United Nations

Global Terrorism Project Report

SQL Project Fall 2019

Columbia University Applied Analytics

Table of Contents

Scenario

According to the Global Terrorism Index, terrorism is defined as “the threatened or actual use of illegal force and violence by a non-state actor to attain a political, economic, religious, or social goal through fear, coercion, or intimidation.”

Terrorism is the largest human-oriented criminal activity that threatens the safety of many countries around the globe. As a result, terrorism has become one of the main peace and security agendas of the United Nations. Countering terrorism requires a collective global effort from multinational governments, international organizations and private sectors. To provide leadership on counter-terrorism mandates, the United Nations must first systematically evaluate and analyze past terrorist activities. As a team, we are hired by the United Nations to analyze the pattern of terrorist attacks and ultimately create an opinion on the relatively risky regions and target terrorist organizations.

Team Contract

1. Responsibilities

- Find suitable datasets (All Team Members)
- Data cleaning (Terry Yang)
- Normalization Plan (All Team Members)
- ER diagram (Boyan He)
- Extract, Transform, Load process (All Team Members)
- Explore Metabase and conduct different analysis (All Team Members)
- Developing plans for the client (All Team Members)
- Report (All Team Members)
- Presentation (All Team Members)

2. Timeline of the Most Challenging Tasks

- Project Checkpoint 4: due Nov 25th, 2019
    - Data transformation
    - Data insertion
    - Plan Explanation
- Project Checkpoint 5: due Dec 2nd, 2019
    - Plan of how clients will interact with the database
    - Visualization

    - Interactive dashboard
- Project Report and Presentation: due Dec 8th, 2019

Proposal

Recently, the Islamic State confirmed the death of its founder and leader Abu Bakr al-Baghdadi and announced a successor. It has always been difficult to figure out what are the most important trends in terrorism globally and therefore, to fight terrorism most efficiently. However, as data scientists, we could leverage data analytics for mining clues. With a well-designed database and appropriate SQL queries, our team can extract useful information and run the sophisticated analysis. Our analysis will be more than beneficial for Anti-Violence organizations, such as the United Nations, to explain the most important trends in terrorism globally. Additionally, our analysis will help the United Nations to look beneath the surface of terrorism, understanding the major motivations behind terrorism. Our major goal is to build a sophisticated global terrorism database system.

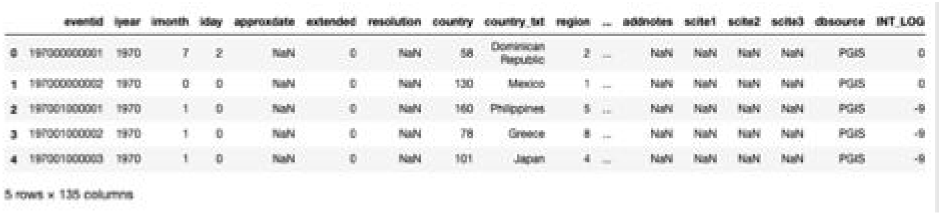
Here are the specific benefits of the database system and processes that we committed to implementing:

- The global terrorism database holds a great chance to serve as a splendid resource of terrorism attacks. Countries that have been suffering from terrorism, such as Iraq and Afghanistan, could utilize this database to get deeper insights and therefore, fight terrorism most efficiently.
- Upgrade or replace the United Nations' current global terrorism database. United Nations will be able to retrieve or update global terrorism data consistently since our database is a real-time database system.
- With a solid understanding of terrorism data, organizations such as the United Nations could prevent & combat terrorism in a more efficient way. United Nations could allocate budget & peacekeeping military force in the right places.
- After careful evaluations, we retrieved a dataset from Kaggle as our original dataset. The original dataset is named as The Global Terrorism Database, which contains 181,691 records of the terrorist attacks worldwide from 1970 to 2017 (except 1993). Here are the main reasons we chose this dataset:
- The dataset has 135 columns of detailed terrorist attacks information such as eventid, year, month, day, on location, tactics, perpetrators, targets, and outcomes.
- The Global Terrorism Database (GTD) is an open-source database including information on terrorist attacks around the world from 1970 through 2017.

- The GTD includes systematic data on domestic as well as international terrorist incidents that have occurred during this period and no. The database is maintained by researchers at the National Consortium for the Study of Terrorism and Responses to Terrorism (START), headquartered at the University of Maryland.
- The database comes with a codebook. From this codebook, our team can first know the data collection methodology, the definition of terrorism in this database and inclusion criteria. Second, the codebook outlines the variables that constitute the GTD and defines the possible values of the variables.

Link to the full dataset: https://www.kaggle.com/START-UMD/gtd

Sample of the dataset



	eventid	iyear	imonth	iday	approxdate	extended	resolution	country	country_txt	region	...	addnotes	scite1	scite2	scite3	dbsource	INT_LOG
0	197000000001	1970	7	2	NaN	0	NaN	58	Dominican Republic	2	...	NaN	NaN	NaN	NaN	PGIS	0
1	197000000002	1970	0	0	NaN	0	NaN	130	Mexico	1	...	NaN	NaN	NaN	NaN	PGIS	0
2	197001000001	1970	1	0	NaN	0	NaN	160	Philippines	5	...	NaN	NaN	NaN	NaN	PGIS	-9
3	197001000002	1970	1	0	NaN	0	NaN	78	Greece	8	...	NaN	NaN	NaN	NaN	PGIS	-9
4	197001000003	1970	1	0	NaN	0	NaN	101	Japan	4	...	NaN	NaN	NaN	NaN	PGIS	-9

5 rows × 135 columns

Database Schema

1. Original Dataset

We've chosen a data set named The Global Terrorism Database, which contains 181,691 records of the terrorist attacks worldwide from 1970 to 2017 (except 1993). The dataset has 135 columns of terrorist attacks information such as eventid, iyear, imonth, iday, on location, tactics, perpetrators, targets, and outcomes.

The Global Terrorism Database (GTD) is an open-source database including information on terrorist attacks around the world from 1970 through 2017. The GTD includes systematic data on domestic as well as international terrorist incidents that have occurred during this time period and no. The database is maintained by researchers at the National Consortium for the Study of Terrorism and Responses to Terrorism (START), headquartered at the University of Maryland. We download this data set through Kaggle: https://www.kaggle.com/START-UMD/gtd

2. Normalization Plan

2.1 Before Normalization

The first step in designing schema is normalization. However, before normalization, some feature selection and data transformation need to be done. Since we have a huge data set, We had to choose our variables wisely. By using the data set codebook provided by GTD, we chose 23 variables that are highly related to our mission. Since some data are not suitable for future analysis, we transformed those data. For example, we have three columns recording date information, which is the year, month and day. To benefit our future analysis, we concatenated these three variables into a new variable, date.

2.2 1NF

After choosing our variables and conducting data transformation, we started to normalize our data set. The first step to normalize our data set is to build the First Normal Form (1NF) of this data set. In 1NF, variables in the data set should be atomic and there cannot be repeating attributes.

In our case, each domain in our data set has already been atomic. However, there are lots of columns that store information about the same thing. For example, we have repeating columns like weapon_type1, weapon_type 2 and weapon_type 3. Therefore, we have to remove the weapon_type2 and weapon_type3 columns, and then rename the weapon_type1 columns as weapon_type, which means this variable will store all information about the weapon. We also will create new rows for weapon_type to store information that is used to store in weapon_type2 and weapon_type3. After we remove all repeating columns. We have a huge events table that contains all the information about each terrorist activity. This table doesn't have repeating attributes and variables in it are atomic. Therefore it meets 1NF.

2.3 2NF

The criteria of 2NF are that tables must be in 1NF and every non-key attribute must be fully dependent on the key. In our table, non-key attribute weapon_type_txt(description of weapons) is dependent on non-key attribute weapon_type (code for weapons). Therefore, our data set doesn't meet 2NF.

To further normalize our data set, we keep all integer and boolean variables, which are variables that only depend on eventid. in our events table and built table for each different entity including attack type, target type, weapon type, and location. Then we insert id for each entity table unless they already had one in the original data set. Those id acts are the primary keys of their table. This action made sure that each entity table meets 2NF. After that, we built relation tables for the events table and the other entity tables. Relation tables link events table and other entity tables. Each relation table consists of eventid (primary key of event table) and primary key of the entity tables which the relation table aims to connect. Therefore, all relation tables meet 2NF as well.

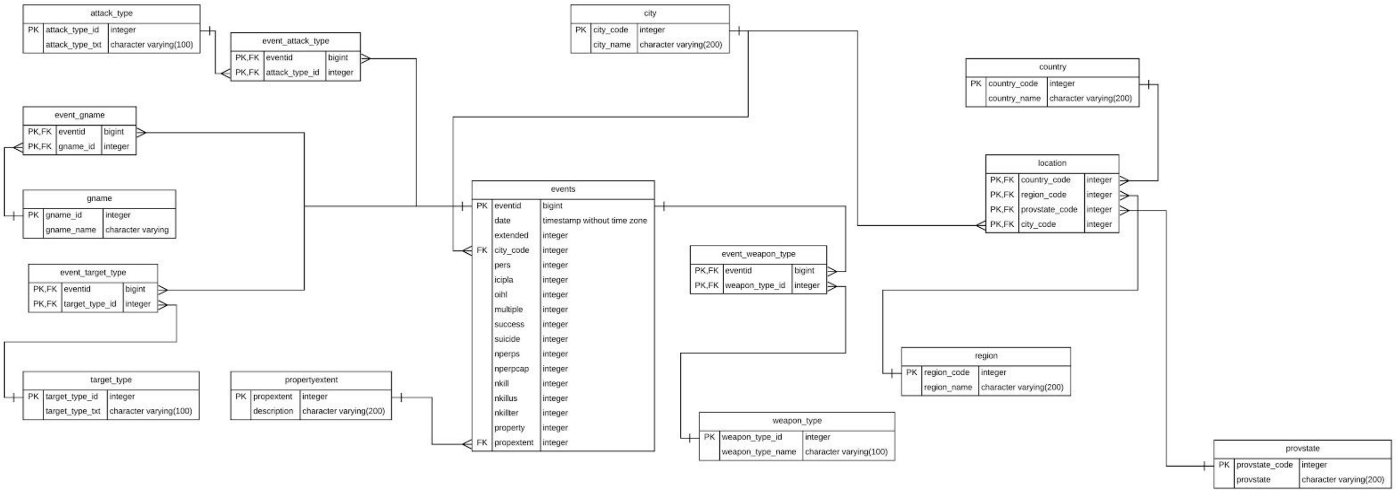
After our normalization, both our entity tables and relation tables meet 2NF and we have a data set that meets 2NF.

2.4 3NF

The criteria for 3NF is that all attributes must only depend on the key. Our data has already met 3NF. No more action needed.

