

Università "Mediterranea" di Reggio Calabria
Facoltà di Ingegneria

Corso di Laurea in Ingegneria Informatica e dei
Sistemi per le Telecomunicazioni

Sistemi Informatici Aziendali - Anno Accademico 2016/2017

Analisi e rappresentazione dei dati del *Global Terrorism Database* tramite i tool QlikSense, Tableau e MongoDB

Candidati:

**Maria Contursi
Cristina Ferrandello
Manuele Santo Foti
Francesca Pillari**

Indice

1	Il Global Terrorism Database	7
1.1	Cosa è il Global Terrorism Database	7
1.2	I dati del Global Terrorism Database	7
1.3	Gli obiettivi prefissati	8
1.3.1	OB1 - Analisi del numero di attacchi terroristici subiti per zona geografica	9
1.3.2	OB2 - Cambiamento delle tipologie di attacco nel corso del tempo	9
1.3.3	OB3 - Studio delle caratteristiche delle organizzazioni terroristiche	9
1.3.4	OB4 - Calcolo dei danni subiti per zona geografica	9
1.3.5	OB5 - Analisi statistica del numero di attacchi per data	10
1.3.6	OB6 - Analisi sul tipo di armi utilizzate durante gli attacchi terroristici	10
1.3.7	OB7 - Andamento e cambiamento temporale delle cause scatenanti degli attacchi terroristici	10
1.3.8	OB8 - Analisi delle varie modalità di attacco	10
1.3.9	OB9 - Calcolo sull'ammontare delle richieste di riscatto	10
1.3.10	OB10 - Stima del numero di vittime	10
1.4	Il modello a stella del GTD	11
1.4.1	Fact Table - Attacco Terroristico	12
1.4.2	Dimension Table - Arma	13
1.4.3	Dimension Table - Danno	13
1.4.4	Dimension Table - Data	13
1.4.5	Dimension Table - Obiettivo	14
1.4.6	Dimension Table - Organizzazione Terroristica	14
1.4.7	Dimension Table - Tipo Attacco	14
2	Manipolazione dei dati del GDT	12
2.1	ETL: Extract, Trasform and Load	12
2.2	Operazioni di ETL nel GDT	13
2.2.1	Riduzione del numero di campi	13
2.2.2	Riempimento dei valori "null"	13
2.2.3	Rename dei campi	13

2.2.4	Slowly changing dimension	13
3	QlikSense e il GTD	16
3.1	QlikSense	16
3.2	Gli obiettivi del GTD in QlikSense	16
3.2.1	OB1 - Analisi del numero di attacchi terroristici subiti per zona geografica	16
3.2.2	OB2 - Cambiamento delle tipologie di attacco nel corso del tempo	18
3.2.3	OB3 - Studio delle caratteristiche delle organizzazioni terroristiche	18
3.2.4	OB4 - Calcolo dei danni subiti per zona geografica	22
3.2.5	OB5 - Analisi statistica del numero di attacchi per data	22
3.2.6	OB6 - Analisi sul tipo di armi utilizzate per gli attacchi terroristici	22
3.2.7	OB7 - Andamento e cambiamento temporale delle cause scatenanti degli attacchi terroristici	26
3.2.8	OB8 - Analisi delle varie modalità di attacco	26
3.2.9	OB9 - Calcolo sull'ammontare delle richieste di riscatto	26
3.2.10	OB10 - Stima del numero di vittime	30
4	Tableau e il GTD	33
4.1	Tableau	33
4.2	Gli obiettivi del GTD in Tableau	33
4.2.1	OB1 - Analisi del numero di attacchi terroristici subiti per zona geografica	34
4.2.2	OB2 - Cambiamento delle tipologie di attacco nel corso del tempo	34
4.2.3	OB3 - Studio delle caratteristiche delle organizzazioni terroristiche	34
4.2.4	OB4 - Calcolo dei danni subiti per zona geografica	34
4.2.5	OB5 - Analisi statistica del numero di attacchi per data	42
4.2.6	OB6 - Analisi sul tipo di armi utilizzate per gli attacchi terroristici	42
4.2.7	OB7 - Andamento e cambiamento temporale delle cause scatenanti degli attacchi terroristici	42
4.2.8	OB8 - Analisi delle varie modalità di attacco	42
4.2.9	OB9 - Calcolo sull'ammontare delle richieste di riscatto	42
4.2.10	OB10 - Stima del numero di vittime	48
4.2.11	OB11 - Previsione sull'andamento futuro del numero di attacchi terroristici	48
5	MongoDB e il GTD	52
5.1	MongoDB	52
5.2	Gli obiettivi del GTD in MongoDB	52

INDICE

5

5.2.1	OB1 - Analisi del numero di attacchi terroristici subiti per zona geografica	53
5.2.2	OB2 - Cambiamento delle tipologie di attacco nel corso del tempo	57
5.2.3	OB3 - Studio delle caratteristiche delle organizzazioni terroristiche	57
5.2.4	OB4 - Calcolo dei danni subiti per zona geografica	57
5.2.5	OB5 - Analisi statistica del numero di attacchi per data	57
5.2.6	OB6 - Analisi sul tipo di armi utilizzate per gli attacchi terroristici	62
5.2.7	OB7 - Andamento e cambiamento temporale delle cause scatenanti degli attacchi terroristici	62
5.2.8	OB8 - Analisi delle varie modalità di attacco	62
5.2.9	OB9 - Calcolo sull'ammontare delle richieste di riscatto	62
5.2.10	OB10 - Stima del numero di vittime	62

6 Commenti e conclusioni

69

Capitolo 1

Il Global Terrorism Database

1.1 Cosa è il Global Terrorism Database

Il *Global Terrorism Database* (GTD) è un progetto che ha conosciuto i suoi albori nel 2001 ed è tutt'oggi portato avanti dal *National Consortium for the Study of Terrorism and Responses of Terrorism*, operante presso l'Università del Maryland. Il GTD è oggi il principale riferimento per le misure, le statistiche e l'impatto che il terrorismo assume in ogni angolo del mondo.

Le informazioni in esso contenute sono il riassunto di numerosi report reperibili sul web e redatti da diversi media. Inoltre, tutte le sorgenti di informazioni utili sono state verificate e classificate come attendibili prima che qualunque dato da esse proveniente fosse inserita nel GTD.

I dati sono accessibili al seguente indirizzo: <https://www.start.umd.edu/gtd>.

Nel presente elaborato si andrà ad usare il GTD per analizzare ogni aspetto riguardante le attività terroristiche, analizzandole dai vari punti di vista, offertici dalle grandi varietà e dettaglio di informazioni che il dataset contiene.

1.2 I dati del Global Terrorism Database

Come detto, il GTD raccoglie al suo interno dati relativi ad attacchi terroristici. Nella fattispecie, la collezione di dati ha registrato, con una minuzia di dettaglio apprezzabile, tutti gli attacchi terroristici e le insurrezioni di vario genere avvenute tra il 1970 e il 2015, conteggiando un numero di occorrenze pari a 156.773, specificandone per ognuna di esse ben 146 campi contenenti tutti i dettagli.

Tali campi, che vedremo più dettagliatamente nei capitoli successivi, esprimono, per ogni singolo attacco terroristico, informazioni riguardanti la data, la zona geografica, l'arma usata, il target, i danni calcolati, la criticità dell'impatto, l'organizzazione terroristica, le modalità di attacco e molto altro ancora.

Ovviamente, visto la grande quantità di dati a disposizione, è stato necessario valutare bene, parallelamente agli obiettivi del presente lavoro, quali informazioni potessero risultare utili e quali no.

1.3 Gli obiettivi prefissati

Dopo aver illustrato la realtà di interesse che abbiamo deciso di analizzare, esaminiamo ora gli obiettivi che il nostro gruppo di lavoro si è prefissato all'inizio del progetto.

Innanzitutto, è corretto sottolineare che, avendo a che fare con dati che registrano eventi di difficile previsione e, dunque, completamente casuali e slegati da ogni logica predittiva, gli obiettivi che ci siamo posti sono, per la maggior parte, dei riassunti statistici, delle linee guida che mirano ad evidenziare il fenomeno degli attacchi terroristici da vari punti di vista. Sarà dunque una sorta di prevenzione dettata dalla pura statistica, che va interpretata, quindi, come una sorta di tendenza di quel determinato fenomeno e non come una verità assoluta.

Dopo questa breve delucidazione riguardo gli obiettivi, possiamo ora elencare quelli che il nostro gruppo di lavoro si è prefissato per il presente progetto:

- OB1 - Analisi del numero di attacchi terroristici subiti per zona geografica;
- OB2 - Cambiamento delle tipologie di attacco nel corso del tempo;
- OB3 - Studio delle caratteristiche delle organizzazioni terroristiche;
- OB4 - Calcolo dei danni subiti per zona geografica;
- OB5 - Analisi statistica del numero di attacchi per data;
- OB6 - Analisi sul tipo di armi utilizzate durante gli attacchi terroristici;
- OB7 - Andamento e cambiamento temporale delle cause scatenanti degli attacchi terroristici;
- OB8 - Analisi delle varie modalità di attacco;
- OB9 - Calcolo sull'ammontare delle richieste di riscatto;
- OB10 - Stima del numero di vittime.

Per esprimerne meglio il significato, analizziamo nel dettaglio ogni singolo obiettivo.

1.3.1 OB1 - Analisi del numero di attacchi terroristici subiti per zona geografica

La presente analisi mira ad evidenziare l'impatto degli attacchi terroristici, da un punto di vista prettamente geografico-quantitativo, nel periodo storico contemplato dal GTD. Nella fattispecie, vengono elencati il numero di attacchi subiti per area geografica, applicando un drill-down per zone, partendo da una suddivisione per continente, fino ad arrivare ad una granularità che esprime un conteggio di attacchi per città.

1.3.2 OB2 - Cambiamento delle tipologie di attacco nel corso del tempo

Il presente obiettivo mira a mettere in evidenza come le modalità di attacco (come, ad esempio, *Assassinio* o *Esplosione*) siano diventate più o meno frequenti nel tempo. L'analisi ha previsto la suddivisione dell'arco temporale del GTD in ulteriori sottointervalli. Lo studio è stato effettuato nelle seguenti porzioni temporali:

- 1970 - 1980;
- 1980 - 1990;
- 1990 - 2015.

Per ognuno di questi intervalli è stato elaborato lo studio precedentemente descritto.

1.3.3 OB3 - Studio delle caratteristiche delle organizzazioni terroristiche

La presente analisi mira ad evidenziare le caratteristiche principali delle organizzazioni terroristiche, classificate per importanza in base al numero di attacchi perpetrati. Nello studio vengono evidenziate le seguenti caratteristiche:

- il numero di attacchi perpetrati;
- tramite un drill-down, gli anni di maggiore attività di ogni singola organizzazione terroristica;
- la qualità dell'organizzazione terroristica, espressa tramite il rapporto tra il numero di attacchi e il numero di attacchi riusciti.

1.3.4 OB4 - Calcolo dei danni subiti per zona geografica

Il presente obiettivo ha il compito di esprimere la quantità di danni subiti, quantificati in migliaia di dollari. Anche in questo caso è presente un drill-down che scende nel dettaglio dell'analisi, partendo, come prima, da una granularità a livello continentale fino ad arrivare ad una suddivisa per città.

1.3.5 OB5 - Analisi statistica del numero di attacchi per data

Il presente obiettivo è una pura analisi statistica. Quest'ultimo infatti esprime, in termini prettamente statistici, quali sono stati l'anno, il mese e il giorno con più attacchi terroristici. L'analisi effettuata può essere ulteriormente elaborata, nel caso di mese e giorno, per tirare fuori una sorta di probabilità che l'evento si verifichi.

1.3.6 OB6 - Analisi sul tipo di armi utilizzate durante gli attacchi terroristici

La presente analisi mira ad evidenziare il tipo di armi usate da ogni organizzazione terroristica. Ovviamente, il raggruppamento viene fatto per organizzazione terroristica, appunto, ed inizialmente si ha il semplice conteggio delle armi utilizzate. Successivamente, tramite un drill-down, si può accedere alla informazioni sul genere e ai sottogeneri di armi più utilizzate.

1.3.7 OB7 - Andamento e cambiamento temporale delle cause scatenanti degli attacchi terroristici

Tale obiettivo mira ad evidenziare, tramite un opportuno andamento temporale, quali sono state e come sono mutate nel tempo le cause politiche, sociali o criminali, che sono state diretta conseguenza o sono sfociati in un attacco terroristico.

1.3.8 OB8 - Analisi delle varie modalità di attacco

Il presente obiettivo mira ad evidenziare, da due punti di vista differenti, gli attacchi terroristici che hanno visto modalità di perpetrazione di tipo multiplo e di tipo suicida. Nella fattispecie, il primo punto di vista è espresso dall'andamento temporale di questi attacchi, mentre il secondo conteggia le due modalità, classificandole per anno.

1.3.9 OB9 - Calcolo sull'ammontare delle richieste di riscatto

Questo obiettivo interessa solo quegli attacchi terroristici che hanno previsto un riscatto. Scendendo più nel dettaglio, tale analisi mostra graficamente, per ogni anno, il rapporto tra il numero di riscatti e la somma delle quantità di denaro richieste per quest'ultimi.

1.3.10 OB10 - Stima del numero di vittime

L'ultimo degli obiettivi rappresenta una stima del numero di vittime. Nel dettaglio, il raggruppamento viene fatto per organizzazione terroristica, per ognuna

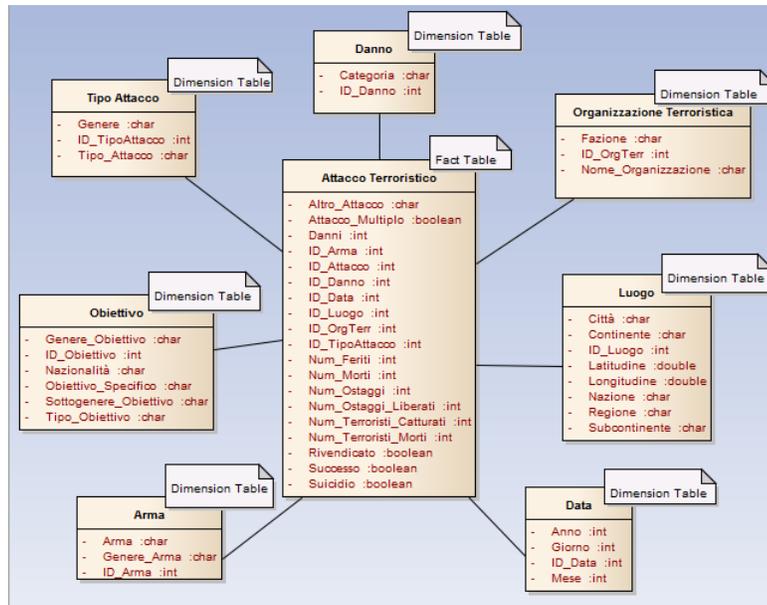


Figura 1.1: Modello Evento - Dimensioni

delle quali vengono conteggiati il numero di vittime tra i civili e quelle tra i terroristi, esprimendo dunque, tramite il rapporto di queste due misure, il grado di qualità/affidabilità dell'organizzazione stessa.

1.4 Il modello a stella del GTD

Dopo aver prefissato gli obiettivi del presente lavoro, è stato opportuno, per avere graficamente una visione più chiara del sistema complessivo, sviluppare un modello che metta in evidenza le relazioni tra *fatti* e *dimensioni*.

