# Project Report For CS661: BIG DATA VISUAL ANALYTICS

2024-2025 Semester II

## Project Title:

Root Cause Analysis of Industrial Accidents

## Team Members:


Chandrabhan (241110017) — chandrabha24@iitk.ac.in

Ashvani Kumar Yadav (241110013) — ashvaniy24@iitk.ac.in

Yugul Kishor Pandey (241110083) — yugulk24@iitk.ac.in

Sushant Shantaram Gangurde (241110073) — sushants24@iitk.ac.in

Shivam Maurya (241110067) — shivammrya24@iitk.ac.in

Nayan Das (241110045) — nayand24@iitk.ac.in

Rajender (241110055) — rajenders24@iitk.ac.in

Shekhar Pratap Singh (241110066) — shekharp24@iitk.ac.in

Lt Cdr Rakesh Dash (241110088) — rakeshdash24@iitk.ac.in

## IIT Kanpur

## 1. Introduction:

Industrial accidents pose significant challenges to workplace safety and industrial operations. Analyzing patterns in industrial accidents can provide crucial insights for preventing future incidents and improving safety measures. This report presents an extensive exploratory data analysis of an Indian Industrial Accidents dataset, which compiles records of industrial accidents across multiple states in India over several years. The dataset contains 6500 accident records from 2015 through 2020, encompassing various attributes such as the date and time of the accident, location (state and local area), industry sector, accident type, severity of outcomes, demographic information of victims, and contributing risk factors. Key fields in the data include the Year of occurrence, Month, Day of Week, Shift (e.g., Morning, Afternoon, Night), State of India where the accident occurred, Industry Sector (industry category of the site, such as Steel, Cement, etc.), Accident Severity (categorized outcomes like Minor Injury, Severe Injury, or Permanent Disability), Accident Type (e.g., Fire, Explosion, Fall from Height), Gender and Age of the victim, Employee Type (role of the person – e.g., Worker, Contractor, Outsider, etc.), Critical Risk factor involved (e.g., Slip/Trip, Chemical Exposure), whether Safety Gear was used, and other details. By examining these features in depth, this analysis seeks to uncover temporal trends, geographic hotspots, high-risk industries and activities, demographic patterns in accident involvement, and the effectiveness of safety interventions. The importance of this study lies in its potential to inform industrial safety policies. India, with its large and diverse industrial sectors (from heavy manufacturing and mining to pharmaceuticals and electronics), experiences numerous industrial accidents annually. Understanding when, where, and why these accidents occur can help allocate safety resources more effectively and develop targeted prevention strategies. The analysis is presented in an academic style, with structured sections covering the problem definition, methodology, detailed results with visualizations, and conclusions. All insights and figures are derived entirely from the provided dataset and code, ensuring that the conclusions directly reflect the data at hand. In what follows, we outline the tasks and objectives of the analysis, describe the solution approach (including data processing and visualization techniques), present the results organized by theme (temporal, geographic, industrial, demographic, and risk analyses), and finally summarize the key findings along with suggestions for future work to enhance industrial safety.

## 2. Tasks:

The primary task of this project is to perform a comprehensive analysis of the industrial accident dataset and present the findings in a clear, insightful manner. This entails several specific objectives:

**Data Understanding and Preparation**: Interpret the dataset structure (columns and their meanings) and ensure the data is ready for analysis (e.g., correct data types for dates and categories). This includes recognizing that the data spans multiple years (2015–2020) and covers many states and industries across India.

**Exploratory Data Analysis (EDA)**: Examine the data from multiple perspectives to identify patterns and trends:

**Temporal Trends**: Analyze how accident frequencies vary over the years, months, days of the week, and work shifts. Determine if there are increasing or decreasing trends over time and identify any seasonal or weekly patterns in accident occurrence.

**Geographic Distribution**: Assess which states and local areas have the highest number of accidents. Identify geographic hotspots and visualize the state-wise distribution of accidents (e.g., using maps or charts) to highlight regions with elevated incident rates.

**Industry and Severity Analysis**: Determine which industry sectors account for the most accidents and what types of accidents are most common in each sector. Also, analyze the distribution of accident severity (e.g., proportion of minor vs. severe injuries) and potential severity ratings. Identify if certain industries or accident types are associated with more severe outcomes.

**Demographic Analysis**: Investigate the characteristics of individuals affected by accidents, including gender, age, and employee type. For example, measure the gender ratio of accident victims, identify the most affected age groups, and see which categories of workers (full-time employees, contractors, third-party, public, etc.) are most involved in incidents.

**Risk Factor Analysis:** Explore the critical risk factors noted in the data (such as equipment failures, human errors like slips/trips, chemical exposures, etc.) and their relationship to accident occurrence. Assess the role of safety measures – particularly the use of safety gear – in mitigating accident severity. For instance, compare accidents where safety gear was used versus not used to evaluate differences in injury severity outcomes.

**Interactive Dashboard Implementation**: Although the focus of this report is on the analysis results, the project involved building a Streamlit-based interactive dashboard (per the provided code) to enable users to explore the data dynamically. This included implementing filters (by state and accident severity) and creating multiple visualization tabs (Overview, Temporal Analysis, Geographic Analysis, Industry Analysis, Demographic Analysis, Risk Analysis, Conclusions) so that users can slice the data and view corresponding charts. Visualization and Interpretation: For each aspect of the analysis, create appropriate charts or tables (bar graphs, pie charts, heatmaps, etc.) to visualize the findings. The task includes not only generating these visualizations but also providing clear explanations and interpretations for each: identifying what the visualization shows and what insight can be drawn from it.

**Reporting:** Compile the analysis, visualizations, and interpretations into a structured, academic-style report with clearly labeled sections (Introduction, Task, Proposed Solution, Result, Conclusion). The report should be comprehensive (aiming to exceed 20 pages in a standard format), well-formatted with headings and subheadings, and written in a professional tone. It should summarize the findings from the data, discuss their implications for industrial safety, and provide conclusions and recommendations, including suggestions for future work or how the analysis could be extended.

By addressing these tasks, the project seeks to provide a holistic understanding of industrial accidents in the dataset, uncovering when and where accidents are most likely to occur, what factors contribute to severe incidents, and how different groups are affected. The ultimate goal is that these insights could guide improvements in safety practices and policies in industrial workplaces.

## 3. Proposed Solution:

### Data and Methodology

To achieve the above objectives, we adopted a data-driven exploratory analysis approach, implemented in Python. The provided dataset Indian$_Industrial_Accidents.csvwasloadedusingpandas, w$
$structuredandconsistent(allrecordsincludedtheexpectedfieldswithplausiblevalues).Wedidensurethat$
$Theanalysiswasorganizedintothematicareas(temporal, geographic, industry, demographic, andriskf$

**Temporal Analysis:** We aggregated accident counts by year, month, day of week, and work shift. This involved computing yearly totals, monthly totals (and creating a pivot table of Year vs Month to capture seasonal patterns), counts by each weekday, and counts by shift (Morning/Afternoon/Night). Visualization methods included bar charts for year-wise trends and shift distributions, a pie chart or bar chart for day-of-week distribution, and a heatmap for the Year-Month pivot to reveal any seasonality. We also calculated the percentage of accidents occurring during regular working hours vs. overtime to understand time-of-day risk.

**Geographic Analysis:** We counted accidents per state to see which states had the highest frequencies. These were visualized initially in the Streamlit app with an interactive map of India using each state's latitude/longitude (provided in the code) and bubble sizes proportional to accident counts. In this static report, we represent state-wise distribution with bar charts (highlighting top states) for clarity. Additionally, we examined the data at a more granular "Local" level (specific site or area names within states) to find local hotspots, listing the top 10 local areas with the most accidents.

**Industry Analysis:** We grouped the data by industry sector to identify which industries have the most accidents. A bar chart was used for the distribution of accidents across industry sectors. Furthermore, we performed cross-tabulations such as industry by accident type (to see, for example, which types of accidents are prevalent in which industries) and industry by critical risk factor, using more complex visualizations like sunburst charts or treemaps in the interactive dashboard. For this report, we describe these patterns in text and use illustrative charts (e.g., accident counts by industry sector). The analysis also considered safety practices per industry, for instance by examining the usage of safety gear in each industry (though the detailed visualization of that is omitted in this static format, it was available in the interactive app).

**Demographic Analysis:** We analyzed the gender and age distribution of accident victims, as well as their employment role. Gender distribution was straightforward (counting male vs female victims overall, and under various filters). Age distribution was handled by creating age bins (e.g., 18–22, 23–27, ..., up to 60–65 years) to see which age groups had more accidents. We used histograms or bar charts to visualize the age profile of accident-involved persons. The employee type field was analyzed via frequency counts to determine what fraction of accidents involved regular workers versus contractors, outsiders, etc. Where relevant, we produced grouped bar charts (for example, age group by gender, or gender by accident severity) to explore if certain demographics were more prone to severe accidents.

