

# **SAFEX.PYBOTS**

**Exchanging of Industrial Safety using NLP based Python Chat Bots**

**TEAM NAME:** SAFEX.PYBOTS



**TEAM MEMBERS:**

- 1) APOORV DIXIT
- 2) DIVYASHREE C
- 3) PRINCE
- 4) SRINATH YASODA
- 5) SRINAYANI

## **Table of Contents:**

<b>1. Project Summary</b>	.....	<b>03</b>
<b>2. Data Interpretation</b>	.....	<b>05</b>
<b>3. Univariate Analysis</b>	.....	<b>06</b>
<b>4. Bivariate Analysis</b>	.....	<b>25</b>
<b>5. Groupby Analysis</b>	.....	<b>61</b>
<b>6. Word Cloud Analysis</b>	.....	<b>63</b>
<b>7. Inference</b>	.....	<b>69</b>

## **INTERIM REPORT OF CAPSTONE PROJECT ON NLP2 – CHATBOT**

### **PROJECT OBJECTIVE:**

Design and build Chabot by employing ML/DL techniques which can help the professionals in determining the potential level and accident level involved in any accident and to highlights the safety risk as per the incident description.

### **PROBLEM STATEMENT AND ABSTRACT:**

For enhancing one's self-esteem, wellbeing, and social mobility, work is very important. However, during work activities if any accidents occurred leads to impairments to worker's health and which in-turn leads to serious social and economic repercussions.

Globally, it is estimated between the range of 1.8% to 6.0% cost of work-related accidents and ill-health out of gross domestic product. It is also estimated as, around 2.3 million people will die in a year around the world due to work-related activities. Specifically, in Brazil, around 2500 such deaths per year which corresponds to one death for every 3.5hr.

Since human activity is involved, accidents & injuries are common despite all the safety measures and precautions put in place. Such injuries can also prove fatal. Industrial accidents can turn depending on the type of industry. For example, a mere spark in a firecracker factory can burn the whole plant leading to loss of lives and property. Workplace injuries are a big concern for both workers and management. It is imperative to classify industrial incidents into different categories and determine whether the event was merely an accident, due to negligence or by incompetence. This avoids reoccurrences, reduce frequency of occurrence & severity and minimize the effects. To achieve this, we employ exploratory data analysis on a dataset from one of the biggest Brazilian industries and find out the top reasons for industrial accidents, nature of accidents, type of employees being injured and so on. We also aim to develop a chatbot application using natural language processing to classify the accident into various critical risks by looking at the description of the accident.

### **DOMAIN:**

Industrial safety. NLP based Chatbot.

### **DATA DESCRIPTION:**

This database is basically records of accidents from 12 different plants in 03 different countries which every line in the data is an occurrence of an accident.

### **Columns description:**

- Data: timestamp or time/date information
- Countries: which country the accident occurred (anonymised)
- Local: the city where the manufacturing plant is located (anonymised)
- Industry sector: which sector the plant belongs to
- Accident level: from I to VI, it registers how severe was the accident (I means not severe but VI means very severe)
- Potential Accident Level: Depending on the Accident Level, the database also registers how severe the accident could have been (due to other factors involved in the accident)
- Genre: if the person is male or female
- Employee or Third Party: if the injured person is an employee or a third party
- Critical Risk: some description of the risk involved in the accident
- Description: Detailed description of how the accident happened

### **LINK FOR THE DATASET**

[Industrial Safety and Health Analytics Database | Kaggle](#)

## DATA INTERPRETATION

Based on the entire dataset -

### 1. Shape and Top 5 details

```
In [2]: data=pd.read_csv('IHMStefanini_industrial_safety_and_health_database_with_accidents_description.csv')
print("Shape of the dataset is : ",data.shape)
data.head()
```

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

The dataset contains 425 instances and 11 attributes

### 2. Missing Values

```
In [4]: data.isnull().sum()
```

```
Out[4]: Data          0
Countries      0
Local          0
Industry Sector 0
Accident Level 0
Potential Accident Level 0
Genre          0
Employee or Third Party 0
Critical Risk   0
Description     0
dtype: int64
```

There are NO NULL values

### 3. Checking of Duplicates and Drop if any-

```
: print("Shape of the dataset before duplicates deletion is : ",data.shape)
print('Number of duplicates in the dataset : ',data.duplicated().sum())
data.drop_duplicates(inplace=True)
print("Shape of the dataset after duplicates deletion is : ",data.shape)
```

```
Shape of the dataset before duplicates deletion is : (425, 10)
Number of duplicates in the dataset : 7
Shape of the dataset after duplicates deletion is : (418, 10)
```

#### 4. Checking of dtypes and data info

```
print('*****Checking the dtypes*****\n')
print(data.dtypes)
print('-----')
print('\n *****Checking the data info***** \n')
print(data.info())

*****Checking the dtypes*****  
  
Data          object  
Countries     object  
Local         object  
Industry Sector    object  
Accident Level    object  
Potential Accident Level    object  
Genre          object  
Employee or Third Party    object  
Critical Risk      object  
Description     object  
dtype: object  
-----  
  
*****Checking the data info*****  
  
<class 'pandas.core.frame.DataFrame'>  
Int64Index: 418 entries, 0 to 424  
Data columns (total 10 columns):  
 #   Column            Non-Null Count  Dtype     
 ---  --     
 0   Data              418 non-null    object    
 1   Countries        418 non-null    object    
 2   Local             418 non-null    object    
 3   Industry Sector  418 non-null    object    
 4   Accident Level   418 non-null    object    
 5   Potential Accident Level 418 non-null    object    
 6   Genre             418 non-null    object    
 7   Employee or Third Party 418 non-null    object    
 8   Critical Risk     418 non-null    object    
 9   Description       418 non-null    object    
dtypes: object(10)  
memory usage: 35.9+ KB  
None
```

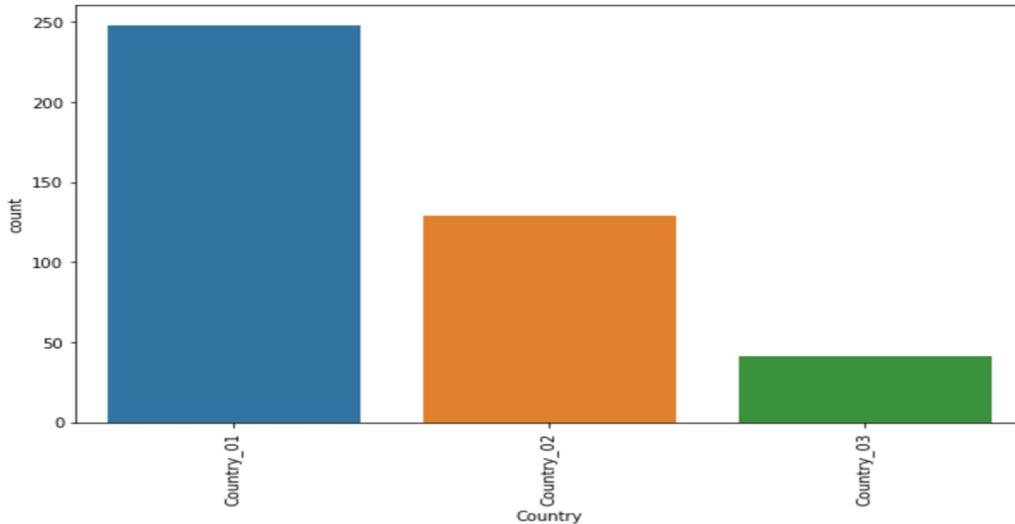
## UNIVARIATE ANALYSIS:

### 1. COUNTRY

#### COUNT PLOT:

It is used to show the counts of observations in each categorical bin using bars. For instance, the count plot () method is used to display the number of accidents happened in country\_01, country\_02 and country\_03 for the country variable.

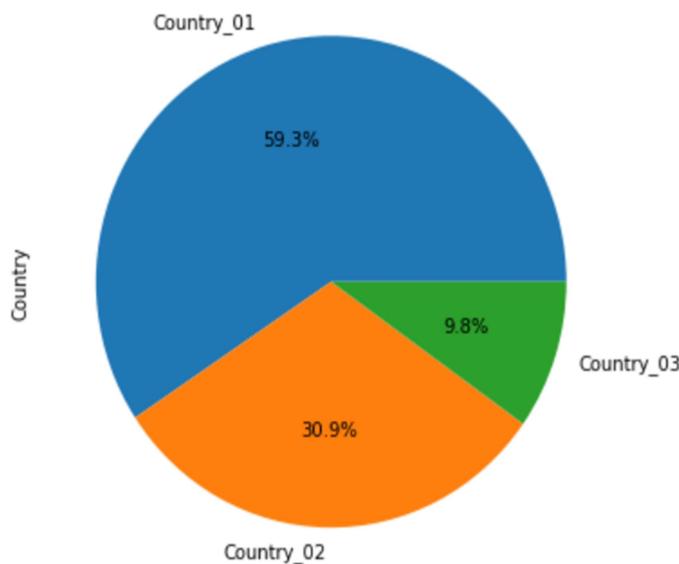
The following are the observations that are made for the country count plot ()-



1. About 250 accidents have happened in the country \_01.
2. About 125 accidents have happened in the country\_02 and about 50 accidents happened in country\_03.
3. The maximum number of accidents happened in Country\_01 and least number of accidents happened in Country\_03.

#### **PIE CHART:**

A pie chart represents data in a circular graph containing slices of different colours. It is used to study the proportion of numerical data. It shows the proportion of data as a percentage of a whole. For instance, the pie chart for the country variable gives the number and percentage of accidents that have occurred in the country\_01, country\_02 and country\_03.

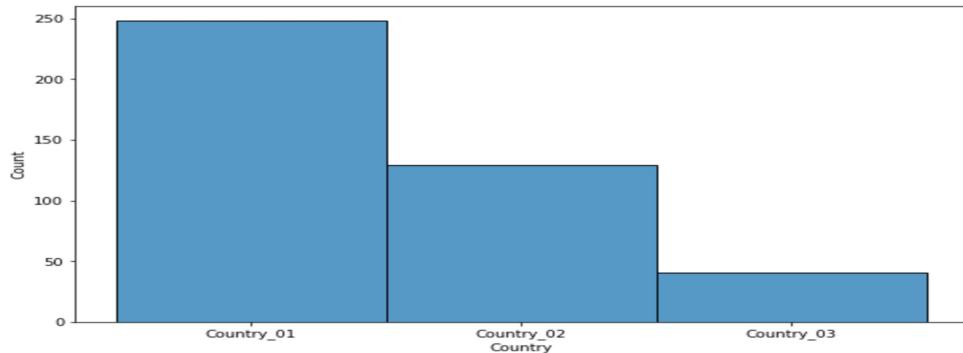


## OBSERVATIONS-

1. The number of maximum accidents taken place in country\_01. i.e.- 59.3%.
2. The least number of accidents took place in country\_03. i.e- 9.8%

## HISTPLOT:

It is used to show the distribution of the datasets. For instance, for the variable country it displays the number of accidents that happened in the country\_01, country\_02 and country\_03.

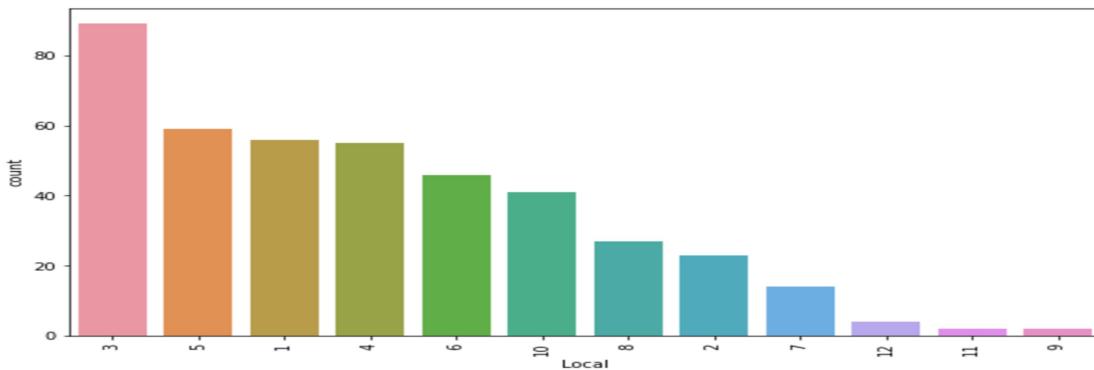


OBSERVATIONS- It can be noted that country\_01 has the maximum number of accidents with a count of 248 and the minimum number of accidents has happened in country\_03 with a count of 41.

## 2. LOCAL:

### COUNT PLOT:

The local variable determines the region of the accident. The countplot() method helps to calculate the number of accidents happened in region wise.

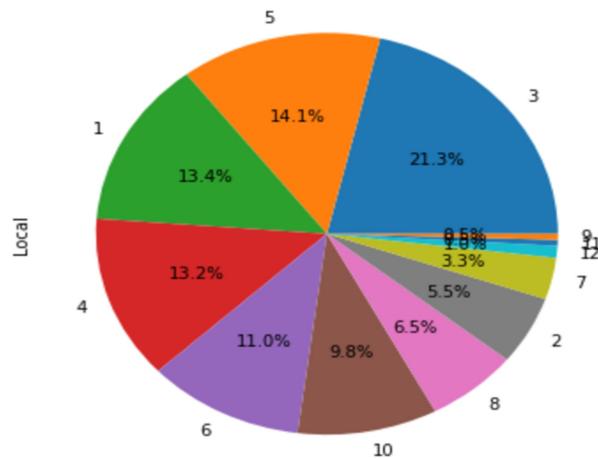


## OBSERVATION-

1. It is observed that in the local\_03 has the maximum number of accidents with a count of about 90.
2. The least number of accidents happened in the local\_09 region.

## PIECHART-

The pie chart for the local variable gives the number and percentage of accidents that have occurred in the local\_01, local\_02 and so on upto local\_11

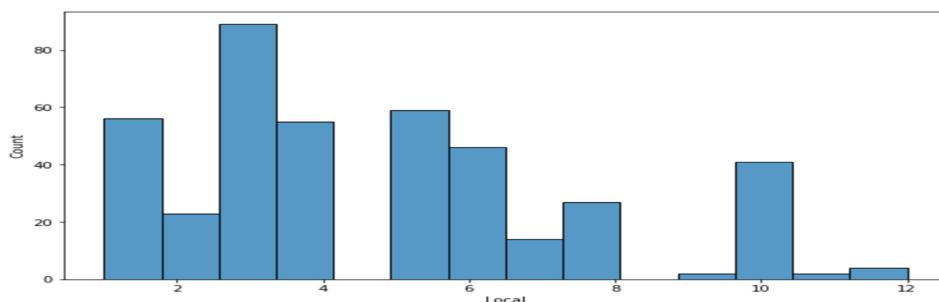


## OBSERVATIONS:

1. From the pie chart it can be observed that at the local\_03 maximum numbers of accidents have taken place with about 21.3%.
2. The least number of accidents have happened in local\_09. i.e-0.5%

## HISTPLOT-

The histplot() of the local variable is useful to determine the distribution.



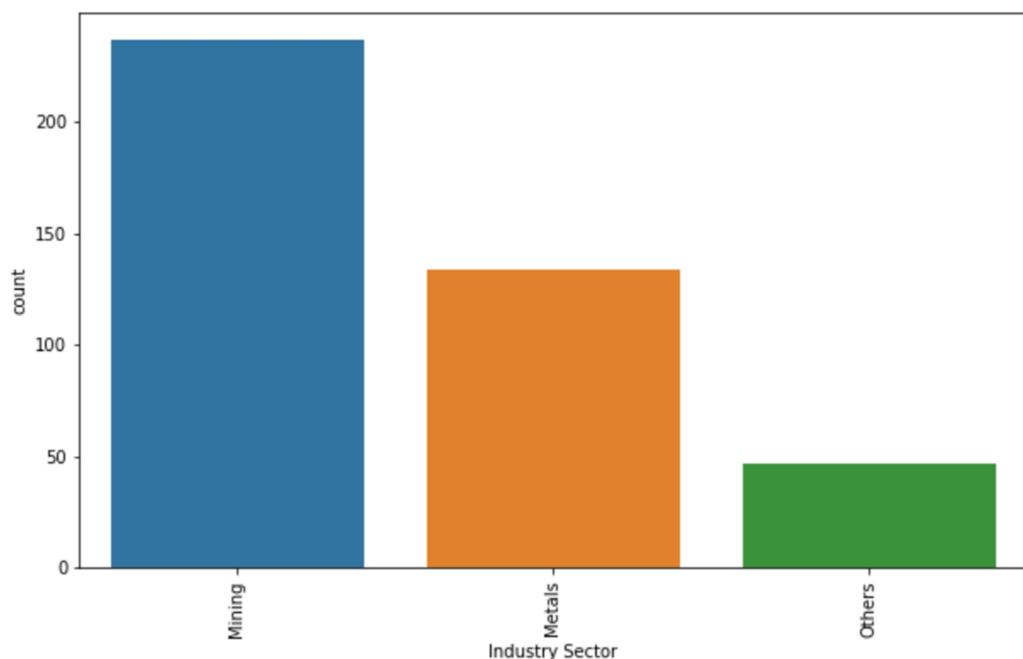
## OBSERVATIONS-

1. It is observed that the number of accidents at local\_03 is about 90.

## 3. INDUSTRY SECTOR-

### COUNTPLOT-

A countplot() method in industrial sector is used to determine the number of accidents that had happened due to different industrial sectors such as mining, metals and others.

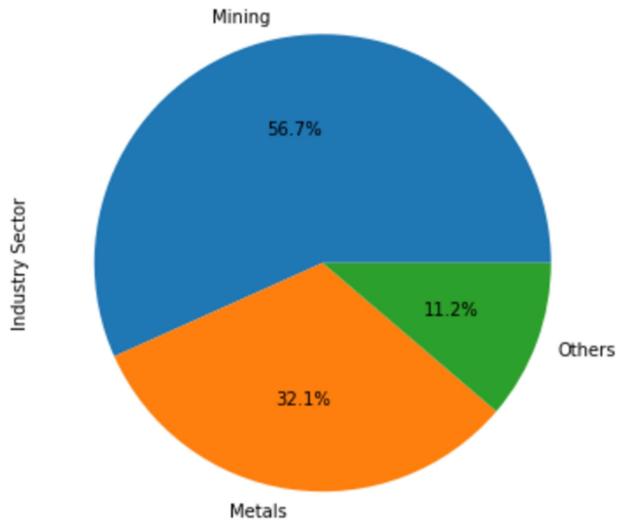


## OBSERVATIONS-

It can be determined that maximum number of accidents happened due to mining sector (237) than metals and others.

