ASSIGNMENT 4

```python
#Q1. Explore the dataset and identify the trends in fatalities over
time. Identify any significant
#changes, spikes, or declines in the number of fatalities.
import pandas as pd
import matplotlib.pyplot as plt

file = pd.read_csv("fatalities.csv")
file
file.info()#WE USED IT TO SEE THE INFO OF OUR DATA


file['date_of_death'] =
pd.to_datetime(file['date_of_death'])#CONVERTED TO THE DATE FORMAT

file['yearlyfatality'] = file['date_of_death'].dt.to_period('Y')

yearly_fatalities = file.groupby('yearlyfatality').size()

#we can observe that every year lives are lost and in 2014 most
fatalities are happened
plt.figure(figsize=(14, 8))
plt.plot(yearly_fatalities.index.astype(str),
yearly_fatalities.values, marker='*', linestyle='--', color='red')
plt.title('YEARLY FATALITES / DEATHS')
plt.xlabel('YEAR')
plt.ylabel('NUMBER OF DEATHS')
plt.show()




import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
#here we can see that mostly people are killed in the year 2014 and
majority are males , and every year they targeted the male majority
#every year the deaths of male ratio is higher than the female
deaths , throughout from year 2000 to 2023 you can see that every year
#males death ratio in comparison with female is highest ,so we can
understand that they focused on eradicating the males to make the
nation weak

file = pd.read_csv('fatalities.csv')
file['date_of_death'] = pd.to_datetime(file['date_of_death'])
file['year_of_death'] = file['date_of_death'].dt.year
sns.set_theme(style="darkgrid")
```

```python
plt.figure(figsize=(14, 8))
sns.countplot(x='year_of_death', hue='gender', data=file,
palette='colorblind')
plt.title('FATALITIES BY YEAR OF DEATH AND GENDER', fontsize=28)
plt.xlabel('YEAR OF DEATH', fontsize=19)
plt.ylabel('NUMBER OF FATALITIES', fontsize=19)
plt.legend(title='Gender', title_fontsize=29)
plt.show()


file = pd.read_csv('fatalities.csv')
file['date_of_death'] = pd.to_datetime(file['date_of_death'])
file['year_of_death'] = file['date_of_death'].dt.year
sns.set_theme(style="whitegrid")
plt.figure(figsize=(14, 8))
sns.countplot(x='year_of_death', hue='citizenship',data=file,
palette='dark')
plt.title('FATALITIES BY YEAR OF DEATH AND CITIZENSHIP', fontsize=28)
plt.xlabel('YEAR OF DEATH', fontsize=19)
plt.ylabel('NUMBER OF FATALITIES', fontsize=19)
plt.legend(title='Gender', title_fontsize=29)
plt.show()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11124 entries, 0 to 11123
Data columns (total 16 columns):
 #   Column                        Non-Null Count  Dtype
---  ------                        --------------  -----
 0   name                          11124 non-null  object
 1   date_of_event                 11124 non-null  object
 2   age                           10995 non-null  float64
 3   citizenship                   11124 non-null  object
```

```
 4    event_location                  11124 non-null  object
 5    event_location_district         11124 non-null  object
 6    event_location_region           11124 non-null  object
 7    date_of_death                   11124 non-null  object
 8    gender                          11104 non-null  object
 9    took_part_in_the_hostilities    9694 non-null   object
 10   place_of_residence              11056 non-null  object
 11   place_of_residence_district     11056 non-null  object
 12   type_of_injury                  10833 non-null  object
 13   ammunition                      5871 non-null   object
 14   killed_by                       11124 non-null  object
 15   notes                           10844 non-null  object
dtypes: float64(1), object(15)
memory usage: 1.4+ MB
```
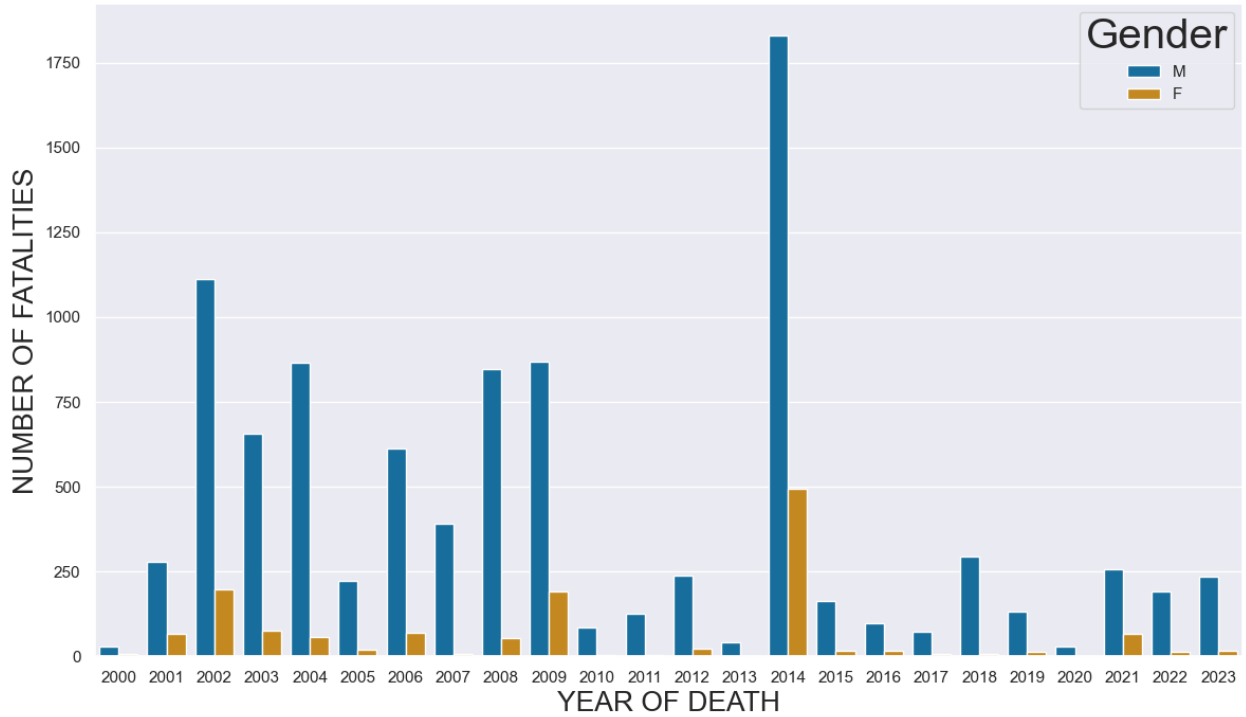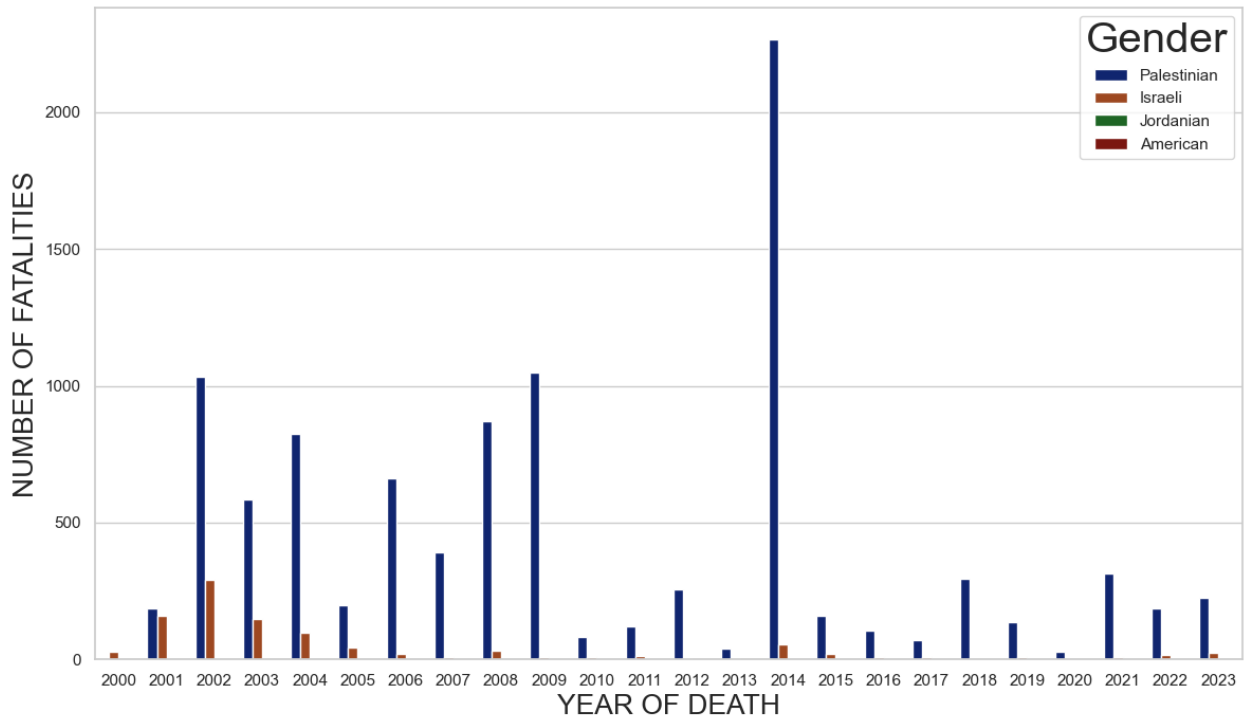

