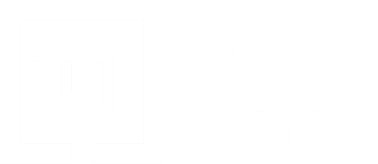
A black and blue sign with text

Description automatically generated

|  |  |
| --- | --- |
| Compte Rendu de Projet ADC  Violence Armée aux États-Unis (2013-2018) | |
| Encadré par : Élaborer par : | |
| Pr. CHEMLAL Yman | ERFOUDI Anass  MOUSDIK Ismail  DAIF Abdellah  EDDAOUDI Mohamed |



Contents

[I. Contexte du Projet 3](#_Toc184473269)

[II. Objectifs du Projet 4](#_Toc184473270)

[III. Description des Données 5](#_Toc184473271)

[1. Introduction et Choix du Jeu de Données 5](#_Toc184473272)

[Question 1 : 5](#_Toc184473273)

[Question 2 : 5](#_Toc184473274)

[2. Importation et Exploration Préliminaire des Données 6](#_Toc184473275)

[Question 1 : 6](#_Toc184473276)

[Question 2 : 6](#_Toc184473277)

[IV. Étapes Réalisées 7](#_Toc184473278)

[3. Netoyage et Préparation des Données 7](#_Toc184473279)

[Question 1 : 7](#_Toc184473280)

[Question 2 : 8](#_Toc184473281)

[4. Analyse Exploratoire 11](#_Toc184473282)

[5. Visualisations Dynamiques dans Power BI 14](#_Toc184473283)

[V. Résultats Principaux 15](#_Toc184473284)

[VI. Applications Pratiques 16](#_Toc184473285)

[VII. Limitations 17](#_Toc184473286)

[VIII. Recommandations 18](#_Toc184473287)

[IX. Conclusion 19](#_Toc184473288)

# Contexte du Projet

La violence armée représente une problématique critique aux États-Unis, avec des conséquences sociales, économiques et humaines profondes. Ce projet s’inscrit dans une démarche d’analyse approfondie pour mieux comprendre ce phénomène à travers plusieurs dimensions clés :

1. **Les États les plus dangereux** : Identifier les régions où la violence armée est la plus fréquente et meurtrière, en mettant en évidence les États et villes les plus touchés.
2. **Les dangers de la violence armée** : Évaluer l'impact de ces incidents en termes de pertes humaines (morts et blessés) et leurs implications sur la sécurité publique.
3. **Les périodes les plus dangereuses de l'année** : Analyser les tendances saisonnières pour repérer les mois ou les saisons où les incidents sont les plus fréquents.
4. **Les groupes démographiques impliqués** : Étudier les caractéristiques des victimes et des suspects (âge, genre, rôle dans l’incident) pour mieux comprendre les populations affectées.

Le projet repose sur une analyse multi-dimensionnelle, combinant :

* **Analyse géographique** : Pour localiser les zones les plus touchées et cartographier les incidents.
* **Analyse temporelle** : Pour identifier les variations saisonnières et annuelles.
* **Analyse démographique** : Pour comprendre les profils des individus impliqués (victimes ou suspects).

L’objectif est d’apporter des éclairages pertinents pour guider les politiques publiques, renforcer la prévention, et sensibiliser aux dangers liés à la violence armée.

# Objectifs du Projet

* Comprendre les tendances temporelles des incidents de violence armée.
* Identifier les zones géographiques les plus affectées.
* Explorer les caractéristiques des incidents (nombre de victimes, type d’armes impliquées, etc.).
* Créer un tableau de bord interactif pour visualiser les résultats et faciliter l'interprétation des données.

# Description des Données

## Introduction et Choix du Jeu de Données

Question 1 : **Quel jeu de données avez-vous choisi, et pourquoi ce sujet est-il pertinent ou intéressant pour l’analyse ?**  
 Le jeu de données choisi concerne la violence armée aux États-Unis entre 2013 et 2018. Ce sujet est pertinent car la violence armée représente un problème critique aux États-Unis, affectant de nombreux aspects de la société, notamment la sécurité publique, la santé et les politiques sociales. Analyser ces données peut aider à identifier des tendances, des zones sensibles, et les facteurs associés à ces incidents. Cela permet également de fournir des insights pour orienter les politiques publiques et les stratégies de prévention.

Question 2 : **Quelles sources de données avez-vous explorées pour identifier un jeu de données pertinent et répondant aux critères (au moins 500 lignes) ?**  
Pour identifier un jeu de données pertinent, plusieurs bases de données publiques ont été explorées, notamment **Kaggle**, Data.gov et Gun Violence Archive. Le jeu de données final provient de Gun Violence Archive, une source fiable qui compile les incidents liés aux armes à feu à partir de rapports des forces de l'ordre, des médias et des bases de données publiques. Ce jeu de données a été choisi car il contient plus de **260 000** lignes, satisfaisant largement les critères de taille minimale, et offre une couverture détaillée des incidents.