### PIECHART-

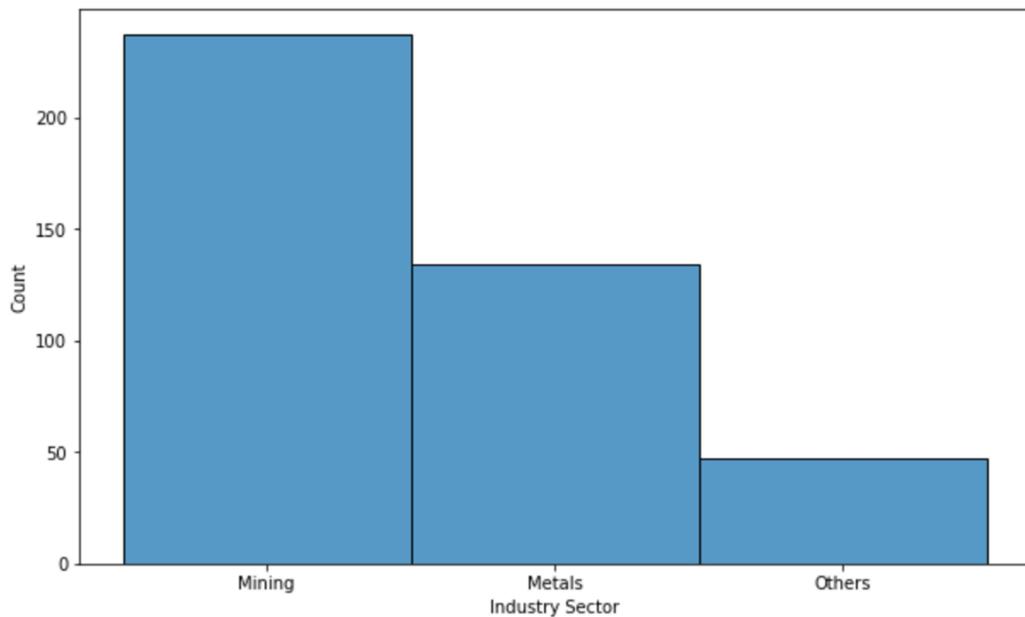
A pie chart is used to determine the percentage of accidents that were caused due to mining, metals and other industrial sector.



**OBSERVATIONS-** It can be concluded that the maximum number of accidents happened due to mining i.e- 56.7% of the total accidents.

### HISTOGRAM-

A histplot is used to determine the distribution of different industrial sectors.

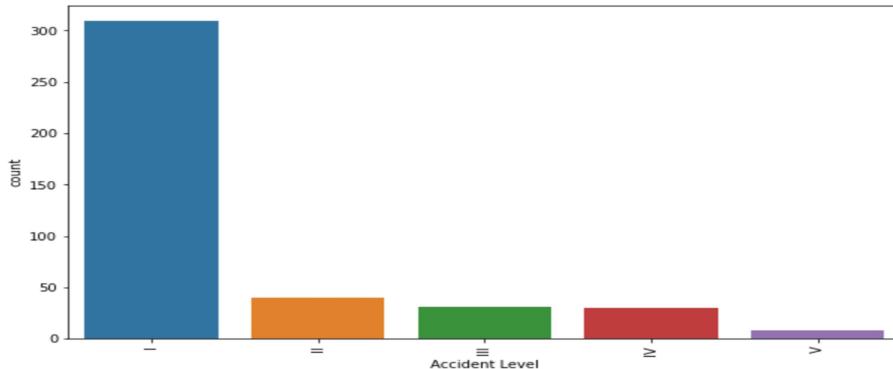


**OBSERVATIONS-** It can be noted that maximum number of accidents happened due to mining sector.

#### 4. ACCIDENT LEVEL-

The accident level provided the information about the severity of the accident. Where, I represents less severity and V represents high severity.

**COUNTPLOT**- The countplot() of the accident level determines the number of accidents that has happened at level I, II, III, IV, and V.

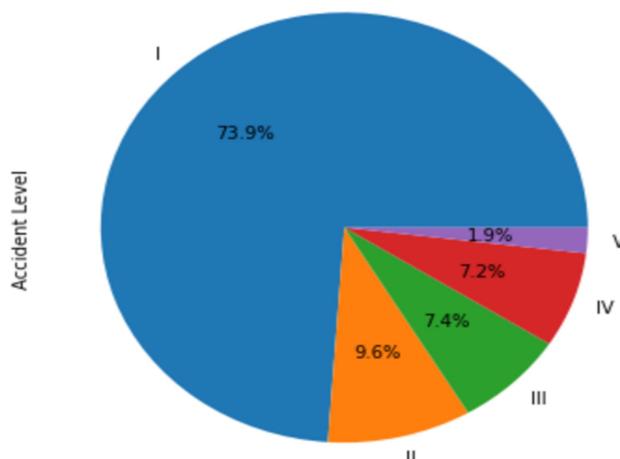


#### OBSERVATIONS-

1. It can be observed that the number of accidents happened at level I (300) are higher than level V (less than 50).
2. Hence, it can be concluded that maximum number of less severe accidents took place.

#### PIE CHART-

The pie chart helps to find out the percentage of accident level I, II, III, IV and V accidents out of all the accidents.

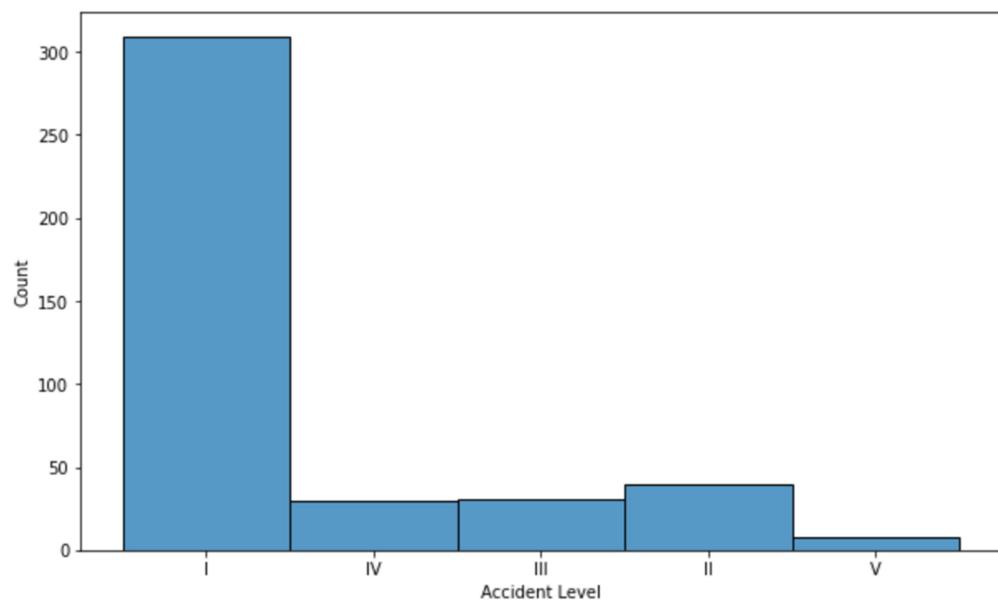


## OBSERVATIONS-

1. It can be noted that maximum percentage of less severe accidents took place. i.e accident level I with 73.9%.

## HISTOGRAM-

The histogram plot could be used to plot the distribution of accident level I, II, III, IV and V.

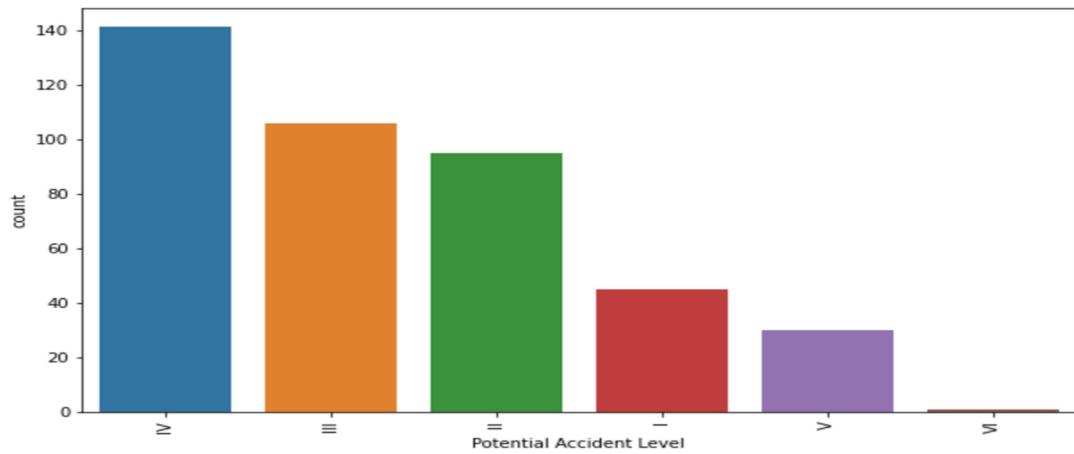


OBSERVATIONS- The number of accidents with less severity(309) are higher than more severity(8) with a huge difference.

## POTENTIAL ACCIDENT LEVEL-

The potential accident level is based on the accident level with other factors. Here, Potential accident level I is less severe and VI is more severe.

**COUNT PLOT-** It determines the number of potential accident level caused by all levels.

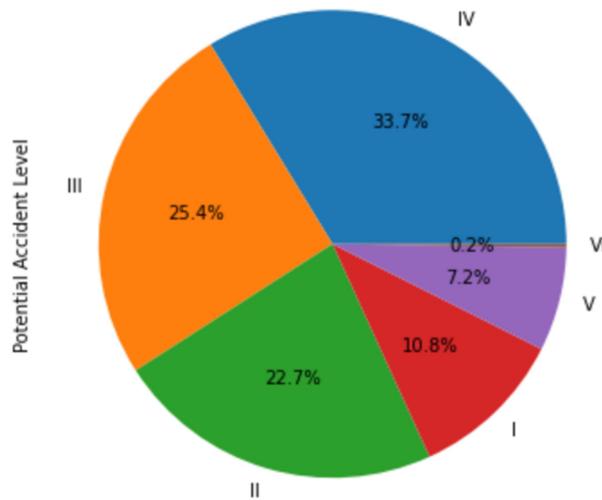


#### OBSERVATIONS-

The number of accidents due to potential accident level IV are maximum with a count of 141.

#### PIECHART-

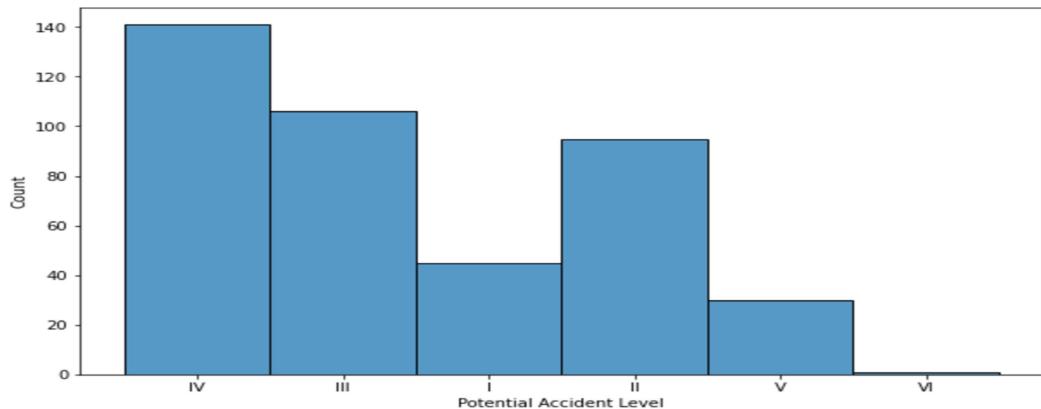
The pie chart determines the percentage of potential accidents caused by level I to VI.



OBSERVATIONS- It can be determined that the percentage and number of accidents caused due to potential accident level IV are maximum.

#### HISTOGRAM-

It gives the distribution of the potential accident level.

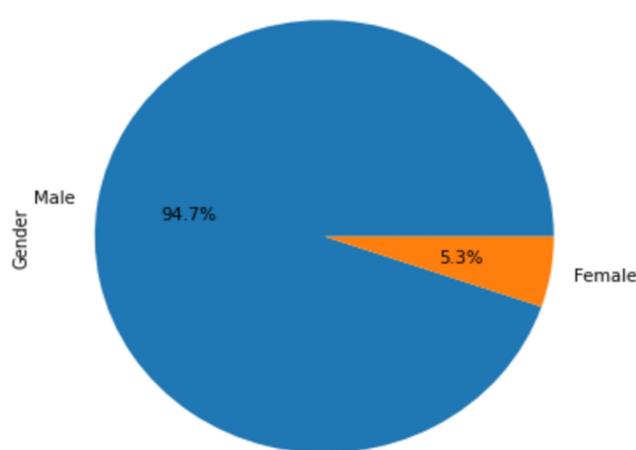
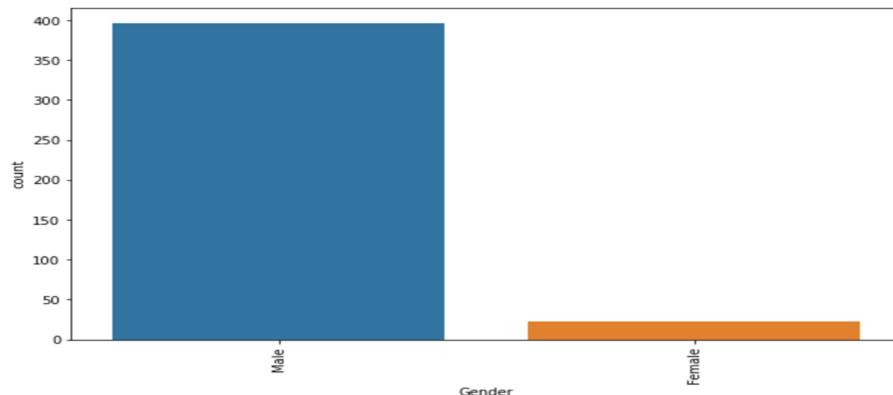


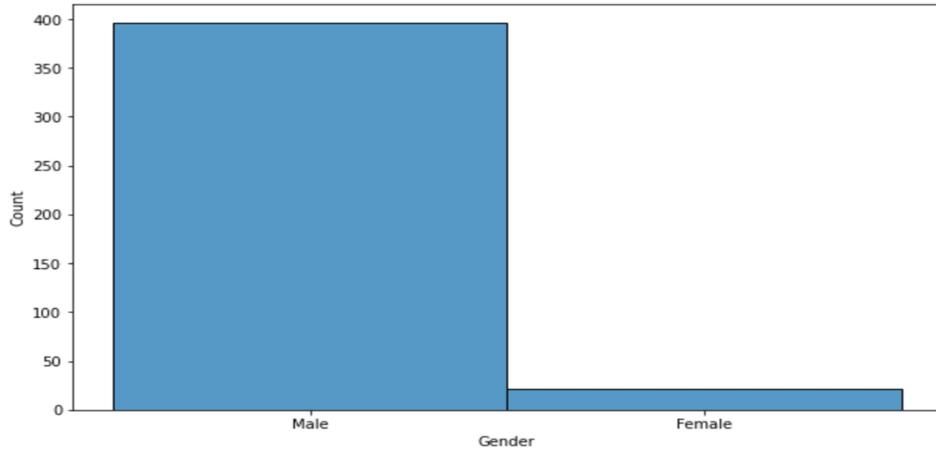
#### OBSERVATIONS-

It can be determined that the number of accidents caused by potential accident level IV are maximum.

#### 5. GENDER-

The gender is used to determine the number of accidents caused to both female and male.

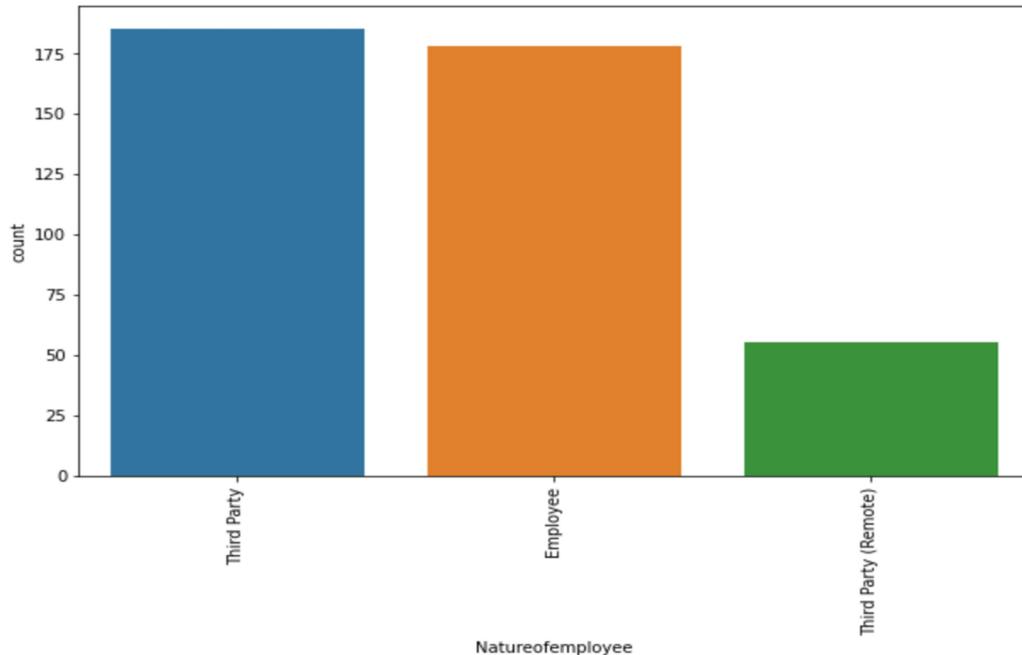




## 6. NATURE OF EMPLOYEE-

The nature of employee is divided into 3 types that is either third party, engineering and third party (remote).