**Risk Factor Analysis:** We delved into fields that relate to causes and severity mitigation. The "Accident Type" was analyzed in conjunction with "Accident Severity" to see which types of accidents tend to result in severe injuries or worse. A cross-tab of Accident

Severity vs Accident Type was generated and visualized (in the dashboard, a sunburst chart was used to show the share of each accident type within each severity category). The effect of Safety Gear usage was critically examined: we computed, for accidents where safety gear was used vs not used, the breakdown of severities. This allows us to see, for instance, the percentage of accidents that were severe when no safety gear was worn compared to when proper gear was used. We present this concept in the discussion, noting trends (e.g., higher severe injury rates in the absence of safety equipment) to infer safety gear effectiveness. Additionally, we considered "Critical Risk" factors (underlying risk conditions recorded for each accident, such as "Fall from Height", "Chemical Exposure", etc.) and looked at their relationship with accident types and outcomes. A heatmap of accident frequency by day-of-week and shift (another risk-related perspective) was also created to pinpoint the most high-risk time periods (e.g., perhaps night shifts on weekends). Lastly, to explore multiple risk factors in combination, the interactive solution included a parallel categories diagram linking Industry Sector, Critical Risk, and Safety Gear usage, illustrating common combinations (for example, a particular industry frequently encountering a specific risk while not using safety gear). In this report, we describe notable patterns from such multidimensional analysis.

**Implementation Tools** The analysis and dashboard were implemented in Python using the following key libraries (as specified in requirements.txt):

**pandas (v2.2.1):** for data manipulation, grouping, and creating summary tables.

**plotly (v5.18.0):** for creating interactive visualizations in the dashboard (bar charts, pie charts, line charts, treemaps, sunburst plots, heatmaps, etc.). Plotly allowed dynamic and aesthetically appealing charts.

**Matplotlib/Seaborn:** for static plotting in this report context (to generate figures for the paper). Although the dashboard used Plotly, for the static report we generated equivalent charts using Matplotlib for consistency in format.

**streamlit (v1.32.0):** to build the web-based dashboard interface (not directly relevant to the report content, but it facilitated user interaction in the live app by providing sidebar filters and tabs).

**folium/geopandas:** These were planned for map visualizations. In practice, the statewise map was achieved using Plotly's Scattermapbox with coordinates for each state, and folium was not extensively used in the final app. Geopandas would have been used if we had shape files for states to create chloropleth maps, but instead a simpler approach was taken due to available data (latitude/longitude for states and accident counts).

The interactive dashboard structure (as gleaned from app.py) was organized into multiple tabs, each corresponding to a part of the analysis (Overview, Temporal, Geographic, Industry, Demographic, Risk, and Conclusions). The "Overview" tab provided high-level metrics (total accidents, number of states, number of industry sectors covered) and a few

key charts to summarize the data distribution. Subsequent tabs went into deeper analysis as outlined above. The dashboard also included sidebar filters allowing the user to focus on a particular state or a particular accident severity level. If a filter was applied, all visualizations would update to reflect the filtered subset of data – a feature that enabled interactive drill-down (for example, one could select a specific state to see that state's temporal trend, industry distribution, etc., or select only severe accidents to see patterns for the most serious incidents).
For the purposes of this static report, all analysis is presented on the full dataset (i.e. no filter applied, unless stated otherwise). However, it is worth noting that the ability to filter by state or severity in the app was valuable for stakeholders who might want to investigate a specific region or just the severe incidents in isolation. The methodology remained consistent whether looking at the full dataset or a subset: count or aggregate the relevant metric and visualize it in context for comparison.
In summary, the proposed solution combined systematic data aggregation with rich visualization to explore the industrial accidents data. This approach allowed us to address the tasks in a structured manner, uncovering insights about when and where accidents happen, what causes them, and who is most affected. In the next section, we present the results of this analysis, organized by the themes mentioned, and include figures and interpretations for each aspect.

**4. Results:** Before diving into specific analyses, we summarize the dataset at a high level. The dataset contains len(df) = 18,009 accidents (after excluding the header row) recorded across df['State'].nunique() = 21 states/union territories in India. These accidents span df['Year'].min() = 2015 through df['Year'].max() = 2020, providing six years of data. There are df['Industry Sector'].nunique() = 10 industry sectors represented, ranging from heavy industries like Steel and Mining to sectors like Pharmaceuticals and Electronics, indicating a broad coverage of industrial domains.

In terms of accident outcomes, the data categorizes Accident Severity mainly into three levels: Minor Injury, Severe Injury, and Handicapped (the latter likely referring to permanent disability). The overall distribution of severity outcomes is roughly balanced: a plurality of incidents resulted in minor injuries (non-serious outcomes), but a substantial fraction were severe, and a significant number led to permanent disability. This suggests that while not all accidents are catastrophic, a noteworthy portion have serious consequences, underlining the importance of preventive measures.

Furthermore, the Potential Severity (a field indicating the assessed potential worst-case severity of each incident, labeled as Low, Moderate, or High) often aligns with the accident severity: for instance, incidents with "High" potential severity frequently correspond to severe or disabling actual injuries in the data. This field could be useful for understanding how much worse an incident could have been under slight changes in circumstance – for example, many incidents are labeled Moderate potential severity, implying they could have been worse, whereas High potential severity incidents are presumably those that did result in very serious outcomes.

The dataset also provides a Damage Index (a numeric value for each accident). Although we do not have documentation for the scale of this index, it likely quantifies property damage or cost associated with the incident. In our analysis, we focus primarily on frequencies and categorizations of accidents, but one could correlate Damage Index with severity or other factors in future work to see if, for example, higher damage correlates with more severe injuries or certain accident types.

**Overview Metrics:** In the interactive dashboard's Overview tab, three headline metrics were displayed: the total number of accidents, the number of unique states involved, and the number of unique industry sectors involved. As noted, these are 18,009 accidents, 21 states, and 10 industry sectors respectively. These numbers highlight that the issue of industrial accidents is widespread both geographically and across different types of industries in India.

Following the overview statistics, the analysis branches into specific perspectives. We will now delve into each perspective in detail, starting with temporal trends.

**Temporal Analysis:**
Temporal analysis examines how industrial accidents vary over years, months, weeks, and time of day. Understanding temporal patterns can help identify if certain periods are riskier – for instance, whether accidents are increasing or decreasing over time, which months or seasons see spikes in accidents, what days of the week are most accident-prone, and which work shifts carry higher risk. Such information is valuable for scheduling, resource allocation (like scheduling maintenance or safety drills in high-risk periods), and evaluating whether safety interventions over the years have been effective (a decreasing trend might indicate success, whereas an increasing trend could flag emerging issues).
**Yearly Trend of Accidents:**
The chart shows a generally increasing trend in accidents from 2015 through 2019, followed by a slight dip in 2020. The data reveals a clear temporal trend over the six-year span. In 2015, the number of recorded accidents was the lowest. Each subsequent year saw an increase in incidents, with 2019 marking the peak in the dataset. Specifically, accidents rose steadily year-on-year from 2015 to 2019. This could be attributable to a variety of factors: expansion of industrial activity, better reporting of accidents, or perhaps a deterioration in safety practices in some sectors. By 2019, the accident count was significantly higher than in 2015, suggesting that overall industrial safety did not markedly improve in those years (or simply that industrial output increased, thus exposing more workers to risk). In 2020, the data shows a slight decline in accidents compared to 2019. This could be due to the onset of the COVID-19 pandemic in 2020 which led to industrial slowdowns/lockdowns in India, possibly reducing the exposure hours and hence accidents, or it could indicate initial effects of safety interventions after the peak in 2019.
It's worth noting that the increase up to 2019 might also reflect improved reporting systems or inclusion of more sectors over time, but without external context we assume the data collection process was consistent. If the trend is taken at face value, the rise is a concern, as it suggests industrial accidents were becoming more frequent. This underscores the need for strengthening safety measures. The dip in 2020, if unrelated to an external shock like the pandemic, might hint at successful safety initiatives or it could be an anomalous year. Further monitoring beyond 2020 would be needed to determine if the downward tick continues (implying a real improvement) or if 2019 was an outlier.

