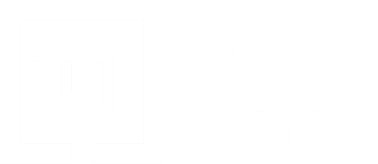
A black and blue sign with text

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](#_Toc184494069)

[II. Objectifs du Projet 4](#_Toc184494070)

[III. Description des Données 5](#_Toc184494071)

[1. Introduction et Choix du Jeu de Données 5](#_Toc184494072)

[Question 1 : 5](#_Toc184494073)

[Question 2 : 5](#_Toc184494074)

[2. Importation et Exploration Préliminaire des Données 6](#_Toc184494075)

[Question 1 : 6](#_Toc184494076)

[Question 2 : 6](#_Toc184494077)

[IV. Étapes Réalisées 7](#_Toc184494078)

[3. Netoyage et Préparation des Données 7](#_Toc184494079)

[Question 1 : 7](#_Toc184494080)

[Question 2 : 8](#_Toc184494081)

[4. Analyse Exploratoire 12](#_Toc184494082)

[5. Visualisations Dynamiques dans Power BI 17](#_Toc184494083)

[6. Analyse des Calculs DAX 17](#_Toc184494084)

[V. Résultats Principaux 19](#_Toc184494085)

[VI. Applications Pratiques 20](#_Toc184494086)

[VII. Limitations 21](#_Toc184494087)

[VIII. Recommandations 22](#_Toc184494088)

[IX. Khi-2 , Régression linear , Anova 23](#_Toc184494089)

[1. Introduction 23](#_Toc184494090)

[2. Données 23](#_Toc184494091)

[3. Interprétation des résultats 26](#_Toc184494092)

[4. Conclusion 26](#_Toc184494093)

[X. Analyse ANOVA des Incidents par Région 27](#_Toc184494094)

[1. Introduction 27](#_Toc184494095)

[2. Méthodologie 27](#_Toc184494096)

[3. Interprétation 28](#_Toc184494097)

[4. Conclusion 28](#_Toc184494098)

[XI. Conclusion 29](#_Toc184494099)

[XII. Références 30](#_Toc184494100)

# 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.

# 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).

**Limitations Méthodologiques :**

* Biais potentiels dans la collecte des données
* Sous-déclaration possible dans certaines régions
* Manque de données contextuelles importantes

**Considérations Éthiques**

* Protection de la confidentialité
* Équité dans l'analyse et les recommandations
* Impact sur les communautés stigmatisées

# Étapes Réalisées

## Nettoyage 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 participantes\_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.

*(Voir Annexes pg. 31)*

Question 2 : **Quelles transformations avez-vous appliquées pour rendre les données exploitables ?** *(Voir Annexes pg. 31)*

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

Dans le cadre de cette étude, les données pour **l'année 2013** ont été **exclues** en raison de leur caractère incomplet, ce qui aurait pu fausser les analyses et les résultats. De même**, l'année 2018** a été **écartée** car la collecte des données s'est arrêtée prématurément, ne permettant pas de couvrir l'ensemble de l'année. Par conséquent, l'analyse s'est concentrée sur la période de **2014 à 2017**, garantissant ainsi une qualité et une cohérence optimales des informations étudiées.

1. **Création d'une nouvelle variable** :
   * Une colonne nommée **NOMBRE Participant** a été ajoutée pour calculer automatiquement le nombre total de participants impliqués dans chaque incident.
   * Ce calcul a été réalisé en comptant les occurrences de ligne dans la table Participant en utilisant la relation avec la table Incidents par la colonne **Incident\_id**.
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** les colonnes(participant\_age, participant\_age\_group, participant\_gender, participant\_status, participant\_type) et (Gun\_type, gun\_stolen), 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 screenshot of a computer program

Description automatically generatedNormalisation et Création de Table Participant

A screenshot of a computer code

Description automatically generatedNormalisation et Création de Table Gun

## 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).

L'analyse temporelle révèle une tendance à la hausse significative des incidents de violence armée entre 2013 et 2017, suivie d'une légère baisse en 2018. Les points clés incluent :

