ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

BUSINESS ANALYTICS
Master of Science

Data Management and Business Intelligence

Global Terrorism Data Warehouse

More than 180,000 terrorist attacks worldwide, 1970-2017

Team Members

Konstantina Georgiopoulou
(p2822004)

Anastasios Theodorou
(p2822007)

Stavros Kasiaris
(p2822022)

Contents

The kind of target application

We are trying to build a data warehouse referring to the terrorist attacks around the world. We aim to understand the motives of the perpetrators and analyze the consequences of these actions. To do this, we found in the kaggle the Global Terrorism Database (GTD) in which more than 180,000 Terrorist Attacks information was included from 1970 through 2017.

Information about the Dataset

To understand the structure of the data and extract the information that we are going to use, it is very important to mention details about the way this dataset was constructed.

To be more specific, The GTD is maintained by the National Consortium for the Study of Terrorism and Responses to Terrorism (START) at the University of Maryland. In the dataset, we have the following categories: incident date, incident location, incident information, attack information, target/victim information, perpetrator information, perpetrator statistics, claims of responsibility, weapon information, casualty information, consequences, kidnapping/hostage taking information, additional information, and source information. In our analysis, not all of the above categories are included.

The Origins of the GTD

In order to acknowledge how the data are structured , we have to be informed about the way this data set was created. In other words, the current GTD is the product of several phases of data collection efforts, each relying on publicly available, unclassified source materials. These include media articles and electronic news archives, and to a lesser extent, existing data sets, secondary source materials, such as books and journals, and legal documents. The original set of incidents that comprise the GTD occurred between 1970 and 1997 and were collected by the Pinkerton Global Intelligence Service (*PGIS*)—a private security agency. After *START* completed digitizing of these handwritten records in 2005, they collaborated

with the Center for Terrorism and Intelligence Studies (*CETIS*) to continue data collection beyond 1997 and expand the scope of the information recorded for each attack. CETIS collected GTD data for terrorist attacks that occurred from January 1998 through March 2008, after which ongoing data collection transitioned to the Institute for the Study of Violent Groups (*ISVG*). ISVG continued as the primary collector of data on attacks that occurred from April 2008 through October 2011. Beginning with cases that occurred in November 2011, all ongoing GTD data collection is conducted by START staff at the University of Maryland.

Dates of GTD Attacks	PGIS	CETIS	ISVG	START
1/1/1970 – 12/31/1997	X			X
1/1/1998 – 3/31/2008		X		X
4/1/2008 – 10/31/2011			X	X
11/1/2011 – 12/31/2018 (ongoing)				X

Table: GTD Data Collection Phases by Collection Institution

Accuracy of the data

An event in order to be recorded in the GTD, it must be documented by at least one high-quality source. Events that are only documented by distinctly biased or unreliable sources are not included in the GTD. However, the GTD does include certain information from potentially biased sources, such as perpetrator claims of responsibility or details about the motive of the attack.

GTD Definition of Terrorism and Inclusion Criteria

The GTD defines a terrorist attack 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. In practice, this means to consider an incident for inclusion in the GTD, all three of the following attributes must be present:

- ❏ The incident must be intentional,
- ❏ The incident must entail some level of violence or immediate threat of violence and
- ❏ The perpetrators of the incidents must be sub-national actors

In addition, at least two of the following three criteria must be present for an incident to be included in the GTD:

- Criterion 1: The act must be aimed at attaining a political, economic, religious, or social goal.
- Criterion 2: There must be evidence of an intention to coerce, intimidate, or convey some other message to a larger audience (or audiences) than the immediate victims.
- Criterion 3: The action must be outside the context of legitimate warfare activities

For all the above analysis, we can assume that the informations of the dataset are accurate and they were collected with carefulness. So, we are more than certain of their integrity and we can safely include them in our analysis, in order to draw conclusions about terrorist attacks over the years.

Problems that we encountered

The first problem that we faced is that not all the columns were needed in our analysis. To be clear, from the 135 columns that the initial data set had, we concluded to have only 41 in our fact table. That means, we worked a lot to clean the data, in a way to understand what the columns present and if we needed them or not.

The second and most crucial problem was that we observed differences in encoding. For instance, in the city column, the unknown was written in two different ways, as "unknown" and "Unknown". That means that we had duplication of this content with the same value.

	lat_id	lat
1140	1140	Zamboanga City
1223	1223	Zamboanga city

Table: An example of false encoding

Also, we had to handle the missing values, which are important for our analysis.

Details about the analysis

We want to do descriptive analysis, in other words, we want to use the data to understand past events. To be more specific, we built the data warehouse to answer the following questions:

- Are terrorist attacks increasing or decreasing over the years?
- Are there months or even days that the number of terrorist attacks are significantly more?
- In which locations do the terrorist attacks mainly take place?
- What are the main types of the attacks?
- What weapons are used more in each attack?
- What are the consequences of each attack type?
- Is there any difference in the consequences across the regions?
- What has happened to our country across the years concerning terrorist attacks?

Description of the relational design of the fact and dimension tables, import methods, cleaning/transformation procedures

The first thing that we had to do was design the ETL process.

Extract

To build the data warehouse for the Global Terrorism, we had a CSV file in which was included the information for all the attacks (date, location, attack type, weapon type, consequences, etc).

Data Transformation

We had to deal with the following issues to transform our data:

1st Issue: **Filtering-** What columns should we keep to have a better analysis? Not all the columns were useful, so we had to keep the columns that help us answer the questions.

Solution: We read all the documentation to understand the definition of each variable and according to our objectives, we examined if we had to keep them or not in our analysis.

Type	Column name	Comment
I. GTD ID and Date		
	eventid	We keep it with another encoding. For further explanation, we use an ascending id to keep all the ids instead of having as id the date as event id.
	iyear/imonth/iday	We keep them and we merge them into one column.
	approxdate	We do not want it because we want to have a more general overview of our analysis.
II. Incident Information		
	crit1, crit2, crit3	These variables record which of the inclusion criteria are met. We want and we convert them into one column.
III. Incident Location		
	country,region, city	We want to include the Location.
IV. Attack Information		
	attacktype1, attacktype2, attacktype3	We keep only the attacktype1 because the other 2 types have many missing values and we want to be more focused on the mean type of incident.
V. Weapon Information		
	weaptype1,weapsubtype1\ weaptype2,weapsubtype2\ weaptype3,weapsubtype3	We keep only the main weapon type and subtype.
VI. Target/Victim Information		
	targtype1/targsubtype1	We want to include the information about victims in our analysis.

In the above table, we see whether or not a variable was included in our analysis, why, and how we handled it.

Through this procedure, we ended having 41 columns in our fact table.

Some of the 135 columns in our first data set.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
eventid	iyear	imonth	iday	approxdate	extended	resolution	country	country_txt	region	region_txt	provstate	city	latitude	longitude	specificity	vicinity	location	summary
1,97E+11	1970	7	2		0		58	Dominican Republic	2	Central America & Caribbean		Santo Domingo	18.456.792	-69.951.164	1	0		
1,97E+11	1970	0	0		0		130	Mexico	1	North America	Federal	Mexico city	19.371.887	-99.086.624	1	0		
1,97E+11	1970	1	0		0		160	Philippines	5	Southeast Asia	Tarlac	Unknown	15.478.598	120.599.741	4	0		
1,97E+11	1970	1	0		0		78	Greece	8	Western Europe	Attica	Athens	3.799.749	23.762.728	1	0		
1,97E+11	1970	1	0		0		101	Japan	4	East Asia	Fukouka	Fukouka	33.580.412	130.396.361	1	0		
1,97E+11	1970	1	1		0		217	United States	1	North America	Illinois	Cairo	37.005.105	-89.176.269	1	0		1/1/1970: Unknown African American assailants fired several b
1,97E+11	1970	1	2		0		218	Uruguay	3	South America	Montevideo	Montevideo	##########	-56.187.214	1	0		
1,97E+11	1970	1	2		0		217	United States	1	North America	California	Oakland	37.791.927	-122.225.906	1	0	Edes Substation	1/2/1970: Unknown perpetrators detonated explosives at the Pa
1,97E+11	1970	1	2		0		217	United States	1	North America	Wisconsin	Madison	43.076.592	-89.412.488	1	0		1/2/1970: Karl Armstrong, a member of the New Years Gang, thre
1,97E+11	1970	1	3		0		217	United States	1	North America	Wisconsin	Madison	4.307.295	-89.386.694	1	0		1/3/1970: Karl Armstrong, a member of the New Years Gang, brol
1,97E+11	1970	1	1		0		217	United States	1	North America	Wisconsin	Baraboo	434.685	-89.744.299	1	0		
1,97E+11	1970	1	6		0		217	United States	1	North America	Colorado	Denver	39.758.968	-104.876.305	1	0		1/6/1970: Unknown perpetrators threw a Molotov cocktail into
1,97E+11	1970	1	8		0		98	Italy	8	Western Europe	Lazio	Rome	41.890.961	12.490.069	1	0		
1,97E+11	1970	1	9		0		217	United States	1	North America	Michigan	Detroit	42.331.685	-83.047.924	1	0		1/9/1970: Unknown perpetrators set off a firebomb at the Packa
1,97E+11	1970	1	9		0		217	United States	1	North America	Puerto Rico	Rio Piedras	18.386.932	-66.061.127	1	0	Caparra Shopping Center	1/9/1970: The Armed Commandos of Liberation claimed credit f
1,97E+11	1970	1	10		0		499	East Germany (GDR)	9	Eastern Europe	Berlin	Berlin	5.250.153	13.401.851	1	0		
1,97E+11	1970	1	11		0		65	Ethiopia	11	Sub-Saharan Africa	Unknown	Unknown			5	0		
1,97E+11	1970	1	12		0		217	United States	1	North America	New York	New York City	40.697.132	-73.931.351	1	0	Brooklyn	1/12/1970: Unknown perpetrators threw a pipe bomb into the va
1,97E+11	1970	1	12		0		217	United States	1	North America	Puerto Rico	Rio Grande	18.379.998	-65.830.948	1	0		1/12/1970: Unknown perpetrators detonated a bomb at a Genera
1,97E+11	1970	1	13		0		217	United States	1	North America	Washington	Seattle	47.610.786	-122.331.306	1	0		1/13/1970: Unknown perpetrators firebombed Fuson's Departme
1,97E+11	1970	1	14		0		217	United States	1	North America	Illinois	Champaign	40.116.748	-8.823.927	1	0	Champaign Police Department	1/14/1970: Suspected Black militants threw two firebombs into t
1,97E+11	1970	1	15		0		218	Uruguay	3	South America	Montevideo	Montevideo	##########	-56.187.214	1	0		
1,97E+11	1970	1	19		0		217	United States	1	North America	Washington	Seattle	47.610.786	-122.331.306	1	0	Seattle University	1/17/1970: Three African Americans were suspected of detonatin
1,97E+11	1970	1	19		0		217	United States	1	North America	Washington	Seattle	47.610.786	-122.331.306	1	0		1/17/1970: Silas and Judith Bissell placed an explosive device in
1,97E+11	1970	1	19	January 19-20, 1970	0		217	United States	1	North America	New Jersey	Jersey City	40.717.892	-74.067.467	1	0	Front of building	1/19/1970: Unknown perpetrators ignited a fire at the Black Pan
1,97E+11	1970	1	20		0		83	Guatemala	2	Central America & Caribbea	Guatemala	Guatemala City	14.622.869	-90.529.068	1	0		
1,97E+11	1970	1	21		0		160	Philippines	5	Southeast Asia	Metropolitan Manila	Quezon City	1.467.428	121.057.495	1	0		
1,97E+11	1970	1	22		0		222	Venezuela	3	South America	Caracas	Caracas	10.482.834	-66.962.128	1	0		
1,97E+11	1970	1	22		0		217	United States	1	North America	Nebraska	South Sioux City	4.247.031	-96.413.949	1	0		1/22/1970: Unknown perpetrators detonated an explosive devic
1,97E+11	1970	1	25		0		217	United States	1	North America	Mississippi	West Point	3.360.651	-88.650.419	1	0		1/25/1970: Unknown perpetrators burnt down a building that w
1,97E+11	1970	1	25		0		217	United States	1	North America	New York	New York City	40.697.132	-73.931.351	1	0	Manhattan	1/25/1970: Responding to what authorities believe was a fake re
1,97E+11	1970	1	26		0		217	United States	1	North America	Mississippi	West Point	3.360.651	-88.650.419	1	0		1/26/1970: Unknown African American assailants, unsatisfied v
1,97E+11	1970	1	26		0		217	United States	1	North America	New York	New York City	40.697.132	-73.931.351	1	0	Queens	1/26/1970: Police officer Joseph Blik was on foot patrol in Quee
1,97E+11	1970	1	27		0		217	United States	1	North America	Ohio	Norwalk	41.241.996	-82.615.241	1	0		1/27/1970: Unknown perpetrators broke into the Selective Servic
1,97E+11	1970	1	28		0		499	East Germany (GDR)	9	Eastern Europe	Berlin	Berlin	5.250.153	13.401.851	1	0		
1,97E+11	1970	1	28		0		217	United States	1	North America	Washington	Seattle	47.610.786	-122.331.306	1	0	Front of building	1/28/1970: Unknown perpetrators detonated an explosive devic
1,97E+11	1970	1	30		0		217	United States	1	North America	Nebraska	South Sioux City	42.479.999	-96.413.046	1	0	210 East 10 St.	1/30/1970: In a series of two incidents that were possibly relate
1,97E+11	1970	1	30		0		217	United States	1	North America	Nebraska	South Sioux City	4.246.634	-9.641.405	1	0	2503 Dakota Ave.	1/30/1970: In a series of two incidents that were possibly relate
1,97E+11	1970	1	30		0		217	United States	1	North America	Florida	Coral Gables	25.720.851	-80.277.857	1	0		1/20/1970: Unknown perpetrators firebombed the premises of th
1,97E+11	1970	1	31		0		160	Philippines	5	Southeast Asia	Tarlac	Bamban	15.675.051	120.331.618	1	0		
1,97E+11	1970	1	31		0		217	United States	1	North America	Colorado	Denver	39.758.968	-104.876.305	1	0	2105 Decatur Street	1/31/1970: Unknown perpetrators bombed the Police Band Build
1,97E+11	1970	2	1		0		217	United States	1	North America	Oregon	Portland	45.511.795	-122.675.629	1	0	North Portland neighborhood	2/1/1970: An unknown perpetrator broke into the Northeast Ore
1,97E+11	1970	2	1		0		217	United States	1	North America	Illinois	Cairo	37.005.105	-89.176.269	1	0	Pyramid Courts	2/1/1970: Unknown perpetrators fired shots at an Illinois State
1,97E+11	1970	2	2		0		217	United States	1	North America	New York	New York City	40.697.132	-73.931.351	1	0	Manhattan	2/2/1970: An incendiary device was discovered at Mays departm

Through Cleaning

We ended having 41 columns in Fact Table

Column Name	Data Type	Allow Nulls
eventid	int	☐
date	smallint	☑
extended	smallint	☑
crit_id	smallint	☑
country	smallint	☑
region	smallint	☑
prov_id	smallint	☑
city_id	int	☑
vicinity	smallint	☑
latitude	int	☑
longitude	int	☑
attacktype	smallint	☑
success	smallint	☑
suicide	smallint	☑
weaptype	smallint	☑
weapsubtype	smallint	☑
targtype	smallint	☑
targsubtype	smallint	☑
perp_id	smallint	☑
gsub_id	smallint	☑
guncertain	smallint	☑
individual	smallint	☑
nperps	smallint	☑
nperpcap	int	☑
nkill	smallint	☑
nkillter	smallint	☑
nwound	smallint	☑
nwoundte	smallint	☑
property	smallint	☑
propextent	smallint	☑
propvalue	bigint	☑
ishostkid	smallint	☑
nhostkid	smallint	☑
nhours	smallint	☑
ndays	smallint	☑
ransom	smallint	☑
ransomamt	int	☑
ransompaid	int	☑
hostkidoutcome	smallint	☑
nreleased	smallint	☑
summary	varchar(2435)	☑

2nd Issue: **Merge multiple columns into a single column-** We wanted to have fewer columns in the fact table for better management and analysis.

Solution: We gathered together the columns that had the same meaning or they could be grouped through a certain pattern. For example, we grouped the date to create a hierarchy and also we grouped the criteria. To do this, we imported the CSV file in R, and we applied the following function:

```
#read excels
install.packages("readxl")
library(readxl)
year <- read_excel(choose.files())
month <- read_excel(choose.files())
date <- read_excel(choose.files())
fact <- read_excel(choose.files())

#combine excels
for(i in 1:nrow(fact)){
  for (j in 1:nrow(year)){
    if (fact$iyear[i]==year$iyear[j]){
      fact$iyear[i] <- year$id_year[j]
    }
  }
}
#create a new column in the main data frame and combine the dates
fact$date <- paste(fact$iyear,"_", fact$imonth, "_" , fact$iday)
#how many unique days we have
x <- unique(fact$date)
#create a new dataframe with all the unique dates
date_dim <- data.frame(matrix(ncol = 2, nrow=length(x)))
colnames(date_dim) <- c ('id', 'un_date')
#fill the data frame
for(i in 1:length(unique(fact$date))){
  date_dim$id[i] <-i
  date_dim$un_date[i] <- x[i]
}
#mapping the ids in date dimension with the dates in main fact table
for(i in 1:nrow(fact)){
  for (j in 1:nrow(date_dim)){
    if (fact$date[i]==date_dim$un_date[j]){
      fact$date[i] <- date_dim$id[j]
    }
  }
}
#export csv
write.csv(date_dim,"C:/Users/konst/OneDrive/Υπολογιστής/D.csv", row.names = FALSE)
write.csv(fact,"C:/Users/konst/OneDrive/Υπολογιστής/F.csv", row.names = FALSE)
write.csv(year,"C:/Users/konst/OneDrive/Υπολογιστής/Y.csv", row.names = FALSE)
write.csv(month,"C:/Users/konst/OneDrive/Υπολογιστής/M.csv", row.names = FALSE)
write.csv(date,"C:/Users/konst/OneDrive/Υπολογιστής/Da.csv", row.names = FALSE)
```

With this way we created the hierarchy:

Date dimension < date < month < year

which are tables that have been connected through foreign keys and the Date Dimension is connecting with the Fact table.