YEARLY FATALITES / DEATHS
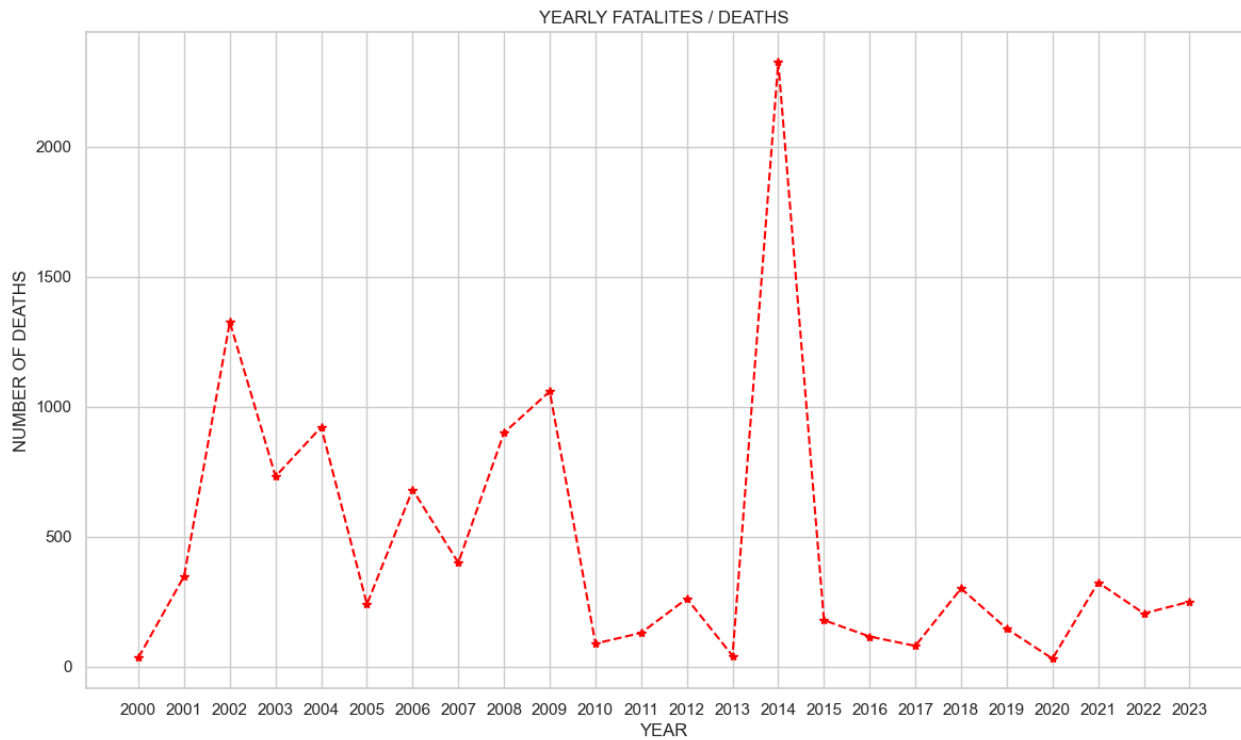
FATALITIES BY YEAR OF DEATH AND GENDER

FATALITIES BY YEAR OF DEATH AND CITIZENSHIP

Q1

We can observed from the graph we plotted is a line plot grpah which displays the number of deaths from the year 2000 to 2023 and in year 2014 we observed the most deaths

In second graph we can observe that the males are mostly targeted and get killed , in comparison to females the male ratio is really high in deaths and in year 2014 it was the highest from 2000 to 2023

In third graph we can observe that the death ratio in Palestinian citizens are high than the other nationalities , Palestinians are targeted the most.

so from these graph we observed that most fatalities happened in year 2014 and the gender which is targeted from 2000 to 2023 are MALES and their nationality is Palestinians.

```python
#Q2. Conduct an analysis by examining the age, gender, and citizenship
of the individuals killed.
#Determine if there are any notable patterns or disparities in the
data.

#cleaning the data
#as there are null values in age
pr =file.isnull().sum()
print(pr)
file['age']=file['age'].fillna(file['age'].median() )
#the null values are covered with median , as the data is skwed

#we are giving the age of the individuals which were killed and by
observing the histogram we can observe that mostly youth is targeted
file = pd.read_csv('fatalities.csv')
sns.set_theme(style="whitegrid")
plt.figure(figsize=(14, 8))
sns.histplot(file['age'], bins=20, color='green', edgecolor='black')
plt.title('AGE OF PEOPLE KILLED')
plt.xlabel('AGE')
plt.ylabel('NUMBER OF DEATHS')
plt.grid(axis='y')
```
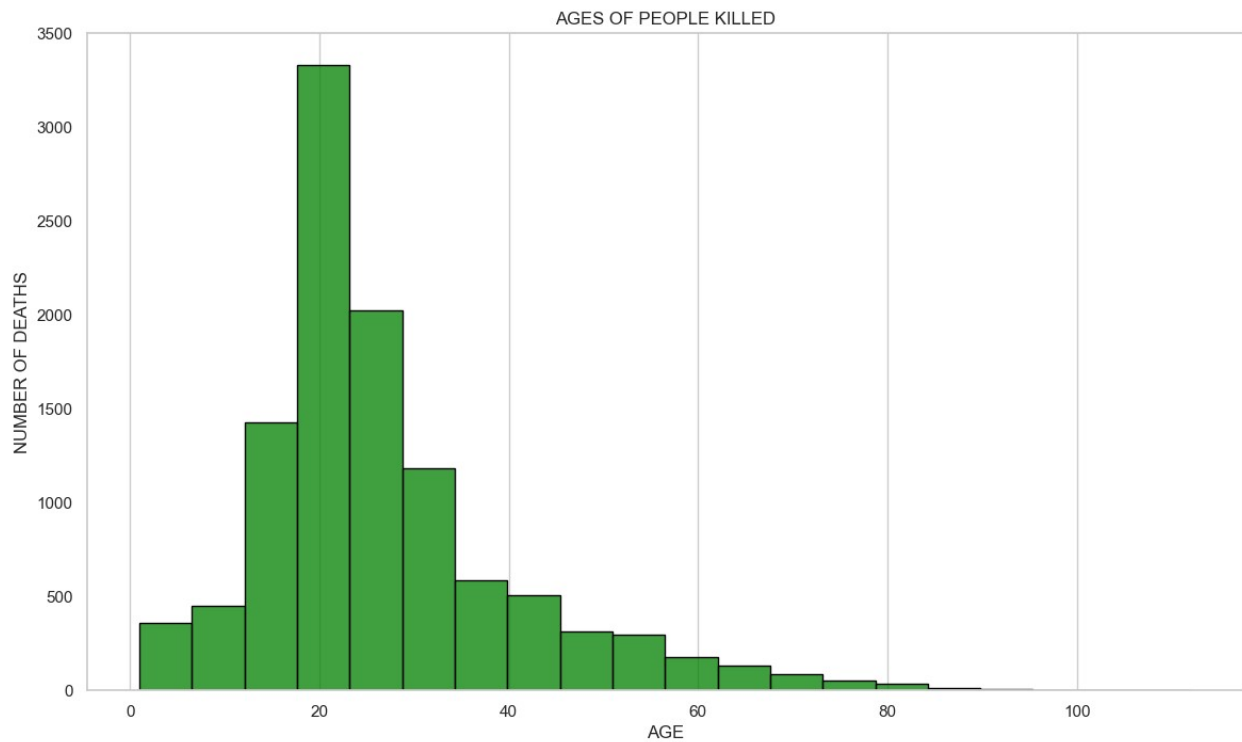
```python
plt.show()