Innanzitutto, si definisce *fatto*, come dice il nome stesso, un evento di importanza centrale all'interno della realtà d'analisi; mentre, con il termine *dimensione* si intende un particolare punto di osservazione dell'evento stesso, o comunque qualcosa ad esso strettamente correlato.

Per comprendere meglio questi due concetti, andiamo ad analizzare il modello a stella elaborato per il Global Terrorism Database. (Fig.1).

Il modello in questione ha previsto l'inserimento di un unico evento, ovvero *Attacco Terroristico*, che è ovviamente il cuore del Global Terrorism Database. A quest'ultimo sono state collegate ben sette dimensioni, le quali sono di seguito elencate:

Attacco Terroristico	
-	Altro_Attacco :char
-	Attacco_Multiplo :boolean
-	Danni :int
-	ID_Arma :int
-	ID_Attacco :int
-	ID_Danno :int
-	ID_Data :int
-	ID_Luogo :int
-	ID_OrgTerr :int
-	ID_TipoAttacco :int
-	Num_Feriti :int
-	Num_Morti :int
-	Num_Ostaggi :int
-	Num_Ostaggi_Liberati :int
-	Num_Terroristi_Catturati :int
-	Num_Terroristi_Morti :int
-	Rivendicato :boolean
-	Successo :boolean
-	Suicidio :boolean

Figura 1.2: Fact Table *Attacco Terroristico*

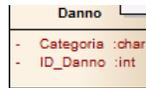
- Arma;
- Danno;
- Data;
- Luogo;
- Obiettivo;
- Organizzazione Terroristica;
- Tipo Attacco.

Si hanno, dunque, utilizzando il gergo adatto, una *fact table* (FT) collegata ad altre sette *dimension table* (DT), ognuna delle quali presenta le proprie caratteristiche e *misure*. Quest'ultime rappresentano, invece, gli aspetti quantitativi del modello, a differenza delle *dimensioni* che invece inquadrano più che altro l'aspetto qualitativo.

Andiamo adesso ad analizzare singolarmente ognuna delle tabelle precedentemente nominate.

1.4.1 Fact Table - Attacco Terroristico

La tabella *Attacco Terroristico* è l'unica a rappresentare un fatto per il modello tratto per il Global Terrorism Database. Rappresenta, dunque, il cuore pulsante del progetto, come dimostrano tutti gli identificativi ad essa collegate, i quali mettono in risalto la correlazione della presente tabella con tutte le altre.

Figura 1.3: Dimension Table *Arma*Figura 1.4: Dimension Table *Danno*

Oltre agli identificativi, si notano vari attributi di tipo *int* e *boolean*, come ad esempio *NumMorti* e *AttaccoMultiplo*, che rappresentano le misure quantitative, prettamente additive, dalle quali poi tirar fuori le analisi elencate ed esplicate nei paragrafi precedenti.

1.4.2 Dimension Table - Arma

La prima tra le dimension table è *Arma*. In essa, oltre al suo personale identificativo, sono inseriti due attributi di tipo *string*, *Genere Arma* e *Arma*, utili per le analisi di alcuni obiettivi.

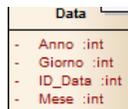
1.4.3 Dimension Table - Danno

La presente dimension table è forse la più semplice come struttura tra quelle elencate, in quanto presenta solo un identificativo e la categoria di danno, ovvero se è un danno a proprietà private o a strutture pubbliche, ad esempio.

Dopo attente considerazioni del caso, il nostro team di lavoro ha deciso che il quantitativo effettivo, espresso in migliaia di dollari, del danno reale è stato opportuno inserirlo nella fact table *Attacco Terroristico*, per evitare eccessiva diversificazione delle tuple della presente tabella.

1.4.4 Dimension Table - Data

La presente dimension table esprime la data in cui è avvenuto un determinato attacco terroristico, specificandone l'anno, il mese e il giorno. Anche in questo caso gli attributi elencati possono essere elaborati per redigere alcune statistiche, come espresso, ad esempio, per l'obiettivo OB5.

Figura 1.5: Dimension Table *Data*

Obiettivo
- Genere_Obiettivo :char
- ID_Obiettivo :int
- Nazionalità :char
- Obiettivo_Specifico :char
- Sottogenere_Obiettivo :char
- Tipo_Obiettivo :char

Figura 1.6: Dimension Table *Obiettivo*

Organizzazione Terroristica
- Fazione :char
- ID_OrgTerr :int
- Nome_Organizzazione :char

Figura 1.7: Dimension Table *Organizzazione Terroristica*

1.4.5 Dimension Table - Obiettivo

La presente dimension table mira ad evidenziare tutte le caratteristiche proprie dell'obiettivo di un generico attacco terroristico. Nella fattispecie, vengono elencati, in maniera via via più specifica, degli attributi utili per il raggiungimento di alcuni obiettivi prefissati e che inquadrano in maniera dettagliata il target dell'attentato.

1.4.6 Dimension Table - Organizzazione Terroristica

La dimension table in questione raggruppa al suo interno alcune informazioni (che non sono altro che gli attributi della tabella) relative ad ogni organizzazione terroristica. Entrando nel dettaglio, vengono specificate sia il nome dell'organizzazione, sia quello del sottogruppo o fazione.

1.4.7 Dimension Table - Tipo Attacco

L'ultima dimension table ha il compito di rappresentare le informazioni, e quindi gli attributi, relativi al tipo di attacco compiuto, arrivando a specificare anche il genere o il motivo scatenante (come *Insurrezione* o *Rappresaglia*, ad esempio).

Tipo Attacco
- Genere :char
- ID_TipoAttacco :int
- Tipo_Attacco :char

Figura 1.8: Dimension Table *Tipo Attacco*

Capitolo 2

Manipolazione dei dati del GDT

2.1 ETL: Extract, Trasform and Load

Dopo la prima fase di presentazione dei dati, il secondo capitolo è dedicato ad una prassi informatica molto importante per le operazioni di business intelligence e big data analytics, ovvero la fase di *Extract, Transform and Load*, più comunemente etichettata col suo acronimo ETL.

Questa è una fase di cruciale importanza per il dataset che viene poi dato in input ai tool di analisi; essa prevede, infatti, come si potrebbe anche intuire dal nome, una sorta di pulizia, di preparazione e di ordinamento dei dati, rendendoli più chiari, completi ed uniformi.

Possiamo dire, dunque, che le operazioni di ETL consistono in un prologo essenziale alla fase di analisi vera e propria, senza della quale gli obiettivi da perseguire potrebbero non essere raggiunti o, ancora peggio, potrebbero essere raggiunti con un risultato differente da quello reale.

Ovviamente, a livello teorico, la procedura di ETL prevede tre fasi, che sono appunto di estrazione, trasformazione e caricamento dei dati.

La fase di estrazione dei dati (*Extract*) è una sorta di selezione di tutti quei dati o campi che, all'interno del dataset, necessitano di una modifica per essere più comprensibili e consistenti.

La fase di trasformazione dei dati (*Transform*) è la parte in cui si ha la vera normalizzazione delle informazioni presenti nel dataset, operazione grazie alla quale si ha la sostituzione dei vecchi dati sporchi o incompleti con quelli nuovi. Infine, durante la fase di caricamento dei dati (*Load*) si ha l'aggiornamento e l'ottenimento dell'intero dataset, ottenendo un archivio di informazioni più pulito e uniforme, ma, soprattutto, pronto per essere analizzato.

2.2 Operazioni di ETL nel GDT

Analizziamo adesso le varie operazioni di ETL che il nostro team di lavoro ha ritenuto opportuno applicare sul Global Terrorism Database, sfruttando in primo momento *Excel* e successivamente le funzioni di modifica di *MongoDB*.

2.2.1 Riduzione del numero di campi

Il dataset del GTD racchiude un numero davvero considerevole di informazioni, molte delle quali risultavano superficiali o addirittura inutili per gli obiettivi prefissati per il presente elaborato.

Occorreva, dunque, eliminare tutte queste informazioni, in modo tale che le operazioni di analisi fossero più rapide, non solo a livello visivo per il nostro team di lavoro, ma soprattutto a livello computazionale, riducendo di molto il numero di campi da analizzare.

Nella fattispecie, si è passati da un numero di colonne pari a **146** ad un numero di soli, per così dire, **54** campi.

2.2.2 Riempimento dei valori "null"

Con valori *null* si intendono tutte quelle celle, identificate dall'intersezione di una riga e una colonna, di cui non è stato specificato alcun valore. Ciò non è buono nè a livello visivo, nè a livello semantico, in quanto, spesso e volentieri, un valore *null* indica che la fonte di quel particolare campo, per una particolare istanza, è sconosciuto, ma ciò rappresenta comunque una conoscenza che va assolutamente considerata nell'analisi per il conseguimento degli obiettivi prefissati.

A ragion per cui, entrando nel dettaglio, i valori *null* sono stati riempiti uniformandoli tal volta ad altri valori presenti (come nel caso dei *boolean* in cui erano presenti solo gli 1), altre volte attingendo dalla documentazione fornitaci insieme al dataset, in cui viene specificato come le celle vuote rappresentino, ad esempio, solo per citarne a paio, quando un *Unknown*, quando un *Undefined*.

2.2.3 Rename dei campi

Durante questa fase di ETL ci si è premurati, per una questione di pura leggibilità, per intuire più facilmente il significato intrinseco delle informazioni di un campo, quello di rinominare ogni campo, traducendo l'iniziale digitura inglese in una più comprensibile in lingua italiana.

Solo per fare un esempio, il campo iniziale *suicide* è stato rinominato come *AttaccoSuicida*.

2.2.4 Slowly changing dimension

Con *slowly changing dimension* si intende, per l'appunto, il lento cambiamento di una certa dimensione nel tempo. Tale cambiamento, che spesso può passare

inosservato, ha invece importanza cruciale ed è corretto considerarne le opportune modifiche, affinché gli obiettivi prefissati vengano raggiunti con la massima precisione possibile.

Per esprimere meglio il concetto, consideriamo un esempio con il quale il nostro team di lavoro si è interfacciato in prima persona. Infatti, il GTD, in quanto contenitore degli eventi terroristici avvenuti tra il 1970 e il 2015, è stato spettatore del lento mutamento di quella dimensione che è rappresentata dai domini di ogni nazione. Nella fattispecie, alcune di esse si sono fuse in un unico stato, altre invece si sono rese protagoniste di tremende scissioni; ad ogni modo, per rendere le analisi il più attuali e precise possibile, è opportuno considerare le disposizioni geografiche odierne, a ragion per cui sono state necessarie le seguenti modifiche:

- East Germany → Germany;
- West Germany → Germany;
- North Vietnam → Vietnam;
- South Vietnam → Vietnam;
- Yugoslavia → Albania;
- Yugoslavia → Bosnia Herzegovina;
- Yugoslavia → Croatia;
- Yugoslavia → Kosovo;
- Yugoslavia → Macedonia;
- Yugoslavia → Montenegro;
- Yugoslavia → Serbia.

Nel caso in cui vi è stata la fusione di due o più nazioni, l'operazione effettuata è stato semplicemente quella di sostituire a ognuna di esse il nome della nuova macronazione. Mentre, nel caso in cui vi è stata una scissione di una nazione in altri sottostati, è stata effettuata una ricerca per città, per verificare a quale stato appartiene oggi quest'ultima.

Capitolo 3

QlikSense e il GTD

La parte di reale analisi dei dati inizia in questo capitolo, dentro il quale spiegheremo, appunto, come il nostro team di lavoro ha perseguito gli obiettivi prefissati utilizzando il primo tra i tool proposti: *QlikSense*.

3.1 QlikSense

QlikSense è un potente software per la visualizzazione dei dati. Questo tool offre una vasta gamma di grafici, tramite i quali è possibile rappresentare i propri dati in maniera semplice, intuitiva e, soprattutto, con un aspetto grafico molto chiaro e di impatto. QlikSense cura molto l'aspetto estetico della rappresentazione dei propri dati, garantendo all'utente una possibilità di modellazione dei grafici abbastanza ampia, per rendere le elaborazioni delle informazioni più gradevoli. QlikSense, più nello specifico, è soprattutto un tool che permette di tirar fuori dai propri dati alcune relazioni o informazioni che non si credeva ci potessero essere. È, dunque, una fonte di ulteriori informazioni, ottenute tramite l'analisi delle misure e delle dimensioni dei dati mandatigli come input.

3.2 Gli obiettivi del GDT in QlikSense

Analizziamo singolarmente gli obiettivi che il nostro team di lavoro si è prefissato, andando questa volta a rappresentarli tramite il tool *QlikSense*.

3.2.1 OB1 - Analisi del numero di attacchi terroristici subiti per zona geografica

Nella Figura 3.1 è presente uno screenshot dei grafici tratti per il presente obiettivo. Nel dettaglio, sono presenti due grafici, uno a barre e uno a torta, e un contatore, il quale evidenzia il numero totale di attacchi presenti nel dataset. Per quanto concerne i due grafici, essi rappresentano la stessa informazione, - in maniera differente a livello visivo: viene fatto un conteggio del numero di

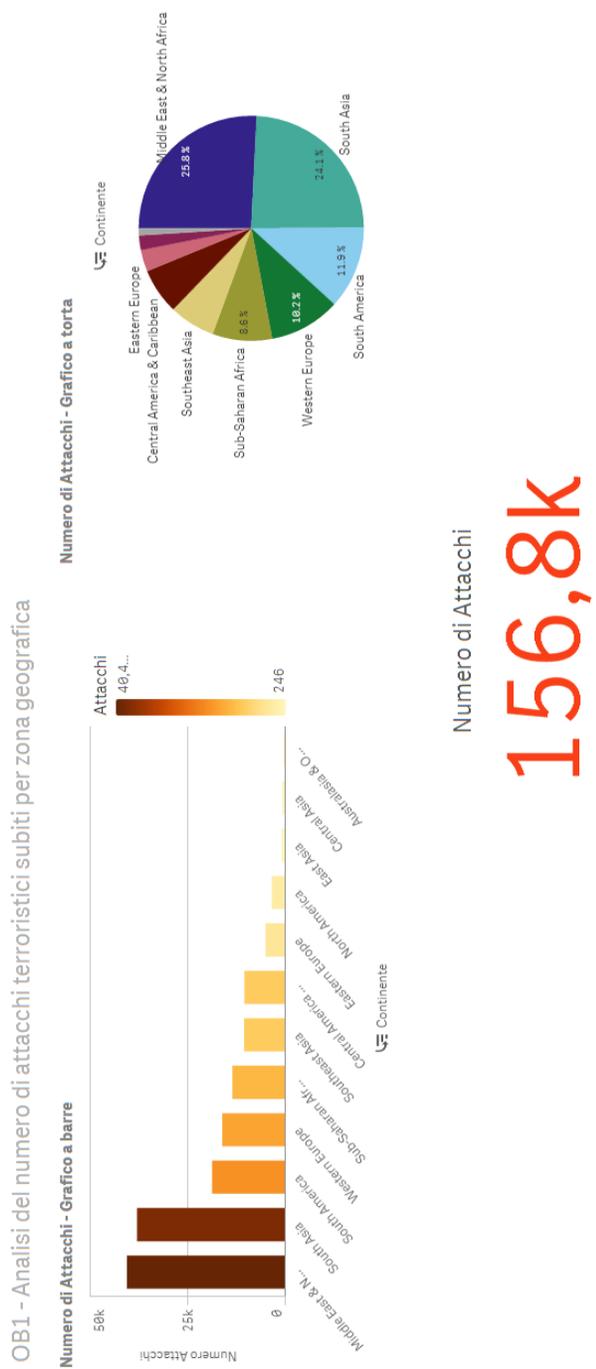


Figura 3.1: Grafico a barre e grafico a torta per OB1

attacchi subiti per area geografica, partendo da una granularità continentale. In Figura 3.2 invece viene evidenziato la possibilità di effettuare un drill-down dei dati, ovvero la possibilità di scendere più nel dettaglio, fino ad effettuare un conteggio di attacchi subiti per città. Evidenziando una parte dell'uno o dell'altro grafico, si ha l'immediata ed automatica modifica degli altri, visualizzando istantaneamente e in maniera intuitiva i dati elaborati per una determinata sottoparte.