Also, with the above code we handled the criteria dimensions

```
#read excels
library(readxl)
crit <- read_excel(choose.files())
#combine criteria
crit$comb <- paste(crit$crit1,",", crit$crit2, "," , crit$crit3)
#how many unique criteria we have
x <-unique(crit$comb)
#create a new dataframe with all the unique criteria
y<- data.frame(matrix(ncol = 2, nrow=length(x)))
colnames(y) <- c ('crit_id', 'crit')
#fill the data frame
for(i in 1:length(unique(crit$comb))){
  y$crit_id[i]<- i
  y$crit[i] <- x[i]

}
#mapping
for(i in 1:nrow(crit)){
  for (j in 1: nrow(y)){
    if (crit$comb[i]==y$crit[j]){
      crit$comb[i]<-y$crit_id[j]
    }
  }
}
#export csv
write.csv(y,"C:/Users/konst/OneDrive/Υπολογιστής/crit_dimm.csv", row.names = FALSE)
write.csv(crit,"C:/Users/konst/OneDrive/Υπολογιστής/crit_dff.csv", row.names = FALSE)
```

With this, we created one column in the Fact Table in which criteria were represented, and one Criteria Dimension (Junk Dimension), which keeps all the possible combinations of the criteria.

3rd Issue: **Encoding of the city:** We have the country_id, the region_id, but not the city in the same encoding.

Solution: We had to do the same procedure in order to create the City Dimension.

```
#read excels
library(readxl)
library(foreign)
pc <- read_excel("Prov_City_ID.xlsx", col_types = "text")
pc <- as.data.frame(pc)
state <-  read_excel("State_ID.xlsx")
state <- as.data.frame(state)

city <-  read_excel("City_ID.xlsx")
city <- as.data.frame(city)

#combine excels
pcity <- tolower(pc$city)
ccity <- tolower(city$city)
pstate <- tolower(pc$provstate)
sstate <- tolower(state$state)

for (i in 1:nrow(pc)){
  if (is.na(pc$provstate[i])){
    pc$prov_id[i] <- 0
  } else {
    for (j in 1:nrow(state)){
      if (pstate[i]==sstate[j]){
        pc$prov_id[i] <- state$state_id[j]
        break
      }
    }
  }
}
for (i in 1:nrow(pc)){
  if (is.na(pc$city[i])){
    pc$city_id[i] <- 0
  } else {
    for (j in 1:nrow(city)){
      if (pcity[i]==ccity[j]){
        pc$city_id[i] <- city$city_id[j]
        break
      }
    }
  }
}
write.csv(pc,"C:/Users/tasso/Desktop/State_City_ID.csv", row.names = FALSE)
```

4th Issue: **Different spelling of the same city.**

Solution: When we inserted the values in the R we were obliged to convert them to lower letters through the tolower() function, because there were different names for the same value (e.g Mexico City and Mexico city).

5th Issue: **Encoding of the Perpetrator and Different spelling.** We had the same problems and with the city column.

Solution: We inserted that excels in R code and converted their values to lower letters in order to create the dimension table.

```
#read excels
library(readxl)
library(foreign)

perp <- read_excel("Perp_Dimention_R.xlsx", col_types = "text")
perp <- as.data.frame(perp)
perp_id <-  read_excel("Perp_name_id.xlsx")
perp_id <- as.data.frame(perp_id)
sub_id <-  read_excel("Perp_sub_id.xlsx")
sub_id <- as.data.frame(sub_id)

pperp <- tolower(perp$gname)
sperp <- tolower(perp_id$gname)
psub <- tolower(perp$gsubname)
ssub <- tolower(sub_id$gsubname)

for (i in 1:nrow(perp)){
  for (j in 1:nrow(perp_id)){
    if (pperp[i]==sperp[j]){
      perp$perp_id[i] <- perp_id$un_id[j]
      break
    }
  }
}

for (i in 1:nrow(perp)){
  if (is.na(perp$gsubname[i])){
    perp$gsub_id[i] <- 0
    next
  } else {
    for (j in 1:nrow(sub_id)){
      if (psub[i]==ssub[j]){
        perp$gsub_id[i] <- sub_id$gsub_id[j]
        break
      }
    }
  }
}

write.csv(perp,"C:/Users/tasso/Desktop/perp_id.csv", row.names = FALSE)
```

6th Issue: **Using rules and lookup tables for data standardization**. In other words, we had to create the Dimension Tables and did the mapping between Fact Table and Dimension Table.

Solution: We created CSV files for every categorical or binary variable with the help of the Excel in order to keep the numeric variables, such as

longitude and latitude in the Fact Table, and in this way, when we inserted the files in the database we designed the snowflake schema.

For instance, we had a variable named *extended*, which is defined as binary, and it means whether or not an event lasts more than 24 hours. So, we created a Binary Dimension Table in which all the binary variables including the *extended* to be connected with foreign keys, 0 for No, 1 for Yes.

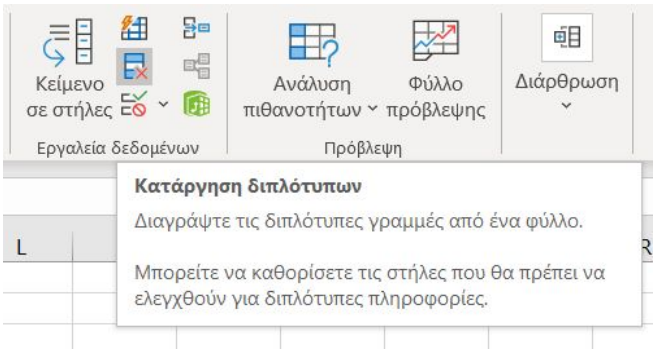
Another example is this of the *attack type*. This field captures the general method of the attack and often reflects the broad class of tactics that are used. It consists of nine categories, which are defined below.

Attack Type Hierarchy:

1. Assassination
2. Hijacking
3. Kidnapping
4. Barricade / Incident
5. Bombing/Explosion
6. Armed Assault
7. Unarmed Assault
8. Facility/Infrastructure Attack
9. Unknown

The steps that we followed are these:

attacktype	attacktype_txt
	Assassination
	Hostage Taking (Kidnapping)
	Assassination
	Bombing/Explosion
	Facility/Infrastructure Attack
	Armed Assault
	Assassination
	Bombing/Explosion
	Facility/Infrastructure Attack
	Facility/Infrastructure Attack
	Bombing/Explosion
	Facility/Infrastructure Attack
	Hijacking
	Facility/Infrastructure Attack
	Facility/Infrastructure Attack
	Bombing/Explosion
	Unknown
	Bombing/Explosion
	Bombing/Explosion
	Facility/Infrastructure Attack
	Facility/Infrastructure Attack
	Armed Assault



attacktype	attacktype_txt
1	Assassination
2	Hostage Taking (Kidnapping)
3	Bombing/Explosion
4	Facility/Infrastructure Attack
5	Armed Assault
6	Hijacking
7	Unknown
8	Unarmed Assault
9	Hostage Taking (Barricade Incident)

We created a new blank Excel sheet and after that we copied and pasted the column of "attacktype_txt" in order to select the duplication removal (the above picture in greek language)to delete all the values that are repeated. Finally, we added numbers in the rows and with that we ended up with the 9 unique attack types. So, we created a Dimension Table called Attack Type, in which all this information was included and after that we did the mapping.

The *extended* and *attack type* were 2 examples. We did this job for many other variables.

7th Issue: **Missing Values.** How do we deal with this issue?

Solution: In R, the empty values are encoding as NA. In this step (Data Transformation) we kept them as such. When we inserted them in the database we did the transformation there and we converted them to null with the following SQL code:

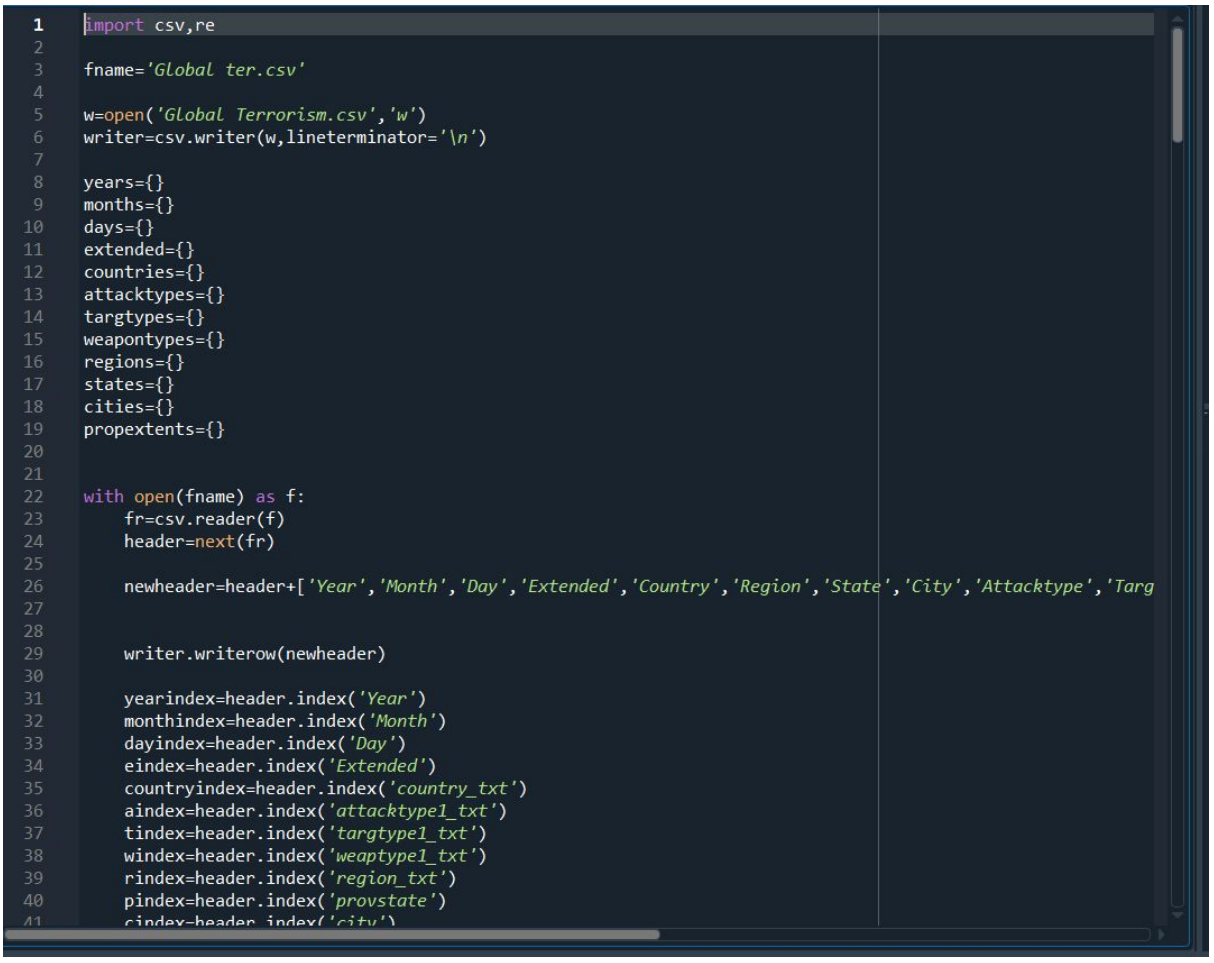
update [Terrorist_DB].[dbo].[Main_Fact_Table] set [nkillter] = NULL where [nkillter] = 'NA'

8th Issue: **Replacing values.** For example, we had the -99, which means unknown and we were enforced to replace it with 0, in order to do calculations such as sum.

Solution: We wrote SQL code and we replaced these values with 0

update [Terrorist_DB].[dbo].[Main_Fact_Table] set [nhostkid] = 0 where [nhostkid] = -99

Another approach for the data transformation is by Python. For some cases it was helpful to use code in order to read the dataset and create some of the dimensional tables.



```
import csv,re

fname='Global ter.csv'

w=open('Global Terrorism.csv','w')
writer=csv.writer(w,lineterminator='\n')

years={}
months={}
days={}
extended={}
countries={}
attacktypes={}
targtypes={}
weapontypes={}
regions={}
states={}
cities={}
propextents={}


with open(fname) as f:
    fr=csv.reader(f)
    header=next(fr)

    newheader=header+['Year','Month','Day','Extended','Country','Region','State','City','Attacktype','Targ


    writer.writerow(newheader)

    yearindex=header.index('Year')
    monthindex=header.index('Month')
    dayindex=header.index('Day')
    eindex=header.index('Extended')
    countryindex=header.index('country_txt')
    aindex=header.index('attacktype1_txt')
    tindex=header.index('targtype1_txt')
    windex=header.index('weaptype1_txt')
    rindex=header.index('region_txt')
    pindex=header.index('provstate')
    cindex=header.index('city')
```

```
    windex=header.index('weaptype1_txt')
    rindex=header.index('region_txt')
    pindex=header.index('provstate')
    cindex=header.index('city')
    nindex=header.index('nkill')
    propertyindex=header.index('property')
    xindex=header.index('propextent_txt')
    woundedindex=header.index('nwound')
    perps=header.index('nperps')

    print(header)

    for row in fr:

        my_region=row[rindex]
        my_year=row[yearindex]
        my_month=row[monthindex]
        my_day=row[dayindex]
        my_extended=row[eindex]
        my_country=row[countryindex]
        my_attacktype=row[aindex]
        my_targtype=row[tindex]
        my_weaptype=row[windex]
        my_nkill=row[nindex]
        my_city=row[cindex]
        my_state=row[pindex]
        my_propextent_txt=row[xindex]
        my_wounded=row[woundedindex]
        my_perps=row[perps]
        my_property=row[propertyindex]
```

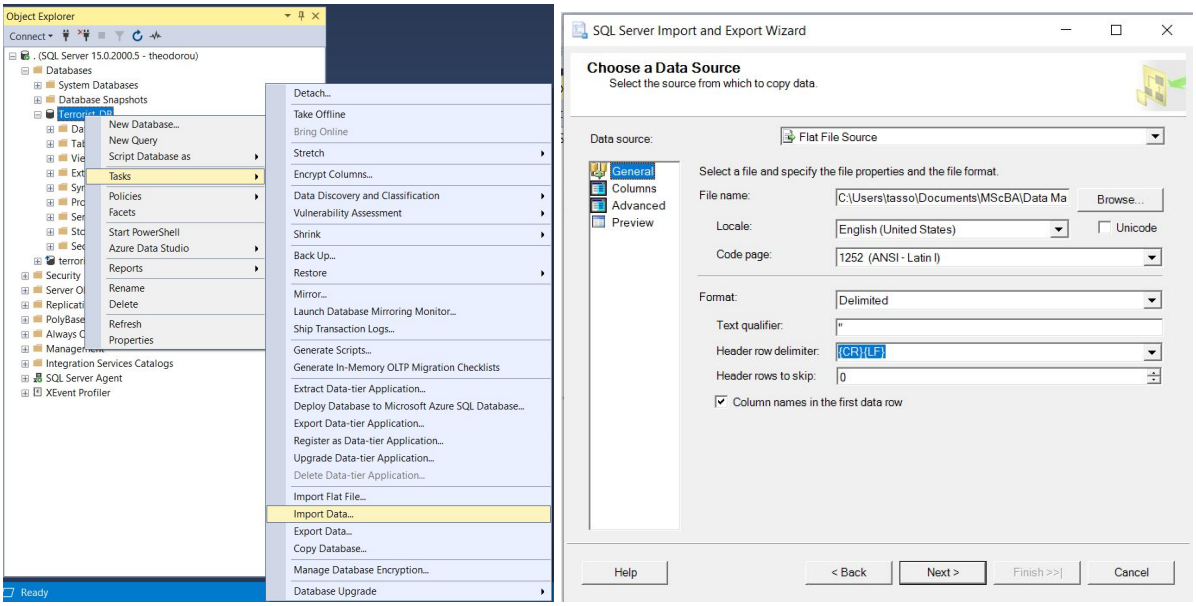
```

    if my_targtype not in targtypes:
        targtypes[my_targtype]=len(targtypes)

    tnum=targtypes[my_targtype]

    if my_attacktype not in attacktypes:
        attacktypes[my_attacktype]=len(attacktypes)

    anum=attacktypes[my_attacktype]


    if my_region not in regions:
        regions[my_region]=len(regions)

    ynum=regions[my_region]


    if my_propextent_txt not in propextents :
        propextents[my_propextent_txt]=len(propextents)

    xnum=propextents[my_propextent_txt]


    if my_state not in states:
        states[my_state]=len(states)

    pnum=states[my_state]


    if my_city not in cities:
        cities[my_city]=len(cities)

    cnum=cities[my_city]

    writer.writerow(row+[yearnum,monthnum,daynum,exnum,countrynum,ynum,pnum,cnum,anum,tnum,wnum,my_propert
```

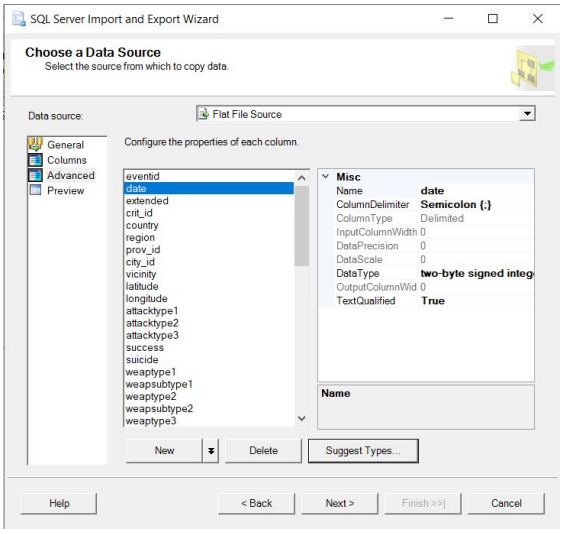
Data Loading

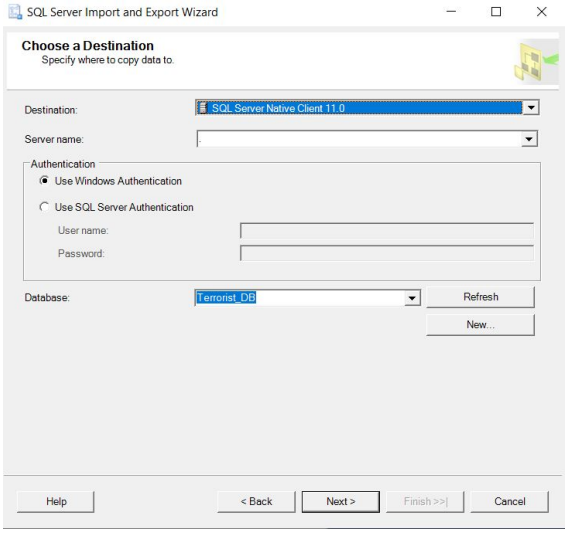
After the creation of the CSV files that we wanted to include in our analysis, in this step we had to load the data in the MSSQL.