## Importation et Exploration Préliminaire des Données

Question 1 : **Quelles sont les principales caractéristiques de votre jeu de données (nombre de colonnes, types de variables) ?**  
Le jeu de données comporte **29 colonnes** couvrant une variété d'informations, notamment :

* **Variables temporelles** : date (date de l'incident).
* **Variables géographiques** : state, city\_or\_county, latitude, longitude.
* **Variables sur les victimes** : n\_killed (nombre de morts), n\_injured (nombre de blessés).
* **Caractéristiques des incidents** : incident\_characteristics (ex. fusillades de masse, possession illégale d'armes).
* **Détails des participants** : participant\_age, participant\_gender, participant\_type.

Le jeu de données est hétérogène, incluant des variables qualitatives, quantitatives, et des données textuelles structurées à l’aide de séparateurs (||).

Question 2 : **Avez-vous identifié des données manquantes ou des valeurs aberrantes lors de votre première exploration ?**

**Si oui, quelles sont-elles ?**  
Oui, plusieurs problèmes ont été identifiés lors de l'exploration préliminaire :

* **Données manquantes** : Certaines colonnes, incident\_url, source\_url, incident\_url\_fields\_missing et location\_description, contiennent des valeurs nulles ou absentes pour de nombreux enregistrements.
* **Valeurs aberrantes** : Dans certaines colonnes, comme n\_guns\_involved, des valeurs sont soit absentes, soit incohérentes (par exemple, des incidents enregistrés avec zéro armes impliquées).
* **Colonnes avec des séparateurs** : Des colonnes comme participant\_age et incident\_characteristics utilisent le séparateur || pour combiner plusieurs valeurs, ce qui complique leur traitement direct.

# Étapes Réalisées

## Netoyage et Préparation des Données

Question 1 : **Quelles étapes avez-vous effectuées pour nettoyer les données (traitement des valeurs manquantes, suppression des doublons, transformation des formats) ?**

Lors de la phase de nettoyage des données, les étapes suivantes ont été réalisées :

1. **Suppression des colonnes inutiles** :
   * Les colonnes suivantes ont été supprimées car elles ne contribuent pas à l'analyse :
     + incident\_url, source\_url, incident\_url\_fields\_missing (liens vers des sources externes).
   * Les colonnes participant\_relationship, gun\_stolen, et gun\_type ont également été supprimées, car elles contiennent des valeurs vides dans presque tous les enregistrements.
2. **Vérification des doublons** :
   * Une analyse des doublons a été effectuée sur la base des colonnes principales (incident\_id, date, state). Aucune duplication n’a été identifiée, ce qui garantit l’unicité des enregistrements.
3. **Traitement des formats** :
   * Les formats de la colonne date ont été standardisés au format AAAA-MM-JJ.
   * Les valeurs textuelles dans des colonnes comme incident\_characteristics ont été uniformisées pour éviter les variations dues à des erreurs de typographie.
4. **Gestion des valeurs manquantes** :
   * Les colonnes ayant des valeurs manquantes importantes et peu exploitables (comme latitude et longitude) ont été conservées pour une analyse partielle géographique.

Question 2 : **Quelles transformations avez-vous appliquées pour rendre les données exploitables ?**

Les transformations suivantes ont été appliquées :

1. **Catégorisation et normalisation des données sur les participants** :
   * Les colonnes relatives aux participants (participant\_age\_group, participant\_gender, participant\_status, et participant\_type) ont été normalisées et catégorisées :
     + **participant\_age\_group** : Groupes d’âge comme Enfant (<18 ans) et Adulte (18+ ans).
     + **participant\_gender** : Normalisation des valeurs (Male, Female).
     + **participant\_status** : Statut des participants (e.g., Killed, Injured, Arrested).
     + **participant\_type** : Types comme Victim, Subject-Suspect.

A screenshot of a computer

Description automatically generated

1. **Création d'une nouvelle variable** :
   * Une colonne nommée **n\_participant\_involved** a été ajoutée pour calculer automatiquement le nombre total de participants impliqués dans chaque incident.
   * Ce calcul a été réalisé en comptant les occurrences dans les colonnes relatives aux participants (participant\_age\_group, etc.).
2. **Création d’une table calendrier :**
   * Une table calendrier a été générée pour enrichir l’analyse temporelle. Cette table contient des colonnes spécifiques telles que :
     + **Année** : Pour faciliter l’analyse des tendances annuelles.
     + **Mois** : Pour identifier les variations saisonnières et mensuelles.
   * La table calendrier a été liée à la colonne date du dataset principal, permettant d'effectuer des analyses temporelles plus précises et d’améliorer la cohérence des données.

Cette approche a permis de structurer les données temporelles de manière efficace, tout en rendant l’analyse plus flexible et intuitive.

