Corentin MAPELLA & Glodie TANDU

**Projet EDD**

**Direction Générale de la Sécurité Intérieur (DGSI)**

Prévenir du terrorisme

2ème Année de Master MIAGE - 2018



# Information du projet :

Ce document comporte le cahier des charges du projet relatif à Prévenir du Térrorisme. Il inscrit les objectifs et les besoins attendus de l’application ainsi qu’une présentation des fonctionnalités, les spécifications et le détail des développements de celui-ci.



|  |  |
| --- | --- |
| **Nom du projet :** | **PREVENIR DU TERRORISME** |
| **Groupe :** | Corentin MAPELLA – Glodie TANDU |
| **Version du document :** | 1.0 |
| **Statut :** | Version Finale |

# Information groupe :

|  |  |
| --- | --- |
| **NOM** | **RÔLE** |
| Corentin MAPELLA | Rédacteur du cahier des charges  Analyse des données  Utilisateur de Power BI |
| Glodie TANDU | Approbateur du cahier des charges  Analyse des données  Utilisateur de Power BI |

# Révision du document :

|  |  |  |
| --- | --- | --- |
| **VERSION** | **NOM** | **DESCRIPTION** |
| 0.1 | gtandu – cmapella | Création du document |
| 0.2 | gtandu – cmapella | Rédaction de la présentation du sujet |
| 0.3 | gtandu – cmapella | Ajout du contexte et de l’objectif |
| 1.0 | gtandu – cmapella | Version Finale |

Table des matières

[I. Introduction 4](#_Toc515305128)

[1. Contexte 4](#_Toc515305129)

[2. Présentation du sujet 5](#_Toc515305130)

[3. Objectifs 5](#_Toc515305131)

[4. Données utilisées 6](#_Toc515305132)

[5. Indicateurs à mesurer 6](#_Toc515305133)

[II. Spécifications 7](#_Toc515305134)

[1. Modèle en flocon 7](#_Toc515305135)

[2. Technologies 8](#_Toc515305136)

[3. Mise en place du DWH 8](#_Toc515305137)

# Introduction

Dans le cadre de notre Master 2 MIAGE Apprentissage, un projet d'entrepôt de données doit être réalisé. Ce projet consiste à collecter et stocker des données pour en faire une analyse afin d’aider une entreprise à prendre des décisions : Nous parlerons **d’analyse décisionnelle**.

Nous allons utiliser un Data Warehouse comme outil d’aide à la décision. Pour rappel, le Data Warehouse, ou entrepôt de données, est une **base de données dédiée au stockage de l'ensemble des données** utilisées dans le cadre de la prise de décision et de l'analyse décisionnelle.

Le Data Warehouse est exclusivement réservé à cet usage. Il est alimenté en données depuis les bases de production grâce notamment aux outils d'ETL Extract Transform Load.

|  |  |
| --- | --- |
| **Étude de besoins spécifiques** | **Spécifications** |
| Contexte  Présentation du sujet  Objectifs  Données utilisées (Sources)  Indicateurs à mesurer | Modélisation  Dimensions (Axes)  Faits (Mesures) |

## Contexte

Dans une époque où le mot « **terrorisme** » est sur les lèvres de chacun et dans laquelle les craintes sont de plus en plus nombreuses, nous avons souhaité comprendre et analyser toutes les attaques référencées.

Tout le monde comprend ce mot mais personne ne sait l’expliquer :

Le terrorisme est l'emploi de **la terreur à des fins politiques, religieuses ou idéologiques.**

Les multiples définitions varient sur :

* L'usage de la **violence** (certaines comprennent des groupes n'utilisant pas la violence mais ayant un discours radical), les techniques utilisées, la nature du sujet (mettant à part le terrorisme d'État)
* L'usage de la **peur**
* Le niveau d'organisation
* L’**idéologie**
* Etc.

Dans nombre de définitions intervient aussi **le critère de la victime** du terrorisme (civile, désarmée, innocente, attaque contre un État démocratique qui aurait permis au terroriste de s'exprimer légalement, etc.).

Un grand nombre d'organisations politiques ou criminelles ont cependant recouru au terrorisme pour faire avancer leur cause ou en retirer des profits. Des partis de gauche comme de droite, des groupes nationalistes, religieux ou révolutionnaires, voire des États, ont commis des actes de terrorisme.