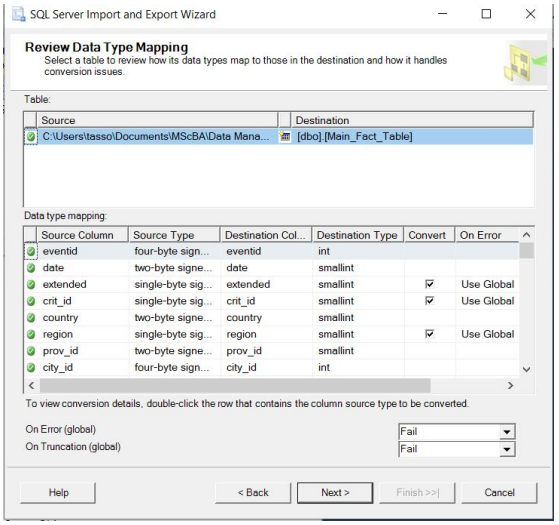
Following the typical procedure, we inserted the files, we did the mapping with the foreign keys and we ended with the below snowflake schema.

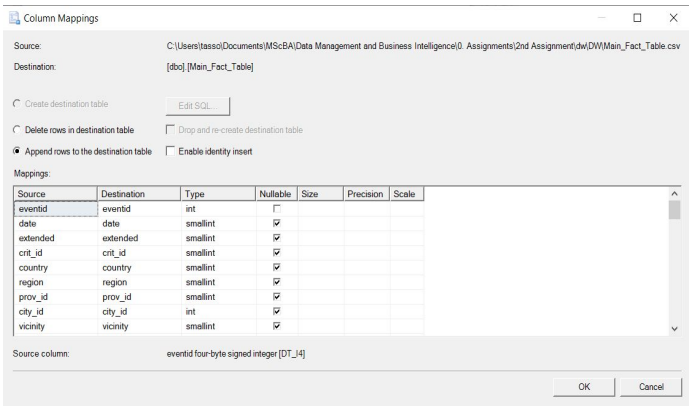
Insert files:

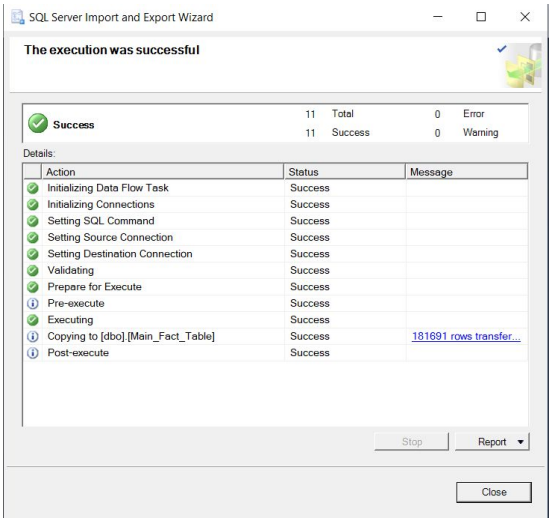






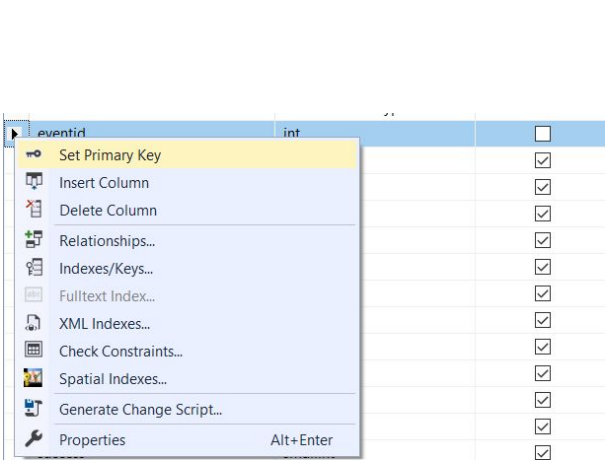


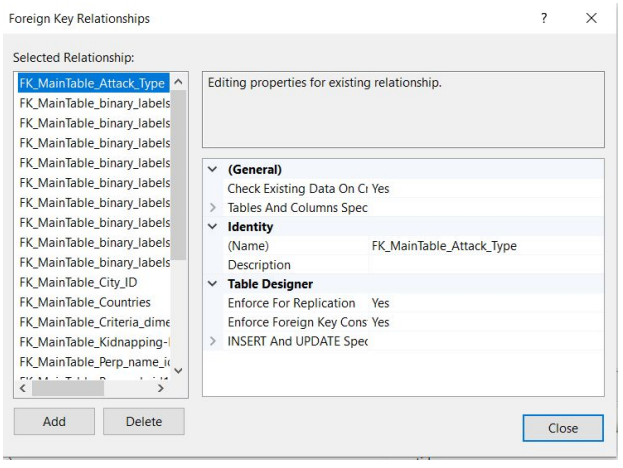




Successful Inserting! We did that process for every CSV file.

Set Primary and foreign keys:





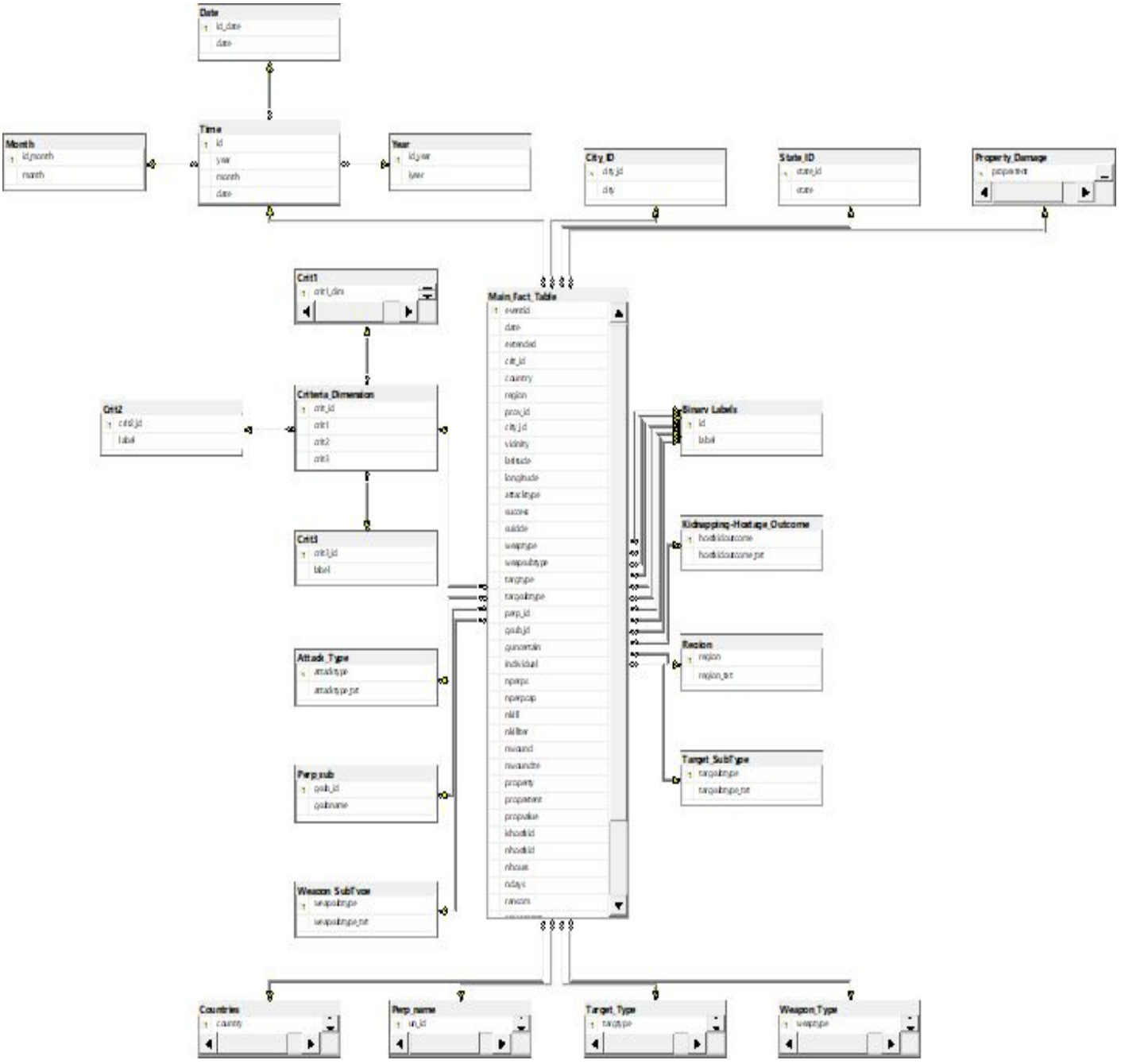
Missing Values:

```
update [Terrorist_DB].[dbo].[Main_Fact_Table] set [nkillter] = NULL where [nkillter] = 'NA'
update [Terrorist_DB].[dbo].[Main_Fact_Table] set [nwound] = NULL where [nwound] = 'NA'
update [Terrorist_DB].[dbo].[Main_Fact_Table] set [nwoundte] = NULL where [nwoundte] = 'NA'
update [Terrorist_DB].[dbo].[Main_Fact_Table] set [propvalue] = NULL where [propvalue] = 'NA'
update [Terrorist_DB].[dbo].[Main_Fact_Table] set [nhostkid] = NULL where [nhostkid] = 'NA'
update [Terrorist_DB].[dbo].[Main_Fact_Table] set [ishostkid] = NULL where [ishostkid] = 'NA'
update [Terrorist_DB].[dbo].[Main_Fact_Table] set [nhours] = NULL where [nhours] = 'NA'
update [Terrorist_DB].[dbo].[Main_Fact_Table] set [ndays] = NULL where [ndays] = 'NA'
update [Terrorist_DB].[dbo].[Main_Fact_Table] set [ransom] = NULL where [ransom] = 'NA'
update [Terrorist_DB].[dbo].[Main_Fact_Table] set [ransomamt] = NULL where [ransomamt] = 'NA'
update [Terrorist_DB].[dbo].[Main_Fact_Table] set [ransompaid] = NULL where [ransompaid] = 'NA'
update [Terrorist_DB].[dbo].[Main_Fact_Table] set [hostkidoutcome] = NULL where [hostkidoutcome] = 'NA'
update [Terrorist_DB].[dbo].[Main_Fact_Table] set [nreleased] = NULL where [nreleased] = 'NA'
update [Terrorist_DB].[dbo].[Main_Fact_Table] set [propextent] = 0 where [propextent] = 'NA'
```

```
update [Terrorist_DB].[dbo].[Main_Fact_Table]  set [nhostkid] = 0 where [nhostkid] = -99
update [Terrorist_DB].[dbo].[Main_Fact_Table]  set [nperpcap] = 0 where [nperpcap] = -99
update [Terrorist_DB].[dbo].[Main_Fact_Table]  set [nhours] = 0 where [nhours] = -99
update [Terrorist_DB].[dbo].[Main_Fact_Table]  set [ndays] = 0 where [ndays] = -99
update [Terrorist_DB].[dbo].[Main_Fact_Table] set [nperps] = 0 where [nperps] = -99
update [Terrorist_DB].[dbo].[Main_Fact_Table] set [ransomamt] = 0 where [ransomamt] = -99
update [Terrorist_DB].[dbo].[Main_Fact_Table] set [ransompaid] = 0 where [ransompaid] = -99
update [Terrorist_DB].[dbo].[Main_Fact_Table]  set [nreleased] = 0 where [nreleased] = -99
update [Terrorist_DB].[dbo].[Main_Fact_Table]  set [propvalue] = 0 where [propvalue] = -99

select[eventid]
```

The code that we wrote in order to handle missing values

And as we completed that we exported the following Snowflake Schema:


Date
id_date
date
Month
id_month
month
Time
id
year
month
date
Year
id_year
year
City_ID
city_id
city
State_ID
state_id
state
Property_Damage
propertext
Crit1
crit1_des
Main_Fact_Table
eventid
date
extended
crit_id
country
region
provi_id
city_id
vicinity
latitude
longitude
attacktype
success
suicide
weapntype
weapsubtype
targtype
targsubtype
perp_id
gsub_id
guncertain
individual
nperps
nperpcap
nkill
nkillter
nwound
nwoundte
property
propextent
propvalue
ishostkid
nhostkid
nhours
ndays
ransom
Crit2
crit2_id
label
Criteria_Dimension
crit_id
crit1
crit2
crit3
Crit3
crit3_id
label
Binary Labels
id
label
Kidnapping-Hostage_Outcome
hostkidoutcome
hostkidoutcome_txt
Region
region
region_txt
Target_SubType
targsubtype
targsubtype_txt
Attack_Type
attacktype
attacktype_txt
Perpsub
gsub_id
gsubname
Weapon SubType
weapsubtype
weapsubtype_txt
Countries
country
Perp_name
un_id
Target_Type
targtype
Weapon_Type
weapntype

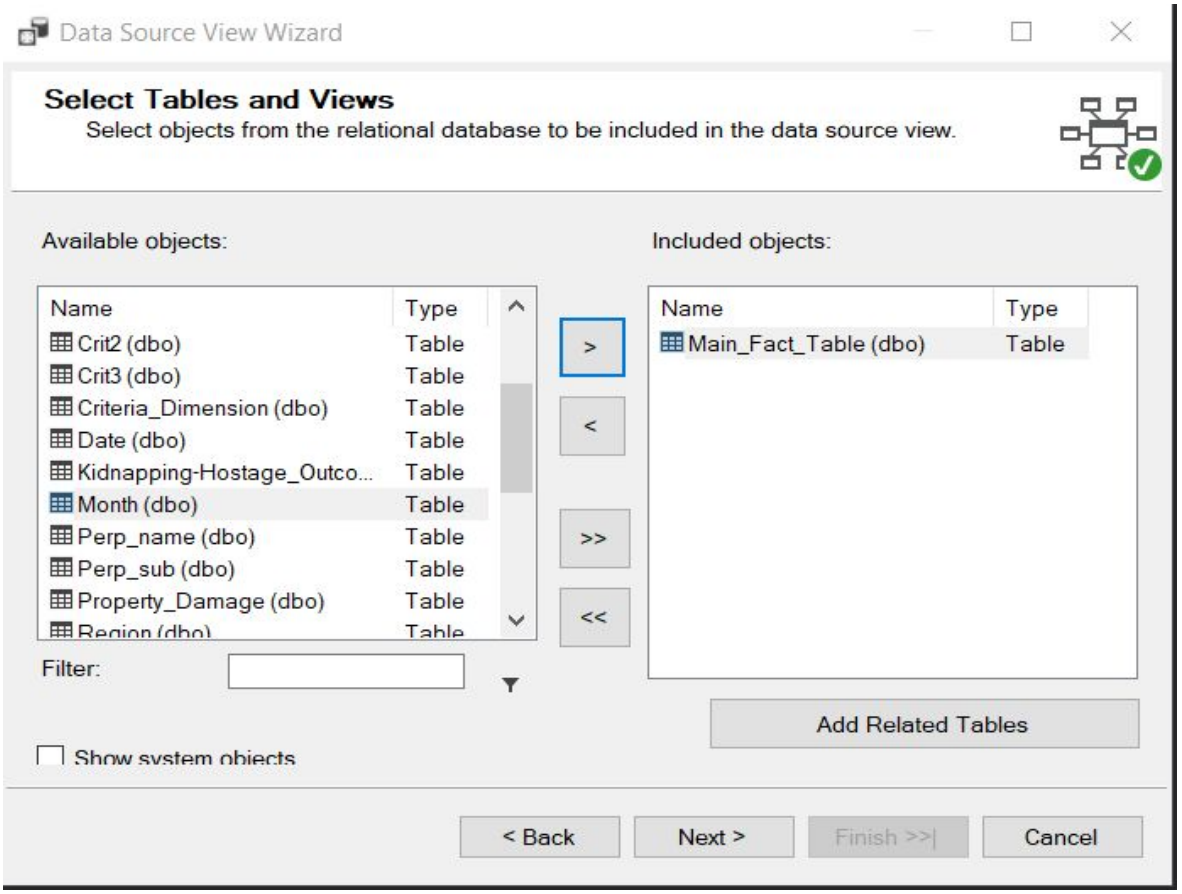
Define a multi-dimensional model over the schema and show some OLAP reports

For the multi-dimensional model, we used the SQL Server Analysis Services (SSAS) and especially the Visual Studio (SSDT) application.

