Name: Yousef Alaa Elden Shaker Turky

Track: Data engineering

Location: Ismailia

Project: EDA of terrorism dataset using python

# Preprocessing

**Selecting relevant columns**: Only the important columns related to the analysis were retained in the Data Frame.

**Handling missing values**: Missing values in the 'nkilled', 'nwounded', 'provstate', 'city', and 'motive' using median for numeric coulmns and The word ‘Unknown’ for non-numeric values.

**Checking for null values**: The Data Frame was checked for null values, and it was found that there were no null values remaining after cleaning, filtering and replace values.

# Descriptive statistics

A screenshot of a black screen

Description automatically generatedA black and white screen with white text

Description automatically generatedCountries with most attacks Regions with most attacks

* Number - Cities with most attacks

A screenshot of a cell phone

Description automatically generated of attacks per year



* Kills per year - Wounds per year

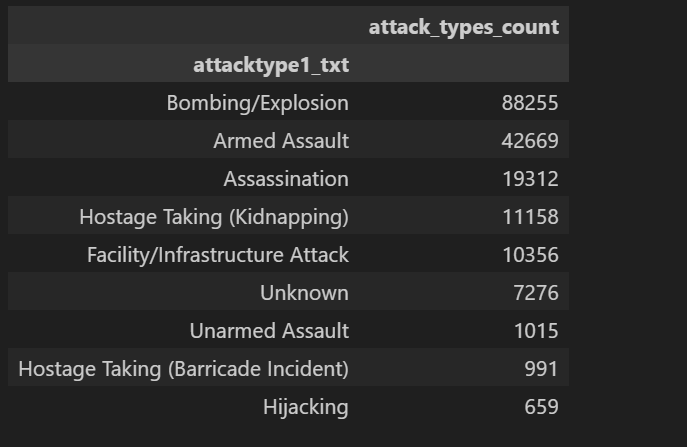
A screenshot of a cell phone

Description automatically generatedA screenshot of a black screen

Description automatically generated

* Groups with most attacks



* Most common attack types
* Most type of people affected from terrorists attacks

A screenshot of a black and white screen

Description automatically generated

# Data Visualization

A graph with blue lines

Description automatically generated

Terrorism Attacks over the year

A graph of a number of attacks

Description automatically generatedAttack per region

A graph of the number of countries/regions with the number of attacks

Description automatically generated

Attack per country top 30 countries

A screenshot of a graph

Description automatically generated

Correlation between numeric columns

A graph of a scatter plot

Description automatically generated

The relationship between the number of casualties and the type of attack

* Insights derived from data:
  + Most attacks happen in the middle east region
  + Most attacks happen in Iraq country
  + Most attacks happen in Baghdad city
  + Group with most terrorist attack is Taliban
  + The most affected group from terrorism is Private Citizens & Property
  + Most used weapon is Explosives
  + Most common attack Types is Bombing/Explosion

Performance comparison between Dask and pandas :

* Reading data
  + Using Dask: 0.11 sec
  + Using pandas: 2.68 sec
* making group by region to get the sum
  + using Dask: 1.35 sec
  + Using Pandas: 4.57 sec