Data Analysis for Safety Management

ZOUAIRI Saim

Institute of Maintenance and Industrial Safety, University of Oran2 Mohamed Ben Ahmed; P.B 1015 El M'naouer 31000 Oran Algeria

E-mail: zouairi.saim@univ-oran2.dz,

Abstract: Data analysis plays an important role in safety management. It can be seen as an important process involving the examination and interpretation of data to identify patterns, trends and recommendations related to safety in a particular environment or context. It helps organizations or individuals to make clear decisions and take the necessary action to improve safety measures and prevent accidents or incidents. One of the most powerful tools in the field of data analysis for decision-making is data visualization. The purpose of data visualization in the data analysis process for safety management is to present complex safety data in a visual format that is easy to understand and interpret. Our work aims to highlight the use of exploratory data analysis, and predictive data analysis in the field of occupational safety and health in order to improve safety levels.

Keywords: predictive maintenance, data analysis , exploratory data analysis, safety management, decision making.

1. INTRODUCTION

In 1996, the American Society of Safety Engineers (ASSE) published a pamphlet entitled Scope and Functions of the Professional Safety Position. This pamphlet provides a superb presentation of why a safety and health professional is essential to the arena of OSH by explaining why the safety and health professional must anticipate, recognize, evaluate, control, and communicate safety and health as their main function (*Professional Safety*, n.d.).

The safety and health professional need to develop methods for anticipating and predicting hazards based on their experience, historical data, and other pertinent sources of information. This will allow for identifying and recognizing hazards in existing and future systems, equipment, products, software, facilities, processes, operations, and procedures during the life expectancy of these various facets of the workplace (Li Vigni et al., 2013).

Data analysis plays an important role in safety management. It can be seen as an important process involving the examination and interpretation of data to identify patterns, trends and recommendations related to safety in a particular environment or context. It helps organizations or individuals to make clear decisions and take the necessary action to improve safety measures and prevent accidents or incidents (Wang et al., 2019).

Data visualization plays a crucial role in enhancing decision-making processes by making data more accessible and understandable, aiding in exploring, analyzing, and selecting optimal decisions (Kharakhash, 2023).

While most visualization tools excel in displaying decision criteria and alternatives, there is a scarcity of tools supporting all decision-making stages effectively. Advanced predictive and investigative insights derived from data visualization can help users understand future trends, reasons behind anomalies, and recommend actions for optimizing outcomes (Rony et al., 2023).

One of the most powerful tools in the field of data analysis for decision-making is data visualization. The purpose of data visualization in the data analysis process for safety management is to present complex safety data in a visual format that is easy to understand and interpret.

In recent years, data analysis has become one of the most important tools in occupational safety and health, due to the technological development that touched all fields and industries, which contributed to the availability of data. The work of (Patriarca et al., 2021) focuses on the importance of learning from previous incidents to improve industrial processes. The study utilizes the Major Hazard Incident Data Service (MHIDAS) database, which contains more than 9000 reports from the 1950s to the 1990s. The goal is to provide a better understanding of industrial incidents and explore ways to analyze them through descriptive and quantitative analyses. To achieve this, the paper suggests using Business Intelligence (BI) tools for data visualization and Machine Learning (ML) algorithms for knowledge extraction. By integrating BI and ML, the study aims to facilitate a thorough investigation of reported safety events in the process industry and highlight the potential for lessons learned.

In the paper (Vaiana et al., 2021) The European Union's road safety policy, based on Directive 2019/1936/EC, emphasizes the importance of road safety inspections (RSI) in reducing accidents and improving road networks. The European Transport Safety Council recommends extending these measures to urban and rural roads. In a study of a high-risk rural road in Southern Italy, the RSI approach identified infrastructure-related features with poor safety conditions. It also found a significant correlation between accidents and road markings gaps, as well as a high density of driveways contributing to crash frequency. Mathematical models were used to analyze the impact of various infrastructure variables on accidents, with a prediction calibrated model proposed as a result.

