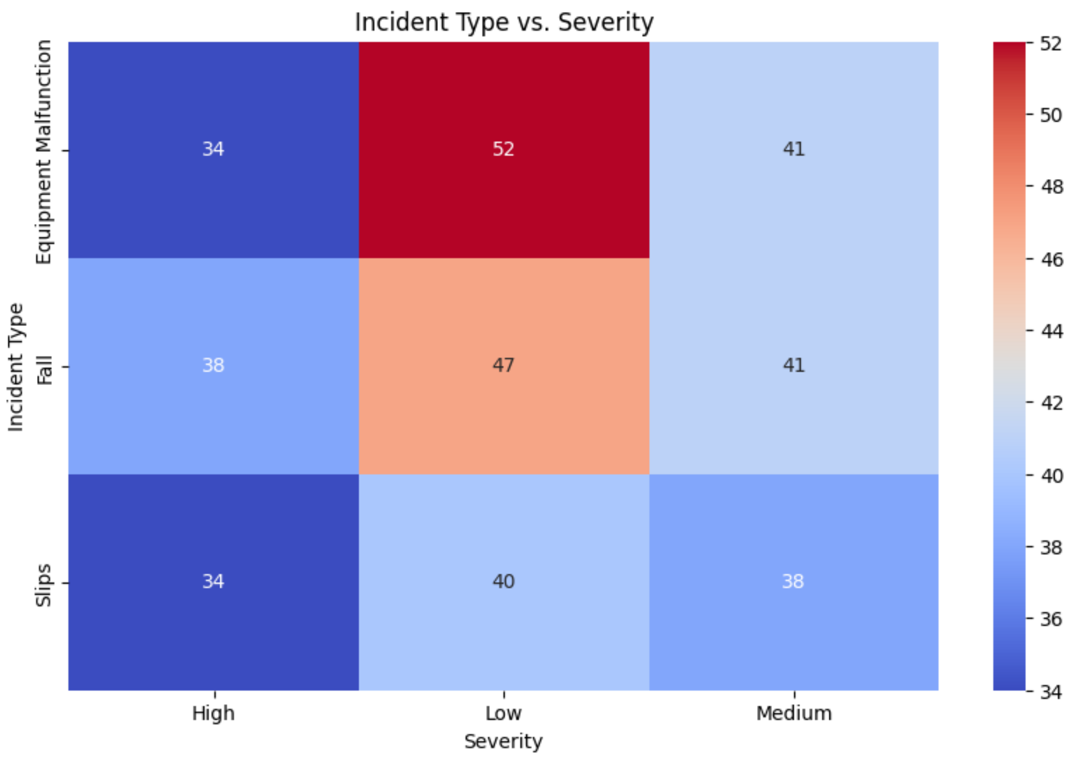

Figure 4: Incident Type vs. Severity Seaborn Map

```
import pandas as pd
readingthedata = pd.read_csv("accident.csv")
severitymapping = {"Low": 1, "Medium": 2, "High": 3}
readingthedata['Severity'] = readingthedata['Severity'].map(severitymapping)
meanseverity = readingthedata['Severity'].mean()
print("Mean Severity:", meanseverity)
```

```
Mean Severity: 1.9095890410958904
```

Figure 5: Mean Severity Analysis

workplace safety. Organizations can proactively enhance safety protocols by understanding incident patterns, distribution, and severity, thereby reducing incident risks and creating a safer and more efficient work environment.

The ”Additional Insights and Implications” section delves deeper into the implications of these findings and highlights opportunities for further research and action.

```
highseverityincidents = readingthedata[readingthedata['Severity'] == 'High']
print(highseverityincidents)
highseverityincidents.to_csv("highseverity.csv", index=False)
print("Data has been saved successfully as highseverity.csv")
```

```
          Date        Incident Type Severity
2     1/3/2023  Equipment Malfunction     High
4     1/5/2023                 Fall     High
6     1/7/2023  Equipment Malfunction     High
7     1/8/2023                Slips     High
14   1/15/2023                Slips     High
..         ...                  ...      ...
341  12/8/2023                Slips     High
342  12/9/2023  Equipment Malfunction     High
344 12/11/2023                 Fall     High
354 12/21/2023                 Fall     High
356 12/23/2023                 Fall     High

[106 rows x 3 columns]
Data has been saved successfully as highseverity.csv
```

Figure 6: High Severity Incidents

5 ADDITIONAL INSIGHTS AND IMPLICATIONS

The project has also uncovered several additional insights and implications that are crucial for increasing workplace safety and efficiency in industrial and systems engineering. These include:

a) **Temporal Trends in Incidents** Upon reviewing the dataset over time, it became apparent that incidents were not distributed evenly throughout the year. The number of incidents increased during certain periods, which could be attributed to various external factors, including seasonal changes, workload fluctuations, or equipment maintenance schedules. Having an understanding of these temporal patterns can assist in resource allocation and scheduling in order to minimize incident risks during times of high risk.

b) **Incident Severity Variation** The analysis of incident severity levels revealed that the majority were classified as having a ”Low” severity level. However, a significant proportion of incidents were classified as ”High” in severity. The observation of high-severity incidents prompts further investigation into their root causes. It is possible to reduce the overall impact on workplace safety and productivity by identifying and addressing these root causes.

c) **Incident Type and Severity Correlation** According to the heatmap visualization, certain incident types, such as ”Equipment Malfunction” and ”Fall,” were more frequently associated with high severity. These correlations indicate that targeted preventive measures and safety protocols are necessary for these specific types of incidents. It may be possible to mitigate the risks associated with these incidents through engineering interventions or training programs.

d) **Mean Severity Analysis** The calculation of the mean severity across all incidents provides a quantitative measure of the overall safety at work. The ability to monitor changes in mean severity over time can serve as an early warning system, alerting organizations to potential safety issues before they escalate. Increasing mean severity may indicate deterioration in safety conditions that requires immediate attention.

e) **Predictive Modeling Opportunities** Despite the fact that this project focused on descriptive analysis, the insights gained can be applied to predictive modeling in the future. On the basis of historical incident data, machine learning algorithms could be trained to predict the likelihood and severity of future incidents. Organizations can use this proactive approach to prevent workplace incidents before they occur by taking preemptive measures.

f) **Policy and Procedural Revisions** According to the findings of this study, workplace safety policies and procedures should be reviewed and revised periodically. The analysis of incident types, severity, and contributing factors can lead to the development of safer workplace policies and safety training programs.

6 CONCLUSION

The purpose of this project was to conduct a thorough investigation of workplace incidents, focusing on principles of Occupational Health and Safety. Utilizing data generation, exploration, and visualization techniques, critical insights have been achieved into incident frequency, patterns, and severity. Notably, ”Equipment Malfunction,” ”Falls” and ”Slips” occur year-round with a tendency to be of ”Low” severity. Safety management must be proactive in the event of high-severity incidents. An important contribution has been the identification of ”Equipment Malfunction” and ”Fall” as incident types prone to ”High” severity, thus emphasizing the importance of targeted preventative measures. The ongoing monitoring of severity revealed fluctuations, emphasizing the need for ongoing safety assessments. A key result of this project is the demonstration of the potential of data-driven decision-making in shaping safety practices, particularly in the context of ”Equipment Malfunction", and the development of predictive models based on this information.

Overall, these data-driven insights offer valuable guidance for enhancing workplace safety in industrial and systems engineering. Organizations can proactively create safer work environments by understanding incidents and their implications. The results of these studies provide practical guidance for safety officers, policymakers, and other professionals seeking to elevate workplace safety standards and practices.