Nella fattispecie, nell'esempio di Figura 3.2 viene selezionato nel grafico a torta lo spicchio relativo all'Est Europa e, subito, il grafico a barre va a modificarsi andando a conteggiare gli attacchi subiti, raggruppati questa volta per nazione, di questo continente. Analogamente si potrebbe conteggiare il numero di attacchi per città semplicemente evidenziando una barra del grafico a barre.

3.2.2 OB2 - Cambiamento delle tipologie di attacco nel corso del tempo

Nella Figura 3.3 viene rappresentato il secondo obiettivo che il nostro team di lavoro si è prefissato, ovvero quello di evidenziare come sono cambiate le modalità di attacco nel corso del tempo. Per fare ciò, è stato suddiviso l'arco temporale trattato dal GTD in tre sottointervalli:

- 1970 - 1980;
- 1980 - 1990;
- 1990 - 2015.

Il metodo di rappresentazione per ogni sottointervallo è stato quello del grafico a barre orizzontale, grazie ai quali è stato possibile evidenziare una minore tendenza all'attacco terroristico per *assassinio*, ma una maggiore tendenza invece per quelli ad *assalto armato*.

La costante temporale, sempre leader indiscussa tra le tipologie di attacco, rimane quella di genere *esplosivo*.

Inoltre, nel grafico a torta, per avere un quadro dell'obiettivo più generale e completo, è stata fatta un'analisi, ovvero un conteggio degli attacchi per tipologia, su tutto l'arco temporale.

3.2.3 OB3 - Studio delle caratteristiche delle organizzazioni terroristiche

La Figura 3.4 mette in risalto alcune delle caratteristiche delle varie organizzazioni terroristiche tramite dei grafici a barre. Nella fattispecie, nei due grafici in alto viene effettuato un conteggio degli attacchi perpetrati da ogni singola organizzazione, mentre nei rimanenti due quest'ultimo conteggio viene raffrontato, invece, con il numero di attacchi riusciti, escludendo dal conteggio generale quelli che non sono andati a buon fine.

È stata effettuata una diversificazione anche in senso verticale, nel senso che

OB2 - Cambiamento delle tipologie di attacco nel corso del tempo

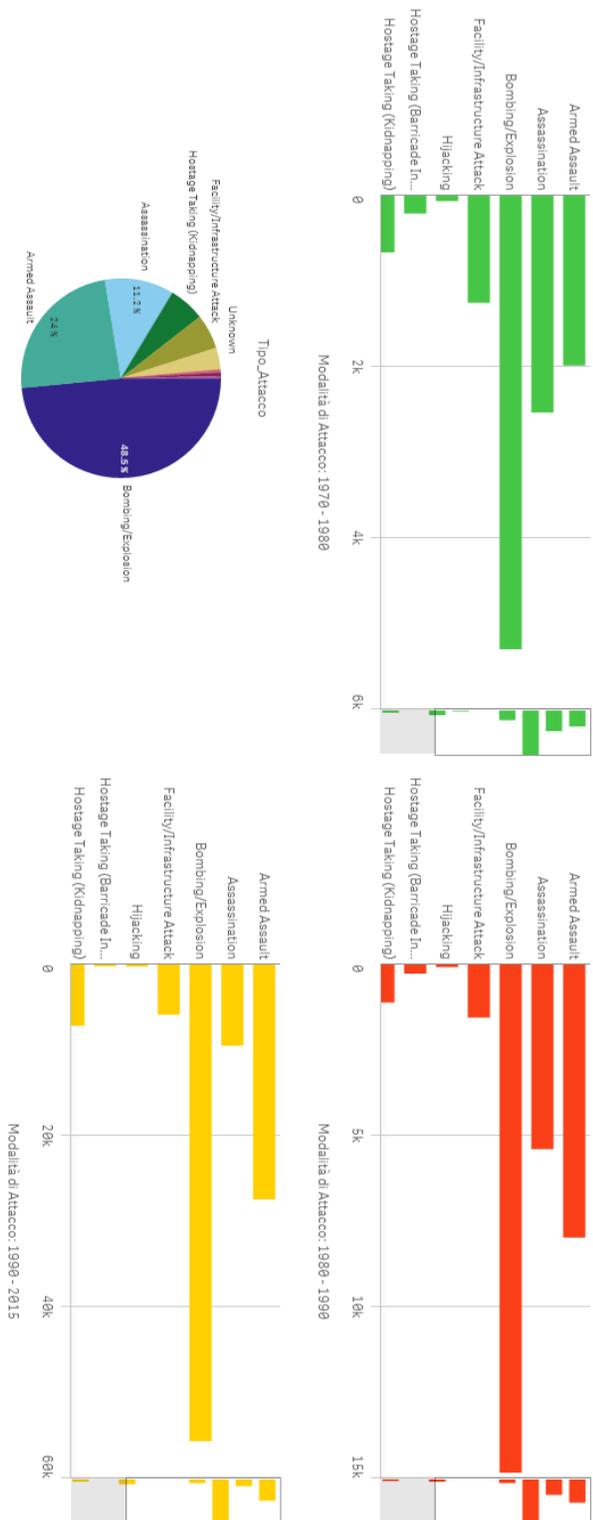


Figura 3.3: Grafici a barre e grafico a torta per OB2

i grafici di sinistra contengono anche gli attacchi con matrice sconosciuta, che rappresentano la stragrande maggioranza e anche quelli con più insuccessi, mentre in quelli di destra sono stati conteggiati solo gli attentati rivendicati da una qualche organizzazione terroristica, facendo emergere che la più attiva tra esse corrisponde al gruppo dei *talebani*.

3.2.4 OB4 - Calcolo dei danni subiti per zona geografica

Nel OB4 è stata prodotta un'analisi sui danni prodotti dagli attacchi terroristici, quantificati in migliaia di dollari e raggruppati per zona geografica.

Come nel caso di OB1, anche in questo obiettivo è previsto un drill-down che, ovviamente, riguarda la granularità per area geografica, partendo dal continente fino ad arrivare ad una stima per città.

Dalle analisi, mostrate in Figura 3.5, è emerso che:

1. il continente con più danni è il *Medio Oriente e Nord Africa*;
2. la nazione con più danni è l'*Iraq*;
3. la città con più danni è *Baghdad*.

Ovviamente, come volevasi dimostrare, ad un numero maggiore di attacchi subiti corrisponde anche un numero maggiore di danni.

3.2.5 OB5 - Analisi statistica del numero di attacchi per data

I grafici a barre della Figura 3.6 mettono in evidenza, stilando una classifica, l'anno, il mese e il giorno in cui è avvenuto il numero maggiore di attacchi.

Dalle analisi è emerso che nel corso degli ultimi anni il numero di attacchi terroristici è aumentato drasticamente, consegnando il primato al 2014, seguito a ruota dal 2015. L'anno più calmo invece è il 1971 con soli 470 attacchi terroristici.

Per quanto riguarda la probabilità di attacco a livello mensile, essa è pressochè equiprobabile, con un valore leggermente maggiore per il primatista Maggio, fino ad un graduale ma lieve discendimento fino al meno probabile Dicembre.

A livello di giorni la situazione è ancora più equilibrata, registrando solo un lieve vantaggio per i giorni 15 e 1; il 31, ovviamente, presentandosi con cadenza dimezzata rispetto agli altri, risulta il giorno meno probabile per un attacco terroristico.

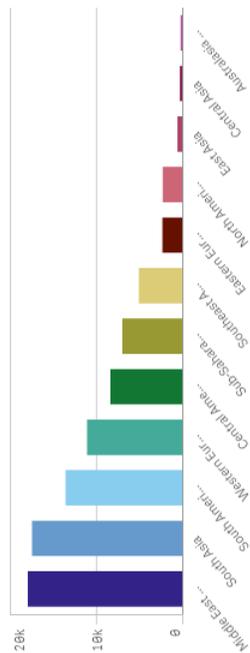
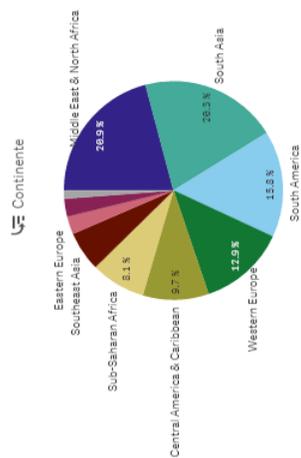
3.2.6 OB6 - Analisi sul tipo di armi utilizzate per gli attacchi terroristici

La Figura 3.7 mira a rappresentare alcune caratteristiche delle organizzazioni terroristiche, più nello specifico il numero di armi usate e quelle maggiormente utilizzate. Ciò è possibile grazie ad un drill-down attuato su ogni organizzazione

OB4 - Calcolo dei danni subiti per zona geografica

Danni Subiti

Continente -> Nazione -> Provincia -> Città



Quantità di Danni calcolati tra il 1970 e il 2015

85,83k

Figura 3.5: Grafico a barre e grafico a torta per OB4

OB5 - Analisi statistica del numero di attacchi per data

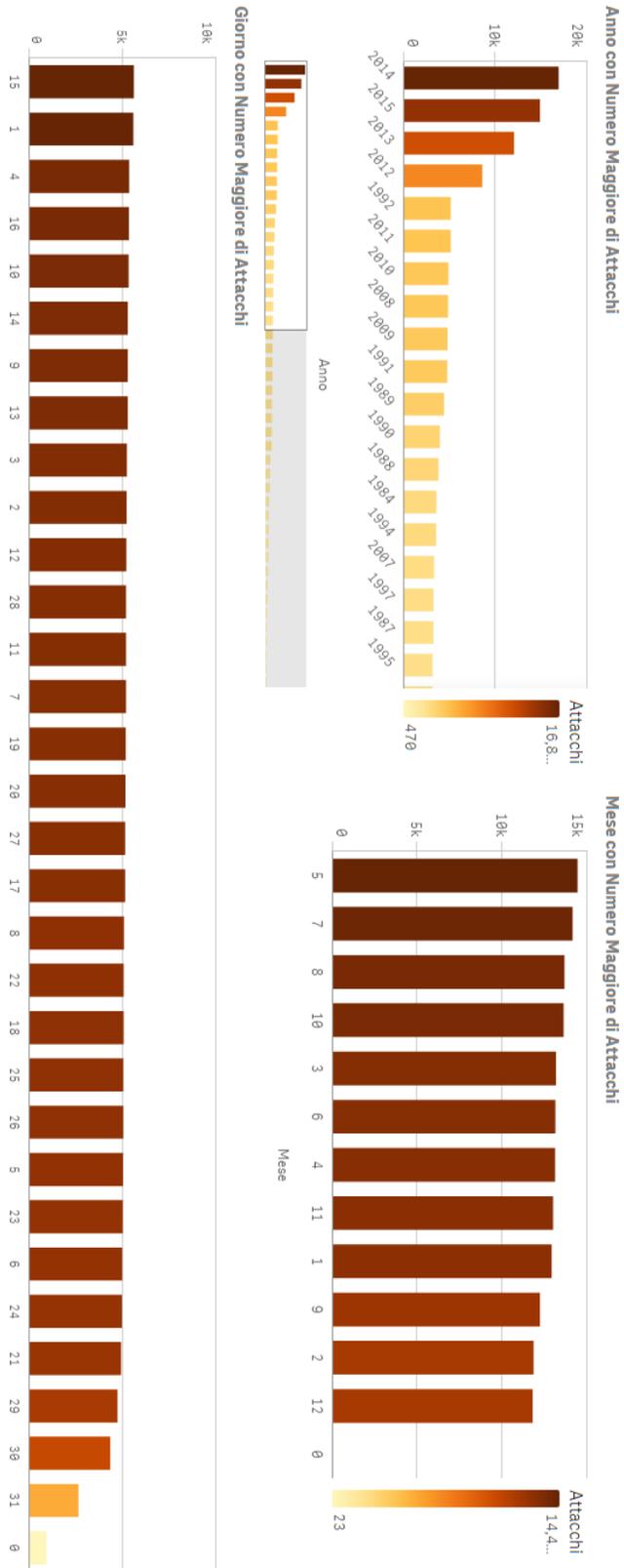


Figura 3.6: Grafici a barre per OB5

terroristica che, evidenziando la colonna che la rappresenta, come mostrato in OB1, fa un conteggio per genere di arma e arma specifica, ordinandole in modo decrescente, evidenziando dunque quelle più utilizzate durante gli attacchi terroristi perpetrati.

3.2.7 OB7 - Andamento e cambiamento temporale delle cause scatenanti degli attacchi terroristici

In Figura 3.8 vengono mostrati per la prima volta i grafici lineari. Essi descrivono l'andamento di una certa misura nel corso di una ben specifica dimensione. Nel nostro caso la dimensione trattata è il tempo e all'interno dei grafici vengono specificati gli andamenti temporali delle ragioni scatenanti degli attacchi terroristici. Più nello specifico, abbiamo notato che quasi tutte le ragioni scatenanti hanno avuto un notevole incremento negli ultimi anni (dovuto al fatto che il numero di attacchi in sé è aumentato molto) e che solo le ragioni per *crimini generici* o per *conflitti tra organizzazione terroristiche* hanno avuto un andamento leggermente più variegato.

È inoltre presente un grafico a torta che inquadra la situazione complessiva, in cui viene esplicitato che le maggiori cause scatenanti di attacchi terroristici sono le *insurrezioni*.

3.2.8 OB8 - Analisi delle varie modalità di attacco

La Figura 3.9 mostra i grafici che rappresentano l'analisi dell'obiettivo OB8. Per questa analisi si è pensato di analizzare le più famose modalità di attacco terroristico perpetrate, ovvero quella di *attacco multiplo* e *attacco suicida*.

Entrambe le misure sono rappresentate, in alto, in un grafico lineare, nel quale oltre al notevole e previsto incremento esponenziale degli ultimi anni, ha conosciuto un notevole aumento anche agli inizi del nuovo millennio.

Nel grafico in basso, quello a barre, vengono invece incolonnate e messe a confronto entrambe le misure, quantificandole e ordinandole per anno in senso decrescente.

3.2.9 OB9 - Calcolo sull'ammontare delle richieste di riscatto

La Figura 3.10 mostra l'unico grafico a dispersione utilizzato tra gli obiettivi prefissati. In questo tipo di grafico vengono implementate due misure, una per ognuna dei due assi, basandosi su una determinata dimensione. Da queste due misure vi si otterranno delle proiezioni su piano di quella specifica dimensione. Per capire meglio il concetto riferiamoci al nostro caso reale.