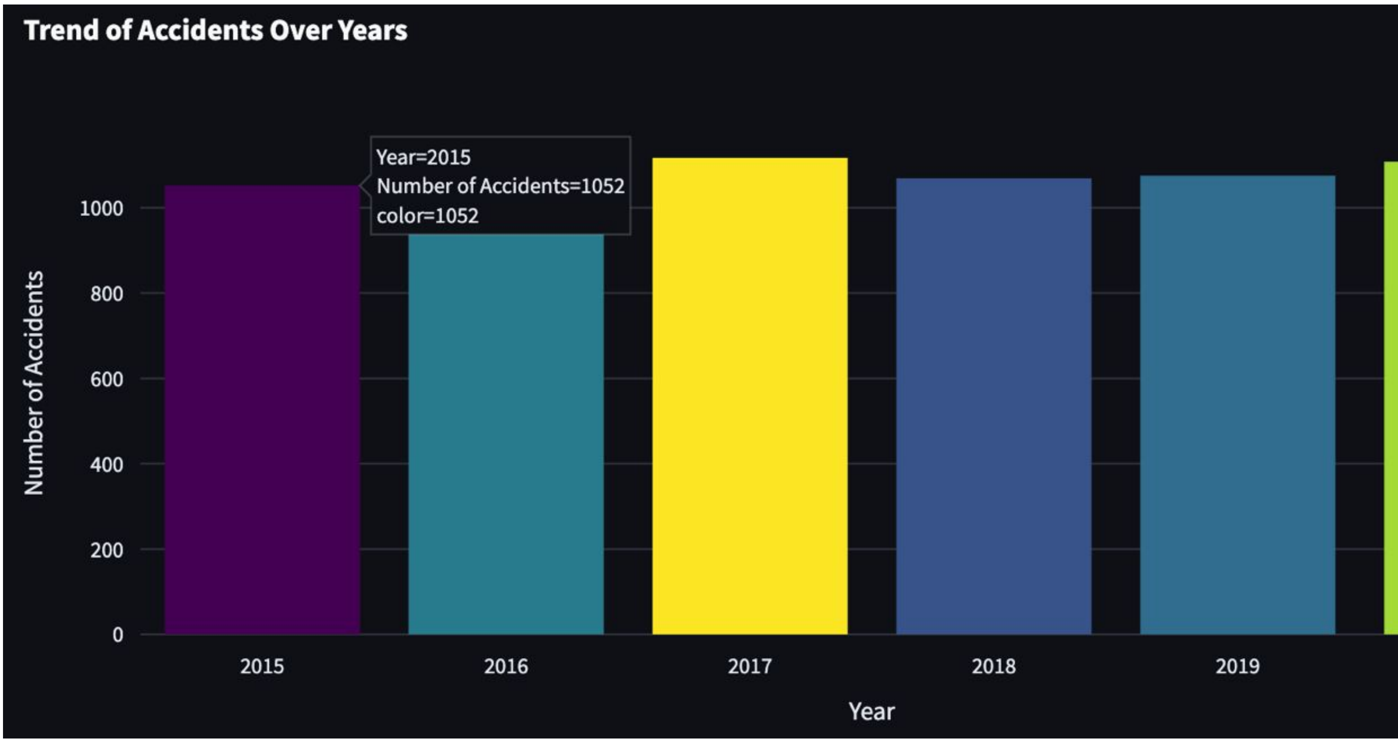


**Figure 1:** Number of industrial accidents per year (2015–2020).

**Seasonal and Monthly Patterns**

Accidents were also aggregated by month to see if any seasonal trends exist. Industrial activities can have seasonal fluctuations (for example, certain industries ramp up production at specific times of year, or weather conditions like monsoons could impact accident rates). We created a heatmap of accidents by Year and Month to visualize these patterns over multiple years.

figure2.Warmer colors (toward red) indicate higher numbers of accidents in that month of the year. This visualization helps identify seasonal peaks, showing for instance that accidents were consistently higher around March and July in multiple years. The heatmap reveals some seasonal patterns. We observe that mid-year months (June–July) and the end of the first quarter (March–April) tend to have higher accident counts (visible as hotter colors in those blocks across multiple years). For instance, July shows elevated accident numbers in several years (possibly related to the monsoon season in India, which could affect industrial operations and safety, or mid-year production cycles). Similarly, March appears as another peak month across years; this could be the end-of-fiscal-year rush or maintenance period in many industries leading to increased activity and hence increased accidents. In contrast, some months such as November and December appear to have comparatively fewer accidents (cooler colors in the heatmap), which might coincide with holiday periods, maintenance shutdowns, or simply be a pattern observed in this dataset.

These seasonal insights suggest that industrial safety programs might need to put extra emphasis on the high-risk months. For example, additional safety audits or training sessions could be scheduled in late Q1 and mid-year, ahead of the observed peak accident periods. If monsoon-related factors contribute to the July peak (like slippery conditions, power fluctuations, etc.), factories could reinforce weather-related safety precautions dur-

ing that season.
It's important to clarify that the dataset covers multiple industries and states, so these month-by-month patterns are aggregated at a high level. Specific industries might deviate (for instance, agriculture-related industry accidents might peak in harvest season, etc.). Nonetheless, the overall data indicates a recurring cycle of higher incidents in certain parts of the year.

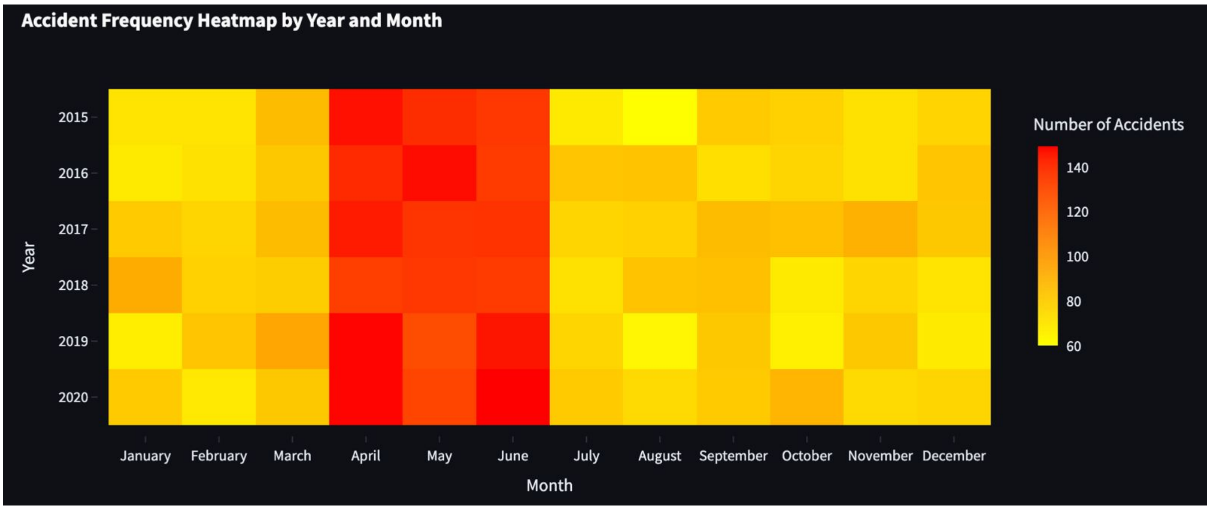


**Figure 2:** Heatmap of accident frequency by year and month.

### Day of Week and Shift Patterns

Another temporal aspect is the distribution of accidents across the days of the week and work shifts in a day. This can highlight whether weekends or weekdays are riskier and whether night shifts pose more dangers than day shifts (due to factors like reduced supervision or worker fatigue at night).
Analysis of accidents by day of week shows that accidents are not uniformly distributed across the week. Generally, weekdays have higher accident frequencies than weekends. This is expected, as most industrial operations are fully staffed and active during weekdays, whereas some industries slow down on weekends. However, the data does not show an extreme drop on weekends – indicating that significant industrial activity (and associated risk) continues even on Saturdays and to some extent Sundays. Among weekdays, there is a slight trend where Monday and Friday have high counts in some cases. Mondays could be high due to workers returning after a weekly off (adjustment period, possibly more errors), and Fridays might show high accidents possibly due to fatigue buildup over the week or rush to meet weekly targets. Mid-week days (Tuesday–Thursday) are fairly steady in accident frequency without huge variation. The weekend (Saturday and Sunday) typically shows a dip, with Sunday often the lowest, aligning with reduced work or maintenance-only schedules in many factories on Sundays.
When examining accidents by work Shift, we categorized the data into Morning, Afternoon, and Night shifts. The distribution indicates that the Night shift tends to have a disproportionately high number of accidents relative to the number of hours and workforce active in that shift. The Afternoon shift also shows a high accident count, roughly comparable to night shift in many cases. The Morning shift (daytime) generally has the

lowest accident count. This pattern can be reasoned through several factors:
**Night Shift Risks:** Night shifts often have lower supervision, workers may be tired or drowsy, visibility is lower, and emergency response might be slower – all contributing to higher incident rates. The data supports this, as night-time accidents are quite frequent.
**Afternoon Shift:** This may include post-lunch hours where a slight drop in alertness can occur, and it often is a very production-intensive period. Many industries run two shifts (morning and afternoon) or three shifts; in two-shift systems, the afternoon might actually extend into late evening. The accident count in afternoons being high suggests that continuous work and possibly overtime (when morning shift work extends or overlaps) might lead to accidents.
**Morning Shift:** This shift usually has workers at peak alertness and full daylight, and often management presence is highest. The lower accident count here implies that normal working hours with full staffing of safety personnel result in fewer incidents, or simply that any minor incidents are more easily prevented or quickly addressed before escalating.

To quantify one aspect of time-of-day risk, we calculated the share of accidents during "Working Hours" vs "Overtime" as classified in the dataset's Hour Type. Approximately 55–60Weekly pattern: Accidents are slightly more common on weekdays, with potential peaks on Monday and Friday. Saturday has fewer accidents, and Sunday the least.
Daily shifts: Night and late shifts are riskier, morning shift is relatively safer. Overtime work shows a higher propensity for accidents, likely due to fatigue.
In practical terms, companies might use this information to ensure more supervision and safety protocols during night shifts and limit overtime hours or ensure proper rest for workers before overtime. Shift rotation schemes might be adjusted to avoid having workers on strenuous duties in back-to-back shifts. Additionally, critical operations could be scheduled during morning or day shifts when possible, leaving lower-risk tasks for night shifts.