- Kindly visit this link below for:
    - Schema code: https://github.com/bh2709/Team-8-Create-Tables/blob/master/table%20creation

ER Diagrams



Link to Lucidchart:
http://sg.app.lucidchart.com/wf/click?upn=4wgIOpld3pNauqZlzpyPEd2TF-2Bn4I5w1OnJCIMs67c2sc1EhCKC79ggcTFigzYSo8cd4Vogfaf-2FplW8jXRTlEafhNXt53m4vLmzmeTHwmAJxR-2FejXTRoVEi95F59ryO8wj-2BCgNfAD-2FDht9VZn3vAEQNKcTWABS26Jyf

KWuYKJwcrbwb3Ag-2BzeY-2BrXV8L5hDpf7Y5jYj94hDtwySZX4hvi3FSxDwP49Wt-2FhuDoGfyg5fSnmgVTjpvUEum-2Fw7BwcKf9ryD52n-2BltDwATOKK182sUqiUbWYdWoTPcWP79j5ziTfkeTpvk78eNSqzCD3kDi0_uda6uzIhbO3zgsIvEFi1g701ALem70WBInzSW9qupXbyVAh2DnDxoCkgca-2FcjBdLljUhSN-2BxbCAI3lqxtOlYHZoj4MRkR6ewoEt4X0N27kcnRFuW-2BXh5we3EeQrp1D-2F1V-2FXT-2B5wSUWW-2Fo8zVhPyc-2FTaUmcWXZMkPrhSQmTVXkhXCd6LvnLo0Ys7arwMvPYwzQPCHxo5NYzqaqEbxGOmXLJPeIRuR28E13jOrpVuCPEhHj65ecQBbK8sZA0NkOgEUCWJKv73amVWs74FeVlQwt5VOnbKY-2B2nerlUXniu-2B9H5Nq28ior1SB5Wj8soBrjqXu-2Flj9loaxjr879HDAk3-2F6CeLWWvmhUX0vF907cdmTtRAKF7u1pH0SroVeiaqFtjNsBFUrBczg9L9Wmu4HPrYeMp6jCE79iw-2BSDdYV5BSKJdL7sK7SoJiqFE5E-2FV4Sj40GQjutVLnJyFSrVuVgXn0SHmw3uC0qeFK9t3fiJWMKp6b8cwbY-2B2Ghw2zVgyQsg2CJzGTvIxI43fRCLldZMW6NdpnOUHnqby7DdI4cDq1i3Sh7IbpoDyoslQUCEU7iOJD-2Bgd-2Fiawf6aCWTEMtoydjpQ-3D-3D

ETL Process

The detailed ETL Code has been stored in Github. Please kindly find the Github link attached below:
https://github.com/bh2709/Project/blob/master/Team%208%20Global%20Terrorism%20Analytics%20Project

ETL in details:

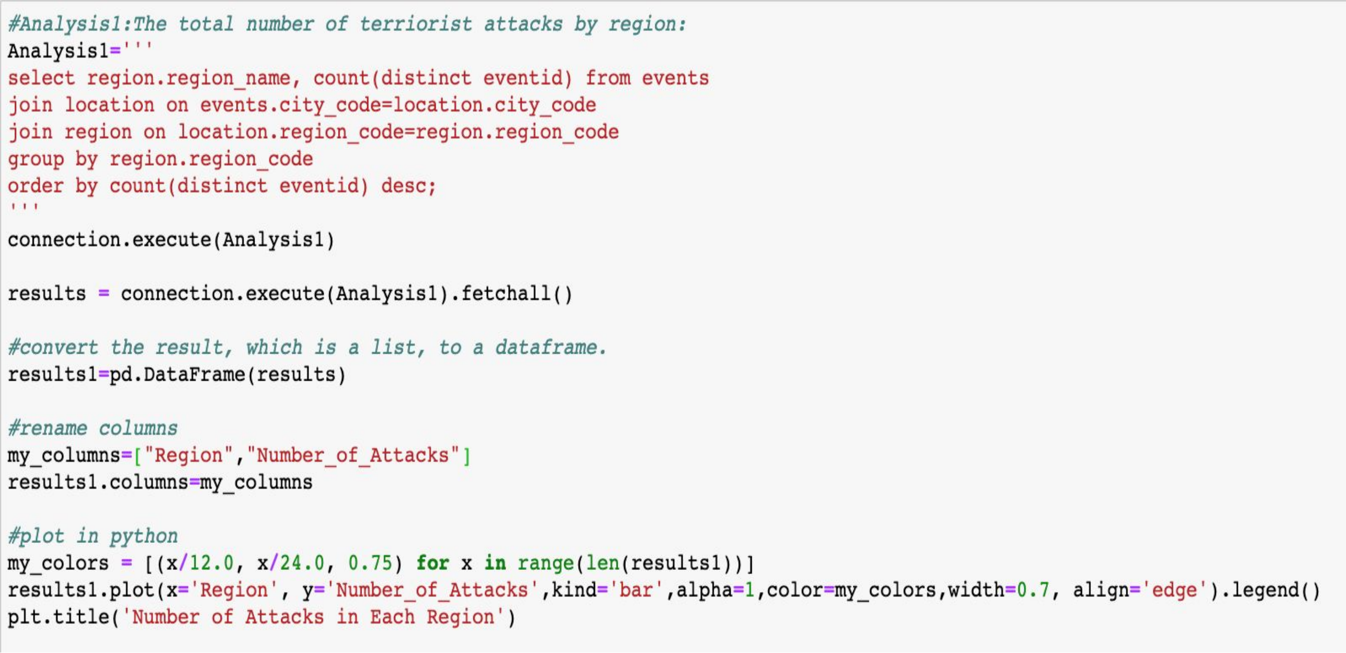
• Extract:
o Remove duplicate columns
o Design appropriate names (rename columns)
o Make a draft for constraints
o Extract useful columns, drop columns with too many null values, and drop duplicated data

• Transform:
o check for duplicated data
o data munging
o push data properly based on our database schema

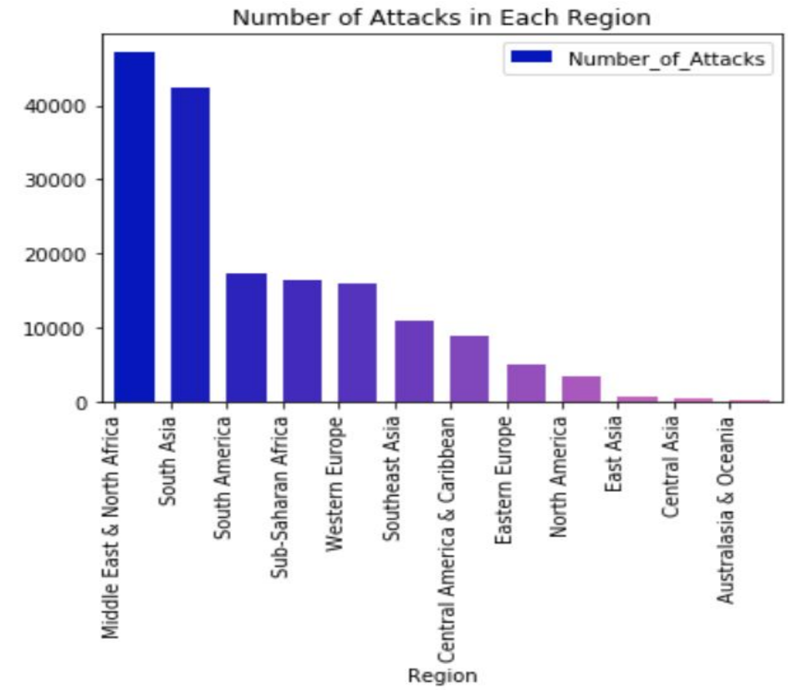
• Load
o Load data
o Check for data loading accuracy by running SQL queries in PostgreSQL database

Analytical Procedures

1. What is the total number of attacks in each region?

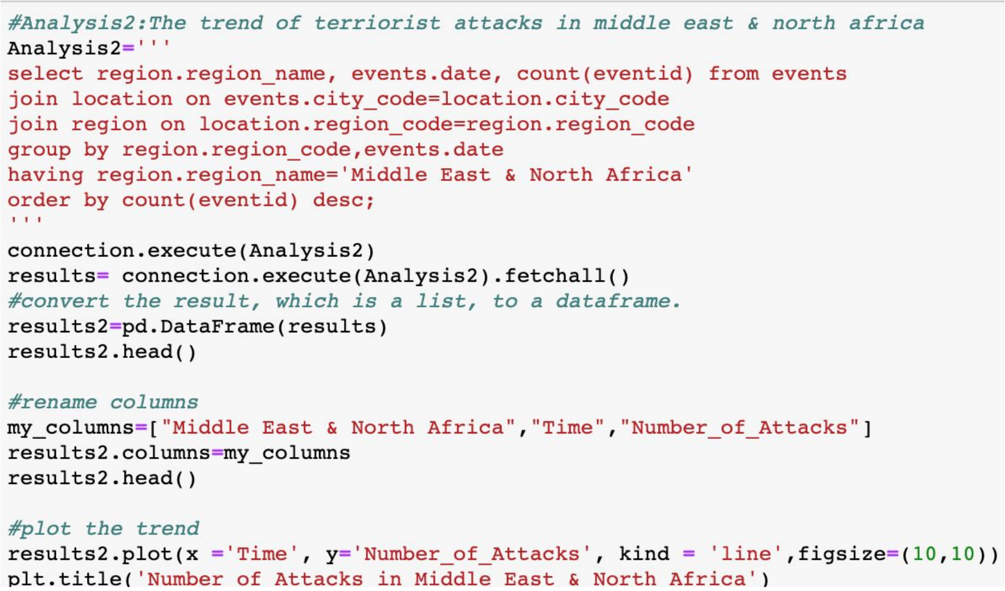


```
#Analysis1:The total number of terriorist attacks by region:
Analysis1='''
select region.region_name, count(distinct eventid) from events
join location on events.city_code=location.city_code
join region on location.region_code=region.region_code
group by region.region_code
order by count(distinct eventid) desc;
'''
connection.execute(Analysis1)

results = connection.execute(Analysis1).fetchall()

#convert the result, which is a list, to a dataframe.
results1=pd.DataFrame(results)

#rename columns
my_columns=["Region","Number_of_Attacks"]
results1.columns=my_columns

#plot in python
my_colors = [(x/12.0, x/24.0, 0.75) for x in range(len(results1))]
results1.plot(x='Region', y='Number_of_Attacks',kind='bar',alpha=1,color=my_colors,width=0.7, align='edge').legend()
plt.title('Number of Attacks in Each Region')
```



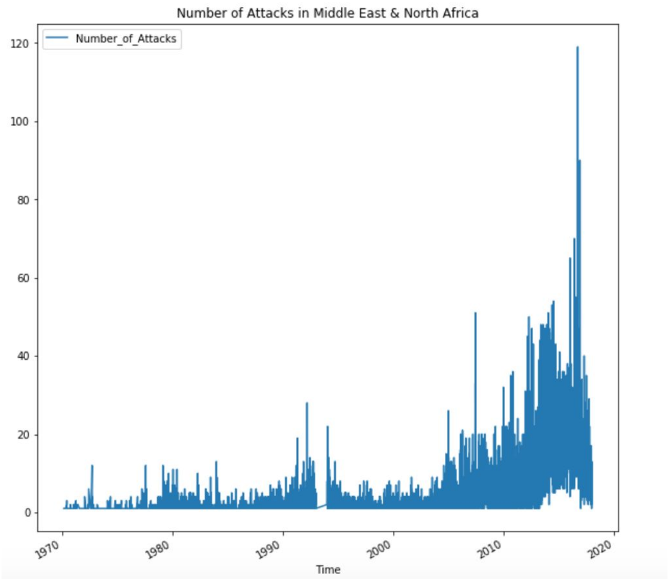
2. The trend of terrorist attacks in the Middle East & North Africa？

Based on our analysis 1, the Middle East & North Africa area suffers from terrorist attacks the most. Therefore, we took a closer look at this area by drawing the overall trend for this particular area.