La dimensione trattata è quella temporale, più specificatamente quella della rappresentazione per *anno*; mentre le misure contemplate sono:

OB7 - Andamento e cambiamento temporale delle cause scatenanti degli attacchi terroristici

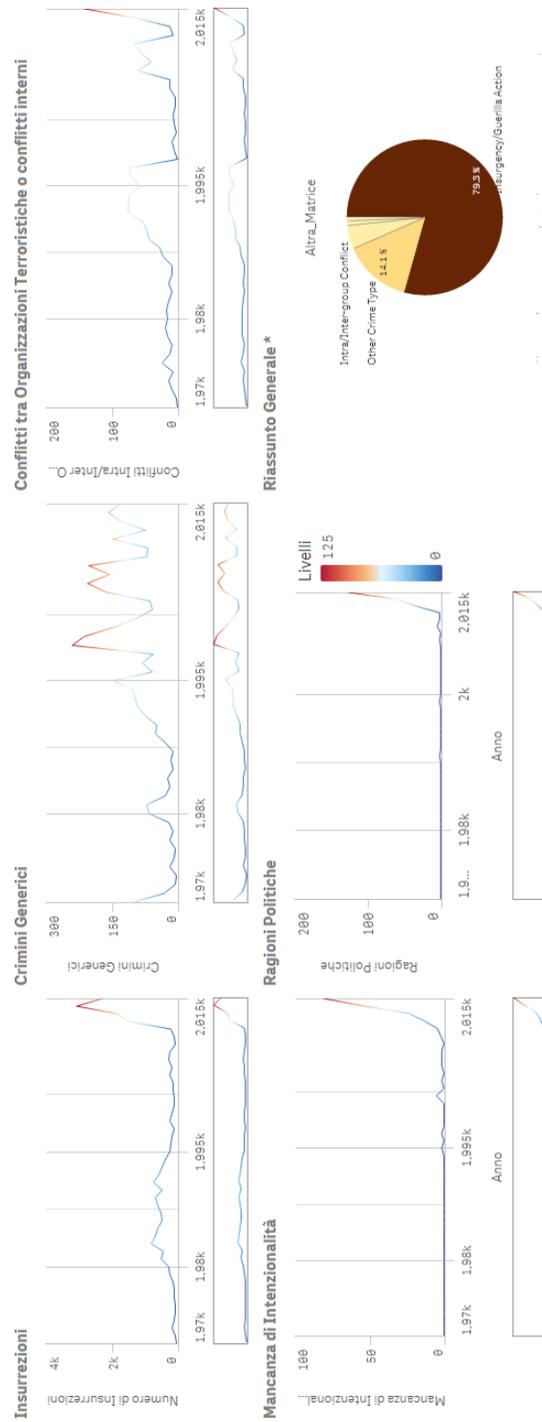


Figura 3.8: Grafici lineari per OB7

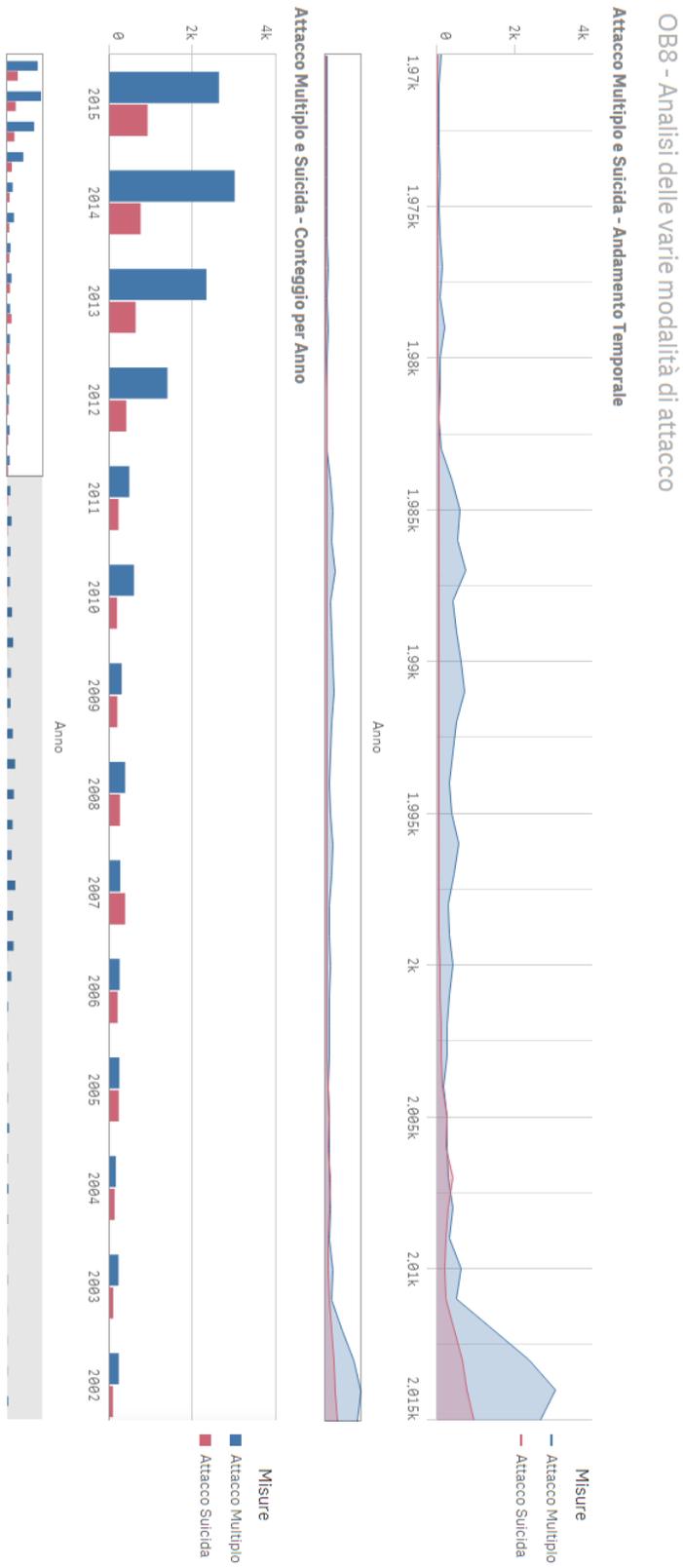


Figura 3.9: Grafico lineare e grafico a barre per OB8

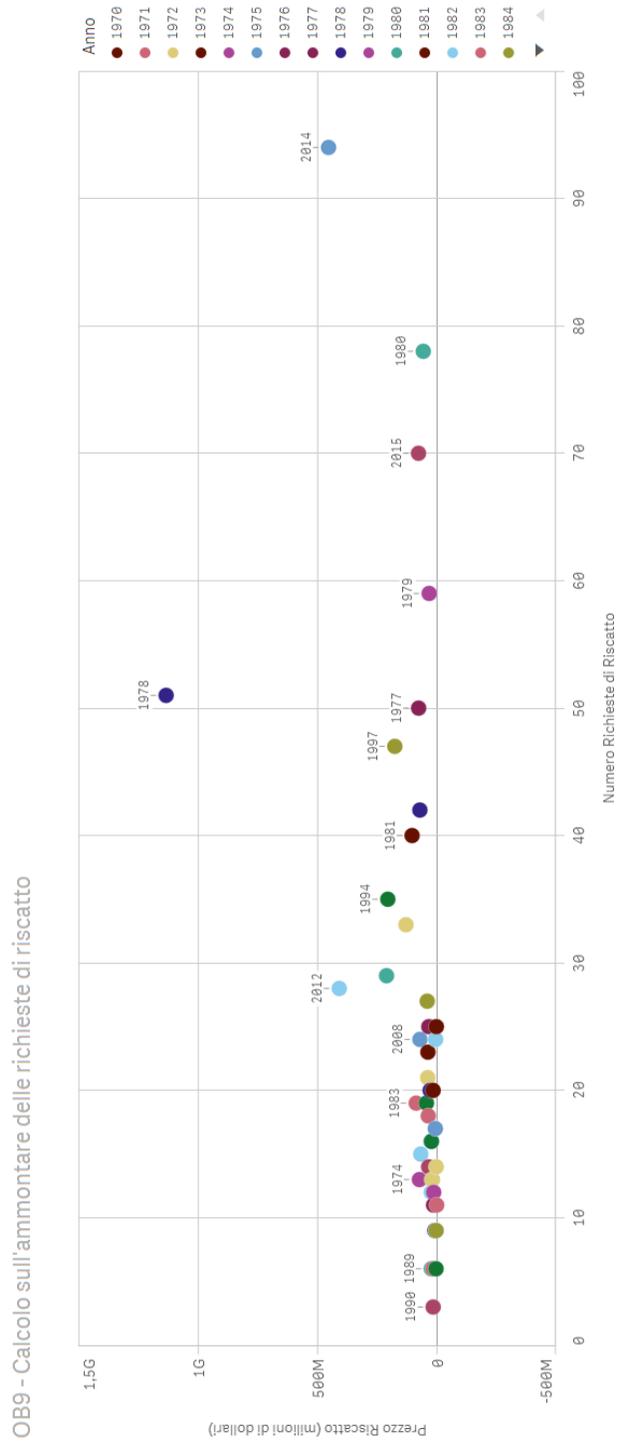


Figura 3.10: Grafico a dispersione per OB9

- sull'asse orizzontale troviamo la misura *Numero di Richieste di Riscatto*, che esprime tutti gli attacchi terroristici di un determinato anno che hanno previsto una richiesta di riscatto;
- sull'asse verticale si trova la somma, espressa in milioni di dollari, di tutti i riscatti richiesti in un determinato anno.

L'anno che ha visto un numero di richieste di riscatto maggiore è stato il 2014, mentre quello che ha avuto l'ammontare di richieste di riscatto più alto è stato il 1978.

3.2.10 OB10 - Stima del numero di vittime

L'ultimo degli obiettivi prefissati, OB10, è rappresentato coi grafici di Figura 3.11.

Nel grafico a barre viene effettuato un conteggio, per organizzazione terroristica, su due misure:

1. la prima colonna indica l'ammontare dei morti causati durante tutti gli attacchi perpetrati da una determinata organizzazione terroristica;
2. la seconda colonna indica l'ammontare dei terroristi morti durante tutti gli attacchi perpetrati da una determinata organizzazione terroristica.

Il grafico a torta invece esprime una sorta di qualità di ogni organizzazione terroristica, effettuando un rapporto tra le due misure descritte precedentemente. Ovviamente, tanto più è alta la percentuale, tanto più è alto il dislivello, tanto maggiore sarà dunque l'affidabilità di quella organizzazione. Entrando nel dettaglio espresso dal grafico a torta, l'organizzazione terroristica più affidabile, quella con perdite minore rapportate al numero di morti causati, è la *Nicaraguan Democratic Force*.

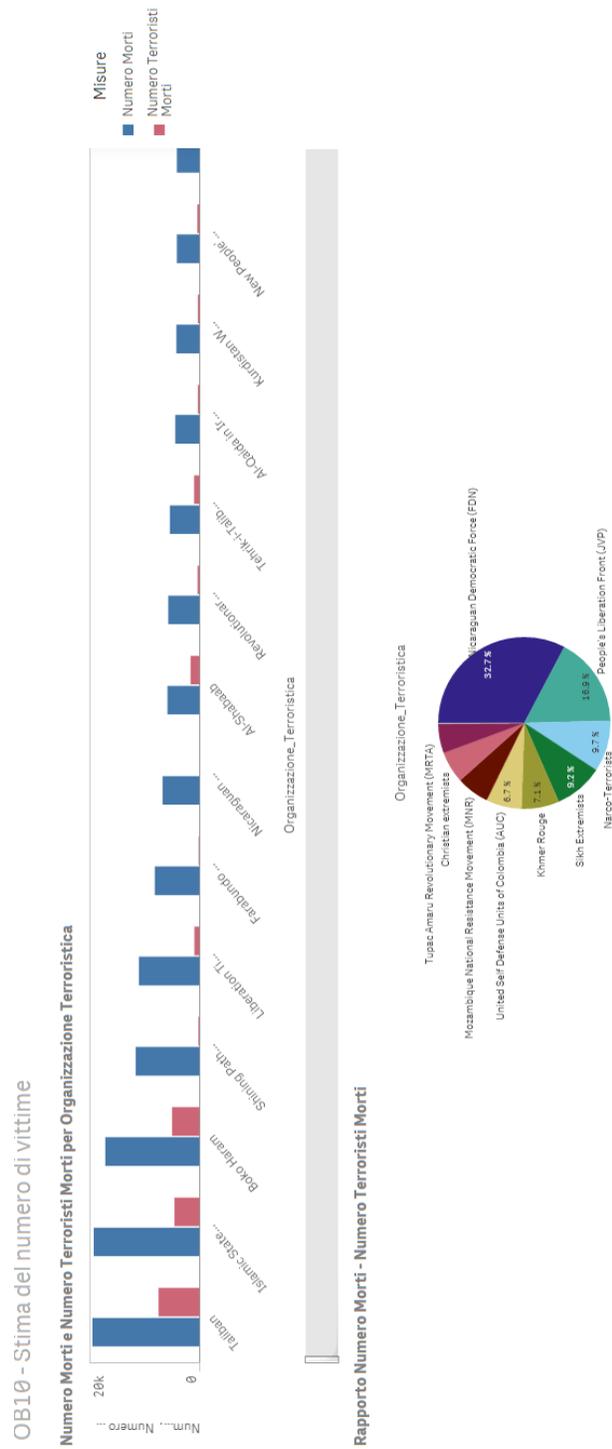


Figura 3.11: Grafico a barre e grafico a torta per OB10

Capitolo 4

Tableau e il GTD

Nel presente capitolo il nostro team di lavoro ha espresso gli obiettivi prefissati avvalendosi di un altro programma per la rappresentazione e l'analisi dei dati: *Tableau*.

4.1 Tableau

Tableau è uno dei software più potenti tra quelli di data visualization; è, inoltre, il principale competitor di QlikSense, l'altro tool visto nel dettaglio nel capitolo precedente.

Tableau mette anch'esso a disposizione dell'utente una vasta gamma di grafici per la rappresentazione di dati, aggiungendo rispetto a QlikSense la possibilità di una visualizzazione tramite delle mappe geografiche. C'è da dire che, rispetto al suo competitor, Tableau risulta essere meno intuitivo, magari anche meno dettagliato e gradevole a livello estetico, ma decisamente più potente in termini di possibilità di analisi e manipolazione dati.

Inoltre, l'approccio che si ha coi grafici è diverso rispetto a quello di QlikSense; infatti, se in quest'ultimo ottenevamo direttamente una dashboard complessiva con tutti i grafici già elaborati, in Tableau ad ogni grafico corrisponde un worksheet, e solo dall'unione di più worksheet legati semanticamente tra di loro nasce poi la dashboard.

4.2 Gli obiettivi del GTD in Tableau

Analizziamo singolarmente gli obiettivi che il nostro team di lavoro si è prefissato, andando questa volta a rappresentarli tramite il tool *Tableau*.

È opportuno segnalare come i risultati ottenuti tramite Tableau corrispondano esattamente a quelli ricavati con QlikSense.

4.2.1 OB1 - Analisi del numero di attacchi terroristici subiti per zona geografica

La Figura 4.1 rappresenta il primo obiettivo del nostro team di lavoro analizzato tramite Tableau. Rispetto a OB1 del capitolo su QlikSense, si possono notare subito sostanziali differenze.

Innanzitutto, la rappresentazione del numero di attacchi per zona geografica può essere effettuata non solo con grafici a barre (come si può notare nella parte di destra della Figura 4.1), ma addirittura tramite delle mappe di calore opportunamente tarate con misure e dimensioni. Le misure adottate per il conteggio degli attacchi terroristici subiti per zona geografica sono le stesse di QlikSense, aggiungendovi a queste, ovviamente, le misure di *latitudine* e *longitudine*, le quali riescono ad associare poi un risultato della query di conteggio con una ben determinata zona geografica della mappa.

Anche in questo caso è possibile applicare un drill-down semplicemente cliccando su un continente/nazione/regione/città di interesse, scendendo nel dettaglio fino ad un conteggio di attacchi terroristici subiti per città. Nella Figura 4.2 viene mostrato il numero di attacchi conteggiati in Calabria.

