Data Driven Governance Final Project

Topic: Causes of accidents in different ages -2018

In [2]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.simplefilter(action='ignore', category=FutureWarning)
import psycopg2 as ps
```

In [3]:
```
data1=pd.read_csv("C:\sem-2 (DSSA-22-24)\DDG/NCRB-ADSI-2018-Table-1.7.csv")
data=data1.copy()   # for copying data set into another variable
data
```

Out[3]:

	Sl. No.	Cause	Below 14 years - Male	Below 14 years - Female	Below 14 years - Transgender	Below 14 years - Total	14 and Above - Below 18 years - Male	14 and Above - Below 18 years - Female	14 and Above - Below 18 years - Transgender	14 and Above - Below 18 years - Total	...	45 and Above - Below 60 years - Total	60 years & Above - Male	60 years & Above - Female	60 years & Above - Transgender	60 years & Above - Total	Total - Male
0	1	Air Crash	0	0	0	0	0	0	0	0	...	11	0	0	0	0	13
1	2	Ship Accidents	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0
2	3	Collapse of Structure (Total)	101	77	0	178	96	49	0	145	...	399	107	80	0	187	1483
3	3	3.1 Collapse of Dwelling House/Residential Bui...	80	60	0	140	54	35	0	89	...	244	78	62	0	140	862
4	3	3.2 Collapse of Official/ Commercial Building	3	0	0	3	1	0	0	1	...	15	3	0	0	3	47
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
57	19	Suffocation	61	35	0	96	59	60	0	119	...	442	134	39	0	173	1422
58	20	Drug Overdose	9	7	0	16	14	10	0	24	...	206	70	25	0	95	720
59	21	Other than above Causes	1452	1068	1	2521	1721	878	0	2599	...	14477	5973	1804	1	7778	45003
60	22	Causes Not Known	295	161	2	458	555	280	0	835	...	3597	1267	350	0	1617	11500
61	Total	Total	10917	5488	3	16408	19039	6082	0	25121	...	87319	31178	9625	1	40804	325767

62 rows × 31 columns

In [30]:
```
data.size
data.shape
```

Out[30]:
```
(62, 31)
```

In [40]:
```
cause=data.iloc[:,1]
cause
```

Out[40]:
```
0                                        Air Crash
1                                    Ship Accidents
2                          Collapse of Structure (Total)
3     3.1 Collapse of Dwelling House/Residential Bui...
4         3.2 Collapse of Official/ Commercial Building
                            ...
57                                      Suffocation
58                                    Drug Overdose
59                               Other than above Causes
60                                      Causes Not Known
61                                                Total
Name: Cause, Length: 62, dtype: object
```

In [4]:
```
df=data.copy()
df.drop([2],axis=0,inplace=True)
df.drop([8,13,18,26,34,39,43,48,61],axis=0,inplace=True)
df   #used for cleaning purpose
```

Out[4]:

	Sl. No.	Cause	Below 14 years - Male	Below 14 years - Female	Below 14 years - Transgender	Below 14 years - Total	14 and Above - Below 18 years - Male	14 and Above - Below 18 years - Female	14 and Above - Below 18 years - Transgender	14 and Above - Below 18 years - Total	...	45 and Above - Below 60 years - Total	60 years & Above - Male	60 years & Above - Female	60 years & Above - Transgender	60 years & Above - Total
0	1	Air Crash	0	0	0	0	0	0	0	0	...	11	0	0	0	0
1	2	Ship Accidents	0	0	0	0	0	0	0	0	...	0	0	0	0	0
3	3	3.1 Collapse of Dwelling House/Residential Bui...	80	60	0	140	54	35	0	89	...	244	78	62	0	140
4	3	3.2 Collapse of Official/ Commercial Building	3	0	0	3	1	0	0	1	...	15	3	0	0	3
5	3	3.3 Collapse of Dam	0	0	0	0	1	0	0	1	...	0	0	0	0	0

Exploratory Data Analysis(EDA)

In [5]:
```
df.head()     # for showing first five rows of dataset
```

Out[5]:

	Sl. No.	Cause	Below 14 years - Male	Below 14 years - Female	Below 14 years - Transgender	Below 14 years - Total	14 and Above - Below 18 years - Male	14 and Above - Below 18 years - Female	14 and Above - Below 18 years - Transgender	14 and Above - Below 18 years - Total	...	45 and Above - Below 60 years - Total	60 years & Above - Male	60 years & Above - Female	60 years & Above - Transgender	60 years & Above - Total	Total - Male	T Fe
0	1	Air Crash	0	0	0	0	0	0	0	0	...	11	0	0	0	0	13	
1	2	Ship Accidents	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	
3	3	3.1 Collapse of Dwelling House/Residential Bui...	80	60	0	140	54	35	0	89	...	244	78	62	0	140	862	
4	3	3.2 Collapse of Official/ Commercial Building	3	0	0	3	1	0	0	1	...	15	3	0	0	3	47	
5	3	3.3 Collapse of Dam	0	0	0	0	1	0	0	1	...	0	0	0	0	0	6	

5 rows × 31 columns

In [6]:
```
data.tail()   #for calculating last five rows of datasets
```

Out[6]:

	Sl. No.	Cause	Below 14 years - Male	Below 14 years - Female	Below 14 years - Transgender	Below 14 years - Total	14 and Above - Below 18 years - Male	14 and Above - Below 18 years - Female	14 and Above - Below 18 years - Transgender	14 and Above - Below 18 years - Total	...	45 and Above - Below 60 years - Total	60 years & Above - Male	60 years & Above - Female	60 years & Above - Transgender	60 years & Above - Total	Total - Male	Tot Fem
57	19	Suffocation	61	35	0	96	59	60	0	119	...	442	134	39	0	173	1422	4
58	20	Drug Overdose	9	7	0	16	14	10	0	24	...	206	70	25	0	95	720	1
59	21	Other than above Causes	1452	1068	1	2521	1721	878	0	2599	...	14477	5973	1804	1	7778	45003	130
60	22	Causes Not Known	295	161	2	458	555	280	0	835	...	3597	1267	350	0	1617	11500	31
61	Total	Total	10917	5488	3	16408	19039	6082	0	25121	...	87319	31178	9625	1	40804	325767	791

5 rows × 31 columns

In [7]:
```
df.isnull()  #for checking at which place null value is
```

Out[7]:

	Sl. No.	Cause	Below 14 years - Male	Below 14 years - Female	Below 14 years - Transgender	Below 14 years - Total	14 and Above - Below 18 years - Male	14 and Above - Below 18 years - Female	14 and Above - Below 18 years - Transgender	14 and Above - Below 18 years - Total	...	45 and Above - Below 60 years - Total	60 years & Above - Male	60 years & Above - Female	60 years & Above - Transgender	60 years & Above - Total	Total - Male	Total - Female
0	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
5	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
6	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
7	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
9	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
10	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
11	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
12	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
14	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
15	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
16	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
17	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
19	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
20	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
