Write a program for performing industrial data analysis using relevant tools and techniques

IIOT LAB - assignment 9

Perform data analysis on industrial safety and health analysis:

1. Show correlation using heatmap on all predictors.
2. Which country has most and least number of accidents?
3. Which industrial sector has most and least number of accidents ?
4. Among employee and third party who is most likely to get into accident ?
5. Which gender is most likely to get into accident ?
6. Perform visualizations like bar plot, count plot, etc.

In [1]:
```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

In [2]:
```
data = pd.read_csv('IHMStefanini_industrial_safety_and_health_database_with_
data['Gender'] = data['Genre'].map({'Male': 0, 'Female': 1})
data.head(5)
```

Out[2]:

	Unnamed: 0	Data	Countries	Local	Industry Sector	Accident Level	Potential Accident Level	Genre	Employee or Third Party
0	0	2016-01-01 00:00:00	Country_01	Local_01	Mining	I	IV	Male	Third Party
1	1	2016-01-02 00:00:00	Country_02	Local_02	Mining	I	IV	Male	Employee
2	2	2016-01-06 00:00:00	Country_01	Local_03	Mining	I	III	Male	Third Party (Remote)
3	3	2016-01-08 00:00:00	Country_01	Local_04	Mining	I	I	Male	Third Party
4	4	2016-01-10 00:00:00	Country_01	Local_04	Mining	IV	IV	Male	Third Party

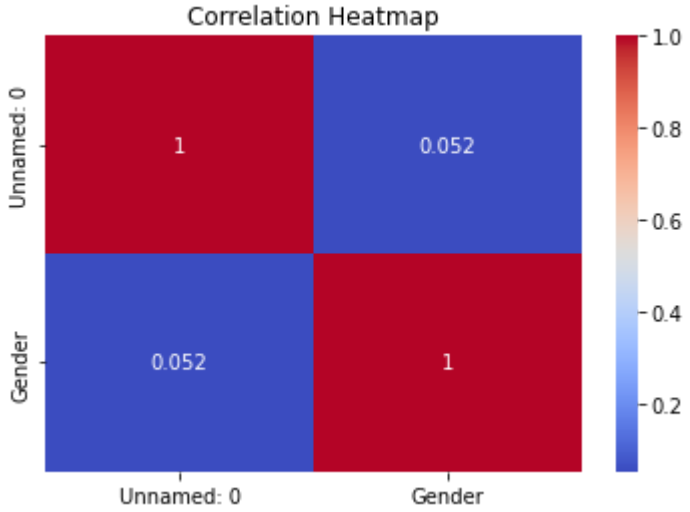
1. Show correlation using heatmap on all predictors.

In [3]:
```
# Select only numeric columns for correlation analysis
numeric_data = data.select_dtypes(include='number')
# Calculate the correlation matrix
correlation_matrix = numeric_data.corr()
correlation_matrix
```

Out[3]:

	Unnamed: 0	Gender
Unnamed: 0	1.000000	0.051851
Gender	0.051851	1.000000

In [4]:
```
# Create a heatmap
plt.figure(figsize=(6, 4))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap')
plt.show()
```



2. Which country has most and least number of accidents?

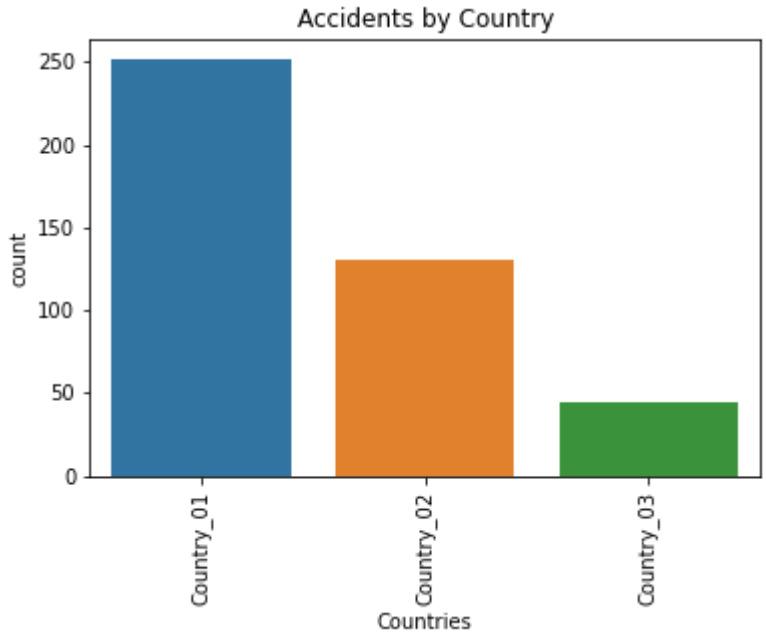
In [5]:
```
most_accidents_country = data['Countries'].value_counts().idxmax()
least_accidents_country = data['Countries'].value_counts().idxmin()
print("Country with the most accidents:", most_accidents_country)
print("Country with the least accidents:", least_accidents_country)
```