4.2.2 OB2 - Cambiamento delle tipologie di attacco nel corso del tempo

La Figura 4.3 mette in evidenza, tramite un grafico a barre, il cambiamento delle tipologie di attacco nel corso del tempo. I risultati ottenuti e le conseguenti considerazioni sono identiche a quelle fatte per QlikSense, anche a causa del fatto che i risultati ottenuti sono ovviamente i medesimi.

L'unica cosa da segnalare è la differenza grafica tra i due tool di analisi dati, diverse ma ugualmente vevole e chiare.

4.2.3 OB3 - Studio delle caratteristiche delle organizzazioni terroristiche

Questo paragrafo mette in luce i risultati ottenuti per OB3 utilizzando il tool Tableau. I risultati ottenuti conteggiando il numero di attacchi per organizzazione terroristica, visualizzabili in Figura 4.4, e quelli in cui si rapportano quest'ultimi al conteggio degli attacchi riusciti, sempre per organizzazione terroristica, mostrati in Figura 4.5, sono i medesimi di quelli ottenuti in QlikSense.

4.2.4 OB4 - Calcolo dei danni subiti per zona geografica

In questa analisi il nostro team di lavoro si è concentrato sui danni subiti per zona geografica, ma, rispetto a QlikSense, ha analizzato l'OB4 anche da altri punti di vista.

Nella figura 4.6, oltre alla mappa di calore che riesce a quantificare i danni per zona geografica semplicemente collocando il puntatore sulla zona interessata, è presente anche un grafico a torta che quantifica l'ammontare dei danni in base

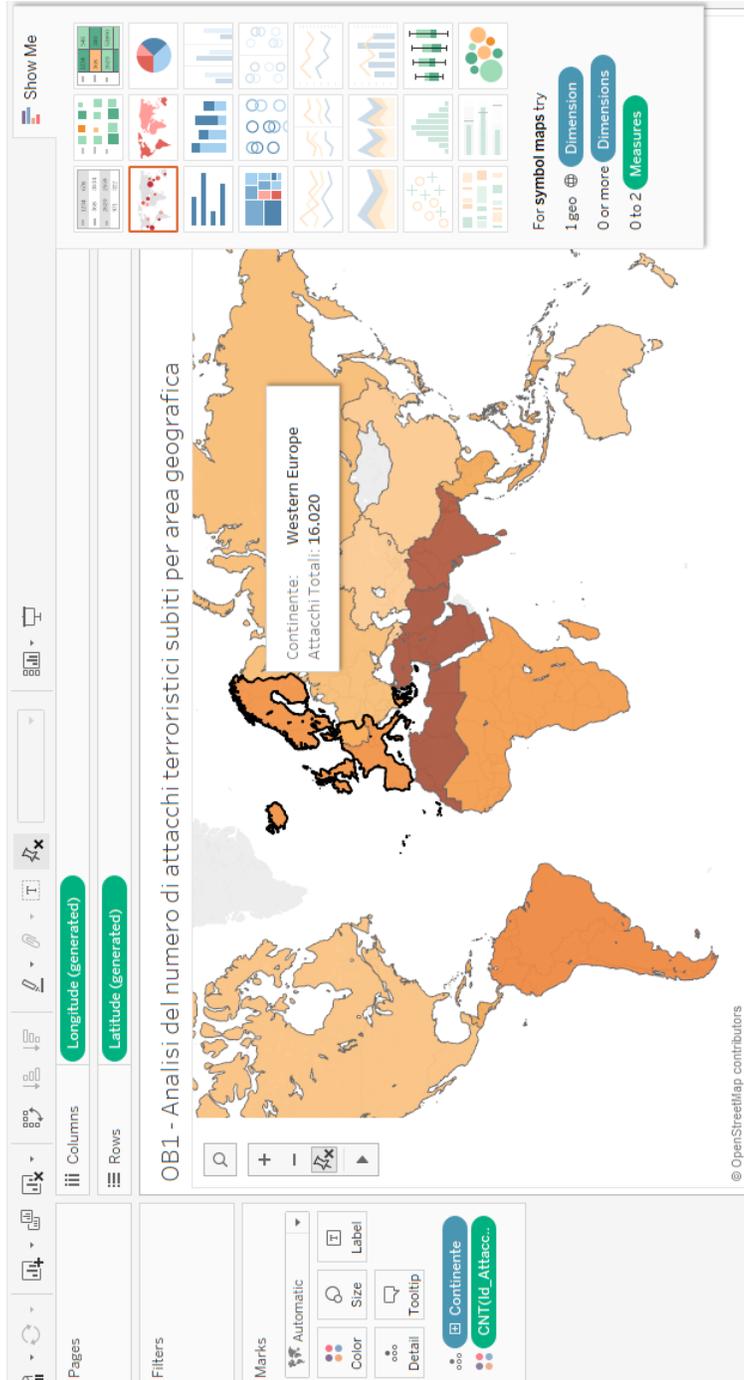


Figura 4.1: Mappa di calore (granularità per continente) per OB1

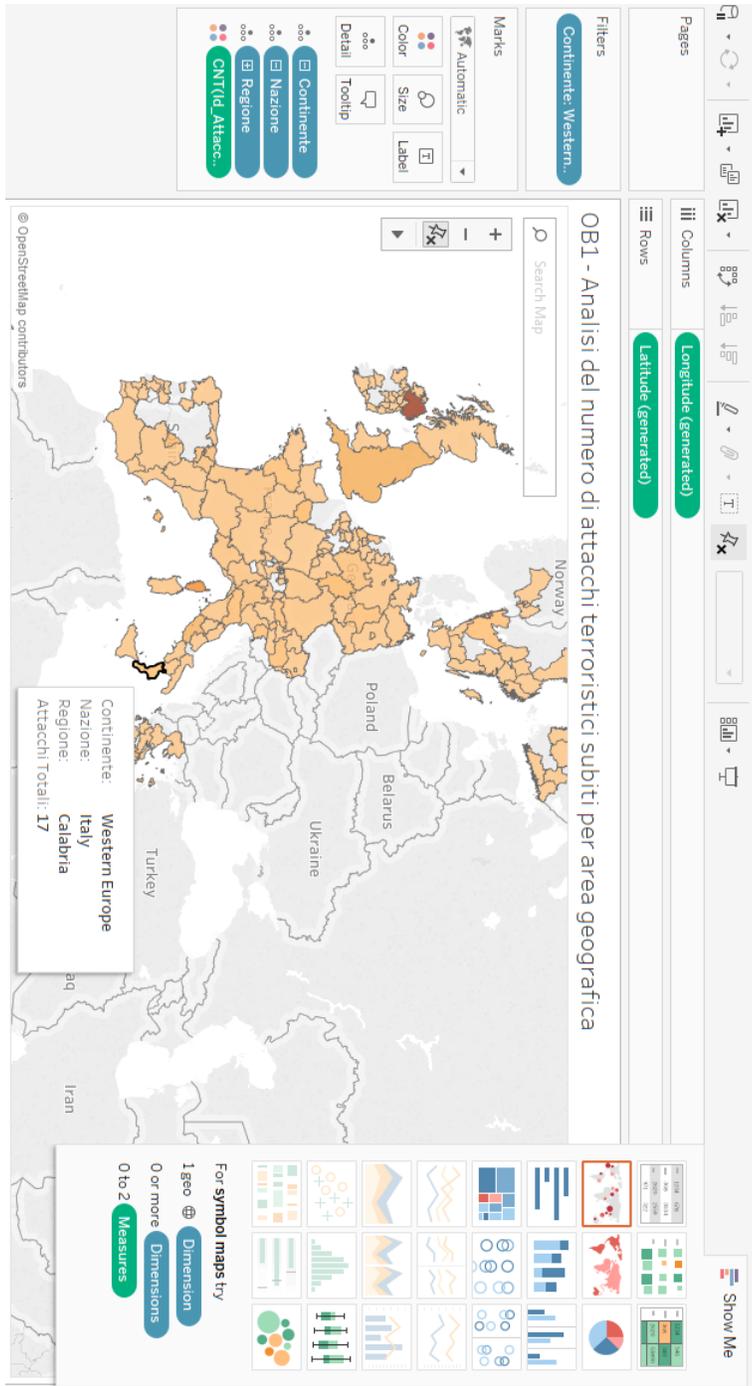


Figura 4.2: Mappa di calore (granularità per regione) per OB1

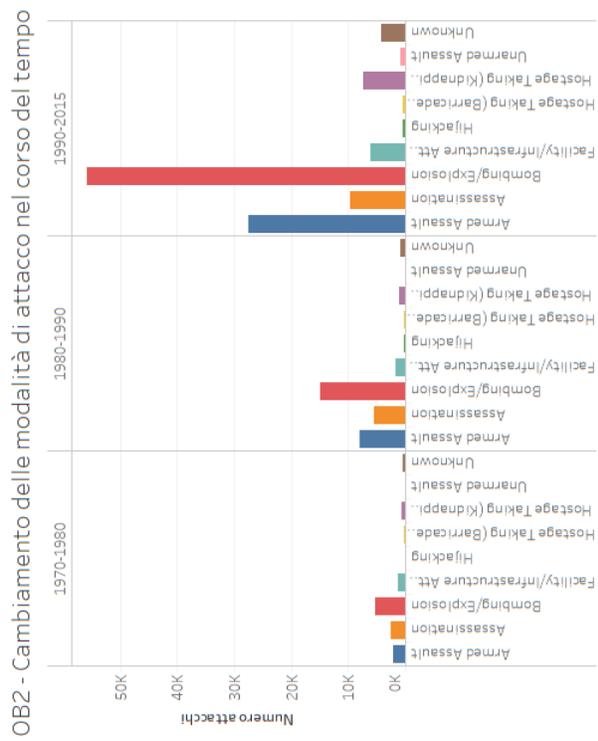


Figura 4.3: Grafico a barre per OB2

OB3 - Studio delle caratteristiche delle organizzazioni terroristiche (grafico quantitativo)

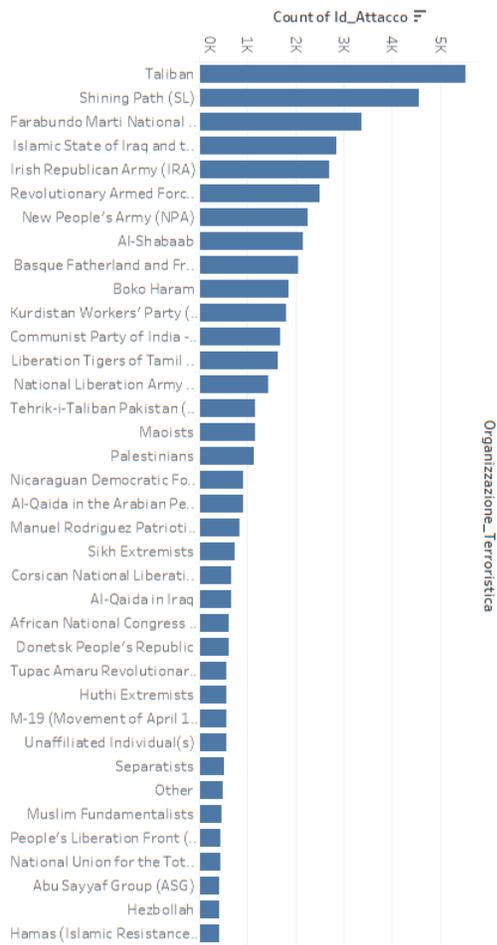


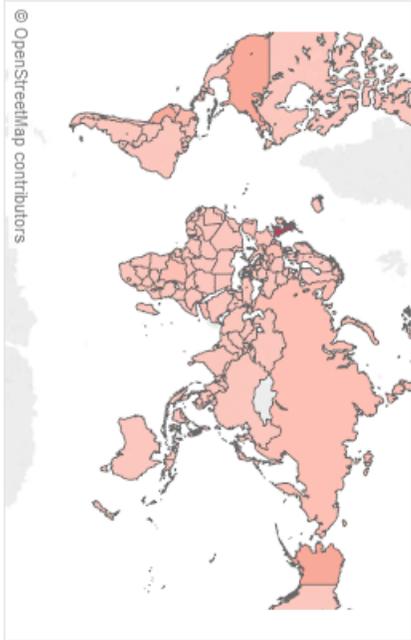
Figura 4.4: Grafico a barre quantitativo per OB3

OB3 - Studio delle caratteristiche delle organizzazioni terroristiche (grafico qualitativo)



Figura 4.5: Grafico a barre qualitativo per OB3

Danni subiti per area geografica



Quantità di danni (in migliaia di dollari)

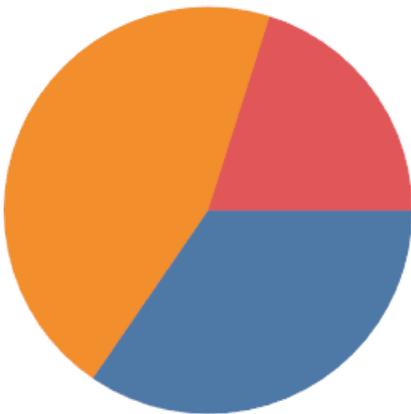


Figura 4.6: Mappa di calore e grafico a torta per OB4

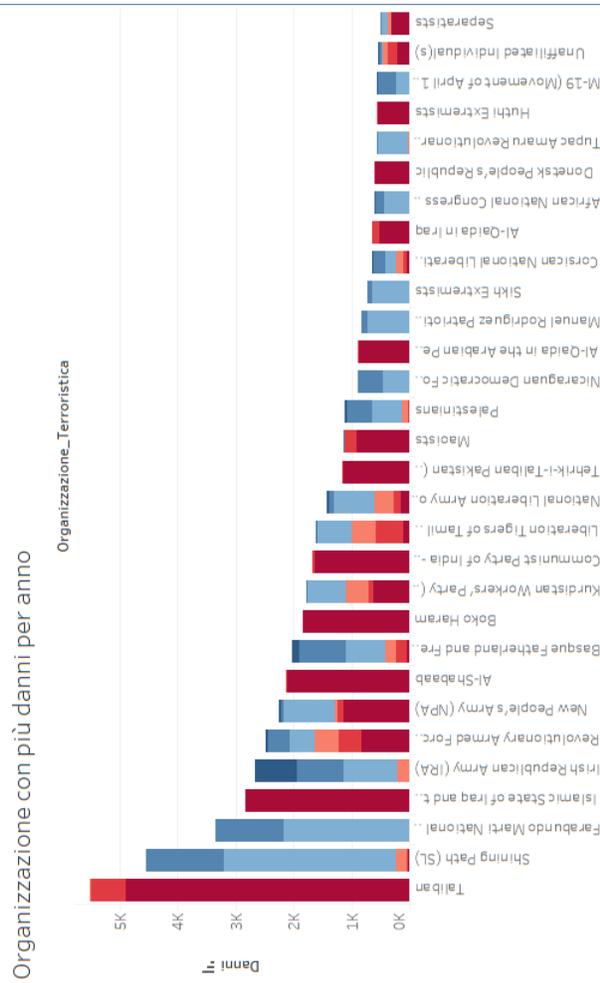


Figura 4.7: Grafico a barre per OB4

alla dimensione dell'attacco terroristico, classificando il tutto in tre spicchi che rappresentano, in ordine di grandezza, attacchi terroristici di tipo *Major*, *Catastrophic* e *Minor*.

Inoltre, in Figura 4.7 è stato stilato un grafico a barre che rappresenta l'ammontare di danni per ogni anno da parte di ogni organizzazione terroristica.

