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

Countries with most attacks

	country_txt	attack_per_country
0	Iraq	24636
1	Pakistan	14368
2	Afghanistan	12731
3	India	11960
4	Colombia	8306
5	Philippines	6908
6	Peru	6096
7	El Salvador	5320
8	United Kingdom	5235
9	Turkey	4292
10	Somalia	4142
11	Nigeria	3907
12	Thailand	3849
13	Yemen	3347
14	Spain	3249

Regions with most attacks

	region_txt	attack_per_region
0	Middle East & North Africa	50474
1	South Asia	44974
2	South America	18978
3	Sub-Saharan Africa	17550
4	Western Europe	16639
5	Southeast Asia	12485
6	Central America & Caribbean	10344
7	Eastern Europe	5144
8	North America	3456
9	East Asia	802
10	Central Asia	563
11	Australasia & Oceania	282

- Number of attacks per year

iyear	Attacks_Per_year
2014	16903
2015	14965
2016	13587
2013	12036
2017	10900
2012	8522
2011	5076
1992	5071
2010	4826
2008	4805
2009	4721
1991	4683
1989	4324
1990	3887
1988	3721

- Cities with most attacks

	city	attack_per_city
1	Baghdad	7589
2	Karachi	2652
3	Lima	2359
4	Mosul	2265
5	Belfast	2171
6	Santiago	1621
7	Mogadishu	1581
8	San Salvador	1558
9	Istanbul	1048
10	Athens	1019
11	Bogota	984
12	Kirkuk	925
13	Beirut	918
14	Medellin	848
15	Benghazi	840

- Kills per year

iyear	kills_per_year
2014	44490.0
2015	38853.0
2016	34871.0
2017	26445.0
2013	22273.0
2012	15497.0
2007	12824.0
1997	10924.0
1984	10450.0
1992	9742.0
1983	9444.0
2006	9380.0
2009	9273.0
2008	9157.0
1991	8429.0

- Wounds per year

iyear	wounded_per_year
2015	44043.0
2014	41128.0
2016	40001.0
2013	37688.0
2012	25445.0
2017	24927.0
2001	22774.0
2007	22524.0
2009	19138.0
2008	18998.0
2010	15947.0
2006	15550.0
2011	14659.0
1995	14292.0
2005	12784.0

- Groups with most attacks

	gname	Group_with_most_attacks
1	Taliban	7478
2	Islamic State of Iraq and the Levant (ISIL)	5613
3	Shining Path (SL)	4555
4	Farabundo Marti National Liberation Front (FMLN)	3351
5	Al-Shabaab	3288
6	New People's Army (NPA)	2772
7	Irish Republican Army (IRA)	2671
8	Revolutionary Armed Forces of Colombia (FARC)	2487
9	Boko Haram	2418
10	Kurdistan Workers' Party (PKK)	2310
11	Basque Fatherland and Freedom (ETA)	2024
12	Communist Party of India - Maoist (CPI-Maoist)	1878
13	Maoists	1630
14	Liberation Tigers of Tamil Eelam (LTTE)	1606
15	National Liberation Army of Colombia (ELN)	1561

