VISUALIZATION

BAR PLOTS:

A picture containing chart

Description automatically generated

The bar chart displays the distribution of hazard types in the dataset. Each bar represents a specific hazard type, and the height of the bar indicates the count or frequency of that hazard type in the dataset. The x-axis represents the different hazard types, while the y-axis represents the count of hazards.

From the chart, we can observe that the most frequent hazard type is 'Stop & Correct', followed by 'Unsafe Condition' and 'Quality'. The counts gradually decrease for the remaining hazard types. This information provides an overview of the relative occurrence of different hazard types in the dataset.

The bar chart helps identify the most common hazards and provides insight into potential areas of concern. It can assist in prioritizing risk mitigation strategies, allocating resources, and identifying patterns or trends in hazard occurrences. By visualizing the hazard types' distribution, stakeholders can make informed decisions and take appropriate actions to address the identified hazards effectively.

Chart

Description automatically generated

From the chart, we can observe the distribution of hazards across different types. The most prevalent hazard type appears to be 'Stop & Correct', followed by 'Unsafe Condition' and 'Quality'. The counts gradually decrease for the remaining hazard types. This visualization provides a clear overview of the relative occurrence of different hazard types in the dataset.

The bar chart offers valuable insights into the frequency and distribution of hazards based on their types. It aids in identifying the most common hazard types, which can assist in prioritizing risk management strategies and implementing targeted preventive measures. Additionally, it allows stakeholders to focus on specific hazard types that require more attention or mitigation efforts.

Chart, bar chart

Description automatically generated

The bar chart shows how hazards are distributed based on the time it takes to complete them. Each bar represents a specific completion time bucket, and its height represents the count of hazards in that bucket. This chart helps us understand the distribution of hazards across different completion time ranges. ***It is particularly important because it represents the target variable for your machine learning model.*** By analyzing this chart, we can gain insights into how completion time affects the count of hazards, which can help in predicting completion time and prioritizing urgent hazards. It provides valuable information for developing strategies to manage and mitigate hazards effectively based on their completion time patterns.

Chart, histogram

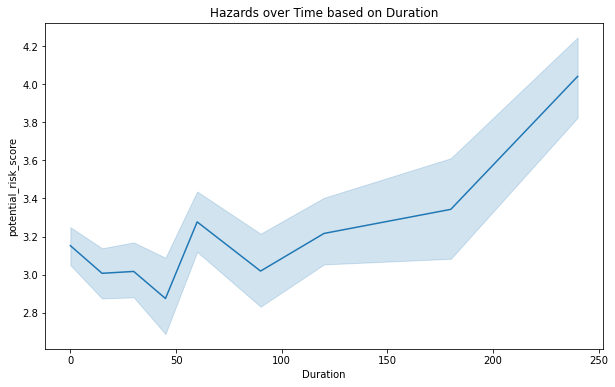
Description automatically generated

Shape

Description automatically generated with medium confidence

The bar chart shows the number of hazards categorized by gender. Each bar represents a gender category, and its height represents the count of hazards associated with that gender. This chart helps us understand the distribution of hazards across different genders. By analysing this chart, we can identify if there are any differences or patterns in hazard occurrences based on gender. It provides insights that can be used to promote workplace safety and address any gender-related issues. This information is valuable for developing targeted safety measures, training programs, or policies to ensure a safe working environment for everyone, regardless of gender.

LINE CHART: Line charts are useful to visualize trends over time. We can use them to see the trend of hazards over time based on the **Duration** column.



The line chart displays the potential risk score of hazards over time based on their duration. The x-axis represents the duration of the hazards, while the y-axis represents the potential risk score.

From the chart, we can observe how the potential risk score of hazards changes over different durations. The line shows the trend of potential risk scores as the duration of hazards increases or decreases. By analyzing this visualization, we can identify any patterns or trends in potential risk scores over time.

This line chart helps in understanding the relationship between the duration of hazards and their potential risk. It provides insights into whether hazards with longer durations tend to have higher or lower potential risks. This information can be crucial for prioritizing hazard mitigation efforts and allocating resources accordingly. By examining the line chart, stakeholders can make informed decisions to effectively manage hazards and reduce potential risks over time.

BOX PLOTS: Box plots are useful to visualize the distribution of continuous data. We can use them to see the distribution of hazard scores based on different columns such as **Hazard\_type**, **potential\_risk**, **further\_action\_required**, etc.

Chart, box and whisker chart

Description automatically generated

The box plot represents the distribution of hazard identification scores based on different hazard types. Each box corresponds to a specific hazard type, and its vertical length indicates the range of scores for that type. The horizontal line within the box represents the median score, while the upper and lower boundaries of the box indicate the interquartile range (IQR), which contains the middle 50% of the scores. The whiskers extend to the minimum and maximum scores within 1.5 times the IQR. From the box plot, we can observe the variation in hazard identification scores across different hazard types. The height and spread of the boxes provide insights into the range and distribution of scores for each type. Comparing the boxes, we can identify hazard types with higher or lower median scores and those with greater score variability.

This visualization helps in understanding the differences in hazard identification scores among different hazard types. It can provide insights into the effectiveness of hazard identification for each type and highlight potential areas for improvement. Stakeholders can use this information to prioritize efforts in enhancing hazard identification practices for specific hazard types, ultimately leading to more effective risk management and mitigation.

Graphical user interface, application

Description automatically generated

HISTOGRAM: Histograms are useful to visualize the distribution of continuous data. We can use them to see the distribution of hazard scores based on different columns such as **hazard\_identification\_score**, **potential\_risk\_score**, **immediate\_action\_score**, etc.

Chart, histogram

Description automatically generated

* This histogram shows the distribution of scores obtained for hazard identification.
* The x-axis represents the score range, and the y-axis represents the frequency or count of hazards falling within each range.
* It helps us understand the distribution and concentration of hazard identification scores in the dataset.

Chart, histogram

Description automatically generated

* This histogram shows the distribution of scores related to potential risk assessment.
* Similar to the previous histogram, the x-axis represents the score range, and the y-axis represents the frequency or count of hazards falling within each range.
* It provides insights into the distribution and concentration of potential risk scores in the dataset.

Chart, histogram

Description automatically generated

* This histogram displays the distribution of scores assigned to immediate actions taken.
* The x-axis represents the score range, and the y-axis represents the frequency or count of hazards falling within each range.
* It gives an overview of the distribution and concentration of immediate action scores in the dataset.

Chart, histogram

Description automatically generated

* This histogram illustrates the distribution of scores assigned to completed actions.
* Again, the x-axis represents the score range, and the y-axis represents the frequency or count of hazards falling within each range.
* It provides insights into the distribution and concentration of completed action scores in the dataset.

CORRELATION PLOT:

Chart, treemap chart

Description automatically generated

EDA

A screenshot of a computer

Description automatically generated

Figure 1 Description of data

Text

Description automatically generated

Figure 2 Data Info (all columns included)

A screenshot of a computer

Description automatically generated

Figure 3 Columns contains null values

A screenshot of a computer

Description automatically generated

Figure 4 columns who has <1000 null values