Data Preprocessing:

Adjusting the Categorical Values

replace\_values **=** {'Local\_01': 1, 'Local\_02': 2, 'Local\_03': 3, 'Local\_04': 4, 'Local\_05': 5, 'Local\_06': 6, 'Local\_07': 7, 'Local\_08': 8, 'Local\_09': 9, 'Local\_10': 10, 'Local\_11': 11, 'Local\_12': 12}

data['Local'] **=** data['Local']**.**map(replace\_values)

replace\_values **=** {'I': 0, 'II': 1, 'III': 3, 'IV': 4, 'V': 5}

data['Accident Level'] **=** data['Accident Level']**.**map(replace\_values)

replace\_values **=** {'I': 0, 'II': 1, 'III': 3, 'IV': 4, 'V': 5, 'VI': 6}

data['Potential Accident Level'] **=** data['Potential Accident Level']**.**map(replace\_values)

**del** replace\_values

A screenshot of a computer

Description automatically generated with medium confidence

Checking for Null Values – There are no null values in any of these below columns

Date 0

Country 0

Local 0

Industry Sector 0

Accident Level 0

Potential Accident Level 0

Gender 0

Employee Type 0

Critical Risk 0

Description 0

dtype: int64

* Added additional columns to define the Year, Month , WeekofYear , Day , Quarter of the incident

Graphical user interface, application

Description automatically generated

* The Country that is most effected is Country\_01
* Mining seems to be the most affected Industry sector when it comes to safety and health issues

Chart

Description automatically generated

* Among different types of employees – Third Party employees working onsite seems to have been affected the most
* Potential accidents level seems to be the highest for Sev 4

4 143

3 106

2 95

1 49

5 31

Chart, bar chart

Description automatically generated

* Male seems have affected more than females in this data set

Male 403

Female 22

A picture containing shape

Description automatically generated

Icon

Description automatically generated

A picture containing text, clipart

Description automatically generatedA map of the world

Description automatically generated with low confidence

Accident Level vs Potential Accident Level analysis :-

1. While potential accidents levels were very severe, the actual accidents that have happened are less severe. Probably there were some mitigation actions that were taken, which are not mentioned in the data.

Chart

Description automatically generated

1. Always the actual Accident Level has been less than the potential one. So looks like some  mitigation steps have been taken

Chart, bar chart

Description automatically generated

1. Graphical user interface

   Description automatically generated with low confidence

* Quarterly analysis of Accidents – More accidents have happened in the later quarters – Q1 and Q2 compared to Q3 and Q4
  + Q2 had the highest Accident Level of severity-I.
  + No severity VI (highest severity) accidents have happened in the Q4

Chart, bar chart

Description automatically generated

# Gender vs Accident Level / Potential Accident Level analysis

* Looking at the graph more Male are prone to accidents compared to females. A picture containing chart

  Description automatically generated
* Looking at the graph below male have higher potential accident levels compared to females

Chart

Description automatically generated

# Employee type vs Accident and Potential Accident Level analysis

* Third party employees are more prone to Accidents than other two types.
* Third Party (Remote) employees are least prone to Accidents.

Chart, bar chart

Description automatically generated

* Employees (Full time) are more prone to accidents compared other two for Accident Severity level I (1).
* For rest of the Accident Severity levels, Third Party employees have hit into accidents, more than rest of the two employee categories on an average.
* Third Party (Remote) working employees are less prone to accidents across all the accident level severities.

Graphical user interface, chart, bar chart

Description automatically generated

# Industry Sector vs Accident and Potential Accident Level analysis

# Chart, bar chart Description automatically generated

* Mining industries always had higher accidents in all the accident severity levels.
* Most fatal accident i.e. Level 5 has happened only in Mining and Metals category only. Other industries did not see fatal accidents of level 5 yet.

Chart, bar chart

Description automatically generated

* Mining industry seems to have maximum potential for severity 4 accidents
* Other than Mining , none of the other industries had potential of highest severity incident

# Country vs Accident and Potential Accident Level analysis

* Country 1 has the highest number of accidents that have happened in the past.
* The most fatal accidents have happened in Country 1 only.
* Country 1 and 2 share the same number of Sev2 accidents

Chart, bar chart

Description automatically generated

* Country 1 has highest potential of Sev 4 accidents.
* The most fatal Sev 6 level accidents can happen in country 1 only.
* The potential of Sev 1 accidents is more in country 3
* The potential of Sev1 accidents in Country 1 is much lesser still the actual accidents count is more.

Chart, bar chart

Description automatically generated

# Month wise accident level analysis

Chart, bar chart

Description automatically generated

* Most of the accidents have happened in the initial 6 months. From July onwards the number and the severity has been reduced
* Hence this is an important field to influence the potential accident level
* None of the fatal accidents have happened from August onwards.

A screenshot of a cell phone

Description automatically generated with low confidence

* Similarly , the potentials of accidents is higher in the first 6 months of the year.
* Month of Feb shows the highest number of Sev 2 category accident’s potential.
* Month plays an important role in Potential Accident target level prediction

# Weekday vs Accident level analysis

Chart, bar chart

Description automatically generated

* There is less potential of Sev 5 accidents on Monday and Sunday.
* The most fatal accidents can happen only on Wednesday. (Looking back at the data – there is only one record in the data sheet on 25th Jan 2017 i.e. Wednesday on which the Potential of accident severity was 5

# Analyze Country and Employee Type

Chart, bar chart

Description automatically generated

* Country 1 has most third party contractors
* Country 3 has very less full time employees and no third party remote workers. Since there are overall less number of workers in the given sectors from this country, the percentage of accidents reported in less 10%

A map of the world

Description automatically generated with low confidence

* Country 2 has very less Thirdparty workers working on site. Most number of full time employees

# Analyse Country vs Industry Sector

Chart, bar chart

Description automatically generated

* Country 1 has the highest Mining industries compared to others. Also Mining industry has 56 % of accidents reported from the report. Thus no wonder if Country 1 has the highest number of incidents reported
* Country 3 has no Mining and Metal industries. Other industries have 11 % of accidents. Country 3 has 10 % of accidents from the reported ones.

# Analyze Country vs Gender ratio

Chart, bar chart

Description automatically generated

* In all the countries male workers are the most. Hence no wonder if the accidents are happening more for Males
* Country 3 does not have any female worker belonging to the given industry

# Analyze Employee type and Gender

Chart, bar chart

Description automatically generated

* In all the three employment categories there are more males and those are the most affected ones during these accidents

# Analyze Year , Industry sector and accidents …

Graphical user interface, chart

Description automatically generated

* Year 2016 saw more incidents than 2017
* Sev 3 category accidents were less than sev 4 in 2016
* Mining industry saw more incidents compared to other industries.

# Analyze the description data

Stop word analysis: Below are the most frequent stop words

Chart

Description automatically generated

Non stop word analysis:

There is an hyphen in the non stop words

Chart

Description automatically generated

N Gram analysis

Chart, funnel chart

Description automatically generated

Chart, funnel chart

Description automatically generated

Text

Description automatically generated