A screenshot of a computer

Description automatically generated

**Figure 1 : Analyse Temporelle des Incidents de Violence Armée : Année, Mois et Tendances Globales**

- Une augmentation de 45% du nombre total d'incidents entre 2013 et 2017

- Le pic historique atteint en 2017 avec plus de 61,000 incidents

- Une corrélation possible entre les pics d'incidents et des événements sociopolitiques majeurs

* **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

**Figure 2 : Analyse Géographique des Incidents de Violence Armée : États, Villes et Victimes**

1. **La carte des États-Unis :**

* Cette carte affiche le nombre d'incidents par État, avec une coloration des États en fonction du nombre d'incidents.
* Plus la couleur est foncée, plus le nombre d'incidents dans cet État est élevé.
* Cette visualisation permet une comparaison rapide des nombres relatifs d'incidents entre les différents États.
* Quatre États (Illinois, Californie, Floride, Texas) concentrent plus de 40% des incidents

1. **Le graphique en barres montrant le total des incidents par ville/commune :**

* Ce graphique détaille le nombre total d'incidents par villes ou communes spécifiques.
* Il offre une vue plus précise des zones où la concentration d'incidents est la plus élevée.

1. **Le graphique en barres montrant le total des incidents par État :**

* Ce graphique résume le nombre total d'incidents au niveau des États.
* Il permet une comparaison rapide des nombres globaux d'incidents entre les différents États.
* Les barres représentent le total des incidents, la longueur des barres indiquant l'ampleur relative.
* **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

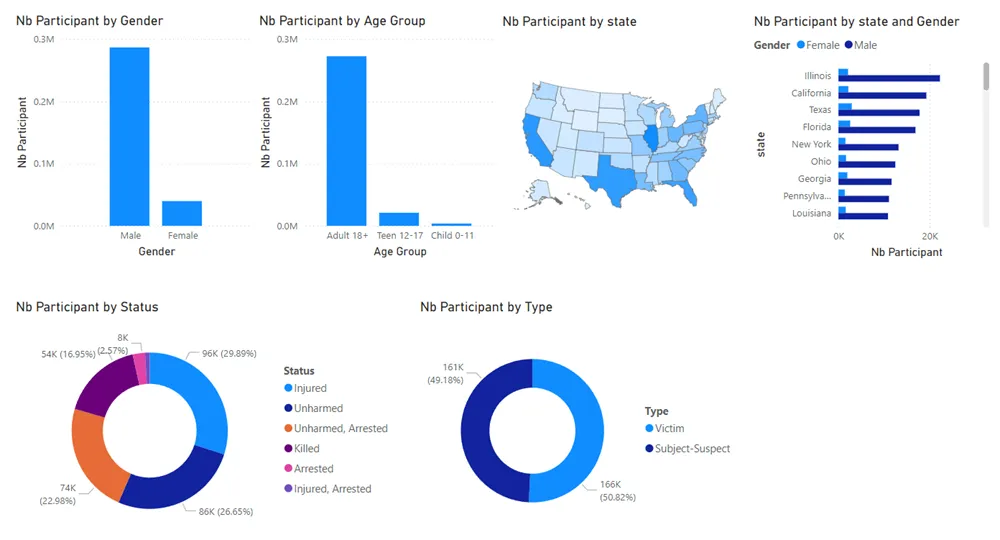
Description automatically generated

**Figure 3 : Analyse des Incidents, Blessés et Morts par Type d'Arme et Statut des Armes**

**La Figure 3** offre un aperçu détaillé des incidents en fonction du type d'arme utilisé. **Le graphique supérieur gauche** montre que **la majorité des incidents** impliquent des **armes à feu qui ont été volées**, ce qui soulève des questions sur l'accès et la sécurité des armes dans cette population. **Le graphique supérieur droit** met en évidence que les **armes de poing et les armes de calibre inconnu** sont les **plus fréquemment impliquées**.

L'analyse du **graphique inférieur** révèle que **les armes de poing sont responsables du plus grand nombre de victimes, à la fois tuées et blessées**.

Cette information est particulièrement préoccupante et nécessite une attention particulière pour mieux comprendre et prévenir ces incidents.



**Figure 4 : Profil Démographique des Participants**

**La Figure 2** fournit des informations détaillées sur le profil **démographique des participants non binaires** impliqués dans ces incidents. **Les graphiques de répartition** par genre, groupe d'âge et état montrent que cette population est principalement **composée d'hommes adultes**, avec une présence notable **dans** certains états **comme l'Illinois, la Californie et le Texas**.