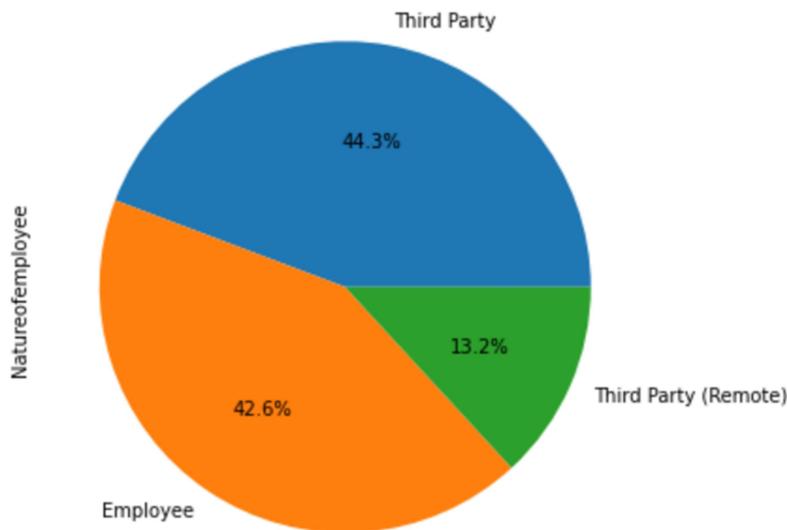
**COUNT PLOT()**- It is used to determine the number of accidents happened according to the nature of the employee such as third party, engineering and third party (remote).



## OBSERVATIONS-

It can be noted that the third party type of employee has faced many accidents and third party(remote) has faced the least accidents.

### PIE CHART-

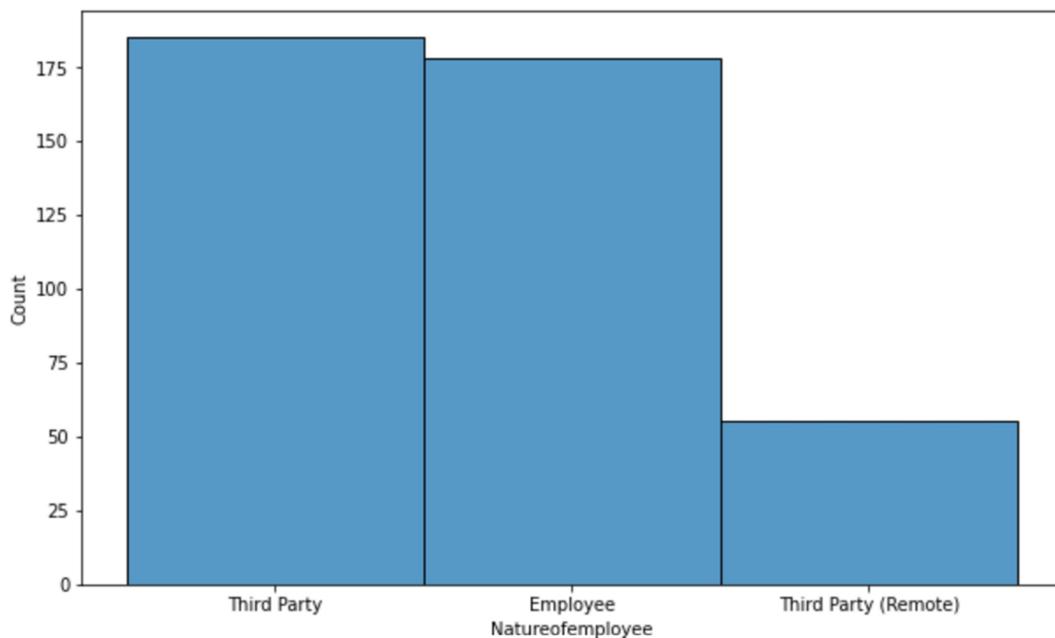


### OBSERVATIONS-

It can be noted that the third party employee has happened to face most of the percentage of accidents. i.e- 44.3%.

### HISTOGRAM-

It is used to show the distribution of the nature of employee such as third party, employee and third party(remote).

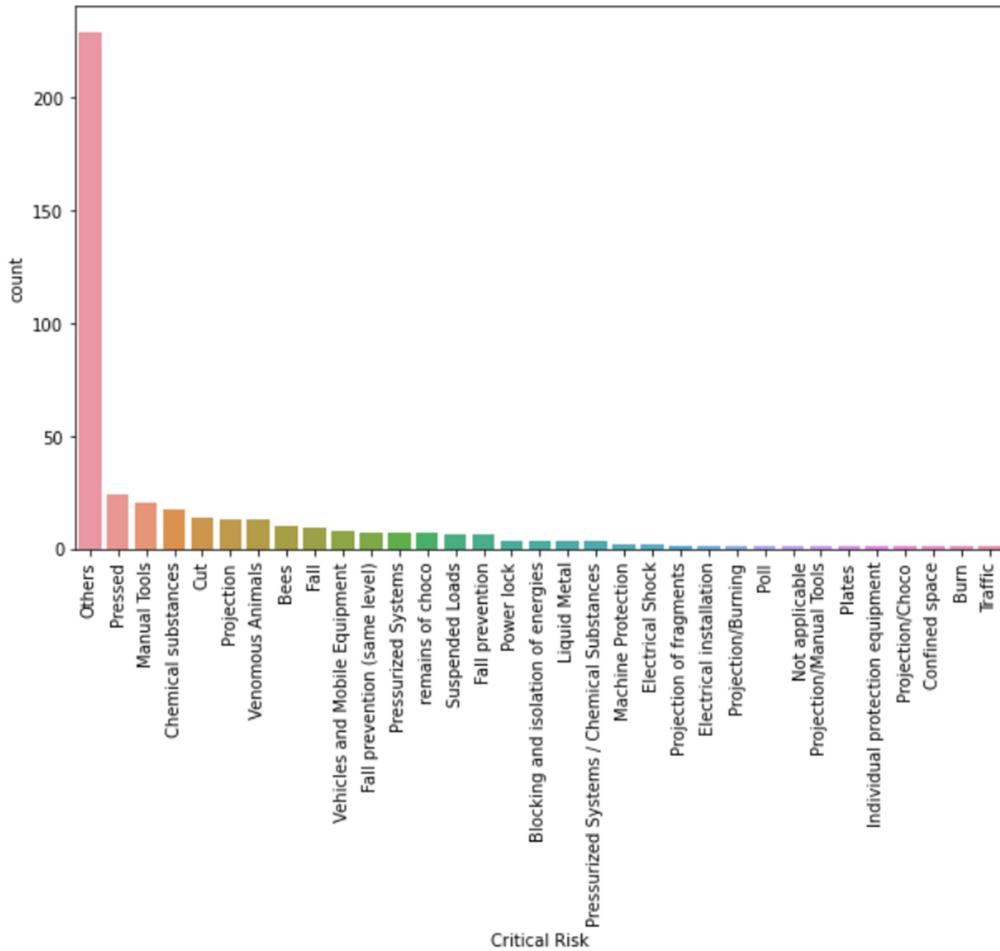


## 7. CRITICAL RISK-

It is used to determine the number of accidents caused by different critical risks such as pressed, pressurized system, electric shock and others.

### COUNT PLOT-

The countplot() method is used to give the count of the number of accidents caused by different critical risks

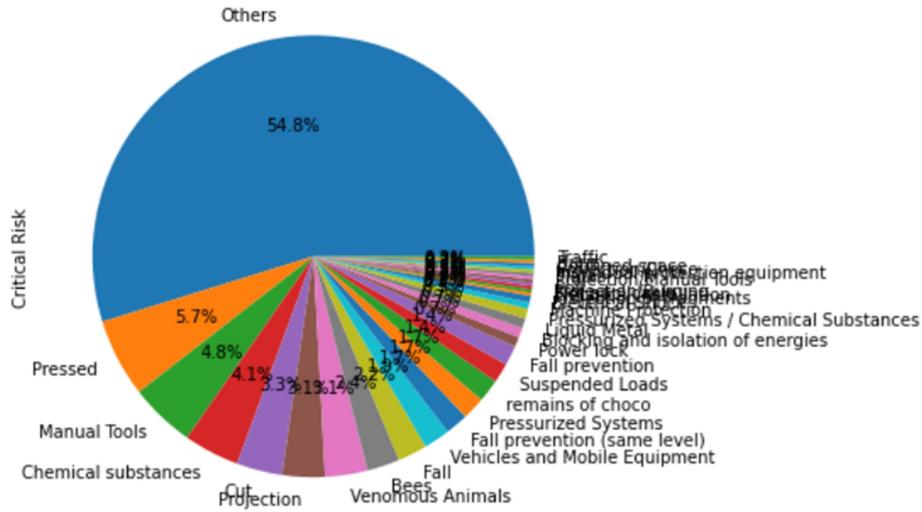


### OBSERVATONS-

It can be noted that the maximum number of accidents have taken place with the critical risk “others” and minimum number of accidents took place with risk poll or traffic.

### PIE CHART-

It is used to show the distribution of the different critical risks.

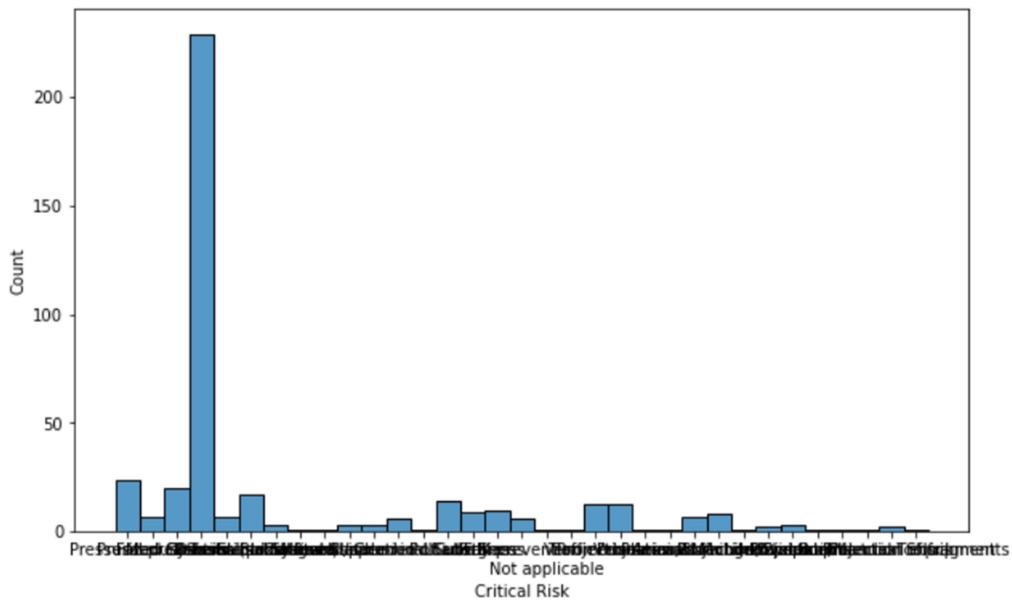


#### OBSERVATIONS –

It can be noted that most of the percentage of accidents have taken place with the critical risk “others”.

#### HISTPLOT-

It is used to show the distribution of the critical risks.



```

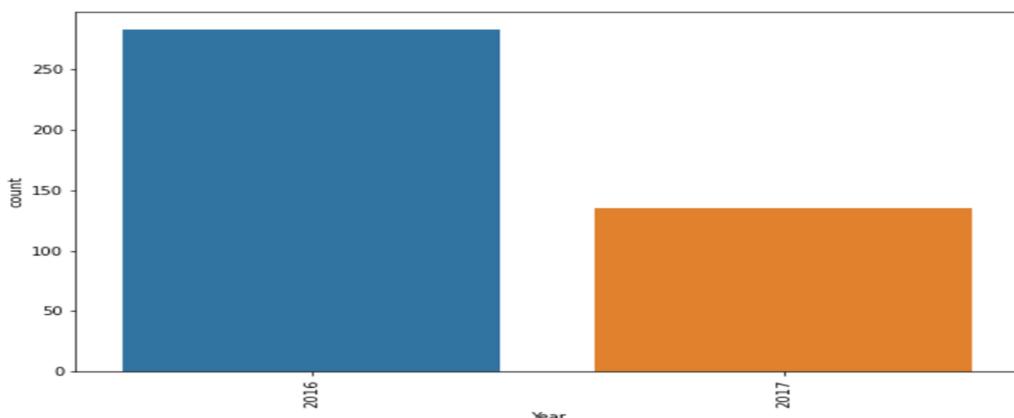
-----
Value Counts for feature: Critical Risk
-----
Others                                229
Pressed                               24
Manual Tools                          20
Chemical substances                   17
Cut                                    14
Projection                            13
Venomous Animals                     13
Bees                                    10
Fall                                    9
Vehicles and Mobile Equipment        8
Fall prevention (same level)          7
Pressurized Systems                  7
remains of choco                      7
Suspended Loads                      6
Fall prevention                       6
Power lock                            3
Blocking and isolation of energies   3
Liquid Metal                          3
Pressurized Systems / Chemical Substances 3
Machine Protection                    2
Electrical Shock                      2
Burn                                    1
Projection of fragments              1
Electrical installation              1
Projection/Burning                   1
Poll                                    1
\nNot applicable                     1
Projection/Manual Tools              1
Plates                                 1
Individual protection equipment     1
Projection/Choco                      1
Confined space                        1
Traffic                                1
Name: Critical Risk, dtype: int64

```

## 8. YEAR-

The year attribute is used to determine the number of accidents that had happened according to the years .i.e- 2016 and 2017.

**COUNT PLOT-** The countplot() method is used to predict the number of accidents taken place in 2016 and 2017

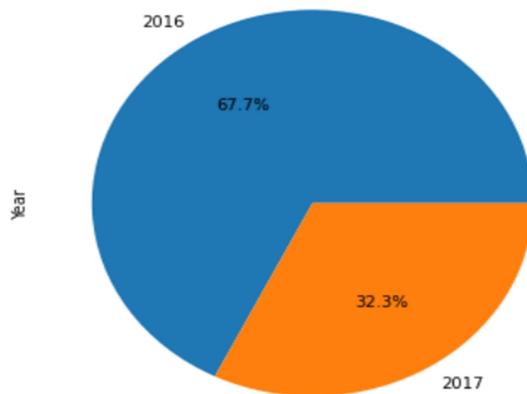


## OBSERVATIONS –

It can be noted that the number of accidents that happened in 2016 (above 250) are higher than 2017(150).

### PIE CHART-

It is used to show the percentage of accidents that had taken place according to the years.

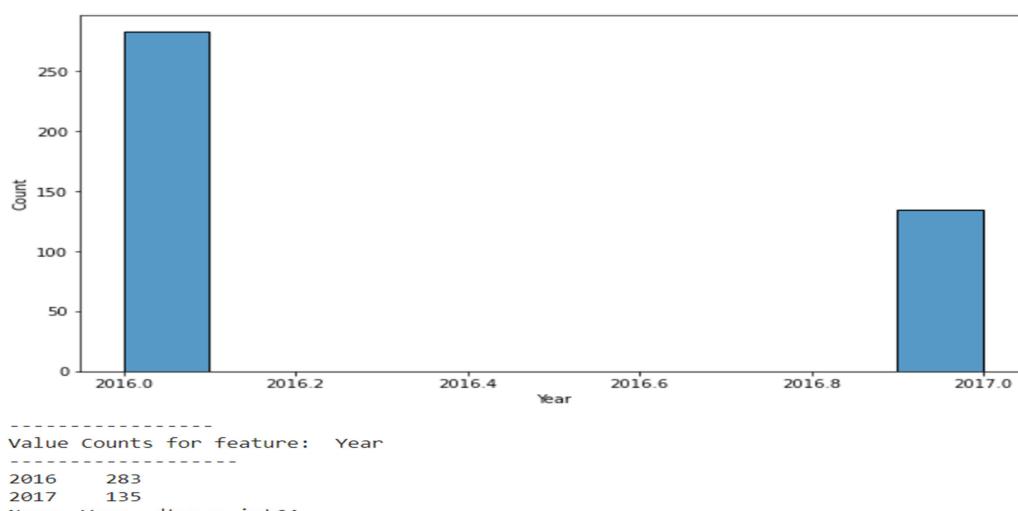


## OBSERVATIONS-

It can be noted that the percentage of accidents occurred in 2016 are maximum

### HISTPLOT-

It shows the distribution of the accidents that had taken place according to the years

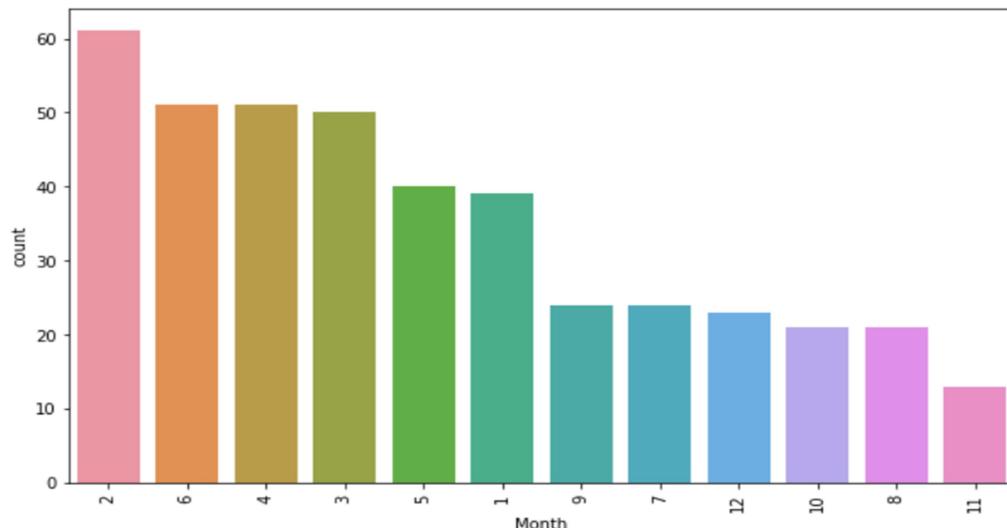


## **10. MONTH-**

It is used to determine the number of accidents that had occurred in all the months of the years.