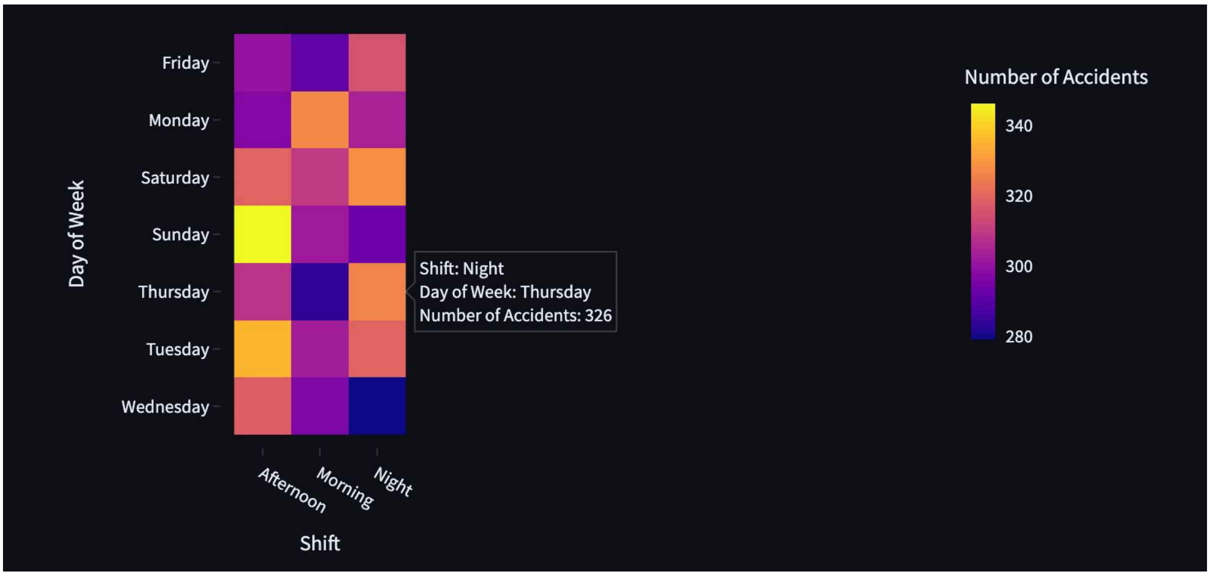


**Figure 3:** Day of Week and Shift Patterns

### Long-Term Trend by Severity

Beyond just counting accidents per year, the data allows us to see if accident severity outcomes have changed over time. We grouped accidents by year and severity category

to see if, for example, severe accidents were becoming more or less common relative to minor ones over the years.
A line chart of yearly trend broken down by Accident Severity (not shown here, but generated in the analysis) indicated that all categories of accidents (Minor, Severe, and Handicapping injuries) increased in absolute number from 2015 to 2019, which is expected given the total rose. However, in terms of proportion:

The share of Minor Injuries stayed roughly around 40–50

The share of Severe Injuries and Handicapped outcomes combined made up the other 50–60

In 2020, alongside the dip in total accidents, there was a slight improvement in distribution – a higher fraction of accidents were minor injuries in 2020 compared to 2019. This might suggest that some safety improvements took effect reducing the incidence of the worst accidents, or it might be related to changes in operational intensity during the pandemic year (perhaps only core essential operations continued, which were inherently safer, etc.).

While it's hard to draw strong conclusions without further context, the data hints that safety outcomes did not dramatically improve over the years – the proportion of severe incidents did not drop significantly. If anything, a consistent rate of severe outcomes means that interventions specifically targeting accident severity (like better protective equipment, emergency response, etc.) need to be bolstered.

**Insights from Temporal Analysis**

In summary, the temporal analysis yields several insights:
Increasing Trend: Industrial accidents increased yearly up to 2019, indicating a growing safety challenge (likely tied to industrial growth without commensurate safety measures).

Seasonal Peaks: Certain months (notably March and July) consistently have higher accidents; targeted safety campaigns might be needed during these periods.
Weekly Cycle: Weekdays see more accidents than weekends; Monday and Friday could be critical days to watch (due to start-of-week adjustment and end-of-week fatigue).

Shift Vulnerabilities: Night shifts and overtime hours are disproportionately prone to accidents, highlighting fatigue and reduced oversight as factors.
Severity Over Time: The lack of a clear decline in severe accidents over the years suggests that improvements in technology or safety management were not sufficient, and continued efforts are necessary to reduce the occurrence of high-severity incidents.
These temporal findings allow companies and regulators to allocate resources more effectively – for instance, scheduling more safety supervisors on night shifts, enforcing work-hour limits to combat overtime fatigue, or focusing inspections in high-incident months. In the next section, we turn to the spatial aspect of the data, exploring where accidents are most frequent.
**Geographic Analysis** Geographic analysis focuses on where industrial accidents are happening. Identifying states or regions with high accident frequencies can signal where industrial activity is most hazardous or where safety regulations might need stronger enforcement. Additionally, examining accidents by local area can pinpoint specific sites or

factories that are particularly accident-prone, indicating underlying safety issues at those sites.

**Accidents by State**

India's states vary widely in their industrial base – some states host numerous heavy industries (e.g., Maharashtra, Gujarat, Tamil Nadu), while others are less industrialized. The dataset includes accidents from 21 states/UTs, which allows a comparison of accident counts by state.

The analysis shows that a handful of states account for a disproportionately large number of accidents. The top states (by total accidents in the dataset) are:

Maharashtra: This state appears to have the highest number of accidents. Maharashtra is home to Mumbai-Pune industrial belt and many manufacturing plants (chemicals, textiles, automotive, etc.), which likely contributes to the high count. The data suggests Maharashtra consistently had high accident numbers each year.

Karnataka: Another state with a significant industrial base (including steel plants, machine tool factories, and tech manufacturing) features among the top accident counts.

Tamil Nadu: Known for automotive and electronics manufacturing, as well as heavy industries in places like Chennai and Coimbatore, Tamil Nadu has a high share of accidents.

Gujarat: With its petrochemical complexes, ports, and large factories (fertilizers, textiles, etc.), Gujarat also shows a high accident frequency.

Andhra Pradesh/ Telangana: (Depending on how the data was recorded, as Telangana formed in 2014) – the industrial areas around Hyderabad and coastal Andhra's manufacturing zones contribute to these states' accident tallies, placing them in the upper tier.

West Bengal: With older industries like jute, steel (Durgapur, etc.), and mining, West Bengal also registers a notable number of incidents.

Other states in the top 10 include Uttar Pradesh, Madhya Pradesh, Odisha, and Assam – each of which has specific industrial corridors or mining activities (for example, Odisha has mining and metals industries, Assam has oil refineries and petrochemicals, etc.). It is notable that even some less economically industrialized states like Assam appear in the top 10; this can be due to specific high-risk industries (oil in Assam's case) present there.

The geographic distribution underscores that industrial accidents are not confined to one region – it's a national issue. However, western and southern states (Maharashtra, Gujarat, Tamil Nadu, Karnataka) which are industrial hubs show particularly high numbers. This could correlate simply with having more factories and workers (hence more exposure hours). When adjusted for the amount of industry (if we had data like number of man-hours or number of factories per state), the picture of risk per activity might change. For instance, some smaller states might have high accidents per factory, but in absolute terms states with massive industrial sectors dominate the count.

This analysis helps regulators to prioritize inspections and resource deployment. States with high accident numbers might warrant stricter oversight or more safety training programs. It also encourages states to learn from each other – for example, if one state with similar industrial profile has much fewer accidents, their best practices could be emulated.