Une constante du terrorisme est **l'usage indiscriminé de la violence meurtrière à l'égard de civils** dans le but de promouvoir un groupe, une cause ou un individu, ou encore de pratiquer l'extorsion à large échelle (mafias, cartels de la drogue, etc.).

## Présentation du sujet

Qui sommes-nous ?

La direction générale de la sécurité intérieure est un service actif de la police nationale. Elle est chargée, sur l’ensemble du territoire de la République, de rechercher, de centraliser et d’exploiter le renseignement intéressant la sécurité nationale ou les intérêts fondamentaux de la Nation.

Ses principales missions sont :

* Le contre-espionnage
* La lutte contre le terrorisme et les extrémismes violents
* La protection du patrimoine économique
* Mission de police judiciaire spécialisée

Au service de la Direction Générale de la Sécurité Intérieure, nous souhaitons à travers ce projet pouvoir anticiper et prévenir les actes terroristes sur le territoire français. Nous avons pour cela, un ensemble de données répertoriant les actes terroristes commis depuis 1970 jusqu’à 2016 à travers le monde.

## Objectifs

Ce projet a donc pour finalité de :

* Anticiper et repérer les nouvelles formes d’attaques terroristes.
* Pouvoir mettre en place les mesures de sécurité nécessaire pour se prémunir de ces attaques.
* Obtenir des tableaux de bords et graphiques pertinent, qui seront présentées aux différents responsables qui ont un pouvoir de décision quant à la mise en œuvre des différentes moyens allouées et mesures de sécurité mise en place.

## Données utilisées

Afin de répondre aux besoins de la DGSI, les données ont été récupéré sur Kaggle.com, un site spécialisé dans la science des données et dans le partage de données. Le lien vers notre source de données est ci-dessous.

[**https://www.kaggle.com/ananthreddy/global-terrorism-database/data**](https://www.kaggle.com/ananthreddy/global-terrorism-database/data)

Les données sont fournies **en format .CSV**.

Elles sont composées de **plus de 170 350 lignes** et **plus de 120 colonnes**.

La taille du fichier est égale à **150 Mo**.