21	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
22	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
23	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
24	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
25	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
27	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
28	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
29	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
30	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
31	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
32	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
33	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
35	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
36	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
37	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
38	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
40	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
41	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
42	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
44	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
45	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
46	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
47	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
49	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
50	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
51	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
52	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
53	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
54	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
55	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
56	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
57	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
58	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
59	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False

	Sl. No.	Cause	Below 14 years - Male	Below 14 years - Female	Below 14 years - Transgender	Below 14 years - Total	14 and Above - Below 18 years - Male	14 and Above - Below 18 years - Female	14 and Above - Below 18 years - Transgender	14 and Above - Below 18 years - Total	...	45 and Above - Below 60 years - Total	60 years & Above - Male	60 years & Above - Female	60 years & Above - Transgender	60 years & Above - Total	Total - Male	Total - Female	.
60	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	

52 rows × 31 columns

In [8]:
```
df.isnull().sum()    #for checking how much null values present in data sets
```

Out[8]:
```
Sl. No.                                    0
Cause                                      0
Below 14 years - Male                      0
Below 14 years - Female                    0
Below 14 years - Transgender               0
Below 14 years - Total                     0
14 and Above - Below 18 years - Male       0
14 and Above - Below 18 years - Female     0
14 and Above - Below 18 years - Transgender    0
14 and Above - Below 18 years - Total      0
18 and Above - Below 30 years - Male       0
18 and Above - Below 30 years - Female     0
18 and Above - Below 30 years - Transgender    0
18 and Above - Below 30 years - Total      0
30 and Above - Below 45 years - Male       0
30 and Above - Below 45 years - Female     0
30 and Above - Below 45 years - Transgender    0
30 and Above - Below 45 years - Total      0
45 and Above - Below 60 years - Male       0
45 and Above - Below 60 years - Female     0
45 and Above - Below 60 years - Transgender    0
45 and Above - Below 60 years - Total      0
60 years & Above - Male                    0
60 years & Above - Female                  0
60 years & Above - Transgender             0
60 years & Above - Total                   0
Total - Male                               0
Total - Female                             0
Total - Transgender                        0
Total - Total                              0
Total - Percentage Share                   0
dtype: int64
```

In [9]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 52 entries, 0 to 60
Data columns (total 31 columns):
 #   Column                                        Non-Null Count  Dtype
---  ------                                        --------------  -----
 0   Sl. No.                                       52 non-null     object
 1   Cause                                         52 non-null     object
 2   Below 14 years - Male                         52 non-null     int64
 3   Below 14 years - Female                       52 non-null     int64
 4   Below 14 years - Transgender                  52 non-null     int64
 5   Below 14 years - Total                        52 non-null     int64
 6   14 and Above - Below 18 years - Male          52 non-null     int64
 7   14 and Above - Below 18 years - Female        52 non-null     int64
 8   14 and Above - Below 18 years - Transgender   52 non-null     int64
 9   14 and Above - Below 18 years - Total         52 non-null     int64
 10  18 and Above - Below 30 years - Male          52 non-null     int64
 11  18 and Above - Below 30 years - Female        52 non-null     int64
 12  18 and Above - Below 30 years - Transgender   52 non-null     int64
 13  18 and Above - Below 30 years - Total         52 non-null     int64
 14  30 and Above - Below 45 years - Male          52 non-null     int64
 15  30 and Above - Below 45 years - Female        52 non-null     int64
 16  30 and Above - Below 45 years - Transgender   52 non-null     int64
 17  30 and Above - Below 45 years - Total         52 non-null     int64
 18  45 and Above - Below 60 years - Male          52 non-null     int64
 19  45 and Above - Below 60 years - Female        52 non-null     int64
 20  45 and Above - Below 60 years - Transgender   52 non-null     int64
 21  45 and Above - Below 60 years - Total         52 non-null     int64
 22  60 years & Above - Male                       52 non-null     int64
 23  60 years & Above - Female                     52 non-null     int64
 24  60 years & Above - Transgender                52 non-null     int64
 25  60 years & Above - Total                      52 non-null     int64
 26  Total - Male                                  52 non-null     int64
 27  Total - Female                                52 non-null     int64
 28  Total - Transgender                           52 non-null     int64
 29  Total - Total                                 52 non-null     int64
 30  Total - Percentage Share                      52 non-null     float64
dtypes: float64(1), int64(28), object(2)
memory usage: 13.0+ KB
```