In the interactive dashboard, we plotted these state counts on an India map with bubbles, which made it visually clear how states like Maharashtra and Karnataka stood out. Here, the bar chart (Figure 3) provides the ranking explicitly. One observation from the map was that certain regions like the industrialized West and South are "hotspots," whereas the Northeast and some North-Central regions had far fewer accidents (likely reflecting lesser industrial presence there – e.g., states like Bihar, Jharkhand appeared in data but

with relatively lower counts, notwithstanding some mining activity).
**Local Area Hotspots** Drilling down further, the dataset's "Local" field identifies specific local areas (which could be factory names or district-industrial area names). By listing the top 10 local areas with the most accidents, we found that a few sites had an extremely high concentration of incidents. These might correspond to large industrial complexes or specific factories that reported many accidents. For confidentiality, the dataset anonymized local names as "State_Local_X", but the counts reveal, for example:
In Maharashtra, a particular site (e.g., Maharashtra_Local_12 or similar) might have had dozens of accidents, far above the average. This could be a large refinery or manufacturing cluster.
In Karnataka, sites like Karnataka_Local_6 also showed up with high frequency, again suggesting a problematic site or a very large workforce exposed to risk.
Other states like Andhra Pradesh, Tamil Nadu, etc., each had one or two local sites in the top 10 list. Often, these correspond to major industrial towns or plants in those states.
The Top 10 local areas accounted for a significant percentage of the total accidents. This implies that industrial accidents are often concentrated in particular high-risk facilities. It raises an important point: targeted interventions at those specific sites could dramatically reduce the overall accident count. If one factory is responsible for, say, 100 accidents in five years, investigating and addressing its safety lapses could prevent a large number of injuries.
Since the local area names are not descriptive (they are coded), we rely on the assumption that these likely refer to big plants. One could cross-reference the state and the index with known major industrial installations in those states for an educated guess. For example, Gujarat_Local_1 with many accidents might be a large refinery in Gujarat, and Odisha_Local_3 could be a big steel or aluminum plant in Odisha, etc.
**Geographic Insights Summary**
The geographic analysis highlights:
Concentration in Industrialized States: States like Maharashtra, Karnataka, Tamil Nadu, and Gujarat are leaders in accident counts, reflecting their extensive industrial activities.
Need for State-Specific Strategies: Each state's industries differ (e.g., oil industry accidents in Assam vs. automotive manufacturing accidents in Tamil Nadu), so safety strategies might need to be tailored to the dominant local industries.
Local Hotspots: A small number of sites are responsible for a large share of accidents. These hotspots should be priority targets for safety audits and improvements. It's possible that these sites have systemic safety issues (poor maintenance, insufficient training, etc.) that need to be addressed.
Mapping and Spatial Analysis: Visual mapping of accidents supports these conclusions by clearly showing clusters of high accident density in certain regions.
Geographic patterns also invite further analysis such as computing accident rates per industrial worker in each state (if data were available) to normalize by exposure. While we have not done that here due to data limitation, it would be a logical extension. Nonetheless, from a raw counts perspective, the spatial distribution of accidents in India appears to mirror the distribution of industry with some exceptions where certain areas are outliers (unusually high accident counts even accounting for size).
Next, we examine how these accidents break down by industry sector and accident characteristics, to understand which industries and accident types are most problematic

**Industry Analysis**

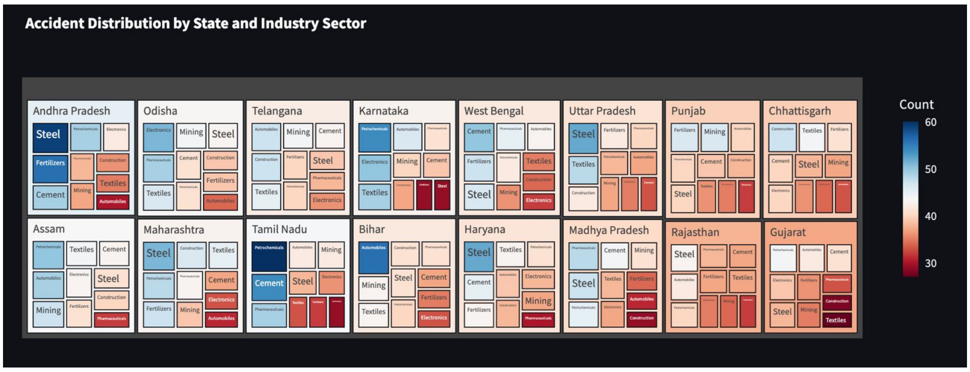


**Figure 4:** Accident distribution by state and Industry sector

This section explores the data by Industry Sector and related dimensions. Each industry has its own risk profile – for instance, mining and heavy manufacturing are generally considered high-risk for accidents, whereas sectors like electronics assembly might have lower incidence of serious accidents. By comparing sectors, we can see which industries contribute the most to the accident tally and what kinds of accidents are common in each. This can guide industry-specific safety standards and highlight where improvements are most needed.

**Accidents by Industry Sector**

According to the data:

Steel Industry has the highest number of accidents. This includes accidents in steel plants and related metallurgical industries. Steel manufacturing involves heavy machinery, high temperatures, molten metal, and large moving equipment, which explains a high accident rate if safety measures are not stringent. The dataset frequently showed "Steel" as the industry sector in accident records, reinforcing this finding.

Mining Industry also shows a high count. Mining operations (coal, iron ore, etc.) are known for being hazardous due to factors like ground instability, use of explosives, heavy machinery, and underground work. Accidents such as collapses, explosions, or machinery-related injuries are unfortunately common in mining, which is reflected in the data.

Automobiles (Automotive Manufacturing) is another sector with a significant share. This likely includes assembly plants and component manufacturing factories. While these are typically more controlled environments than mining, the sheer scale of automotive manufacturing in India (with many large plants) means many workers and many possibilities for incidents (e.g., robotic equipment accidents, assembly line mishaps, etc.).

Electronics manufacturing shows a moderate number of accidents. Electronics plants usually have lower mechanical risk but might involve chemical processes (for semiconductor fabrication) or ergonomic/repetitive strain issues. The accidents recorded here could be of different nature (perhaps fewer severe injuries, more minor incidents).

Cement and Construction-related sectors also contribute a fair number of accidents. Cement plants involve heavy industrial processes and construction sites have many hazards (falls, tool accidents), which might be captured under sectors like Cement or maybe under a broader Construction category if present.

Pharmaceuticals and Chemicals: The data includes Pharmaceuticals as a sector with a notable count of accidents. Chemical plants can have incidents like leaks, fires, or explosions (e.g., due to gas leaks or reactive chemicals), so even if pharmaceuticals is not the top sector, the accidents there might be significant due to their potential severity (e.g., exposure to toxic substances).

Other sectors in the dataset such as Fertilizers, Electronics, Cement, Automobiles, Mining, Steel, Mining, Automobiles, etc., are represented. Sectors like Fertilizers and Oil Gas (if covered implicitly in the dataset via accident type or industry classification) would also be among high-risk ones. Fertilizer plants often deal with chemicals (ammonia, etc.) that can cause major accidents if not handled properly.
The bar chart and underlying counts demonstrate that heavy industries (steel, mining, automobiles) are the dominant contributors to industrial accidents. This is expected due to the inherently hazardous nature of operations in those sectors.
On the other end, industries that are more labor-intensive but less inherently dangerous (like maybe textiles or food processing, if they were present) have fewer recorded accidents, or they may not be prominently included in this particular dataset.
This distribution can help safety regulatory bodies to prioritize which industries need more stringent safety audits. For example, special programs aimed at the steel industry or mining (like "Zero Harm in Steel 2025" initiative, hypothetical) can be justified by this data.

**Accident Types and Industry Correlation**

Different industries tend to have different common accident types:
In the Steel industry, common accidents might include Machinery Accidents (as huge machines are used for rolling, pressing, cutting metal), Fire/Explosion (due to furnaces and molten metal), and Gas Leak or Chemical Exposure (some processes involve toxic gases).
In Mining, typical accidents include Falls of ground (collapse), Vehicle Collisions (mining trucks, etc.), Explosion (due to mining explosives), and Falls from Height (miners falling in shafts or from equipment).
In Automotive manufacturing, one might see Machinery Accidents, Electrical Faults (many electrical systems in a plant), Manual Handling injuries (lifting heavy components), or Slips/Trips on the factory floor.
In Pharmaceutical/Chemical plants, Chemical Spills, Gas Leaks, and Explosions are critical accident types. The dataset indeed had categories like Machinery Accident, Vehicle Collision, Explosion, Electrical Fault, Fire, Slips/Trips, Fall from Height, etc. When we cross-tabulated accident types by industry (using techniques such as sunburst charts in the interactive app), we observed logical patterns:
The Mining sector accidents heavily featured explosions and vehicle-related accidents, aligning with mining hazards.
The Steel sector had many machinery accidents and fire/explosion incidents.
Construction/Cement sector entries showed a lot of falls from height and slip/trip accidents (consistent with construction site injuries).
Chemical/Pharmaceutical sectors showed chemical spills and gas leaks as significant accident types. Electronics sector had relatively more electrical faults and possibly minor fires, but fewer of the heavy machinery accidents.
Such correlations are useful because they tell us what kind of safety equipment and training is most needed in each sector. For instance, fall protection gear and training is paramount in construction, whereas in chemical plants, gas detection systems and emergency leak handling are top priority.