### **COUNTPLOT-**

It is used to determine the number of accidents that had occurred according to their months.

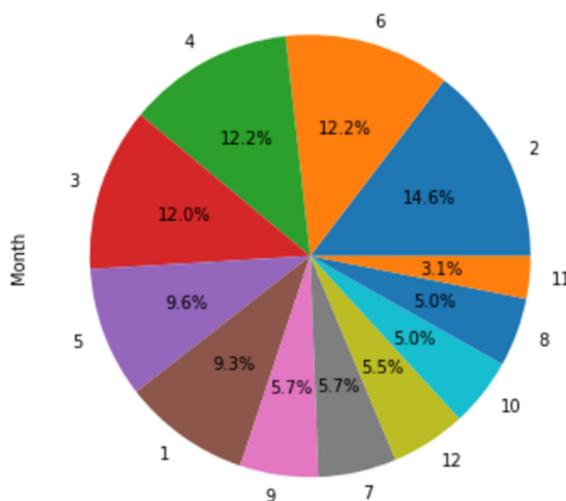


### OBSERVATIONS –

It can be noted that the maximum number of accidents happened in Feb and minimum number of accidents occurred in November.

### **PIECHART-**

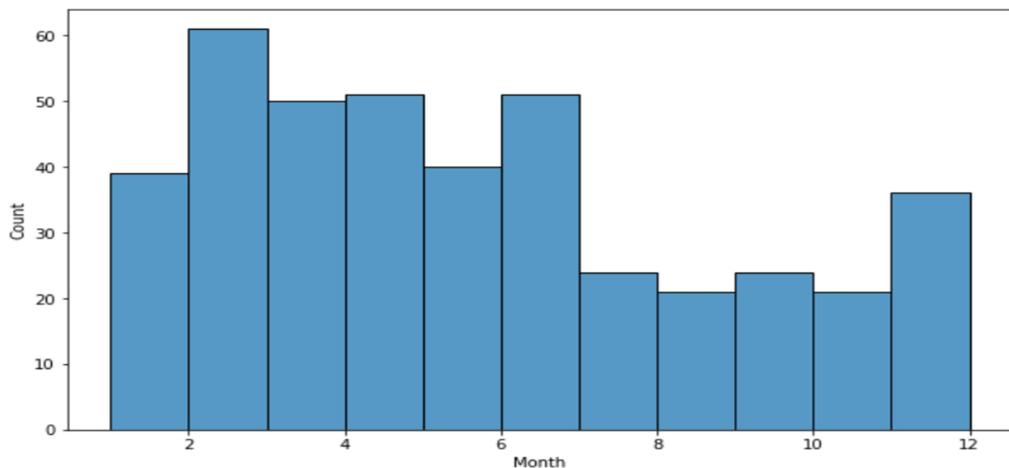
It is used to show the number of accidents that had happened in all the months in terms of percentage.



**OBSERVATION-** It can be noted from the above piechart that the maximum percentage of accidents happened in the month February.

### **HISTOGRAM-**

The histplot is used to determine the distribution of accidents according to the months.

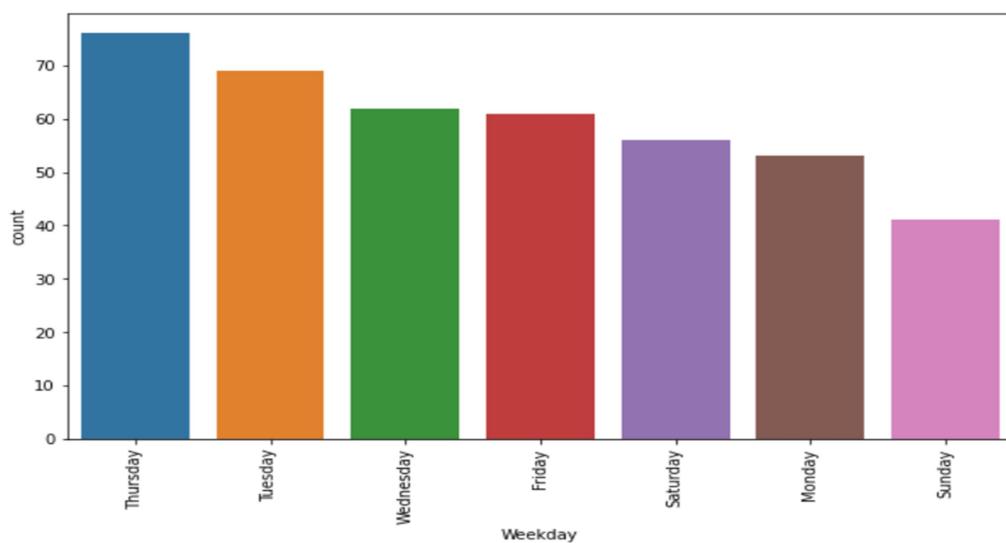


### **11. WEEKDAY-**

It is used to determine the number of accidents on the different days of a week.

### **COUNTPLOT-**

The countplot() method is used to determine the number of accidents that had occurred in the days of the week.

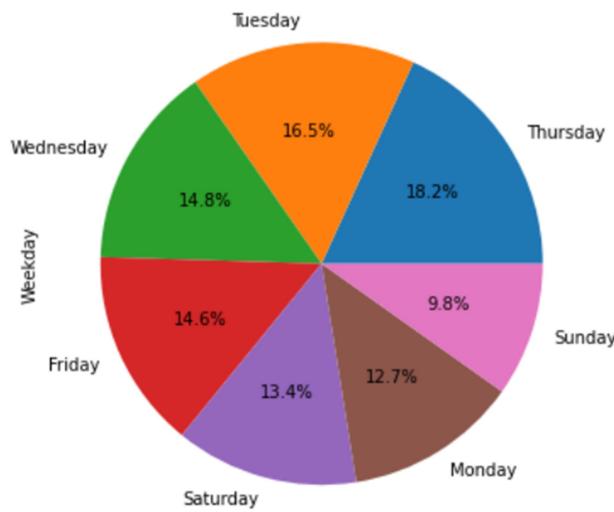


## OBSERVATIONS –

It can be determined that maximum number of accidents have taken place on Thursday with a count of about 76 and least number of accidents took place on Sunday.

## PIECHART-

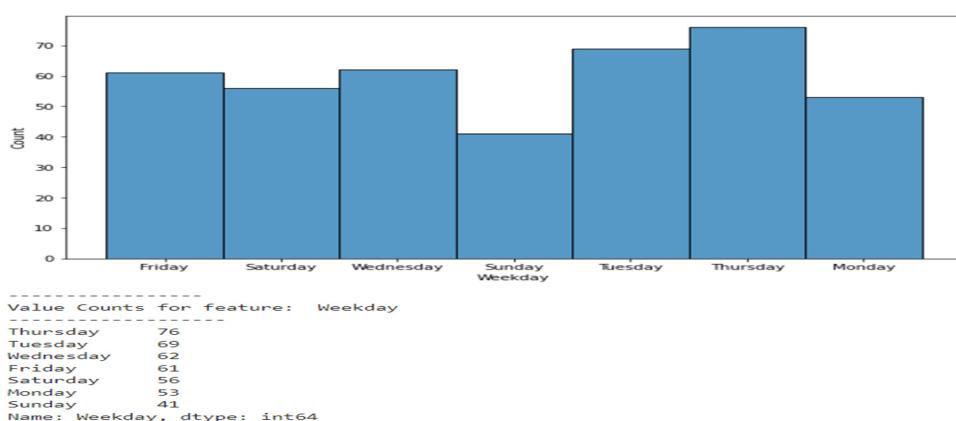
It is used to determine the percentage of accidents that took place in days of a week.



OBSERVATION – It can be observed that maximum percentage of accidents happened on Thursday with 16.5% and least number of accidents happened on Sunday.

## HISTPLOT-

It is used to determine the distribution of accidents according to the days.



OBSERVATION –

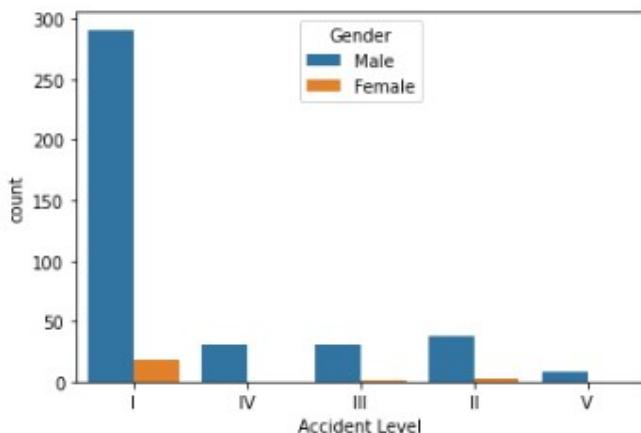
It can be noted that maximum number of accidents have taken place on Thursday

## BIVARIATE ANALYSIS:

### 1 Gender vs RestAll

#### 4.1 Gender vs Accident level

Count Plot:



Cross table Analysis:

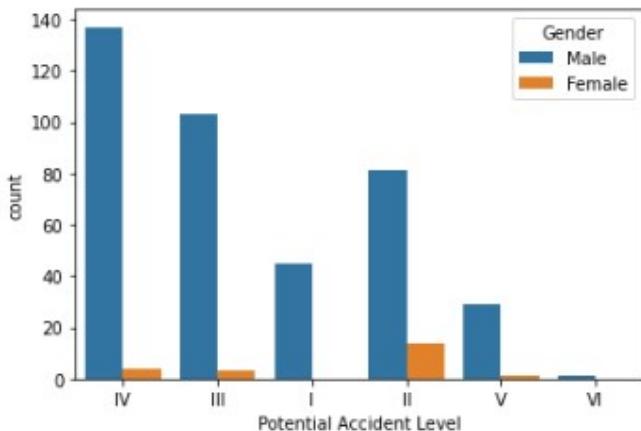
Gender	Female	Male
Accident Level		
I	18	291
II	3	37
III	1	30
IV	0	30
V	0	8

Observation –

From the above plots, it can be determined that the most of the accidents happened at level I with gender male.

## 4.2 Gender Vs Potential Accident Level

Count Plot:



Cross table Analysis:

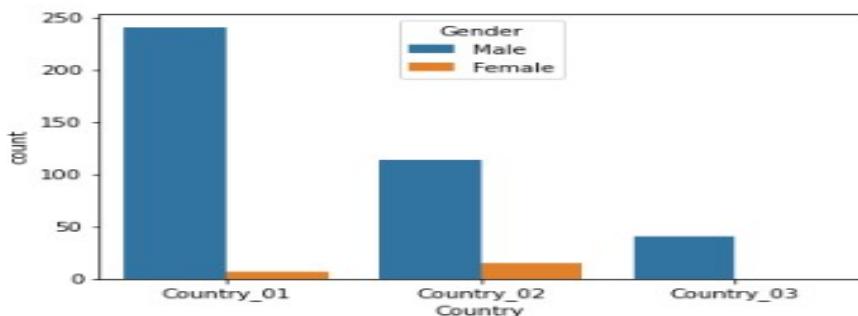
		Gender	Female	Male
		Potential Accident Level		
	I	0	45	
	II	14	81	
	III	3	103	
	IV	4	137	
	V	1	29	
	VI	0	1	

Observation –

From the above, it can be determined that most of the potential level accidents happened to male compared to female, of which Potential Accident Level of IV is dominant.

## 4.3 Gender vs Country:

Count Plot:



Cross table Analysis:

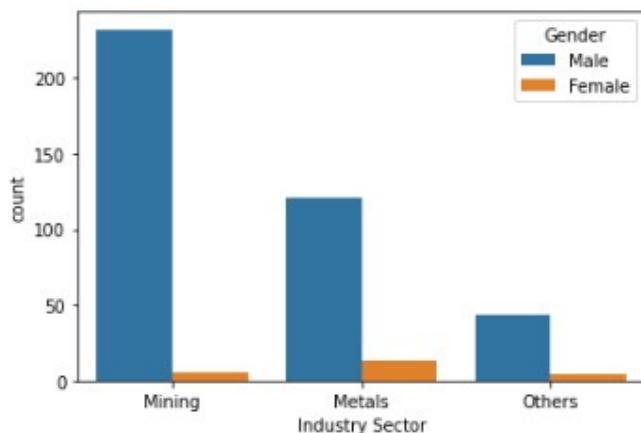
Gender	Female	Male
Country		
Country_01	7	241
Country_02	15	114
Country_03	0	41

Observation –

From the above countplot, it can be determined that the maximum number of accidents took place in country\_01 to males and they are about 241.

#### 4.4 Gender vs Industry Sector:

Count Plot:



Cross table Analysis:

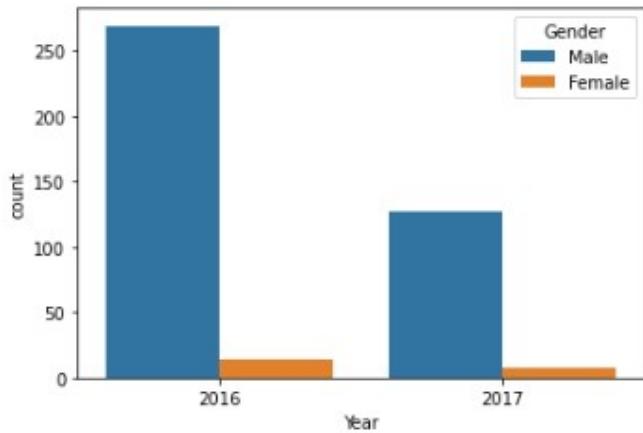
Gender	Female	Male
Industry Sector		
Metals	13	121
Mining	5	232
Others	4	43

Observation –

From the above plots, it is evident that most of the accidents happened to Male in the mining sector, around 232.

## 4.5 Gender vs Year

Count Plot:



Cross table Analysis:

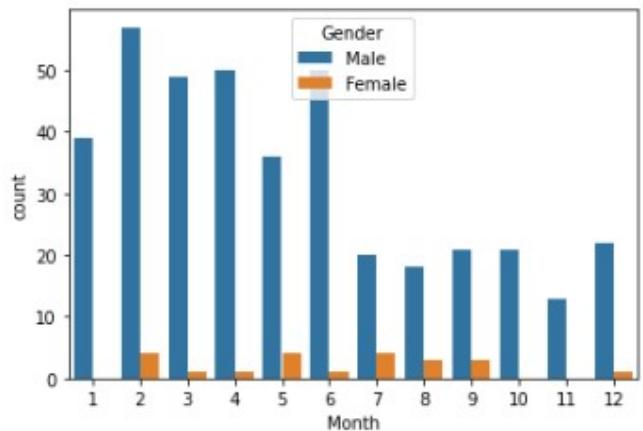
Gender	Female	Male
Year		
2016	14	269
2017	8	127

Observation –

From the above countplot, it is clearly evident that maximum accidents took place in 2016 to the male when compared to female with a count of 269.

## 4.6 Gender vs month:

Count Plot:



## Cross table Analysis:

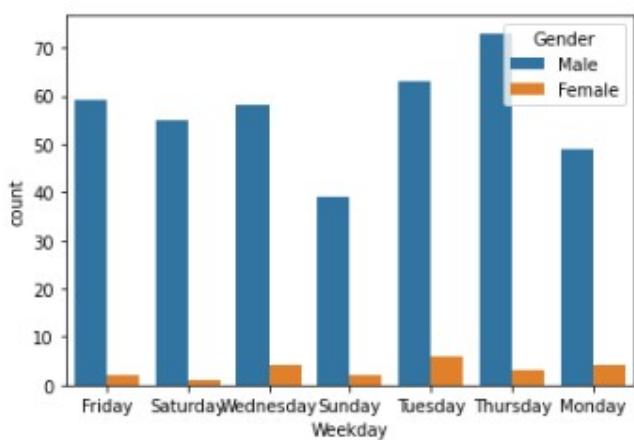
Gender	Female	Male
Month		
1	0	39
2	4	57
3	1	49
4	1	50
5	4	36
6	1	50
7	4	20
8	3	18
9	3	21
10	0	21
11	0	13
12	1	22

Observation –

From the above count plot, it is determined that maximum number of accidents happened to male in the month feb with a count 57.