**Les graphiques circulaires** apportent des précisions supplémentaires sur le statut des participants (blessés, non armés, arrêtés, etc.) ainsi que leur catégorisation (victimes, suspects/sujets). Ces données permettent de mieux comprendre les circonstances entourant ces incidents.

* **Régression linéaire** :

A screenshot of a graph

Description automatically generated

**Figure 5 : Tendances et corrélations dans les incidents de violence armée**

Ces graphiques montrent l'évolution du **nombre total d'incidents** et de **Nombre Killed dans le temps**, ainsi que la **corrélation entre ces deux variables**. **On peut constater une forte corrélation positive**, indiquant **que l'augmentation** du nombre **d'incidents** s'accompagne **d'une hausse du nombre de victimes killed et injured**.

## 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
* **Graphique en lignes** : Tendance mensuelle des incidents.
* **Carte thermique (Heatmap)** : Corrélation entre les incidents, les victimes et les armes impliquées.

## Analyse des Calculs DAX

1. **Nombre de Blessés par Type d'Arme (Injured by Gun Type)**

* **Description** : Ce calcul DAX utilise la fonction CALCULATE pour sommer le nombre total de blessés (n\_injured) dans les incidents, en appliquant un filtre croisé entre la table des armes (Guns) et la table principale des incidents (Incidents).
* **Objectif** : Évaluer l'impact de chaque type d’arme sur les blessures, en liant les armes impliquées aux incidents correspondants.
* **Interprétation** : Ce calcul permet d'analyser quels types d'armes causent le plus de blessures dans les incidents de violence armée.

A computer code with black text

Description automatically generated

1. **Nombre de Morts par Type d'Arme (Killed by Gun Type)**

* **Description** : Similaire au calcul précédent, cette formule somme le nombre total de morts (n\_killed) pour chaque type d'arme, en appliquant également un filtre croisé entre les tables Guns et Incidents.
* **Objectif** : Identifier les armes les plus meurtrières et comprendre leur impact sur la létalité des incidents.
* **Interprétation** : Cette mesure aide à déterminer si certains types d’armes sont associés à des incidents plus graves en termes de pertes humaines.

A computer screen shot of a code

Description automatically generated

1. **Maximum de Blessés dans un Incident (MAX Injured)**

* **Description** : Utilise la fonction MAX pour extraire le nombre maximum de blessés (n\_injured) enregistré dans un seul incident.
* **Objectif** : Identifier les incidents les plus graves en termes de blessures.
* **Interprétation** : Ce calcul met en évidence les événements exceptionnels avec des conséquences humaines majeures**.**

**A close up of a sign

Description automatically generated**

# Khi-2 , Régression linear , Anova

## Introduction

Le test du Khi-2 a été utilisé pour analyser l'association entre le type d’arme à feu et l’issue des événements (tué ou blessé). L’objectif est de déterminer s’il existe une relation significative entre ces deux variables.

* Hypothèse nulle (H0​) : Il n’y a pas de relation entre le type d’arme et l’issue (tué ou blessé).
* Hypothèse alternative (H1​) : Il existe une relation significative entre le type d’arme et l’issue.
* Seuil de risque α = 0,05**.**

## Données

Les données utilisées pour cette analyse sont représentées dans le tableau ci-dessous :