In the Industry area, as in most research fields, the system complexity to be faced is increasing in the safety and health professional concepts. Therefore, exploratory data analysis (EDA) is well suited. EDA is well known in statistics and sciences as that operative approach to data analysis aimed to improve understanding and accessibility of the results. Without forgetting the soundness of statistical models and hypothesis formulation, which is intrinsically connected to the concept of 'analysis' in its scientific meaning, the focus is moved to 'exploration', which, as a word, leads to more exotic thoughts and feelings, such as unravelling mysterious threads or discovering unknown worlds. As a matter of fact, EDA does relate to the process of revealing hidden and unknown information from data in such a form that the analyst obtains an immediate, direct and easy-to-understand representation of it. Visual graphs are a mandatory element of this approach, owing to the intrinsic ability of the human brain to get a more direct and trustworthy interpretation of similarities, differences, trends, clusters and correlations through a picture, rather than a series of numbers. As a matter of fact, our perception of reality is that we believe what we are able to see.

2. DATA ANALYSIS USEFULNESS

Data analysis for safety management is an important process that involves examining and interpreting data to identify patterns, trends, and insights related to safety in a particular environment or context. It helps organizations or individuals make informed decisions and take necessary actions to improve safety measures and prevent accidents or incidents. Here are some key steps involved in data analysis for safety management:

2.1. Data Collection

Gather relevant data related to safety, such as accident reports, near-miss incidents, safety inspections, employee observations, and any other relevant information.

2.2. Data Cleaning:

Ensure that the collected data is accurate, complete, and free from errors. Remove any duplicate or irrelevant data.

2.3. Data Exploration:

Explore the data to understand its characteristics, such as the distribution of accidents or incidents, the frequency of specific types of incidents, and any patterns that may emerge.

2.4. Data Visualization:

Use charts, graphs, and other visual representations to present the data in a meaningful way. This helps in identifying trends and patterns more easily.

2.5. Data Analysis Techniques:

Apply various statistical and analytical techniques to uncover insights from the data. This may include methods such as trend analysis, root cause analysis, correlation analysis, and predictive modeling.

2.6. Identify Safety Issues:

Analyze the data to identify potential safety issues, recurring patterns, or areas of concern. This can help prioritize safety interventions and allocate resources effectively.

2.7. Action Planning:

Based on the insights gained from the data analysis, develop action plans and strategies to address the identified safety issues. This may involve implementing new safety procedures, providing additional training, or making changes to the physical environment.

8. Monitoring and Evaluation:

Continuously monitor the effectiveness of the implemented safety measures and evaluate their impact. Collect new data over time to assess whether the safety management efforts have resulted in improvements and adjust strategies as needed.

3. RESULTS AND CASE STUDY

The database comes from one of the biggest industries in Brazil and in the world. It is an urgent need for industries/companies around the globe to understand why employees still suffer some injuries/accidents in plants. Sometimes they also die in such environment. Table 1 shows the first 5 lines of this data collection.

First, we started with importing the necessary libraries and adjusting the settings options. After that, we followed three steps to conduct our exploratory data analysis:

- Data collection
- Data cleaning
- Univariate analysis

Last but not least, we wrapped up the process with a conclusion, and we mentioned some limitations that faced us. Figure 1 show that Country_01 has the greatest number of accidents beside the other Countries, that is why we should take some investigation in order to find what is wrong so to correct it.

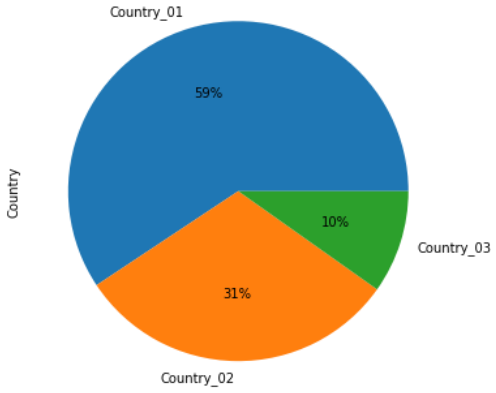

Figure 1. Accidents occurrence by country.

Table 1 the top 5 rows of IHMStefanini industrial safety and health database

	Unnamed: 0	DData	Countries	Local	Industry Sector	Accident Level	Potential Accident Level	Genre	Employee or Third Party	Critical Risk	Description
0	0	2016-01-01 00:00:00	Country_01	Local_01	Mining	I	IV	Male	Third Party	Pressed	While removing the drill rod of the Jumbo 08 f...
1	1	2016-01-02 00:00:00	Country_02	Local_02	Mining	I	IV	Male	Employee	Pressurized Systems	During the activation of a sodium sulphide pum...
2	2	2016-01-06 00:00:00	Country_01	Local_03	Mining	I	III	Male	Third Party (Remote)	Manual Tools	In the sub-station MILPO located at level +170...
3	3	2016-01-08 00:00:00	Country_01	Local_04	Mining	I	I	Male	Third Party	Others	Being 9:45 am. approximately in the Nv. 1880 C...
4	4	2016-01-10 00:00:00	Country_01	Local_04	Mining	IV	IV	Male	Third Party	Others	Approximately at 11:45 a.m. in circumstances t...