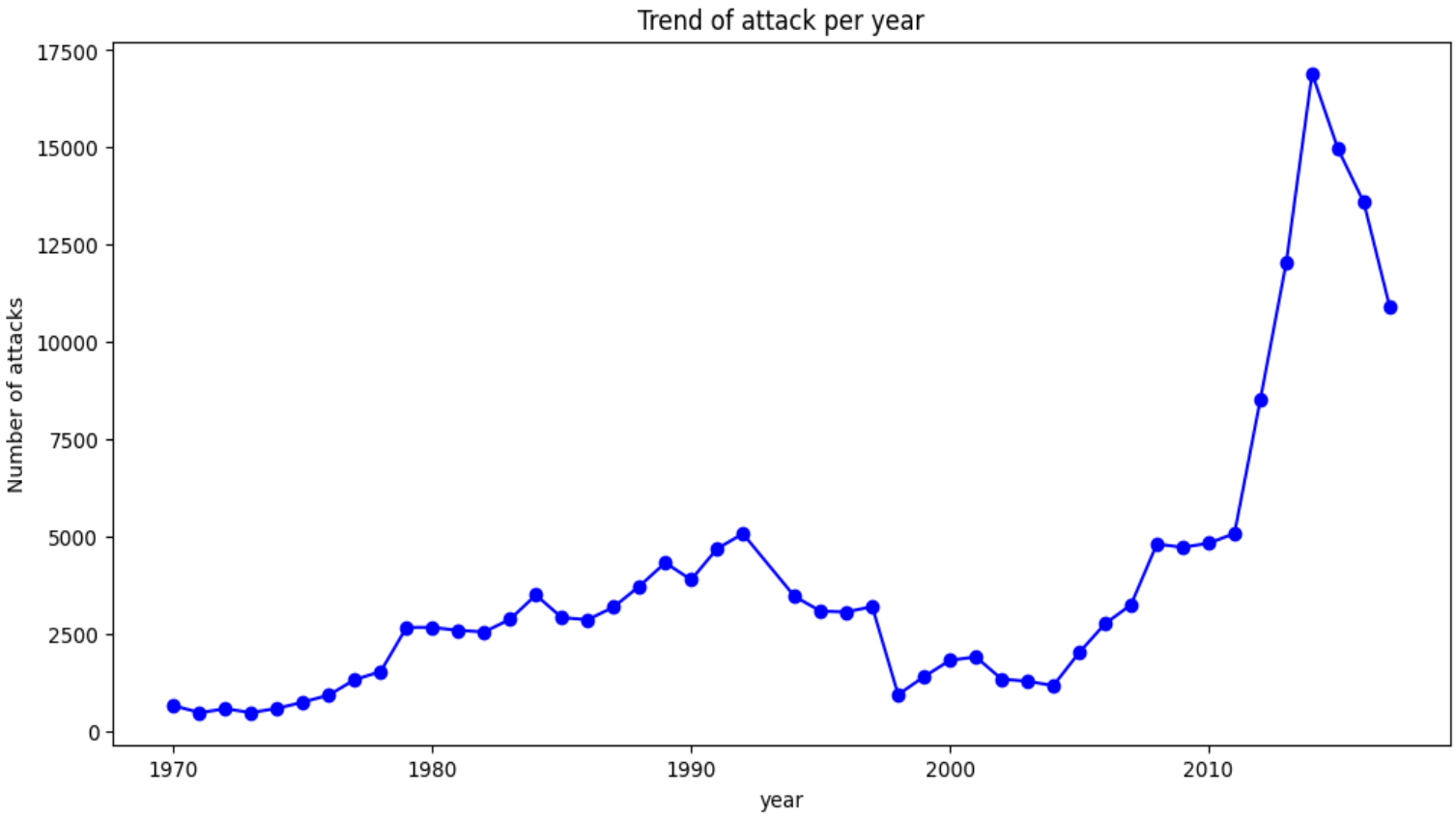
- Most common attack types

attacktype1_txt	attack_types_count
Bombing/Explosion	88255
Armed Assault	42669
Assassination	19312
Hostage Taking (Kidnapping)	11158
Facility/Infrastructure Attack	10356
Unknown	7276
Unarmed Assault	1015
Hostage Taking (Barricade Incident)	991
Hijacking	659