|  |  |  |  |
| --- | --- | --- | --- |
| **gun\_type** | **Killed by Gun Type** | **Injured by Gun Type** | **Totale** |
| **Handgun** | **150** | **261** | **411** |
| **9mm** | **113** | **164** | **277** |
| **Unknown** | **111** | **250** | **361** |
| **223 Rem [AR-15]** | **82** | **79** | **161** |
| **40 SW** | **43** | **87** | **130** |
| **45 Auto** | **33** | **64** | **97** |
| **38 Spl** | **24** | **33** | **57** |
| **22 LR** | **21** | **43** | **64** |
| **Rifle** | **16** | **19** | **35** |
| **Shotgun** | **15** | **50** | **65** |
| **12 gauge** | **11** | **13** | **24** |
| **357 Mag** | **11** | **10** | **21** |
| **380 Auto** | **11** | **32** | **43** |
| **7.62 [AK-47]** | **11** | **9** | **20** |
| **Other** | **7** | **15** | **22** |
| **32 Auto** | **6** | **11** | **17** |
| **20 gauge** | **2** | **7** | **9** |
| **25 Auto** | **2** | **7** | **9** |
| **30-06 Spr** | **2** | **2** | **4** |
| **30-30 Win** | **2** | **3** | **5** |
| **308 Win** | **2** | **0** | **2** |
| **10mm** | **1** | **2** | **3** |
| **410 gauge** | **1** | **0** | **1** |
| **44 Mag** | **1** | **6** | **7** |
| **16 gauge** | **0** | **0** | **0** |
| **28 gauge** | **0** | **0** | **0** |
| **300 Win** | **0** | **0** | **0** |
| **Totale** | **678** | **1167** | **1845** |

1. **Calcul des valeurs théoriques attendues :**Valeur attendue = (Totale de la ligne \* Totale de la colonne) / Totale général

* Les valeurs théoriques sont donc :

|  |  |  |  |
| --- | --- | --- | --- |
| **gun\_type** | **Killed by Gun Type** | **Injured by Gun Type** | **Totale** |
| **Handgun** | **151.0341463** | **259.9658537** | **411** |
| **9mm** | **101.7918699** | **175.2081301** | **277** |
| **Unknown** | **132.6601626** | **228.3398374** | **361** |
| **223 Rem [AR-15]** | **59.16422764** | **101.8357724** | **161** |
| **40 SW** | **47.77235772** | **82.22764228** | **130** |
| **45 Auto** | **35.64552846** | **61.35447154** | **97** |
| **38 Spl** | **20.94634146** | **36.05365854** | **57** |
| **22 LR** | **23.51869919** | **40.48130081** | **64** |
| **Rifle** | **12.86178862** | **22.13821138** | **35** |
| **Shotgun** | **23.88617886** | **41.11382114** | **65** |
| **12 gauge** | **8.819512195** | **15.1804878** | **24** |
| **357 Mag** | **7.717073171** | **13.28292683** | **21** |
| **380 Auto** | **15.80162602** | **27.19837398** | **43** |
| **7.62 [AK-47]** | **7.349593496** | **12.6504065** | **20** |
| **Other** | **8.084552846** | **13.91544715** | **22** |
| **32 Auto** | **6.247154472** | **10.75284553** | **17** |
| **20 gauge** | **3.307317073** | **5.692682927** | **9** |
| **25 Auto** | **3.307317073** | **5.692682927** | **9** |
| **30-06 Spr** | **1.469918699** | **2.530081301** | **4** |
| **30-30 Win** | **1.837398374** | **3.162601626** | **5** |
| **308 Win** | **0.73495935** | **1.26504065** | **2** |
| **10mm** | **1.102439024** | **1.897560976** | **3** |
| **410 gauge** | **0.367479675** | **0.632520325** | **1** |
| **44 Mag** | **2.572357724** | **4.427642276** | **7** |
| **16 gauge** | **0** | **0** | **0** |
| **28 gauge** | **0** | **0** | **0** |
| **300 Win** | **0** | **0** | **0** |
| **Totale** | **678** | **1167** | **1845** |