Figure 2. Manufacturing plants distribution by local

As we can see, in figure 2, the highest manufacturing plants are located in Local_03 city, and the lowest manufacturing plants are located in Local_09 & Local_11. Hence it is suitable to focus on these locals by multiplying the safety measures so the risk level decrease

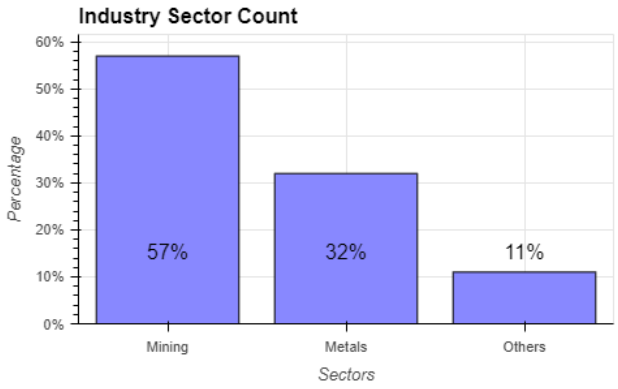

Figure 3. Distribution of "Industry Sector" label

In figure 3, we can see that the highest production rate is in the mining sector (57%), which directly implies that the probability of risk occurring will therefore be higher than in the other sectors.

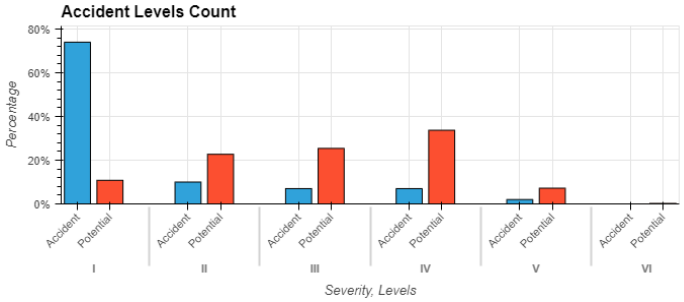

Figure 4. Distribution of "Accident Level" and "Potential Accident Level" label

Figure 3, show that the number of accidents decreases as the Accident Level increases. Whereas the number of accidents increases as the Potential Accident Level increases. Therefore, accidents with level of severity 1 are the most common.

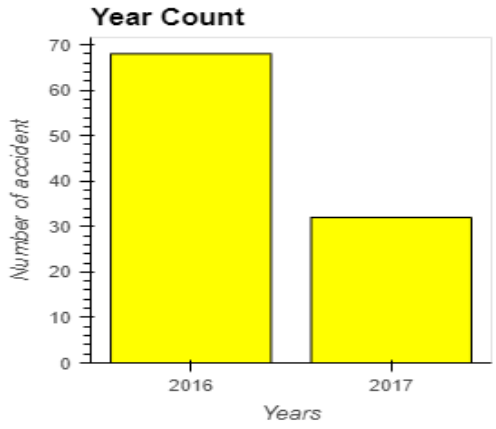

Figure 5. Accidents distribution by years

In figure 5 and figure 6, we can that accidents are recorded from 1st Jan 2016 to 9th July 2017 every month, there is a high number of accidents in 2016 and fewer in 2017.

The number of accidents is high at beginning of the year and it keeps decreasing later.