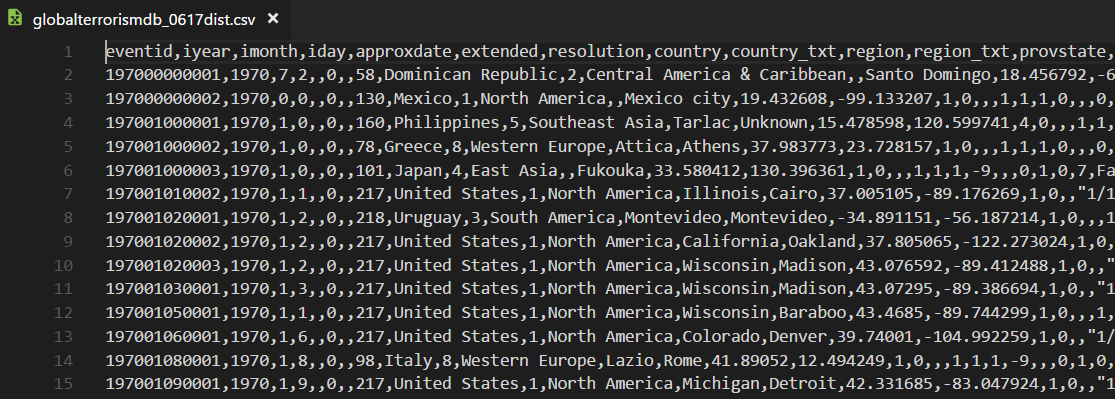
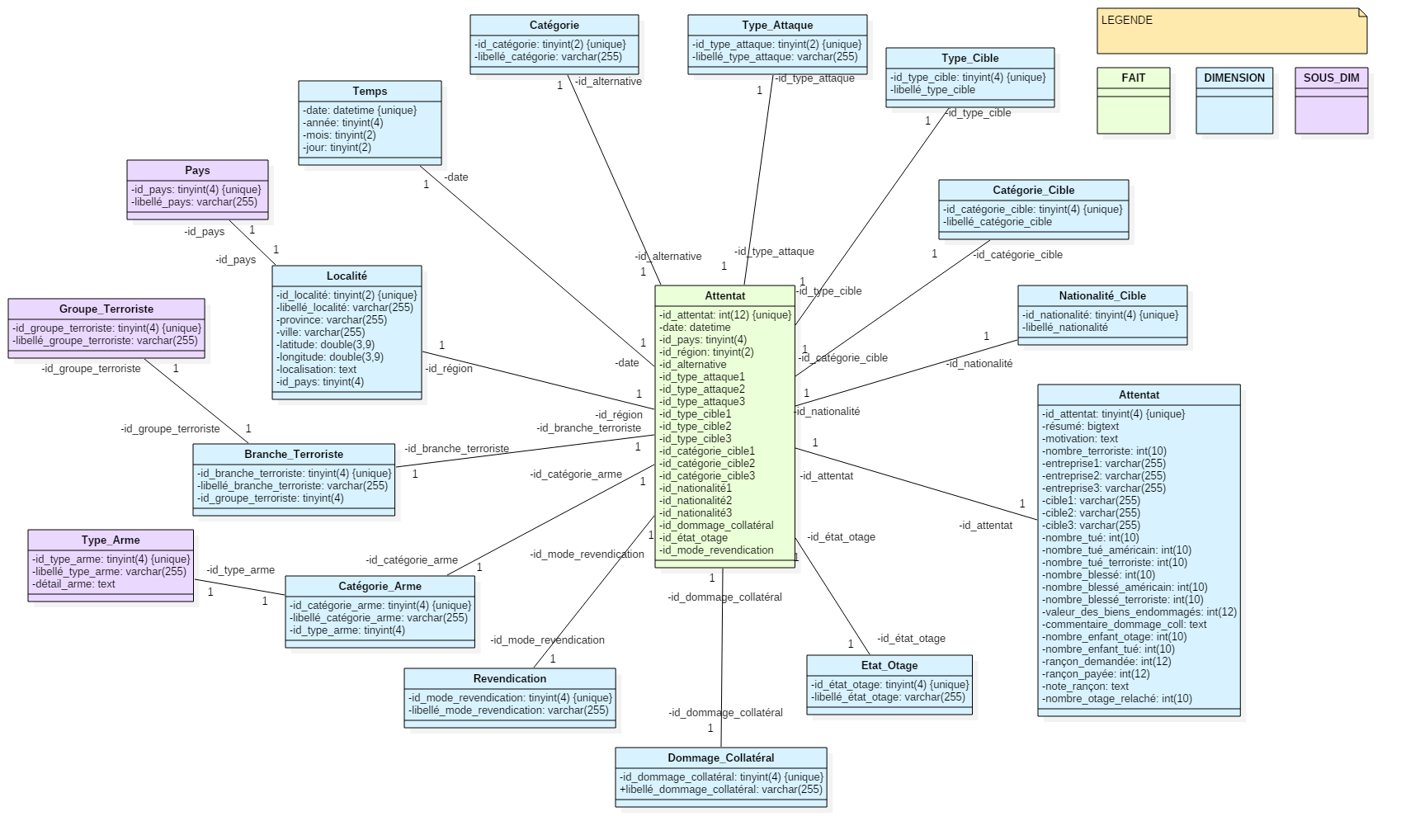


Figure 1 - Jeu de données pour le terrorisme

## Indicateurs à mesurer

# Spécifications

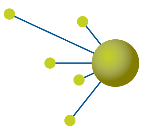
## Modèle en flocon



## Technologies



Pour mener à bien notre mission, nous avons choisi de nous appuyer sur **Power BI**. Cette solution est une **suite d’outils décisionnels** qui permet de **visualiser les données**, de produire des **tableaux de bord dynamique**, **créer des rapports** afin de prendre des décisions stratégiques.



Concernant **l’ETL (Extract Transform and Load),** nous avons opté pour **Talend** qui est l’une des références pour l’intégration de données.