#we have observed with bar chart that mostly male were targeted in the
killing
file = pd.read_csv('fatalities.csv')
file['gender'] = file['gender'].fillna(file['gender'].mode()[0])
sns.set_theme(style="ticks")
plt.figure(figsize=(14, 8))
sns.countplot(x='gender', data=file,hue='gender',
palette='colorblind', edgecolor='black')
plt.title('GENDER OF PEOPLE KILLED')
plt.xlabel('GENDER')
plt.ylabel('NUMBER OF DEATHS')
plt.grid(axis='y')
plt.show()


#HERE WE HAVE GIVEN A PIE CHART TO NDERSTAND THE GENDER DISTRIBUTION
MORE ACCURATELY
file['gender'].value_counts().plot(kind='pie', autopct='%1.1f%%')
plt.title('VICTIMS GENDER')
plt.xlabel('GENDER')
plt.ylabel('')




#we have seen that the majoirty of the citizens which are killed are
palestenians
citizenship_counts = file['citizenship'].value_counts()
plt.figure(figsize=(14, 8))
citizenship_counts.plot(kind='bar',
color=['red','blue','green','yellow'],edgecolor='black')
plt.title('Citizenship Distribution of Individuals Killed')
plt.xlabel('Citizenship')
plt.ylabel('Count')
plt.grid(axis='y')
plt.show()
```
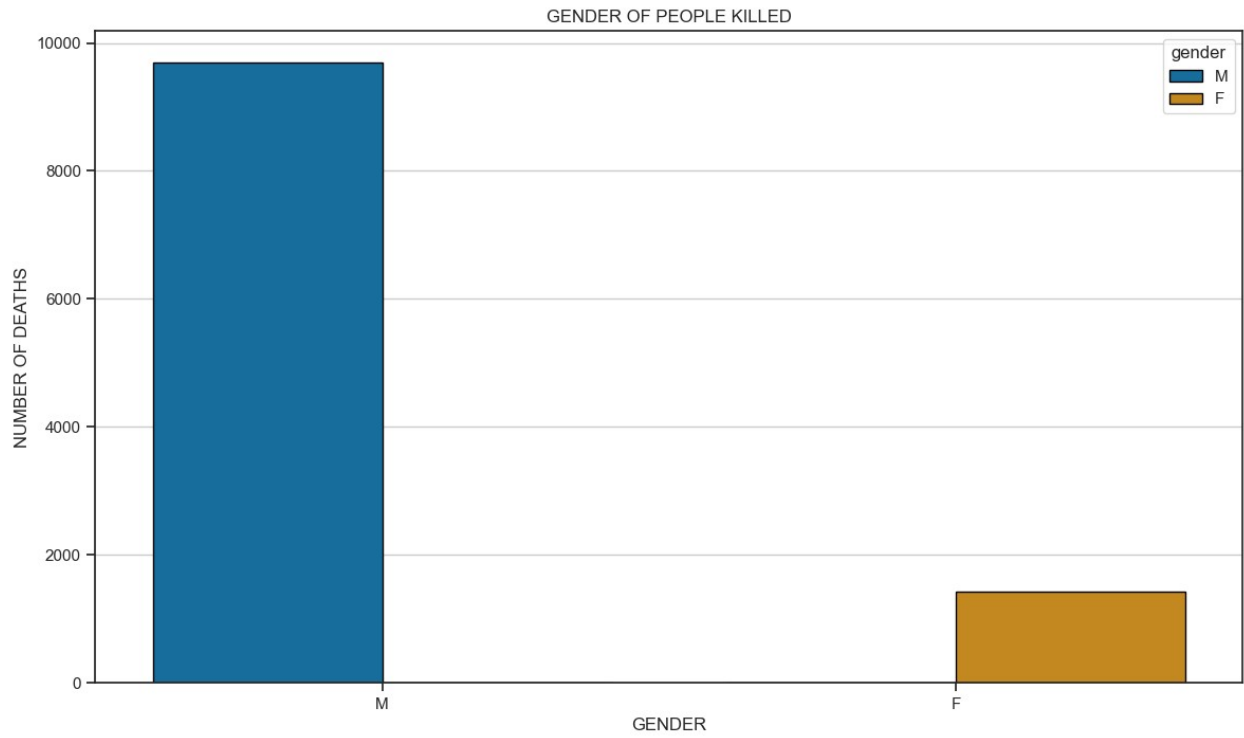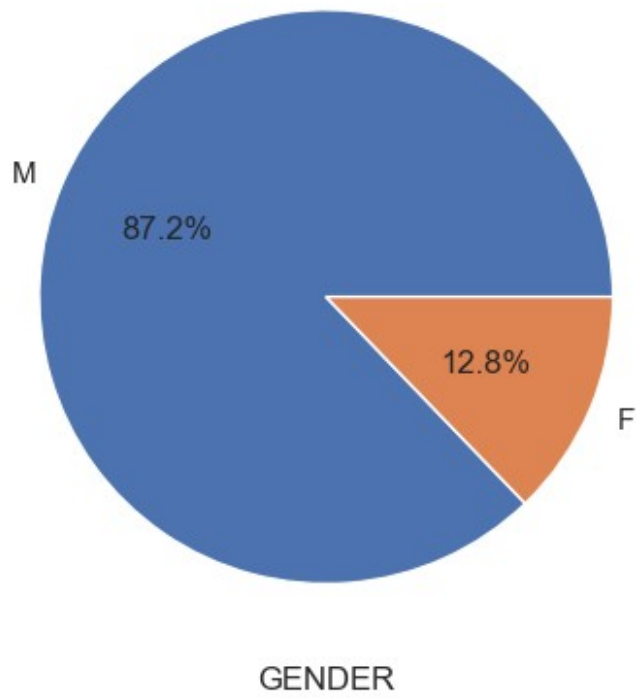
```
name                      0
date_of_event             0
age                     129
citizenship               0
event_location            0
event_location_district   0
```

```
event_location_region                0
date_of_death                        0
gender                               0
took_part_in_the_hostilities      1430
place_of_residence                  68
place_of_residence_district         68
type_of_injury                     291
ammunition                        5253
killed_by                            0
notes                              280
dtype: int64
```
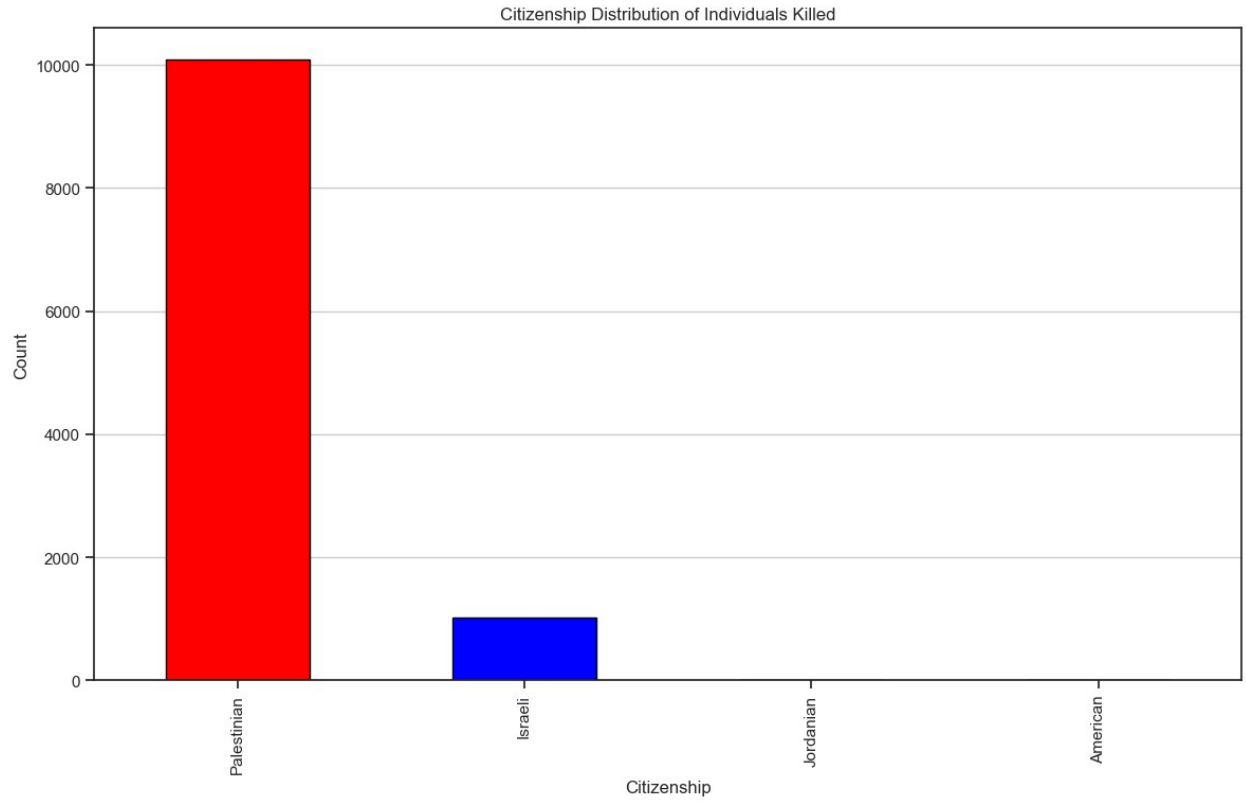