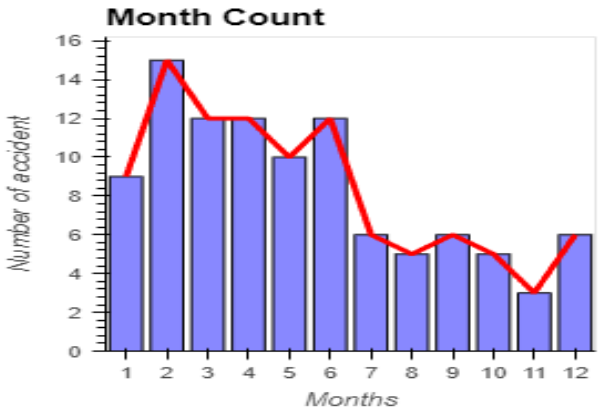

Figure 6. Accidents distribution by months

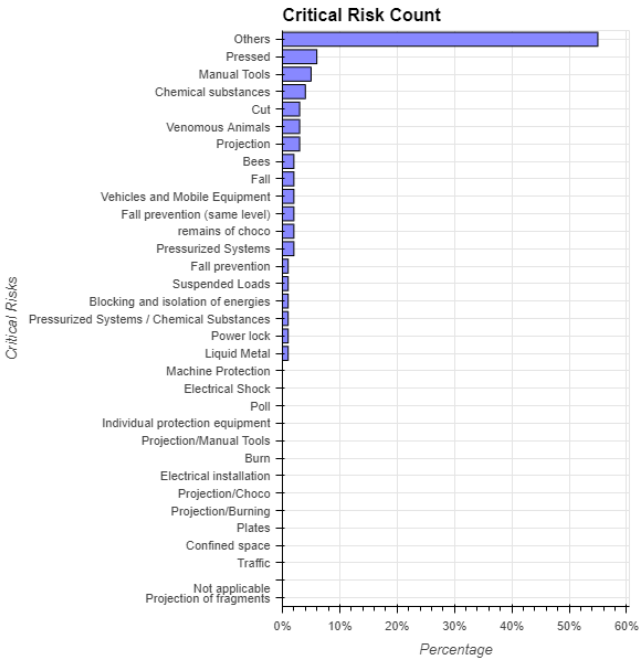

Figure 7. Ratio of critical risk classification

As figure 7 shows, the Critical Risks classified as 'Others' are the most frequent. Therefore, we need to focus on the column "Others", as it represents most of the Critical risks, to closely examine the factors or elements that led to this situation.

4. CONCLUSIONS

We conducted an exploratory analysis of the data using Python; overall, the study was fairly successful as we came up with the following points:

- We cleaned the dataset and make it more useful.
- We had some useful insights and detected patterns in the dataset.

- It was possible to explore the data using Excel, but with Python, the process is easier and fancier, especially for presentations.
- We can do further exploration using NLP analysis. Moreover, even further by applying predictive analysis algorithms.

Data visualization has enabled us to identify the factors which have given rise to undesirable situations, and which need to be verified and followed up with corrections and recommendations to avoid them in the future.

This data representation will facilitate the task of decision-making by providing appropriate instructions and recommendations, while avoiding the complexity of accumulated data in data tables. These instructions will now feed the database of feedback cases, which can be exploited at a later date.

5. LIMITATIONS

- Fewer features are available in the data set.
- Lack of access to richer data.
- The database goes back four years from the time the study was conducted.

REFERENCES

Kharakhash, O. (2023). Data visualization: Transforming complex data into actionable insights. *Automation of Technological and Business Processes*, *15*(2), 4–12.

Li Vigni, M., Durante, C., & Cocchi, M. (2013). Chapter 3—Exploratory Data Analysis. In F. Marini (Ed.), *Data Handling in Science and Technology* (Vol. 28, pp. 55–126). Elsevier. https://doi.org/10.1016/B978-0-444-59528-7.00003-X

Patriarca, R., Di Gravio, G., Antonioni, G., & Paltrinieri, N. (2021). Investigating occupational and operational industrial safety data through Business Intelligence and Machine Learning. *Journal of Loss Prevention in the Process Industries*, *73*, 104608.

Professional Safety. (n.d.). Retrieved June 25, 2024, from https://www.assp.org/publications/professional-safety

Rony, M. M. U., Du, F., Rossi, R., Hoffswell, J., Chhaya, N., Burhanuddin, I., & Koh, E. (2023). *Augmenting Visualizations with Predictive and Investigative Insights to Facilitate Decision Making*. 77–81.

Vaiana, R., Perri, G., Iuele, T., & Gallelli, V. (2021). A comprehensive approach combining regulatory procedures and accident data analysis for road safety management based on the European Directive 2019/1936/EC. *Safety*, *7*(1), 6.

Wang, B., Wu, C., Huang, L., & Kang, L. (2019). Using data-driven safety decision-making to realize smart safety management in the era of big data: A theoretical perspective on basic questions and their answers. *Journal of Cleaner Production*, *210*, 1595–1604. https://doi.org/10.1016/j.jclepro.2018.11.181