Depot Incident Analysis: Project Brief

Project Introduction 2
Dataset information 3
Aim & Hypothesis 4
Step-by-Step Process 5
Total Depot Incidents 5
FWI Per Incident 6
Total FWI by Period 8
Results & Findings 10
Challenges 12
Future Enhancements 13
Conclusion 14

Project Introduction

A working group, known as Depot Working Group (DWG) was formed to investigate workforce harm within railway depots, bringing together an analyst from the Rail Safety and Standards Board (RSSB), track workers, depot workers/managers, and key figures from across the rail industry. The primary focus of this initiative was to gain a deeper understanding of the risks and incidents affecting workers in depots and to determine whether targeted interventions were needed to improve safety.

In June 2023, as part of the DWG's ongoing work, a key meeting led to a GB-wide change in how work is carried out within depots. This resulted in the creation of a new Standard developed by RSSB, which introduced updated safety procedures and risk management protocols aimed at reducing depot-related incidents. These changes were driven by evidence from SMIS data and discussions within the working group, emphasising the importance of data-driven decision-making in shaping industry-wide safety policies.

This project builds on the momentum of these safety improvements, using detailed incident analysis to further understand trends, risks, and necessary interventions to enhance depot worker safety.

Measuring Workforce Harm: The Fatalities and Weighted Injuries (FWI) Index

To assess the severity of depot incidents, the analysis incorporated the Fatalities and Weighted Injuries (FWI) Index, a well-established safety metric in the GB rail industry. The FWI index provides a composite measure of risk and harm, weighting different types of injuries based on severity.

The weighting system used in FWI is as follows:

- Fatality = 1.0
- Specified (serious) injury = 0.125
- Severe injury = 0.01
- Non-severe injury = 0.001
- Shock/trauma = 0.001 or 0.01 (depending on severity)

This framework allows for a quantitative comparison of incident severity, where a fatality is considered 8 times more serious than a specified injury, 100 times more serious than a severe injury, and 1,000 times more serious than a non-severe injury. By applying FWI to depot incidents, the analysis could identify which types of accidents had the highest impact on workforce harm and whether interventions were necessary.

Data-Driven Safety Analysis

This project provided an opportunity for RSSB to leverage its Safety Management Intelligence System (SMIS): the rail industry's central platform for health and safety reporting. SMIS is a comprehensive business intelligence tool that collects and provides access to thousands of safety-related events that occur annually across Britain's rail network. Depot incidents fall within the scope of this dataset, and track workers, depot workers, and managers use SMIS to report any accidents.

As part of this project, I was tasked with analysing SMIS data to better understand the profile of incidents occurring in depots. This involved identifying key trends, categorising incidents, and assessing the primary contributors to workforce harm. By utilising data-driven insights, the objective was to provide actionable recommendations that could help enhance safety measures and reduce risks for those working in depot environments.

Data Quality Considerations

Since SMIS relies on front-end user input, the data is susceptible to data quality issues due to variations in how incidents are reported. The system collects information through forms filled out by depot workers, track workers, and managers, which can lead to inconsistent, incomplete, or misclassified data.

During the analysis, missing data was identified, particularly in fields such as personal accident type, requiring careful handling. Some incidents lacked sufficient detail, while others contained inconsistencies in categorisation. To ensure reliable insights, data cleaning and preprocessing were necessary, including:

- Merging related fields to fill in missing details.
- Standardising incident categories for consistency.
- Grouping low-frequency events to avoid fragmentation of data.

These steps helped refine the dataset, allowing for a more accurate representation of depot incidents and ensuring that the findings were based on robust and meaningful insights.

Dataset information

The dataset used in this analysis is an adaptation of data from the Safety Management Intelligence System (SMIS). To ensure confidentiality and compliance with data protection policies, the data has been randomised and anonymised before analysis. The original dataset was extracted using SQL and exported as CSV files for processing.

Data Structure & Key Columns

The dataset contains information on total incidents in depots across various categories. Key columns include:

- **SMIS Reference:** A unique identifier assigned to each reported event.
- **Event Date:** The date when the incident occurred.
- **Period:** A time-based grouping following ‘yyyypp’ format (e.g. 2020 period 1 = 202001) to analyse trends over time.
- **Place:** A location descriptor indicating where the incident happened.
- **Report Title:** A summary of the incident type, providing a high-level categorisation.
- **Personal Accident Type:** If applicable, specifies the type of personal accident involved (e.g., slips, trips, falls, awkward body movement).
- **Description:** A textual field detailing what happened during the incident.