AGES OF PEOPLE KILLED

GENDER OF PEOPLE KILLED

NUMBER OF DEATHS

GENDER

gender
M
F

VICTIMS GENDER

M
87.2%

12.8%
F

GENDER

Citizenship Distribution of Individuals Killed

Q2

(we have cleaned the data by taking in null places in gender col and median in age col because the data was skwed there)

In the first graph of bar chart we have observed that mostly the youth is targeted which is from the age of 20 to 25 and their death number is above 3000 , so they can finish the future of a nation

In second graph we observe that the male's death ratio is super high as comapre to the females , the male deaths are more than 9000

In third pie chart grah we observe that the male death percenatge is 87.2% which is really high and female death ratio is 12.8%

In forth graph we can see that majority of the citizens who have lost their lives are palestinians and after them there are israelis

from all these graphs we have observed that youth is targeted and mostly male have lost their lives and about 87% males have died and the nationality which is facing all this are mostly Palestinian citizens

```python
#Q3. Visualize the distribution of fatalities and identify areas that have experienced higher
#levels of violence.

#we can observe that the most traget area is the Gaza strip and after that it is West abnk and these are the areas where
#palestenians are in population
plt.figure(figsize=(14,8))
citizenship_counts = file['event_location_region'].value_counts()
citizenship_counts.plot(kind='bar',
color=['brown','goldenrod','steelblue'], edgecolor='black')
plt.title('AREAS WITH HIGH FATALITIES RATE IN REGION')
plt.xlabel('\nREGION')
plt.ylabel('NUMBER OF FATALITIES')
```

```python
plt.grid(axis='y')
plt.show()


#graph after observing we understood that mostly palestinians are