4.2.5 OB5 - Analisi statistica del numero di attacchi per data

La Figura 4.8 esprime l'obiettivo OB5 prefissato dal nostro team di lavoro. I risultati ottenuti, ordinati in senso decrescente, sono gli stessi di quelli ottenuti per QlikSense. L'unica differenza è rappresentata dal fatto che non si hanno grafici a barre, ma delle semplici raffigurazioni in cui ad una determinata data viene affiancato il numero di occorrenze per quest'ultima.

4.2.6 OB6 - Analisi sul tipo di armi utilizzate per gli attacchi terroristici

La Figura 4.9 rappresenta, tramite un grafico a barre, le armi maggiormente utilizzate da ogni organizzazione terroristica per perpetrare i propri attentati. In ogni colonna sono rappresentate le varie tipologie di arma, caratterizzata ognuna da un colore differente.

4.2.7 OB7 - Andamento e cambiamento temporale delle cause scatenanti degli attacchi terroristici

La Figura 4.10 è quella relativa ad OB7. Essa mette in mostra l'andamento e il cambiamento temporale delle cause che hanno scatenato gli attacchi terroristici. I risultati ottenuti sono i medesimi di quelli visti per QlikSense, solo che in questo caso si è preferito inserire tutti gli andamenti in un unico grafico.

4.2.8 OB8 - Analisi delle varie modalità di attacco

Anche in questo caso, come mostrato in Figura 4.11, il nostro team di lavoro ha preferito inserire l'andamento temporale delle misure considerate, *AttaccoMultiplo* e *AttaccoSuicida*, in un unico grafico lineare, viceversa di quanto fatto in QlikSense.

4.2.9 OB9 - Calcolo sull'ammontare delle richieste di riscatto

La Figura 4.12 rappresenta tramite un grafico lineare l'ammontare delle richieste di riscatto catalogate per anno. Tableau in questo caso, al contrario di QlikSense, non ci ha dato la possibilità di inserire un grafico a dispersione, a ragion per cui è stato necessario virare su uno lineare.

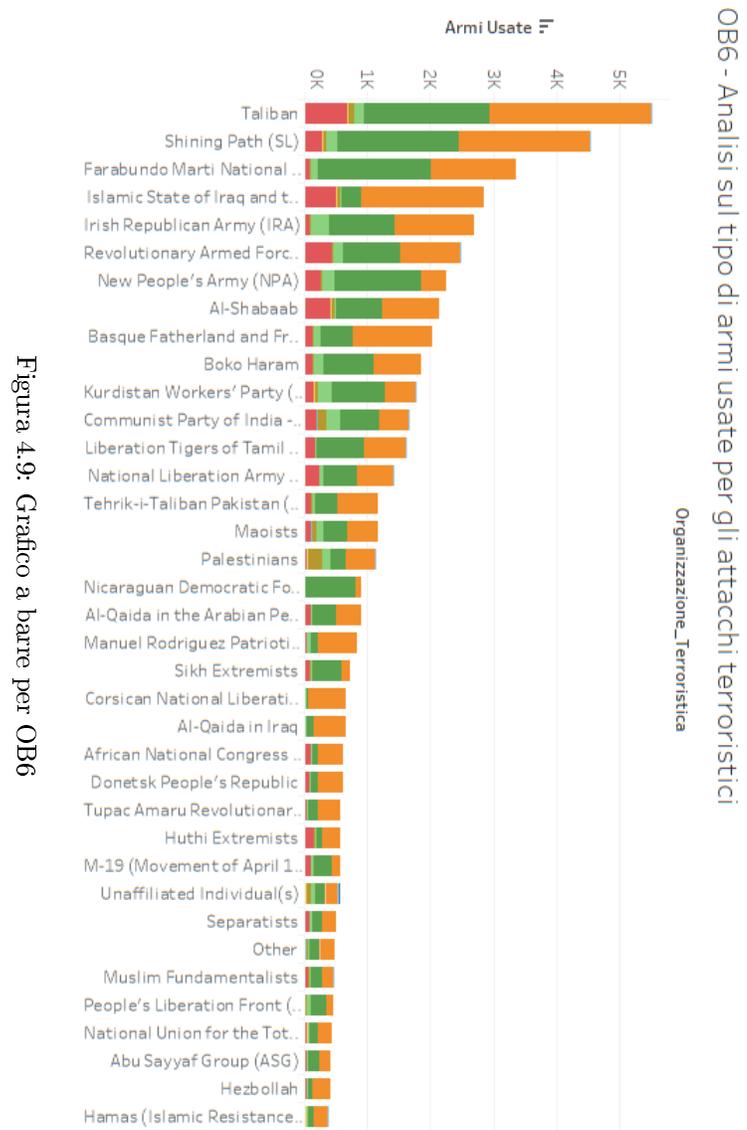


Figura 4.9: Grafico a barre per OB6

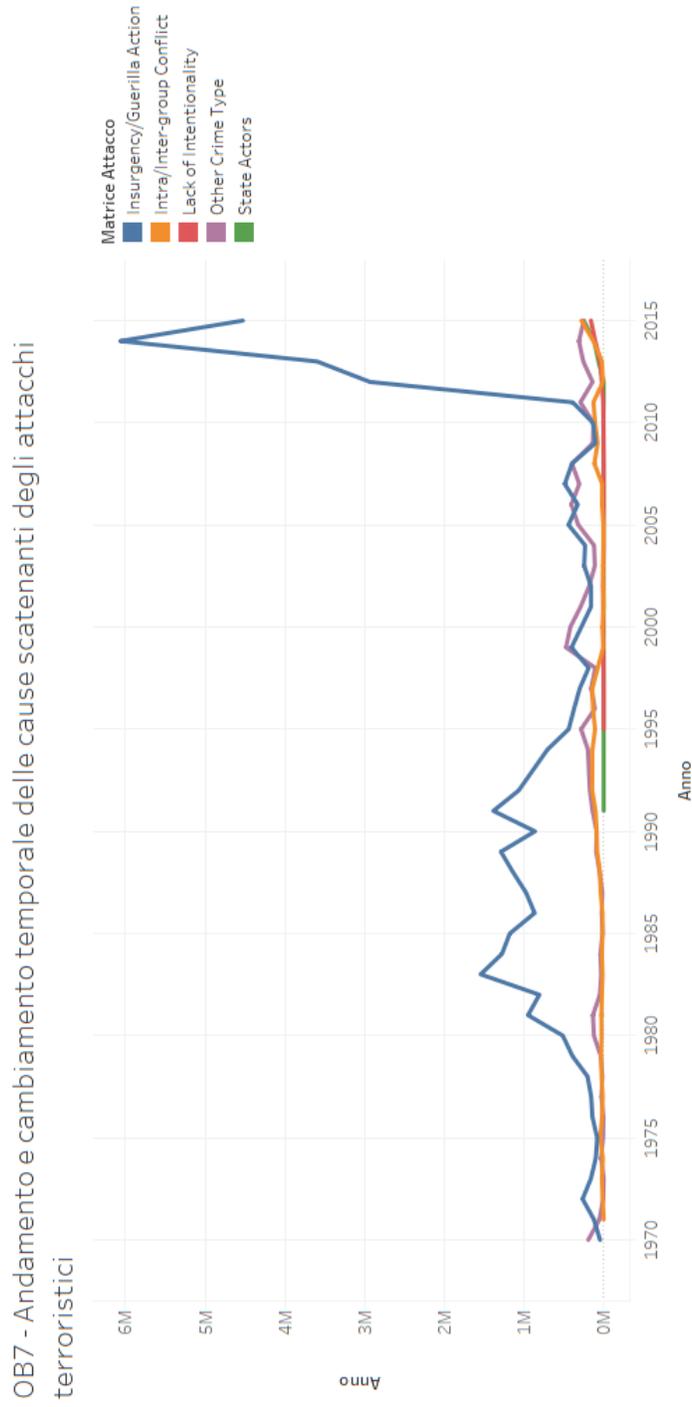


Figura 4.10: Grafico lineare per OB7

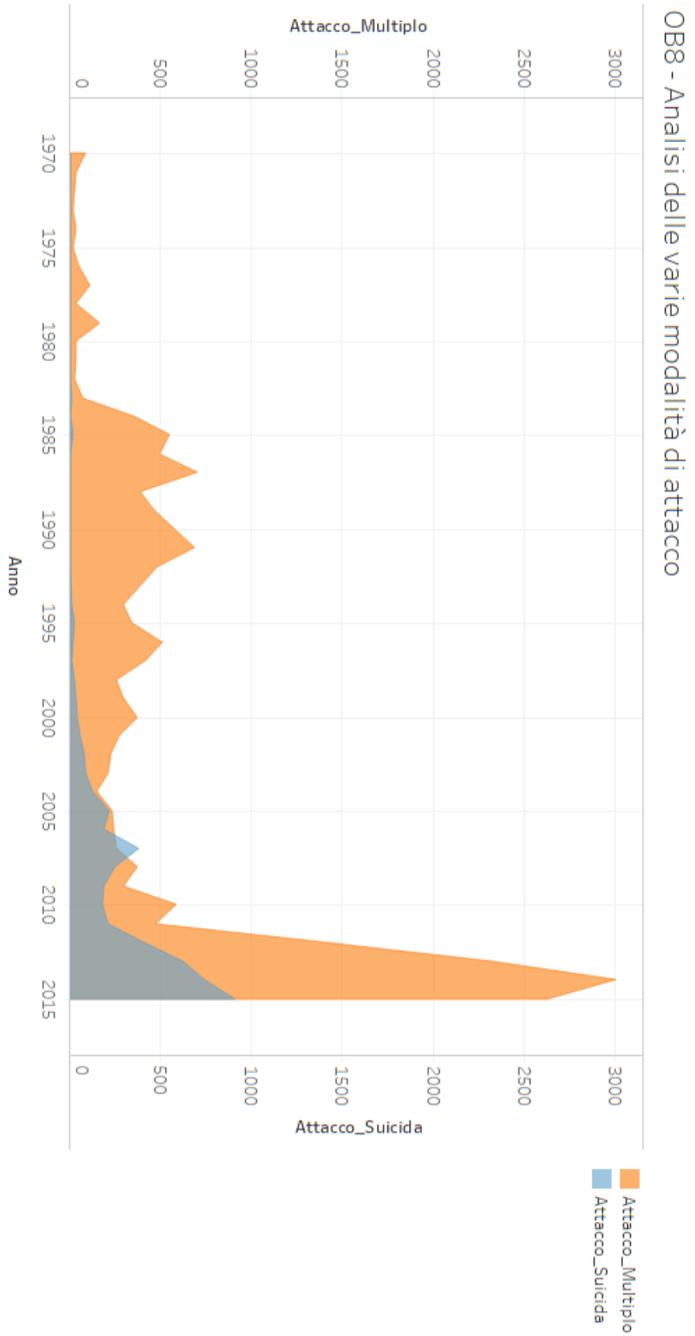


Figura 4.11: Grafico lineare per OB8

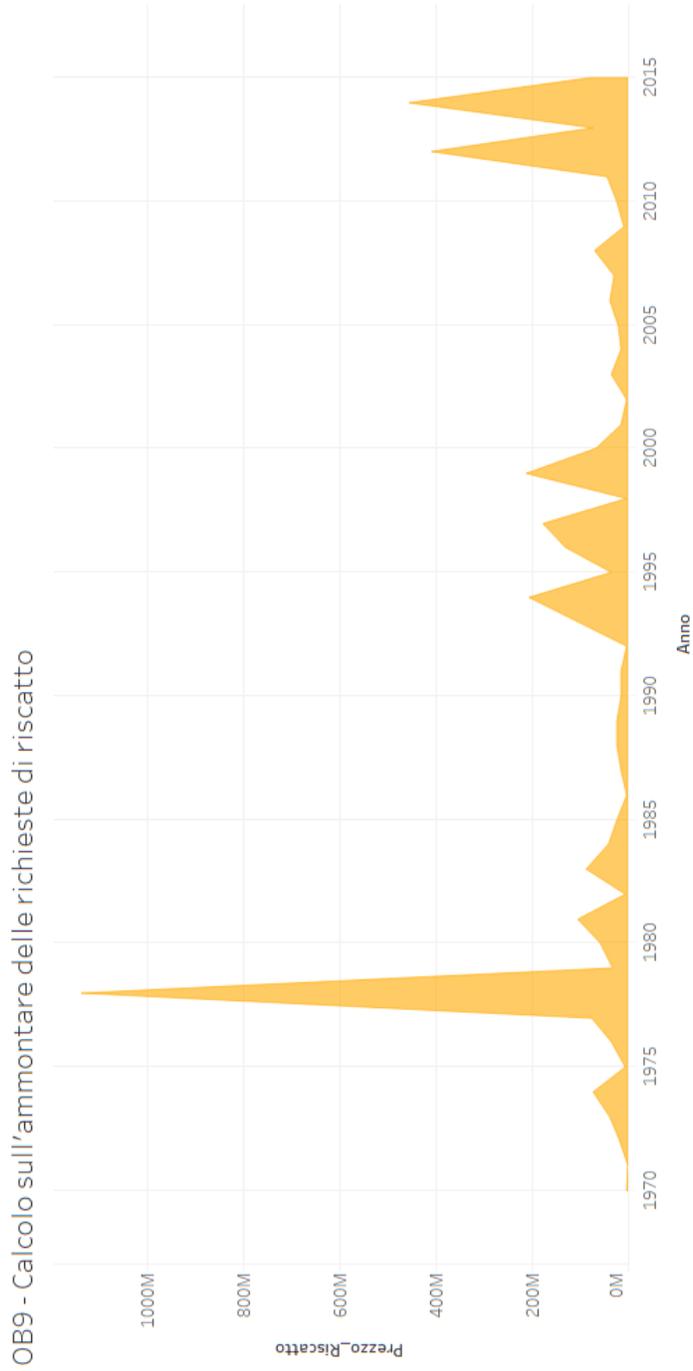


Figura 4.12: Grafico lineare per OB9

4.2.10 OB10 - Stima del numero di vittime

In Tableau si è preferito, rispetto ai grafici a barre di QlikSense, inserire un grafico a dispersione per stimare il numero di vittime civili e numero di vittime attentatori raggruppate per organizzazione terroristica. I risultati sono mostrati in Figura 4.13.

4.2.11 OB11 - Previsione sull'andamento futuro del numero di attacchi terroristici

Tableau, a differenza di QlikSense, consente di utilizzare dei particolari modelli di previsione statistica come *trendline* e *forecast*. In particolare, *forecast* consente di fare delle stime future prendendo in input una misura e una dimensione. Nel nostro caso abbiamo deciso di provare questa funzione scegliendo come misura il numero di attacchi per anno, che rappresenta nient'altro che OB1. Il risultato ottenuto, ovviamente, non può essere considerato molto attendibile perché privo di numerosi fattori che potrebbero influire sul risultato finale. Dunque, con i soli strumenti a nostra disposizione, ovvero i dati grezzi del GTD, abbiamo ottenuto da questa analisi che il numero di attacchi continuerà a crescere in maniera consistente nei prossimi anni, come si può notare dalla Figura 4.14.