Debian est un **système d'exploitation libre**. Un système d'exploitation est une suite de programmes de base et d’utilitaires permettant à un ordinateur de fonctionner. Il est fourni avec plus de 51000 paquets **dont Maria DB.**

MariaDB est un **système de gestion de base de données** édité sous **licence GPL**. Il s'agit **d'un fork** **communautaire de MySQL** : la gouvernance du projet est assurée par la fondation MariaDB9, et sa maintenance par la société Monty Program AB, créateur du projet10. Cette gouvernance confère au logiciel **l’assurance de rester libre**.

## Mise en place du DWH

La première étape fut de préparer les données, c’est-à-dire, rendre les données compréhensibles, **supprimer les informations non pertinentes** dans notre analyse et **obtenir des données dites « valides ».**

Pour cela, nous avons utilisé « Talend Data Preparation », qui est un outil de préparation des données permettant de **nettoyer les données** via un navigateur web. C’est un outil très puissant qui permet par exemple de **calculer le pourcentage de cellule vide**, **en erreur** et **valide** de chaque colonne.

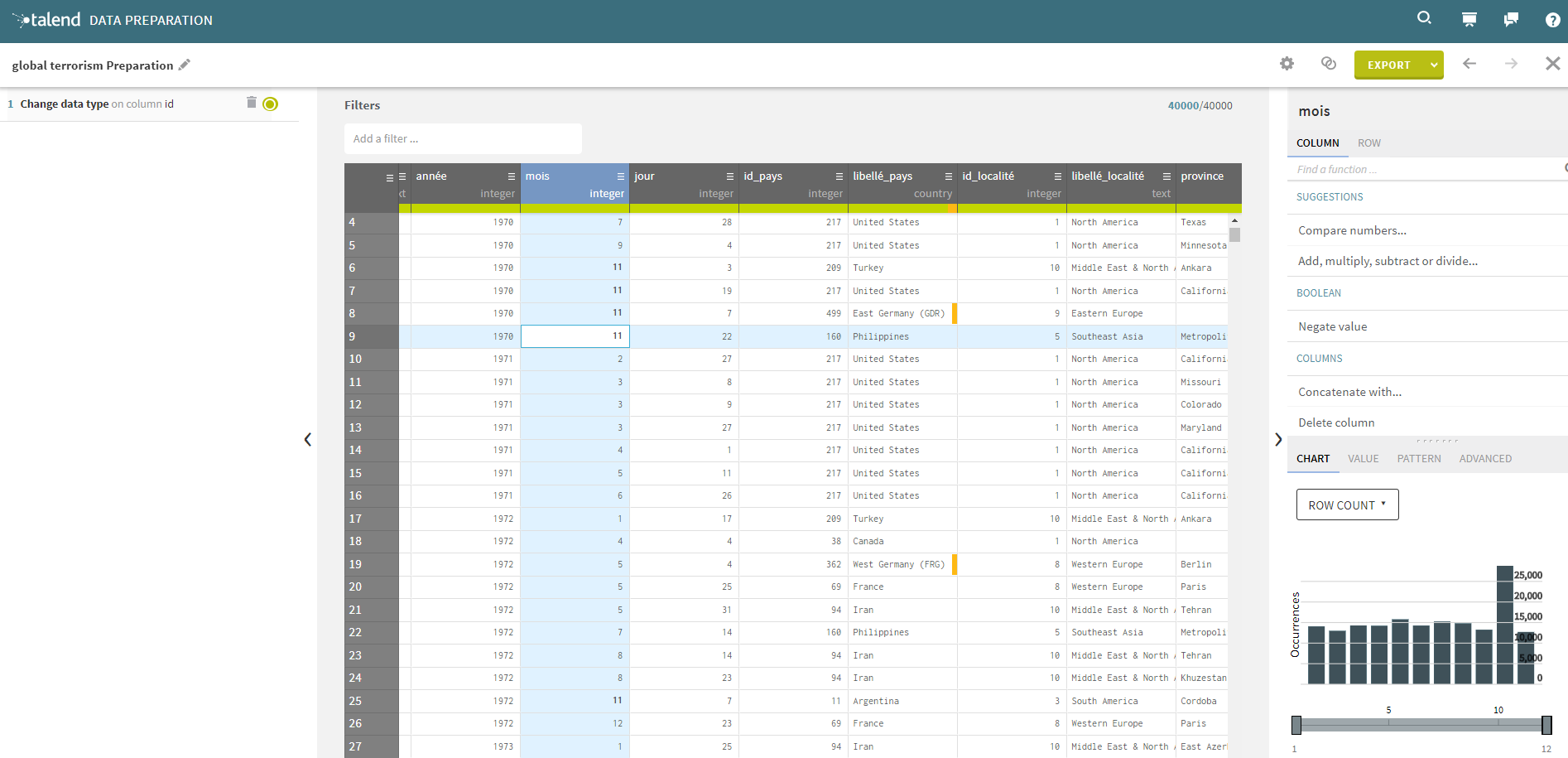


Figure 2 – Talend Data Preparation

La deuxième étape consiste à créer nos différentes dimensions conformément à notre modèle.