```
Country with the most accidents: Country_01
Country with the least accidents: Country_03
```

In [6]:
```
data['Countries'].value_counts()
```

Out[6]:
```
Countries
Country_01    251
Country_02    130
Country_03     44
Name: count, dtype: int64
```

In [7]:
```
plt.figure(figsize=(6, 4))
sns.countplot(x='Countries', data=data)
plt.title('Accidents by Country')
plt.xticks(rotation=90)
plt.show()
```



3. Which industrial sector has most and least number of accidents ?

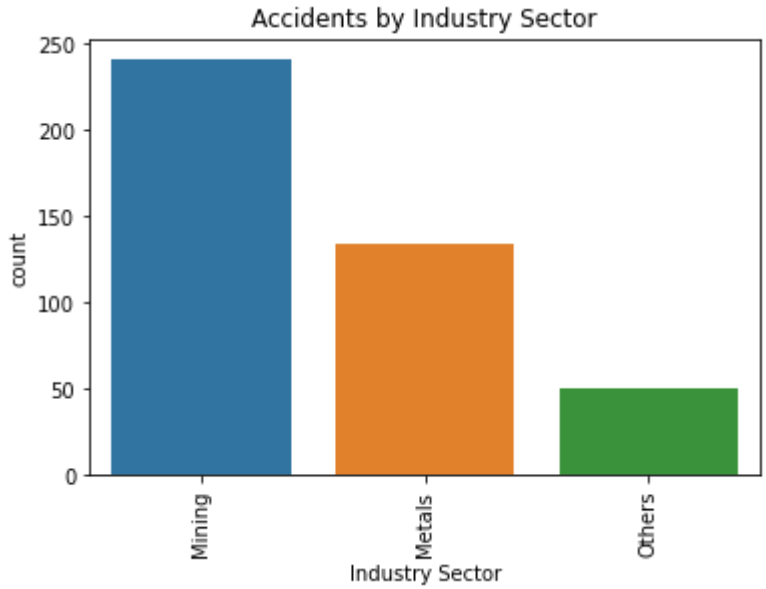
In [8]:
```
most_accidents_sector = data['Industry Sector'].value_counts().idxmax()
least_accidents_sector = data['Industry Sector'].value_counts().idxmin()
print("Sector with the most accidents:", most_accidents_sector)
print("Sector with the least accidents:", least_accidents_sector)
```