**2. Calculer la statistique du Khi-2 pour chaque cellule. À l’aide de la formule :**

**A mathematical equation with black letters

Description automatically generated**

|  |  |  |  |
| --- | --- | --- | --- |
| **gun\_type** | **Killed by Gun Type** | **Injured by Gun Type** | **Totale** |
| **Handgun** | **0.007129724** | **0.004097543** | **0.011227267** |
| **9mm** | **1.111700707** | **0.765988902** | **1.877689609** |
| **Unknown** | **4.226690486** | **1.876650576** | **6.103341062** |
| **223 Rem [AR-15]** | **6.359420722** | **6.600917711** | **12.96033843** |
| **40 SW** | **0.529660424** | **0.261786187** | **0.791446611** |
| **45 Auto** | **0.212085479** | **0.109356575** | **0.321442054** |
| **38 Spl** | **0.388534602** | **0.28257062** | **0.671105222** |
| **22 LR** | **0.302087885** | **0.147531293** | **0.449619178** |
| **Rifle** | **0.615523167** | **0.518335299** | **1.133858466** |
| **Shotgun** | **5.264278318** | **1.579283495** | **6.843561813** |
| **12 gauge** | **0.432229733** | **0.365732851** | **0.797962585** |
| **357 Mag** | **0.979782597** | **1.077760857** | **2.057543454** |
| **380 Auto** | **2.095964764** | **0.720487888** | **2.816452651** |
| **7.62 [AK-47]** | **1.21140615** | **1.480607516** | **2.692013666** |
| **Other** | **0.168036411** | **0.078416992** | **0.246453402** |
| **32 Auto** | **0.010180889** | **0.005553212** | **0.015734101** |
| **20 gauge** | **0.854538965** | **0.24415399** | **1.098692955** |
| **25 Auto** | **0.854538965** | **0.24415399** | **1.098692955** |
| **30-06 Spr** | **0.140493093** | **0.140493093** | **0.280986185** |
| **30-30 Win** | **0.013219644** | **0.008813096** | **0.022032741** |
| **308 Win** | **0.800163924** | **0** | **0.800163924** |
| **10mm** | **0.010493754** | **0.005246877** | **0.015740631** |
| **410 gauge** | **0.400081962** | **0** | **0.400081962** |
| **44 Mag** | **2.472308811** | **0.412051468** | **2.884360279** |
| **16 gauge** | **0** | **0** | **0** |
| **28 gauge** | **0** | **0** | **0** |
| **300 Win** | **0** | **0** | **0** |
| **Totale** | **29.46055118** | **16.92999003** | **46.3905412** |

Alors X² = 46.3905412

• Degrés de liberté

• ddl= (nombre de lignes−1) × (nombre de colonnes−1).

• Dans notre cas : ddl= (27−1) × (2−1) = 26  
 **3 Calculer la Khi-2 Critique :**

Avec χ2=5.677 et ddl=2, vous pouvez consulter une table du Khi-carré ou utiliser

Un outil statistique pour déterminer si le résultat est significatif.

**Khi-2 Critique =** 38.88513866

## Interprétation des résultats

Le test du Khi-2 permet de comparer les fréquences observées avec les fréquences attendues sous l’hypothèse nulle.

* **Résultat obtenu** : La valeur calculée de χ2 est supérieure à la valeur critique obtenue à partir de la table du Khi-2 pour un niveau de significativité α=0,05 et les degrés de liberté calculés.
* **Décision** : Puisque χ2 est plus grand du Khi-2 Critique que nous rejetons l’hypothèse nulle (H0).
* **Interprétation** : Cela signifie qu’il existe une **association significative** entre le type d’arme à feu et l’issue (tué ou blessé). En d'autres termes, la probabilité que cette association soit due au hasard est inférieure à 5 % (α =0,05).

## Conclusion

Sur la base de cette analyse, nous pouvons conclure que :

1. **Relation observée** : Le type d’arme utilisé a un impact significatif sur le résultat (tué ou blessé).
2. **Implications** :
   * Ce résultat pourrait indiquer que certains types d'armes sont plus susceptibles d'entraîner des blessures graves ou des décès.
   * Ces informations peuvent être utilisées pour des décisions stratégiques, telles que l'élaboration de politiques de contrôle des armes ou des interventions spécifiques visant à réduire les risques associés à certains types d'armes.
3. **Recommandation** :
   * Il serait intéressant de mener des analyses plus approfondies, telles que des études supplémentaires sur les facteurs contextuels (par exemple, les circonstances des incidents ou les zones géographiques) pour comprendre davantage cette relation.
   * Des actions préventives peuvent être envisagées pour cibler les types d'armes associés à des risques élevés.