## 4.7 Gender vs weekday:

Count Plot:



Cross table Analysis:

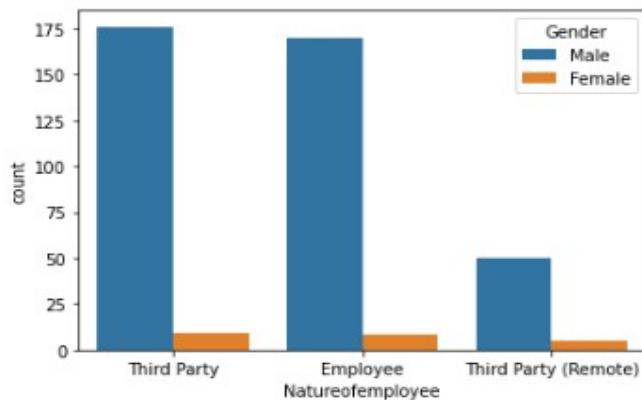
Gender	Female	Male
Weekday		
Friday	2	59
Monday	4	49
Saturday	1	55
Sunday	2	39
Thursday	3	73
Tuesday	6	63
Wednesday	4	58

Observation –

Max accidents happened to male on Thursday with a count of more than 73

#### 4.8 Gender vs Nature of Employee:

Count Plot:



Cross table Analysis:

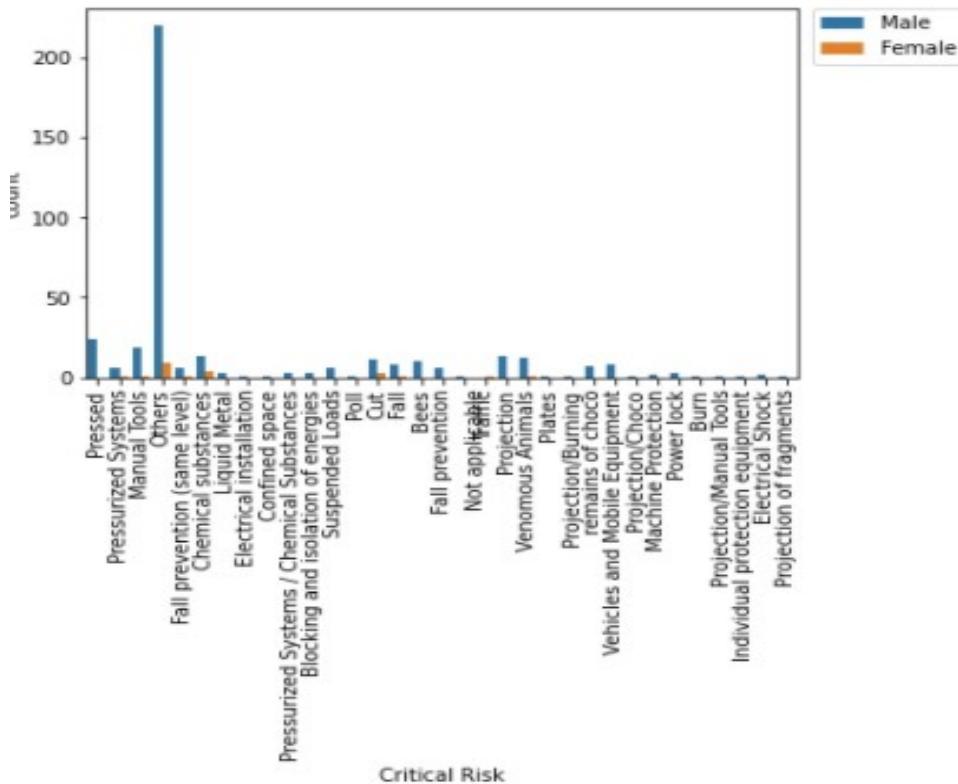
Gender	Female	Male
Nature of employee		
Employee	8	170
Third Party	9	176
Third Party (Remote)	5	50

Observation –

From the above output, it is clearly evident that maximum accidents happened to third party male employees. i.e- 176.

## 4.9 Gender vs Critical Risk:

Count Plot:



Cross table Analysis:

Critical Risk	Female	Male
InNot applicable	0	1
Bees	0	10
Blocking and isolation of energies	0	3
Burn	0	1
Chemical substances	4	13
Confined space	0	1
Cut	3	11
Electrical Shock	0	2
Electrical installation	0	1
Fall	1	8
Fall prevention	0	6
Fall prevention (same level)	1	6
Individual protection equipment	0	1
Liquid Metal	0	3
Machine Protection	0	2
Manual Tools	1	19
Others	9	220

Plates	0	1
Poll	0	1
Power lock	0	3
Pressed	0	24
Pressurized Systems	1	6
Pressurized Systems / Chemical Substances	0	3
Projection	0	13
Projection of fragments	0	1
Projection/Burning	0	1
Projection/Choco	0	1
Projection/Manual Tools	0	1
Suspended Loads	0	6
Traffic	1	0
Vehicles and Mobile Equipment	0	8
Venomous Animals	1	12
remains of choco	0	7

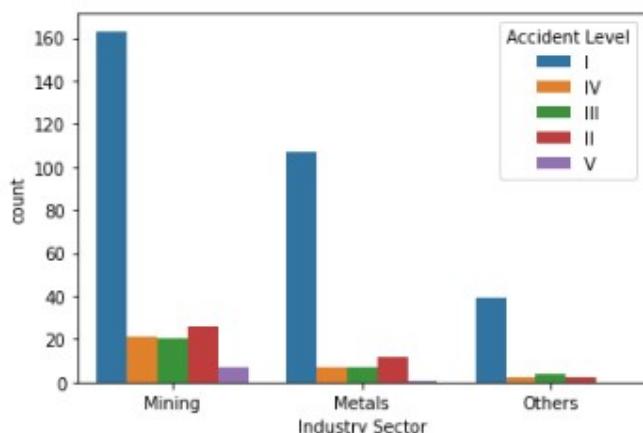
Observation –

Critical Risk of type "Others" is dominant across both Male and Female Genders.

## 2 Industry sector vs RestAll:

### 5.1 Industry Sector Vs Accident Level:

Count Plot:



Cross table Analysis:

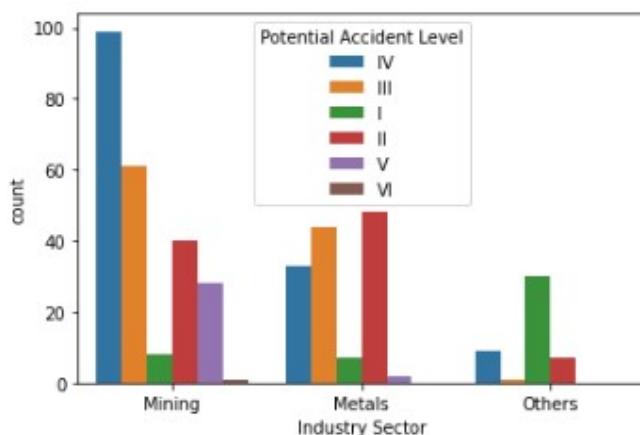
Accident Level	I	II	III	IV	V
Industry Sector					
Metals	107	12	7	7	1
Mining	163	26	20	21	7
Others	39	2	4	2	0

Observation –

Maximum number of accidents happened in the mining sector with accident Level I. i.e- 163.

## 5.2 Industry sector vs potential accident level

Count Plot:



Cross table Analysis:

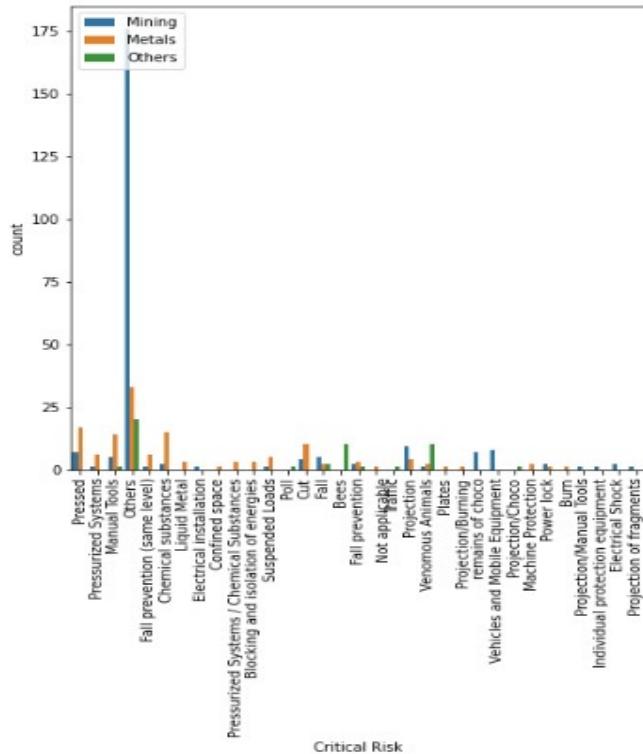
Potential Accident Level	I	II	III	IV	V	VI
Industry Sector						
Metals	7	48	44	33	2	0
Mining	8	40	61	99	28	1
Others	30	7	1	9	0	0

Observation –

Maximum number of accidents happened in the potential accident level 4 and mining sector with a count 99. Minimum number of accidents took place in the mining sector at a potential accident level 6.

## 5.3 Industry Sector vs Critical Risk¶

Count Plot:



Cross table Analysis:

Industry Sector	Metals	Mining	Others
Critical Risk			
\nNot applicable	1	0	0
Bees	0	0	10
Blocking and isolation of energies	3	0	0
Burn	1	0	0
Chemical substances	15	2	0
Confined space	1	0	0
Cut	10	4	0
Electrical Shock	0	2	0
Electrical installation	0	1	0
Fall	2	5	2
Fall prevention	3	2	1
Fall prevention (same level)	6	1	0
Individual protection equipment	0	1	0
Liquid Metal	3	0	0
Machine Protection	2	0	0
Manual Tools	14	5	1
Others	33	176	20
Plates	1	0	0
Poll	0	0	1
Power lock	1	2	0
Pressed	17	7	0
Pressured Systems	6	1	0
Pressured Systems / Chemical Substances	3	0	0
Projection	4	9	0

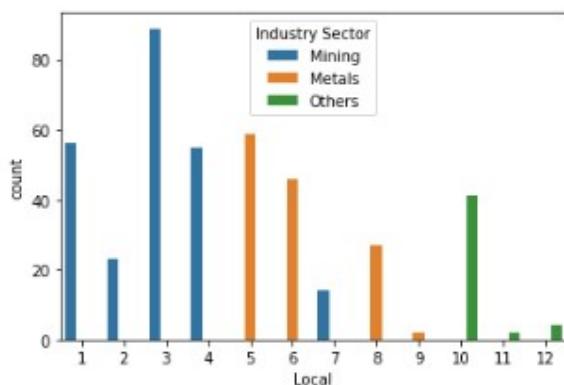
Projection of fragments	0	1	0
Projection/Burning	1	0	0
Projection/Choco	0	0	1
Projection/Manual Tools	0	1	0
Suspended Loads	5	1	0
Traffic	0	0	1
Vehicles and Mobile Equipment	0	8	0
Venomous Animals	2	1	10
remains of choco	0	7	0

Observation –

From the above count plot, it is evident that maximum number of accidents happened in mining with a critical risk of others. i.e about 175.

#### 5.4 Industry sector vs Local:

Count Plot:



Cross table Analysis:

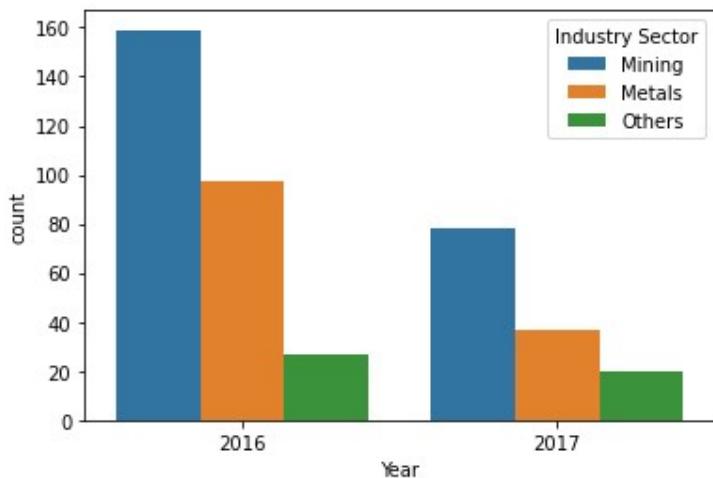
Industry Sector	Local	1	2	3	4	5	6	7	8	9	10	11	12
Metals	0	0	0	0	59	46	0	27	2	0	0	0	0
Mining	56	23	89	55	0	0	14	0	0	0	0	0	0
Others	0	0	0	0	0	0	0	0	0	41	2	4	0

Observation –

Many accidents happened with a local 3 and industrial sector mining. i.e- more than 80. Least accidents took place with local 11 and industrial sector others.

## 5.5 Industry Sector Vs Year:

Count Plot:



Cross table Analysis:

Industry Sector	Year	2016	2017
Metals		97	37
Mining		159	78
Others		27	20

Observation –

From the above plot, the following could be determined

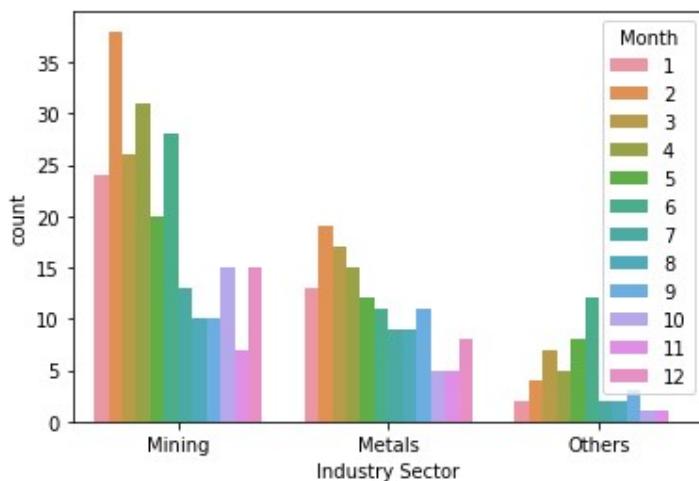
- The number of accidents taken place in year 2016 for mining sector is 160.
- The number of accidents taken place in year 2016 wrt metals sector is about 100.
- The number of accidents taken place in the year 2016 wrt others sector is about 30. Hence, it can be determined that maximum accidents took place in mining sector in the year 2016.
- The number of accidents taken place in the year 2017 wrt mining sector is 80.
- The number of accidents taken place in the year 2017 wrt metals sector is about 40.

- The number of accidents taken place in the year 2017 wrt others sector is 20.

Hence, it can be determined that max accidents took place in mining sector in the year 2017.

## 5.6 Industry Sector Vs Month:

Count Plot:



Cross table Analysis:

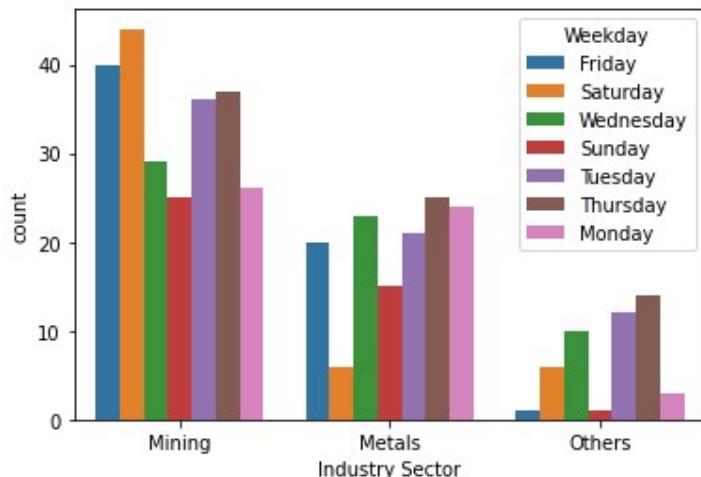
Month	1	2	3	4	5	6	7	8	9	10	11	12
Industry Sector												
Metals	13	19	17	15	12	11	9	9	11	5	5	8
Mining	24	38	26	31	20	28	13	10	10	15	7	15
Others	2	4	7	5	8	12	2	2	3	1	1	0

Observation –

Maximum number of accidents happened in the month feb and mining sector. The least number of accidents took place in the others sector and month december..

## 5.7 Industry sector vs weekday:

Count Plot:



Cross table Analysis:

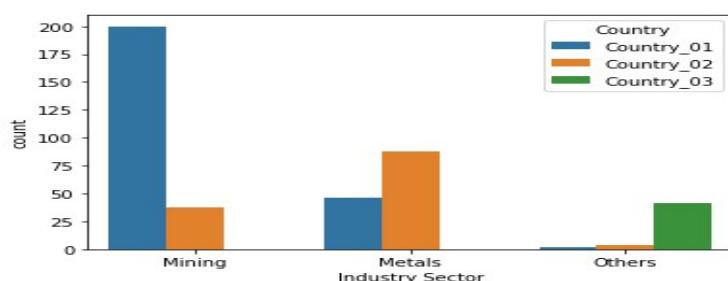
	Weekday	Friday	Monday	Saturday	Sunday	Thursday	Tuesday	Wednesday
Industry Sector								
Metals		20	24	6	15	25	21	23
Mining		40	26	44	25	37	36	29
Others		1	3	6	1	14	12	10

Observation –

Maximum number of accidents happened on the day saturday in the mining sector. i.e- more than 40. The least number of accidents happened on the day sunday in the others sector.

## 5.8 Industry sector vs Country:

Count Plot:



Cross table Analysis:

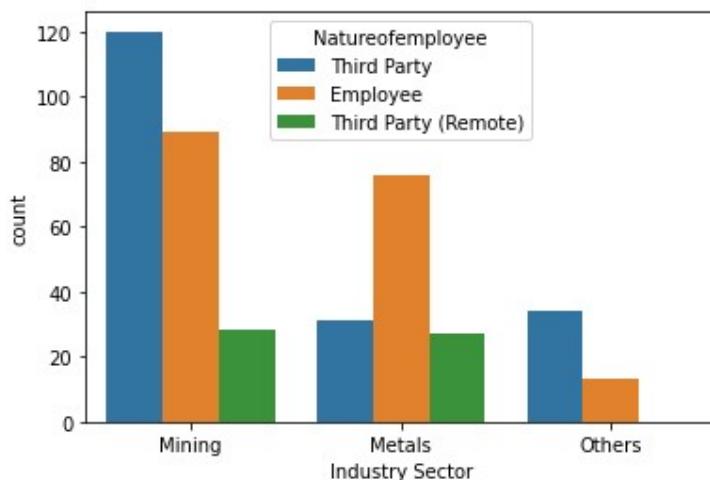
	Country	Country_01	Country_02	Country_03
Industry Sector				
Metals	46	88	0	
Mining	200	37	0	
Others	2	4	41	

Observation –

From the above count plot, it is evident that the maximum number of accidents took place in country\_01 and mining sector.i.e- 200. The least number of accidents took place in country \_01 and others sector.

## 5.9 Industry Sector Vs nature of employee:

Count Plot:



Cross table Analysis:

	Natureofemployee	Employee	Third Party	Third Party (Remote)
Industry Sector				
Metals	76	31	27	
Mining	89	120	28	
Others	13	34	0	

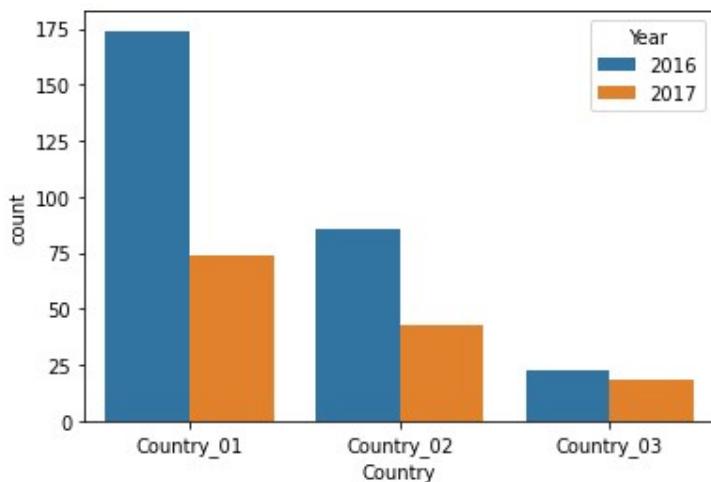
Observation –

From the above count plot, it is clearly evident that the maximum accidents took place in the mining sector with the third party employee type. i.e- about 120. The least number of accidents took place in the others sectors with the nature of employee as employee.

