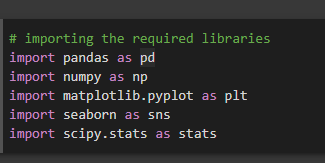
**MDSC-102P-ESE-FINAL LAB REPORT**

***Dataset :*** *Fatalities in the Israeli-Palestinian* - This dataset records fatalities in the Israeli-Palestinian conflict from 2000 to 2023.

* We start by importing the required libraries:



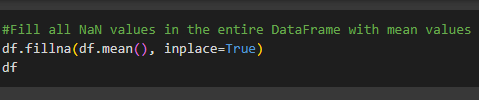
* Reading the dataset. There are a total of 16 features in the dataset.



The features are **:** name**,** date\_of\_event**,** age**,** citizenship**,** event\_location**,** event\_location\_district**,** event\_location\_region**,** date\_of\_death**,** gender**,** took\_part\_in\_the\_hostilities**,** place\_of\_residence**,** place\_of\_residence\_district**,** type\_of\_injury**,** ammunition**,** killed\_by**,** notes.

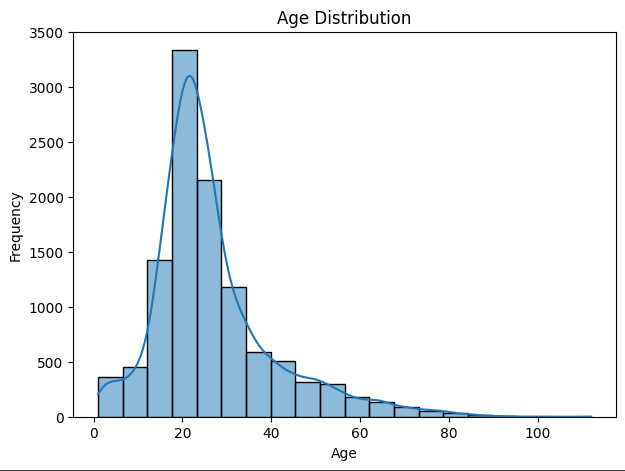
***Data Pre-processing:***

* Filling all the missing values with the mean of their respective columns.



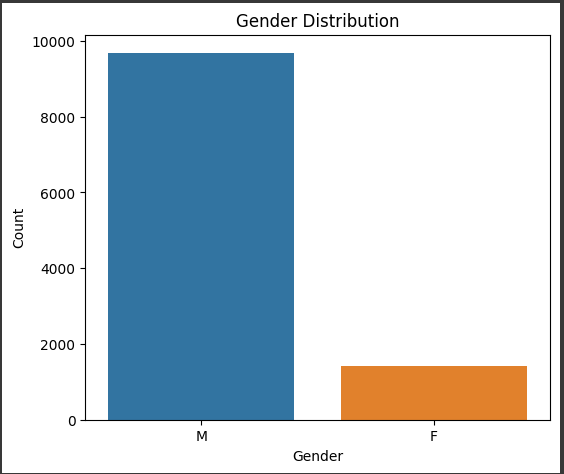
***Data Visualization:***

1. Age distribution of the victims-



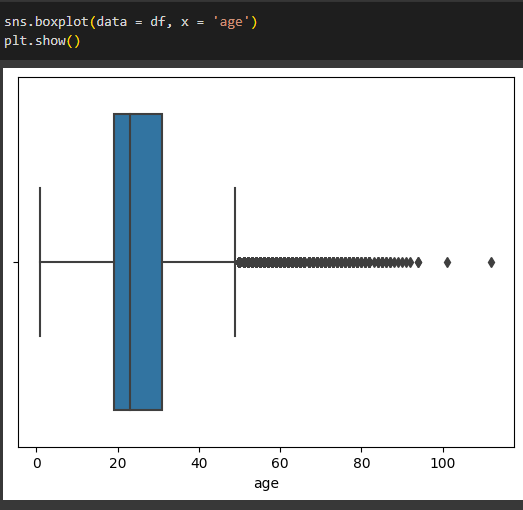
It is evident that a large portion of the casualties were around the age group of 15 to 35. This means that most of the killings/murders happened in places, where the youth gathered more. The age of individuals in the dataset varies widely, with a mean age of approximately 26.44 years. The majority of individuals fall in the 19-31 age range.

1. Gender distribution of casualties-



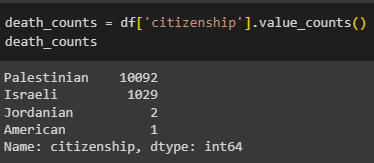
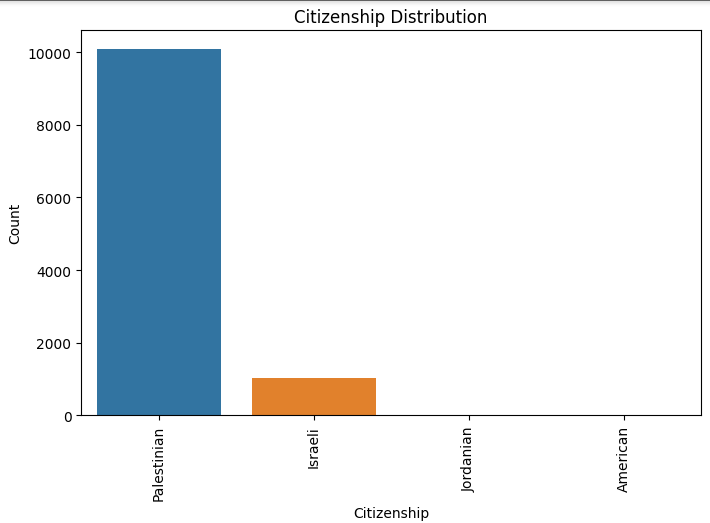
Exceptionally large number of males were slaughtered/killed compared to females. The dataset is skewed toward males, with a higher number of male casualties compared to females.

1. Finding skewness in age of the casualties-



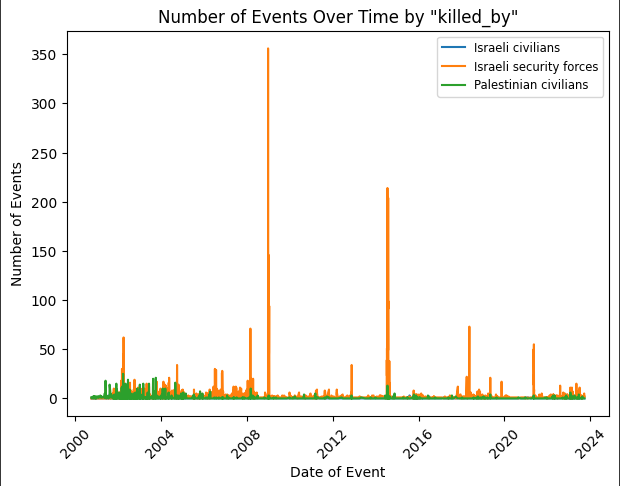
The boxplot of age data is right-skewed, indicating that most individuals in the dataset are younger, with a few older individuals considered outliers. The median age is relatively lower.

1. Citizenship-

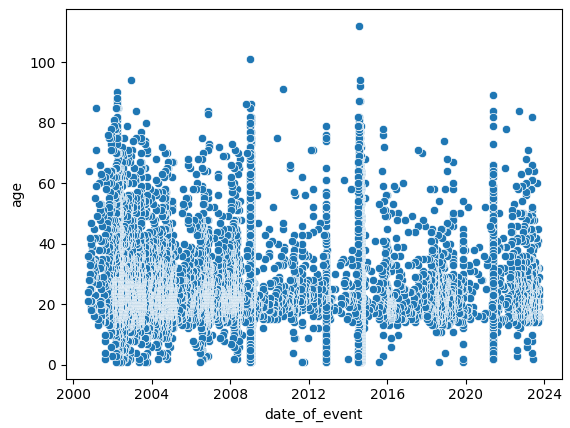
Palestinian citizens outnumber Israeli citizens in the dataset, indicating a higher number of Palestinian casualties. Jordanian and American citizens are very less affected compared to the others.

1. Calculating count of events for each combination-



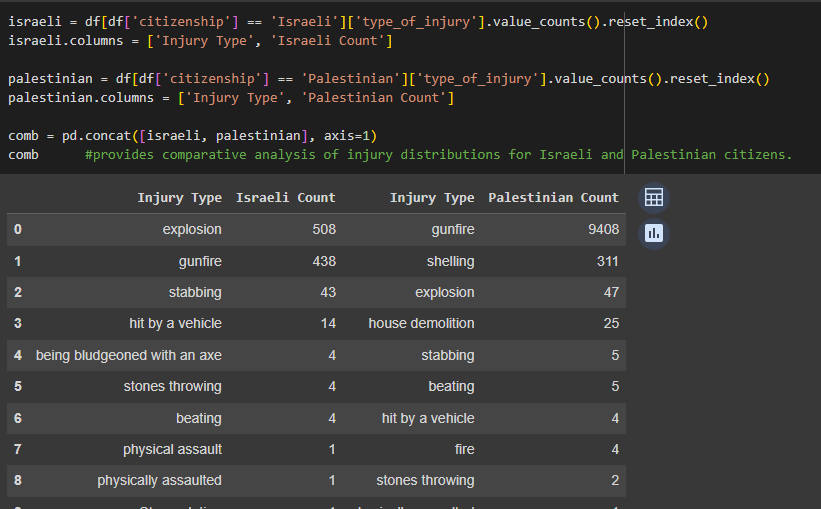
Majority of the victims were killed by Israeli security forces and then by Palestinian civilians. And comparatively less number of people were killed by Israeli civilians. The number of events involving Israeli security forces appears to increase in 2008-2009 and again in 2015.

1. Scatterplot for age of casualties-



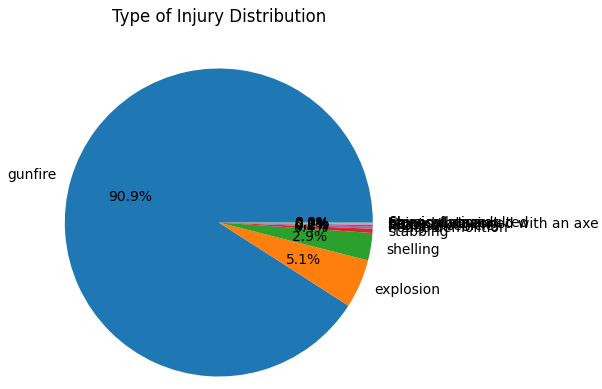
Major portion of casualties were between the age groups of 15 and 40 years. Even many senior citizens were killed. There was a huge massacre between the years 2002 and 2009. Later they reduced comparatively, and were again increased after 2022.

1. Concatenating two fields for comparison -



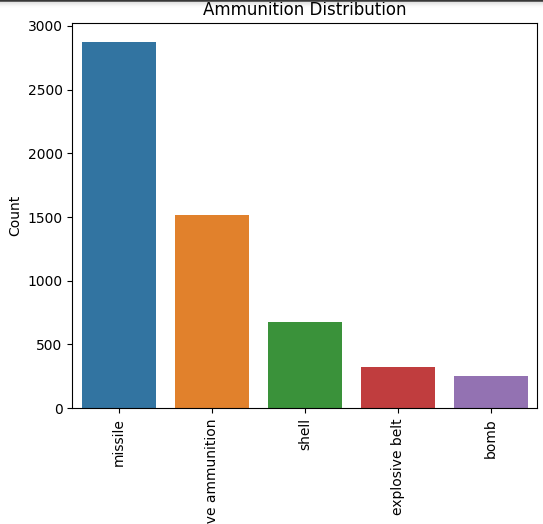
Provides comparative analysis of injury distributions for Israeli and Palestinian citizens.

1. Pie Plot for Type of injury -



A very large portion of casualties were injured by gunfire, and then by explosion and shelling and stabbings.

1. Ammunition Distribution -



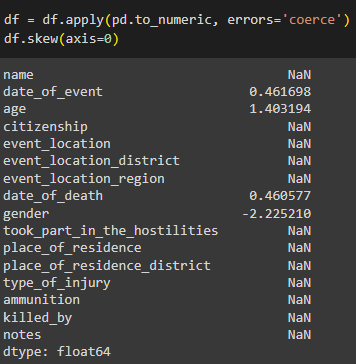
Here are the top five ammunition used for killing. Exceptionally large count of missiles were used for killing. Even live ammunition is used a lot to kill the civilians. Shells, explosive belts and bombs also contributed well to the murders. The majority of events involve unidentified weapons, followed by Qassam rockets and bombs. In 2015, Grad rockets and missiles were also used.

***Testing Hypothesis :***

1. Converting values of gender to 1’s and 0’s -

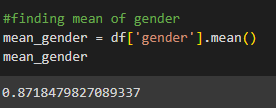
****

1. Converts all columns to numeric type and then calculates skewness -

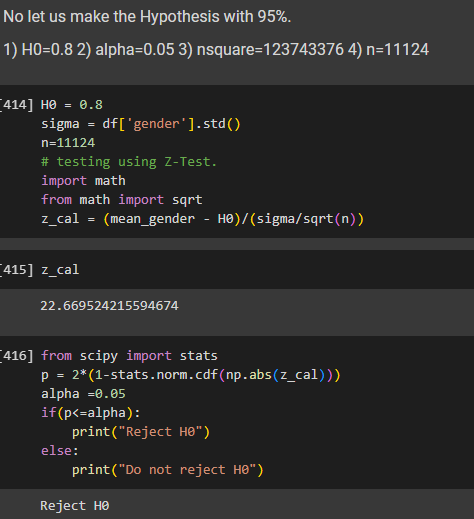
****

***Z - Test -***

1. Finding mean of gender -



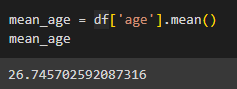
1. Calculating z-value and then testing hypothesis -



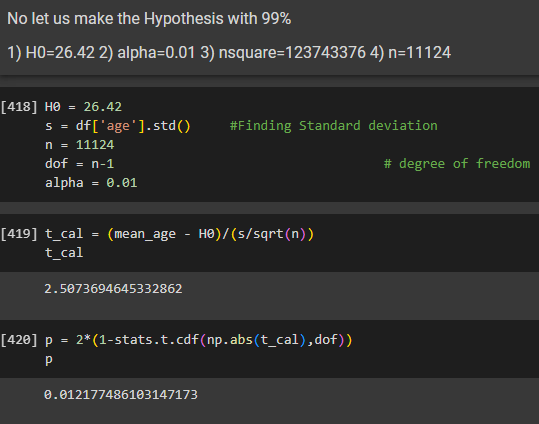
Therefore it suggests that the sample mean of the 'gender' significantly differs from 0.8 and it is evident to reject the null hypothesis at the 0.05 significance level.

***T - Test -***

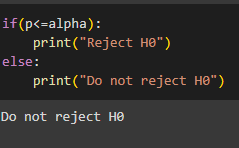
1. Finding mean of age -



1. Calculating t-value -



1. Testing hypothesis -



Therefore, the inference is that there is not enough evidence to conclude that the population mean of 'age' is different from 26.42 within the given sample, at a 99% confidence level and hence it is evident not to reject the null hypothesis at the 0.01 significance level.

***Inferences:***

* The age of individuals in the dataset varies widely, with a mean age of approximately 26.74 years. The majority of individuals fall in the 19-31 age range.
* The dataset is skewed toward males, with a higher number of male casualties compared to females.
* Palestinian citizens outnumber Israeli citizens in the dataset, indicating a higher number of Palestinian casualties.
* Over 4000 individuals were not involved in hostilities, while more than 3000 but less than 4000 were involved. Approximately 1000 casualties were Israeli, and nearly 200 were targeted killings. Another 2000 are unidentified.
* Event counts were notably high in 2008-2009 and again in 2015-2016.
* The number of events involving Israeli security forces appears to increase in 2008-2009 and again in 2015.
* The majority of events involve unidentified weapons, followed by Qassam rockets and bombs. In 2015, Grad rockets and missiles were also used.
* The top 5 event locations are Gaza City, Rafah, Khan Yunis, Jabalya RC, and Beit Lahiya, with Gaza City having the highest count.
* The distribution of place of residence mirrors the event locations, with Gaza City being prominent.
* Israeli security forces are responsible for a significant number of casualties, while Palestinian civilians are responsible for fewer casualties.

**—SAIRAM—**