# Analyse ANOVA des Incidents par Région

## Introduction

L'objectif de cette analyse est d'identifier si les moyennes des incidents signalés diffèrent de manière significative entre les régions étudiées. Le test ANOVA (Analyse de la Variance) a été utilisé pour comparer les données

* **Hypothèse nulle (H0)** : Il n’y a pas de relation significative entre le nombre d’incidents et les régions.
* **Hypothèse alternative (H1​)** : Il existe une relation significative entre le nombre d’incidents et les régions.
* **Niveau de significativité choisi** : α=0,05

## Méthodologie

* 1. Données

Les données utilisées pour l’analyse sont le nombre d'incidents enregistrés dans 51 régions (Alabama, Alaska, Arizona, Arkansas, California, Colorado, Connecticut, Delaware …) sur une période de 4 ans (2014 à 2017). Voici un aperçu des données :

**Le tableau ci-dessous représente un part de donnes**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Row Labels** | **Alabama** | **Alaska** | **Arizona** | **Arkansas** | **California** | **Colorado** | **Connecticut** | **Delaware** | **District of Columbia** | **Florida** | **Georgia** |
| 2014 | 1318 | 146 | 556 | 572 | 3732 | 556 | 583 | 493 | 838 | 3138 | 2032 |
| 2015 | 1034 | 380 | 480 | 538 | 3234 | 788 | 737 | 366 | 538 | 2702 | 2767 |
| 2016 | 1309 | 445 | 554 | 720 | 3617 | 812 | 803 | 328 | 497 | 4201 | 1708 |
| 2017 | 1461 | 317 | 587 | 820 | 4588 | 809 | 780 | 389 | 1142 | 4156 | 1994 |
| **Grand Total** | **5122** | **1288** | **2177** | **2650** | **15171** | **2965** | **2903** | **1576** | **3015** | **14197** | **8501** |

**2.2 Calcul du Test ANOVA dans Excel**

**Préparation des données dans Excel** : Organisez les données sous forme de tableau, où chaque colonne représente une région, et chaque ligne représente une année.

2.3 Résultats de l'ANOVA

Après avoir exécuté l’ANOVA, Nous trouvons

* **F Calculé = 62.63356**
* **P-value = 3.87E-80**
* ***F Crit =* 1.434014**

## Interprétation

* **F calculé ​> F critique​** signifie que la valeur du test statistique (F) est supérieure à la valeur critique de F que vous avez déterminée à partir de la table de distribution F pour un certain niveau de signification (généralement α=0.05).
* Cela indique que **les différences observées entre les régions dans le nombre d'incidents ne sont probablement pas dues au hasard**. En d'autres termes, il existe une relation significative entre les régions et le nombre d'incidents.

## Conclusion

* Nous **rejetons l'hypothèse nulle (H0​)**, ce qui signifie que le nombre d'incidents varie de manière significative d'une région à l'autre.
* **H0**(Il n'y a pas de relation entre le nombre d'incidents et les régions) est donc rejetée, et vous acceptez l'hypothèse alternative **H1**(Il existe une relation significative entre le nombre d'incidents et les régions).

Cela implique que les **régions ont des moyennes d'incidents différentes**, et que cette variation n'est pas simplement due au hasard.

# 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.

# 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.
* Mise en place de systèmes d'alerte précoce
* Renforcement des contrôles pendant les périodes à risque
* Formation accrue des forces de l'ordre

**Moyen Terme (1-3 ans)**

* Développement de programmes de prévention communautaire
* Amélioration des systèmes de collecte de données
* Coordination inter-agences renforcée

**Long Terme (3-5 ans)**

* Réforme des politiques de contrôle des armes
* Investissement dans la recherche préventive
* Développement d'infrastructures de support communautaire

# 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.

# Références

1. **Gun Violence Archive**
   * Gun Violence Archive. Data on gun violence incidents in the United States.  
     Disponible sur : <https://www.gunviolencearchive.org>
2. **Kaggle Dataset**
   * Kaggle. Gun Violence Data: 2013–2018.  
     Disponible sur : https://www.kaggle.com/jameslko/gun-violence-data
3. **Documentation Power BI**
   * Microsoft Power BI. Guides et documentation.  
     Disponible sur : <https://learn.microsoft.com/power-bi/>