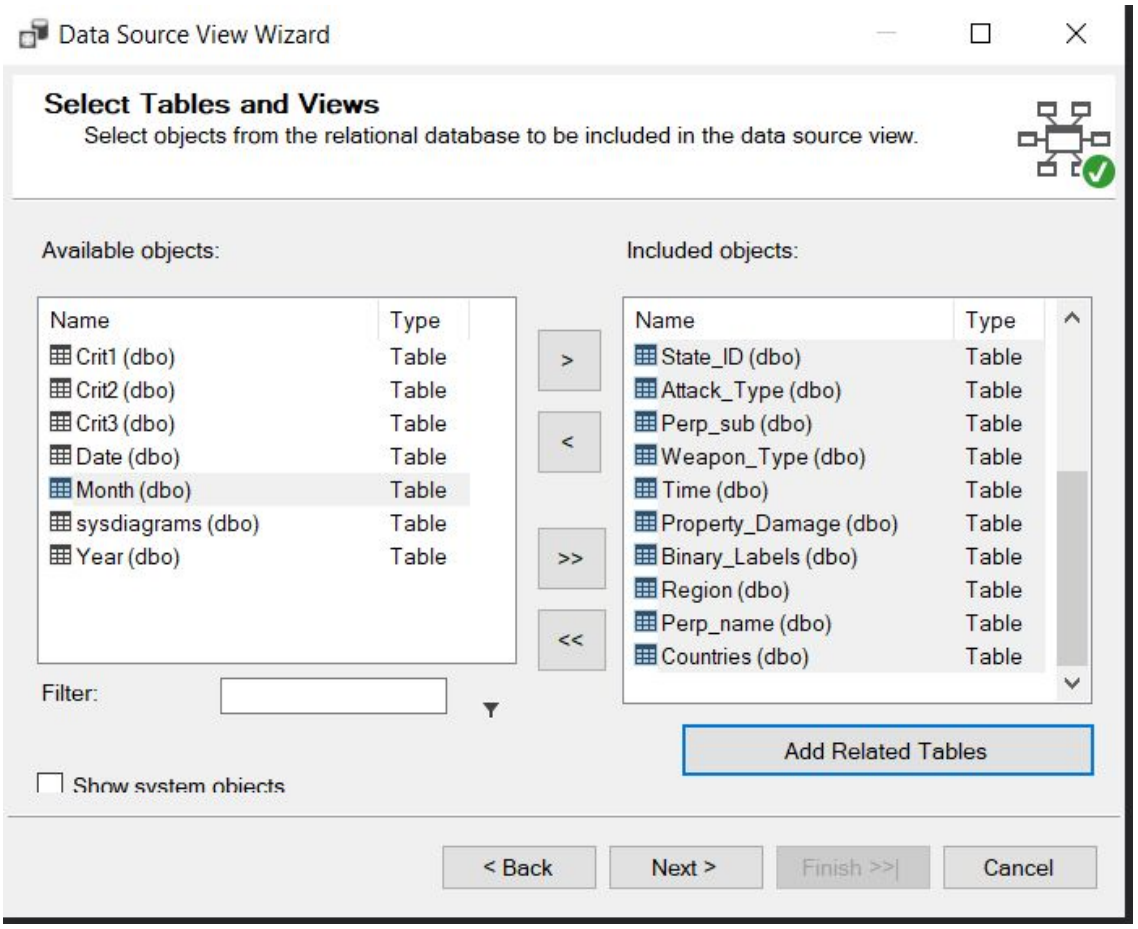
Designing Cubes

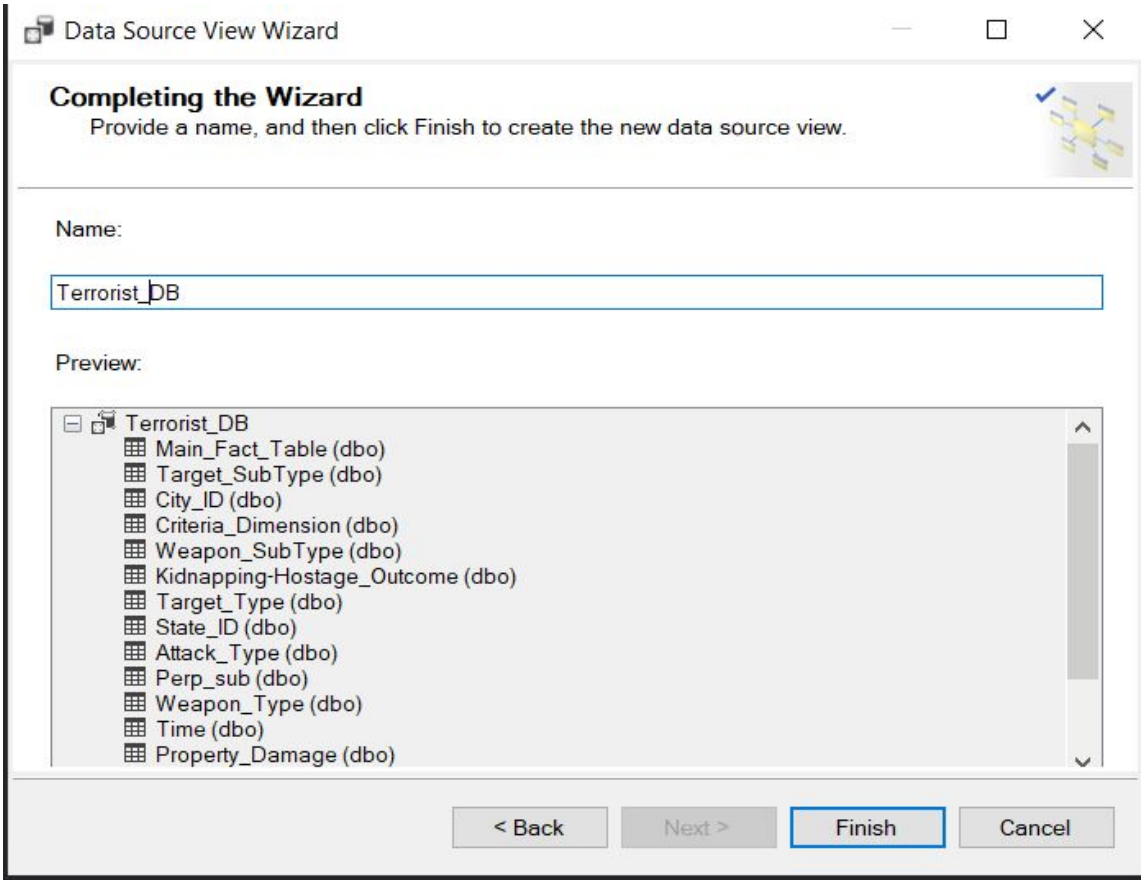
For the purpose of this project, we created two small and simple cubes and one bigger with multiple dimensions and calculated measures. The process we followed is the below:

After the creation of the project we built a new data source based upon the database in the SQL Server Management Studio (SSMS):

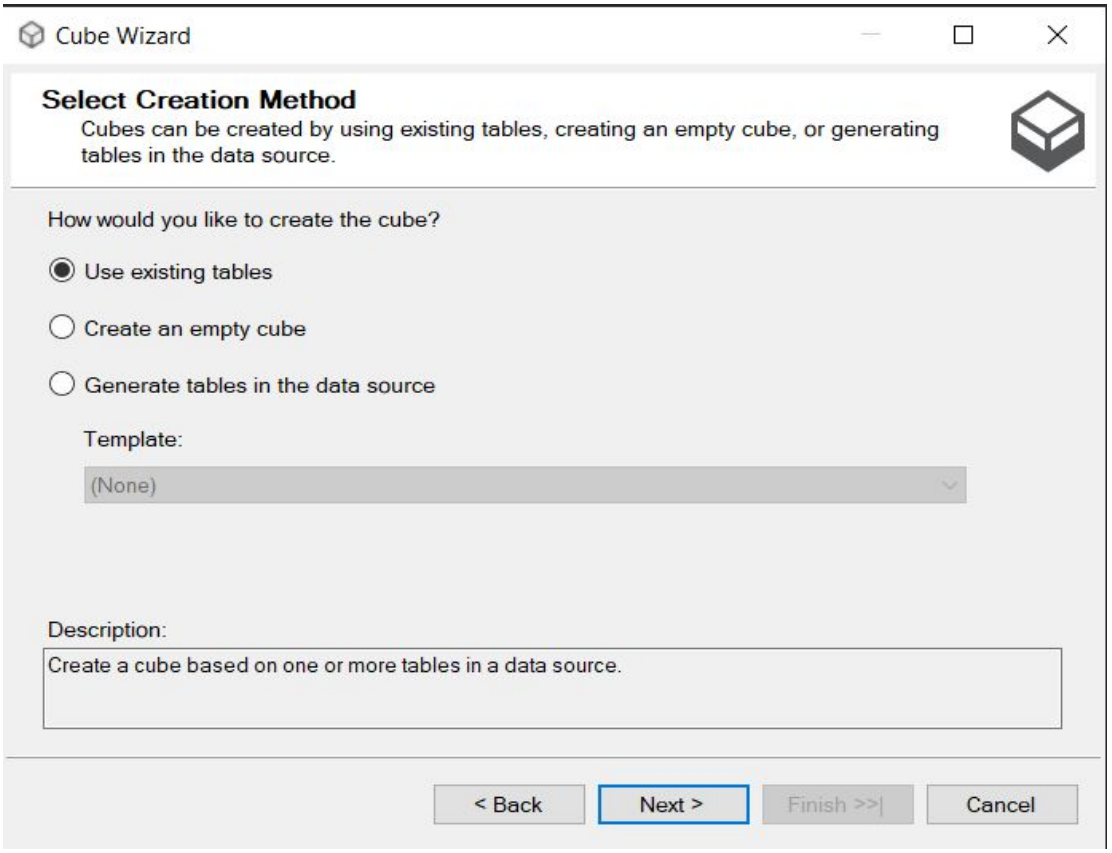


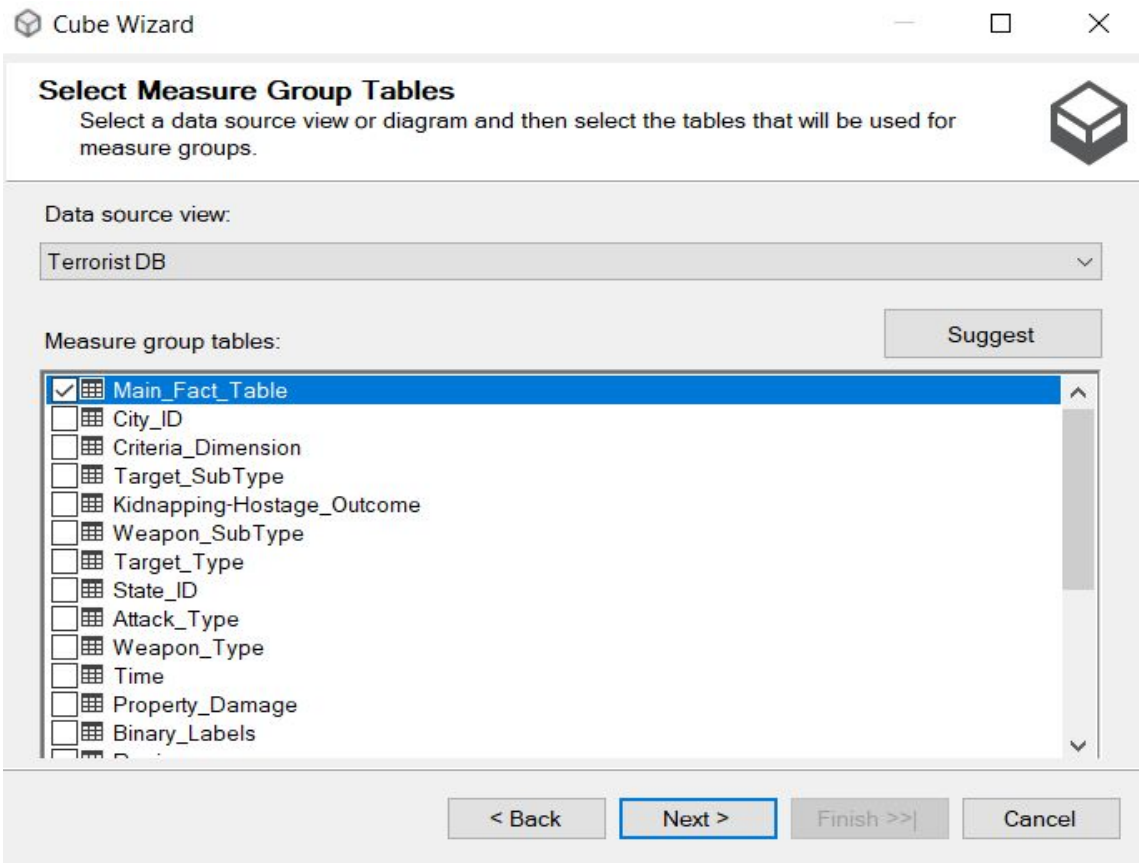
We added all the Related Tables with our main Fact Table (based on the foreign keys):

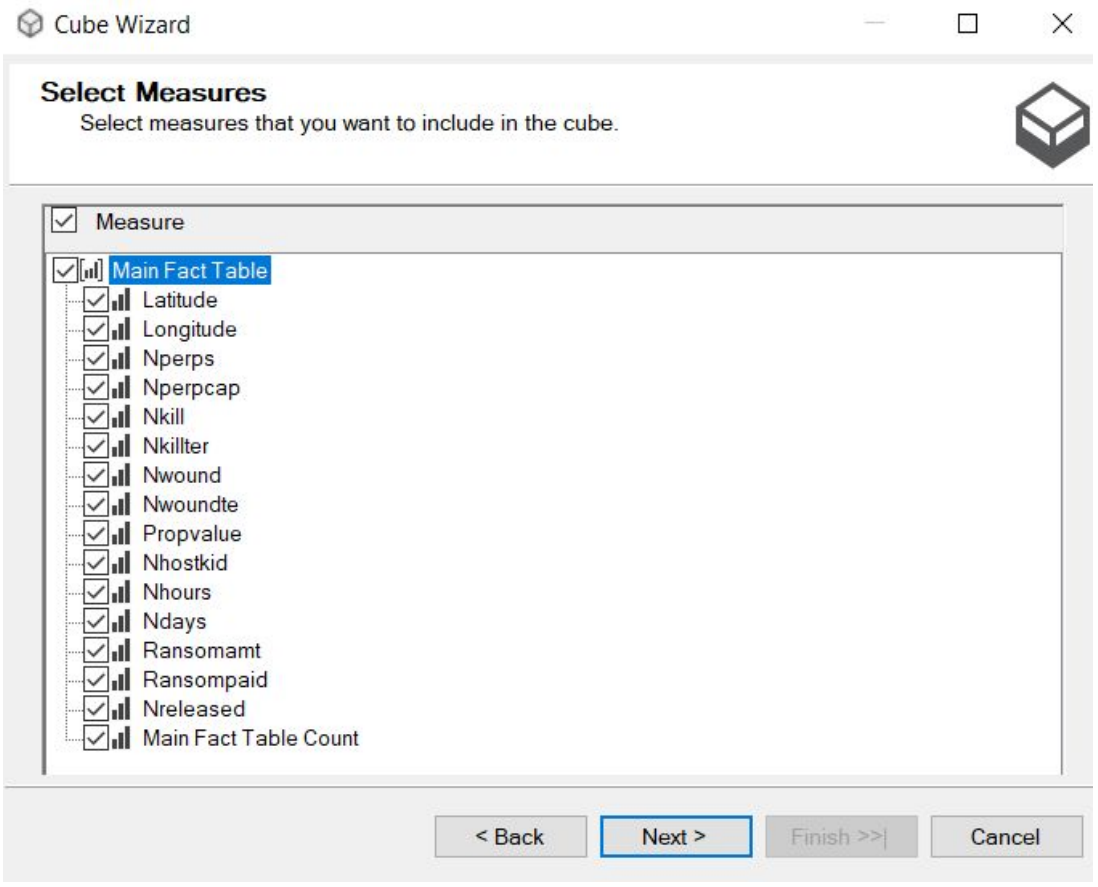

Data Source View Wizard
Select Tables and Views
Select objects from the relational database to be included in the data source view.
Available objects:
Name
Type
Crit1 (dbo)
Table
Crit2 (dbo)
Table
Crit3 (dbo)
Table
Date (dbo)
Table
Month (dbo)
Table
sysdiagrams (dbo)
Table
Year (dbo)
Table
>
<
>>
<<
Included objects:
Name
Type
State_ID (dbo)
Table
Attack_Type (dbo)
Table
Perp_sub (dbo)
Table
Weapon_Type (dbo)
Table
Time (dbo)
Table
Property_Damage (dbo)
Table
Binary_Labels (dbo)
Table
Region (dbo)
Table
Perp_name (dbo)
Table
Countries (dbo)
Table
Filter:
Add Related Tables
Show system objects
< Back
Next >
Finish >>|
Cancel


Data Source View Wizard
Completing the Wizard
Provide a name, and then click Finish to create the new data source view.
Name:
Terrorist_DB
Preview:
Terrorist_DB
Main_Fact_Table (dbo)
Target_SubType (dbo)
City_ID (dbo)
Criteria_Dimension (dbo)
Weapon_SubType (dbo)
Kidnapping-Hostage_Outcome (dbo)
Target_Type (dbo)
State_ID (dbo)
Attack_Type (dbo)
Perp_sub (dbo)
Weapon_Type (dbo)
Time (dbo)
Property_Damage (dbo)
< Back
Next >
Finish
Cancel

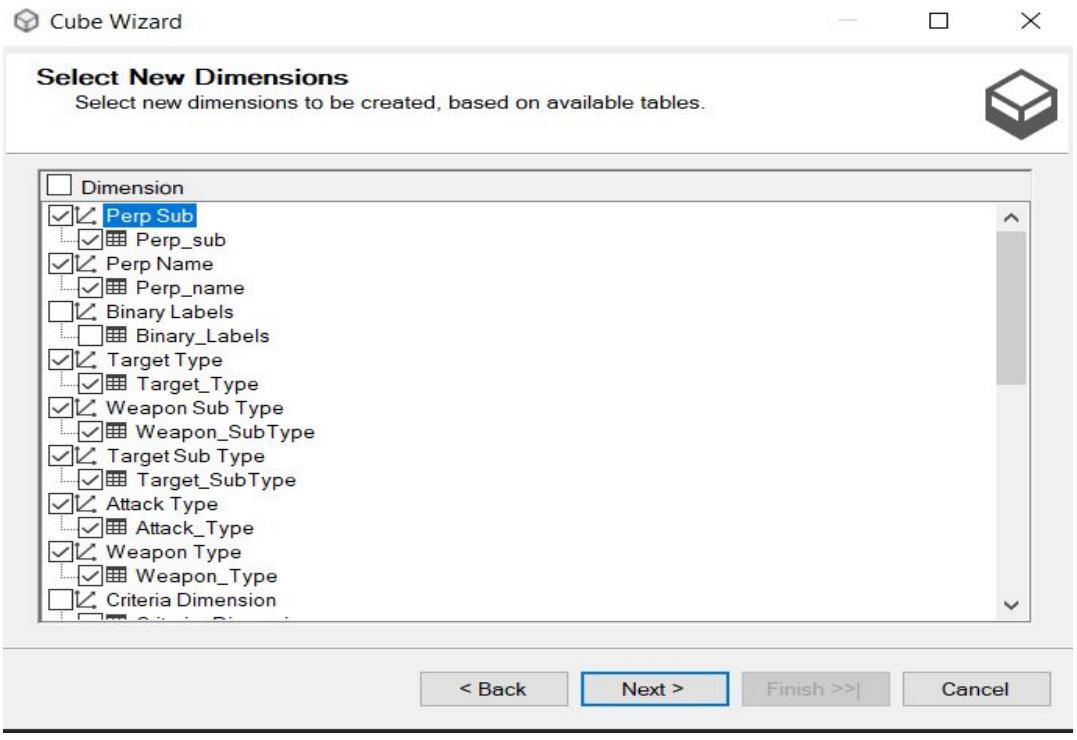
After that we created the cubes with the below wizard:

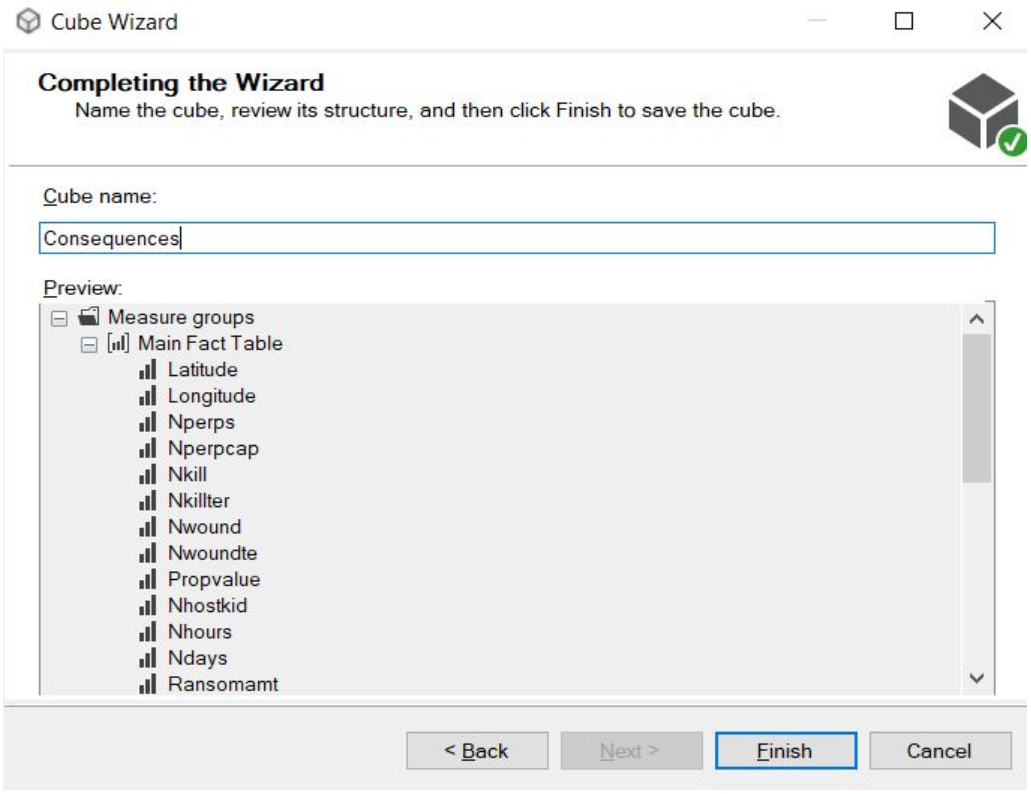




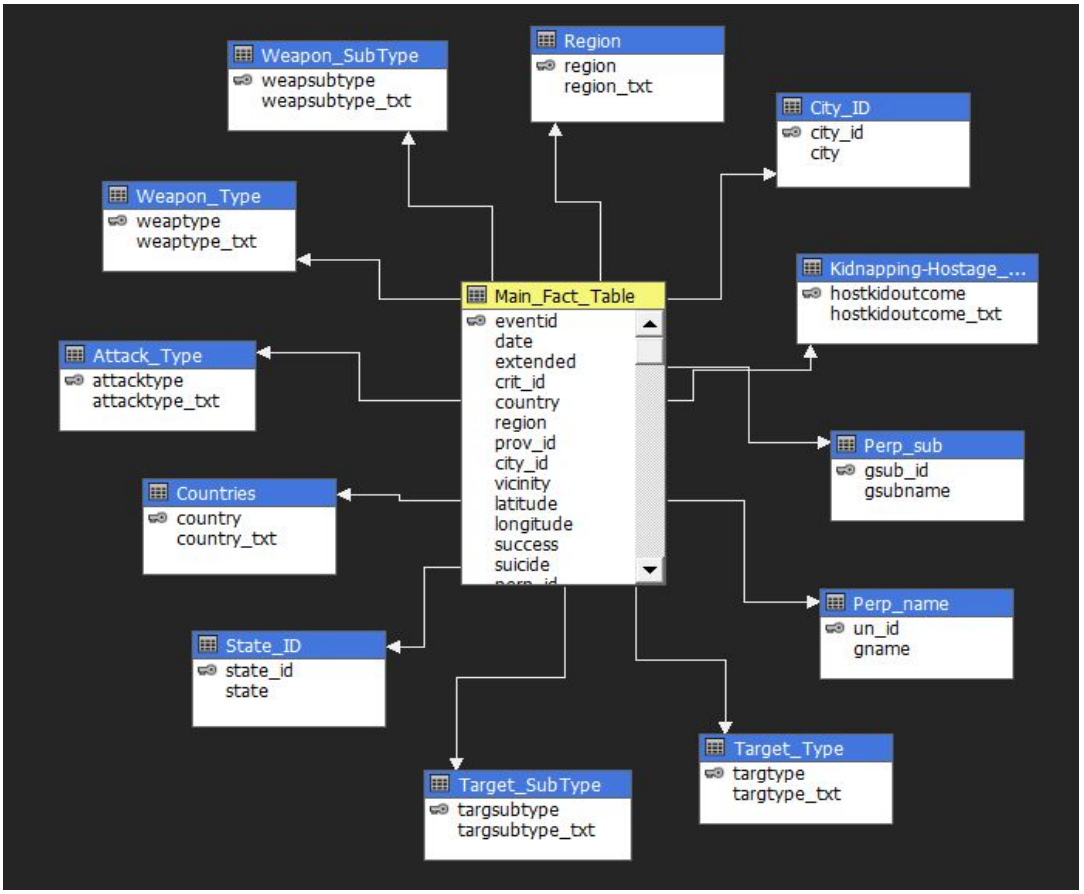


We selected the measures we wanted (in our case all the measures that exist in the fact table - 15 measures) and then we picked the dimensions that we wanted the cube to have:

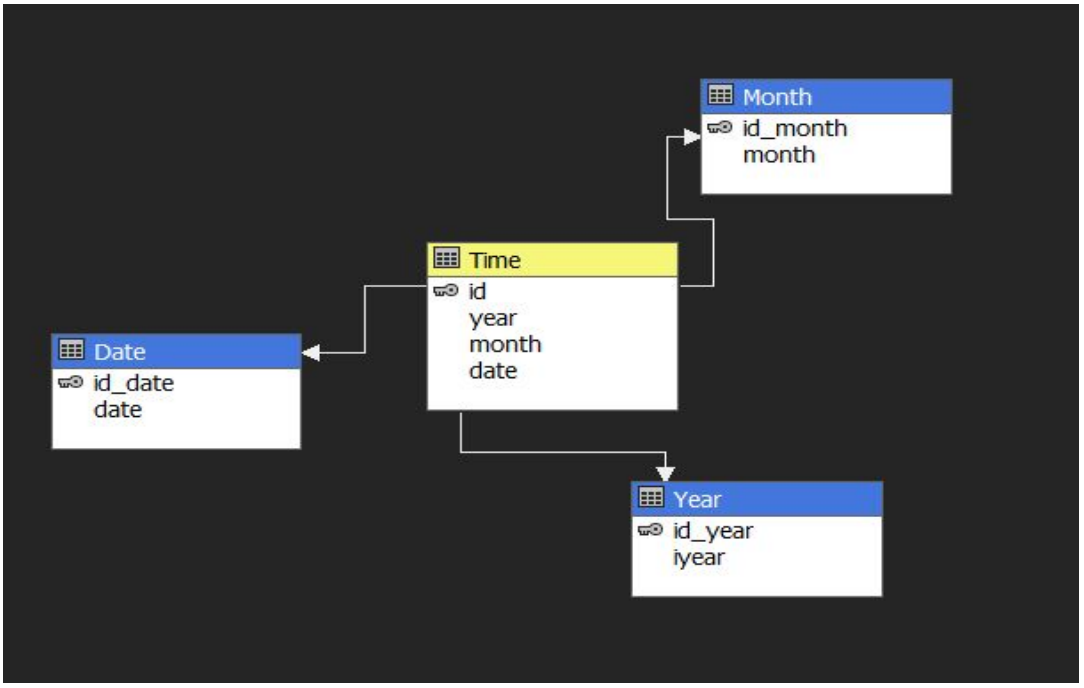


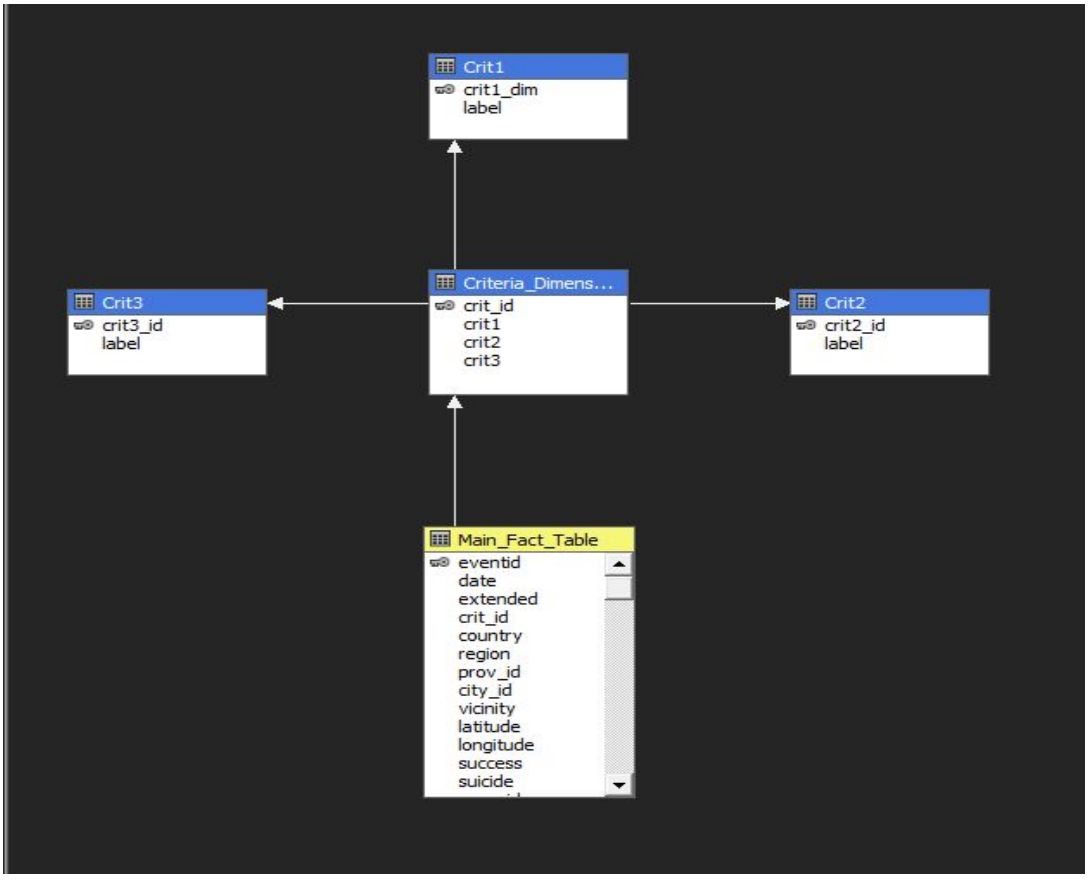


We named the cube as "Consequences" and then after deploying it we had the below diagram with 12 dimensions:



We followed the same procedure and with the other two simple cubes:

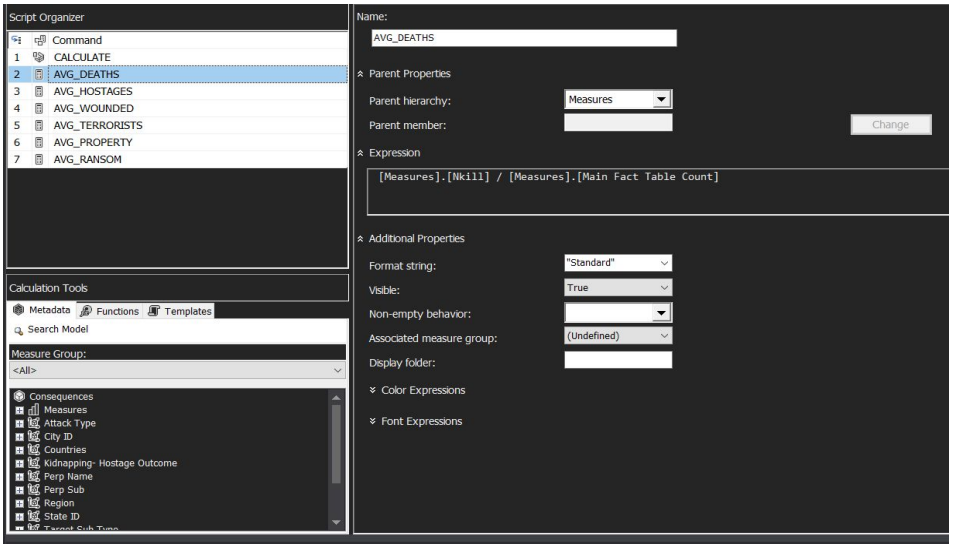




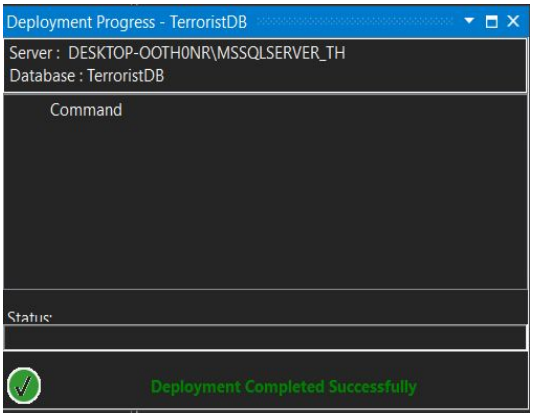
We are mainly interested in editing and presenting the "Consequences" Cube, that is why the following "Print Screen" images will refer to this. It is

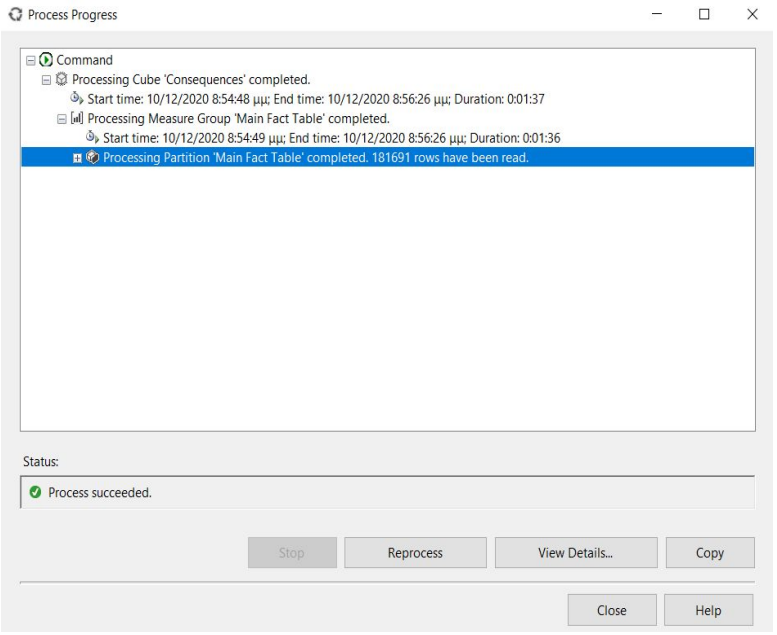
worth mentioning that the Time Cube has three dimensions (Date, Month, Year) and one measure (the one that is default by the system, the count). On the contrary, the Criteria Cube has one dimension (Criteria) and three sub-dimensions (Crit1, Crit2 and Crit3). The measures of the latter cube are all the numeric variables from the fact table plus the default count measure.

As far as the bigger cube is concerned, we created some new measures that calculate the average deaths, wounded, hostages, property damage (in US Dollars), ransom that is being paid, and the number of terrorists per terrorist attack. For example,

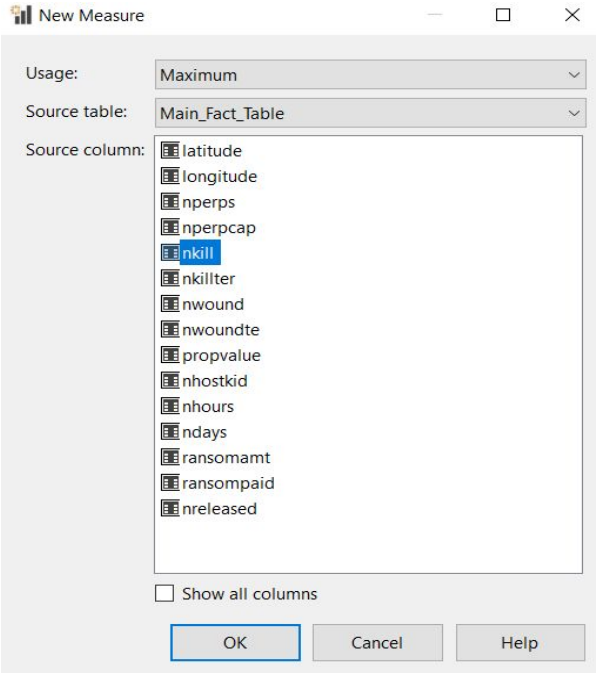


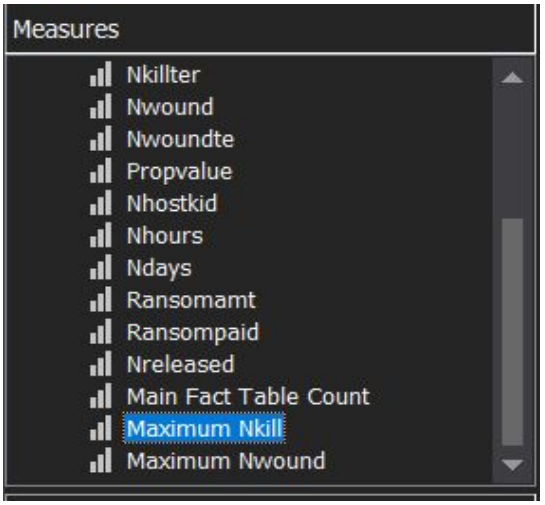
We deployed the cube in every step we took





and finally we added two more measures to use them in future presentation (maximum deaths and wounded):





Browsing the cubes & OLAP Reports

After the above procedure, we exported the cubes from SQL Server Analysis Services (SSAS) to excel in order to draw some OLAP (Online Analytical Processing) Reports with the help of filters that the latter application provides.