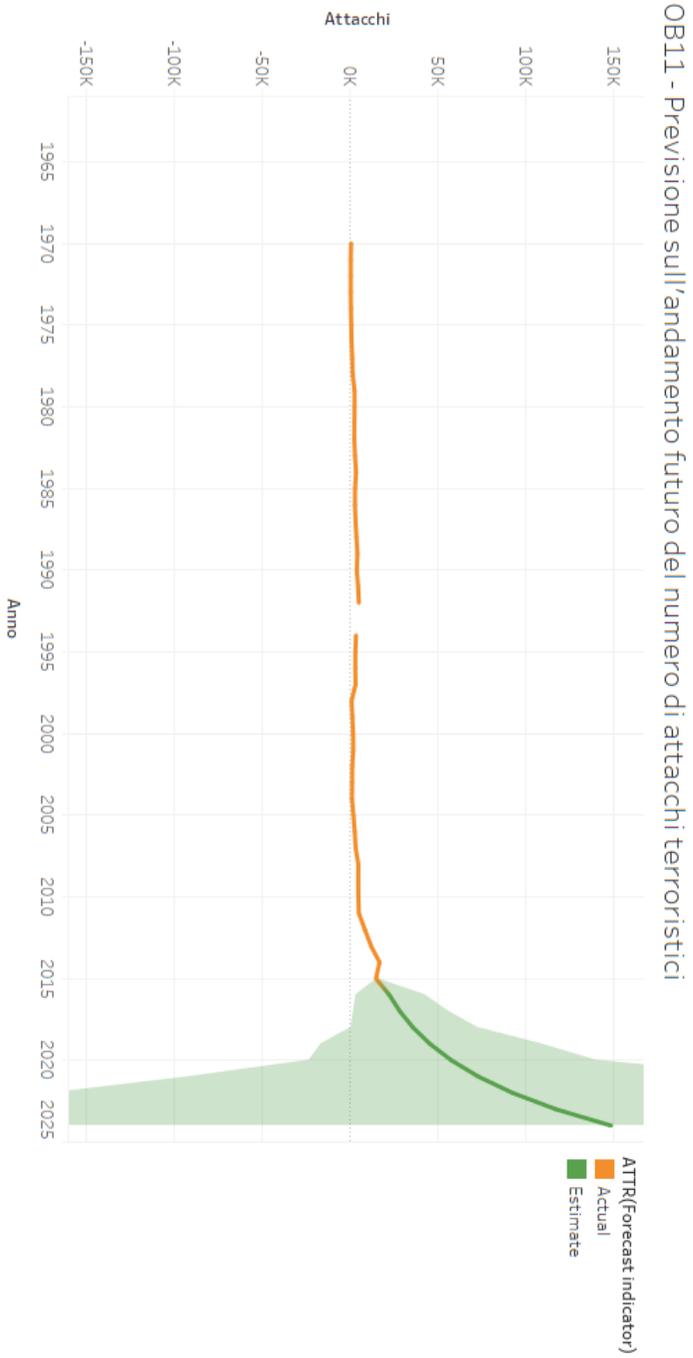


Figura 4.14: Grafico lineare per OB11

Capitolo 5

MongoDB e il GTD

L'ultimo tool che il nostro team di lavoro ha utilizzato per trattare l'analisi, la manipolazione e la rappresentazione dei dati presenti all'interno del Global Terrorism Database è stato *MongoDB*.

5.1 MongoDB

MongoDB è il leader tra i database NoSql. Esso conferisce all'utente una vastità di opzioni tramite le quali l'analizzatore dei dati può semplificare ma anche ampliare la sua esperienza.

La maggiore rapidità in termini computazionali rispetto ai database Sql, la possibilità di gestire grandi quantità di documenti, agendo su di essi con le funzioni di *aggregazione* e *MapReduce*, la possibilità di interfacciarsi con vari linguaggi di programmazione (Java su tutti), sono solo alcuni esempi delle grandi potenzialità che ci offre MongoDB.

Per lavorare con MongoDB è stato necessario creare un server in locale su cui caricare i nostri dati del GTD e, inoltre, installare lato client il software *MongoChef* tramite la cui *Intellishell* è stato possibile analizzare i dati e strutturare le query secondo il linguaggio di programmazione proprio di MongoDB.

5.2 Gli obiettivi del GTD in MongoDB

Analizziamo singolarmente gli obiettivi che il nostro team di lavoro si è prefissato, andando questa volta a rappresentarli tramite il tool *MongoDB*.

È opportuno segnalare come i risultati ottenuti tramite MongoDB corrispondano esattamente a quelli ricavati con QlikSense e Tableau.

5.2.1 OB1 - Analisi del numero di attacchi terroristici subiti per zona geografica

Il primo tra gli obiettivi che il nostro team di lavoro si è prefissato è, come già visto nei capitoli precedenti, il conteggio del numero di attacchi terroristici subiti per zona geografica.

Nel caso di MongoDB ci si è resi conto che l'operazione di drill-down non è realizzabile, a ragion per cui sono state necessarie delle opportune divisioni dell'obiettivo principale in vari sottobiettivi. In questo caso, sono stati creati tre script ad hoc, ognuno per ogni suddivisione geografica: uno a livello continentale, uno a suddivisione per nazione e uno, invece, a raggruppamento per città.

La Figura 5.1 mostra, nella parte alta, lo script relativo all'interrogazione, alla query vera e propria, mentre la parte in basso rappresenta la visualizzazione dei risultati ottenuti. Soffermiamoci un attimo su quest'ultimo aspetto, ovvero sulla visualizzazione del risultato ottenuto.

MongoDB mette a disposizione tre tipi di visualizzazione dei dati ottenuti:

- *tree view* o visualizzazione ad albero (Figura 5.2);
- *table view* o visualizzazione a tabella (Figura 5.3);
- *JSON view* o visualizzazione in formato (Figura 5.1).

Il nostro team di lavoro ha comunemente scelto in maniera del tutto democratica che la sua visualizzazione preferita è quella tabellare, a ragion per cui le successive figure verranno illustrate con questa tipologia di rappresentazione.

Dopo queste necessarie delucidazioni sui tipi di visualizzazione, spostiamoci alla parte relativa a quella dello script, prendendo in considerazione l'immagine di Figura 5.3.

In tutti gli script, essendo gli obiettivi proposti pressochè delle mere analisi statistiche, ci si è avvalsi unicamente della funzione *aggregate*. Essa consente di fare una selezione all'interno del database e, successivamente, ordinare e manipolare i dati in base alle necessità dell'utente.

Nella fattispecie, la funzione *aggregate* è incentrata sull'utilizzo di altre tre funzioni dedicate:

1. *match* consente di andare a selezionare, in base alla condizione esplicitata dall'utente, un ben determinato elenco di tuple;
2. *group* consente di raggruppare i risultati matchati in base ad un determinato attributo;
3. *sort* indica il tipo di ordinamento, in cui al valore 1 corrisponde un ordinamento crescente, mentre a -1 un ordinamento decrescente.

```
1 //OB1 - Analisi del numero di attacchi terroristici subiti per continente
2
3 db.gdt.aggregate(
4 [
5   {$group: {_id: "$Continente", Numero_Attacchi_Subiti: {$sum: 1}}}
6 ]
7 )
8
```

Aggregate 50 Documents 1 to 12 JSON View

```
1 {
2   "_id" : "Southeast Asia",
3   "Numero_Attacchi_Subiti" : NumberInt(10360)
4 }
5 {
6   "_id" : "Central America & Caribbean",
7   "Numero_Attacchi_Subiti" : NumberInt(10337)
8 }
9 {
10  "_id" : "Western Europe",
11  "Numero_Attacchi_Subiti" : NumberInt(16020)
12 }
13 {
14  "_id" : "North America",
15  "Numero_Attacchi_Subiti" : NumberInt(16020)
16 }
```

Figura 5.1: Script e risultati, in visualizzazione JSON, ottenuti per OB1 - Raggruppamento per Continente

The screenshot displays the MongoDB Studio interface. The top pane shows a JavaScript aggregate script for analyzing terrorist attacks by country. The bottom pane shows the results in a tree view, where each country is a key and its corresponding data is a document containing an array of attack details.

```

1 //OB1 - Analisi del numero di attacchi terroristici subiti per nazione
2
3 db.gdt.aggregate(
4 [
5   {$group: {_id: "$Nazione", Numero_Attacchi_Subiti: {$sum: 1}}},
6   {$sort: {_id: 1}}
7 ]
8 )
9
10
11

```

The results pane shows a tree view with 10 documents. Each document has a key representing a country and a value representing an array of attack details. The countries listed are Afghanistan, Albania, Algeria, Andorra, Angola, Antigua and Barbuda, Argentina, Armenia, Australia, and Austria.

Key	Value	Type
(1) {_id: Afghanistan}	{ 2 fields }	Document
(2) {_id: Albania}	{ 2 fields }	Document
(3) {_id: Algeria}	{ 2 fields }	Document
(4) {_id: Andorra}	{ 2 fields }	Document
(5) {_id: Angola}	{ 2 fields }	Document
(6) {_id: Antigua and Barbuda}	{ 2 fields }	Document
(7) {_id: Argentina}	{ 2 fields }	Document
(8) {_id: Armenia}	{ 2 fields }	Document
(9) {_id: Australia}	{ 2 fields }	Document
(10) {_id: Austria}	{ 2 fields }	Document

Figura 5.2: Script e risultati, in visualizzazione ad albero, ottenuti per OB1 - Raggruppamento per Nazione

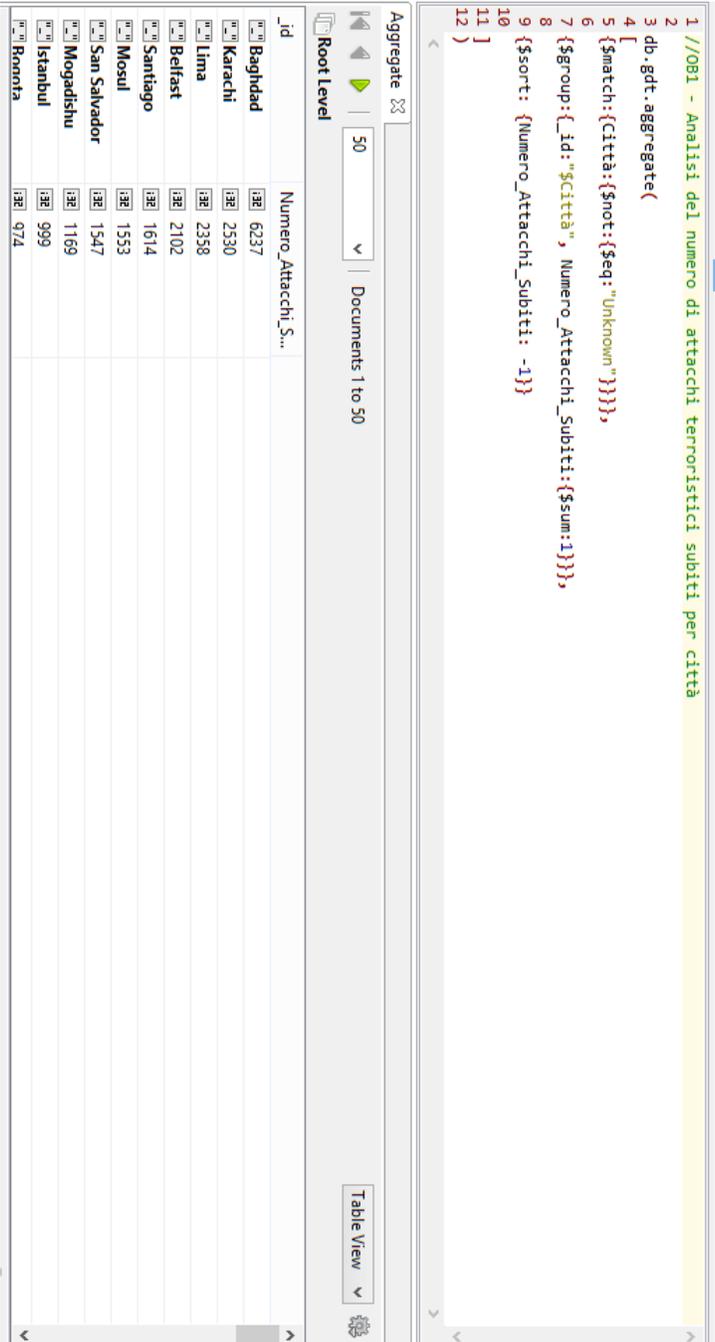


Figura 5.3: Script e risultati, in visualizzazione a tabella, ottenuti per OBI - Raggruppamento per Città

Nel nostro caso di Figura 5.3, il *match* va a prelevare dal database tutte quelle tuple per le quali l'attributo *Città* non è uguale ad *Unknown*. La funzione *group* raggruppa i risultati ottenuti basandosi sull'attributo *Città* ed effettua una somma (*sum: 1* equivale ad un *count* vero e proprio) sul numero di attacchi terroristici subiti, appunto, da ognuna città. Infine, l'ordinamento viene fatto in senso decrescente, andando ad impostare con valore -1 la funzione *sort*.

5.2.2 OB2 - Cambiamento delle tipologie di attacco nel corso del tempo

La Figura 5.4 mette in mostra una parte dei risultati ottenuti per OB2. Nella fattispecie, in questo caso sono state conteggiate le tipologie di attacco per gli anni maggiori del 1990 e minori del 2016. Sono state poi raggruppate per tipologia e ordinate in senso decrescente. Gli script per gli intervalli temporali 1970-1980 e 1981-1990 sono i medesimi con l'unica differenza della condizione posta nella funzione di matching.

5.2.3 OB3 - Studio delle caratteristiche delle organizzazioni terroristiche

La Figura 5.5 illustra i risultati ottenuti per OB5. In questo caso, a differenza dei casi precedenti, sono state create due variabili e ad ognuna delle quali è stata associata una somma. *NumeroAttacchi* indica il numero di attacchi perpetrati da ogni organizzazione terroristica, mentre *NumeroAttacchiRiusciti* rappresenta tutti quelli andati a segno.

I risultati sono stati ordinati in senso decrescente.

Inoltre, in questo caso non è stato necessario utilizzare la funzione *match* perchè non è stato necessario andare ad effettuare un filtraggio sulle tuple.

5.2.4 OB4 - Calcolo dei danni subiti per zona geografica

Anche nel caso di OB4 è stato necessario effettuare una suddivisione dell'obiettivo in varie parti, per l'impossibilità di applicare dei drill-down agli script. A ragion per cui viene mostrato in Figura 5.6 solo il calcolo dei danni subiti per città, tenendo in considerazione che gli script in cui vengono espressi i danni per continente e nazione sono i medesimi, con l'unico cambiamento dell'attributo di raggruppamento.