This structured dataset enables the identification of trends, key contributors to harm, and potential areas for safety improvement within depots.

Aim & Hypothesis

Aim

The aim of this analysis was to investigate incidents occurring in railway depots and assess whether recent changes to safety procedures have led to measurable improvements. By monitoring trends over time, and identifying the primary contributors to harm for depot workers, this study aims to provide data-driven insights into workplace safety risks. By analysing reported incidents, the goal was to determine whether specific types of accidents, particularly accidents like slips, trips, falls, and awkward body movements pose a significant risk and whether additional safety measures are required to mitigate these risks and enhance worker safety.

Hypothesis

This analysis was guided by the following hypotheses:

1. There has been a decrease in depot incidents over time.

2. Certain incident types still contribute disproportionately to harm and require targeted interventions.

By testing these hypotheses, the analysis aimed to provide data-driven insights to support decision-making around depot safety improvements.

Step-by-Step Process

Total Depot Incidents

Data Preparation

To prepare the data, a new column, ‘combined_report_title’, was created by merging two existing columns from the dataset:

- ‘personal_accident_type’, which was preferred due to its level of detail, and
- ‘report_title’, which was used when the ‘personal_accident_type’ was missing.

Additionally, incidents were categorised, with low-frequency events grouped under an ‘Other’ category to improve the overall clarity and organisation of the dataset. Duplicate SMIS event entries were also removed in order to prevent double counting.

Incident Categorisation

The incidents were then classified into the following key categories:

Table 1 Combined_report_title categorisation for different incident types

Slips, trips, and falls	Awkward body movement
Interaction with plant, machine, equipment, or tool	Railway operating incident and SPADs
Train derailment, failures, and faults	Train striking or struck by an object, vehicle, or animal
Public behaviour & prohibited area incidents	Non-rail vehicle collisions
Irregular signal aspect sequence	Other (low-frequency/unclassified events)

Data Analysis & Visualisations

Total Incident Count Over Time

To analyse trends in depot incidents, the data was grouped by the ‘period’ column, allowing for a structured view of how incidents fluctuated over time on a period-by-period basis. A new column, **‘TotalDepotIncidents’**, was created to sum the number of reported incidents per period. This provided a **clear time-series representation** of depot incident occurrences.

Trend Analysis

To account for short-term fluctuations and highlight long-term patterns, a 13-period annual moving average was applied. This smoothing technique helped identify sustained increases or decreases in depot incidents, reducing the impact of random variability in the dataset.

Statistical Confidence Intervals

To assess whether observed changes in incident counts were statistically significant, 95% confidence intervals were calculated. These provided upper and lower bounds, allowing for a better understanding of whether fluctuations were due to random variation or indicative of a meaningful long-term trend.

Visualisation: Stacked column Chart

To visually represent incident trends, a stacked column chart was used, with a dashed line for confidence bounds and a line for the annual moving average. This chart displayed incident categories over time, making it easier to identify which types of incidents were increasing or decreasing. By categorising depot incidents in this way, the visualisation helped pinpoint emerging risks and areas requiring further investigation or intervention.

FWI Per Incident

Data Preparation

To prepare the data, a new column, event_type_combined’, was created by merging two existing columns from the dataset:

- ‘personal_accident_type’, which was preferred due to its level of detail, and
- ‘sub_event’, which was used when the ‘personal_accident_type’ was missing.
- If neither ‘personal_accident_type’ or ‘sub_event’ had a value, it would then be assigned the value ‘Other’

Additionally, incidents were categorised, with low-frequency events grouped under an ‘Other’ category to improve the overall clarity and organisation of the dataset. Duplicate SMIS event entries were also removed in order to prevent double counting.

Incident Categorisation

The incidents were then classified into the following key categories:

Table 2 event_type_combined categorisation

Slip, trip, fall	Interaction with plant, machine, equip or tool
Hazardous susbstance[1]	Awkward body movement
Interaction with train	Other (not in scope)

Data Analysis & Visualisations

Grouping and Summation

Events were grouped by period and category, and the FWI values were summed for each group. This provided the basis for a breakdown of total harm per incident type over time.