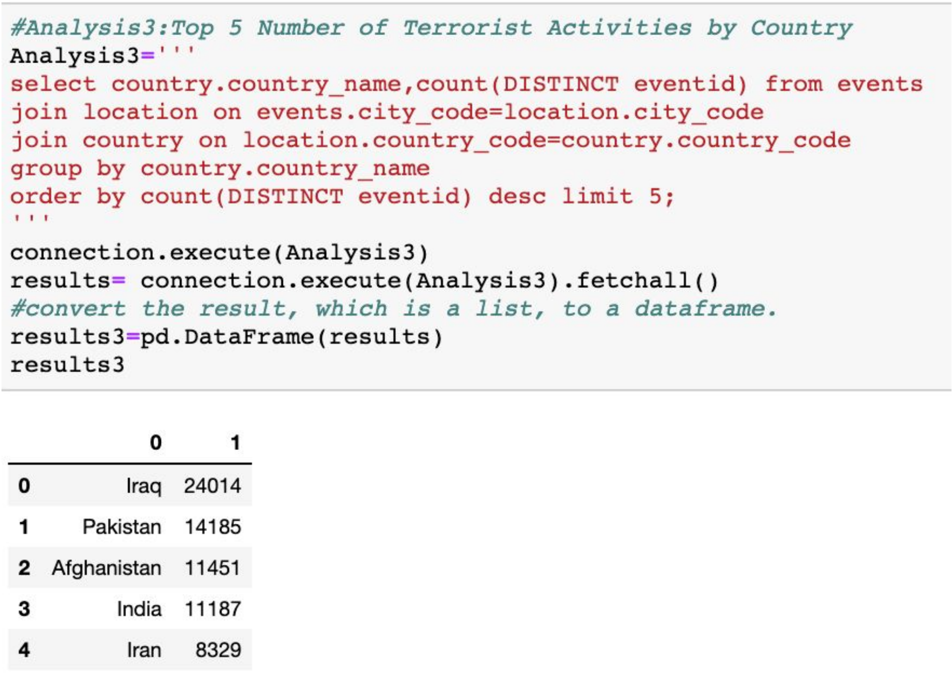
```
#Analysis2:The trend of terriorist attacks in middle east & north africa
Analysis2='''
select region.region_name, events.date, count(eventid) from events
join location on events.city_code=location.city_code
join region on location.region_code=region.region_code
group by region.region_code,events.date
having region.region_name='Middle East & North Africa'
order by count(eventid) desc;
'''
connection.execute(Analysis2)
results= connection.execute(Analysis2).fetchall()
#convert the result, which is a list, to a dataframe.
results2=pd.DataFrame(results)
results2.head()

#rename columns
my_columns=["Middle East & North Africa","Time","Number_of_Attacks"]
results2.columns=my_columns
results2.head()

#plot the trend
results2.plot(x ='Time', y='Number_of_Attacks', kind = 'line',figsize=(10,10))
plt.title('Number of Attacks in Middle East & North Africa')
```

```
Text(0.5, 1.0, 'Number of Attacks in Middle East & North Africa')
```



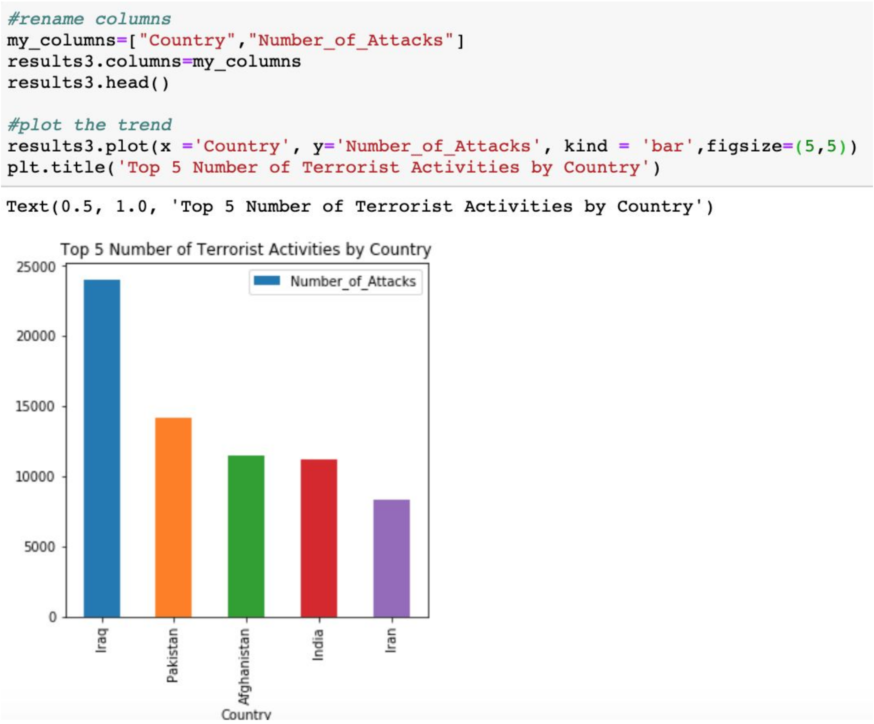
3. Top 5 Number of Terrorist Activities by Country？

Besides regions, we want to get a better understanding of global terrorism by country.



```
#Analysis3:Top 5 Number of Terrorist Activities by Country
Analysis3='''
select country.country_name,count(DISTINCT eventid) from events
join location on events.city_code=location.city_code
join country on location.country_code=country.country_code
group by country.country_name
order by count(DISTINCT eventid) desc limit 5;
'''
connection.execute(Analysis3)
results= connection.execute(Analysis3).fetchall()
#convert the result, which is a list, to a dataframe.
results3=pd.DataFrame(results)
results3
```

	0	1
0	Iraq	24014
1	Pakistan	14185
2	Afghanistan	11451
3	India	11187
4	Iran	8329

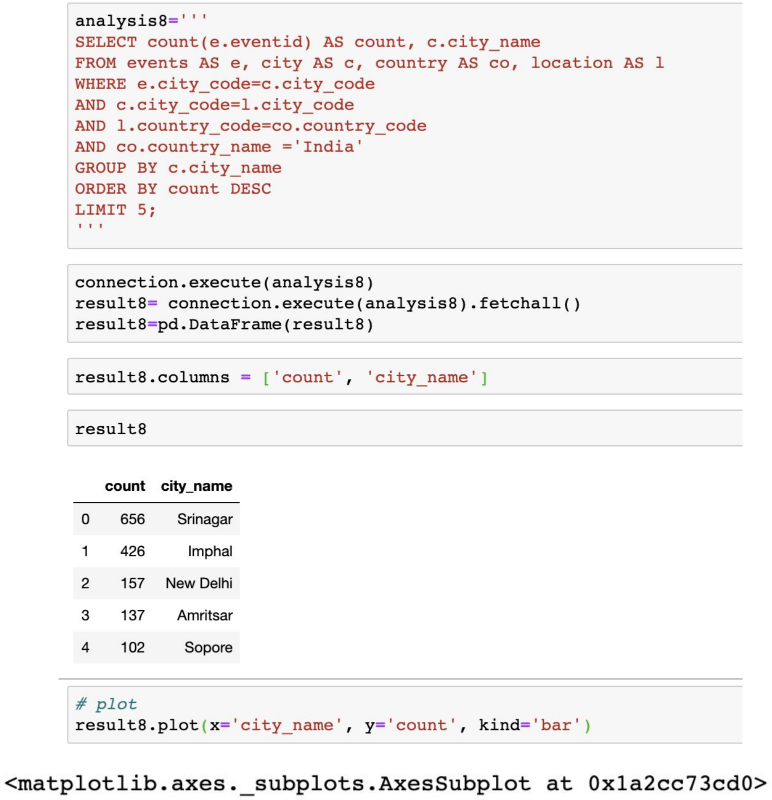


```
#rename columns
my_columns=["Country","Number_of_Attacks"]
results3.columns=my_columns
results3.head()

#plot the trend
results3.plot(x ='Country', y='Number_of_Attacks', kind = 'bar',figsize=(5,5))
plt.title('Top 5 Number of Terrorist Activities by Country')
```

```
Text(0.5, 1.0, 'Top 5 Number of Terrorist Activities by Country')
```

4. Top 5 Indian Cities who have seen the terrorism the most？

Obviously, the direct SQL query can be easily used in our database, which is one of the most simple ways to analyze data, especially for analysts.



```
analysis8='''
SELECT count(e.eventid) AS count, c.city_name
FROM events AS e, city AS c, country AS co, location AS l
WHERE e.city_code=c.city_code
AND c.city_code=l.city_code
AND l.country_code=co.country_code
AND co.country_name ='India'
GROUP BY c.city_name
ORDER BY count DESC
LIMIT 5;
'''
```

```
connection.execute(analysis8)
result8= connection.execute(analysis8).fetchall()
result8=pd.DataFrame(result8)
```

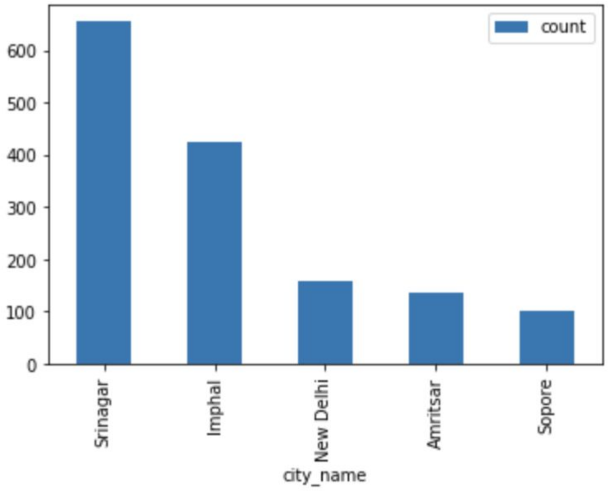
```
result8.columns = ['count', 'city_name']
```

```
result8
```

	count	city_name
0	656	Srinagar
1	426	Imphal
2	157	New Delhi
3	137	Amritsar
4	102	Sopore

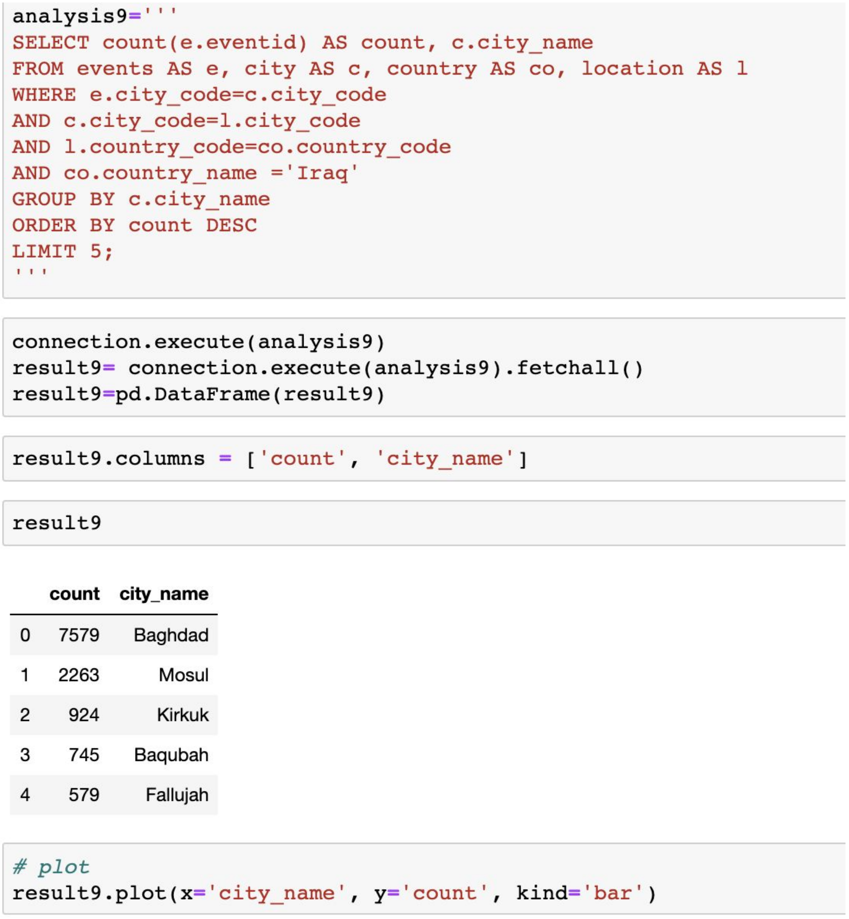
```
# plot
result8.plot(x='city_name', y='count', kind='bar')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x1a2cc73cd0>
```



The result is Srinagar, Imphal, New Delhi, Amritsar ,and Sopore. These places are all in the north part of India.