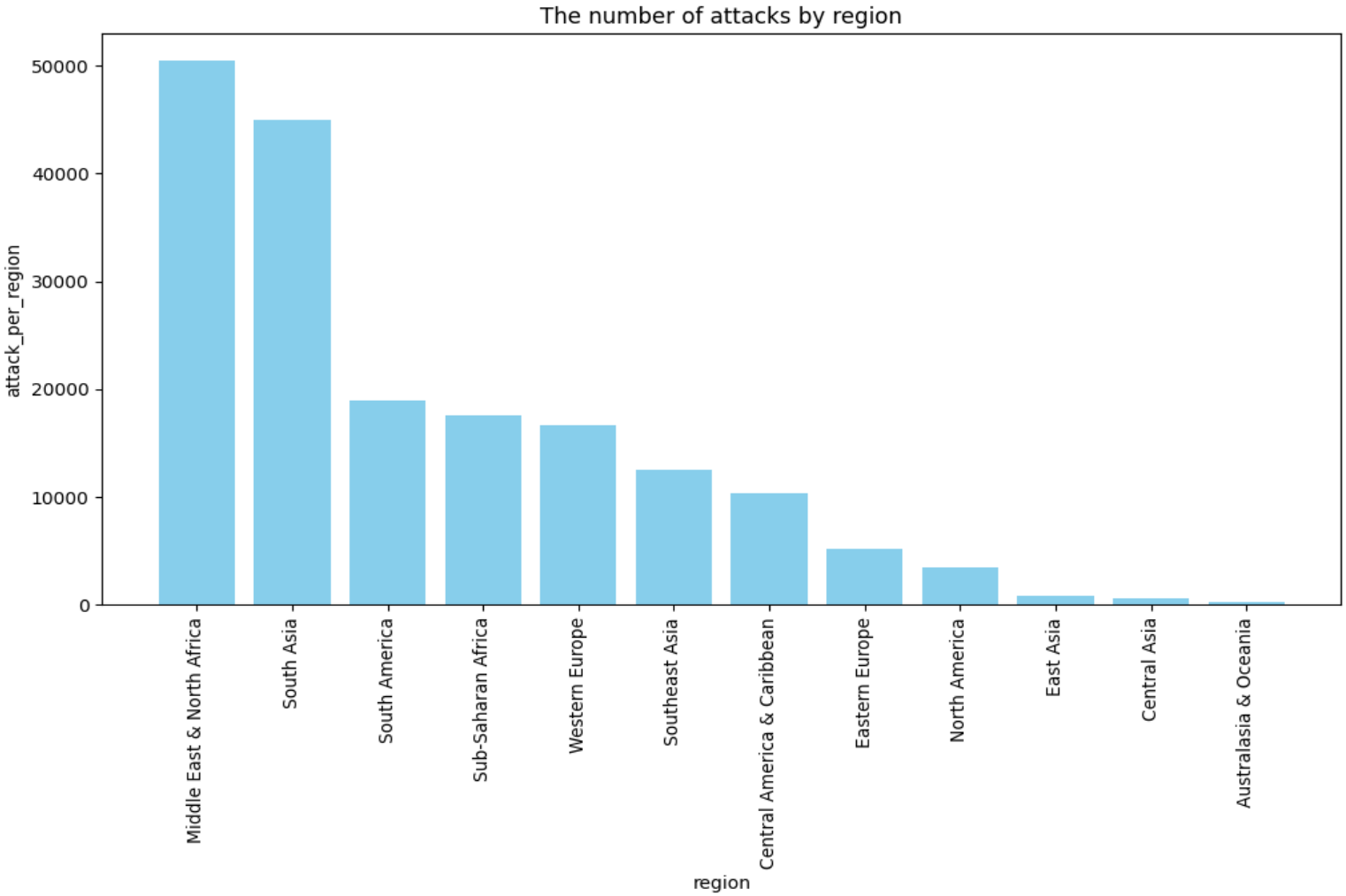
- Most type of people affected from terrorists attacks

targtype1_txt	Most_target_affected
Private Citizens & Property	43511
Military	27984
Police	24506
Government (General)	21283
Business	20669
Transportation	6799
Utilities	6023
Unknown	5898
Religious Figures/Institutions	4440
Educational Institution	4322