### 3 Country vs RestAll:

#### 6.1 Country vs Year

Count Plot:



Cross table Analysis:

Year	2016	2017
Country		
Country_01	174	74
Country_02	86	43
Country_03	23	18

Observation –

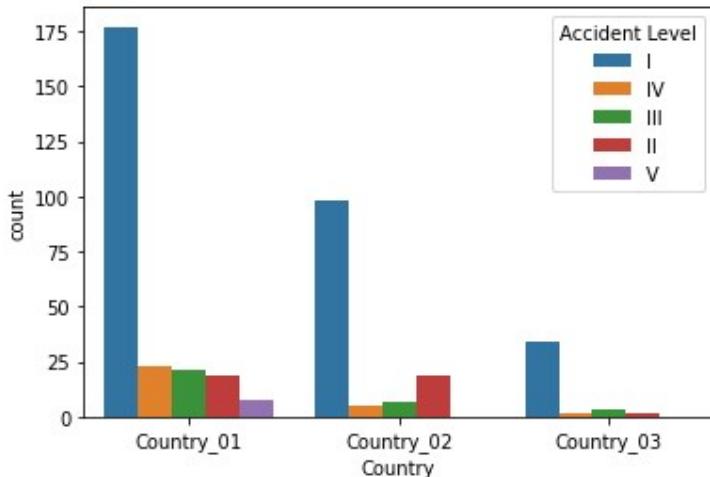
From the above output, the following can be determined-

- The number of accidents taken place in country\_01 and year 2016 is 174.
- The number of accidents taken place in country\_01 and year 2017 is about 74.

- The number of accidents taken place in country\_02 and year 2016 is more than 86.
- The number of accidents taken place in country\_02 and year 2017 is about 43.
- The number of accidents taken place in country\_03 and year 2016 is about 23.
- The number of accidents taken place in country\_03 and year 2017 is about 18.

## 6.2 Country Vs accident level:

Count Plot:



Cross table Analysis:

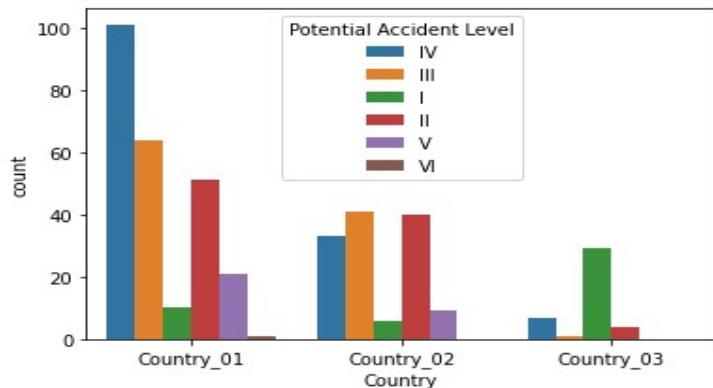
Accident Level	I	II	III	IV	V
Country					
Country_01	177	19	21	23	8
Country_02	98	19	7	5	0
Country_03	34	2	3	2	0

Observation –

From the above count plot, it is clearly evident that the maximum number of accidents took place in accident level 1 and country\_01.

### 6.3 Country Vs Potential Accident Level:

Count Plot:



Cross table Analysis:

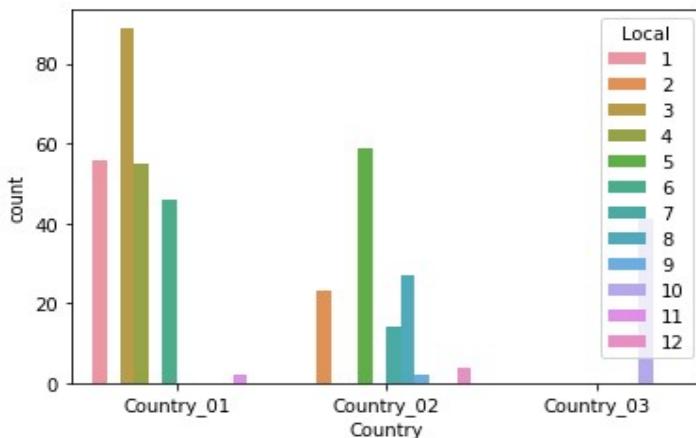
Potential Accident Level	I	II	III	IV	V	VI
Country						
Country_01	10	51	64	101	21	1
Country_02	6	40	41	33	9	0
Country_03	29	4	1	7	0	0

Observation –

From the above plot, it is evident that the maximum accidents occurred in country\_01 and potential accident level 3

### 6.4 Country Vs Local:

Count Plot:



Cross table Analysis:

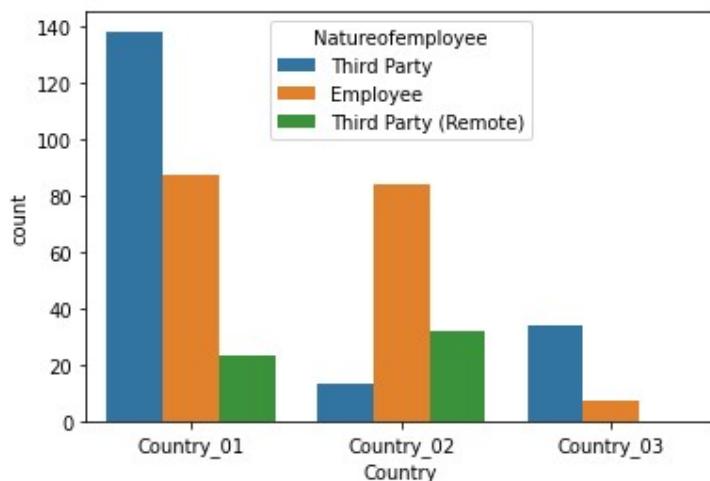
Local	1	2	3	4	5	6	7	8	9	10	11	12
Country												
Country_01	56	0	89	55	0	46	0	0	0	0	2	0
Country_02	0	23	0	0	59	0	14	27	2	0	0	4
Country_03	0	0	0	0	0	0	0	0	0	41	0	0

Observation –

Country 1 is more dominant in local 3 region and least dominant in Local 12.

## 6.5 Country Vs Nature Of Employee

Count Plot:



Cross table Analysis:

Natureofemployee	Employee	Third Party	Third Party (Remote)
Country			
Country_01	87	138	23
Country_02	84	13	32
Country_03	7	34	0

Observation –

Accidents in Country 01 is more dominant in Third Party type of employee, country 03 is least dominant in Third Party (Remote).

## 6.6 Country Vs Critical Risk:

Cross table Analysis:

	Country	Country 01	Country 02	Country 03
	Critical Risk			
\nNot applicable	0	1	0	
Bees	0	0	10	
Blocking and isolation of energies	1	2	0	
Burn	0	1	0	
Chemical substances	4	13	0	
Confined space	0	1	0	
Cut	5	9	0	
Electrical Shock	2	0	0	
Electrical installation	1	0	0	
Fall	6	1	2	
Fall prevention	3	2	1	
Fall prevention (same level)	5	2	0	
Individual protection equipment	1	0	0	
Liquid Metal	0	3	0	
Machine Protection	1	1	0	
Manual Tools	7	12	1	
Others	169	45	15	
Plates	1	0	0	
Poll	0	0	1	
Power lock	3	0	0	
Pressed	9	15	0	
Pressurized Systems	1	6	0	
Pressurized Systems / Chemical Substances	2	1	0	
Projection	9	4	0	
Projection of fragments	1	0	0	
Projection/Burning	0	1	0	
Projection/Choco	0	0	1	
Projection/Manual Tools	1	0	0	
Suspended Loads	3	3	0	
Traffic	0	1	0	
Vehicles and Mobile Equipment	7	1	0	
Venomous Animals	0	3	10	
remains of choco	6	1	0	

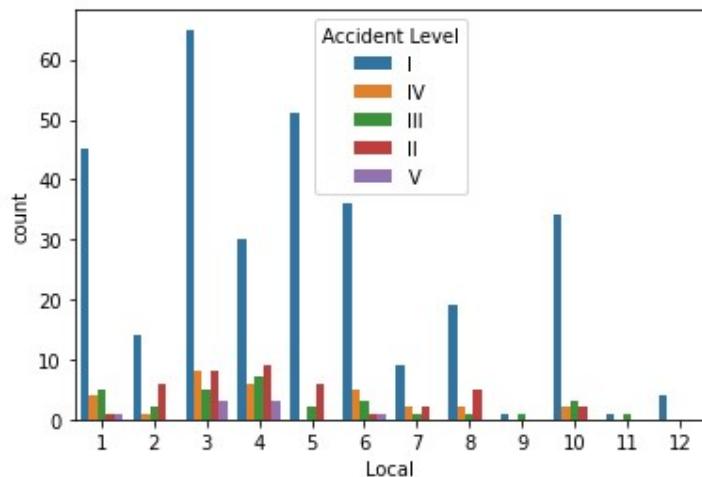
Observation –

Country 01 is more dominant in Others Critical Risk and Critical Risk is least dominant in Country 03.

## 4 Local Vs Rest All

### 7.1 Local Vs Accident Level

Count Plot:



Cross table Analysis:

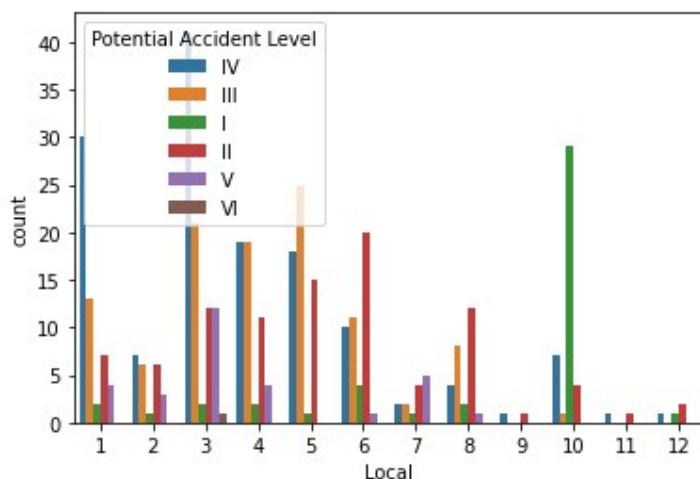
Local	1	2	3	4	5	6	7	8	9	10	11	12
Accident Level												
I	45	14	65	30	51	36	9	19	1	34	1	4
II	1	6	8	9	6	1	2	5	0	2	0	0
III	5	2	5	7	2	3	1	1	1	3	1	0
IV	4	1	8	6	0	5	2	2	0	2	0	0
V	1	0	3	3	0	1	0	0	0	0	0	0

Observation –

Accident level 1 is more dominant in Local 2 region with 65 accidents, while Accident Level V is least across all Locals

## 7.2 Local Vs Potential Accident Level:

Count Plot:



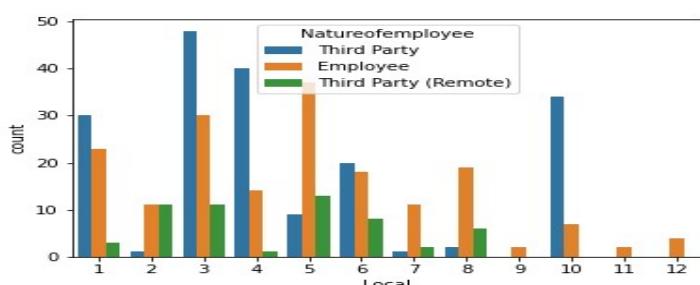
Cross table Analysis:

		Local	1	2	3	4	5	6	7	8	9	10	11	12
		Potential Accident Level	I	II	III	IV	V	VI						
	I	2	1	2	2	1	4	1	2	0	29	0	1	
	II	7	6	12	11	15	20	4	12	1	4	1	2	
	III	13	6	21	19	25	11	2	8	0	1	0	0	
	IV	30	7	41	19	18	10	2	4	1	7	1	1	
	V	4	3	12	4	0	1	5	1	0	0	0	0	
	VI	0	0	1	0	0	0	0	0	0	0	0	0	

Observation: Overall Local 3 is more prone to Multiple potential accidents, while local 12 is the least.

## 7.3 Local Vs Natureofemployee:

Count Plot:



## Cross table Analysis:

Local	1	2	3	4	5	6	7	8	9	10	11	12
Nature of employee												
Employee	23	11	30	14	37	18	11	19	2	7	2	4
Third Party	30	1	48	40	9	20	1	2	0	34	0	0
Third Party (Remote)	3	11	11	1	13	8	2	6	0	0	0	0

## Observation –

Type Employee is more dominant across all Locals, while Type Third Party(Remote) is least dominant across all Locals.

## 7.4 Local Vs Critical Risk:

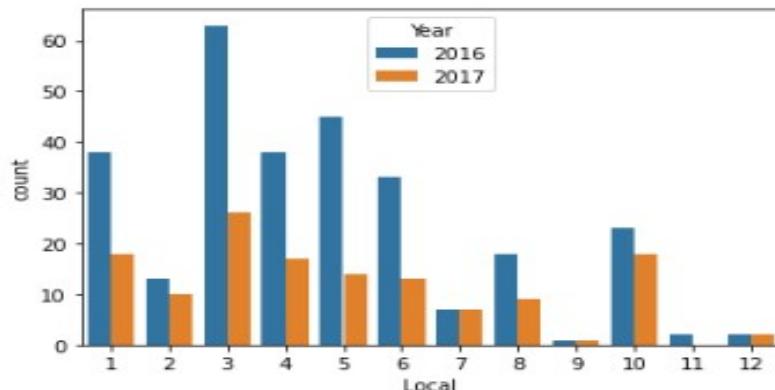
### Cross table Analysis:

Critical Risk	Local	1	2	3	4	5	6	7	8	9	10	11	12
\nNot applicable	0	0	0	0	0	0	0	1	0	0	0	0	0
Bees	0	0	0	0	0	0	0	0	0	10	0	0	0
Blocking and isolation of energies	0	0	0	0	1	1	0	1	0	0	0	0	0
Burn	0	0	0	0	1	0	0	0	0	0	0	0	0
Chemical substances	1	1	0	0	11	3	0	1	0	0	0	0	0
Confined space	0	0	0	0	1	0	0	0	0	0	0	0	0
Cut	0	0	2	1	7	2	1	1	0	0	0	0	0
Electrical Shock	0	0	2	0	0	0	0	0	0	0	0	0	0
Electrical installation	1	0	0	0	0	0	0	0	0	0	0	0	0
Fall	2	0	1	2	1	1	0	0	0	2	0	0	0
Fall prevention	0	0	1	0	1	2	1	0	0	1	0	0	0
Fall prevention (same level)	0	0	0	1	2	4	0	0	0	0	0	0	0
Individual protection equipment	0	0	1	0	0	0	0	0	0	0	0	0	0
Liquid Metal	0	0	0	0	2	0	0	1	0	0	0	0	0
Machine Protection	0	0	0	0	1	1	0	0	0	0	0	0	0
Manual Tools	1	2	1	1	5	4	0	4	1	1	0	0	0
Others	41	16	68	43	10	15	8	8	0	15	2	3	0
Plates	0	0	0	0	0	1	0	0	0	0	0	0	0
Poll	0	0	0	0	0	0	0	0	0	1	0	0	0
Power lock	0	0	2	0	0	1	0	0	0	0	0	0	0
Pressed	2	2	1	2	6	4	0	7	0	0	0	0	0
Pressurized Systems	0	1	0	0	3	1	0	2	0	0	0	0	0
Pressurized Systems / Chemical Substances	0	0	0	0	1	2	0	0	0	0	0	0	0
Projection	3	0	4	0	1	2	2	1	0	0	0	0	0
Projection of fragments	0	0	0	1	0	0	0	0	0	0	0	0	0
Projection/Burning	0	0	0	0	1	0	0	0	0	0	0	0	0
Projection/Choco	0	0	0	0	0	0	0	0	0	1	0	0	0
Projection/Manual Tools	0	0	0	1	0	0	0	0	0	0	0	0	0
Suspended Loads	1	0	0	0	3	2	0	0	0	0	0	0	0
Traffic	0	0	0	0	0	0	0	0	0	0	0	1	0
Vehicles and Mobile Equipment	4	1	2	1	0	0	0	0	0	0	0	0	0
Venomous Animals	0	0	0	0	1	0	1	0	1	10	0	0	0

Observation: Critical Risk of type "Others" is dominant across all Locals

## 7.5 Local Vs Year

Count Plot:



Cross table Analysis:

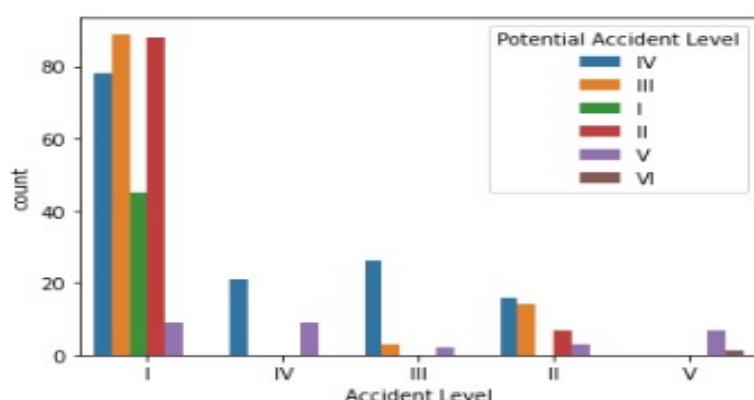
Local	1	2	3	4	5	6	7	8	9	10	11	12
Year												
2016	38	13	63	38	45	33	7	18	1	23	2	2
2017	18	10	26	17	14	13	7	9	1	18	0	2

Observation: Year 2016 has more accidents across all Local regions compared to 2017.