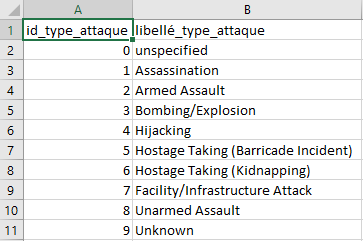


Figure 3 – Dimension type d’attaque

Puis la table de fait :

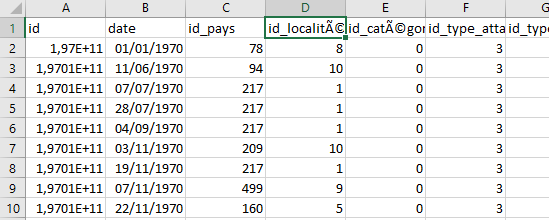


Figure 4 - Table de Fait "Attentat"

La troisième étape consiste à **intégrer toutes les données** triées dans la base Maria DB.

Pour ce faire, nous avons **importé les fichiers .CSV** que nous avions traité en amont et le logiciel s’est occupé de générer les tables.

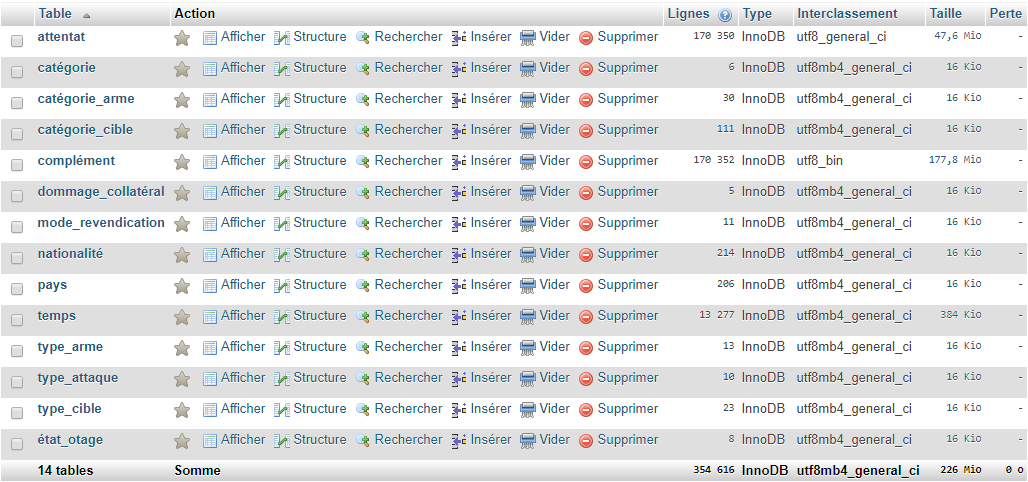


Figure 5 - Tables de la base de données

Cinquième étape

Connexion à notre base de données depuis Power BI. Il faut au préalable installer un connecteur MySQL sur son poste.

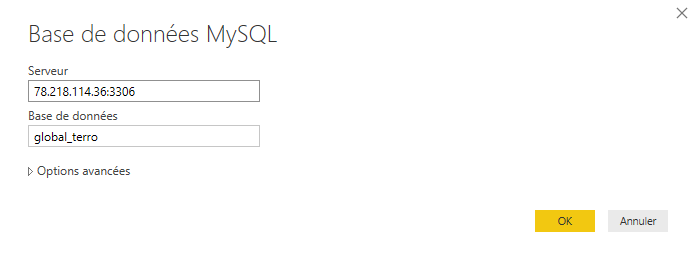


Figure 6 – Connexion base de données MySQL avec Power BI