5. Top 5 Cities from Iraq who have seen the terrorist acts the most？



```
analysis9='''
SELECT count(e.eventid) AS count, c.city_name
FROM events AS e, city AS c, country AS co, location AS l
WHERE e.city_code=c.city_code
AND c.city_code=l.city_code
AND l.country_code=co.country_code
AND co.country_name ='Iraq'
GROUP BY c.city_name
ORDER BY count DESC
LIMIT 5;
'''
```

```
connection.execute(analysis9)
result9= connection.execute(analysis9).fetchall()
result9=pd.DataFrame(result9)
```

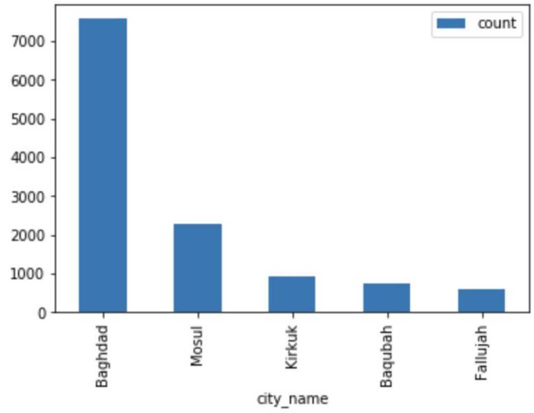
```
result9.columns = ['count', 'city_name']
```

```
result9
```

	count	city_name
0	7579	Baghdad
1	2263	Mosul
2	924	Kirkuk
3	745	Baqubah
4	579	Fallujah

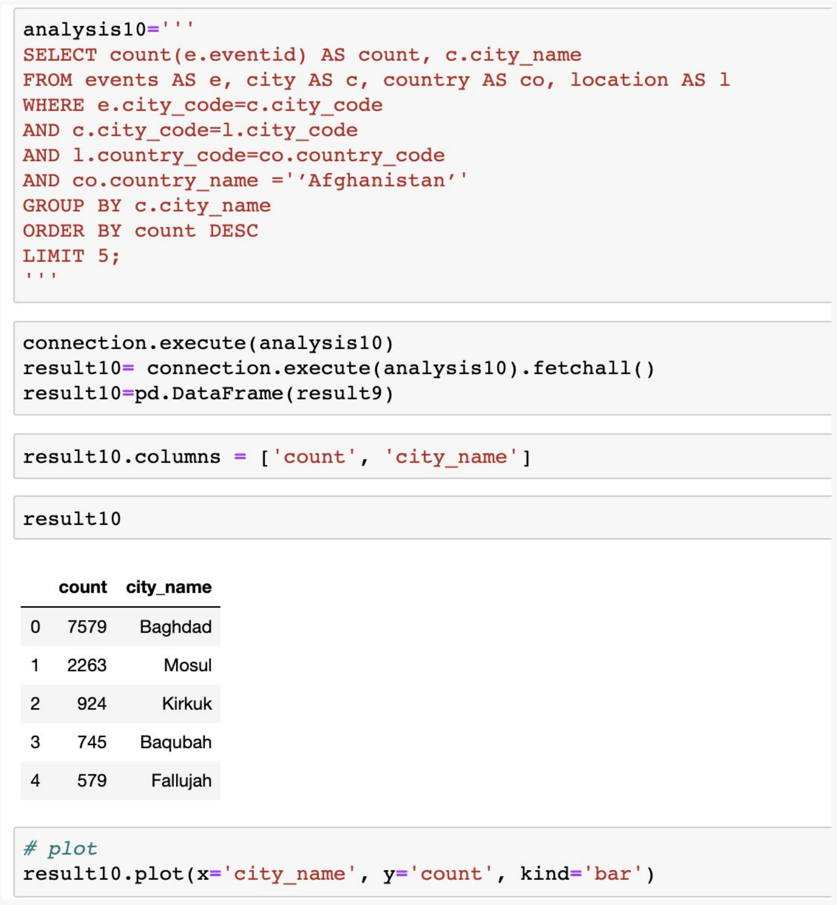
```
# plot
result9.plot(x='city_name', y='count', kind='bar')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x122605890>
```



The result is Baghdad, Mosul, Kirkuk, Baqubah ,and Fallujah.

6. Top 5 Cities from Afghanistan who has seen the terrorism the mos？



```
analysis10='''
SELECT count(e.eventid) AS count, c.city_name
FROM events AS e, city AS c, country AS co, location AS l
WHERE e.city_code=c.city_code
AND c.city_code=l.city_code
AND l.country_code=co.country_code
AND co.country_name ='’Afghanistan’'
GROUP BY c.city_name
ORDER BY count DESC
LIMIT 5;
'''
```

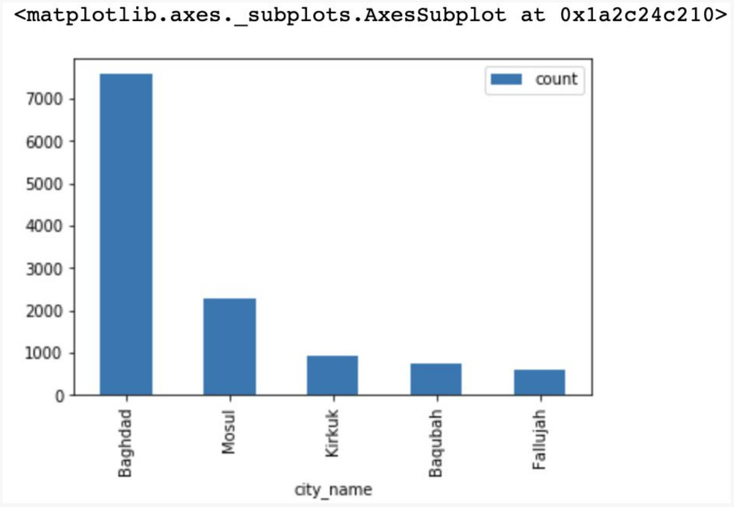
```
connection.execute(analysis10)
result10= connection.execute(analysis10).fetchall()
result10=pd.DataFrame(result9)
```

```
result10.columns = ['count', 'city_name']
```

```
result10
```

	count	city_name
0	7579	Baghdad
1	2263	Mosul
2	924	Kirkuk
3	745	Baqubah
4	579	Fallujah

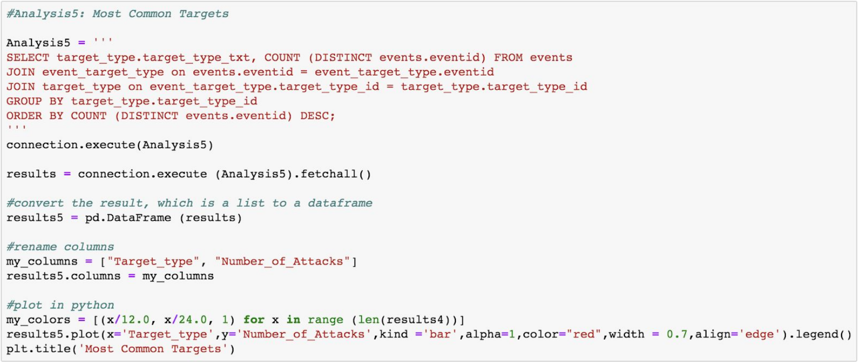
```
# plot
result10.plot(x='city_name', y='count', kind='bar')
```



```
<matplotlib.axes._subplots.AxesSubplot at 0x1a2c24c210>
```

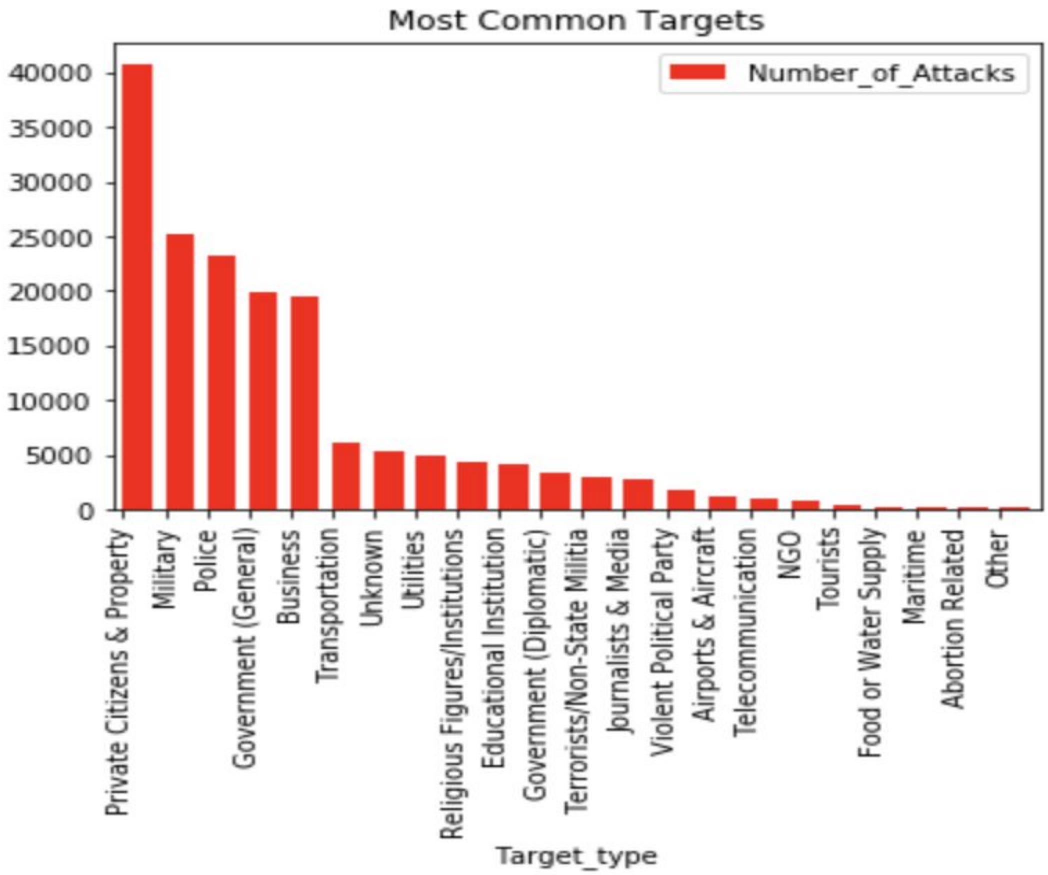
The result is Kabul, Kandahar, Jalalabad, Lashkar Gah ,and Ghazni.