Pour **normaliser** la colonne **incident\_characteristics**, qui contient des informations complexes séparées par des délimiteurs, et structurer les données de manière exploitable, deux tables distinctes ont été créées : une table **Gun** et une table **Participant**. Cette normalisation permet de simplifier l’analyse en séparant les détails des armes impliquées et les caractéristiques des participants. Les codes Python ci-dessous montrent les étapes utilisées pour extraire et organiser ces informations à partir de la colonne d'origine.

A screen shot of a computer program

Description automatically generated

A computer screen shot of a program code

Description automatically generatedA screen shot of a computer program

Description automatically generated

## Analyse Exploratoire

* **Analyse temporelle** :
  1. Calcul du nombre total d’incidents par année et par mois.
  2. Identification des tendances (augmentation ou diminution des incidents).

A screenshot of a computer

Description automatically generated

* **Analyse géographique** :
  1. Distribution des incidents par état et par ville.
  2. Création de cartes interactives pour visualiser les zones les plus touchées.

A screenshot of a computer

Description automatically generated

* **Analyse démographique** :
  1. Répartition des victimes par âge, sexe et rôle (participant\_type).
  2. Comparaison des incidents impliquant différents types d'armes.

A screenshot of a computer

Description automatically generated

A close-up of a graph

Description automatically generated

## Visualisations Dynamiques dans Power BI

Les visualisations suivantes ont été créées :

* **Graphique en barres** : Nombre total d'incidents par année.
* **Carte géographique** : Répartition des incidents dans chaque état.
* **Diagramme circulaire** : Proportion des incidents selon leurs caractéristiques (fusillades de masse, incidents domestiques, etc.).
* **Graphique en lignes** : Tendance mensuelle des incidents.
* **Carte thermique (Heatmap)** : Corrélation entre les incidents, les victimes et les armes impliquées.

# Résultats Principaux

Les résultats obtenus permettent d’apporter des éclairages pertinents pour guider les politiques publiques, renforcer les stratégies de prévention et sensibiliser aux dangers liés à la violence armée :

* **Tendances temporelles :**
  + Le 1er janvier est identifié comme le jour le plus dangereux pour les incidents de violence armée.
  + Les mois de mars et janvier enregistrent le plus grand nombre d'incidents dans l'année, ce qui souligne une concentration saisonnière à la fin et au début de l'année.
* **Tendances géographiques :**
  + Les États les plus touchés par la violence armée sont l'Illinois, la Californie, la Floride et le Texas, représentant une part importante des incidents recensés.
  + Les grandes villes de ces États concentrent la majorité des incidents, reflétant l’impact des densités urbaines sur le phénomène.
* **Caractéristiques des incidents :**
  + Bien que les fusillades de masse constituent une faible proportion des incidents, elles entraînent un nombre disproportionné de victimes, mettant en évidence leur gravité.
  + Les hommes adultes sont les principales victimes et auteurs impliqués dans ces incidents, confirmant un schéma récurrent au sein des violences armées.

Ces résultats mettent en évidence des aspects critiques et récurrents du phénomène de la violence armée, offrant des points d’intervention prioritaires pour les décideurs et les organismes de prévention.

# Applications Pratiques

* **Politiques publiques** :
  + Renforcement des lois sur les armes dans les états les plus touchés.
  + Ciblage des programmes de prévention dans les zones à risque.
* **Sensibilisation** :
  + Campagnes pour informer le public sur les risques liés aux armes à feu.
* **Éclairages pour la recherche** :
  + Analyse approfondie des facteurs contribuant à la violence armée.

# Limitations

* Données incomplètes : Certaines colonnes manquent de données critiques (par exemple, location\_description).
* Difficulté à standardiser certaines variables (comme incident\_characteristics).
* Nécessité de données additionnelles pour évaluer des facteurs contextuels (population, politique d’armement par état).

# Recommandations

* Étendre l’analyse à d’autres facteurs contextuels (données socio-économiques, densité de population).
* Collaboration avec des experts en politiques publiques pour interpréter les résultats.
* Exploration future des tendances post-2018 pour observer les évolutions.

# Conclusion

L’analyse des données sur la violence armée aux États-Unis entre 2013 et 2018 a permis de dégager des tendances importantes et d’identifier des zones critiques nécessitant une attention particulière. Les résultats montrent que certains États, comme l’Illinois, la Californie, la Floride et le Texas, concentrent une part significative des incidents, souvent dans des grandes zones urbaines. Sur le plan temporel, les mois de janvier et mars ainsi que le 1er janvier se distinguent comme des périodes particulièrement dangereuses.

Les fusillades de masse, bien qu’en faible proportion, restent particulièrement dévastatrices en termes de pertes humaines, et les hommes adultes sont les plus souvent impliqués, tant comme victimes que comme auteurs. Ces observations soulignent la nécessité de renforcer les politiques publiques, comme le contrôle des armes à feu, tout en sensibilisant les communautés locales aux périodes et zones à risque.

Ces analyses offrent une base solide pour guider les décideurs dans la conception de stratégies de prévention plus ciblées et efficaces, visant à réduire l'impact de la violence armée sur les populations.