Calculating FWI per Incident

To assess the severity of incidents relative to their frequency, FWI per incident was calculated by dividing the total FWI by the count of events in each period and category. This allowed for a standardised measure of harm per incident, rather than just absolute totals.

Summing for Total FWI

A new column, 'TotalFWI', was created by summing the FWI values across all categories for each period. This provided a high-level view of the total harm occurring in depots over time.

Applying a 13-Period Annual Moving Average

To smooth fluctuations and better identify long-term trends, both FWI per incident and TotalFWI were converted into a 13-period annual moving average. This approach helped reduce short-term variability and provided a clearer picture of whether depot safety was improving or deteriorating over time.

These steps ensured that the FWI per incident chart accurately reflected workforce harm trends, allowing for a data-driven assessment of safety improvements and the need for further interventions.

[1] There was an identified error in the spelling of event_category 'Hazardous susbstance'. The charting of this data was used to correct this.

Visualisation: Line graph

For the analysis of Fatalities and Weighted Injuries (FWI) per incident, a line graph was chosen to represent the time series data. Line graphs are particularly effective for displaying trends over time, making them ideal for tracking changes in safety metrics across multiple periods. The use of a time series allowed for clear visualisation of how FWI values evolved, highlighting trends, fluctuations, and long-term patterns in workforce safety.

The line graph was selected due to its ability to effectively convey the progression of FWI over time, enabling easy identification of both upward and downward trends. By connecting data points in a continuous line, the graph offers a clear view of how safety has varied in the depot environment across different periods. This makes it easier to detect periods of improvement or deterioration, allowing for a more nuanced understanding of safety dynamics.

Total FWI by Period

Data Preparation

To facilitate the analysis of injury data, a series of data preparation steps were undertaken. The primary dataset, smis_df1, underwent transformations to create aggregated views suitable for examining trends across different periods and injury severity levels.

First, the granularity of the injury degree information within the injury_degree_2013 column was simplified. A mapping dictionary, injury_map, was defined to consolidate various descriptions into a more manageable set of categories: 'Fatal', 'Non-severe', 'Severe', 'Shock/Trauma', and 'Specified'. This mapping was then applied to the injury_degree_2013 column, and the resulting simplified categories were stored in a new column named injuries simplified. This step aimed to reduce the complexity of the injury classifications for subsequent analysis.

Next, two aggregated dataframes were created to provide different perspectives on the injury data.

The first aggregation, stored in smis_counts1, focused on the total count of injury incidents within each period. This was achieved by grouping the smis_df1 dataframe by the 'period' column and counting the occurrences within the newly created injuries simplified column. Missing values in the counts were explicitly filled with zero to ensure data completeness.

The second aggregation, resulting in the smis_counts2 dataframe, aimed to quantify the sum of 'fwi_2013' values for each injury category within each period. This involved grouping smis_df1 by both 'period' and the simplified 'injuries simplified' categories. The 'fwi_2013' values were then summed for each combination, and the results were unstacked to have 'injuries simplified' as separate columns. Any missing values resulting from periods with no incidents in a particular injury category were filled with zero. Finally, a 'TotalInjuryIncidents' column was added to smis_counts2 by summing the 'fwi_2013' values across all injury categories for each period. This provides an overall measure of the total 'fwi_2013' associated with injury incidents in each period.

These data preparation steps transformed the initial dataset into formats suitable for analysing the distribution and magnitude of different injury types across various time periods.

Data Analysis & Visualisations

Grouping and Summation

Events were grouped by period and category, and the FWI values were summed for each group. This provided the basis for a breakdown of total harm per incident type over time.

Summing for Total FWI

A new column, ‘TotalInjuryIncidents’, was created by summing the FWI values across all categories for each period. This provided a high-level view of the distribution of total harm occurring in depots over time.