In [10]:
```
print(df.dtypes) #for looking columns and their datatypes
```

```
Sl. No.                                        object
Cause                                          object
Below 14 years - Male                           int64
Below 14 years - Female                         int64
Below 14 years - Transgender                    int64
Below 14 years - Total                          int64
14 and Above - Below 18 years - Male            int64
14 and Above - Below 18 years - Female          int64
14 and Above - Below 18 years - Transgender     int64
14 and Above - Below 18 years - Total           int64
18 and Above - Below 30 years - Male            int64
18 and Above - Below 30 years - Female          int64
18 and Above - Below 30 years - Transgender     int64
18 and Above - Below 30 years - Total           int64
30 and Above - Below 45 years - Male            int64
30 and Above - Below 45 years - Female          int64
30 and Above - Below 45 years - Transgender     int64
30 and Above - Below 45 years - Total           int64
45 and Above - Below 60 years - Male            int64
45 and Above - Below 60 years - Female          int64
45 and Above - Below 60 years - Transgender     int64
45 and Above - Below 60 years - Total           int64
60 years & Above - Male                         int64
60 years & Above - Female                       int64
60 years & Above - Transgender                  int64
60 years & Above - Total                        int64
Total - Male                                    int64
Total - Female                                  int64
Total - Transgender                             int64
Total - Total                                   int64
Total - Percentage Share                      float64
dtype: object
```

In [24]:
```
df.nunique()    #To find unique values in columns
```

Out[24]:
```
Sl. No.                                    22
Cause                                      52
Below 14 years - Male                      34
Below 14 years - Female                    32
Below 14 years - Transgender                3
Below 14 years - Total                     38
14 and Above - Below 18 years - Male       41
14 and Above - Below 18 years - Female     34
14 and Above - Below 18 years - Transgender     1
14 and Above - Below 18 years - Total      41
18 and Above - Below 30 years - Male       46
18 and Above - Below 30 years - Female     39
18 and Above - Below 30 years - Transgender     4
18 and Above - Below 30 years - Total      47
30 and Above - Below 45 years - Male       48
30 and Above - Below 45 years - Female     42
30 and Above - Below 45 years - Transgender     3
30 and Above - Below 45 years - Total      50
45 and Above - Below 60 years - Male       46
45 and Above - Below 60 years - Female     38
45 and Above - Below 60 years - Transgender     1
45 and Above - Below 60 years - Total      49
60 years & Above - Male                    41
60 years & Above - Female                  33
60 years & Above - Transgender              2
60 years & Above - Total                   41
Total - Male                               49
Total - Female                             45
Total - Transgender                         4
Total - Total                              51
Total - Percentage Share                   21
dtype: int64
```

In [12]:
```
print(df.size) #fpr checking size of the data
print(df.shape)  #for checking shape of the data
```

```
1612
(52, 31)
```

In [13]:
```
df.columns
```

Out[13]:
```
Index(['Sl. No.', 'Cause', 'Below 14 years - Male', 'Below 14 years - Female',
       'Below 14 years - Transgender', 'Below 14 years - Total',
       '14 and Above - Below 18 years - Male',
       '14 and Above - Below 18 years - Female',
       '14 and Above - Below 18 years - Transgender',
       '14 and Above - Below 18 years - Total',
       '18 and Above - Below 30 years - Male',
       '18 and Above - Below 30 years - Female',
       '18 and Above - Below 30 years - Transgender',
       '18 and Above - Below 30 years - Total',
       '30 and Above - Below 45 years - Male',
       '30 and Above - Below 45 years - Female',
       '30 and Above - Below 45 years - Transgender',
       '30 and Above - Below 45 years - Total',
       '45 and Above - Below 60 years - Male',
       '45 and Above - Below 60 years - Female',
       '45 and Above - Below 60 years - Transgender',
       '45 and Above - Below 60 years - Total', '60 years & Above - Male',
       '60 years & Above - Female', '60 years & Above - Transgender',
       '60 years & Above - Total', 'Total - Male', 'Total - Female',
       'Total - Transgender', 'Total - Total', 'Total - Percentage Share'],
      dtype='object')
```