killed in gaza strip and west bank , and in israel isrelies
#death ratio is more than the palestinains because GAZA STRIP AND WEST
BANK IS POPULATED BY PALESTINANS AND IN ISRAEL THE RATIO
#OF PALESTINAINS ARE LESS IN COMAPRISON TO ISRAELIES.
sns.set_theme(style="ticks")
plt.figure(figsize=(14, 8))
sns.countplot(x='event_location_region', hue='citizenship', data=file,
palette='deep', edgecolor='black')
plt.title('CITIZENS WITH HIGH DEATH RATE IN REGION')
plt.xlabel('\nREGION')
plt.ylabel('NUMBER OF DEATHS')
plt.grid(axis='y')
plt.show()

#the district which is most effected by the killing is Gaza after it
is North Gaza
locationdistr_counts = file['event_location_district'].value_counts()
plt.figure(figsize=(14, 8))
locationdistr_counts.plot(kind='bar',
color=['blue','orange','green','salmon',],edgecolor='black')
plt.title('DEATH RATE BY DISTRICT')
plt.xlabel('\nDISTRICTS')
plt.ylabel('NUMBER OF DEATHS')
plt.grid(axis='y')
plt.show()


#observing that the location with the most deaths is Gaza city where
palestinians lived
location_counts = file['event_location'].value_counts().head(20)
plt.figure(figsize=(14, 8))
location_counts.plot(kind='bar',
color=['green','blue','red'],edgecolor='black')
plt.title('DEATH RATE BY LOCATION')
plt.xlabel('\nLOCATIONS')
plt.ylabel('NUMBER OF DEATHS')
plt.grid(axis='y')
plt.show()
```
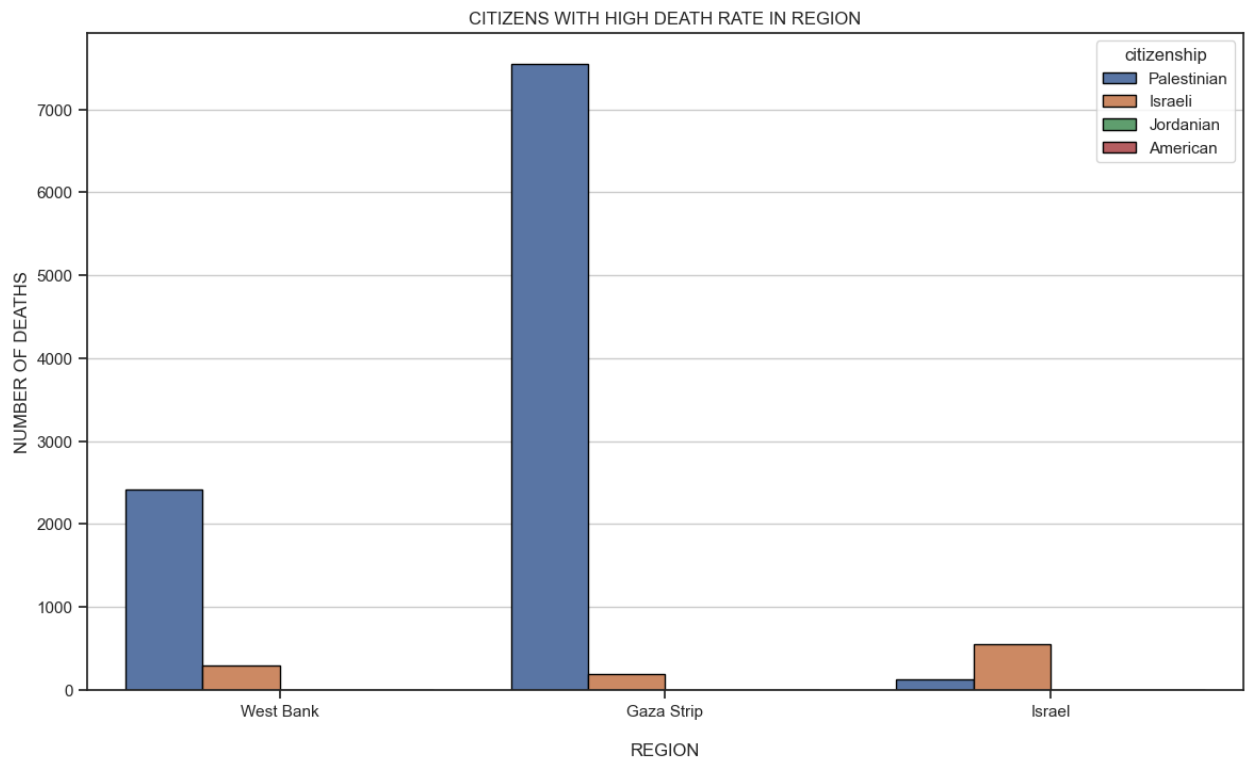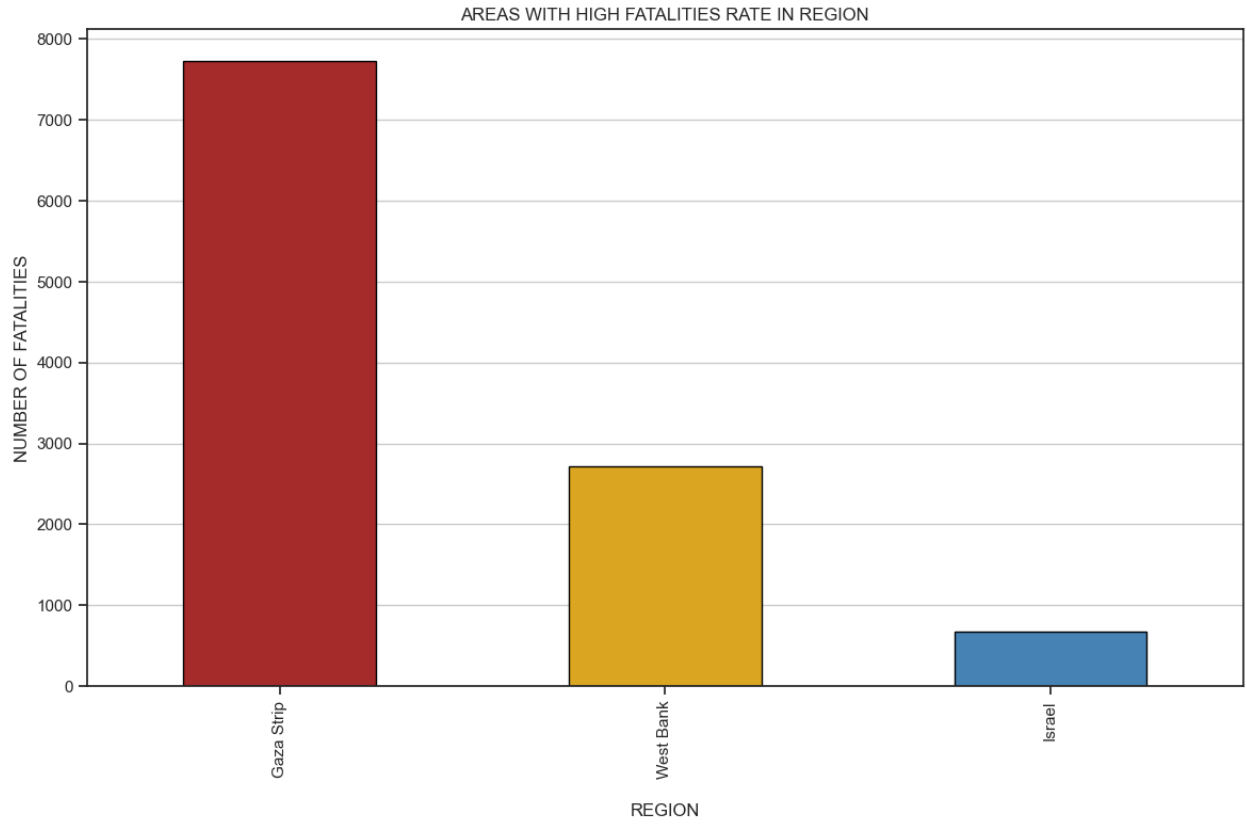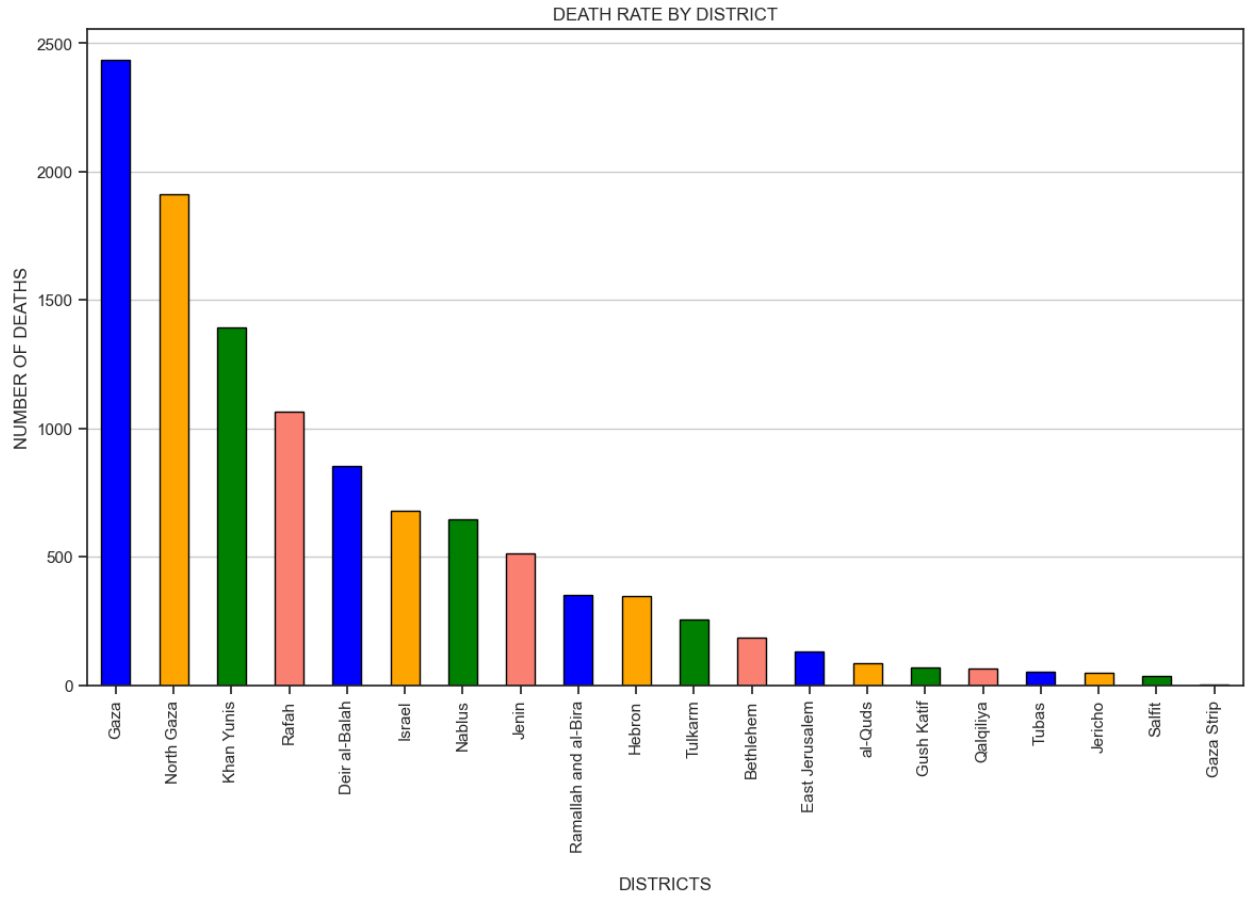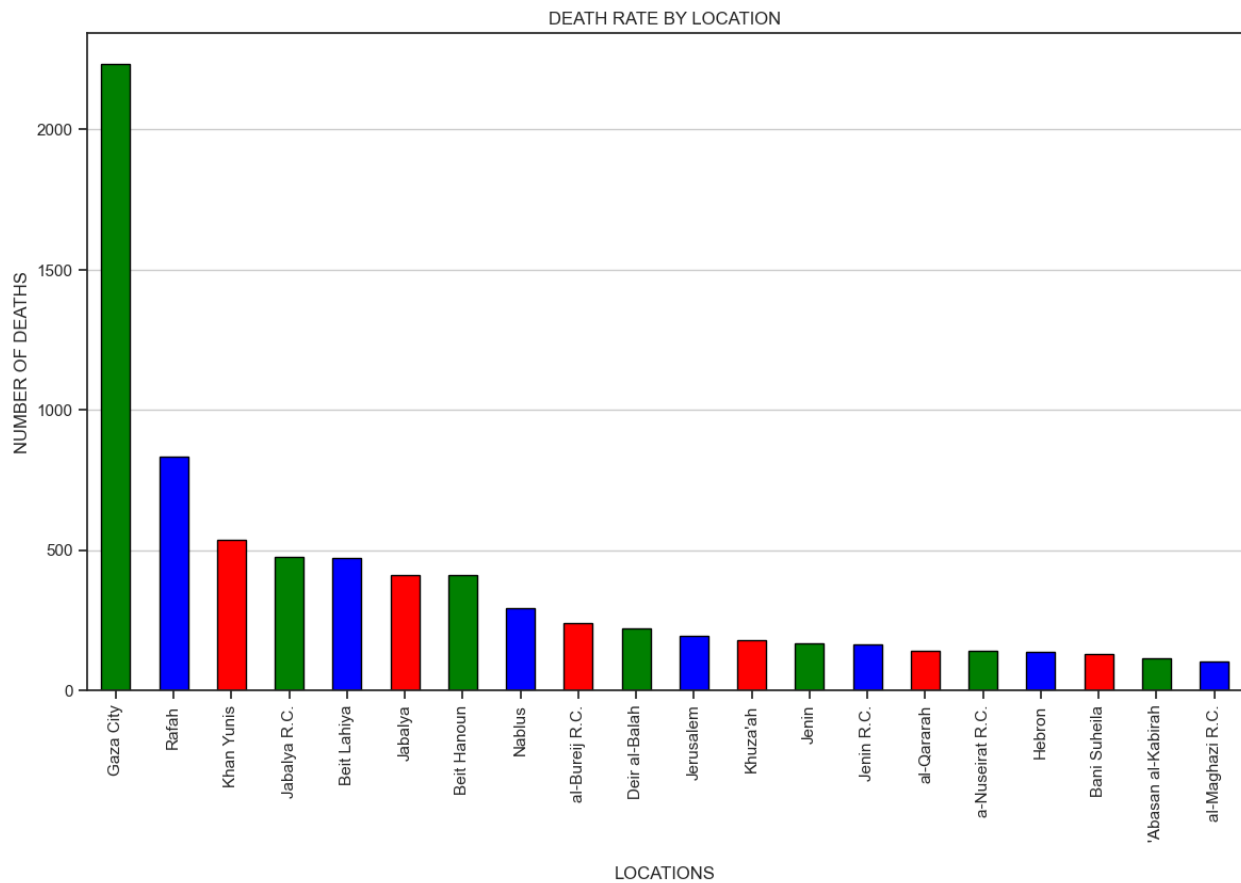
AREAS WITH HIGH FATALITIES RATE IN REGION

NUMBER OF FATALITIES

REGION

CITIZENS WITH HIGH DEATH RATE IN REGION

NUMBER OF DEATHS

REGION

citizenship
Palestinian
Israeli
Jordanian
American

DEATH RATE BY DISTRICT

DEATH RATE BY LOCATION

Q3

In graph we can observed that the region which is most targeted is the Gaza strip and after that West bank .

In second graph we can observe that the Gaza strip and West bank is populated by Palestinians and in Israel the ratio of Palestinians are very less and it is populated by the Israelis

In thrid graph we can observe that the district which face more killings are GAZA , NORTH GAZA and khan yunis which are populated by the Palestinians and death toll here is above 2000

In forth graph we can see that mostly deaths are taking place in Gaza city and after that Rafah which are populated by the Palestinans and the death toll is above 2000

from all these observation we can see that mostly Palestinans places are targeted and they are killed there

```
#Q4. Examine the types of injuries inflicted on individuals. Identify
the most common types of
#injuries and assess their severity.

file['type_of_injury']=file['type_of_injury'].fillna('Not
Specified')#filled with another category of "Not specified"
injured_counts = file['type_of_injury'].value_counts()

plt.figure(figsize=(14, 8))
injured_counts.plot(kind='bar',
color=['steelblue','crimson','orange','green'], edgecolor='black')
plt.title('NUMBER OF PEOPLE FACED INJURIES AND INJURIES TYPE
',fontsize=28)
plt.xlabel('TYPES OF INJURIES',fontsize=19)
plt.ylabel('NUMBER OF PEOPLE INJURED',fontsize=19)
plt.grid(axis='y')
plt.show()
```

```python
sns.set(style="darkgrid")
plt.figure(figsize=(14, 8))
sns.countplot(x="type_of_injury", hue="gender",
palette='dark',data=file)
plt.xlabel("TYPES OF INJURY",fontsize=19)
plt.ylabel("NUMBER OF PEOPLE INJURED",fontsize=19)
plt.title("PEOPLE INJURY TYPE AND THEIR GENDER",fontsize=28)
plt.xticks(rotation=90)
plt.legend(title='GENDER', fontsize=19)
plt.grid(axis='x')
plt.show()


sns.set(style="darkgrid")
plt.figure(figsize=(14, 8))
sns.countplot(x="type_of_injury", hue="citizenship",
palette='bright',data=file)
plt.xlabel("TYPES OF INJURY",fontsize=19)
plt.ylabel("NUMBER OF PEOPLE INJURED",fontsize=19)
plt.title("PEOPLE INJURY TYPE AND THEIR CITIZENSHIP",fontsize=28)
plt.xticks(rotation=90)
plt.legend(title='CITIZENSHIP', fontsize=30)
plt.grid(axis='x')
plt.show()
```
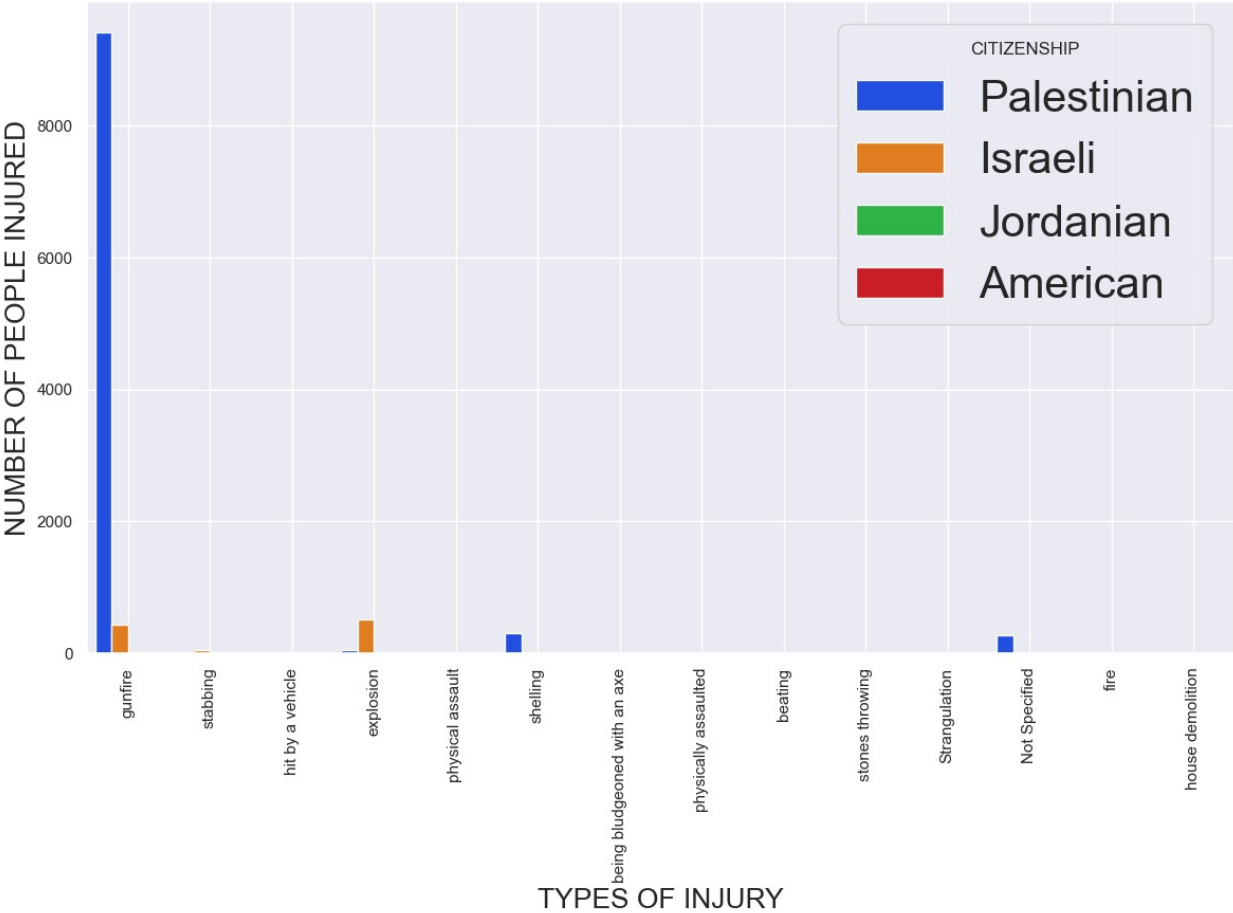
NUMBER OF PEOPLE FACED INJURIES AND INJURIES TYPE
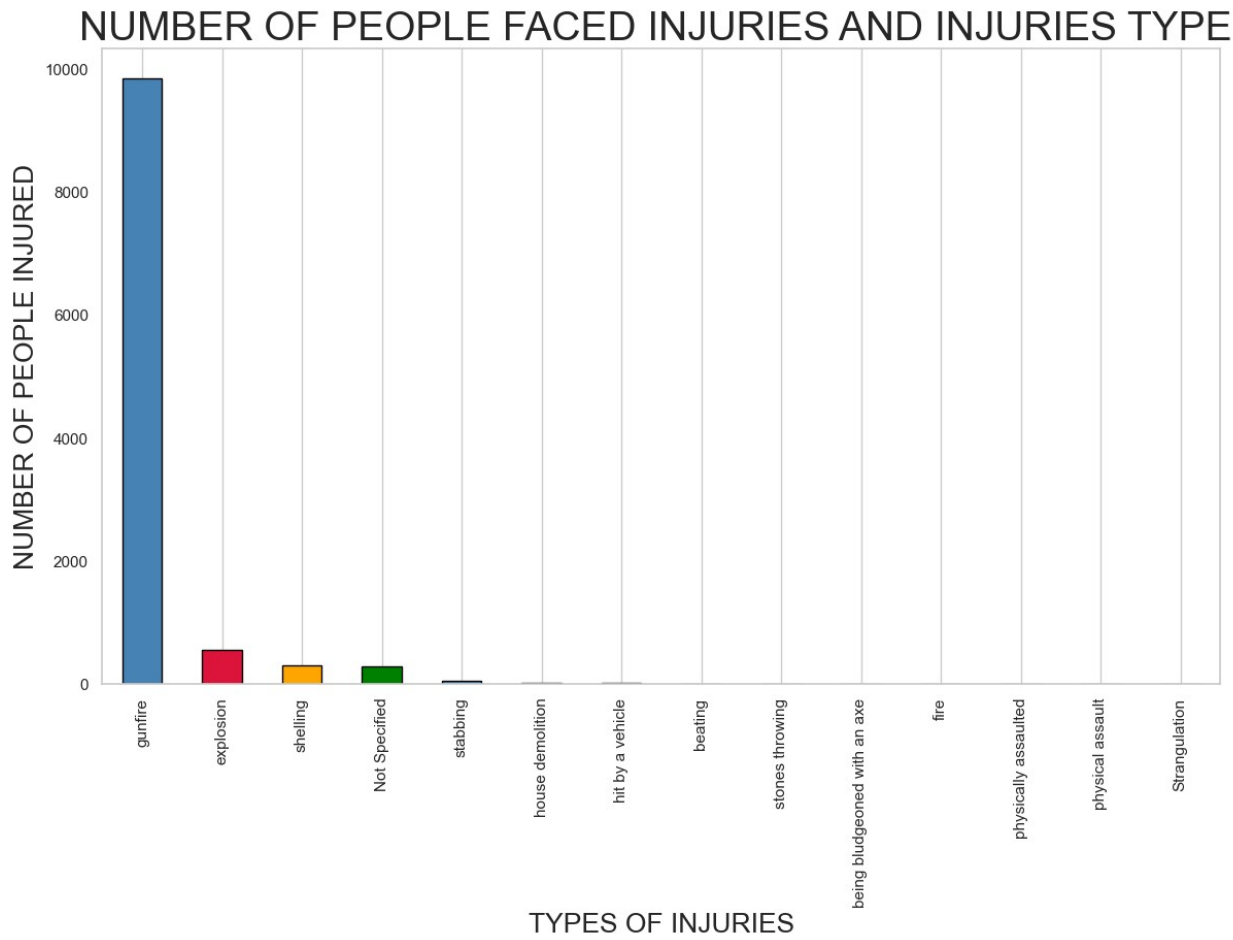
PEOPLE INJURY TYPE AND THEIR GENDER
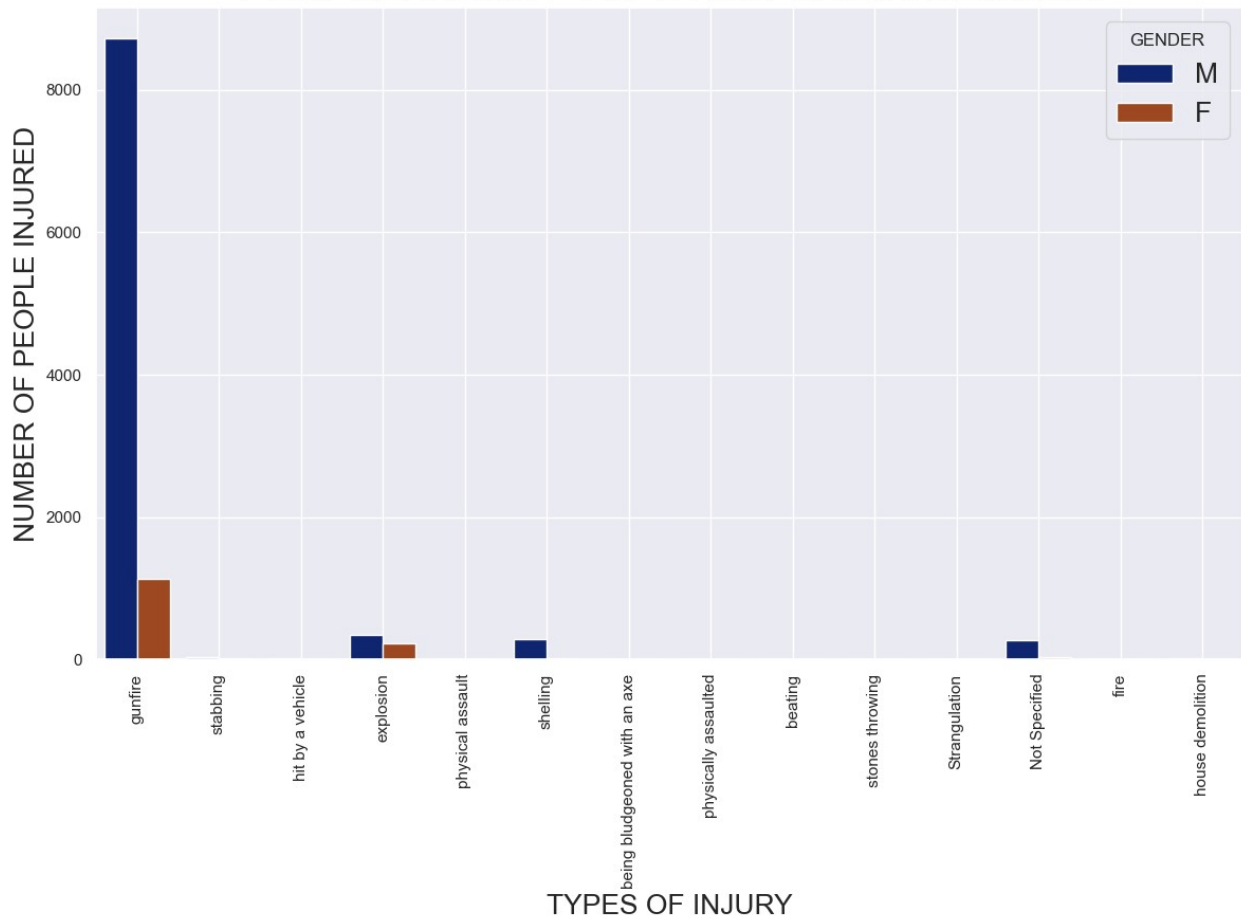
PEOPLE INJURY TYPE AND THEIR CITIZENSHIP

Q4

(we have cleaned the data by entering "Not specified" in null places in types_of_injury col)

We have observed that mostly people are getting injured by the "GUN FIRE" and "EXPLOSION "and their ratio is more than 9000

We have observed that mostly males are injured by the "GUN FIRE" and females are also injured by the "GUN FIRE" but their ratio is less as compare to the males

In third graph we can see that the citizens which are injured the most are Palestinians

so we have observed that Palestinans are targetd and they are mostly injured by the "GUN FIRE" and their injury toll is above 9000

```
#Q5. Analyze the ammunition and means by which the individuals were killed. Determine the
#most frequently used weapons or methods and evaluate their impact.
sns.set(style="darkgrid")
file['ammunition']=file['ammunition'].fillna('Not Specified' )#also finsihed the null values with "Not specified"
ammunitiontyp_counts = file['ammunition'].value_counts()
#WE CAN OBSERVE THAT MOSTLY PEOPLE ARE INJURED AND KILLED BY MISSILES
plt.figure(figsize=(14, 8))
ammunitiontyp_counts.plot(kind='bar',
color=['goldenrod','teal','indigo'], edgecolor='black')
plt.title('KILLED BY THE TYPE OF AMMUNITION',fontsize=28)
plt.xlabel('AMMUNITION TYPE',fontsize=19)
plt.ylabel('NUMBER OF PEOPLE KILLED OR INJURED',fontsize=14)
plt.grid(axis='y')
plt.show()

#here we have observed that the mostly people are killed by the Israeli forces
sns.set(style="darkgrid")
means_of_killing_counts = file['killed_by'].value_counts()
plt.figure(figsize=(12, 7))
means_of_killing_counts.plot(kind='bar',
color=['cadetblue','brown','green'],edgecolor='black')
```
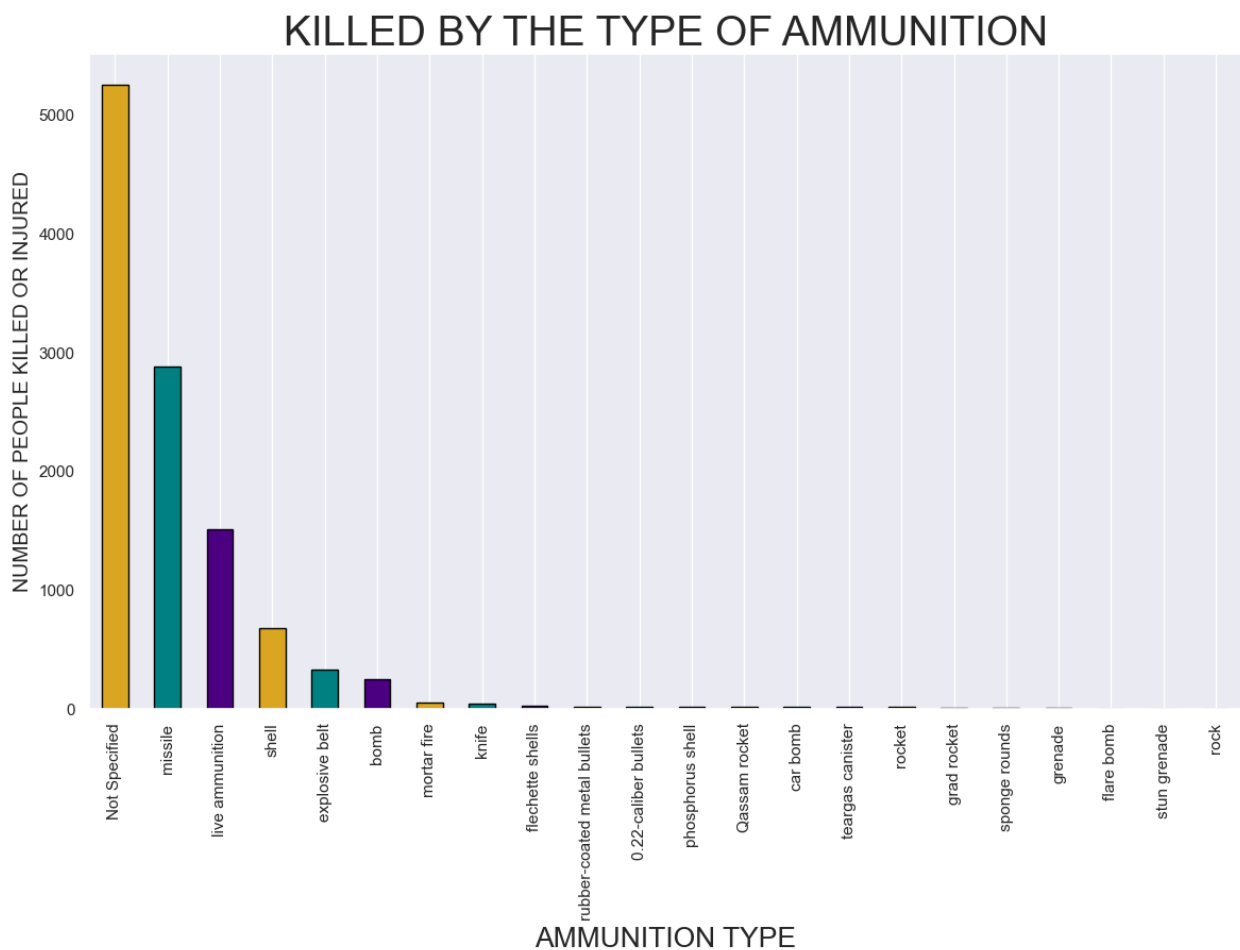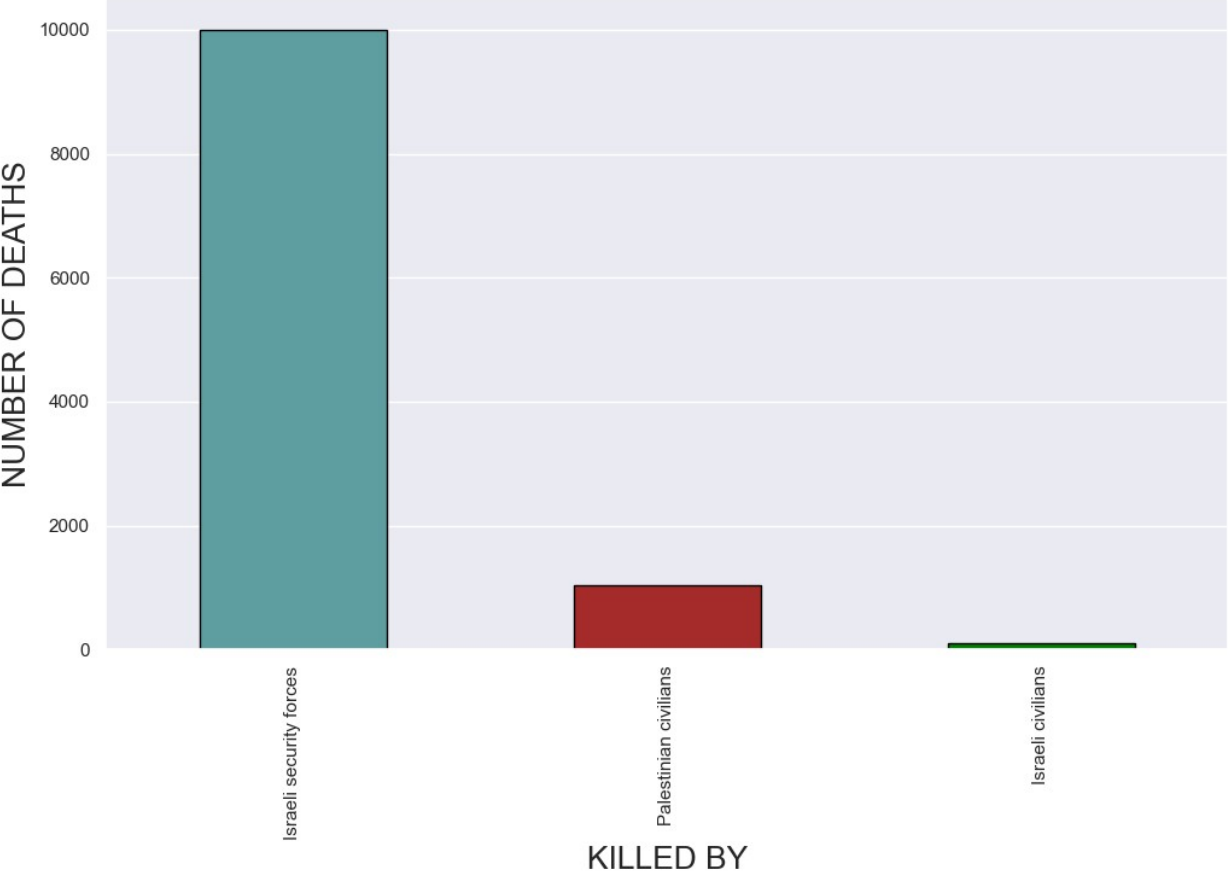
```
plt.title('NUMBER OF PEOPLE KILLED BY',fontsize=28)
plt.xlabel('KILLED BY',fontsize=19)
plt.ylabel('NUMBER OF DEATHS',fontsize=19)
plt.grid(axis='x')
plt.show()


sns.set(style="whitegrid")
plt.figure(figsize=(14, 8))
sns.countplot(x='killed_by', hue='ammunition', data=file,
palette='bright', edgecolor='black')
plt.title('PEOPLE WHO KILLED AND AMMUNITION THEY USED',fontsize=28)
plt.xlabel('KILLED BY',fontsize=19)
plt.ylabel('NUMBER OF DEATHS',fontsize=19)
plt.legend(title='Ammunition',fontsize=9)
plt.grid(axis='x')
plt.show()
```
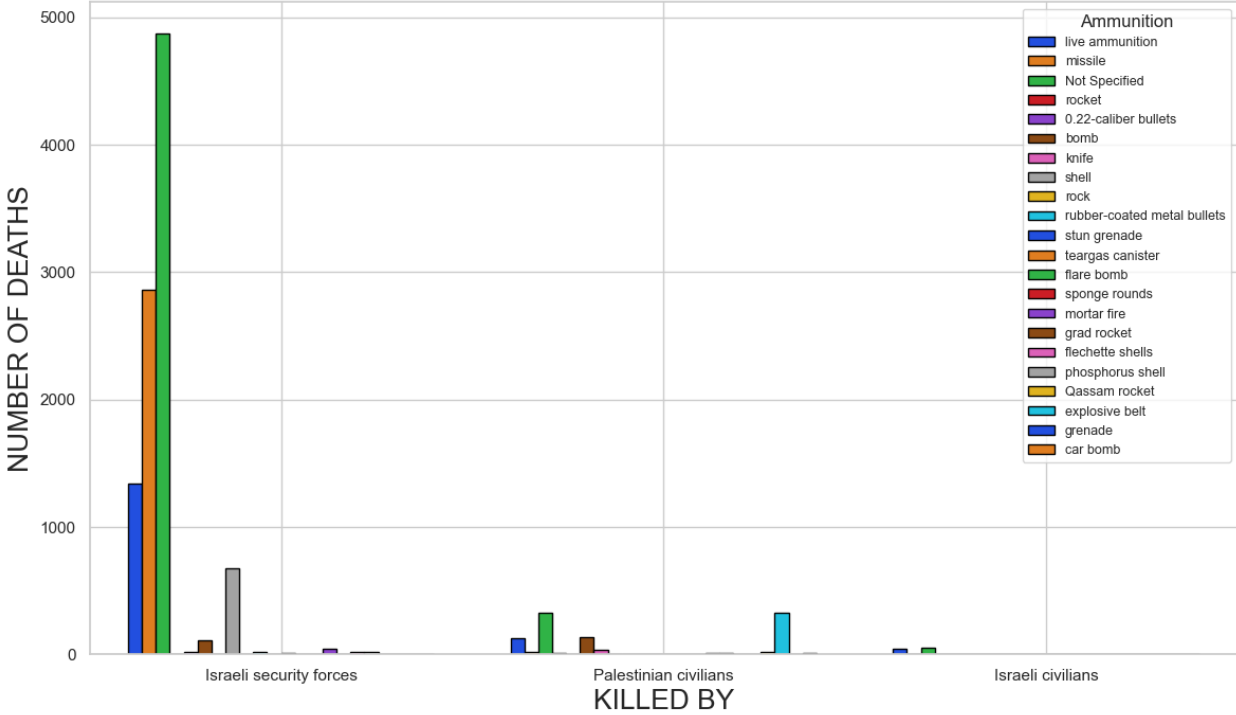