Applying a 13-Period Annual Moving Average

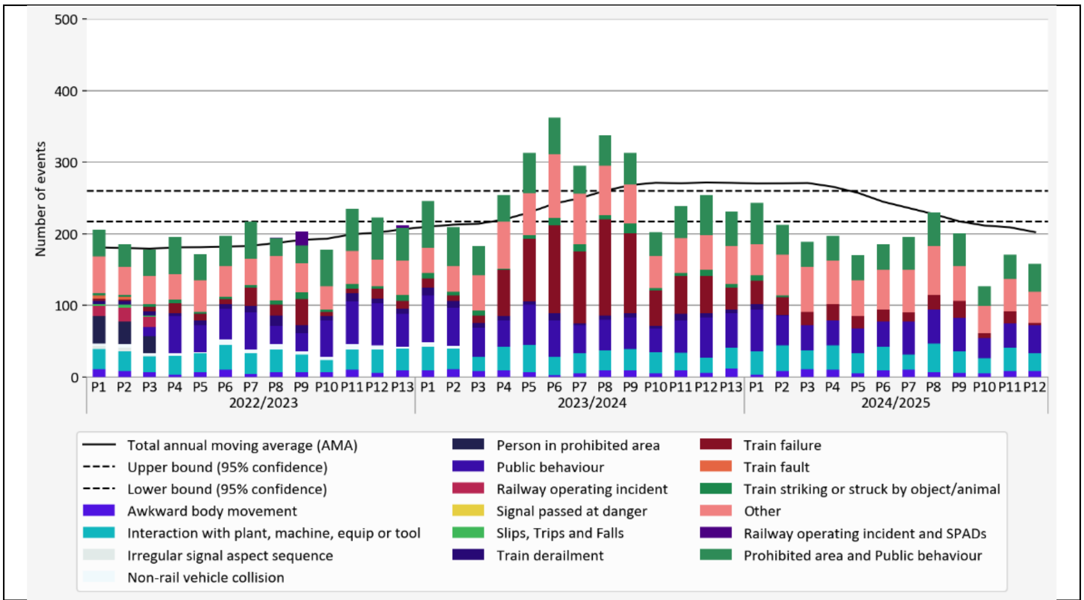
To smooth fluctuations and better identify long-term trends, both FWI per incident and TotalFWI were converted into a 13-period annual moving average. This approach helped reduce short-term variability and provided a clearer picture of whether depot safety was improving or deteriorating over time.

These steps ensured that the total FWI distribution chart accurately reflected long term workforce harm trends, allowing for a data-driven assessment of safety improvements and the need for further interventions.

Visualisation: stacked column chart

A stacked column chart was chosen to visualize the Fatal and Weighted Injuries (FWI) because it effectively displays the contribution of each injury type ('Fatal', 'Specified', 'Severe', 'Non-severe', and 'Shock/Trauma') to the total FWI within each period. This allows for easy comparison of the magnitude of different injury types and how their relative proportions change over time, providing a clear understanding of the factors contributing to the overall FWI trend. The inclusion of the total annual moving average and confidence intervals further contextualizes these contributions within the broader trend and its variability.

Results & Findings

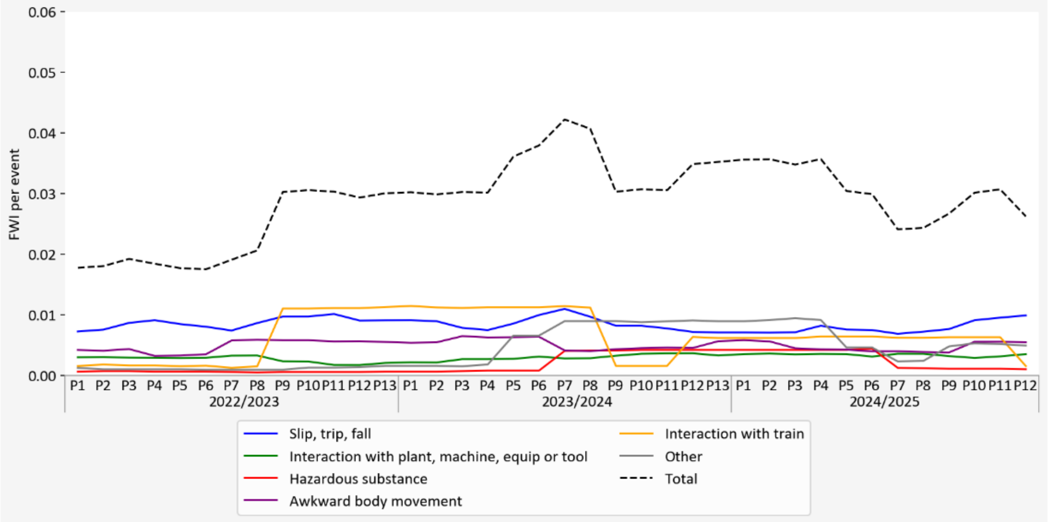


This chart shows the total number of events recorded as happening in depots, broken down by the main category selected for event entry.