```
Sector with the most accidents: Mining
Sector with the least accidents: Others
```

In [9]:
```
data['Industry Sector'].value_counts()
```

Out[9]:
```
Industry Sector
Mining    241
Metals    134
Others     50
Name: count, dtype: int64
```

In [10]:
```
plt.figure(figsize=(6, 4))
sns.countplot(x='Industry Sector', data=data)
plt.title('Accidents by Industry Sector')
plt.xticks(rotation=90)
plt.show()
```



4. Among employee and third party who is most likely to get into accident ?

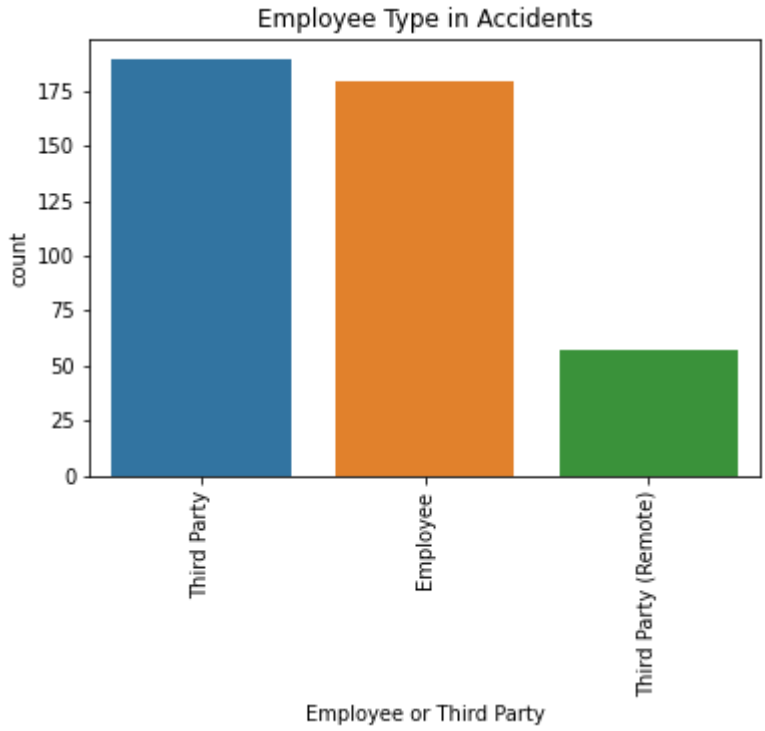
In [11]:
```
most_likely_to_accident = data['Employee or Third Party'].value_counts().id
print("Most likely to get into an accident:", most_likely_to_accident)
```

```
Most likely to get into an accident: Third Party
```

In [12]:
```
data['Employee or Third Party'].value_counts()
```

Out[12]:
```
Employee or Third Party
Third Party             189
Employee                179
Third Party (Remote)     57
Name: count, dtype: int64
```

In [13]:
```
plt.figure(figsize=(6, 4))
sns.countplot(x='Employee or Third Party', data=data)
plt.title('Employee Type in Accidents')
plt.xticks(rotation=90)
plt.show()
```



5. Which gender is most likely to get into accident ?

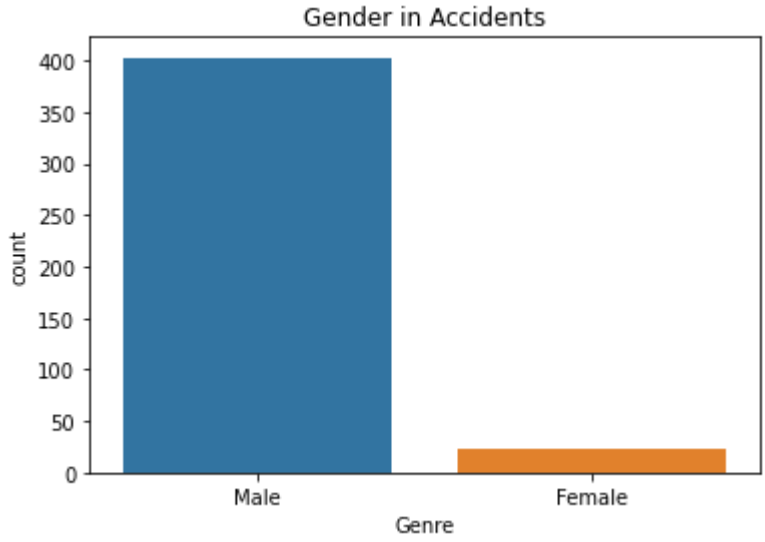
In [14]:
```
most_likely_gender = data['Genre'].value_counts().idxmax()
print("Most likely gender to get into an accident:", most_likely_gender)
```

```
Most likely gender to get into an accident: Male
```

In [15]:
```
data['Genre'].value_counts()
```

Out[15]:
```
Genre
Male      403
Female     22
Name: count, dtype: int64
```

In [16]:
```
plt.figure(figsize=(6, 4))
sns.countplot(x='Genre', data=data)
plt.title('Gender in Accidents')
plt.show()
```



In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]: