# Fatalities in the Israeli-Palestinian Conflict from 2000 - 2023

**Rahardiansyah F.**
**Milestone 1 RMT-027**

***Technology / Tools: Python, Pandas, NumPy, Matplotlib, SciPy, Tableau.***

# Identifying Problems with SMART Framework

HACKTIV8

In light of the ongoing Israeli-Palestinian Conflict, a comprehensive analysis is essential to gain meaningful insights into the dynamics of the conflict. With that said, this analysis will focus three key aspects: demographic, geospatial, and fatality trends.

- **Specific**: The analysis will include demographic. geospatial and `fatality trend
- **Measurable**: Gained insight on correlation between variables mentioned earlier.
- **Achievable**: The dataset contains 11124 entries that is going to help wth the analysis.
- **Relevant**: An analysis is now more relevant than ever to understand the current ongoing conflict.
- **Time-bound**: Data analysis will be done in 2 days.

# 5W + 1H

1. **What** is the mostly used weapon in the conflict?
2. **Where** is the district region with the most conflict?
3. **Who** is the the most affected by the conflict based on age and gender?
4. **Who** is the most responsible for the fatalities based on citizenship?
5. **When** is there spikes or increase in fatalities in the conflict?

## Input:

```
df = pd.read_csv('https://raw.githubusercontent.com/rahardianfatoni/datasets/main/fatalities_isr_pse_conflict_2000_to_2023.csv')
```

```
df.head(15)
```

## Output:

| date_of_event | age | citizenship | event_location | event_location_district | event_location_region | gender | type_of_injury | ammunition | killed_by | datetime |
|---|---|---|---|---|---|---|---|---|---|---|
| 2023-09-24 | 32.0 | Palestinian | Nur Shams R.C. | Tulkarm | West Bank | M | gunfire | live ammunition | Israeli security forces | 2023-09-24 00:00:00 |
| 2023-09-24 | 21.0 | Palestinian | Nur Shams R.C. | Tulkarm | West Bank | M | gunfire | live ammunition | Israeli security forces | 2023-09-24 00:00:00 |
| 2023-09-22 | 16.0 | Palestinian | Kfar Dan | Jenin | West Bank | M | gunfire | live ammunition | Israeli security forces | 2023-09-22 00:00:00 |
| 2023-09-20 | 19.0 | Palestinian | 'Aqbat Jaber R.C. | Jericho | West Bank | M | gunfire | live ammunition | Israeli security forces | 2023-09-20 00:00:00 |
| 2023-09-19 | 15.0 | Palestinian | Jenin R.C. | Jenin | West Bank | M | gunfire | live ammunition | Israeli security forces | 2023-09-19 00:00:00 |
| 2023-09-19 | 29.0 | Palestinian | Jenin R.C. | Jenin | West Bank | M | gunfire | missile | Israeli security forces | 2023-09-19 00:00:00 |
| 2023-09-19 | 24.0 | Palestinian | Gaza City | Gaza | Gaza Strip | M | gunfire | live ammunition | Israeli security forces | 2023-09-19 00:00:00 |
| 2023-09-19 | 25.0 | Palestinian | Jenin R.C. | Jenin | West Bank | M | gunfire | missile | Israeli security forces | 2023-09-19 00:00:00 |
| 2023-09-19 | 23.0 | Palestinian | Jenin R.C. | Jenin | West Bank | M | gunfire | missile | Israeli security forces | 2023-09-19 00:00:00 |
| 2023-09-09 | 15.0 | Palestinian | al-'Arrub R.C. | Hebron | West Bank | M | gunfire | live ammunition | Israeli security forces | 2023-09-09 00:00:00 |
| 2023-09-05 | 16.0 | Palestinian | Argaman | Jericho | West Bank | M | gunfire | live ammunition | Israeli security forces | 2023-09-05 00:00:00 |
| 2023-09-05 | 21.0 | Palestinian | Nur Shams R.C. | Tulkarm | West Bank | M | gunfire | live ammunition | Israeli security forces | 2023-09-05 00:00:00 |
| 2023-09-01 | 45.0 | Palestinian | 'Aqqabah | Tubas | West Bank | M | gunfire | live ammunition | Israeli security forces | 2023-09-01 00:00:00 |
| 2023-08-31 | 41.0 | Palestinian | Beit Sira | Ramallah and al-Bira | West Bank | M | gunfire | live ammunition | Israeli security forces | 2023-08-31 00:00:00 |
| 2023-08-30 | 14.0 | Palestinian | al-Musrarah | East Jerusalem | West Bank | M | stabbing | live ammunition | Israeli security forces | 2023-08-30 00:00:00 |

HACKTIV8

## Input:

```
df.info()
```

## Output:

```
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
```

**Insight**:

- There are 11124 total entries to the dataset, with 15 columns consisting of string values except “age” column.
- The columns that will be used for analysis: `**date_of_event**`, `**age**`, `**citizenship**`, `**event_location**`(s), `**gender**`, `**ammunition**`, and `**killed_by**`. Hence, to minimize cluttering, we will drop the unused columns.
- We shall also remove the null values found in `age` and `type_of_injury` columns. The null values in the `ammunition` column will be assumed as intentional.

# Columns Explanation

- DATA FILTERING AND CLEANING

| Column | Dtype | Explanation |
| --- | --- | --- |
| **date_of_event** | object | Date of the event in YYYY-mm-dd format |
| **age** | float64 | Age when subject perished in Years |
| **citizenship** | object | Citizenship of subject |
| **event_location** | object | Location of event, divided into 3 sub-categories |
| **gender** | object | Gender of subject (M/F) |
| **killed_by** | object | Perpetrator of fatality |

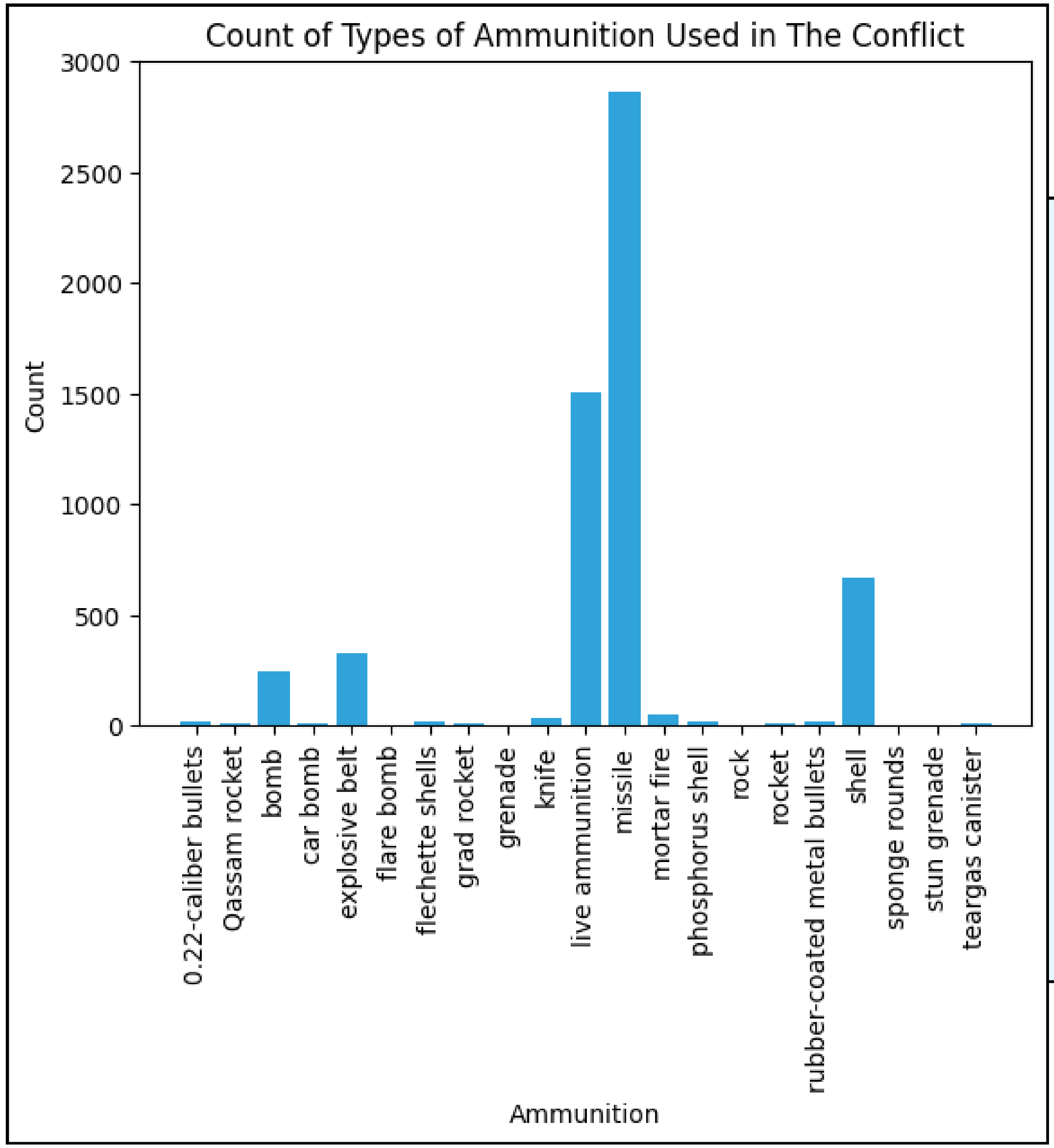
# What is the mostly used weapon in the conflict?

**Input:**

```
df_ammunition =
df.groupby(['ammunition']).size().reset_inde
x(name='count')
```

```
plt.bar(df_ammunition['ammunition'],
df_ammunition['count'], color = "#30A3DA")
plt.xticks(rotation=90)
plt.xlabel('Ammunition')
plt.ylabel('Count')
plt.title('Count of Types of Ammunition
Used in The Conflict')
plt.show()
```

**Output:**




- **Primary Ammunition Type**: The insight indicates that missiles are the main ammunition used in the conflict. This implies that the conflict involves advanced weaponry capable of long-range attacks. Missiles are generally sophisticated and can be employed in various military scenarios, including air-to-ground or ground-to-ground operations.
- **Frequency of Use**: The graph shows that the number of occurrences of missile usage (2861) almost doubles that of live ammunition (1510). This suggests that missiles are extensively employed in the conflict compared to live ammunition.

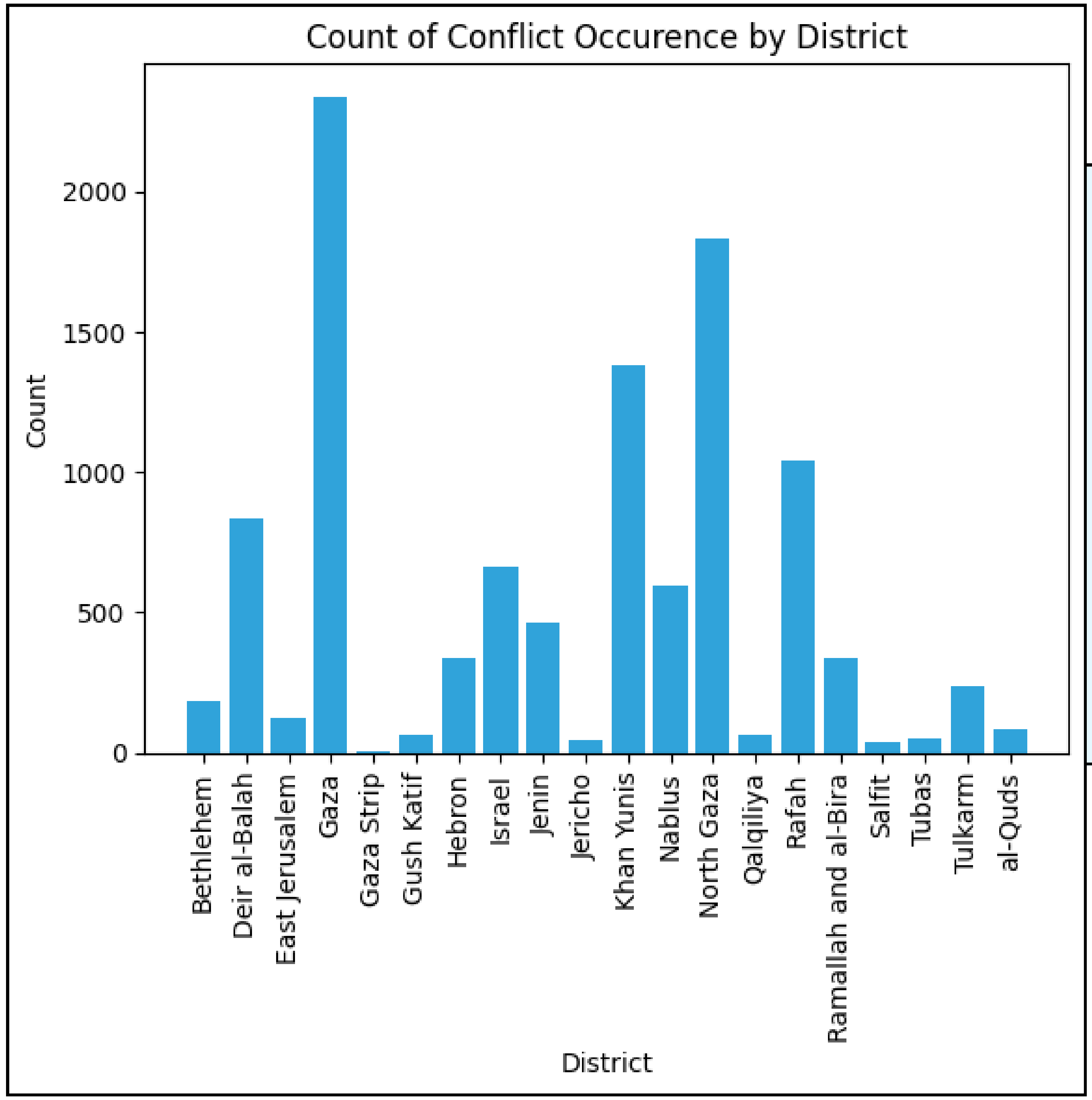
# Where is the district region with the most conflict?

**Input:**

```
df_ammunition =
df.groupby(['event_location_district']).size().
reset_index(name='count')
```

```
plt.bar(df_ammunition['ammunition'],
df_ammunition['count'], color = "#30A3DA")
plt.xticks(rotation=90)
plt.xlabel('District')
plt.ylabel('Count')
plt.title('Count of Conflict Occurence by
District')
plt.show()
```

**Output:**




- Based on the graph, we can conclude that most of the conflict happens in Palestinian territory, with the 3 most occurences happening in **Gaza**, **North Gaza** and **Khan Yunis** by order of frequency.

# Who is the the most affected by the conflict based on age and gender?

## Input:

```
print(f"Mean age : {df['age'].mean()}")
print(f"Median age : {df['age'].median()}")
print(f"Mode age : {df['age'].mode()[0]}")
print(f"Standar deviation : {df['age'].std()}")
print(f"Skewness age: {df['age'].skew()}")
```

```
print(f"Number of women affected by the
conflict : {df['gender'][df['gender'] ==
'F'].count()}")

print(f"Number of men affected by the
conflict : {df['gender'][df['gender'] ==
'M'].count()}")
```

## Output:

- Mean age : 26.69143177151391
- Median age : 23.0
- Mode age : 22.0
- Standar deviation : 13.73911107392046
- Skewness age: 1.3853387554332095
- Number of women affected by the conflict : 1369
- Number of men affected by the conflict : 9345

- **Mean Age** (26 years old): The mean age provides an average value and is influenced by extreme values. In this case, the mean age of 26 suggests that, on average, the population affected by the conflict is relatively young.
- **Median Age** (23 years old): The median represents the middle point of the dataset when arranged in ascending order. The fact that the median age (23) is lower than the mean suggests that there are some higher age values pulling the mean up, indicating a positively skewed distribution.
- **Mode Age** (22 years old): The mode is the age that appears most frequently in the dataset. In this case, 22 years old is the most common age among the affected population.
- **Positive Skew** (Skewness over 1): A positive skewness indicates that the distribution of ages is skewed to the right, meaning there are relatively more younger individuals in the population than older ones. The tail of the distribution extends towards higher ages.
- Finally, men are mostly affected by the conflict (~87 %): This coincides with the assumption that men are more likely to be sent or join the conflict voluntarily. Further analysis could be done since Israel also deploy female soldiers meanwhile the Palestinian resistance do not, we shall do a further analysis on whether there is a significant difference between gender and citizenship on the fatalities.

# Conducting a Chi-Squared Hypothesis Test

**H0** → There is no relationship between Gender and Citizenship (Men and Women are the same).

**H1** → There is a relationship between Gender and Citizenship (Men and Women are different).

# Conducting a Chi-Squared Hypothesis Test

**Input:**

```
contingency_table =
pd.crosstab(df['gender'],df['citizenship'])
contingency_table
```

```
res =
stats.chi2_contingency(contingency_table)
print("P-value:",res.pvalue)
```

**Output:**

- Critical value: 0.05
- p-val = 1.8579299277363706e-81 < critical value
- H0 rejected
- Conclusion:

**There is a relationship between gender and citizenship.**

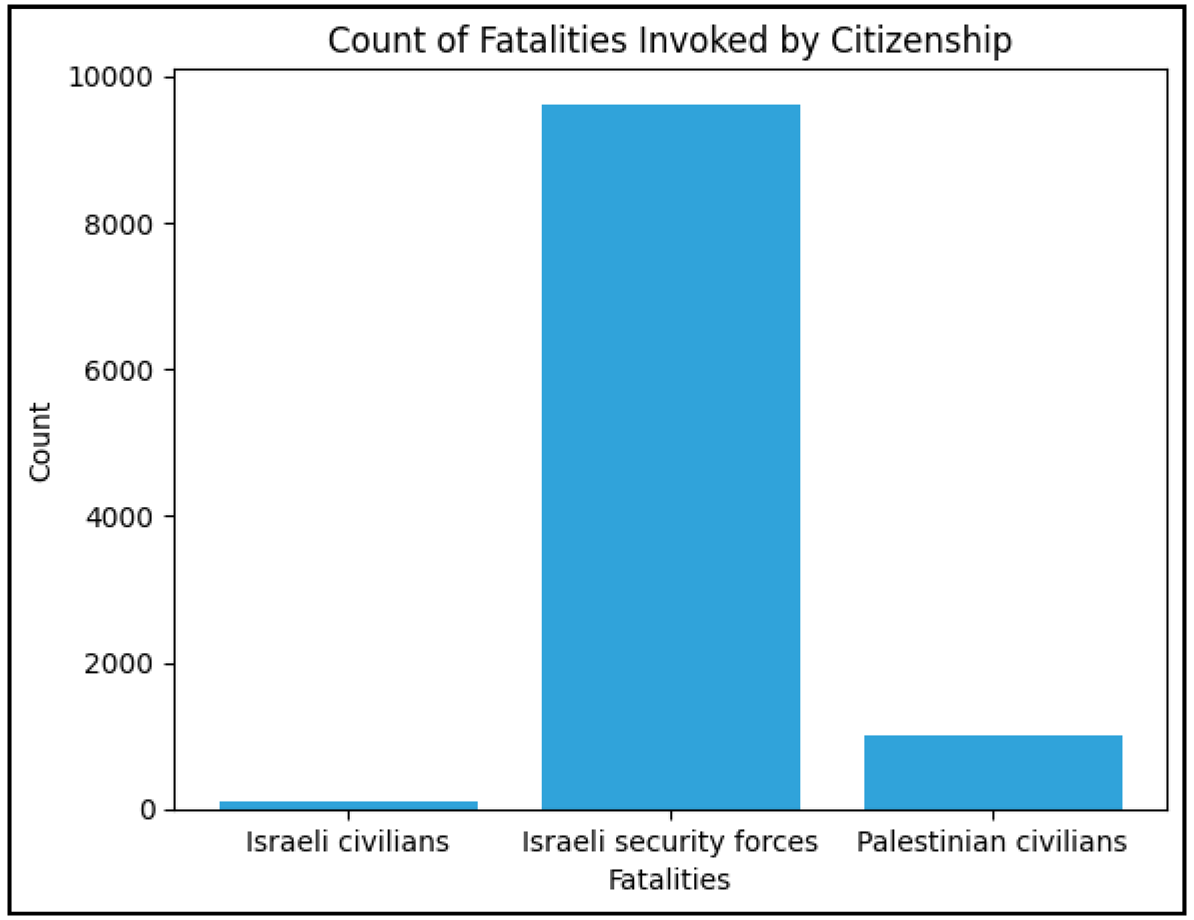
# Who is the most responsible for the fatalities based on citizenship?

**Input:**

```
df_killed_by =
df.groupby(['killed_by
']).size().reset_index(name='count')
```

```
plt.bar(df_killed_by['killed_by'],
df_killed_by['count'], color = "#30A3DA")
plt.xlabel('Fatalities')
plt.ylabel('Count')
plt.title('Count of Fatalities Invoked by
Citizenship')
plt.show()
```

**Output:**




- **The Israeli Security Forces is responsible for 9615 or about ~89% of the fatalities that happened in the conflict.** The responsibility for a large proportion of fatalities by the Israeli Security Forces raises questions about the nature of their involvement, the tactics employed, and the potential impact on civilian populations.

HACKTIV8

# When is there spikes or increase in fatalities in the conflict?

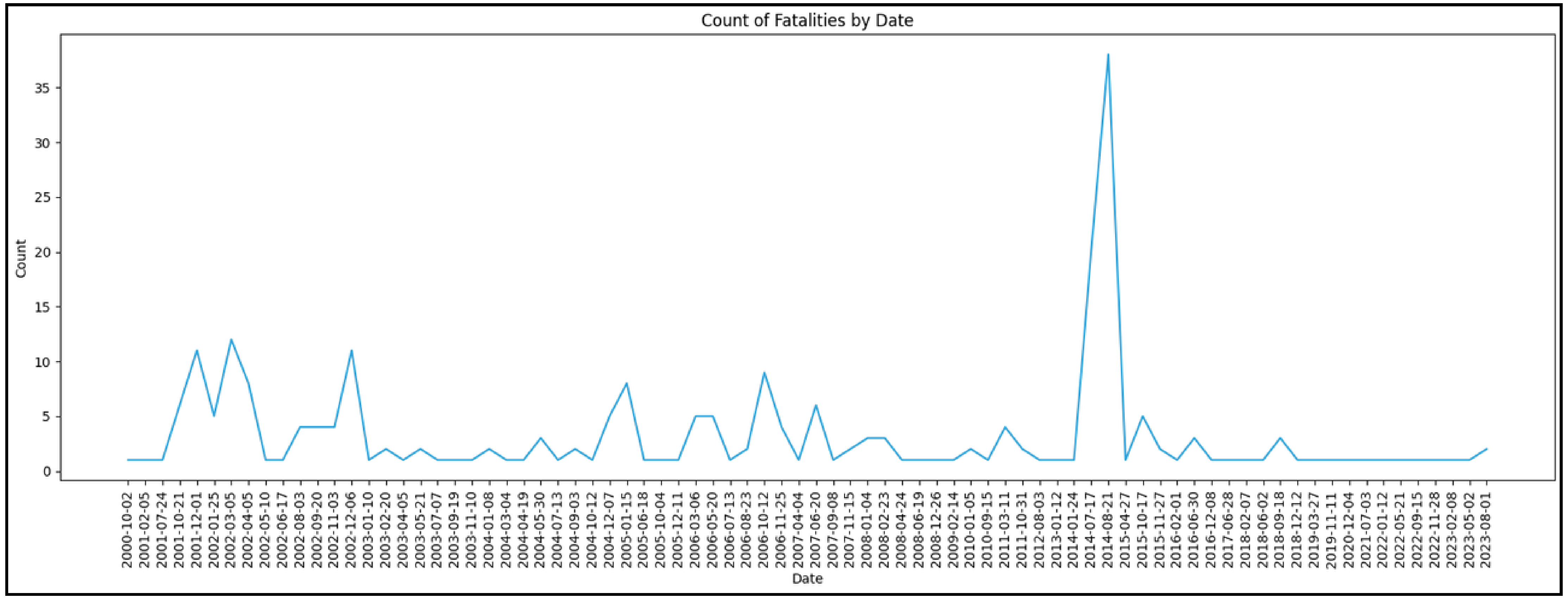
## Input:

```
df['datetime'] =
pd.to_datetime(df['date_of_event'])

df_date =
df.groupby(['date_of_event']).size()
.reset_index(name='count')
```

```
plt.figure(figsize=(20, 6))
plt.plot(df_date['date_of_event'][::30], df_date['count'][::30], color = "#30A3DA")
plt.xticks(df_date['date_of_event'][::30], rotation = 90)
plt.xlabel('Date')
plt.ylabel('Count')
plt.title('Count of Fatalities by Date')
plt.show()
```

## Output:




- The graph identifies one distinct spike in conflict occurrences: between **24th of January 2014 and 27th of April 2015.** These, along with the many smaller spikes indicate the long on-going conflict over the period of 23 years in Palestine.

# Conclusion

- **Fatalities by Demographic Patterns**:

The affected population has a predominantly young demographic, with a mean age of 26 and a mode at 22 years old. The positive skewness suggests a concentration of younger individuals, potentially indicating the impact on a specific age group.

- **Fatalities by Ammunition (weapon) Type**:

Missiles emerge as the primary ammunition used in the conflict, with a significant increase in occurrences compared to live ammunition. This points towards the likelihood of a long-distance artillery clash.

- **Fatalities by District**:

Most conflict occurrences are concentrated in Palestinian territory, with Gaza, North Gaza, and Khan Yunis experiencing the highest frequencies.

- **Fatalities by Citizenship**:

The Israeli Security Forces are responsible for a substantial portion (~89%) of the fatalities, emphasizing their significant role in the conflict.

- **Fatalities by Date**:

Two notable spikes in conflict occurrences are identified, one between 2008-12-26 and 2009-02-14 and another between 2014-07-17 and 2014-08-21. These spikes suggest periods of intensified conflict activity, demanding a closer examination of events and triggers during these timeframes.

# Thank you!

HACKTIV8 | CO