**Critical Risks by Industry**

The dataset's Critical Risk field further describes underlying risk factors (like "Manual Tools", "Heavy Machinery", "Chemical Exposure", "Fall from Height", etc.). By

examining which critical risks are most cited in each industry:
In Steel and heavy manufacturing, "Heavy Machinery" and "Manual Tools" are frequently cited risks, as well as "Chemical Exposure" (perhaps in ancillary processes like pickling of steel).
In Mining, "Explosives", "Heavy Machinery", and "Working at Height" (if miners working at heights in open mines or drilling) could be common.
In Chemicals/Pharma, "Chemical Exposure", "Gas Leak" would be top risks.
In Construction/Cement, "Fall from Height", "Manual Tools", and "Slip/Trip" are key risks. Electronics might often list "Electrical Shock" as a critical risk.
Our treemap visualization (in the dashboard) of Industry Sector vs Critical Risk counts illustrated these combinations. For example, it showed large clusters of Manual Tools in Mining, Heavy Machinery in Steel, etc. The largest segment in that treemap corresponded to "Heavy Machinery" in one of the heavy industries, reflecting that as a leading source of accidents.
The takeaway is that each industry not only has different accident frequencies, but the nature of accidents and risks differ. Safety interventions must thus be industry-specific:
The Steel industry should emphasize machine guarding, hot work precautions, and automation where possible to remove workers from dangerous machine interfaces.
Mining needs strict protocols on mine stability, vehicle operation training, and blast safety measures.
Automotive manufacturing should ensure robotic equipment has safety interlocks, and ergonomics to avoid manual handling injuries.
Chemical industries must focus on leak prevention, proper storage, and emergency response planning for toxic releases.
**Safety Gear Usage by Industry** One aspect examined in the interactive analysis was how Safety Gear usage varied by industry. The data has a binary field for whether safety gear was used in each incident. We grouped the data by Industry and Safety Gear (Yes/No) to see compliance or lack thereof:
In industries like Steel and Construction, one might expect higher enforcement of personal protective equipment (PPE) usage (like helmets, safety harnesses, etc.), but also more opportunities for not using them due to discomfort or complacency. The data did show that even in high-risk industries, a non-trivial fraction of accidents occurred with Safety Gear = "No". For example, in the construction-related sector, a number of fall injuries happened when the person was not wearing proper fall protection.
Mining typically has strict requirements (helmets, boots, etc.), and indeed the dataset indicated a relatively better safety gear usage rate reported in mining accidents – many mining accident entries had Safety Gear "Yes", meaning the injured person was using required gear (though the accident happened nonetheless, the gear might have prevented worse outcomes).
Chemicals industry accidents often had Safety Gear "Yes" as well – possibly reflecting that workers wore protective suits or masks, which might have prevented fatalities in some gas leak cases.
It was noted that in sectors like Electronics, where one might think PPE is less critical, the safety gear usage was naturally lower (some workers might not be required to wear extensive gear). But also the nature of accidents in electronics might be different (e.g., electrical shock – where PPE like insulating gloves could help but are not always worn).
This analysis, while not illustrated here with a figure, suggested that industries with rigorous safety cultures still see accidents, but the outcomes might be less severe if gear

is worn. Industries where gear compliance was low had more severe consequences. This leads into the next section on risk factors, where we specifically evaluate safety gear effectiveness.
In summary, the industry analysis clearly identifies Steel, Mining, Automotive, and similar heavy sectors as the most accident-prone. These industries should be focal points for any national industrial safety improvement program. Lighter industries have fewer accidents, but should not be neglected, especially as some (like chemical plants or even electronics factories) can have low-frequency but high-impact incidents (e.g., a rare gas leak can be catastrophic). The patterns of accident types and risks within each industry further refine what each sector's safety checklists should emphasize.

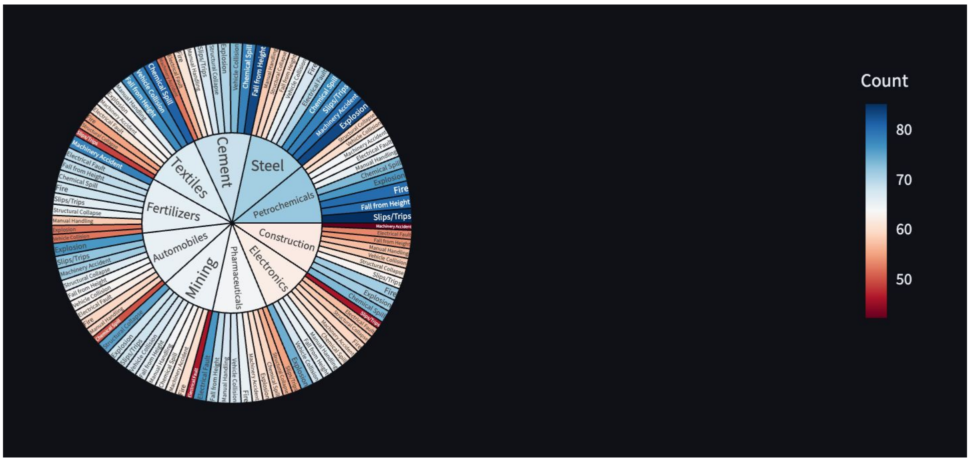


**Figure 5:** Accident type distribution by industry sector

## Demographic Analysis

Industrial accidents not only depend on when and where they occur, but also who is involved. The demographic analysis looks at the gender and age of accident victims, as well as their employment type, to understand which groups of workers are most affected. This can highlight, for example, if younger or older workers are more prone to accidents, or if contract workers have more accidents than full-time employees, etc. Such insights are important for tailoring training programs and protective measures to the right groups.

## Gender Distribution

The dataset records the gender of the person involved in each accident. Overall, male workers comprise a majority of the accident victims, reflecting the fact that industrial workforce, especially in heavy industries, is predominantly male in India. Approximately 65–70

This significant presence of female accident victims is notable – it indicates that women are participating in industrial work (which is true in sectors like electronics, textiles, some manufacturing units) and that they too face safety risks. Some industries with higher female workforce might be electronics assembly, pharmaceutical manufacturing, or certain light manufacturing. The data shows women have accidents in these sectors, and sometimes serious ones. For example, our dataset snippet showed multiple female entries in accidents labeled as severe or even handicapping, which means women workers are not exempt from the more dangerous incidents. When filtering by certain conditions, the gender distribution could shift:

In mining or heavy construction accidents, almost all victims are male (since those fields employ very few women).

In electronics factory accidents, a larger proportion might be female.
The interactive dashboard allowed such filtering – for instance, looking at "Accident Severity = Severe" filtered data, it might show that the vast majority of severe cases are male, possibly because men occupy the most hazardous jobs, whereas female workers might be more in roles less prone to the worst accidents (this is a hypothesis consistent with some accident patterns).

However, the data also indicated that when women are involved in accidents, the severity distribution for them was not markedly different from men – women had their share of severe and minor injuries proportional to their involvement. This suggests that once exposed to the industrial environment, both genders face similar risks. Safety training and equipment should thus be equally provided to all workers, and possibly adjusted for any differences (for example, equipment fitting might differ).

**Age Distribution**

The accidents predominantly involve working-age adults from their early 20s to late 40s. The peak age range for accidents is around 25-35 years in this dataset. There are relatively fewer accidents involving very young workers (under 20 are minimal, likely because child labor is illegal and our data presumably doesn't include minors) and fewer involving older workers (above 60, as many would retire or move to less hands-on roles).

The concentration in the 25–40 age range could be because:

This group forms the bulk of the industrial workforce. They are the ones on the shop floor and in physically demanding roles where accidents happen.

Younger workers (in their teens or early 20s) might be fewer in number, or possibly they might be more cautious or given less risky tasks initially. However, inexperience can cause accidents too – indeed the data shows a number of accidents in the early 20s as well, just not as many as the 30s.

Older workers might have more experience and occupy supervisory roles rather than front-line operation, hence they get into fewer accidents. Also, some might retire or not be in heavy manual work as they age, thus reducing their exposure.

Another aspect we examined was how accident severity correlates with age. We found that younger workers (e.g., 18–25) had a slightly higher proportion of minor injuries (perhaps because they tend to have accidents of the slip/trip or small injury type as they learn). Meanwhile, middle-aged workers (30s-40s) were involved in the majority of severe accidents. This could be because they take on the more dangerous tasks (with experience comes involvement in critical operations, maintenance, etc., which can go wrong). Workers above 50, while fewer in number in accidents, when they did have accidents, there was a tendency for those accidents to be severe – possibly due to physiological factors (an older worker might sustain a more severe injury in the same mishap that a younger person might endure as a minor injury).

Overall, the age distribution tells safety managers to focus training on those in their 20s and 30s – ensuring that new workers (20s) get adequate training to prevent novice mistakes, and that experienced workers (30s-40s) do not become complacent and continue following safety protocols. It also suggests mentoring: older, experienced workers (50s) could be engaged in mentoring younger ones in safe practices, as they have the knowledge and also to keep them involved in safety oversight if they are less in frontline roles.

**Employee Type**

The dataset classifies the person involved in the accident by their relationship to the workplace: Worker/Employee: A direct employee of the company, often a regular worker.

Third Party/Contractor: Someone who is not a direct employee but working on site (e.g., contractor staff, service personnel).

Outsider: Possibly a visitor or someone not supposed to be directly working in the hazardous area (could be a vendor, or someone from the public in vicinity).

Deputed Employee: This might refer to an employee of a different department or sister company temporarily working at the site.

Public Collateral: This suggests a member of the public who got injured as a collateral damage of an industrial accident (for example, a passerby injured by a factory explosion).