7. Most Common Targets？



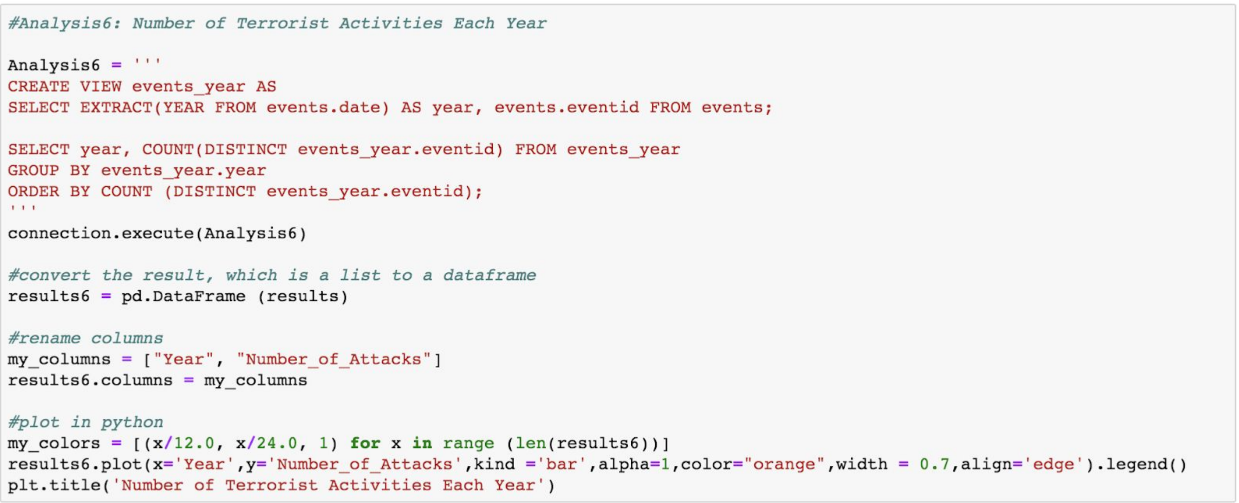
```
#Analysis5: Most Common Targets

Analysis5 = '''
SELECT target_type.target_type_txt, COUNT (DISTINCT events.eventid) FROM events
JOIN event_target_type on events.eventid = event_target_type.eventid
JOIN target_type on event_target_type.target_type_id = target_type.target_type_id
GROUP BY target_type.target_type_id
ORDER BY COUNT (DISTINCT events.eventid) DESC;
'''
connection.execute(Analysis5)

results = connection.execute (Analysis5).fetchall()

#convert the result, which is a list to a dataframe
results5 = pd.DataFrame (results)

#rename columns
my_columns = ["Target_type", "Number_of_Attacks"]
results5.columns = my_columns

#plot in python
my_colors = [(x/12.0, x/24.0, 1) for x in range (len(results4))]
results5.plot(x='Target_type',y='Number_of_Attacks',kind ='bar',alpha=1,color="red",width = 0.7,align='edge').legend()
plt.title('Most Common Targets')
```

```
Text(0.5, 1.0, 'Most Common Targets')
```



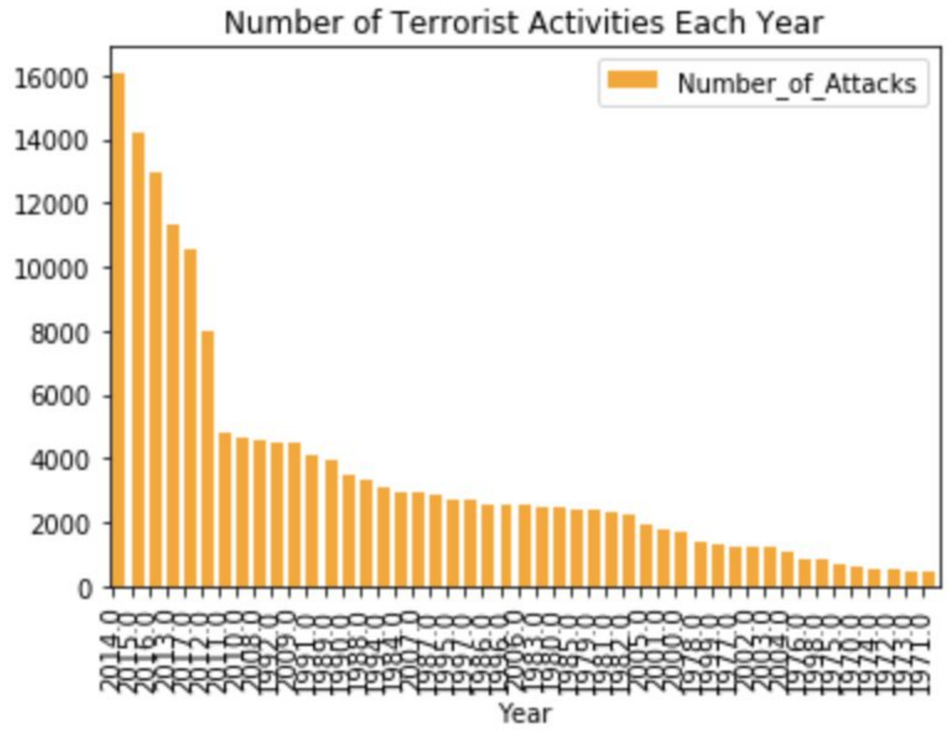
This graph demonstrates that the most common target in a terrorist attack is mainly private citizens & property (civilians), followed by military, police, government, and business, etc.

8. Years with Most Terrorist Attacks



```
#Analysis6: Number of Terrorist Activities Each Year

Analysis6 = '''
CREATE VIEW events_year AS
SELECT EXTRACT(YEAR FROM events.date) AS year, events.eventid FROM events;

SELECT year, COUNT(DISTINCT events_year.eventid) FROM events_year
GROUP BY events_year.year
ORDER BY COUNT (DISTINCT events_year.eventid);
'''
connection.execute(Analysis6)

#convert the result, which is a list to a dataframe
results6 = pd.DataFrame (results)

#rename columns
my_columns = ["Year", "Number_of_Attacks"]
results6.columns = my_columns

#plot in python
my_colors = [(x/12.0, x/24.0, 1) for x in range (len(results6))]
results6.plot(x='Year',y='Number_of_Attacks',kind ='bar',alpha=1,color="orange",width = 0.7,align='edge').legend()
plt.title('Number of Terrorist Activities Each Year')
```

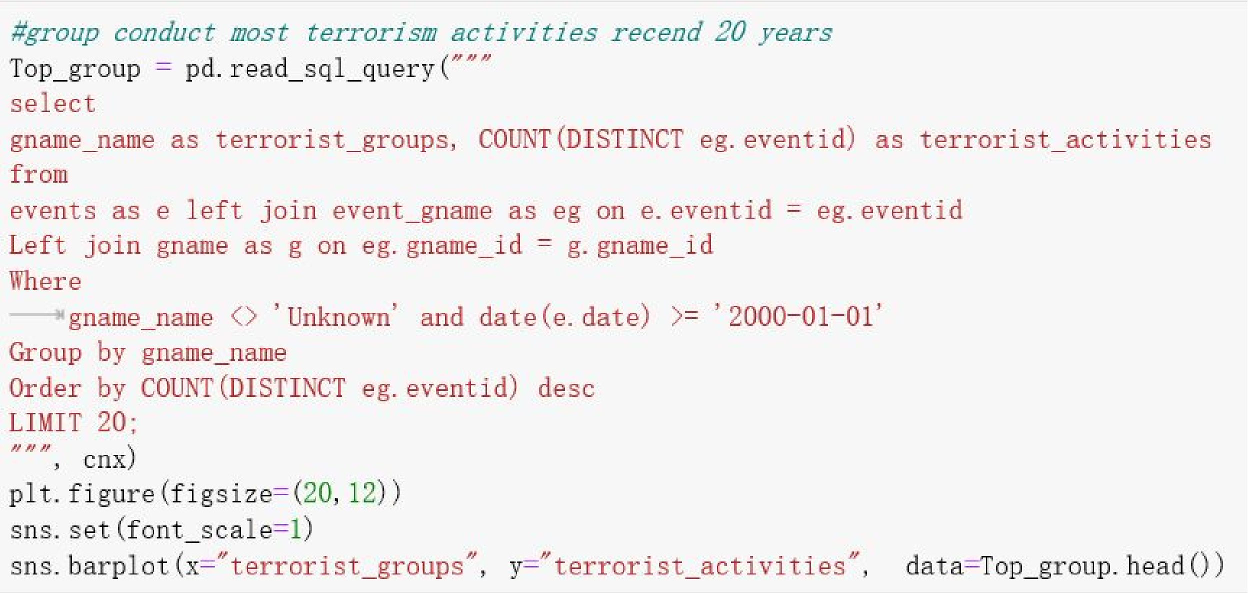
```
Text(0.5, 1.0, 'Number of Terrorist Activities Each Year')
```



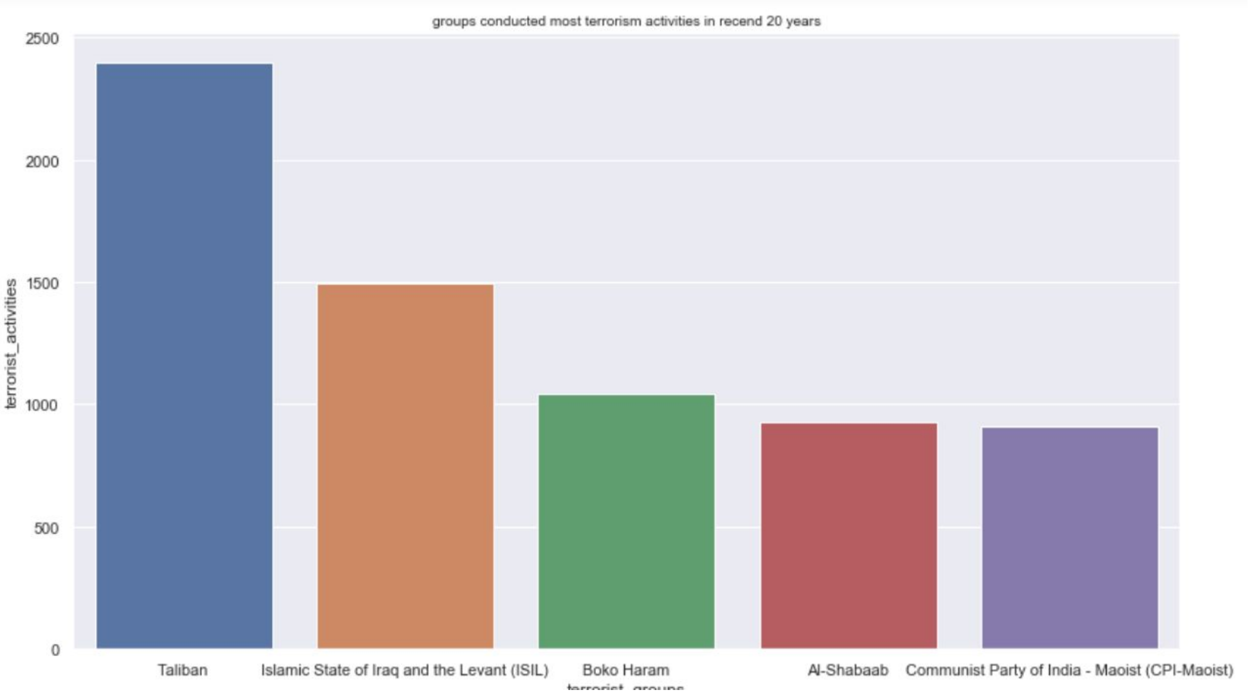
From this graph, we can see that the year 2014 has the most terrorist attacks. An interesting fact is that the top 8 years with most attacks are all post-2010. In other words, since 2010, the number of terrorists grew exponentially until 2014.

9. Top Terrorist Groups

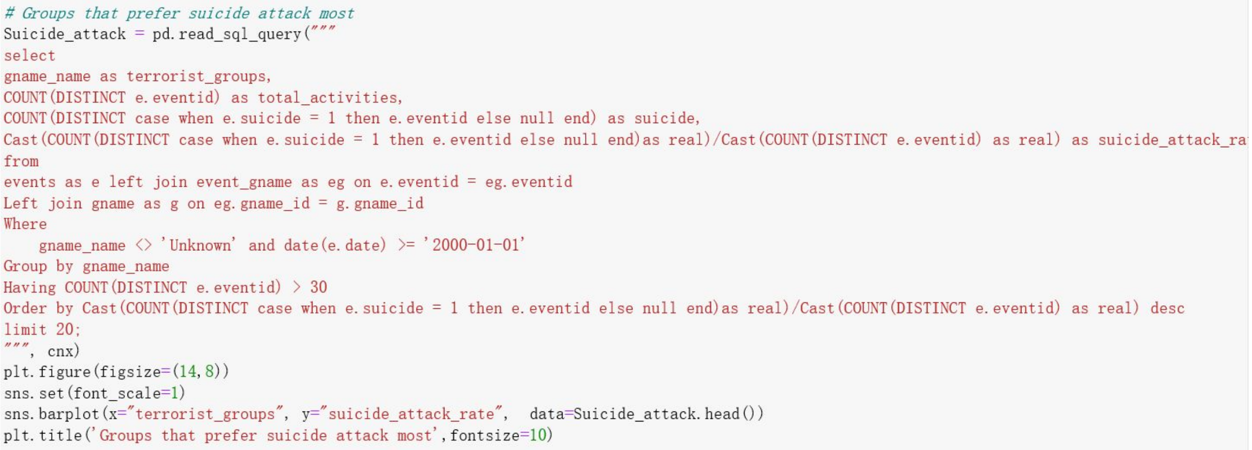
Taliban and ISIL are the two most active terrorist groups in the recent 20 years.