Over the past 9 periods (P4 2024/25) There has been a decrease in the number of reported incidents. Since period 9 (P9) 2024/25, the annual moving average has decreased below the 95% confidence bands compared to the P1 2022/23 (April 2022) to P12 2023/24 (March 2024) baseline, indicating we can be 95% confident there has been a statistically significant decrease.

Data

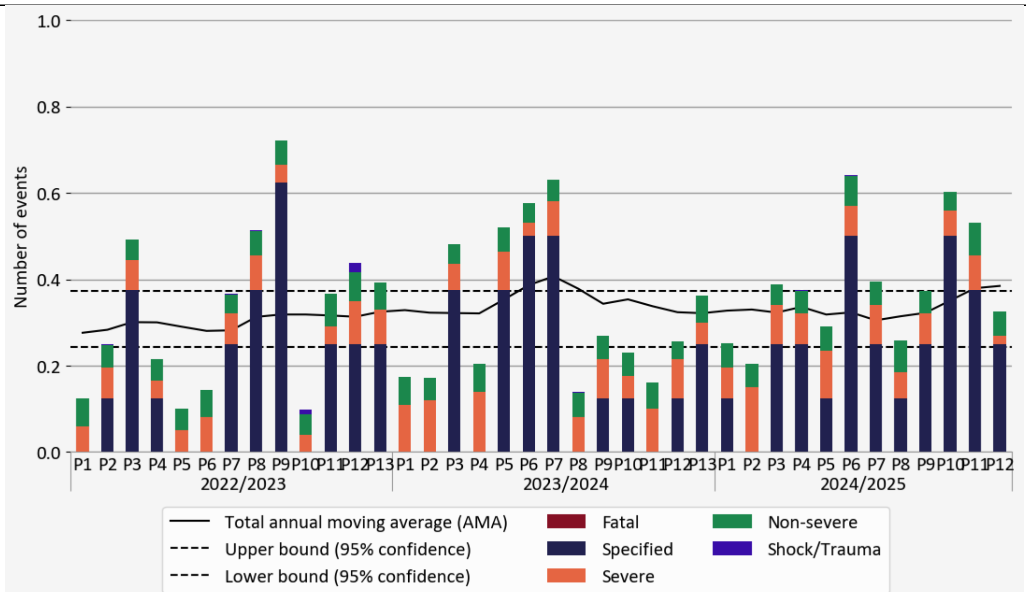
The data in this chart shows all events recorded in SMIS that have occurred in depot locations.



This chart shows the trend in workforce harm (FWI) per event due to personal accidents in depots. This is broken down by the cause of the event over time, spanning from 2022/2023 to 2024/2025. The total FWI per event shows a peak in late 2023/early 2024 and generally fluctuates between 0.02 and 0.04. Since P7/P8 2023/24, the total FWI per incident has decreased, meaning that on average there has been a decrease in the safety consequence (in terms of FWI) of personal accident events recorded in SMIS over the reporting period.

Data

The data in this chart accounts for all events recorded in SMIS that have occurred in depot locations with known injury severities in FWI.



This chart shows the weighted distribution in workforce harm (FWI) by its injury severity due to personal accidents in depots. The chart shows data from 2022/2022 to 2023/2025. The black line represents the total annual moving average (AMA) of FWI, showing a general increasing trend from

2022/2023 to 2024/2025. The chart also includes upper and lower confidence bounds (dashed lines), indicating the variability around the average where between 2023/2024 P6 and P7 there was a statistically significant increase in workforce harm, and 2024/2025 P12 on the P1 2022/23 (April 2022) to P12 2023/24 (March 2024) confidence bounds.

Among the injury types, 'Specified' (dark blue) injuries contribute significantly to the overall FWI, with 'Specified' injuries showing a notable increase in the middle periods of the years 2023/2024. 'Severe' (orange) and ‘Non-severe’ (green) exhibits stable total FWI per period.

Data
The data in this chart shows all events recorded in SMIS that have occurred in depot locations where there has been a reported injury.

Challenges

During the analysis of depot incidents, several challenges were encountered that required data preprocessing and careful methodological considerations. These challenges primarily related to data completeness, categorisation complexity, and variance in reporting.