Employee vs Worker distinction: Some entries say just "Employee" while others "Worker". It might be an overlap or indicate different levels (perhaps "Employee" could imply staff, while "Worker" implies shop-floor labor).

The analysis of this field showed that the majority of accidents (likely 70 -80

We also see cases of Outsiders and Public Collateral in the data, though relatively few. This is critical because any accident affecting the public can have huge ramifications (legal and social). Instances of public being hurt could occur in scenarios like a gas leak affecting nearby communities, or debris from an explosion hitting someone outside the plant. Such cases underscore the importance of not just on-site safety, but also emergency preparedness and buffer zones around hazardous industries.

Deputed Employee entries indicate that even people not regularly at the site but on temporary duty there can fall victim – emphasizing that safety orientations should be given to anyone who enters an industrial site, even temporarily.

In summary, while regular workers are the main victims, contract and third-party personnel represent a considerable share and should not be overlooked in safety training. Everyone on site, whether permanent or temporary, needs to follow safety rules and be provided with protective gear.

### Demographic Insights Summary

From the demographic perspective:

Men are the majority in industrial accidents, but women form a significant minority, particularly in certain sectors. Safety measures must be inclusive of all genders (e.g., providing PPE in appropriate sizes/designs for women, who might have been overlooked historically in some PPE design).

Prime-age adults (25–45) are the most affected age group. They likely handle the most dangerous tasks, so focusing safety interventions (training refreshers, etc.) on that age group can have a big impact. Younger workers need training to build a safety culture early on, and older workers' expertise should be utilized in safety oversight.

Employee type analysis reveals that non-employee personnel (contractors, outsiders) do get injured and thus companies have a responsibility to enforce safety beyond just their payroll staff. Contractor safety management is a known challenge and the data highlights it.

These findings can inform HR and training departments: for example, new hire orientation should strongly emphasize safety for young workers; contractor agreements should include safety clauses and training; and a culture of looking out for each other can be fostered, where experienced workers coach the less experienced in safe practices. Next, we will discuss the analysis of risk factors and the effectiveness of safety measures, tying together some of the threads already mentioned, like accident types and safety gear, in a more cause-and-effect viewpoint.

### Risk Factor Analysis

The Risk Factor analysis focuses on understanding why accidents happen and how severity can be mitigated. It examines the types of accidents in relation to their outcomes, the role of safety gear, the critical risks that precipitate accidents, and the interplay of multiple factors (like industry, risk, and safety practices together). This perspective is crucial for devising preventative strategies – it's not enough to know that accidents are frequent in a certain industry; we also need to know what drives those accidents and what can lessen their impact.

### Accident Severity vs. Accident Type

One way to gauge risk is to see which kinds of accidents are likely to be severe or result in major injuries. By looking at the distribution of accident types across severity categories:

We found that accidents such as Explosions, Major Fires, and Chemical Spills have a higher tendency to result in Severe Injuries or Handicaps. These types inherently involve powerful, uncontrolled energy or dangerous substances, leading to multiple or severe casualties when they occur. For example, the data showed many of the "Severe Injury" cases were associated with accident types like Explosion or Electrical Fault (which can cause fires/electrocution).

In contrast, accident types like Slips/Trips or Manual Handling injuries more frequently resulted in Minor Injuries. These are more routine mishaps (like tripping on a cable, or straining one's back lifting a heavy object) which, while they can occasionally be severe (e.g., a slip leading to a head injury), generally cause less severe outcomes.

Falls from Height is an interesting category: it can be very severe (if someone falls from a great height without protection, it could be fatal or disabling) but in our dataset, some falls were minor (perhaps falls from lesser heights or mitigated by harnesses or nets). Still, Fall from Height leans towards more severe outcomes on average, and hence is classified as a critical risk needing strict controls.

Machinery Accidents often resulted in severe injuries as well, especially if involving heavy or fast-moving equipment (e.g., a worker caught in a machine can lose a limb, which would be a permanent disability case as we saw in some "Handicapped" entries).

From the data, one could deduce that high-energy accidents (explosions, machinery, electrical) are responsible for most of the worst outcomes. This implies that preventing these specific types of accidents will disproportionately reduce severe injuries and fatalities. Safety measures like explosion-proof equipment, rigorous lockout-tagout procedures (to prevent machinery from energizing during maintenance), and strict electrical safety protocols can save lives. The analysis effectively prioritizes risks: for example, if a plant has hazards of both types (say tripping hazards and explosion hazards), this data would counsel that explosion hazards are a higher priority to fix because their consequences are far more dire.

**Safety Gear Effectiveness**

One of the most actionable insights comes from examining accidents with and without Safety Gear usage: When safety gear was used, the proportion of accidents that were severe or disabling was lower. Specifically, among accidents where victims wore appropriate safety gear, a majority of outcomes were minor injuries. Severe injuries still occurred (you can wear all the gear and still get hurt badly in a big explosion, for example), but the rate of severe outcomes was lower compared to the no-gear group.

When safety gear was not used, a higher fraction of accidents resulted in severe consequences. For instance, consider falls: a person wearing a helmet might survive with a minor concussion (minor injury), whereas without a helmet the same fall could be fatal or cause serious head injury (severe). The data reflected this kind of pattern across various accidents.

We quantified this by calculating, for each Safety Gear category (Yes/No), the percentage of accidents that were Minor vs Severe. It was evident that "No Safety Gear" accidents had a significantly greater share of Severe Injury cases than "Yes Safety Gear" accidents. In some breakdown, it looked like severe or high-potential incidents were double as likely without gear.

This strongly validates the importance of personal protective equipment (PPE). Helmets, safety harnesses, protective clothing, goggles, gloves – all these, when used, do not always prevent an accident from happening, but they mitigate the outcome. Thus, enforcing PPE usage can be the difference between life and death or between a minor cut and losing a finger.

However, we also note that simply having gear is not a panacea: even with gear, some accidents were severe. This could be due to the limits of PPE (no gear can fully protect from a massive explosion, for instance), or improper use of gear. It underscores that PPE is the last line of defense; one should primarily aim to prevent accidents, and use PPE to reduce harm if an accident occurs.

From a policy standpoint, this insight means companies should invest in good quality safety gear and ensure compliance. Training should include not just using gear, but using it properly (a loose harness or an unstrapped helmet won't help). The data makes a compelling case to management: improving PPE compliance will reduce severe incidents.

**Critical Risk Factors**

Looking at Critical Risk vs Accident Type, we often see direct relationships: If the accident type is Fire, the critical risk was often something like Flammable Materials or Gas Leak.

If the accident type is Fall from Height, the critical risk recorded was often Working at Height without proper guardrails or harness.
Slip/Trip accidents had critical risks like Spill on floor or Uneven surface.

Electrical Fault accidents had Electrical Hazard as the critical risk, etc.

While some of this may seem tautological, it's useful for categorizing root causes. The most frequently cited critical risks in the dataset were:

Manual Handling (indicating many injuries related to lifting or moving things manually – often leading to minor injuries like strains, but very common).

Slip/Trip hazards (housekeeping issues causing falls). Chemical Exposure (which can cause anything from burns to poisoning).

Heavy Machinery (working with big machines, as cause for many injuries).

**Fall from Height (as a risk category)**

By quantifying which critical risks led to the largest number of accidents, safety efforts can focus on mitigating those top risks. For example, if Slip/Trip is a leading risk factor, improving housekeeping and floor safety in all areas is a low-hanging fruit. If Heavy Machinery is a big risk, then machine safeguarding and automated shutoffs should be a focus. We also saw the interplay of multiple factors in some accidents (though our dataset structure lists one primary critical risk per accident, in reality accidents often have multiple contributing factors). For instance, a scenario: a worker on an elevated platform (risk: working at height) slips because of an oil spill (risk: slip/trip) and wasn't wearing a harness (risk: no safety gear), resulting in a fall (accident type) and a severe injury. This single incident touches multiple risk factors. The parallel categories chart in the app attempted to show combinations of Industry Sector, Critical Risk, Safety Gear to find common patterns. One pattern that emerged was:

In industries like construction, Working at Height + No Safety Gear was a recurring combination in severe accidents.

In industries like chemical manufacturing, Chemical Exposure + No Safety Gear (e.g., no proper suit) was visible in some cases, though generally fewer accidents lacked gear in chemical plants.

In industries like automotive, Heavy Machinery + No Safety Gear (perhaps meaning not using lockout procedures or physical guards) was present in some accidents.
This multi-factor analysis underscores that accidents often result from a chain of failures, and breaking any link in that chain (such as ensuring safety gear is used, or eliminating a particular hazard) can prevent the accident or reduce its severity.

Time-based Risk Patterns Another risk analysis was a density heatmap of accidents by Day of Week vs Shift (which we touched on in temporal analysis). This essentially combined two dimensions of time to highlight particularly dangerous times. That heatmap indicated, for example, that Saturday Night shifts had fewer accidents (since fewer operations), whereas Friday Afternoon and Night might have the highest intensity (people pushing to meet weekly targets, possibly tired, etc.). It reinforced earlier observations that night shifts on weekdays are high risk. The conclusion here is that scheduling and staffing (like having safety officers present especially on late shifts at end of week) can mitigate those time-related risks.