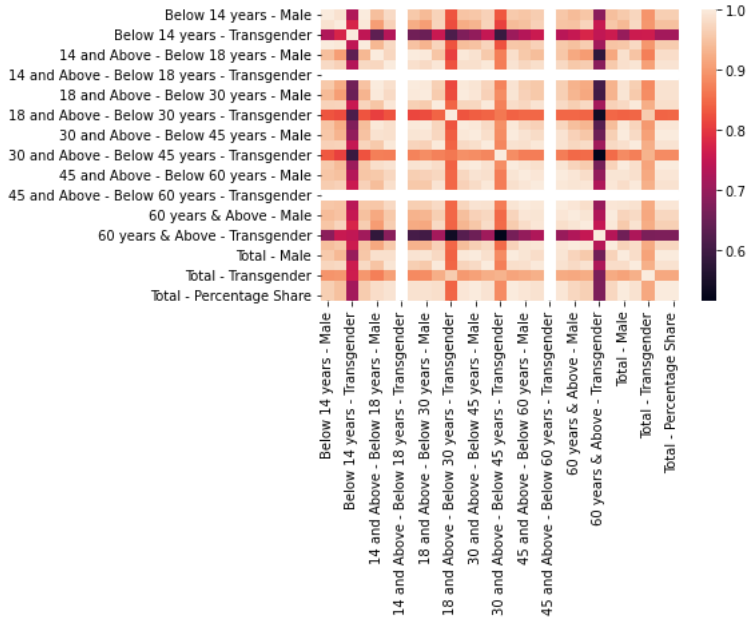
Statistical Analysis

In [14]:
```
data.corr()
```

Out[14]:

	Below 14 years - Male	Below 14 years - Female	Below 14 years - Transgender	Below 14 years - Total	14 and Above - Below 18 years - Male	14 and Above - Below 18 years - Female	14 and Above - Below 18 years - Transgender	14 and Above - Below 18 years - Total	18 and Above - Below 30 years - Male	18 and Above - Below 30 years - Female	...	45 and Above - Below 60 years - Total	60 years & Above - Male	60 years & Above - Female	Tr
Below 14 years - Male	1.000000	0.987834	0.728308	0.998701	0.945226	0.977581	NaN	0.956495	0.939528	0.968136	...	0.954075	0.940878	0.952396	
Below 14 years - Female	0.987834	1.000000	0.774191	0.994475	0.904114	0.967701	NaN	0.922186	0.904578	0.971045	...	0.949915	0.947586	0.970242	
Below 14 years - Transgender	0.728308	0.774191	1.000000	0.745371	0.634289	0.731247	NaN	0.658663	0.649944	0.756114	...	0.737186	0.752605	0.775793	
Below 14 years - Total	0.998701	0.994475	0.745371	1.000000	0.934298	0.976974	NaN	0.947826	0.930605	0.971694	...	0.955281	0.945610	0.960806	
14 and Above - Below 18	0.945226	0.904114	0.634289	0.934298	1.000000	0.975449	NaN	0.998789	0.997487	0.952575	...	0.954033	0.917604	0.902457	

In [15]:
```
sns.heatmap(data.corr())
plt.show()
```



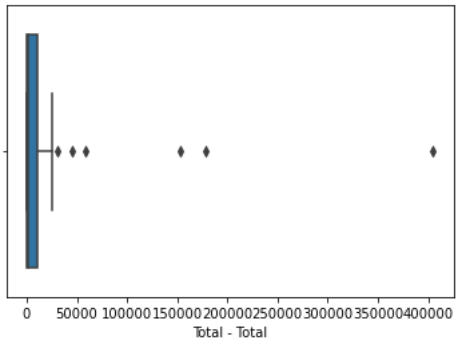
In [16]:
```
df.describe()
```

Out[16]:

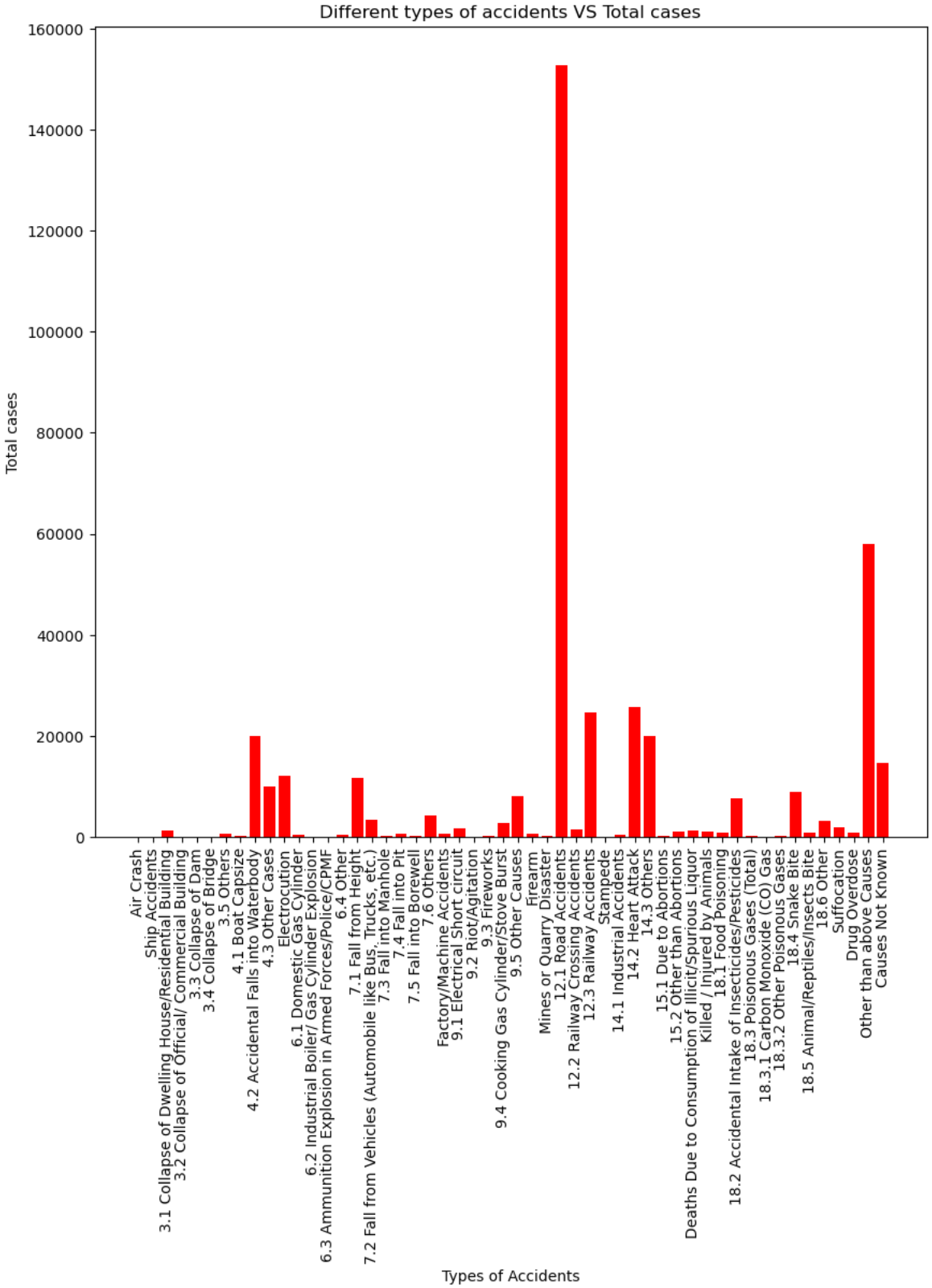
	Below 14 years - Male	Below 14 years - Female	Below 14 years - Transgender	Below 14 years - Total	14 and Above - Below 18 years - Male	14 and Above - Below 18 years - Female	14 and Above - Below 18 years - Transgender	14 and Above - Below 18 years - Total	18 and Above - Below 30 years - Male	18 and Above - Below 30 years - Female	...	45 and Above - Below 60 years - Tota
count	52.000000	52.000000	52.000000	52.000000	52.000000	52.000000	52.0	52.000000	52.000000	52.000000	...	52.000000
mean	210.000000	105.576923	0.057692	315.634615	366.192308	117.038462	0.0	483.230769	1707.730769	405.423077	...	1679.826923
std	548.165425	241.143627	0.307645	779.236907	1424.463023	328.597433	0.0	1742.499800	6233.359942	989.064746	...	4478.318087
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	...	0.000000
25%	3.000000	2.000000	0.000000	5.000000	5.000000	2.750000	0.0	6.000000	24.500000	9.000000	...	22.000000
50%	17.500000	11.000000	0.000000	26.000000	25.500000	12.000000	0.0	36.500000	161.500000	48.000000	...	137.500000
75%	83.250000	60.500000	0.000000	147.000000	141.500000	98.750000	0.0	198.250000	932.500000	413.750000	...	1095.000000
max	3132.000000	1090.000000	2.000000	4222.000000	10112.000000	2177.000000	0.0	12289.000000	44005.000000	6192.000000	...	27229.000000