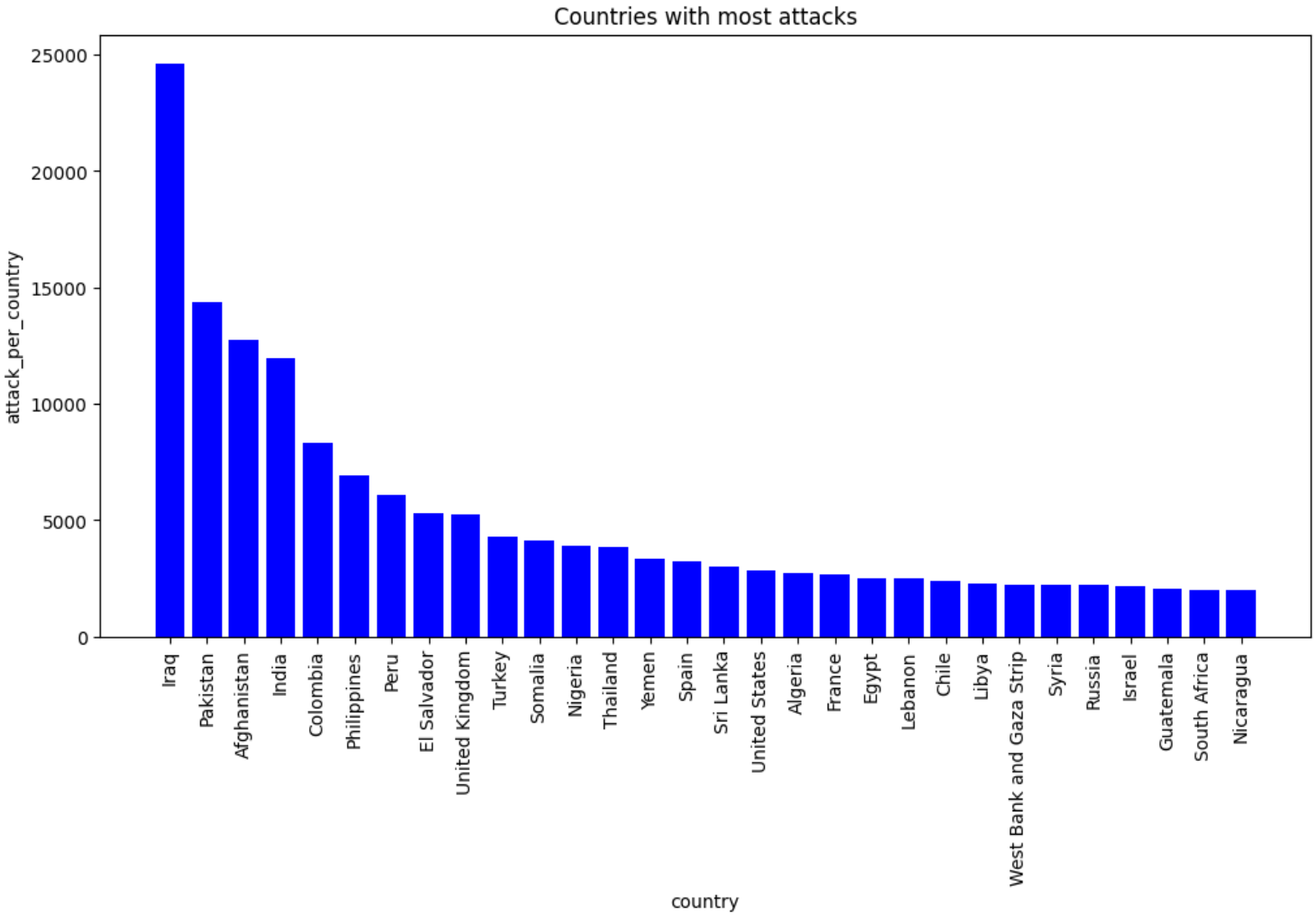
Data Visualization



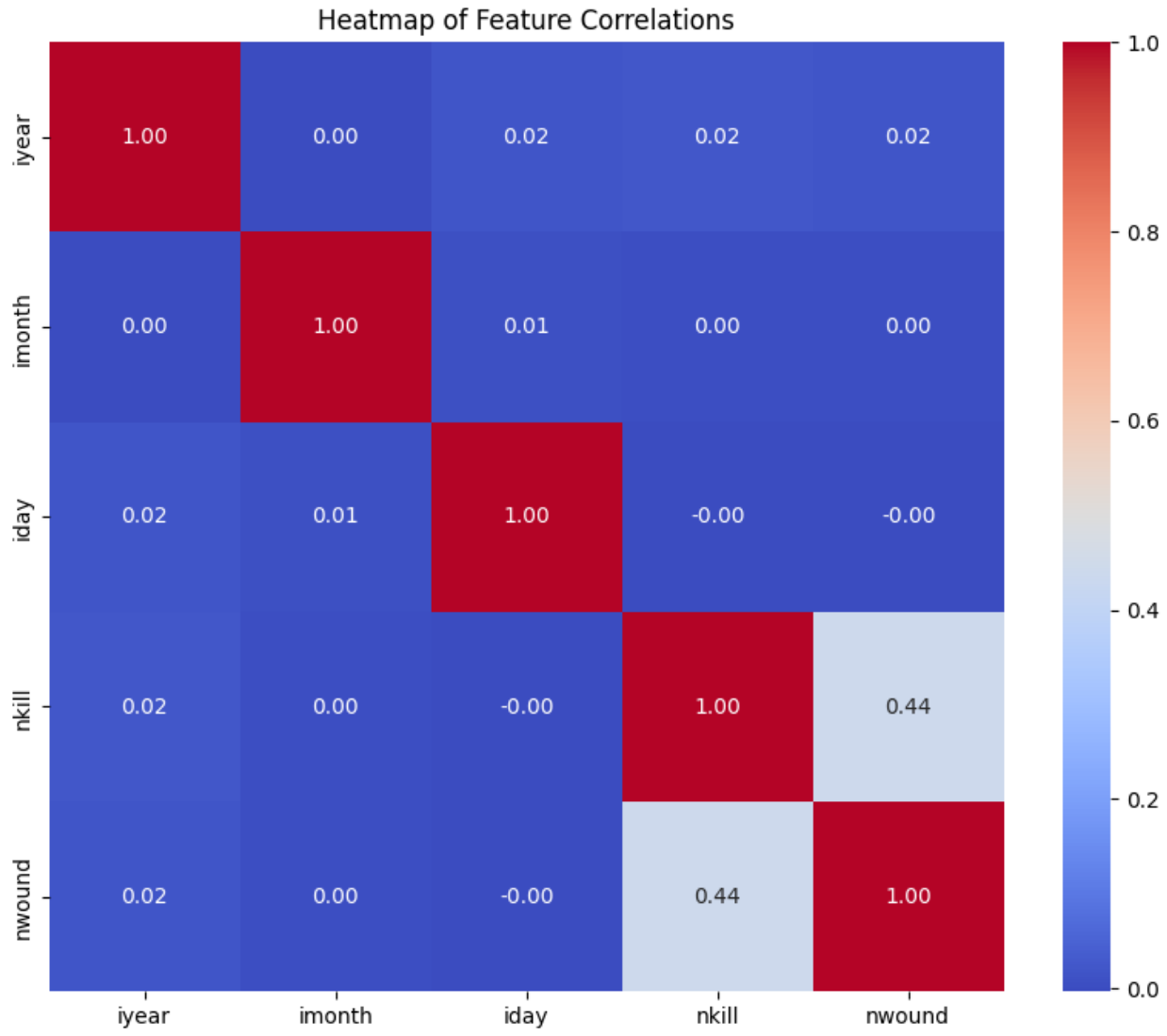