5.2.5 OB5 - Analisi statistica del numero di attacchi per data

La Figura 5.7 mostra l'analisi statistica effettuata sul numero di attacchi effettuati per data, in questo caso prendendo in considerazione i giorni del mese. Le analisi mensile e annuale sono le medesime, con l'unico cambiamento dell'attributo di raggruppamento.

```

1 //OB2 - Cambiamento delle tipologie di attacco nel corso del tempo (1990-2015)
2
3 db.gdt.aggregate(
4 [
5   {$match: {Anno: {$gt:1990, $lt:2016}}},
6   {$group: {_id:"Tipo_Attacco", Numero_Attestati: {$sum: 1}}},
7   {$sort: {Numero_Attestati: -1}}
8 ]
9 )
10

```

_id	Numero_Attestati
Bombing/Explo...	55781
Armed Assault	27553
Assassination	9612
Hostage Taking...	7293
Facility/Infrastr...	6001
Unknown	4206
Unarmed Assault	752
Hijacking	375
Hostage Taking...	316

Figura 5.4: Script e risultati ottenuti per OB2

The screenshot displays the MongoDB Studio interface. The top pane shows an aggregate query script for analyzing terrorist organizations. The bottom pane shows the results in table view, sorted by the number of attacks in descending order.

```

1 //083 - Studio delle caratteristiche delle organizzazioni terroristiche
2
3 db.gdt.aggregate(
4 [
5   {$group: {_id: "$Organizzazione_Terroristica", Numero_Attacchi: {$sum: 1},
6     Numero_Attacchi_Riusciti: {$sum: "$Attacco_Riuscito"}},
7   {$sort: {Numero_Attacchi: -1}}
8 ]
9 )

```

_id	Numero_Attacchi	Numero_Attacchi_...
Unknown	71922	63472
Taliban	5502	5126
Shining Path (SI)	4548	4332
Farabundo Mar...	3351	3317
Islamic State of...	2833	2599
Irish Republica...	2670	2280
Revolutionary ...	2474	2289
New People's A...	2241	2028
Al-Shabaab	2127	1980
Basnie Fatherl...	2024	1721

Figura 5.5: Script e risultati ottenuti per OB3

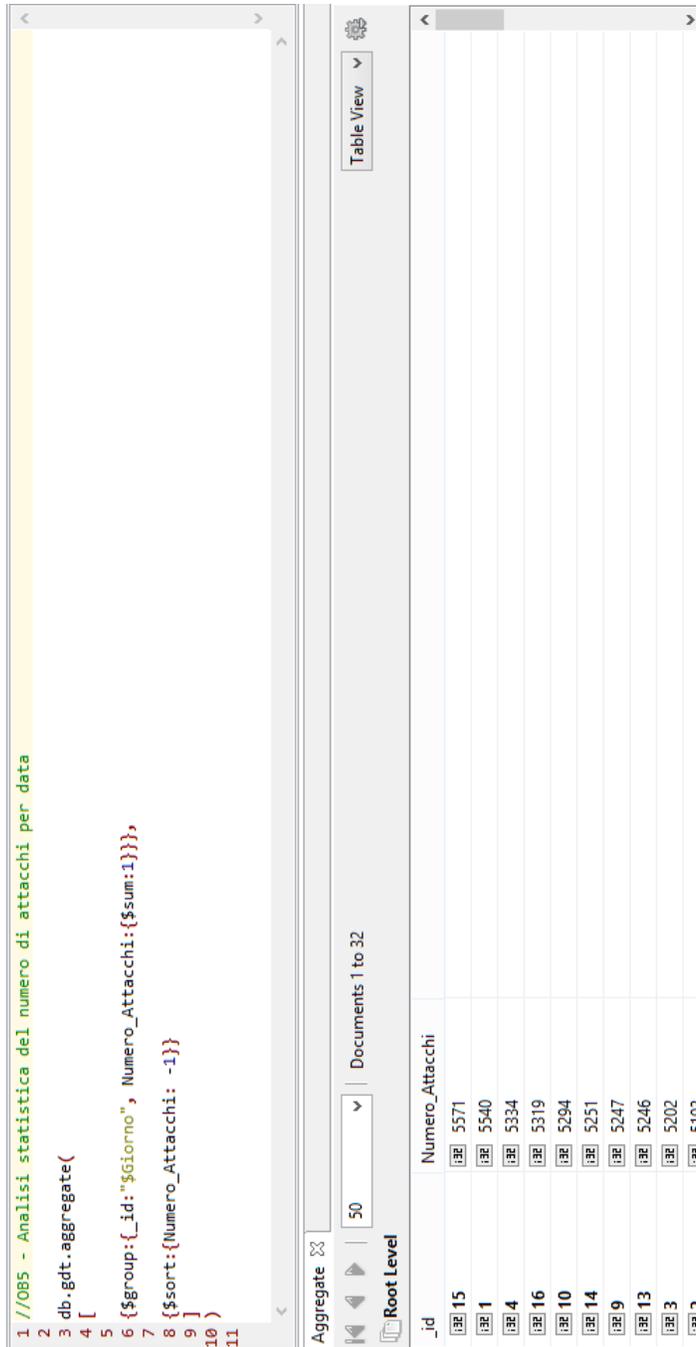
```

1 //OB4 - Calcolo dei danni subiti per città
2
3 db.gdt.aggregate([
4   {$match: {Danni_Proprietà: {$gt: 0}}},
5   {$group: {_id: "$Città", Danni_Totale: {$sum: "$Danni_Proprietà"}}},
6   {$sort: {Danni_Totale: -1}}
7 ])
8

```

_id	Danni_Totale
Unknown	5013
Baghdad	2498
Lima	2060
Santiago	1476
San Salvador	1212
Karachi	1098
Belfast	857
Athens	814
Istanbul	788
Rangoon	698

Figura 5.6: Script e risultati ottenuti per OB4



The screenshot displays the MongoDB Shell interface. The top pane shows the following aggregate command:

```
1 //085 - Analisi statistica del numero di attacchi per data
2
3 db.gdt.aggregate(
4 [
5   {$group: {_id: "$Giorno", Numero_Attacchi: {$sum:1}}},
6   {$sort: {Numero_Attacchi: -1}}
7 ]
8 )
9
10
11
```

The bottom pane shows the results in table view. The table has two columns: `_id` and `Numero_Attacchi`. The results are sorted in descending order of `Numero_Attacchi`.

_id	Numero_Attacchi
15	5571
1	5540
4	5334
16	5319
10	5294
14	5251
9	5247
13	5246
3	5202
2	5193

Figura 5.7: Script e risultati ottenuti per OB5

5.2.6 OB6 - Analisi sul tipo di armi utilizzate per gli attacchi terroristici

La Figura 5.8 mostra una classifica delle armi maggiormente impiegate per gli attacchi terroristici. La classifica è ordinata in senso decrescente e sono stati esclusi dal conteggio, ovviamente, tutte le tuple di cui non era specificata l'arma utilizzata.

5.2.7 OB7 - Andamento e cambiamento temporale delle cause scatenanti degli attacchi terroristici

La Figura 5.9 mette in risalto l'andamento temporale, tra il 1970 e 1980, delle cause scatenanti degli attacchi terroristici. Per analizzarne il cambiamento temporale, invece, è necessario abbinare a questa analisi gli altri due script che contengono il medesimo studio per gli altri intervalli di tempo, contenuti nel software con cui è stato prodotto il presente lavoro e che per evitare inutili ripetizioni di codice non verranno proposti.

5.2.8 OB8 - Analisi delle varie modalità di attacco

La Figura 5.10 nella visualizzazione tabellare dei risultati ottenuti, mette in luce per la prima volta ben quattro colonne, che rappresentano rispettivamente:

- l'anno, ordinato in senso decrescente;
- il conteggio del numero di attacchi in quell'anno;
- il conteggio del numero di attacchi multipli avvenuti in quell'anno;
- il conteggio del numero di attacchi suicida avvenuti in quell'anno.

5.2.9 OB9 - Calcolo sull'ammontare delle richieste di riscatto

La Figura 5.11 illustra il raggiungimento dell'obiettivo OB9 tramite MongoDB, andando ad evidenziare i risultati ottenuti in tre tabelle, ordinate in senso decrescente partendo dall'anno in cui si è avuto il quantitativo (in migliaia di dollari) di richieste di riscatto più alto. A tale quantità viene affiancato, inoltre, il numero di richieste di riscatto avvenute in quel determinato anno.

5.2.10 OB10 - Stima del numero di vittime

La Figura 5.12 mette in evidenza il rapporto che c'è tra i morti causati da una organizzazione terroristica e i terroristi rimasti uccisi mentre perpetrava gli attentati.

The screenshot shows a MongoDB aggregation pipeline script in a code editor and its results in a table view.

Script:

```

1 //OB6 - Analisi sul tipo di armi utilizzate per gli attacchi terroristici
2
3 db.gdt.aggregate(
4 [
5   {$match:{Generere_Arma:{$not:{$eq:"Unknown"}}}},
6   {$group: {_id:"$Generere_Arma", Usata:{$sum:1}}},
7   {$sort: {Usata: -1}}
8 ]
9
10
11 ]
12 )

```

Results Table:

_id	Usata
Explosives/Bo...	79126
Firearms	51802
Incendiary	9812
Melee	3013
Chemical	231
Sabotage Equip...	123
Vehicle (not to ...	104
Other	92
Biological	35
Fake Weapons	33

Figura 5.8: Script e risultati ottenuti per OB6

```

1 //OB7 - Andamento e cambiamento temporale delle cause scatenanti degli attacchi terroristici (1970-1980)
2
3 db.gdt.aggregate(
4 [
5   {$match: {Anno: {$gt:1969, $lt:1981}}},
6   {$group: { "_id": "$Altra_Matrice", "Numero_Attenati": {$sum: 1}}},
7   {$sort: {Numero_Attenati: -1}}
8 ]
9 )
    
```

_id	Numero_Attenati
Unknown	11014
Insurgency/Gu...	1111
Other Crime Ty...	263
Intra/Inter-gro...	112

Figura 5.9: Script e risultati ottenuti per OB7

The screenshot displays the MongoDB Studio interface. The left pane shows an aggregation script for analyzing attack modalities. The right pane shows the results in table view, displaying aggregated data for each year from 2006 to 2015.

```

1 //088 - Analisi delle varie modalità di attacco
2
3 db.gdt.aggregate(
4 [
5   {$group: {_id: "$Anno", Attacchi_Totali: {$sum: 1}, Numero_Attacchi_Multipli: {$sum: "$Attacco_Multiplo"},
6     Numero_Attacchi_Suicida: {$sum: "$Attacco_Suicida"}},
7   {$sort: {Anno: -1}}
8 ]
9 )
10
11

```

_id	Attacchi_Totali	Numero_Attacchi_...	Numero_Attacchi_S...
2015	14806	2624	906
2014	16840	3003	739
2013	11990	2324	623
2012	8498	1385	402
2011	5067	468	208
2010	4821	582	175
2009	4721	290	182
2008	4787	371	242
2007	3241	253	373
2006	2751	238	187

Figura 5.10: Script e risultati ottenuti per OB8

The screenshot shows the MongoDB Aggregator interface. The top pane contains a JavaScript aggregation script:

```

1 //OB9 - Calcolo sull'ammontare delle richieste di riscatto
2
3 db.gdt.aggregate(
4 [
5   {$match:{Riscatto:{$gt:0}, Prezzo_Riscatto:{$gt:1}}},
6   {$group:{$_id: "$Anno", Numero_Richieste:{$sum:1}, Prezzo_Riscatti:{$sum:"$Prezzo_Riscatto"}},
7   {$sort:{$Prezzo_Riscatti: -1}}
8 ]
9 )
    
```

The bottom pane shows the results in table view, displaying 15 documents. The columns are `_id`, `Numero_Richieste`, and `Prezzo_Riscatti`.

_id	Numero_Richieste	Prezzo_Riscatti
1978	29	113609000
2014	61	454412089,8
2012	17	409391832,76
1999	11	211034800
1994	35	205684127
1997	45	176269877
1996	27	129417206
1981	15	103796000
1983	11	86565000
2015	40	76819130,5500...

Figura 5.11: Script e risultati ottenuti per OB9

```

1 //OB10 - Stima del numero di vittime
2
3 db.gdt.aggregate(
4 [
5   {$match: {Num_Morti:{$gt: 0}, Num_Terroristi_Morti:{$gt:0}},
6   {$group: {
7     _id: "$Organizzazione_Terroristica", Numero_Civili_Morti:{$sum: "$Num_Morti"},
8     Numero_Terroristi_Morti:{$sum: "$Num_Terroristi_Morti"}},
9   {$sort: {Numero_Civili_Morti: -1}}
10 ]
11 )
12 )

```

Aggregate 50 Documents 1 to 50

_id	Numero_Civili_Morti	Numero_Terroristi_...
Al-Shabaab	3114.5	1501.0
Al-Qaida in Iraq	2410.5	235.0
Al-Qaida in the ...	1939.0	653.0
Al-Nusra Front	1827.0	563.0
Liberation Tige...	1567	899
Islamic State of...	1216	49
Donetsk Peopl...	770	687
Huthi Extremists	630.0	424.0
Revolutionary ...	610	203
Ishkar-e-Ihan...	601.0	55

Figura 5.12: Script e risultati ottenuti per OB10

Capitolo 6

Conclusioni

Questo ultimo capitolo è dedicato ai commenti, alle conclusioni e alle considerazioni che il team ha espresso riguardo al progetto stesso, alla collaborazione per la realizzazione del presente lavoro e riguardo ai software utilizzati.

Innanzitutto, sviluppare il presente progetto è stato certamente un lavoro che ha richiesto molto tempo ed impegno, ma che è risultato essere sempre piacevole, stimolante, a tratti entusiasmante, perchè si è riusciti davvero ad entrare in una sorta di clima aziendale, sia in termini di progetto che di collaborazione tra i colleghi del team.

Entusiasmante oltremodo per il tema trattato, quello degli attacchi terroristici, un argomento molto attuale e che rappresenta una piaga per l'intera umanità. Analizzando i dati nel dettaglio, ci si è forzatamente addentrati in quella che è la storia, spesso dilaniata, di alcuni paesi, nella forma mentis delle organizzazioni terroristiche, nel reale impatto che questo problema ha su tantissime società ed etnie, e molto altro ancora.

Da un punto di vista prettamente umano, invece, sviluppare in gruppo il presente lavoro ha insegnato ad ognuno dei membri del team cosa voglia dire collaborare all'interno di un progetto. Durante lo sviluppo delle analisi ci sono stati molti confronti, magari anche accesi, ma che portavano comunque ad un punto di convergenza, ovvero alla scelta ottima per un determinato obiettivo prefissato.

Dal punto di vista della coesione del gruppo, invece, non ci sono stati particolari problemi. I compiti, le direttive, la cura dei dettagli, la stesura di un particolare documento o presentazione, sono state smistate e assegnate ai diversi membri senza particolari problemi, mettendo in risalto un vero senso di squadra.

Il progetto sul Global Terrorism Database ha, inoltre, inconsciamente consolidato i rapporti tra i membri del team.

L'approccio che si è avuto col tool *QlikSense* è stato subito entusiasmante: vedere la facilità e le grandi possibilità di rappresentazione dei dati ha destato subito l'interesse del team. *QlikSense* ha trovato riscontri molto positivi da parte di ognuno dei membri anche per le qualità grafiche, per la bellezza estetica con la quale è stato possibile esprimere gli obiettivi prefissati. Questo tool è stato davvero una bella scoperta.

Con *Tableau* l'esperienza avuta è iniziata in un modo più complicato rispetto a *QlikSense*. *Tableau* è meno intuitivo, meno user-friendly, ma analizzando le potenzialità riteniamo possa essere superiore di diverse spanne rispetto al suo competitor.

In sostanza, quello che si perde in grafica con *Tableau* lo si recupera in funzioni messe a disposizione dell'utente.

In definitiva, i due tool rappresentano dei veri e propri trade-off tra grafica e dettaglio di rappresentazione dei dati, e la scelta su cui andare a ricadere dipende dalle esigenze dell'utente analizzatore dei dati, andando a considerare soprattutto a che tipo di cliente sono diretti i grafici prodotti.

MongoDB è stato il software con il quale il team si è trovato inizialmente spiazzato. Questo status è stato provocato da un approccio di interrogazione dei database differente da quelli visti precedentemente durante il percorso universitario. L'impatto con il tipo documentale NoSql ha richiesto qualche tempo per la chiara comprensione del tipo di linguaggio, un tempo necessario per abituarsi ad un altro tipo di logica per quanto riguarda le interrogazioni del database.

Una volta entrati nel meccanismo, quando il linguaggio da utilizzare e i tipi di ragionamenti da fare sono stati chiari, gli obiettivi prefissati sono stati raggiunti con facilità, andando, inoltre, ad apprezzare le notevoli migliorie computazionali che questo genere di database e di interrogazioni hanno portato in questo settore dell'informatica.