8 rows × 29 columns

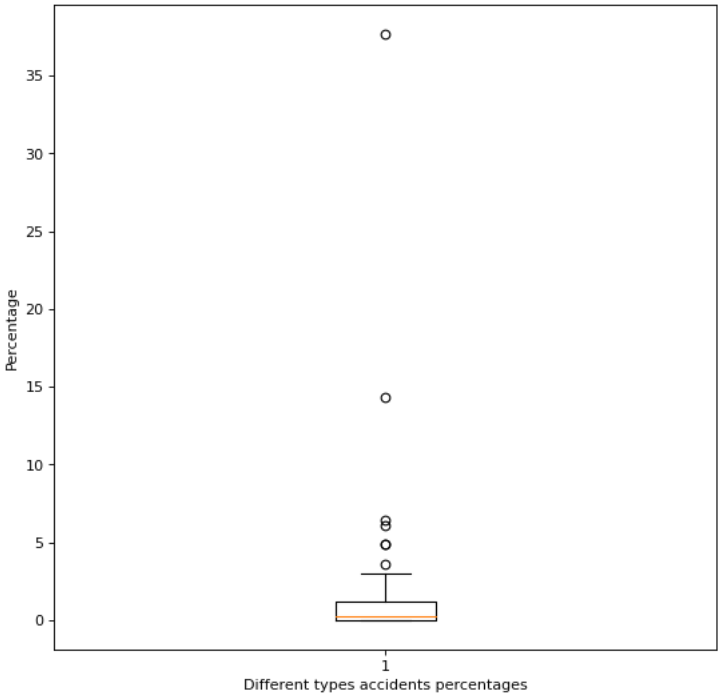
In [17]:
```
sns.boxplot(x=data["Total - Total"])
plt.show()
```



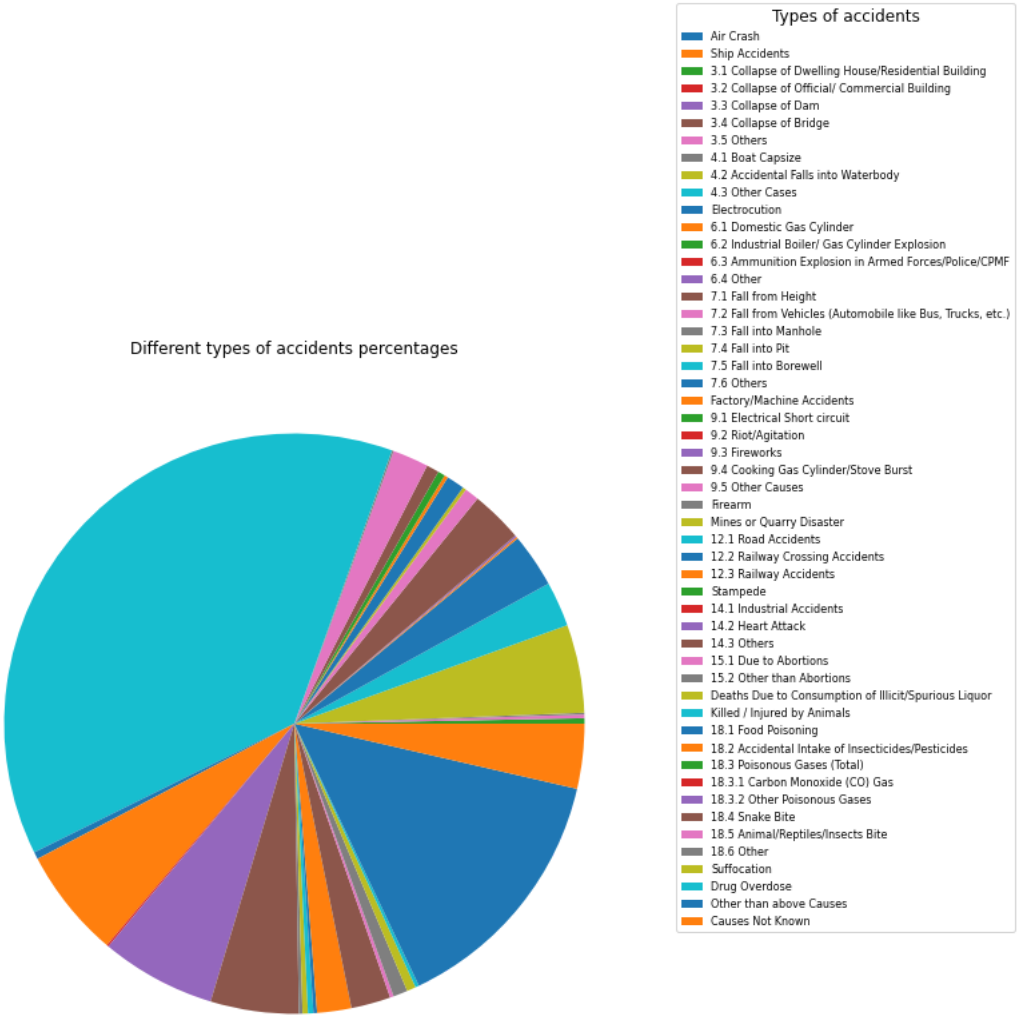
In [18]:
```
plt.figure(figsize=(10,10), dpi=100)
plt.bar(df['Cause'],df['Total - Total'],color="red")
plt.xticks(rotation=90)
plt.title("Different types of accidents VS Total cases")
plt.xlabel("Types of Accidents")
plt.ylabel("Total cases")
plt.show()
```