Data Completeness

One of the key challenges was missing values in the ‘personal_accident_type’ column. In some cases, incidents lacked a specific personal accident classification, making it difficult to analyse accident trends accurately. To address this, missing values were supplemented using the ‘report_title’ column, which provided a general description of the incident. This approach ensured that each incident had a meaningful classification, improving the overall dataset integrity.

Additionally, duplicate event entries were identified based on the SMIS Reference, where the same incident was logged multiple times. These duplicates were removed to ensure that incident counts were not artificially inflated, maintaining data accuracy and preventing skewed results in the analysis.

Categorisation Complexity

Many incidents did not fit neatly into a single category, as some overlapped between multiple classifications. This complexity required careful manual review of some data and categorisation to ensure consistency in the dataset. A standardised mapping process was implemented, grouping related incidents together and consolidating low-frequency events into an 'Other' category to avoid fragmentation.

This issue also highlights a broader challenge for users reporting incidents in SMIS. The complexity of categorisation means that workers, managers, and reporters may struggle to accurately classify incidents, leading to inconsistencies in reporting.

This will be further discussed in the Future Enhancement and Conclusions sections, where recommendations will be made to simplify incident reporting forms to improve data quality and usability.

Variance in Reporting

Differences in how incidents were reported across various periods and by different individuals posed another challenge. Variations in terminology, level of detail, and reporting practices affected the direct comparability of data over time. This required the use of data normalisation techniques to ensure that trends and patterns remained statistically valid and reflective of actual depot safety risks.

By addressing these challenges, the dataset was refined for more accurate trend analysis, ensuring that insights derived from the data were both reliable and actionable for improving depot safety.

Future Enhancements

To further improve the analysis of depot incidents and enhance workforce safety, several key advancements could be implemented:

Refining Incident Categorisation

The current classification process relies on manual grouping, which can introduce inconsistencies and limit deeper insights. Future enhancements could incorporate machine learning clustering techniques to identify hidden patterns within incident data, ensuring that similar events are categorised more accurately. This would help reduce misclassification errors and provide more detailed insights into key risk factors.

Predictive Modelling

By leveraging historical incident data, predictive models could be developed to forecast future trends in depot incidents. Using statistical and machine learning techniques, these models could help proactively identify periods of increased risk, allowing for early intervention and the implementation of targeted safety measures before incidents occur.

Granular Risk Assessment

A more detailed risk assessment approach could be introduced by analysing incidents based on specific variables such as job roles, time of day, and location within the depot. Understanding which roles or locations are most at risk could enable more tailored safety strategies to reduce incidents and improve worker protection.

Automated Reporting Dashboards

The development of real-time visualisation tools could significantly enhance how depot incidents are monitored. By implementing automated reporting dashboards, stakeholders would be able to track emerging trends dynamically without the need for manual data extraction and processing. This would allow decision-makers to respond to safety concerns more efficiently and implement measures in real time.

Conclusion

The Depot Incident Analysis project has provided valuable insights into the safety risks and incidents affecting workers in railway depots. Through the collaborative efforts of the Depot Working Group (DWG) and the implementation of the new safety Standard by the Rail Safety and Standards Board (RSSB), significant strides have been made in enhancing depot safety. The analysis of the Safety Management Intelligence System (SMIS) data has revealed key trends and contributors to workforce harm, highlighting the importance of data-driven decision-making in shaping industry-wide safety policies.

The use of the Fatalities and Weighted Injuries (FWI) Index has allowed for a quantitative assessment of incident severity, identifying the types of accidents with the highest impact on workforce harm. The data-driven approach has provided actionable recommendations for targeted interventions, leading to measurable improvements in depot safety. The analysis has shown a statistically significant decrease in the number of reported incidents and a reduction in the safety consequences of personal accident events over the reporting period.

Despite the progress made, challenges such as data completeness, categorisation complexity, and variance in reporting remain. Addressing these challenges through future enhancements, such as refining incident categorisation, predictive modelling, granular risk assessment, and automated reporting dashboards, will further improve the analysis of depot incidents and enhance workforce safety.

In conclusion, the Depot Incident Analysis project has demonstrated the effectiveness of collaborative efforts and data-driven approaches in improving depot safety. Continued focus on refining data quality and implementing advanced analytical techniques will ensure sustained improvements in the safety and well-being of depot workers.