Drill down (Location): We go from the higher level (regions)to the lower level (countries of Western Europe).

Ετικέτες γραμμής	Nkill	AVG_DEATHS	Nwound	AVG_WOUNDED	Terrorist Attacks
Australasia & Oceania	530	1,88	1020	3,62	282
Central America & Caribbean	22686	2,19	27415	2,65	10344
Central Asia	1505	2,67	1426	2,53	563
East Asia	1690	2,11	1882	2,35	802
Eastern Europe	12344	2,40	16046	3,12	5144
Middle East & North Africa	115338	2,29	145666	2,89	50474
North America	6796	1,97	9262	2,68	3456
South America	41770	2,20	52890	2,79	18978
South Asia	102660	2,28	131670	2,93	44974
Southeast Asia	28244	2,26	34903	2,80	12485
Sub-Saharan Africa	41337	2,36	50856	2,90	17550
Western Europe	36968	2,22	50986	3,06	16639
Γενικό Άθροισμα	**411868**	**2,27**	**524022**	**2,88**	**181691**

Ετικέτες γραμμής	Nkill	AVG_DEATHS	Nwound	AVG_WOUNDED	Terrorist Attacks
Western Europe					
Andorra	0	0,00	0	0,00	1
Austria	239	2,08	323	2,81	115
Belgium	571	3,71	632	4,10	154
Cyprus	293	2,22	233	1,77	132
Denmark	133	3,24	190	4,63	41
Finland	17	0,85	14	0,70	20
France	5613	2,08	6455	2,40	2693
Germany	1678	2,28	1906	2,59	735
Greece	2809	2,20	3887	3,05	1275
Iceland	45	11,25	0	0,00	4
Ireland	515	1,68	605	1,97	307
Italy	3025	1,93	4524	2,89	1565
Luxembourg	119	7,44	26	1,63	16
Malta	16	0,70	15	0,65	23
Netherlands	196	1,51	127	0,98	130
Norway	29	1,53	25	1,32	19
Portugal	253	1,81	428	3,06	140
Spain	7862	2,42	10223	3,15	3249
Sweden	305	2,31	511	3,87	132
Switzerland	202	1,82	304	2,74	111
United Kingdom	12004	2,29	13618	2,60	5235
Vatican City	1	1,00	2	2,00	1
West Germany (FRG)	1043	1,93	6938	12,82	541
Γενικό Άθροισμα	**36968**	**2,22**	**50986**	**3,06**	**16639**

and then we can do slice and dice [select two countries (Cyprus and Greece) and project them]

Ετικέτες γραμμής	AVG_DEATHS	Nkill	AVG_WOUNDED	Nwound	Main Fact Table Count
Cyprus	2,22	293	1,77	233	132
Greece	2,20	2809	3,05	3887	1275
Γενικό Άθροισμα	**2,20**	**3102**	**2,93**	**4120**	**1407**

Drill down (Weapon Type): We go from the higher level (weapon type)to the lower level (weapon sub-type).

Ετικέτες γραμμής	AVG_DEATHS	Nkill	AVG_WOUNDED	Nwound	Main Fact Table Count
Biological	2,26	79	3,91	137	35
Chemical	1,56	501	3,20	1026	321
Explosives	2,27	209899	2,87	265370	92426
Fake Weapons	1,67	55	2,61	86	33
Firearms	2,27	132944	2,86	167254	58524
Incendiary	2,23	24827	2,73	30374	11135
Melee	2,22	8097	2,94	10738	3655
Other	1,46	166	2,00	228	114
Radiological	2,36	33	5,00	70	14
Sabotage Equipment	2,51	354	3,49	492	141
Unknown	2,29	34675	3,16	47858	15157
Vehicle (not to include vehicle-borne explosives, i.e., car or truck bombs)	1,75	238	2,86	389	136
Γενικό Άθροισμα	**2,27**	**411868**	**2,88**	**524022**	**181691**

Ετικέτες γραμμής	AVG_DEATHS	Nkill	AVG_WOUNDED	Nwound	Main Fact Table Count
Biological					
	2,26	79	3,91	137	35
Chemical					
	1,39	241	1,73	300	173
Explosive	3,11	202	9,02	586	65
Poisoning	0,70	58	1,69	140	83
Explosives					
	2,25	27	2,75	33	12
Dynamite/TNT	2,06	2521	2,43	2971	1222
Grenade	2,49	12875	2,94	15169	5167
Landmine	2,11	8966	2,60	11064	4251
Letter Bomb	2,53	1388	3,55	1946	548
Other Explosive Type	2,45	17868	3,06	22324	7304
Pipe Bomb	2,10	1311	2,47	1541	625
Pressure Trigger	2,00	438	2,08	456	219
Projectile (rockets, mortars, RPGs, etc.)	2,23	21940	2,67	26330	9848
Remote Trigger	2,50	6786	3,34	9068	2719
Sticky Bomb	2,02	3214	2,54	4054	1594
Suicide (carried bodily by human being)	2,19	7095	2,88	9353	3245
Time Fuse	2,06	1630	1,81	1432	792
Unknown Explosive Type	2,25	101395	2,85	128287	44980
Unknown Gun Type	0,00	0	6,00	6	1
Vehicle	2,27	22445	3,17	31336	9899

Roll Up: from total terrorist attacks per month we are moving up to total terrorist attacks per year

Ετικέτες γραμμής	Time Count
"April"	1354
"August"	1386
"December"	1355
"February"	1256
"January"	1391
"July"	1383
"June"	1339
"March"	1404
"May"	1394
"November"	1350
"October"	1397
"September"	1339
"Unknown"	11
Γενικό Άθροισμα	**16359**

Ετικέτες γραμμής	Time Count
"1970"	
"April"	26
"August"	25
"December"	16
"February"	22
"January"	23
"July"	26
"June"	26
"March"	26
"May"	24
"November"	19
"October"	25
"September"	21
"Unknown"	1
"1971"	
"April"	24
"August"	16
"December"	15
"February"	22
"January"	22
"July"	18
"June"	17
"March"	26
"May"	23
"November"	20
"October"	21
"September"	20

Slice and Dice: From all the countries we filter and we select these we are interested in.

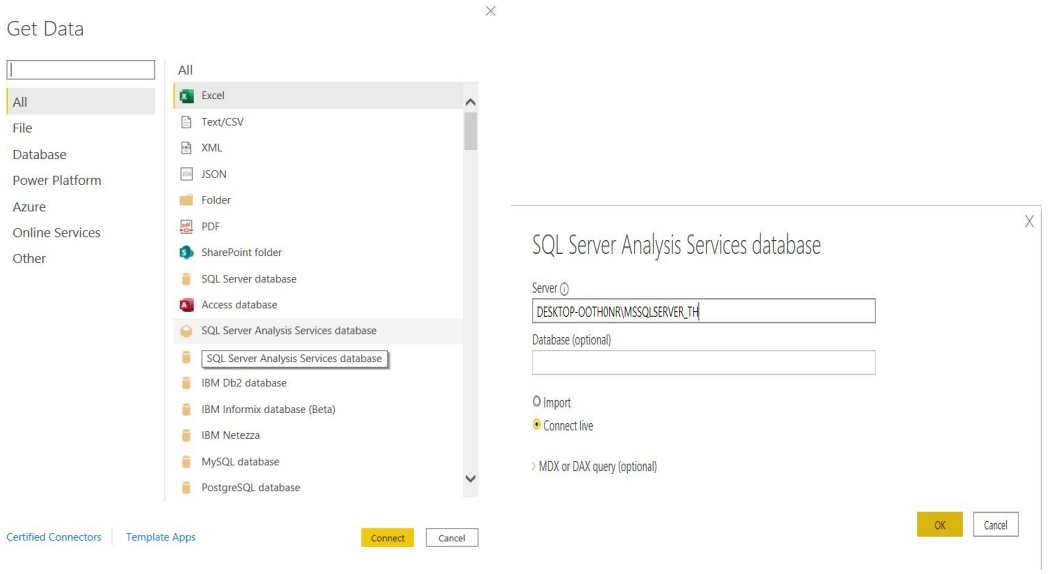
Ετικέτες γραμμής	Nkill	Maximum Nkill	AVG_DEATHS	Nwound	Maximum Nwound	AVG_WOUNDED	Propvalue	AVG_PROPERTY	Terrorist Attacks
Afghanistan	30257	1383	2,38	42656	8191	3,35	257542852	20229,59	12731
Albania	171	25	2,14	156	50	1,95	46474	580,93	80
Algeria	6726	304	2,45	7692	491	2,80	15675231	5714,63	2743
Andorra	0	0	0,00	0	0	0,00	2000	2000,00	1
Angola	1334	52	2,67	1422	100	2,85	839500	1682,36	499
Antigua and Barbuda	5	5	5,00	3	3	3,00	0	0,00	1
Argentina	1605	78	1,97	2379	400	2,92	4824604	5919,76	815
Armenia	32	9	1,33	84	30	3,50	1500	62,50	24
Australia	299	53	2,62	715	300	6,27	8245016	72324,70	114
Austria	239	26	2,08	323	53	2,81	2463000	21417,39	115
Azerbaijan	105	37	2,14	95	20	1,94	50000	1020,41	49
Bahamas	19	17	3,80	0	0	0,00	0	0,00	5
Bahrain	317	30	1,53	291	22	1,41	3066083	14812,00	207
Bangladesh	3799	146	2,31	4176	220	2,53	52983881	32150,41	1648
Barbados	5	2	1,67	10	10	3,33	0	0,00	3
Belarus	70	30	5,38	71	60	5,46	30000	2307,69	13
Belgium	571	240	3,71	632	300	4,10	378357	2456,86	154
Belize	3	2	0,38	4	4	0,50	0	0,00	8
Benin	18	14	2,25	11	7	1,38	0	0,00	8
Bhutan	33	24	5,50	2	2	0,33	0	0,00	6
Bolivia	475	27	1,51	1074	250	3,42	4170740103	13282611,79	314
Bosnia-Herzegovina	367	62	2,31	374	66	2,35	210000	1320,75	159
Botswana	10	3	1,00	7	4	0,70	0	0,00	10
Brazil	457	40	1,67	442	35	1,62	902500	3305,86	273
Brunei	26	18	4,33	10	9	1,67	130000	21666,67	6
Bulgaria	41	5	0,79	99	15	1,90	538450	10354,81	52
Burkina Faso	136	42	2,62	145	43	2,79	0	0,00	52
Cyprus	293	51	2,22	233	65	1,77	413000	3128,79	132
Greece	2809	200	2,20	3887	200	3,05	7344252	5760,20	1275
Γενικό Άθροισμα	**3102**	**200**	**2,20**	**4120**	**200**	**2,93**	**7757252**	**5513,33**	**1407**

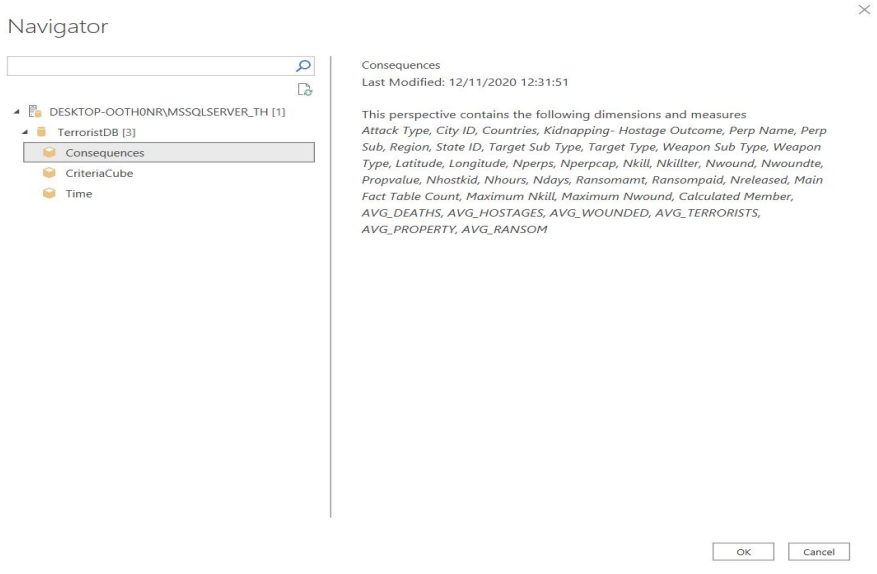
Pivot: reorient the cube to present one of its side (first we put as rows the years and as columns the months and later we transform the table to show the other way around, years as columns and months as rows)

Time Count	Ετικέτες στήλης												
Ετικέτες γραμμής	"April"	"August"	"December"	"February"	"January"	"July"	"June"	"March"	"May"	"November"	"October"	"September"	"Unknown"	Γενικό Άθροισμα
"1970"	26	25	16	22	23	26	26	26	24	19	25	21	1	280
"1971"	24	16	15	22	22	18	17	26	23	20	21	20		244
"1972"	14	24	26	14	17	27	19	22	26	20	17	20	1	247
"1973"	25	19	22	20	25	18	17	22	23	21	20	19	1	252
"1974"	24	19	22	18	23	18	21	28	24	23	19	20	1	260
"1975"	23	27	22	25	18	22	21	23	27	24	28	22	1	283
"1976"	27	27	20	25	29	28	28	24	29	28	27	26	1	319
"1977"	29	27	30	25	27	30	29	30	27	26	31	26	1	338
"1978"	26	31	31	29	29	26	30	30	31	30	30	30	1	354
"1979"	31	32	32	28	32	31	31	32	32	31	32	31	1	376
"1980"	31	32	30	30	31	31	31	30	31	31	32	29		369
"1981"	30	31	32	29	32	31	31	31	31	31	31	31		371
"1982"	31	32	32	27	32	31	30	32	32	28	31	31	1	370
"1983"	31	31	30	29	32	32	30	32	28	31	32	31		369
"1984"	31	32	31	30	32	31	29	32	30	31	32	31		372
"1985"	29	30	32	28	30	30	30	32	32	30	31	30		364
"1986"	31	31	31	27	31	31	30	32	30	29	32	29	1	365
"1987"	30	32	32	28	31	31	30	32	30	30	32	30		368
"1988"	30	31	31	29	31	31	31	31	31	30	31	30		367
"1989"	31	31	32	28	32	31	30	31	31	30	31	30		368
"1990"	31	31	30	28	32	32	30	32	32	31	32	31		372
"1991"	31	32	32	29	32	32	30	31	32	30	31	31		373
"1992"	31	32	32	30	32	31	30	31	32	30	31	31		373
"1994"	31	31	31	28	32	31	30	31	32	30	31	30		368
"1995"	29	31	31	28	32	30	31	31	31	30	31	30		365

Time Count	Ετικέτες στήλης																
Ετικέτες γραμμής	"1970"	"1971"	"1972"	"1973"	"1974"	"1975"	"1976"	"1977"	"1978"	"1979"	"1980"	"1981"	"1982"	"1983"	"1984"	"1985"	"1986"
"April"	26	24	14	25	24	23	27	29	26	31	31	30	31	31	31	29	31
"August"	25	16	24	19	19	27	27	27	31	32	32	31	32	31	32	30	31
"December"	16	15	26	22	22	22	20	30	31	32	30	32	32	30	31	32	31
"February"	22	22	14	20	18	25	25	25	29	28	30	29	27	29	30	28	27
"January"	23	22	17	25	23	18	29	27	29	32	31	32	32	32	32	30	31
"July"	26	18	27	18	18	22	28	30	26	31	31	31	31	32	31	30	31
"June"	26	17	19	17	21	21	28	29	30	31	31	31	30	30	29	30	30
"March"	26	26	22	22	28	23	24	30	30	32	30	31	32	32	32	32	32
"May"	24	23	26	23	24	27	29	27	31	32	31	31	32	28	30	32	30
"November"	19	20	20	21	23	24	28	26	30	31	31	31	28	31	31	30	29
"October"	25	21	17	20	19	28	27	31	30	32	32	31	31	32	32	31	32
"September"	21	20	20	19	20	22	26	26	30	31	29	31	31	31	31	30	29
"Unknown"	1		1	1	1	1	1	1	1	1			1				1
Γενικό Άθροισμα	**280**	**244**	**247**	**252**	**260**	**283**	**319**	**338**	**354**	**376**	**369**	**371**	**370**	**369**	**372**	**364**	**365**

After the above OLAP Reports we proceed to the connection of the cubes from the SSAS with the Power BI Desktop:



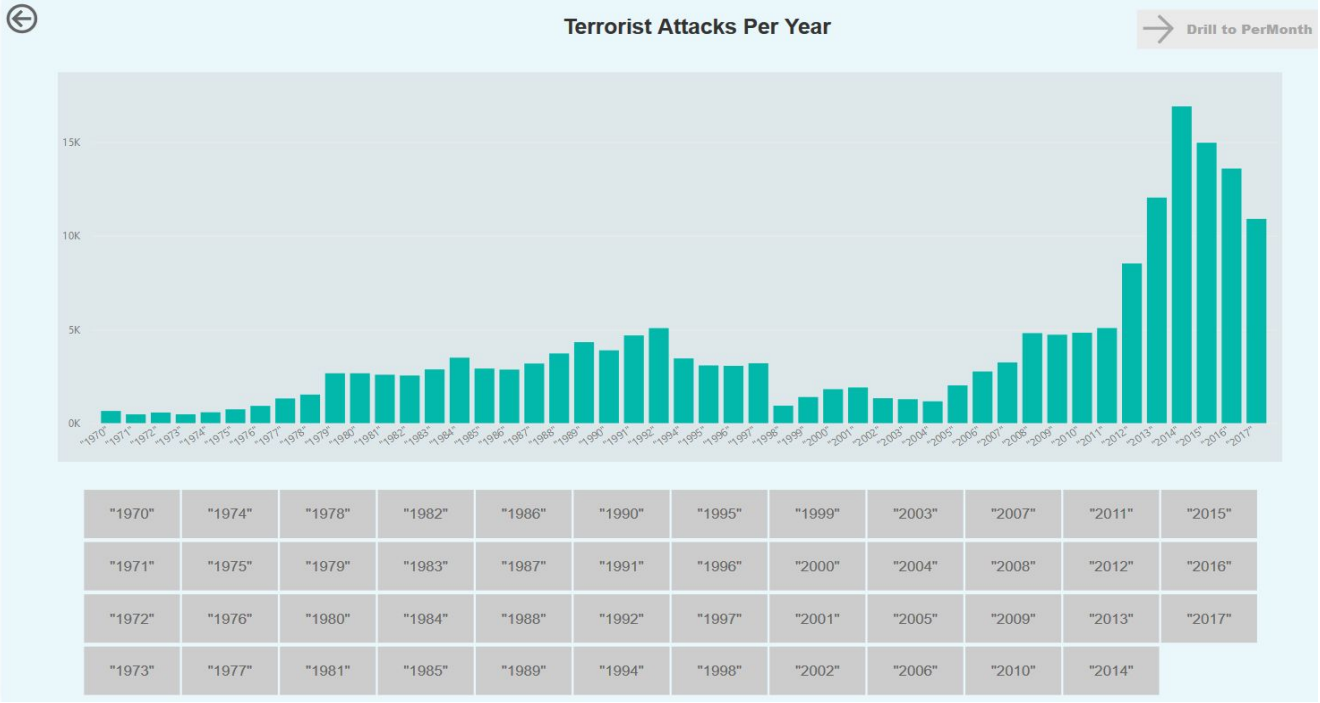


Power BI - visualizations of these reports and description of the visualization

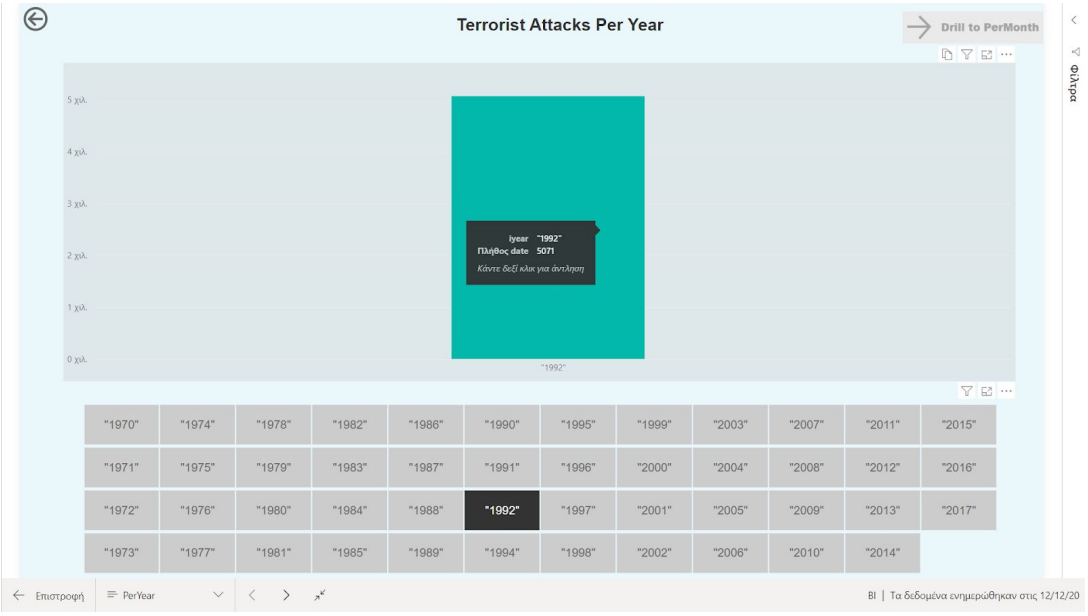
The main aim of this project is to answer the initial questions, analyzing the data.

To do so, we worked with the Microsoft Power BI, in which we connected the cubes through SSAS and we visualized some reports. All the reports have data and pre-calculated measures that have been taken from the cubes.

In the first tab, we see information according to the years. So, going back to the question: "*Are terrorist attacks increasing or decreasing nowadays?*" We can answer that we observe an increasing trend in the Terrorist Attacks in recent years.

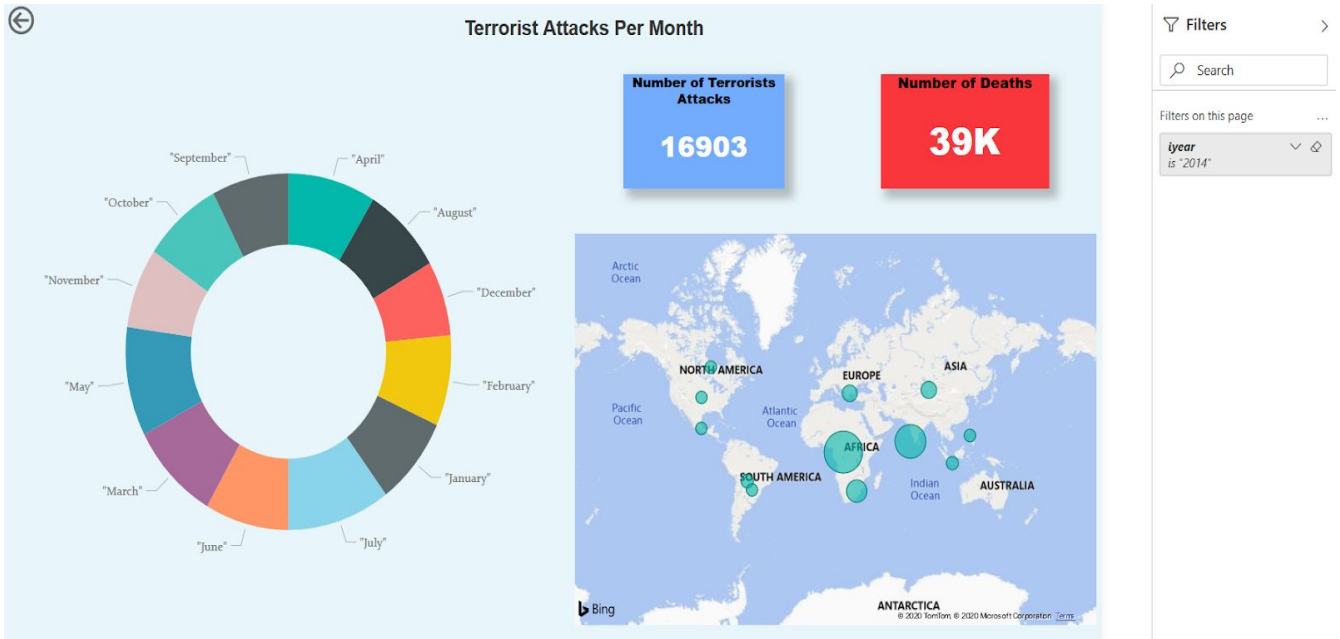


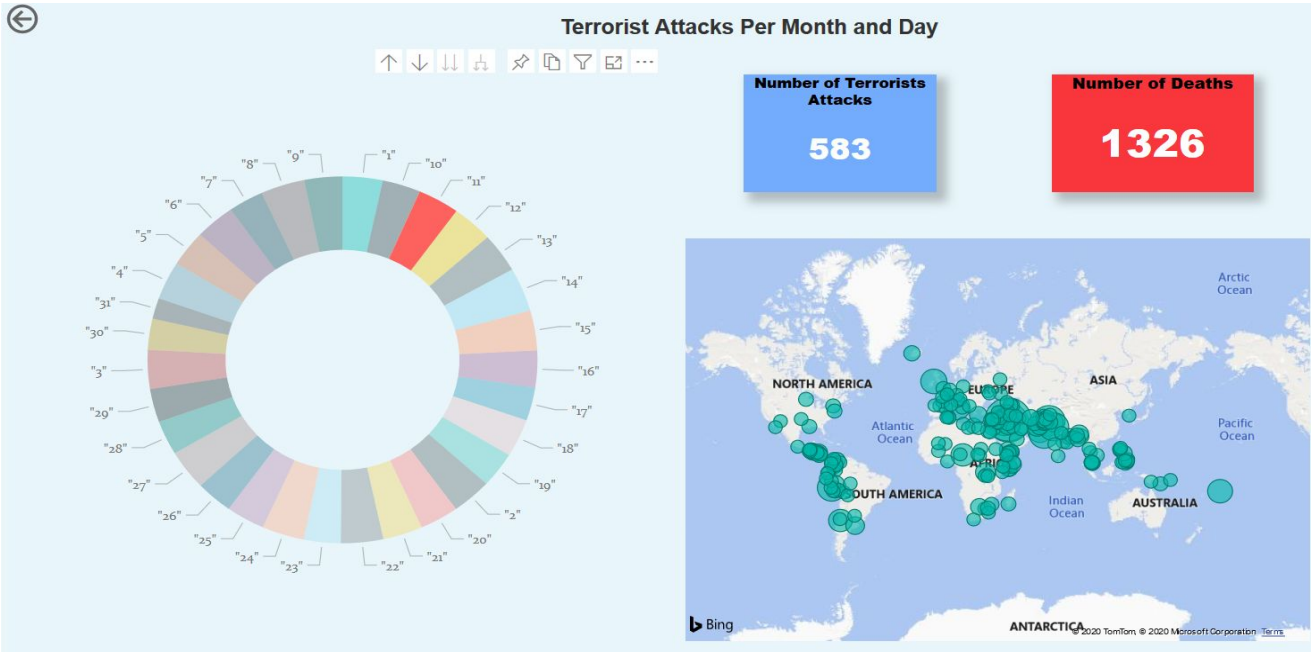
On the bottom of this page we have a filter of the Years, so we can easily choose the Year we are interested in.



Drill Down Report: We can go from year to month, selecting the Year, and then clicking on the "Drill to PerMonth" button.

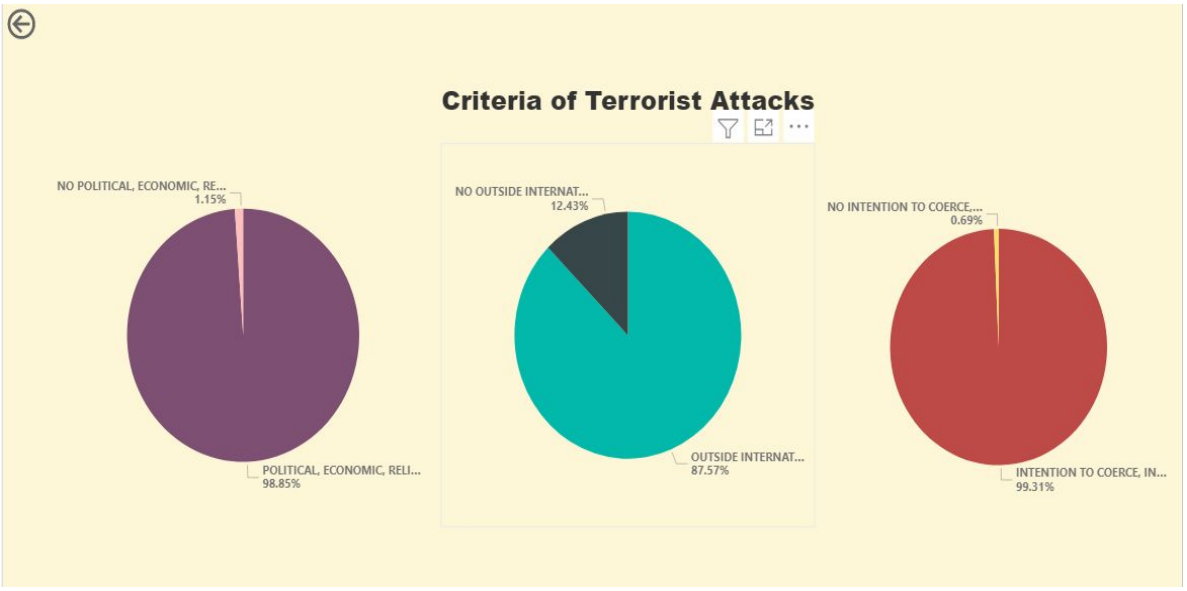
On this page, we have a donut, in which we can observe the months with the most terrorist attacks, an overview of the total number of attacks as well as the number of deaths and the geographic location (region). From this graph we can also drill down further to the days and from the region to the state.





In the above graph, we see all the attacks that we have in our data set and we can understand that there is no significant difference in the months that terrorist attacks occurred. We can assume that they are approximately uniformly distributed across both the months and the days over the years.

The next step is to analyze the criteria.



We can conclude that the aims of the attacks were mainly 2, either political, economic, religious, or social or outside international humanitarian law and

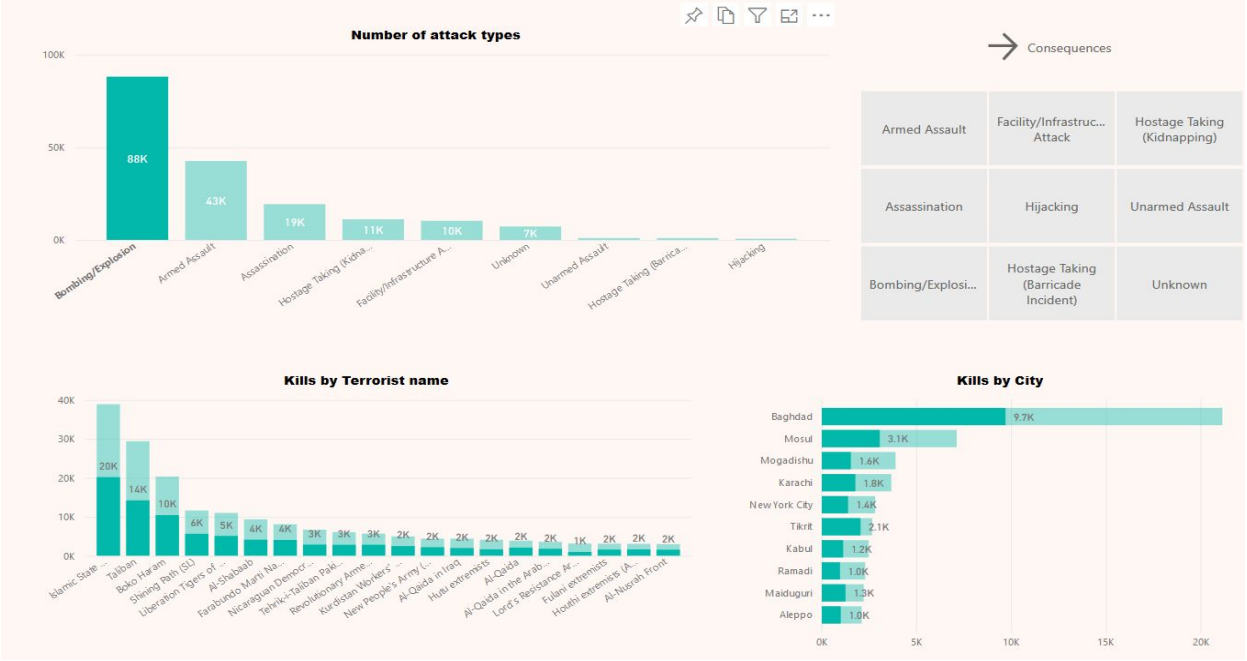
less intention to coerce, intimidate, or convey some other message to a larger audience.

Also, we are interested in analyzing the attack types. As we have mentioned before, we have the following Attack Types:

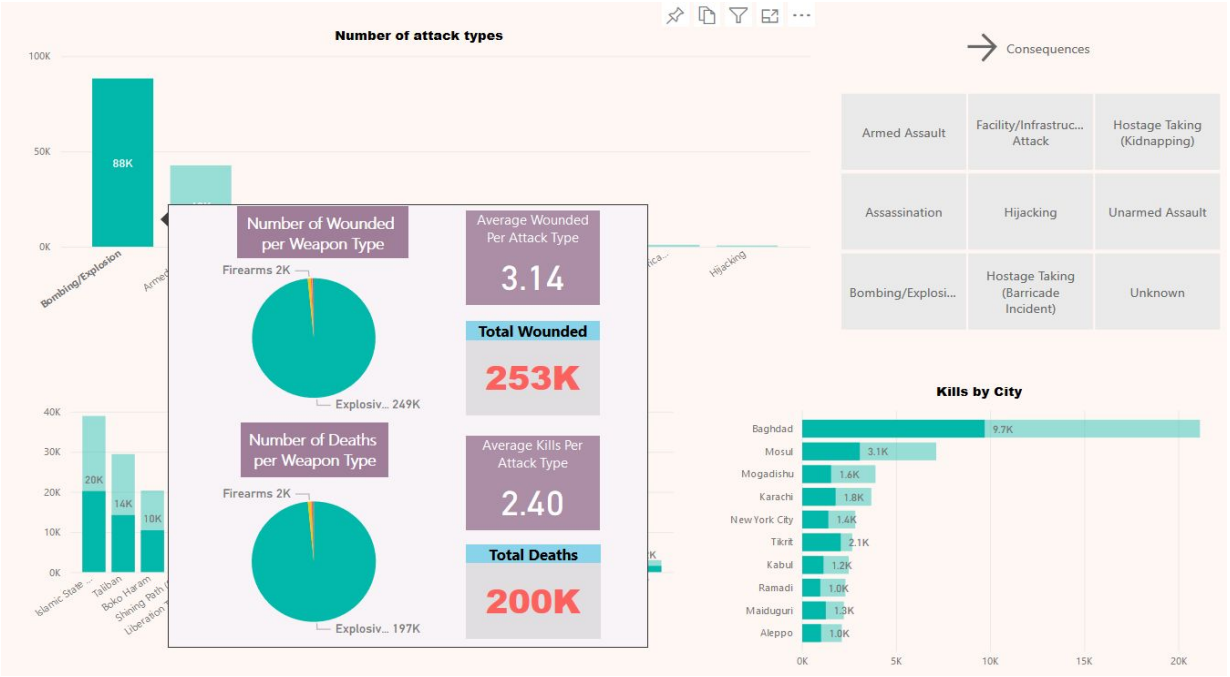
10. Assassination
11. Hijacking
12. Kidnapping
13. Barricade / Incident
14. Bombing/Explosion
15. Armed Assault
16. Unarmed Assault
17. Facility/Infrastructure Attack
18. Unknown

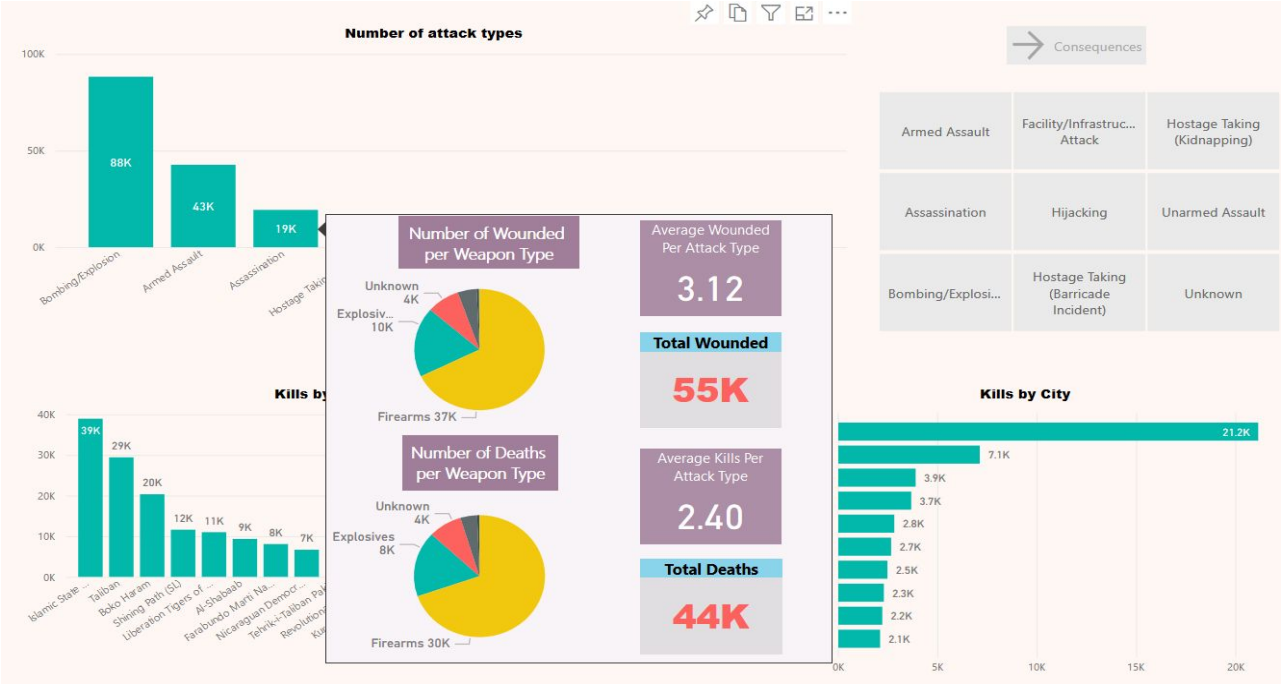
In the above report, we conclude that the most common attack types are *Bombing/Explosion* and *Armed Assault*. Analyzing the kills by the Terrorists' names, *the Islamic State of Iraq and the Levant (ISIL)* and *Taliban* are those terrorists that killed the most people and the cities with the most kills are *Baghdad* and *Mosul*.