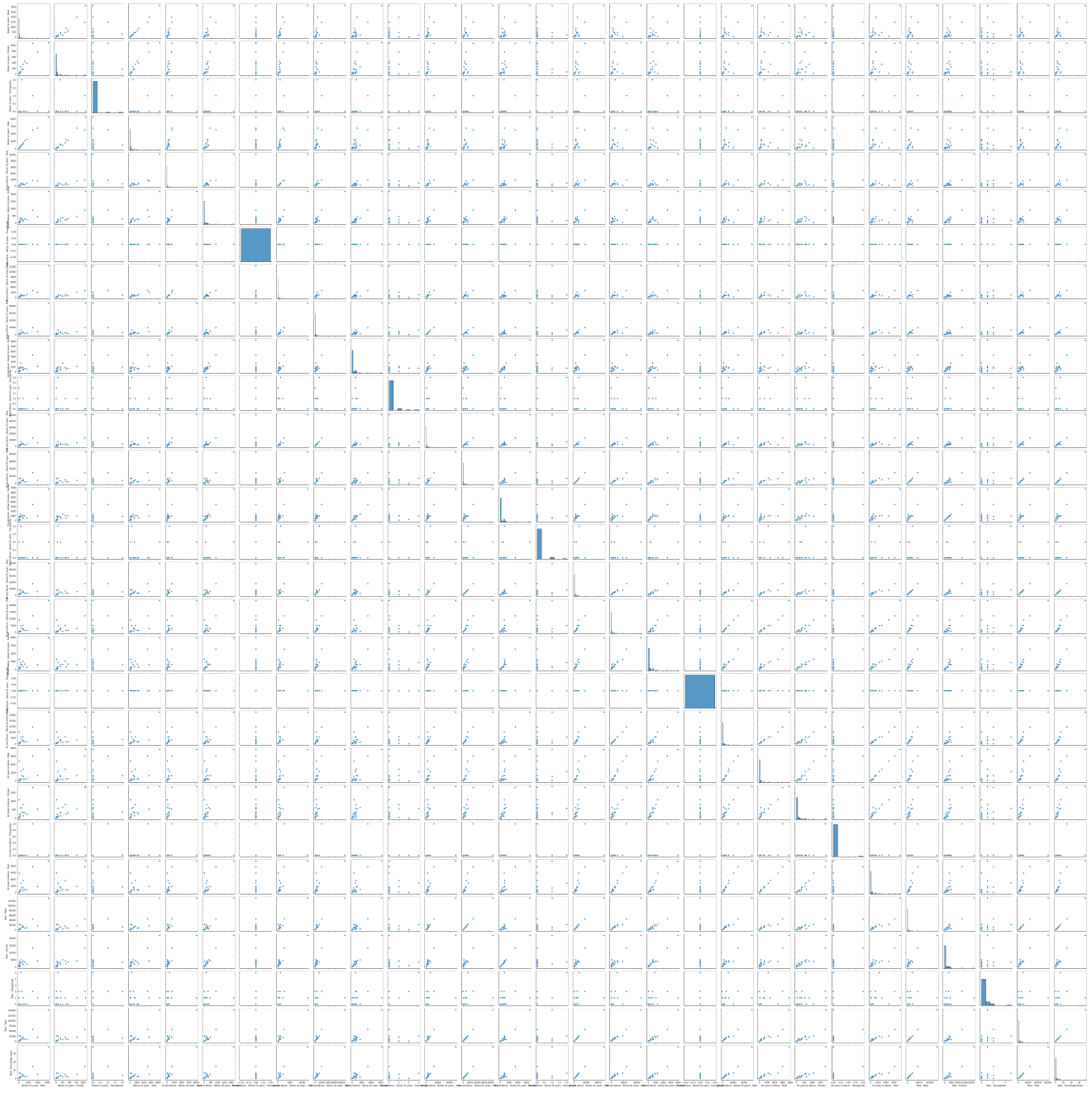
In [19]:
```
plt.figure(figsize=(8,8), dpi=80)
plt.boxplot(df['Total - Percentage Share'])
plt.xlabel("Different types accidents percentages")
plt.ylabel("Percentage")
plt.show()
```



In [20]:
```
fig = plt.figure(1005, figsize=(10,10))
plt.pie(df['Total - Percentage Share'])
plt.title("Different types of accidents percentages")
plt.legend(df['Cause'], loc="best",fontsize=8, bbox_to_anchor=(-0.5,-0.5,2,2),title="Types of accidents",
           title_fontsize="large")
plt.show()
```



In [21]:
```
sns.pairplot(df,kind="scatter")            #to plot subplots of data
plt.show()
```