## 5 Accident Level Vs Rest All:

### 8.1 Accident Level Vs Potential Accident Level:

Count Plot:



## Cross table Analysis:

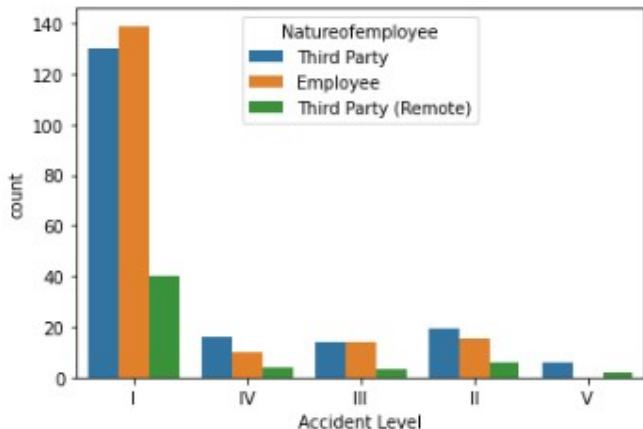
Accident Level	I	II	III	IV	V
Potential Accident Level					
I	45	0	0	0	0
II	88	7	0	0	0
III	89	14	3	0	0
IV	78	16	26	21	0
V	9	3	2	9	7
VI	0	0	0	0	1

Observation –

Accident Level I is more related to Potential Accident levels of I, II, III, IV, V, VI.

## 8.2 Accident Level Vs Natureofemployee:

Count Plot:



Cross table Analysis:

Natureofemployee	Employee	Third Party	Third Party (Remote)
Accident Level			
I	139	130	40
II	15	19	6
III	14	14	3
IV	10	16	4
V	0	6	2

Observation: Accident Level I is more dominant across all Employee types, where Level V is least across all types.

### 8.3 Accident Level Vs Critical Risk:

Cross table Analysis:

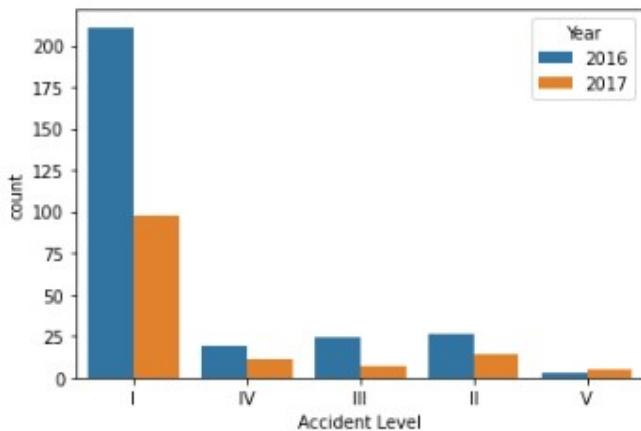
Critical Risk	Accident Level	I	II	III	IV	V
	\nNot applicable	0	0	0	1	0
Bees	Bees	10	0	0	0	0
Blocking and isolation of energies	Blocking and isolation of energies	3	0	0	0	0
Burn	Burn	0	0	1	0	0
Chemical substances	Chemical substances	15	2	0	0	0
Confined space	Confined space	1	0	0	0	0
Cut	Cut	11	2	1	0	0
Electrical Shock	Electrical Shock	2	0	0	0	0
Electrical installation	Electrical installation	0	0	0	1	0
Fall	Fall	6	0	0	2	1
Fall prevention	Fall prevention	5	0	0	1	0
Fall prevention (same level)	Fall prevention (same level)	6	0	0	1	0
Individual protection equipment	Individual protection equipment	0	1	0	0	0
Liquid Metal	Liquid Metal	3	0	0	0	0
Machine Protection	Machine Protection	2	0	0	0	0
Manual Tools	Manual Tools	12	5	3	0	0
Others	Others	169	21	23	13	3
Plates	Plates	1	0	0	0	0
Poll	Poll	0	0	0	1	0
Power lock	Power lock	0	0	0	1	2
Pressed	Pressed	17	1	2	4	0
Pressurized Systems	Pressurized Systems	6	1	0	0	0
Pressurized Systems / Chemical Substances	Pressurized Systems / Chemical Substances	2	1	0	0	0
Projection	Projection	10	2	0	1	0
Projection of fragments	Projection of fragments	1	0	0	0	0
Projection/Burning	Projection/Burning	0	1	0	0	0
Projection/Choco	Projection/Choco	1	0	0	0	0
Projection/Manual Tools	Projection/Manual Tools	1	0	0	0	0
Suspended Loads	Suspended Loads	4	0	1	1	0
Traffic	Traffic	1	0	0	0	0
Vehicles and Mobile Equipment	Vehicles and Mobile Equipment	5	1	0	1	1
Venomous Animals	Venomous Animals	13	0	0	0	0

Observation –

Accident Level I is more dominant with Other critical Risk type

## 8.4 Accident Level Vs Year :

Count Plot:



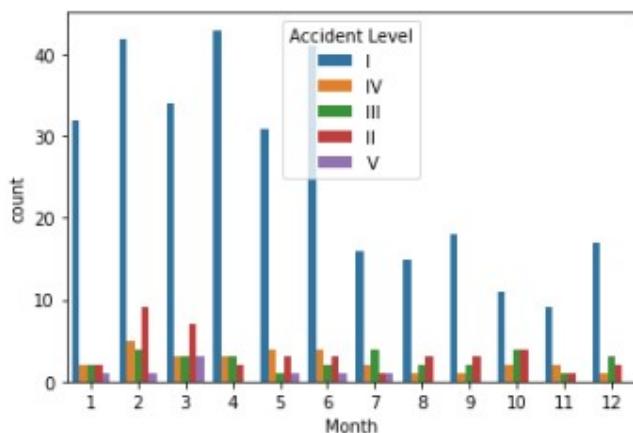
Cross table Analysis:

Accident Level	I	II	III	IV	V
Year					
2016	211	26	24	19	3
2017	98	14	7	11	5

Observation: Accident Level I is more dominant in across 2016 and 2017 years, and Level V is minimum.

## 8.5 Accident Level Vs Month:

Count Plot:



### Cross table Analysis:

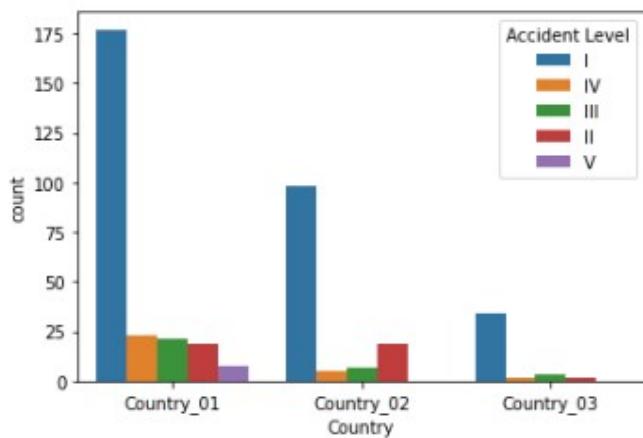
Month	1	2	3	4	5	6	7	8	9	10	11	12
Accident Level												
I	32	42	34	43	31	41	16	15	18	11	9	17
II	2	9	7	2	3	3	1	3	3	4	1	2
III	2	4	3	3	1	2	4	2	2	4	1	3
IV	2	5	3	3	4	4	2	1	1	2	2	1
V	1	1	3	0	1	1	1	0	0	0	0	0

Observation –

Accident Level 1 dominates across all Months while Level V is minimum.

### 8.6 Accident Level Vs Country:

Count Plot:



### Cross table Analysis:

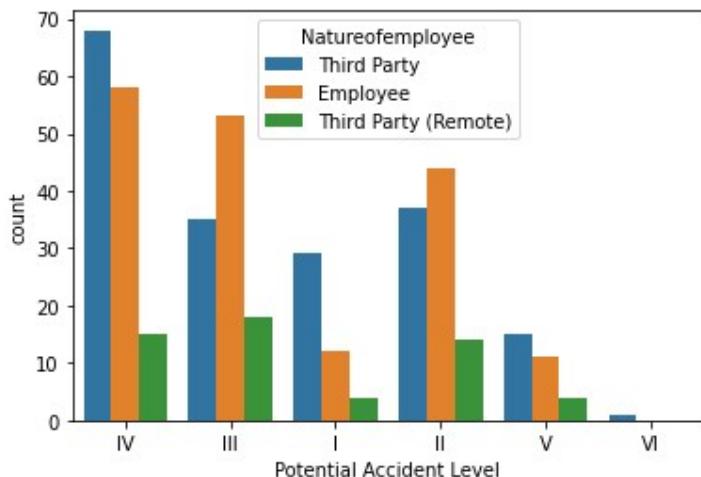
Country	Country_01	Country_02	Country_03
Accident Level			
I	177	98	34
II	19	19	2
III	21	7	3
IV	23	5	2
V	8	0	0

Observation: Accident Level I is more dominant across all Countries, while Accident Level V is least dominant across all countries.

## 6 Potential Accident Level Vs Rest All:

### 9.1 Potential Accident Level Vs Natureofemployee:

Count Plot:



Cross table Analysis:

	Natureofemployee	Employee	Third Party	Third Party (Remote)
Potential Accident Level				
I		12	29	4
II		44	37	14
III		53	35	18
IV		58	68	15
V		11	15	4
VI		0	1	0

Observation:

Potential Accident level IV dominants in ThirdParty, while VI is least dominant in Third Party(Remote) across all

## 9.2 Potential Accident Level Vs Critical Risk:

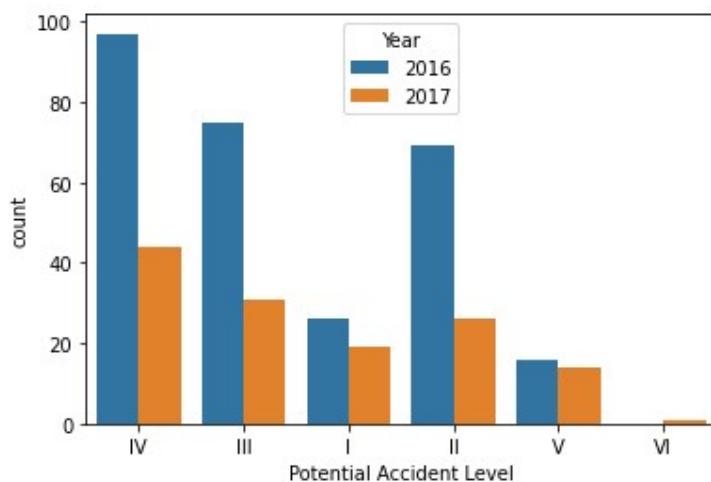
Cross table Analysis:

Potential Accident Level	I	II	III	IV	V	VI
Critical Risk						
\nNot applicable	0	0	0	0	1	0
Bees	10	0	0	0	0	0
Blocking and isolation of energies	0	1	2	0	0	0
Burn	0	0	0	1	0	0
Chemical substances	0	5	8	4	0	0
Confined space	0	0	1	0	0	0
Cut	1	6	5	2	0	0
Electrical Shock	0	0	0	2	0	0
Electrical installation	0	0	0	0	1	0
Fall	1	1	4	2	1	0
Fall prevention	1	0	0	5	0	0
Fall prevention (same level)	1	1	3	2	0	0
Individual protection equipment	0	0	0	1	0	0
Liquid Metal	1	0	0	2	0	0
Machine Protection	0	0	2	0	0	0
Manual Tools	2	5	9	4	0	0
Others	16	60	53	85	15	0
Plates	0	1	0	0	0	0
Others	16	60	53	85	15	0
Plates	0	1	0	0	0	0
Poll	0	0	0	1	0	0
Power lock	0	0	0	0	3	0
Pressed	2	5	9	7	1	0
Pressurized Systems	0	2	3	2	0	0
Pressurized Systems / Chemical Substances	0	1	0	2	0	0
Projection	0	2	2	7	2	0
Projection of fragments	0	0	0	1	0	0
Projection/Burning	0	0	0	1	0	0
Projection/Choco	0	1	0	0	0	0
Projection/Manual Tools	0	0	1	0	0	0
Suspended Loads	0	1	0	5	0	0
Traffic	0	1	0	0	0	0
Vehicles and Mobile Equipment	0	0	2	2	4	0
Venomous Animals	10	2	1	0	0	0
remains of choco	0	0	1	3	2	1

Observation: Among all Critical Risk with Type as "Others" is dominant across all Potential Accident Levels.

### 9.3 Potential Accident Level Vs Year:

Count Plot:



Cross table Analysis:

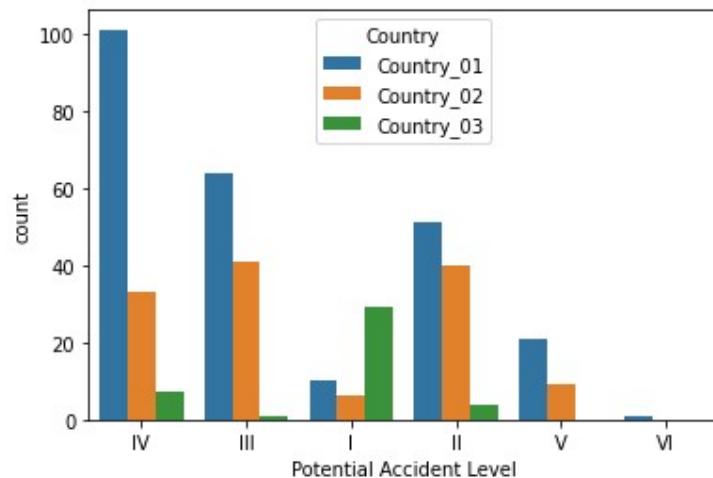
	Year	2016	2017
Potential Accident Level			
I		26	19
II		69	26
III		75	31
IV		97	44
V		16	14
VI		0	1

Observation:

There is Decrease in Number of accidents across all Potential Accident level from 2016 to 2017. Potential Accident level IV is dominant in both 2016 and 2017

## 9.4 Potential Accident Level Vs Country:

Count Plot:



Cross table Analysis:

Country	Country_01	Country_02	Country_03
Potential Accident Level			
I	10	6	29
II	51	40	4
III	64	41	1
IV	101	33	7
V	21	9	0
VI	1	0	0

Observation –

Potential Accident Level IV is dominant across all countries, while with VI least number of accidents happened.

## 7 Nature of employee Vs RestAll:

### 10.1 Nature of employee Vs Critical Risk:

Cross table Analysis:

	Nature of employee	Employee	Third Party	Third Party (Remote)
Critical Risk				
\nNot applicable	1	0	0	
Bees	1	9	0	
Blocking and isolation of energies	2	0	1	
Burn	1	0	0	
Chemical substances	9	2	6	
Confined space	1	0	0	
Cut	8	5	1	
Electrical Shock	0	0	2	
Electrical installation	0	1	0	
Fall	0	5	4	
Fall prevention	2	3	1	
Fall prevention (same level)	3	4	0	
Individual protection equipment	0	1	0	
Liquid Metal	3	0	0	
Machine Protection	1	1	0	
Manual Tools	7	7	6	
Others	99	109	21	
Plates	1	0	0	
Poll	0	1	0	
Power lock	0	0	3	
Pressed	12	7	5	
Pressurized Systems	4	1	2	
Pressurized Systems / Chemical Substances	1	1	1	
Projection	7	6	0	
Projection of fragments	0	1	0	
Projection/Burning	1	0	0	
Projection/Choco	0	1	0	
Projection/Manual Tools	0	1	0	
Suspended Loads	4	1	1	
Traffic	1	0	0	
Vehicles and Mobile Equipment	3	5	0	
Venomous Animals	3	9	1	
remains of choco	3	4	0	

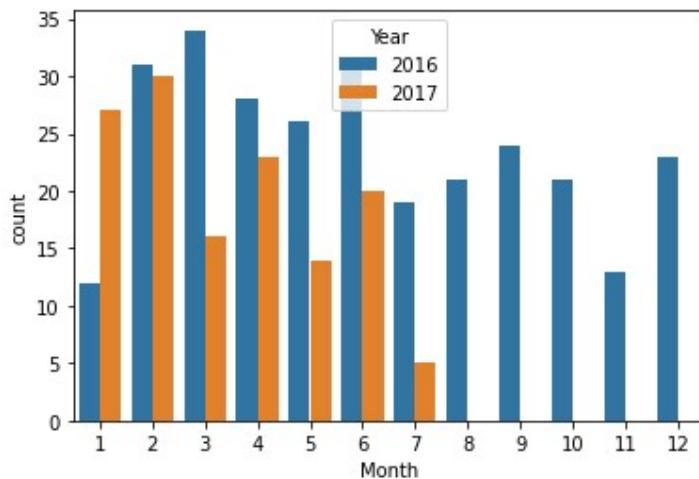
Observation –

Critical Risk of type "Others" is dominant across all Types of Employees.

## 8 Year Vs RestAll:

### 11.1 Year vs month:

Count Plot:



Cross table Analysis:

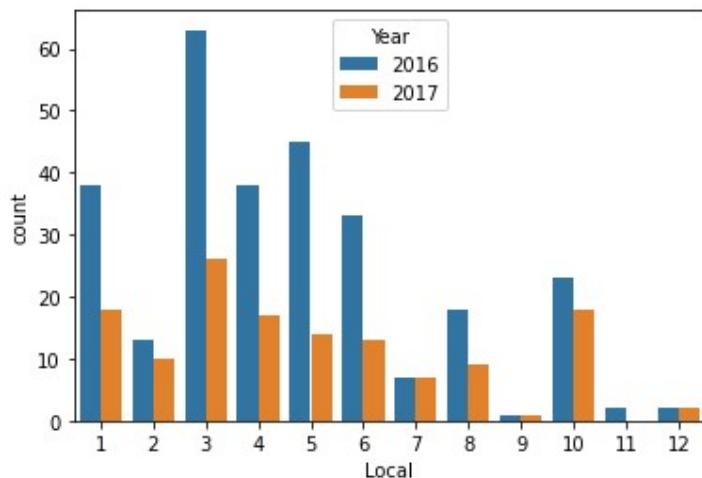
Year	2016	2017
Month		
1	12	27
2	31	30
3	34	16
4	28	23
5	26	14
6	31	20
7	19	5
8	21	0
9	24	0
10	21	0
11	13	0
12	23	0

Observation –

From the above plot, it is evident that the max accidents happened in the year 2016 and march.