KILLED BY THE TYPE OF AMMUNITION

NUMBER OF PEOPLE KILLED BY

NUMBER OF DEATHS

KILLED BY

Israeli security forces

Palestinian civilians

Israeli civilians

PEOPLE WHO KILLED AND AMMUNITION THEY USED

Q5

In first graph The ammuniation which is mostly used is Not specified because it was not in our data set so we cleaned the data and instead of NULL we added 'Not specified' , after that most injuries are by missiles and live ammuniation

In second graph we can observe that most people are killed by the "ISRAELI SECURITY FORCES" , they have killed "10,000" people which is a huge number and the data set we have gotten is of 1100+ people and out of that 10,000 people are killed by them

In third graph what Israeli security forces mostly used are missiles and live ammuniation

from all these observation we have understood that the "ISRAELI SECURITY FORCES" are behind the killing of 10,000 and they have used Missiles and live ammuniation

```python
#Q6. Create profiles of the victims based on the available data such as age, gender, citizenship,
#and place of residence. Identify common characteristics among the victims.
import pandas as pd
import matplotlib.pyplot as plt

plt.figure(figsize=(20, 20))

#HERE WE HAVE MADE PROFILING AND IN THESE SUBPLOTS WE CAN SEE THAT HOW
MANY DEATHS AND INJURY HAS HAPPENED ON THE BASIS OF , THEIR GENDER ,
AGE , THEIR RESIDENCE AND CITIZENSHIP

#IN GENDER WE OBSERVE THAT MOSTLY MALE ARE KILLED

plt.subplot(221)
file['gender'].value_counts().plot(kind='pie', autopct='%1.1f%%')
plt.title('Gender of Victims')
plt.xlabel('Gender')
plt.ylabel('')
```

```python
# THE VICTIMS AGE GRAPHS SHOWS US THAT MOSTLY YOUTH IS TARGETED FROM
AGE 15 TO 25
plt.subplot(222)
plt.hist(file['age'], bins=17, color='green', edgecolor='red')
plt.title('VICTIMS AGE')
plt.xlabel('AGE')
plt.ylabel('FREQUENCY')

# THE RESIDENCE WHICH IS EFFECTED THE MOST IS GAZA CITY WHICH IS
POPULATED BY THE PALESTENIANS
plt.subplot(223)
file['place_of_residence'].value_counts().head().plot(kind='bar',
color='blue', edgecolor='black')
plt.title("VICTIM'S RESIDENCE")
plt.xlabel('PLACE OF RESIDENCE')
plt.ylabel('NUMBER OF PEOPLE')

# THE CITIZENS WHICH ARE MOSTLY INJURED OR KILLED ARE FROM PALESTINE
plt.subplot(224)
file['citizenship'].value_counts().plot(kind='bar', color='red',
edgecolor='black')
plt.title('CITIZENSHIP OF VICTIMS')
plt.xlabel('CITIZENSHIP')
plt.ylabel('NUMBER OF PEOPLE')

# AGE GROUPS BASED ON 10-YEAR INTERVALS
plt.figure(figsize=(10, 6))
age_bins = [0, 25, 35, 45, 55, 65, 75, 85, 95, 105, 115]
age_labels = ['0-25', '25-35', '35-45', '45-55', '55-65', '65-75',
'75-85', '85-95', '95-105', '105-115']
file['age_group'] = pd.cut(file['age'], bins=age_bins,
labels=age_labels)
age_group_counts = file['age_group'].value_counts()
sns.set_theme(style="darkgrid")
sns.barplot(x=age_group_counts.index, y=age_group_counts.values,
palette='dark', order=age_labels)
plt.title('Bar Chart of Age Groups')
plt.xlabel('Age Group')
plt.ylabel('Count')

# CITIZENSHIP BY GENDER
plt.figure(figsize=(10, 6))
sns.set_theme(style="whitegrid")
sns.countplot(x='citizenship', hue='gender', data=file,
palette='bright')
plt.title('Count Plot of Citizenship by Gender')
plt.xlabel('Citizenship')
plt.ylabel('Count')
plt.legend(title='Gender')
```
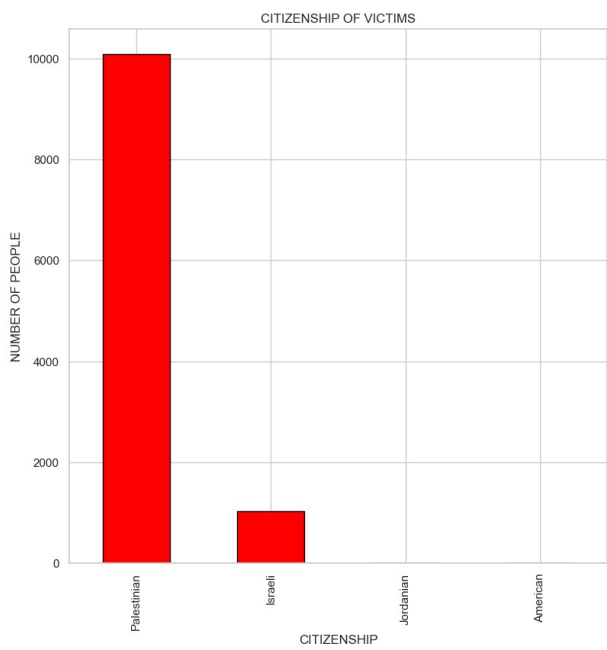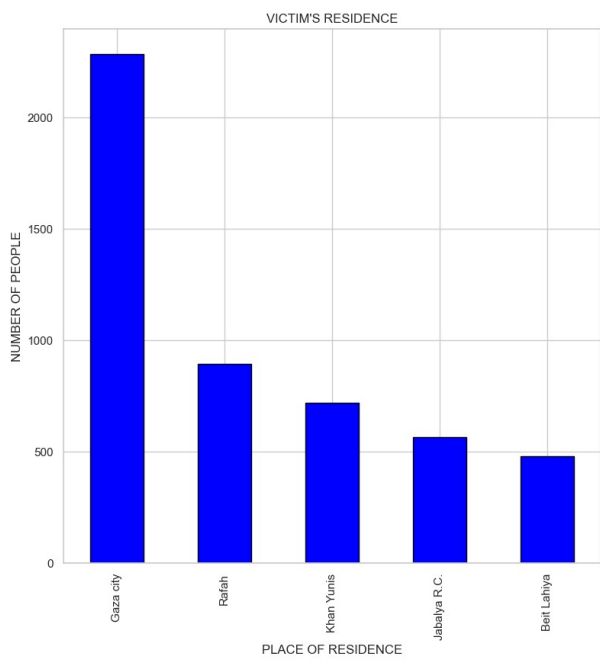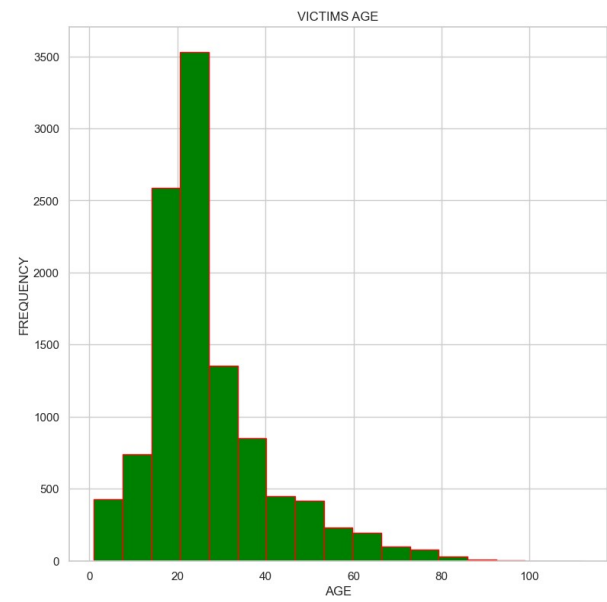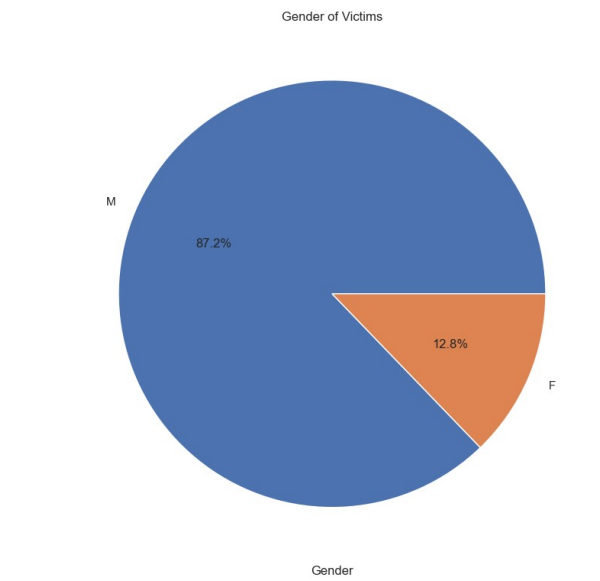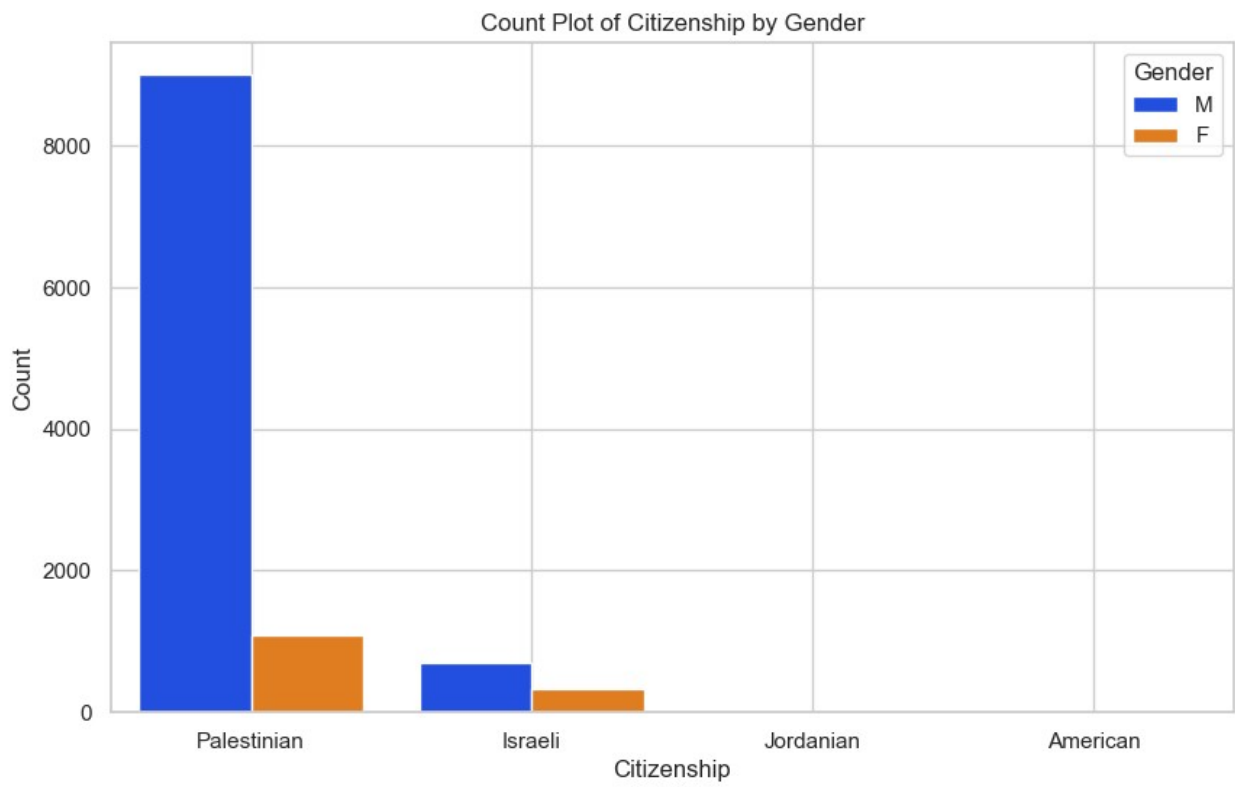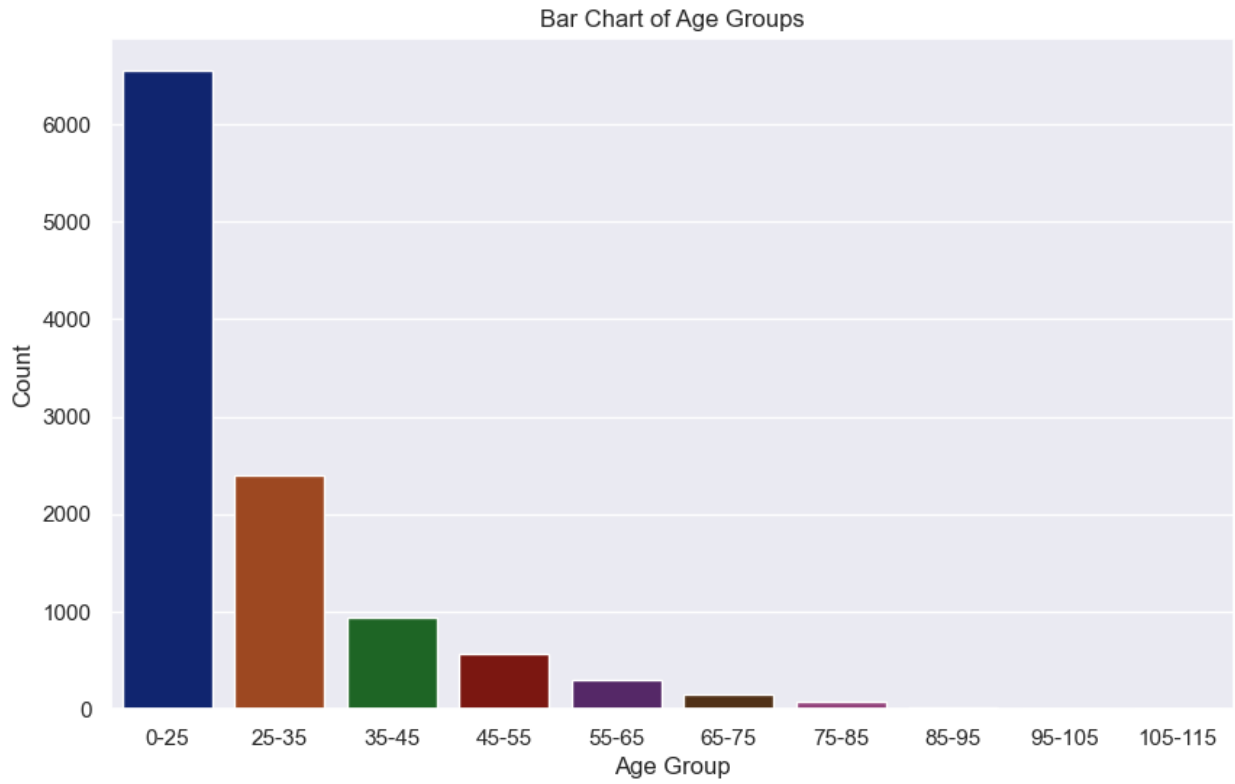
```
plt.grid(axis='x')

plt.show()
```

Bar Chart of Age Groups

Count Plot of Citizenship by Gender

Q 6

In first pie chart grah we observe that the male death percenatge is 87.2% which is really high and female death ratio is 12.8%

In the second graph of bar chart we have observed that mostly the youth is targeted which is from the age of 20 to 25 and their death number is above 3000 , so they can finish the future of a nation

In third graph we can see that mostly deaths are taking place in Gaza city and after that Rafah which are populated by the Palestinans and the death toll is above 2000

In forth graph we can see that the citizens which are injured the most are Palestinians

In fifth graph I have made the groups of age and the group which is most targeted is the youth from age 0-25 and after that 25-35 , so they can finish the future .

In sixth graph we can observe mostly Palestinans are killed and specifically their males