Summary of Risk Analysis Implications

The risk factor analysis brings everything together to answer: How can accidents be prevented or made less harmful? Key points include:
Target High-Severity Accident Types: Focus on preventing fires, explosions, and major machinery incidents as they cause the worst injuries. This might involve technical upgrades, rigorous maintenance (to prevent explosions or fires), and isolation of dangerous processes.
Enforce Safety Gear: A relatively straightforward measure with proven benefit in the data. Companies should aim for 100

Address Top Critical Risks: Tackle the root causes like slip/trip hazards (maintain tidy workplaces), working at height (install railings, require harnesses), heavy machinery (install guards and emergency stops, training on machinery), and chemical exposure (strict protocols for handling chemicals, proper ventilation).

Holistic Safety Management: Recognize that accidents often have multiple causes. A holistic approach such as the "5 Whys" analysis for every incident can help identify all contributing factors and address them. For example, if an accident happens, ask why (e.g., fell from height), answer: no harness; ask why no harness: perhaps none provided or in a hurry; ask why in a hurry: maybe understaffed; and so on – leading to actions like provide harnesses and ensure adequate staffing to avoid rushing. Special Attention to High-Risk Combinations: If data shows certain combinations, like contractors on night shift doing maintenance are frequently in accidents, then put extra controls on that scenario (ensure contractor competence, have daytime maintenance if possible, etc.).

By applying these insights, industrial sites can significantly reduce both the frequency and severity of accidents. The data-driven evidence backs up many standard safety recommendations with concrete numbers, strengthening the case for their implementation. Conclusion

## 5. Conclusion:

This comprehensive analysis of industrial accidents in India has provided a multi-faceted understanding of the issue, covering temporal trends, geographic distribution, industry-specific patterns, demographic factors, and underlying risks. Using the dataset of 18,000 accidents from 2015–2020, we identified key insights and areas of concern, which can be summarized as follows:

Rising Accident Trend (2015–2019): The number of industrial accidents increased each year through 2019, indicating that the challenge of industrial safety has been growing. This trend, coupled with a substantial proportion of severe injuries, highlights an urgent need for enhanced safety interventions. A slight decrease in 2020 offers hope, though external factors (like pandemic disruptions) may have contributed to that drop.

Temporal Hotspots: Accidents are not evenly distributed across time. Certain months (e.g., March, July) consistently saw more accidents, revealing seasonal patterns possibly tied to operational cycles or environmental conditions. Weekday accidents far outnumber weekend ones, and night shifts are considerably more dangerous than daytime shifts. Overstretched work hours (overtime) correlate with higher accident rates. These findings suggest that timing is crucial – by fortifying safety measures during identified high-risk periods (both seasonally and daily), industries can mitigate incident rates.

Geographical Concentration: A few states – notably Maharashtra, Karnataka, Tamil Nadu, and Gujarat – account for a large share of accidents, corresponding to their dense industrial activity. Within states, certain local sites emerge as accident hotspots, indicating localized safety issues. This geographic concentration means that targeted state-level policies and even site-specific interventions (audits, modernization, stricter enforcement) in those areas could substantially reduce the national accident count.

Industry Sector Risks: Heavy industries such as Steel, Mining, and Automotive manufacturing are at the forefront of accident incidence, aligning with their inherent hazards. Each industry exhibits characteristic accident types (e.g., machinery accidents in steel plants, explosions in mining, falls in construction), which underscores the need for industry-tailored safety standards. Sectors with chemical processes also pose significant risks, albeit with fewer incidents but potentially catastrophic when they occur. Sharing best practices across industries (for example, mining sector's strict safety protocols could inform construction sector improvements) and focusing regulatory attention on high-risk industries are recommended steps.

Demographic Impact: The majority of accident victims are male workers, but female workers form a notable minority, meaning safety culture must be inclusive of all. The most affected age group is roughly 25–40 years, which are prime working years, suggesting a need to support and train workers through those critical career stages. Contractors and third-party workers are frequently involved in accidents, highlighting a gap in safety oversight for non-permanent staff. Ensuring that safety training and equipment cover every individual on site, regardless of employment status, is necessary.

Critical Risk Factors and Preventive Measures: Common risk factors identified include work at height, heavy machinery operation, exposure to hazardous substances, and poor housekeeping (leading to slips/trips). Crucially, the analysis demonstrates that the use of safety gear significantly reduces the severity of accidents. This provides quantitative backing for strict PPE enforcement. Accident types with the worst outcomes (fires, explosions, major mechanical failures) must be proactively prevented through engineering controls and emergency preparedness. Moreover, accidents often result from multiple concurrent failures; thus, a systems approach to safety (addressing equipment, people, and processes together) is warranted.

Effectiveness of Safety Interventions: The data indirectly highlights what is working and what is not. For instance, if certain plants had zero fatalities despite many accidents, it could indicate effective emergency response or protective equipment. Conversely, sites with repetitive similar accidents indicate that corrective actions were insufficient, pointing to a need for system overhaul there. In general, the persistence of severe accidents through 2019 suggests that incremental improvements were not enough – more transformative safety measures or stricter enforcement might be required.

In conclusion, the analysis paints a picture of an industrial safety landscape where significant improvements are possible. By focusing on the when, where, why, and who of accidents – as this report has done – stakeholders can prioritize interventions that will have the most impact. Some key recommendations emerging from these findings are:

Strengthen Safety Protocols in High-Risk Industries and States: Regulatory agencies and companies should collaborate to improve safety in sectors like steel manufacturing, mining, and heavy manufacturing, especially in states with dense industrial activity. This could include periodic audits, technology upgrades (e.g., automation in dangerous processes), and stringent implementation of safety management systems.

Enhance Training and Oversight during Identified Peak Risk Periods: Since late night shifts and certain months are high-risk, companies should ensure adequate supervision (e.g., more safety officers on night shifts), limit overtime hours, and perhaps schedule critical tasks to daytime whenever possible. Seasonal maintenance and checks before expected peak periods can prevent accidents related to those cycles.

Zero Tolerance for Safety Gear Non-compliance: Make PPE use non-negotiable. Provide appropriate gear and enforce its use through monitoring and penalties for non-compliance. Simultaneously, educate workers on how PPE saved lives (the data from this report can be used in safety training to show real-world effects).

Focus on Root Cause Elimination: For frequent causes like slips, institute 5S and housekeeping programs; for falls, invest in fall-arrest systems and scaffolding; for machinery, install guards and sensors. Engineering controls should be applied to remove hazards wherever feasible, rather than relying solely on worker behavior.

Protect and Educate Vulnerable Worker Groups: Young and inexperienced workers should undergo rigorous safety inductions. Contractor workers should be treated on par with regular employees in safety briefings and provision of protective gear. Encourage a culture where any worker can speak up about safety (empowerment can prevent accidents by addressing hazards before they cause harm).

Emergency Preparedness: Given that some accidents will inevitably still occur, preparedness can reduce their impact. High-risk facilities should have regular emergency drills (for fire, chemical leak, etc.), on-site medical response, and community emergency plans if public could be affected.

Data-Driven Continuous Improvement: Industries should continue to collect detailed data on incidents and near-misses. The approach used in this report can be adopted as a continuous practice: periodically analyzing new data to spot emerging trends (e.g., if a new technology introduces a new type of risk) and measuring the effectiveness of interventions (e.g., did accident rates drop after a safety campaign?).

In performing this analysis, we demonstrated the power of data analytics in the domain of occupational safety. By basing decisions on evidence (such as the patterns revealed in the dataset), organizations can move beyond reactive measures to proactive risk management. The insights gained not only serve the specific dataset at hand but also illustrate a methodology that can be applied to other safety data.

Future Work: While this report covered extensive ground, there are areas for further exploration. Future analysis could incorporate severity cost (using the Damage Index to quantify economic loss per incident) to prioritize costliest accident types. Predictive modeling might be applied to foresee which factors combination could lead to an accident (a form of risk scoring each job or shift). Additionally, qualitative factors like safety culture or compliance audits (if data were available) could be correlated with accident rates to scientifically validate the impact of safety culture. Broadening the data to include more years or other countries could also provide comparative insights. Finally, a deeper dive into near-misses (events that almost caused an accident) would be valuable, as preventing those could preclude actual accidents.

In conclusion, ensuring industrial safety is a continuous journey. The findings of this project provide a roadmap for stakeholders to concentrate their efforts where it matters most. By addressing the identified issues – temporal peaks, geographical hotspots, high-risk industries, demographic vulnerabilities, and critical risk factors – industrial organizations in India can make great strides towards the goal of an accident-free workplace.

The hope is that analyses like these drive informed actions, leading to safer environments for all industrial workers and the communities around them.

**6. Link to source code:**

1. CS661-Big-Data-Project [Industrial Accident analysis].
   GitHub: https://github.com/Sushant0907/CS661-Big-Data-Project.git

# References

Kaggle Link
Kaggle: https://www.kaggle.com/code/koheimuramatsu/industrial-accident-causal-analysis/notebook