In [22]:
```
df.columns
```

Out[22]:
```
Index(['Sl. No.', 'Cause', 'Below 14 years - Male', 'Below 14 years - Female',
       'Below 14 years - Transgender', 'Below 14 years - Total',
       '14 and Above - Below 18 years - Male',
       '14 and Above - Below 18 years - Female',
       '14 and Above - Below 18 years - Transgender',
       '14 and Above - Below 18 years - Total',
       '18 and Above - Below 30 years - Male',
       '18 and Above - Below 30 years - Female',
       '18 and Above - Below 30 years - Transgender',
       '18 and Above - Below 30 years - Total',
       '30 and Above - Below 45 years - Male',
       '30 and Above - Below 45 years - Female',
       '30 and Above - Below 45 years - Transgender',
       '30 and Above - Below 45 years - Total',
       '45 and Above - Below 60 years - Male',
       '45 and Above - Below 60 years - Female',
       '45 and Above - Below 60 years - Transgender',
       '45 and Above - Below 60 years - Total', '60 years & Above - Male',
       '60 years & Above - Female', '60 years & Above - Transgender',
       '60 years & Above - Total', 'Total - Male', 'Total - Female',
       'Total - Transgender', 'Total - Total', 'Total - Percentage Share'],
      dtype='object')
```

In [23]:
```
df.to_excel('Types of accidents.xlsx',index=False)   #for converting file into excel
```

In [27]:
```
import pandas_profiling as pp
profile=pp.ProfileReport(df)
profile.to_file("output.html")    #for visualization through pandas_profiling
```

```
Summarize dataset:   0%|          | 0/5 [00:00<?, ?it/s]

Generate report structure:   0%|          | 0/1 [00:00<?, ?it/s]

Render HTML:   0%|          | 0/1 [00:00<?, ?it/s]

Export report to file:   0%|          | 0/1 [00:00<?, ?it/s]
```

Connection with postgreSQL database

In [56]:
```
conn = ps.connect("dbname=students user=postgres password=souravk28")    #for connecting
cursor=conn.cursor()          #for cursor/operator connection
cursor.execute("DROP TABLE IF EXISTS accidents")
```

In [57]:
```
df=data.iloc[:,1:31]
df
```

Out[57]:

	Cause	Below 14 years - Male	Below 14 years - Female	Below 14 years - Transgender	Below 14 years - Total	14 and Above - Below 18 years - Male	14 and Above - Below 18 years - Female	14 and Above - Below 18 years - Transgender	14 and Above - Below 18 years - Total	18 and Above - Below 30 years - Male	...	45 and Above - Below 60 years - Total	60 years & Above - Male	60 years & Above - Female	60 years & Above - Transgender	60 years & Above - Total	Tota Ma
0	Air Crash	0	0	0	0	0	0	0	0	2	...	11	0	0	0	0	
1	Ship Accidents	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	
2	Collapse of Structure (Total)	101	77	0	178	96	49	0	145	392	...	399	107	80	0	187	14
3	3.1 Collapse of Dwelling House/Residential Bui...	80	60	0	140	54	35	0	89	209	...	244	78	62	0	140	8
4	3.2 Collapse of Official/ Commercial Building	3	0	0	3	1	0	0	1	15	...	15	3	0	0	3	[illegible]
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
57	Suffocation	61	35	0	96	59	60	0	119	391	...	442	134	39	0	173	14
58	Drug Overdose	9	7	0	16	14	10	0	24	199	...	206	70	25	0	95	7
59	Other than above Causes	1452	1068	1	2521	1721	878	0	2599	9678	...	14477	5973	1804	1	7778	450
60	Causes Not Known	295	161	2	458	555	280	0	835	2645	...	3597	1267	350	0	1617	115
61	Total	10917	5488	3	16408	19039	6082	0	25121	88763	...	87319	31178	9625	1	40804	3257

62 rows × 30 columns

In [58]:
```
cursor.execute("DROP TABLE IF EXISTS accidents")
cursor.execute("create table accidents(causes text, Below14yrsMale float,Below14yrsFemale float,Below14yrsTransgender float,Below
```

In [59]:
```
conn.commit()
```

In [60]:
```
for i in df.index:
    value =[df.at[i,col] for col in list(df.columns)]
    query=("insert into accidents values('%s','%s','%s','%s','%s','%s','%s','%s','%s','%s','%s','%s','%s','%s','%s','%s','%s','%s
    cursor.execute(query)
conn.commit()
```

In [61]:
```
cursor.execute("select * from accidents")
conn.commit()
cursor.fetchall()
```

Out[61]:
```
[('Air Crash',
  0.0,
  0.0,
  0.0,
  0.0,
  0.0,
  0.0,
  0.0,
  0.0,
  2.0,
  0.0,
  0.0,
  2.0,
  2.0,
  0.0,
  0.0,
  2.0,
  9.0,
  2.0,
```

In [62]:
```
cursor.execute("select * from accidents where causes like 'S%'")
conn.commit()
cursor.fetchall()
```

Out[62]:
```
[('Ship Accidents',
  0.0,
  0.0,
  0.0,
  0.0,
  0.0,
  0.0,
  0.0,
  0.0,
  0.0,
  0.0,
  0.0,
  0.0,
  0.0,
  0.0,
  0.0,
  0.0,
  0.0,
  0.0,
  0.0,
  0.0,
```