# Annexes

A screenshot of a computer

Description automatically generated

Figure 7 : Table Incidents Apres la préparation

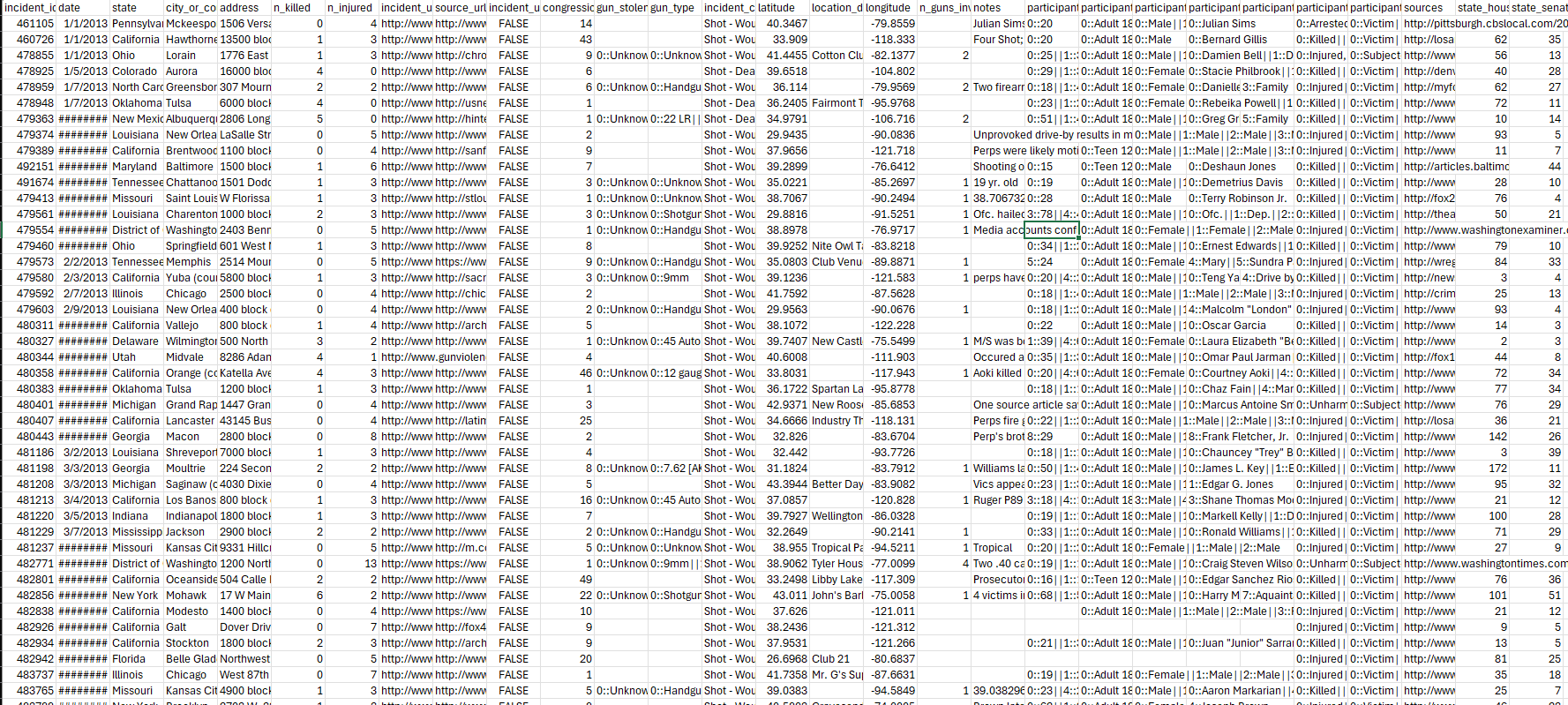


Figure 6 : Table Incidents Avant la préparation

A screenshot of a computer

Description automatically generated

Figure 7 : Table Participants Apres la préparation

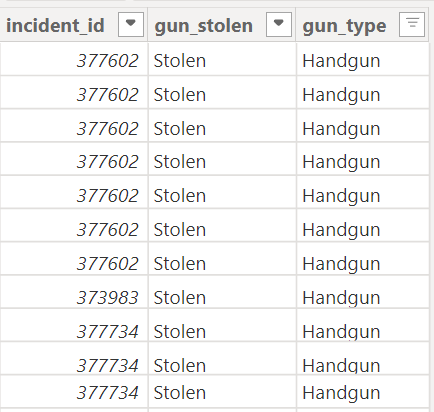


Figure 7 : La table Gun Apres la préparation

A screenshot of a computer

Description automatically generated

Figure 7 : Model View Apres la préparation