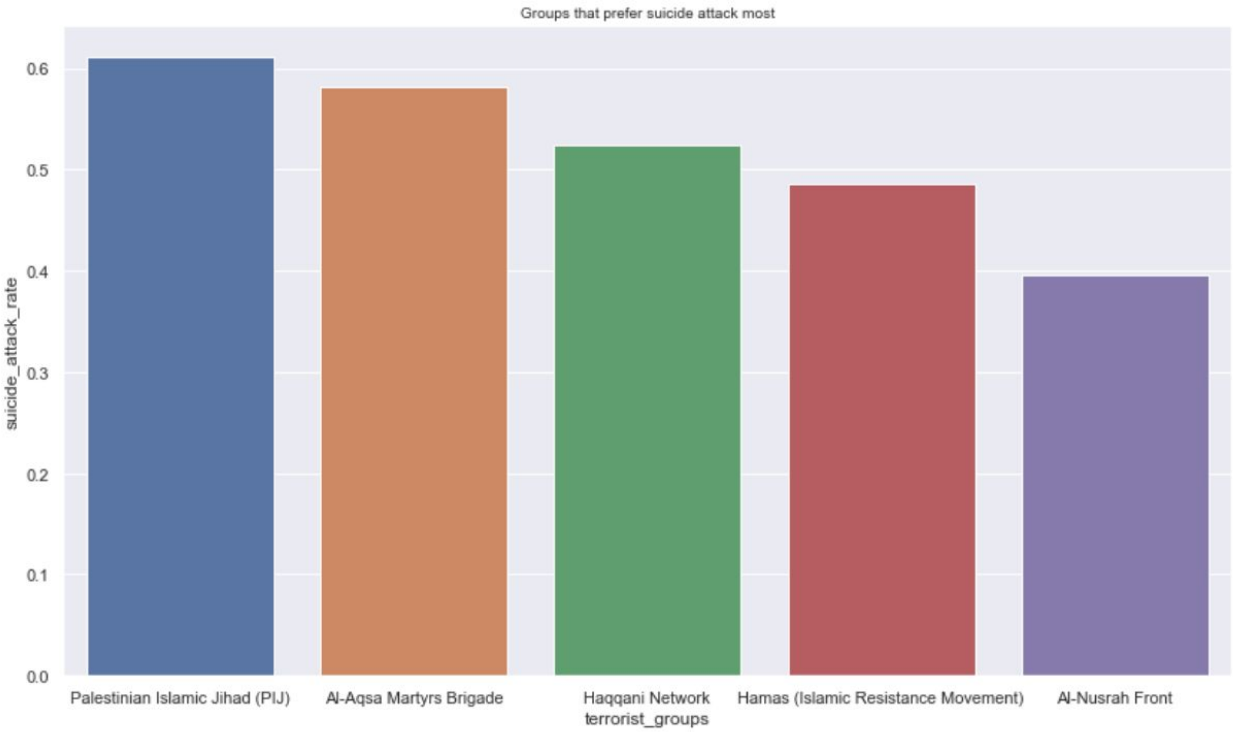
```
#group conduct most terrorism activities recend 20 years
Top_group = pd.read_sql_query("""
select
gname_name as terrorist_groups, COUNT(DISTINCT eg.eventid) as terrorist_activities
from
events as e left join event_gname as eg on e.eventid = eg.eventid
Left join gname as g on eg.gname_id = g.gname_id
Where
	gname_name <> 'Unknown' and date(e.date) >= '2000-01-01'
Group by gname_name
Order by COUNT(DISTINCT eg.eventid) desc
LIMIT 20;
""", cnx)
plt.figure(figsize=(20,12))
sns.set(font_scale=1)
sns.barplot(x="terrorist_groups", y="terrorist_activities",  data=Top_group.head())
```



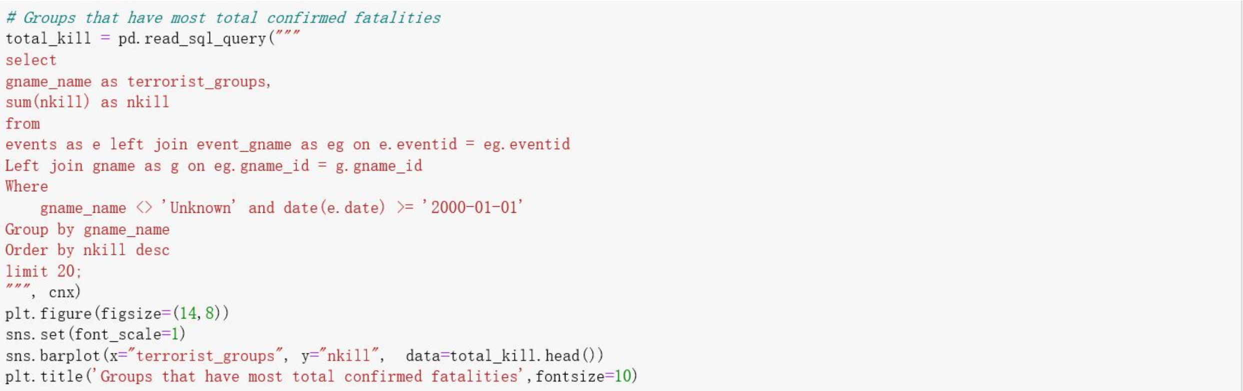
10. Groups that prefer suicide attacks most



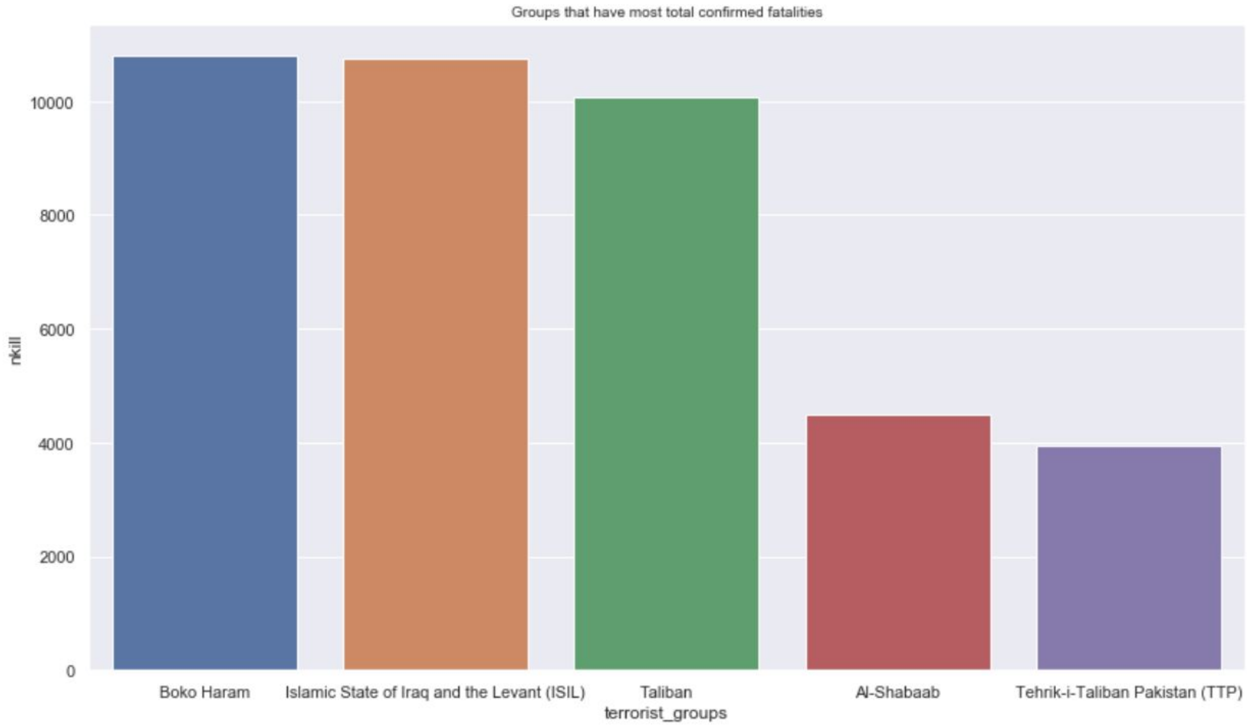
```
# Groups that prefer suicide attack most
Suicide_attack = pd.read_sql_query("""
select
gname_name as terrorist_groups,
COUNT(DISTINCT e.eventid) as total_activities,
COUNT(DISTINCT case when e.suicide = 1 then e.eventid else null end) as suicide,
Cast(COUNT(DISTINCT case when e.suicide = 1 then e.eventid else null end)as real)/Cast(COUNT(DISTINCT e.eventid) as real) as suicide_attack_ra
from
events as e left join event_gname as eg on e.eventid = eg.eventid
Left join gname as g on eg.gname_id = g.gname_id
Where
    gname_name <> 'Unknown' and date(e.date) >= '2000-01-01'
Group by gname_name
Having COUNT(DISTINCT e.eventid) > 30
Order by Cast(COUNT(DISTINCT case when e.suicide = 1 then e.eventid else null end)as real)/Cast(COUNT(DISTINCT e.eventid) as real) desc
limit 20;
""", cnx)
plt.figure(figsize=(14,8))
sns.set(font_scale=1)
sns.barplot(x="terrorist_groups", y="suicide_attack_rate",  data=Suicide_attack.head())
plt.title('Groups that prefer suicide attack most',fontsize=10)
```



11. Groups that have most total confirmed fatalities

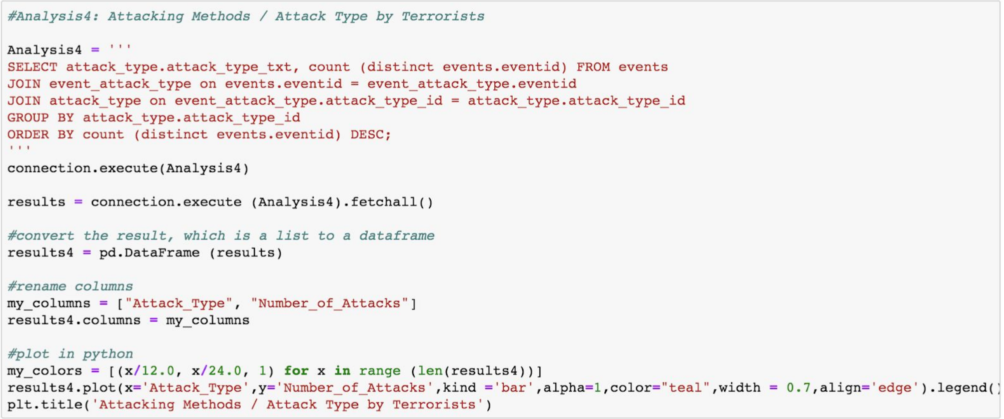


```
# Groups that have most total confirmed fatalities
total_kill = pd.read_sql_query("""
select
gname_name as terrorist_groups,
sum(nkill) as nkill
from
events as e left join event_gname as eg on e.eventid = eg.eventid
Left join gname as g on eg.gname_id = g.gname_id
Where
    gname_name <> 'Unknown' and date(e.date) >= '2000-01-01'
Group by gname_name
Order by nkill desc
limit 20;
""", cnx)
plt.figure(figsize=(14,8))
sns.set(font_scale=1)
sns.barplot(x="terrorist_groups", y="nkill",  data=total_kill.head())
plt.title('Groups that have most total confirmed fatalities',fontsize=10)
```