We can have all this information according to the attack type.



If we are interested in extracting data according to one type of attack, we hover on it.

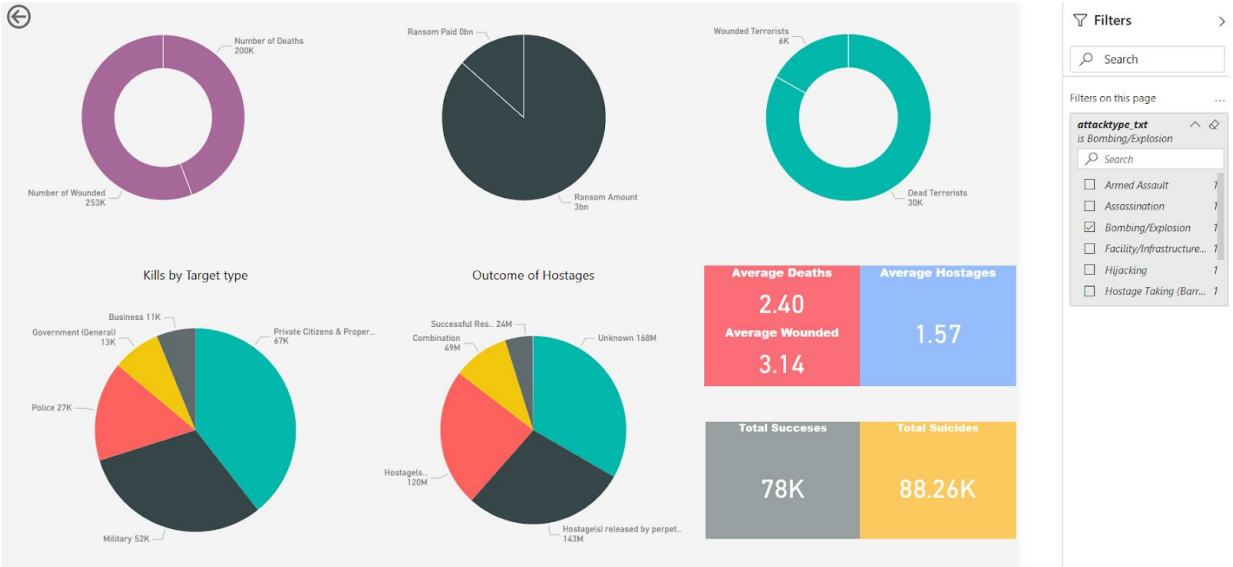




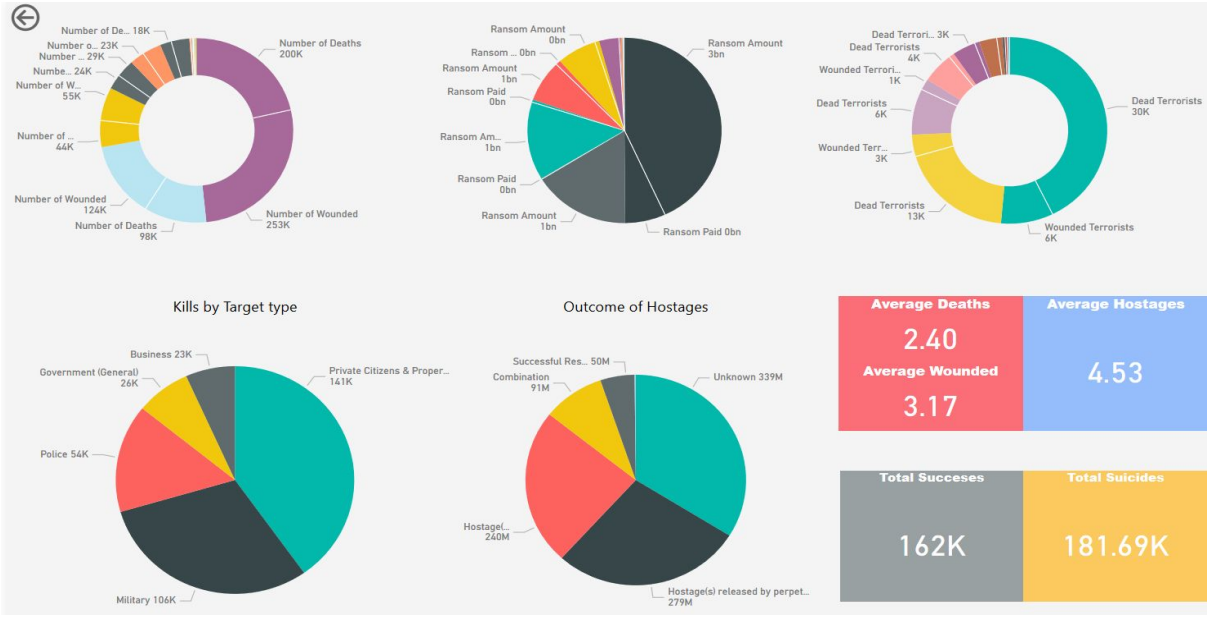
The data we see are

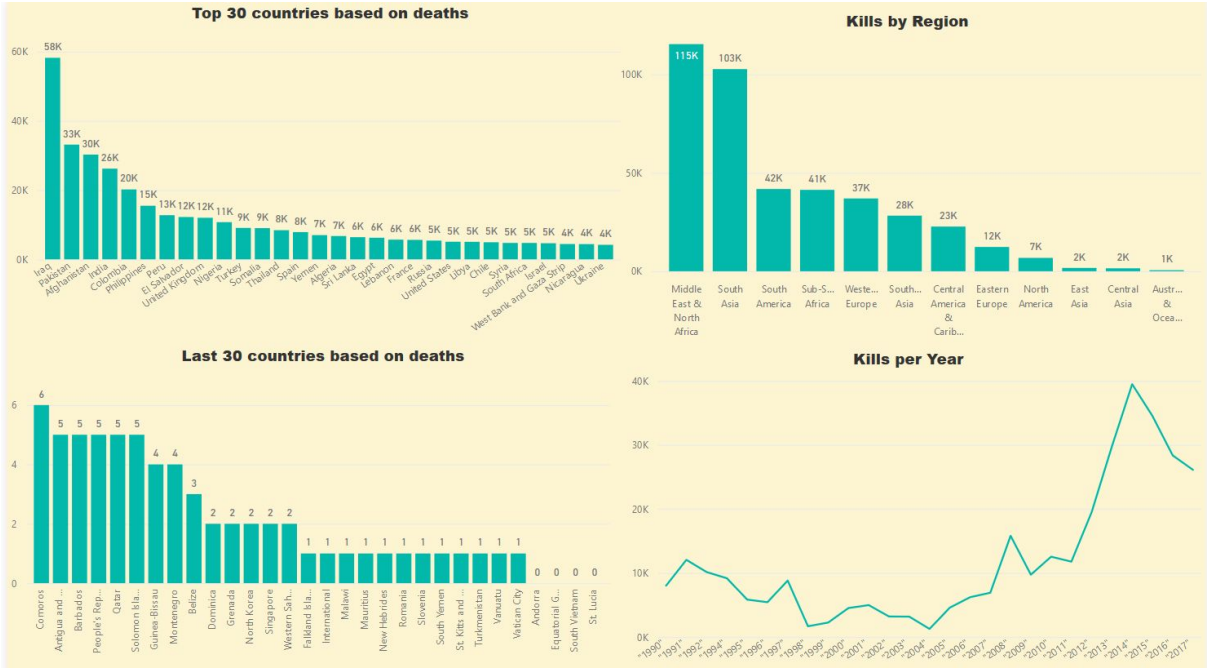
- Number of Wounded per Weapon Type- obviously each and every attack type has different weapons, average wounded per attack type, and the total number.
- Number of Deaths per Weapon Type- obviously each and every attack type has different weapons, average deaths per attack type, and the total number.

<u>Drill Down Report:</u> We can navigate from attack type to consequences, selecting the Type, and then clicking on the "Consequences" button.

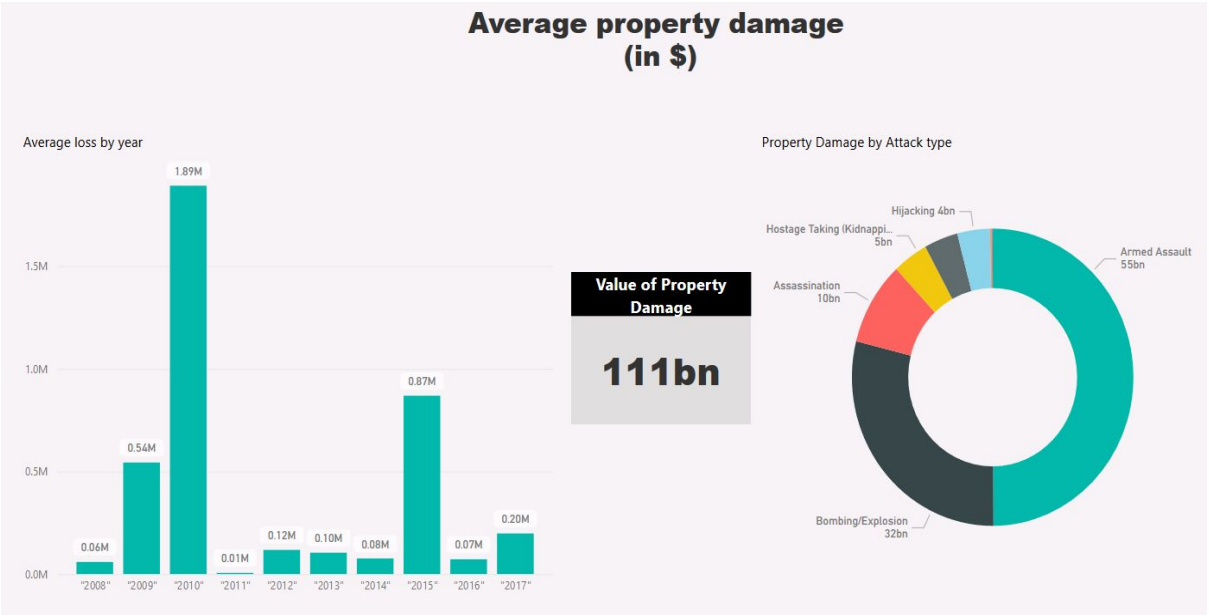


Also, we can have an overall picture about the consequences of the attacks:





The above diagrams show us the kills across counties, kills per region and kills per year. As we can observe, Iraq has the most kills by far (58k). Also, the Middle East & North Africa and South Asia are the two regions with the most kills. In the bottom right diagram we see that during the last 5 years, we have an exponential increase of victims and especially in 2014 where almost 40k kills have been recorded.

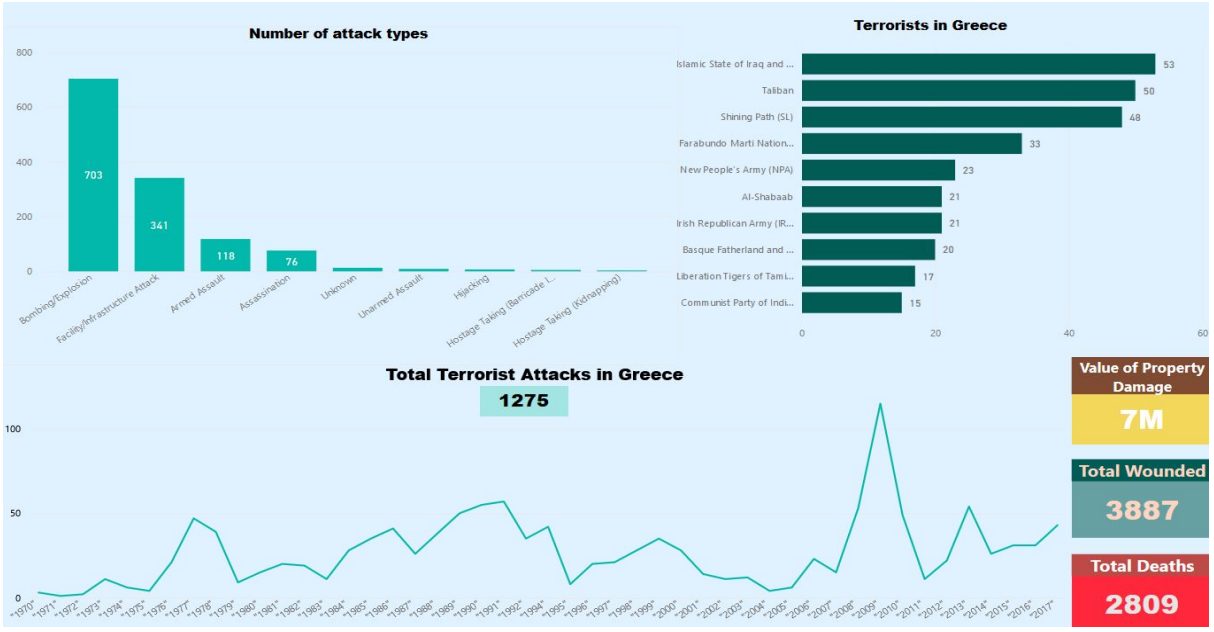


Here, we have depicted the overall value of property damage for the top 10 years. We can conclude that the average loss per year dramatically defers. Most values are relatively small (about 0.2 million) and only 1 of those is very large (2010 - 1.89M).

At the right diagram, we see the property damage categorized by attack type. Armed Assault damage is almost equal with the sum of all the other types of attack combined together. Bombing/Explosion follows.

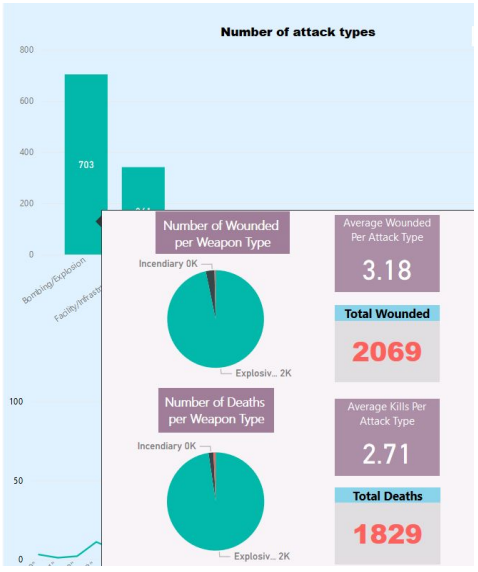
As we have built our graphs in Power BI Desktop we transferred them on the online version of Power BI, so we can easily share them with others and have access no matter the time.

Case Study Greece



This is an overview of the terrorism attacks that occured in Greece. The total number is *1.275*, with *7M* total value of property damage, 3887 number of wounded and 2809 number of deaths. The two most common attack types are Bombing/Exploisoning and facility/infrastructure. We see a peak in the years 2008-2009, more probably because of the death of "Alexis Grhgoropoulos".

Furthermore, if we hover on any attack type, we see more information, such as number of wounded, average kills, etc.



Conclusions

Through this process we can conclude the following:

- Time Analysis
    - The number of attacks is increasing since 2012. The year with the most attacks is 2014, with a total number of 16.903 events.
    - The distribution is uniform across the months and the days, which means that there is no significant difference between the months and days.
- Location Analysis
    - The most Terrorism attacks occurred in the Middle East & North Africa, with a total number of 50.474.
    - In Europe, the total number is 21.783(Eastern Europe-5.144 and Wester 16.639)
    - In Greece, 1.275 attacks were recorded.
- Attack Type
    - The 2 types with most attacks are Bombing/Exploisoning (88k) and Armed Assault (43k)
    - The most effective attack is Armed Assault because in each event the average number of deaths and wounded is bigger than the other attacks (2.44 average number of deaths and 3.18 average number of wounded)
- Kills
    - The Islamic State of Iraq and the Levant (ISIL) and Taliban are those terrorists that has killed the most people
    - Baghdad and Mosul are the cities in where most people have been killed
    - Iraq has the most deaths by far (58k)
    - Middle East & North Africa and South Asia are the two regions with the most kills
    - During the last 5 years, we have an exponential increase of victims and especially in 2014 where almost 40k kills have been recorded
- Property Loss

    - Armed Assault damage is almost equal with the sum of all the other types of attack combined together. Bombing/Explosion follows.
- Consequences
    - Average Deaths per attack is 2.40
    - Average Wounded is 3.17 per attack
    - Average Hostages is 4,53 per attack