Terrorism Attacks over the year



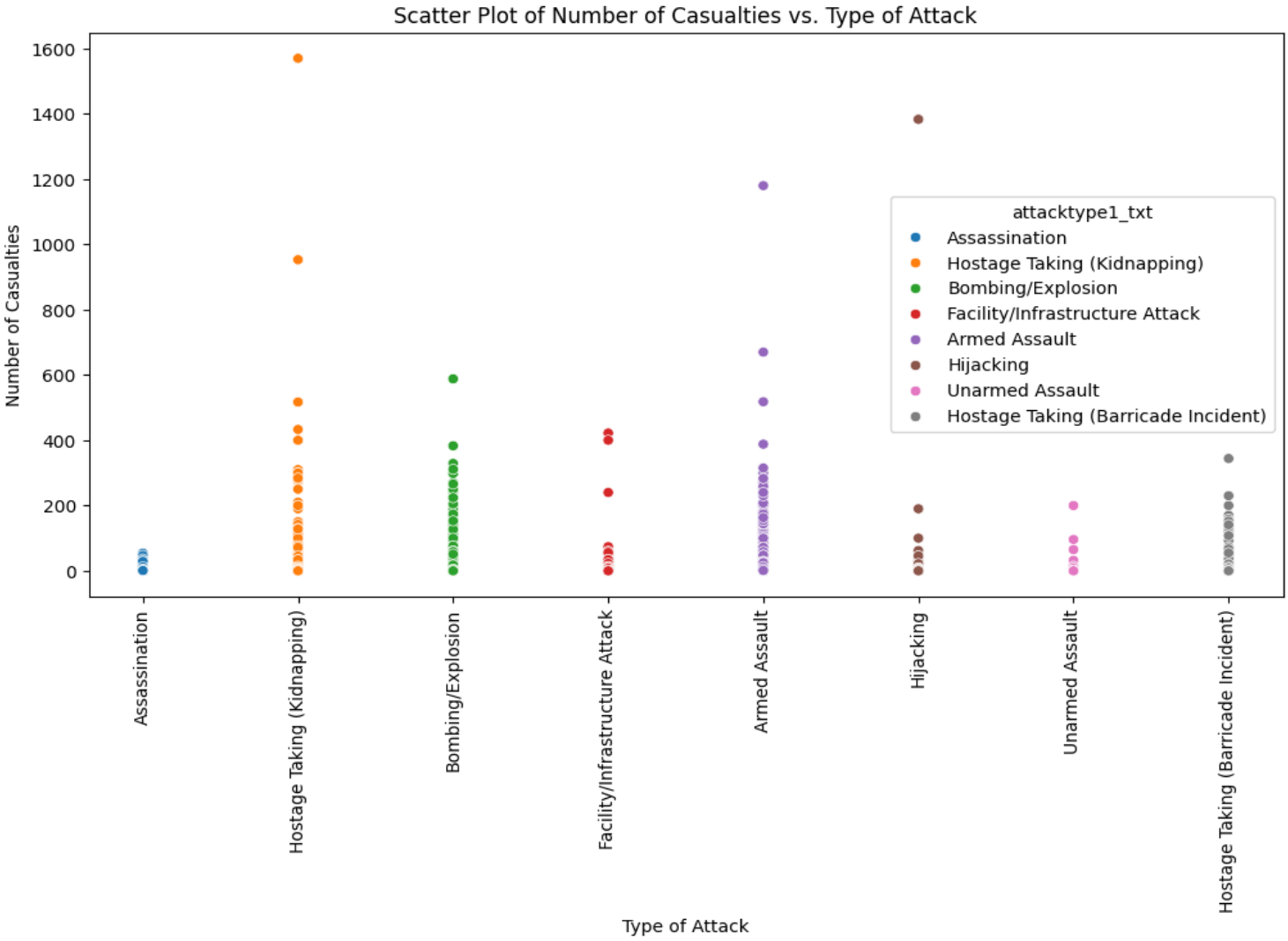
Attack per region



Attack per country top 30 countries



Correlation between numeric columns



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