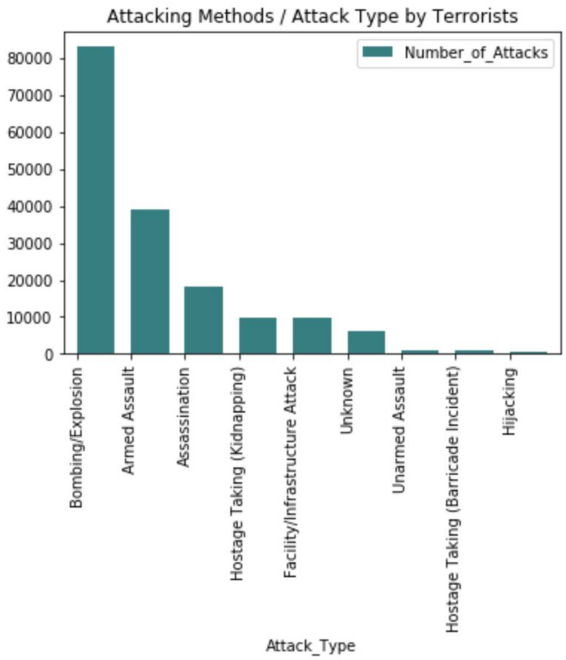
From this graph, we can tell that Boko Haram, ISIL ,and Taliban’s terrorist activities caused lots of casualties。

12. Attacking Methods / Attack Type by Terrorists



```
#Analysis4: Attacking Methods / Attack Type by Terrorists

Analysis4 = '''
SELECT attack_type.attack_type_txt, count (distinct events.eventid) FROM events
JOIN event_attack_type on events.eventid = event_attack_type.eventid
JOIN attack_type on event_attack_type.attack_type_id = attack_type.attack_type_id
GROUP BY attack_type.attack_type_id
ORDER BY count (distinct events.eventid) DESC;
'''
connection.execute(Analysis4)

results = connection.execute (Analysis4).fetchall()

#convert the result, which is a list to a dataframe
results4 = pd.DataFrame (results)

#rename columns
my_columns = ["Attack_Type", "Number_of_Attacks"]
results4.columns = my_columns

#plot in python
my_colors = [(x/12.0, x/24.0, 1) for x in range (len(results4))]
results4.plot(x='Attack_Type',y='Number_of_Attacks',kind ='bar',alpha=1,color="teal",width = 0.7,align='edge').legend(
plt.title('Attacking Methods / Attack Type by Terrorists')
```

```
Text(0.5, 1.0, 'Attacking Methods / Attack Type by Terrorists')
```



From this graph, we can see that bombing/explosion is the most used attack method by terrorists, followed by armed assault, assassination, and hostage-taking, etc.

Database Interactions

Since we have customers from two sides, The United Nations (non-tech business users) and analysts (tech users), there are two different interactions according to their characteristics.

Jupyter notebook scripts, direct SQL queries ,and ER Diagrams are all used in this database for the analysts. They help the analysts to interact with our database system because analysts are already proficient in programming languages so that they can understand our rationale by reading our scripts. These programming languages can be more clear and practical for analysts to operate. On the other hand, ER Diagrams will help analysts to understand the relational database without any confusion.

For as c-level officers, we are using Metabase to create visualizations to demonstrate our findings/insights. The dashboard in Metabase helps the c-level officers to interact with our database system because it is an easy, open source way for everyone to ask questions and learn from data, especially for people who don't have the technical background. Also, Metabase visualizations will help the c-level officers to grasp the gist in a short period, which is efficient for the organization so they can take some actions directly.

Specifically, we will create Metabase dashboards to answer the following questions:

1. The trend in Terrorist Activities

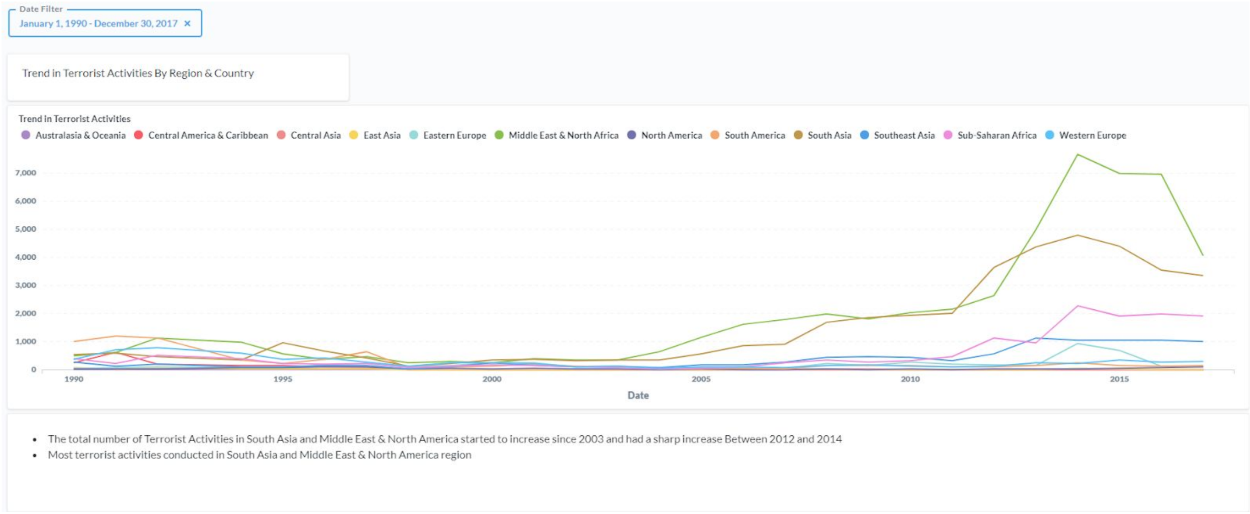
2. What did the Year 2014 bring to the world?

We plan to reduce the redundancy of our data by presenting our data in a historical bias from different aspects such as weapon type, target type and so on.

Metabase Dashboard

1. Trend in Terrorist Activities

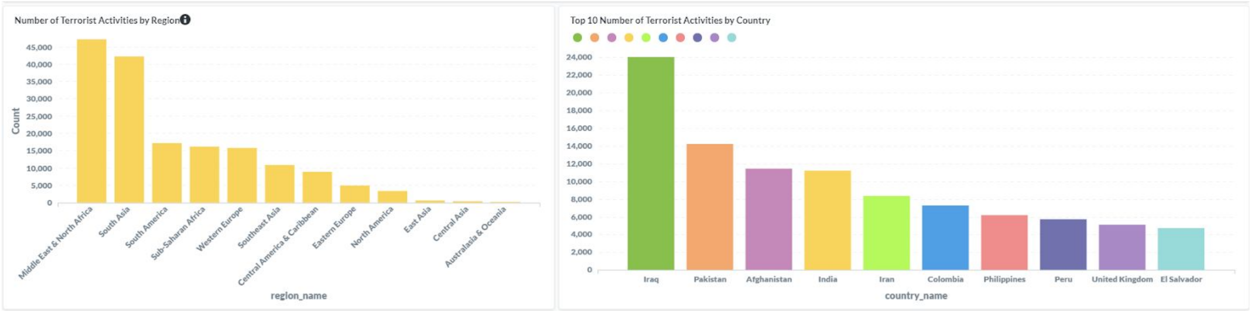
First, we draw the overall trend of global terrorist activities by regions over time. Then, we leveraged the Metabase dashboard to study the overall terrorist activities by region & county. As we can see from the charts, Middle East & North Africa and South Asia are the major areas that suffer from terrorist attacks. Besides, based on the trend chart, we can see that the total number & frequency of terrorist attacks increased dramatically since the start of the millennium.


Date Filter

January 1, 1990 - December 30, 2017 ×

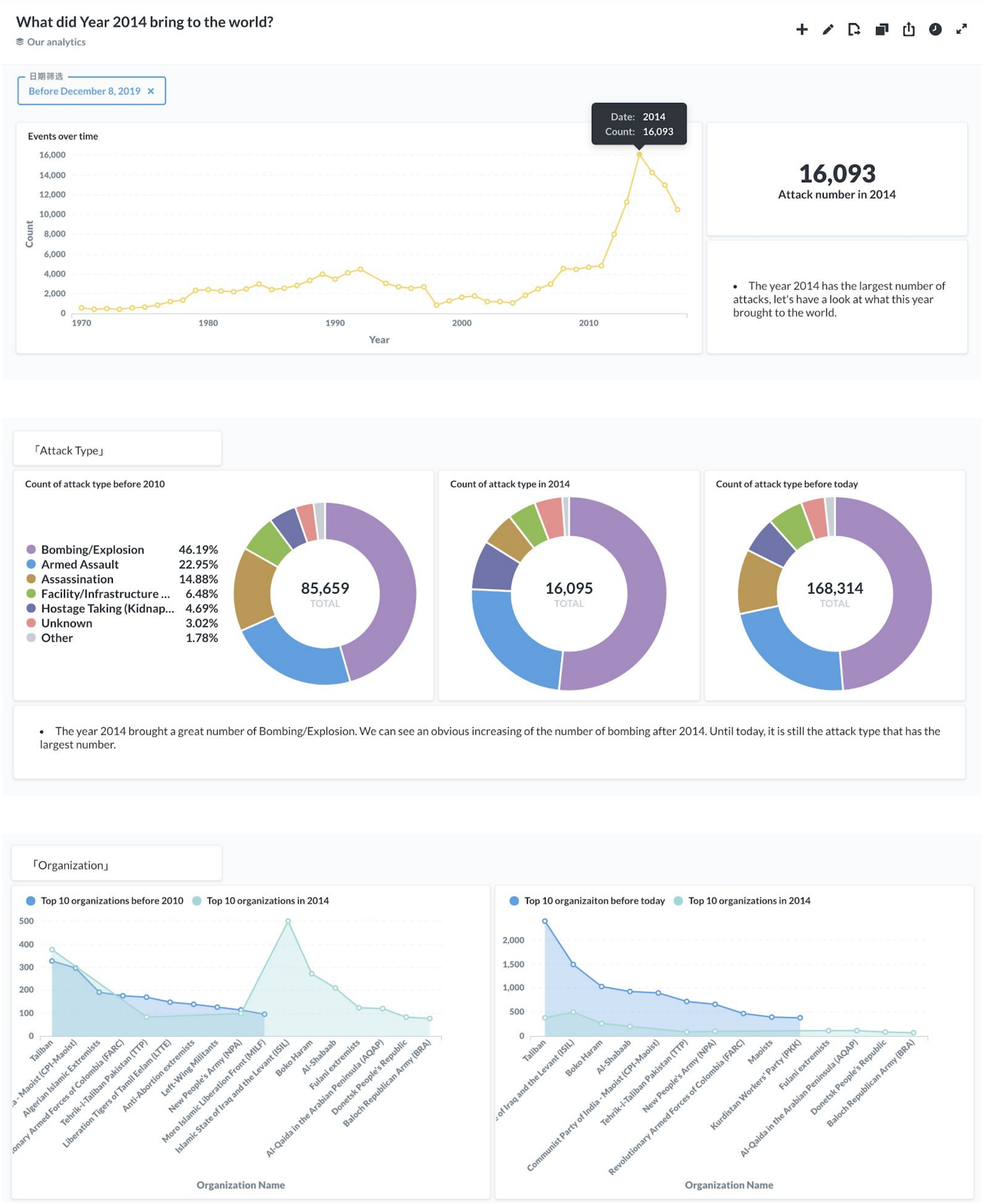
Trend in Terrorist Activities By Region & Country

- The total number of Terrorist Activities in South Asia and Middle East & North America started to increase since 2003 and had a sharp increase Between 2012 and 2014
- Most terrorist activities conducted in South Asia and Middle East & North America region



2. What did the Year 2014 bring to the world?

The year 2014 has the largest number of attacks (16093), so it is necessary to figure out what this year brought to the world. Therefore, we did a dashboard to compare the changes in the attack type and the organizations before and after 2014. So the UN can directly understand the insight.