## 11.2 Year vs Local:

Count Plot:



Cross table Analysis:

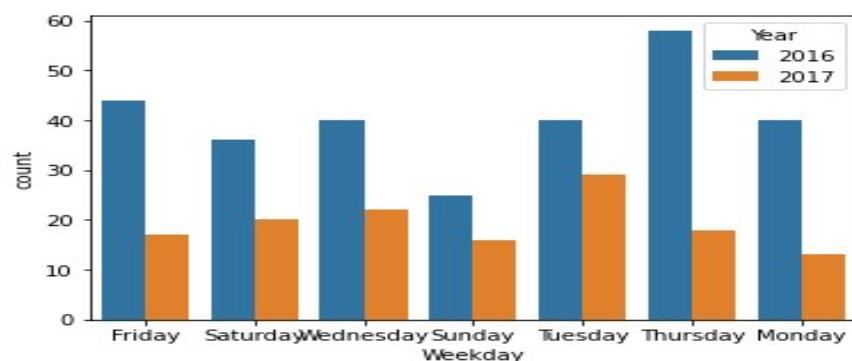
Local	1	2	3	4	5	6	7	8	9	10	11	12
Year												
2016	38	13	63	38	45	33	7	18	1	23	2	2
2017	18	10	26	17	14	13	7	9	1	18	0	2

Observation –

From the above plot, it can be determined that the maximum accidents took place in the local 3 and year 2016.

## 11.3 Year vs Weekday:

Count Plot:



Cross table Analysis:

	Weekday	Friday	Monday	Saturday	Sunday	Thursday	Tuesday	Wednesday
Year								
2016	44	40	36	25	58	40	40	
2017	17	13	20	16	18	29	22	

Observation –

From the above plot, it is clearly evident that maximum number of accidents took place on Thursday and year 2016.

#### 11.4 Year vs Critical Risk:

Cross table Analysis:

	Year	2016	2017
Critical Risk			
\nNot applicable	1	0	
Bees	10	0	
Blocking and isolation of energies	3	0	
Burn	0	1	
Chemical substances	13	4	
Confined space	1	0	
Cut	6	8	
Electrical Shock	0	2	
Electrical installation	1	0	
Fall	2	7	
Fall prevention	1	5	
Fall prevention (same level)	6	1	
Individual protection equipment	0	1	
Liquid Metal	2	1	
Machine Protection	0	2	
Manual Tools	14	6	
Others	189	40	

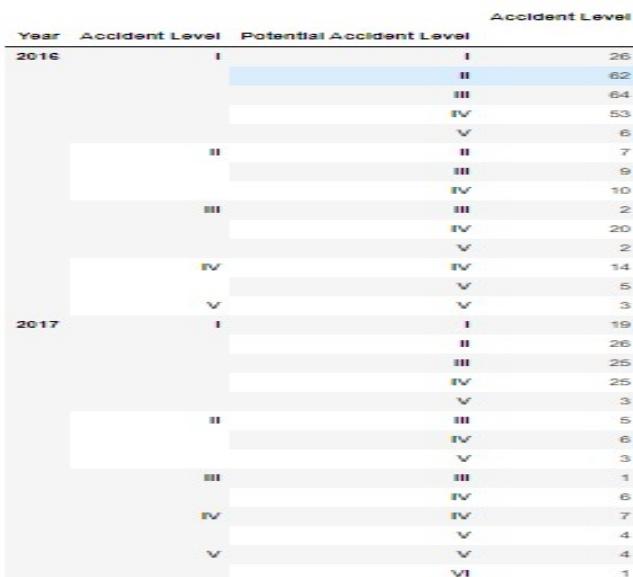
Plates	1	0
Poll	1	0
Power lock	0	3
Pressed	14	10
Pressurized Systems	7	0
Pressurized Systems / Chemical Substances	3	0
Projection	1	12
Projection of fragments	0	1
Projection/Burning	0	1
Projection/Choco	0	1
Projection/Manual Tools	0	1
Suspended Loads	5	1
Traffic	1	0
Vehicles and Mobile Equipment	0	8
Venomous Animals	1	12
remains of choco	0	7

Observation: From the above plot, it is clearly evident that maximum number of accidents took place with "Others" and year 2016.

## Groupby Analysis:

### 1. Year wise distribution of accidents and potential accident levels:

```
data.groupby(['Year','Accident Level','Potential Accident Level'])[['Accident Level']].count()
```



-Year 2016 with Accident Level I has maximum accidents of 64 with Potential Accident Level III and 62 with Potential Accident Level II

- Year 2017 with Accident Level I has maximum accidents of 26 with Potential Accident Level II and 25 with Potential Accident Level III, IV

## 2. Year wise distribution of Industry Sector and accident levels:

```
data.groupby(['Year','Industry Sector','Accident Level'])[['Accident Level']].count()
```

Year	Industry Sector	Accident Level	
		I	II
2016	Metals	I	79
		II	9
		III	4
		IV	5
	Mining	I	112
		II	15
		III	17
		IV	12
		V	3
	Others	I	20
		II	2
		III	3
		IV	2
		V	1
2017	Metals	I	28
		II	3
		III	3
		IV	2
	Mining	I	51
		II	11
		III	3
		IV	9
		V	4
	Others	I	19
		II	1

1. Year 2016 with Industry Sector of Type "Metals" has maximum accidents of 79 with Accident Level I
2. Year 2016 with Industry Sector of Type "Mining" has maximum accidents of 112 with Accident Level I
3. Year 2016 with Industry Sector of Type "Others" has maximum accidents of 20 with Accident Level I
4. Year 2017 with Industry Sector of Type "Metals" has maximum accidents of 28 with Accident Level I
5. Year 2017 with Industry Sector of Type "Mining" has maximum accidents of 51 with Accident Level I
6. Year 2017 with Industry Sector of Type "Others" has maximum accidents of 19 with Accident Level I

### **3. Industry Sector wise distribution of Country and accident levels**

```
data.groupby(['Industry Sector','Country','Accident Level'])[['Accident Level']].count()
```

Industry	Sector	Country	Accident Level	Accident Level	
				I	II
Metals	Country_01		I	36	1
			II	5	1
			III	3	1
			IV	5	1
			V	1	1
	Country_02		I	71	1
			II	11	1
			III	4	1
			IV	2	1
			V	140	1
Mining	Country_01		I	18	1
			II	17	1
			III	18	1
			IV	7	1
			V	23	1
	Country_02		I	8	1
			II	3	1
			III	3	1
			IV	3	1
			V	34	1
Others	Country_01		I	1	1
			II	1	1
			III	4	1
	Country_02		I	2	1
			II	3	1
			III	2	1

1. Metals in Country\_01 has maximum accidents with Level 1 with 36 count
  2. Metals in Country\_02 has maximum accidents with Level 1 with 71 count
  3. Mining in Country\_01 has maximum accidents with Level 1 with 140 count
  4. Mining in Country\_02 has maximum accidents with Level 1 with 23 count
  5. Others in Country\_03 has maximum accidents with Level 1 with 34 count

# **Word Cloud Analysis:**

## ***Accident Level:***

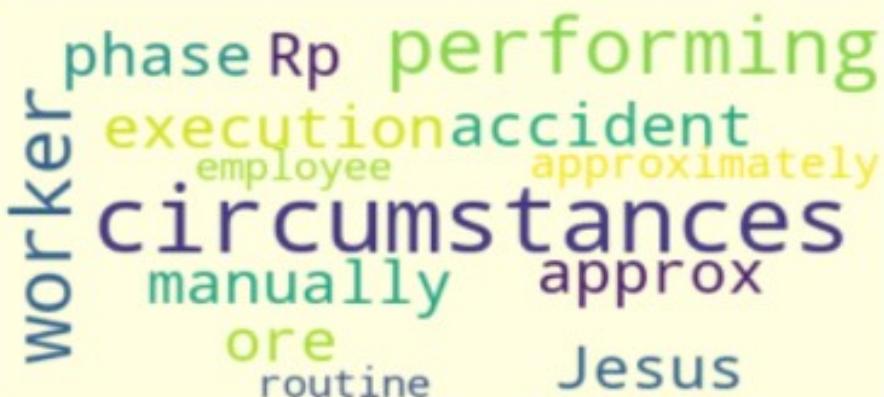
- ## 1. WordCloud for Accident Level : II



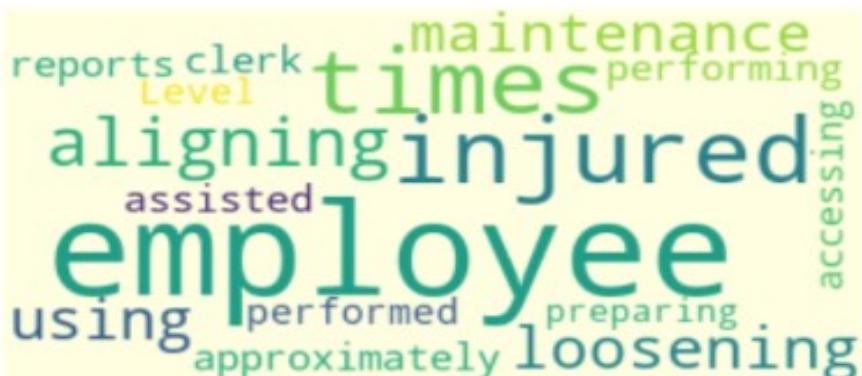
2. WordCloud for Accident Level : IV



3. WordCloud for Accident Level : III



4. WordCloud for Accident Level : II



5. WordCloud for Accident Level : V

approximately  
access  
performing

A word cloud visualization where the words "approximately", "access", and "performing" are the most prominent, displayed in large, bold, purple, green, and blue colors respectively. The background is a light yellow gradient.

**Potential Accident Level:**

1. WordCloud for Potential Accident Level : IV

leaving reports preparing phase plant  
Rp Eriks approx ore operator  
field performing making routine moments  
Jesus auxiliary execution section locomotive maintaining  
auxiliary clerk  
circumstance maintenance approaching  
worker approximately performed tower technician  
removing mine mechanic times  
trying carrying manually aligning completing positioning  
level employee accident entering

A word cloud visualization centered around the word "circumstance". Other prominent words include "reports", "operator", "employee", "worker", "approximately", "performed", "completed", and "times". Smaller words like "auxiliary", "clerk", "mine", "mechanic", "tower", "technician", "aligning", "entering", and "positioning" are also visible.

2. WordCloud for Potential Accident Level : III

conducting approximately Emerson maid discharging manipulating  
employee pit  
times worker clerk injured using  
moment moment loading operator technician moved  
performing installing replacing preparing  
carrying opened moving phase level circumstances reports collaborator

A word cloud visualization centered around the word "employee". Other prominent words include "approximately", "operator", "level", "circumstances", "reports", and "collaborator". Smaller words like "Emerson", "maid", "discharging", "manipulating", "pit", "times", "worker", "clerk", "injured", "using", "moment", "loading", "operator", "technician", "moved", "moving", "phase", "installing", "replacing", "prepared", "opened", "carrying", and "moving" are also present.

3. WordCloud for Potential Accident Level : I



4. WordCloud for Potential Accident Level : II

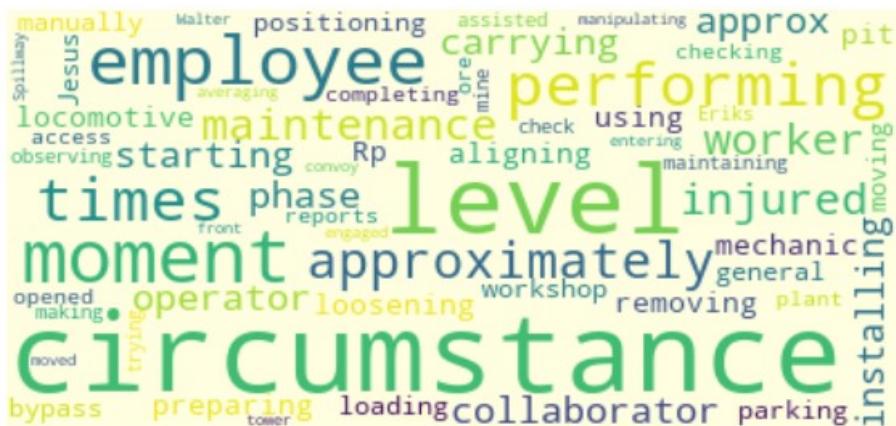


5. WordCloud for Potential Accident Level : V



## *Industry Sector*

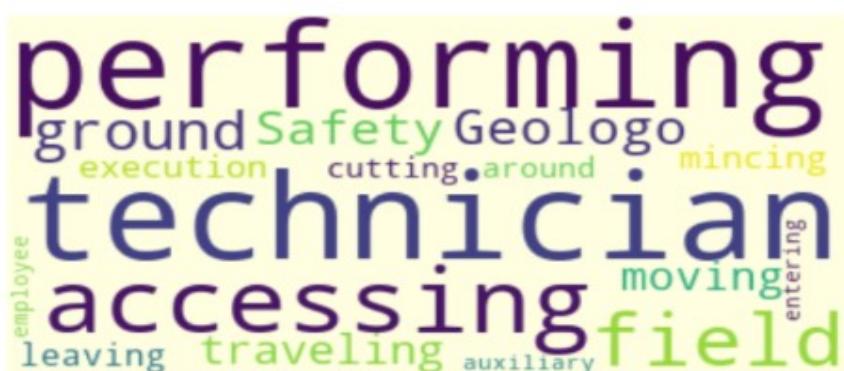
## 1. WordCloud for Industry type : Mining



## 2. WordCloud for Industry type : Metals



### 3. WordCloud for Industry type : Others



### ***Country:***

## 1. WordCloud for Country : Country\_01



## 2. WordCloud for Country : Country\_02



### 3. WordCloud for Country : Country\_03



## **INFERENCE:**

### **Inference from Univariate Analysis:**

- Country\_01 has 59.3% (highest) Proportion of Total number of accidents that took place in 2016 and 2017.
- Local\_03 has 21.3% (highest) Proportion in Total number of accidents that took place in 2016 and 2017.
- Mining has 56.7% (highest) Proportion in Total number of accidents that took place in 2016 and 2017.
- Accidentlevel I has 73.9% (highest) Proportion in Total number of accidents that took place in 2016 and 2017.
- PotentialAccidentlevel IV has 33.7%(highest) Proportion in Total number of accidents that took place in 2016 and 2017.
- Male Gender has 94.7%(highest) Proportion in Total number of accidents that took place in 2016 and 2017.
- Nature of Employee of type "Third Party" has 44.3% (highest) Proportion in Total number of accidents that took place in 2016 and 2017.
- Critical Risk of type "Others" has 54.8% (highest) Proportion in Total number of accidents that took place in 2016 and 2017.
- Year 2016 has 67.7% (highest) Proportion in Total number of accidents that took place among the whole.
- Feb Month has 14.6% (highest) Proportion in Total number of accidents, followed by April and June with 12.2%, that took place in 2016 and 2017.
- Thursday has 18.2% (highest) Proportion in Total number of accidents that took place in 2016 and 2017, followed by Tuesday and Wednesday. While the Lowest seems to be on Sunday.

**Inference from Bivariate Analysis:**

- Accident level I is dominant among Male and Female Genders.
- Potential Accident Level of IV is dominant among Male, while II is dominant among Female.
- Country\_01 is dominant among Male accidents, while Country\_02 is dominant among Female accidents.
- Mining sector is dominant among Male accidents while Metals is dominant among Female accidents.
- Third Party employee type is dominant in accidents among Male and Female Gender
- Accident Level I has maximum number of Accidents happened in the Mining sector.
- Potential accident level 4 has maximum number of Accidents happened in the Mining sector.
- Local 3 has maximum number of Accidents happened in the Mining sector. Local 11 has least number of accidents in industrial sector.
- Mining sector has maximum number of accidents took place in Country\_01 while Others has least number of accidents in Country\_01.
- Maximum number of accidents took place in accident level 1 and country\_01.
- Maximum number of accidents took place in country\_01 and potential accident level 3.
- Country 1 is more dominant among all Local Regions while Country 03 is least dominant.
- Country 1 is more dominant among all Types of Employees while Country 03 is least dominant.
- Overall Local 3 is more prone to Multiple potential Accidents, while Local 12 is the least dominant.
- Employee is more dominant across all Local Regions, while Type Third Party (Remote) is least dominant across all Local Regions.
- Accident Level I is more related to Potential Accident levels of all Levels, while Accident Level V is least related.
- Accident Level I is more dominant across all Employee types and all Countries and all Regions, where Level V is least across all Employee Types, all Countries and all Regions.
- Potential Accident level IV dominants in ThirdParty, while VI is least dominant in Third

Party (Remote) across all.

- Critical Risk of type "Others" is dominant across all Types of Employees, Regions, Industry Types.
- Maximum accidents took place in the Local Region 3 in both 2016 and 2017 while Local Region 12 is the least among both years.

## **Summary:**

1. Employees in Country\_01 are more prone to Accidents. More Precautionary measures and training should be given to avoid similar accidents.
2. Most accidents are of less Severity Comparatively. More Precautionary measures and training will help avoid similar accidents.
3. Study reveals more potential accidents can happen with High Severity of IV. Minimizing the Less Severity accidents can help avoid High Severity Potential Accidents.
4. Mining industry needs special Care and attention by providing Industry level Safety Measures to help Employees avoid accidents.
5. More Study and Analysis is required on what Accident category types that Critical Risk of type "Others" is associated with and necessary Precautionary measures should be implemented.
6. Though there is a decline in Number of Accidents from 2016 and 2017, the less Severity accidents are still dominant. Special attention is required to minimize these from Management.
7. Feb, April, June are Critical Months which are more prone to accidents. Monitoring the Weather, Industrial Environments and Employee Work Mood should be analysed and precautioned with proper arrangements.
8. Thursday is most prone to accidents. Needs to check if proper Senior Technicians are available and monitor the situation with Timely maintenance to avoid Accidents happening due to any failures.