What did Year 2014 bring to the world?

Our analytics

日期筛选

Before December 8, 2019 ×

16,093

Attack number in 2014

- The year 2014 has the largest number of attacks, let's have a look at what this year brought to the world.

「Attack Type」

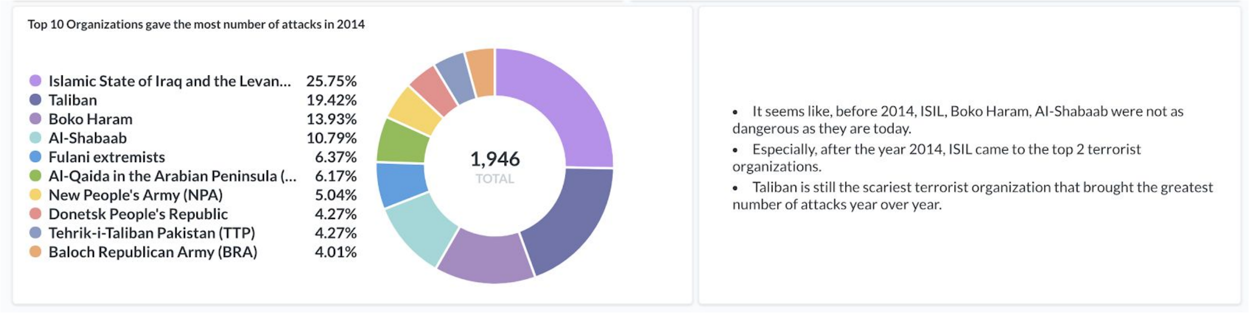
Count of attack type before 2010
85,659
TOTAL
Count of attack type in 2014
16,095
TOTAL
Count of attack type before today
168,314
TOTAL

● Bombing/Explosion	46.19%
● Armed Assault	22.95%
● Assassination	14.88%
● Facility/Infrastructure ...	6.48%
● Hostage Taking (Kidnap...	4.69%
● Unknown	3.02%
● Other	1.78%

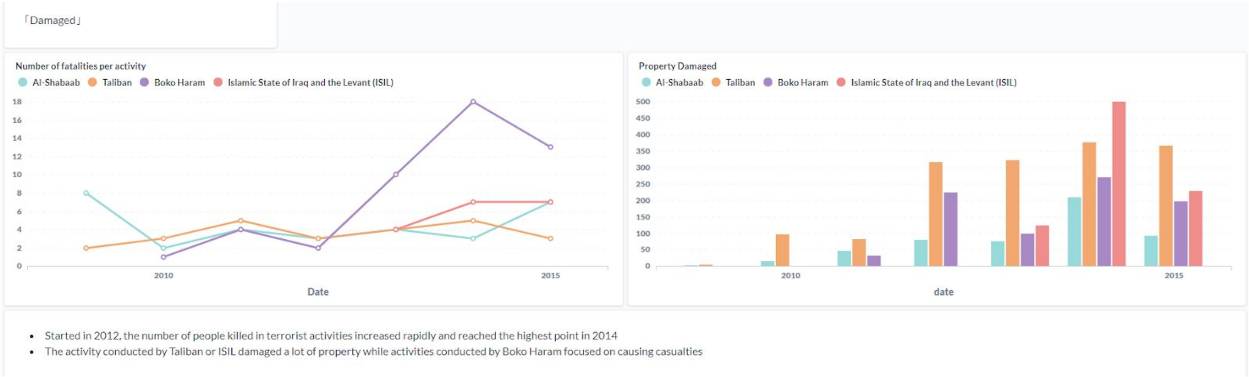
- The year 2014 brought a great number of Bombing/Explosion. We can see an obvious increasing of the number of bombing after 2014. Until today, it is still the attack type that has the largest number.

「Organization」

● Top 10 organizations before 2010
● Top 10 organizations in 2014
500
400
300
200
100
0
Taliban
a - Maoist (CPI-Maoist)
Algerian Islamic Extremists
onary Armed Forces of Colombia (FARC)
Tehrik-i-Taliban Pakistan (TTP)
Liberation Tigers of Tamil Eelam (LTTE)
Anti-Abortion extremists
Left-Wing Militants
New People's Army (NPA)
Moro Islamic Liberation Front (MILF)
Islamic State of Iraq and the Levant (ISIL)
Boko Haram
Al-Shabaab
Fulani extremists
Al-Qaida in the Arabian Peninsula (AQAP)
Donetsk People's Republic
Baloch Republican Army (BRA)
Organization Name
● Top 10 organizaiton before today
● Top 10 organizations in 2014
2,000
1,500
1,000
500
0
Taliban
of Iraq and the Levant (ISIL)
Boko Haram
Al-Shabaab
Communist Party of India - Maoist (CPI-Maoist)
Tehrik-i-Taliban Pakistan (TTP)
New People's Army (NPA)
Revolutionary Armed Forces of Colombia (FARC)
Maoists
Kurdistan Workers' Party (PKK)
Fulani extremists
Al-Qaida in the Arabian Peninsula (AQAP)
Donetsk People's Republic
Baloch Republican Army (BRA)
Organization Name


Top 10 Organizations gave the most number of attacks in 2014
Islamic State of Iraq and the Levan... 25.75%
Taliban 19.42%
Boko Haram 13.93%
Al-Shabaab 10.79%
Fulani extremists 6.37%
Al-Qaida in the Arabian Peninsula (... 6.17%
New People's Army (NPA) 5.04%
Donetsk People's Republic 4.27%
Tehrik-i-Taliban Pakistan (TTP) 4.27%
Baloch Republican Army (BRA) 4.01%
1,946
TOTAL

- It seems like, before 2014, ISIL, Boko Haram, Al-Shabaab were not as dangerous as they are today.
- Especially, after the year 2014, ISIL came to the top 2 terrorist organizations.
- Taliban is still the scariest terrorist organization that brought the greatest number of attacks year over year.



- Started in 2012, the number of people killed in terrorist activities increased rapidly and reached the highest point in 2014
- The activity conducted by Taliban or ISIL damaged a lot of property while activities conducted by Boko Haram focused on causing casualties

Conclusion

In this project, our goal is to analyze the pattern of terrorist attacks and ultimately create an opinion on the relatively risky regions and target terrorist organizations.

Our client, the United Nations, has many archived and unstructured data and needs to store them properly. Our plan is to build a relational database management system to help the client to analyze the international terrorism characteristics. Based on it, the UN can take some actions to different regions in the world to safeguard world peace.

We designed a simple and efficient database with 15 tables. The Global Terrorism Database (GTD) is easily manageable and updatable. We also load all unstructured data from The National Consortium for the Study of Terrorism and Responses to Terrorism (START) to our database via the ETL process.

After the detailed design and analysis, we prepared two deliverables for two types of users who are analysts and C-level users (the UN). Specifically, the analytical procedures give some examples of how analysts can analyze the data through our database, including the direct SQL query and the Python language. From the analytical process, analysts can acquire some in-depth conclusions. In this project, we got the following findings:

1. Even in one country, there are some safe cities while some specific cities are extremely dangerous.
2. Elaborate Points one: even in those most dangerous countries like Iraq, most terrorist activities happened in one-two cities only.
3. The year 2014 has the most terrorist attacks. An interesting fact is that the top 8 years with most attacks are all post-2010. In other words, since 2010, the number of terrorists grew exponentially until 2014.

In terms of the C-level users, the UN, in this case, we are providing the dashboard with the overview and the breakdown of the characteristics of the data through the Metabase, which is visualized and can be easily understood by the non-tech users. Through the Metabase dashboard, we have generated some useful insights:

1. The year 2014 brought a great number of Bombing/Explosion. We can see an obvious increase in the number of bombing after 2014. Until today, it is still the attack type that has the largest number.
2. Before 2014, ISIL, Boko Haram, Al-Shabaab were not as dangerous as they are today.
3. Especially, after the year 2014, ISIL came to the top 2 terrorist organizations.
4. Taliban is still the scariest terrorist organization that brought the greatest number of attacks year over year.
5. Started in 2012, the number of people killed in terrorist activities increased rapidly and reached the highest point in 2014
6. The activity conducted by Taliban or ISIL damaged a lot of property while activities conducted by Boko Haram focused on causing casualties

We are confident that our client, the United Nations, will benefit from our database and take actions to different areas based on the understanding of the terrorism characteristics so that they can safeguard the world peace better.

Appendix: The summary of all the links

Original data set through Kaggle: https://www.kaggle.com/START-UMD/gtd

GTD Database Codebook:https://start.umd.edu/gtd/downloads/Codebook.pdf

Timeline - Project Management through Mondays (we also downloaded it as a pdf file in attachment in case that you don't have the account):
https://columbia154.monday.com/boards/372674143/

Schema code through Github:
https://github.com/bh2709/Team-8-Create-Tables/blob/master/table%20creation

ETL Process code through GIthub:

https://github.com/bh2709/Project/blob/master/Team%208%20Global%20Terrorism%20Analytics%20Project

ER Diagram through Lucidchart:
http://sg.app.lucidchart.com/wf/click?upn=4wgIOpld3pNauqZlzpyPEd2TF-2Bn4I5w1OnJCIMs67c2sc1EhCKC79ggcTFigzYSo8cd4Vogfaf-2FplW8jXRTlEafhNXt53m4vLmzmeTHwmAJxR-2FejXTRoVEi95F59ryO8wj-2BCgNfAD-2FDht9VZn3vAEQNKcTWABS26JyfKWuYKJwcrbwb3Ag-2BzeY-2BrXV8L5hDpf7Y5jYj94hDtwySZX4hvi3FSxDwP49Wt-2FhuDoGfyg5fSnmgVTjpvUEum-2Fw7BwcKf9ryD52n-2BltDwATOKK182sUqiUbWYdWoTPcWP79j5ziTfkeTpvk78eNSqzCD3kDi0_uda6uzIhbO3zgsIvEFi1g701ALem70WBInzSW9qupXbyVAh2DnDxoCkgca-2FcjBdLljUhSN-2BxbCAI3IqxtOIYHZoj4MRkR6ewoEt4X0N27kcnRFuW-2BXh5we3EeQrp1D-2F1V-2FXT-2B5wSUWW-2Fo8zVhPyc-2FTaUmcWXZMkPrhSQmTVXkhXCd6LvnLo0Ys7arwMvPYwzQPCHxo5NYzqaqEbxGOmXLJPeIRuR28E13jOrpVuCPEhHj65ecQBbK8sZA0NkOgEUCWJKv73amVWs74FeVIQwt5VOnbKY-2B2nerlUXniu-2B9H5Nq28ior1SB5Wj8soBrjqXu-2FIj9loaxjr879HDAk3-2F6CeLWWvmhUX0vF907cdmTtRAKF7u1pH0SroVeiaqFtjNsBFUrBczg9L9Wmu4HPrYeMp6jCE79iw-2BSDdYV5BSKJdL7sK7SoJiqFE5E-2FV4Sj40GQjutVLnJyFSrVuVgXn0SHmw3uC0qeFK9t3fiJWMKp6b8cwbY-2B2Ghw2zVgyQsg2CJzGTvIxI43fRCLIdZMW6NdpnOUHnqby7DdI4cDq1i3Sh7IbpoDyosIQUCEU7iOJD-2Bgd-2Fiawf6aCWTEMtoydjpQ-3D-3D

Metabase Dashboard 1:
http://f19server.apan5310.com:3208/public/dashboard/320483fc-0745-4881-a528-b47ccfa40dab
Metabase Dashboard 2:
http://f19server.apan5310.com:3208/public/dashboard/96e285f1-7659-47af-88c0-12fab0bc8cc7