

Document de conception

Ce projet vise à analyser les profils, les comportements et les tendances des tueurs en série à travers différents pays, à partir d'une base de données relationnelle construite pour répondre à des questions précises. Le but principal est d'identifier les corrélations possibles entre divers paramètres (âge, méthode de meurtre, situation géographique, état mental, etc.) afin d'en tirer des enseignements comportementaux et sociaux. La portée de cette base de données couvre des informations individuelles (nom, âge, dates d'activité, diagnostic de santé mentale, nombre de victimes,...), géographiques (pays, législation sur les armes, taux de criminalité et de pauvreté,...), temporelles (date d'arrestation, année de jugement,...) ainsi que les méthodes de meurtre utilisées. De plus, la base de données a été construite à la fois grâce à des données réelles, récupérées dans un dataset Kaggle, ainsi que de données fictives inventées/générées grâce à un LLM.

En termes d'optimisation, la base est normalisée de manière à éviter la redondance et à simplifier les mises à jour. Cependant, certaines limitations subsistent. Le volume de données est restreint, ce qui limite la portée statistique de certaines conclusions ainsi que leurs validité car beaucoup de données ont été inventées afin d'étayer la base de données. Par ailleurs, certaines valeurs (comme les diagnostics ou les descriptions de cibles) dépendent fortement de la qualité des mots-clés, rendant l'analyse sémantique approximative.

Malgré ces contraintes, la base permet d'aborder des analyses pertinentes telles que la répartition des méthodes de meurtre selon l'âge, la comparaison de profils psychologiques, ou encore les tendances géographiques liées à la criminalité et à la pauvreté. Elle constitue ainsi un outil rigoureux et adaptable à des études criminologiques exploratoires.

Portée de la Base de Données

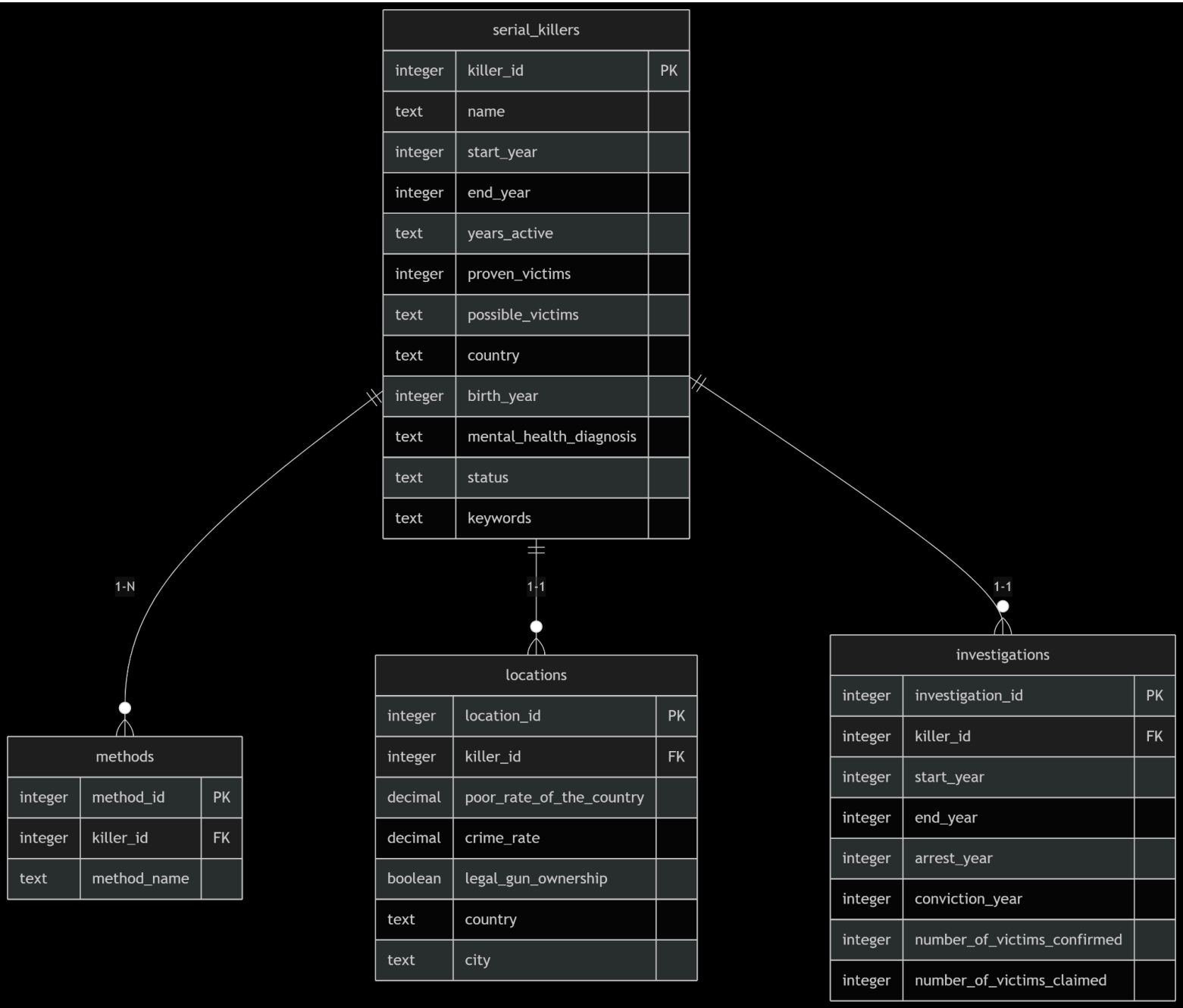
La base de données a été conçue pour :

- Centraliser des informations sur les tueurs en série les plus prolifiques.
- Analyser leurs méthodes, localisations et profils psychologiques.
- Faciliter les requêtes croisées (ex : corrélation entre diagnostic mental et méthode de meurtre).
-

Limites de la portée:

- Nombre de tueurs limités : Ne prends que les 34 les plus prolifiques
- S'appuie sur une vraie base de données reflétant la réalité, mais possèdent beaucoup de données inventés
- Période couverte : Principalement 20e et 21e siècles (peu de cas historiques avant 1900).
- Données victimes : Seul le nombre est stocké, pas de détails individuels (âge, sexe, etc.).

Diagramme de relation d'entités



Description des entités :

- Table serial_killers

- killer_id (PK) : Identifiant unique du tueur (clé primaire)
- name : Nom du tueur en série (*données réelles*)
- start_year : Année de début de ses crimes (*données réelles*)
- end_year : Année de fin de ses crimes (*données réelles*)
- years_active : Période d'activité (format texte) (*données réelles*)
- proven_victims : Nombre de victimes confirmées (*données réelles*)
- possible_victims : Estimation des victimes potentielles (*données réelles*)
- country : Pays d'activité principal (*données réelles*)
- birth_year : Année de naissance du tueur (*données inventées/générées*)
- mental_health_diagnosis : Troubles mentaux diagnostiqués (*données inventées/générées*)
- status : Statut (emprisonné, décédé, etc.) (*données inventées/générées*)
- keywords : Petite description concernant le tueur (*données réelles*)

- Table methods

- method_id (PK) : Identifiant unique de la méthode (clé primaire)
- killer_id (FK) : Lien vers le tueur (clé étrangère)
- method_name : Méthode utilisée pour tuer (*données réelles*)

- Table locations

- location_id (PK) : Identifiant du lieu (clé primaire)
- killer_id (FK) : Lien vers le tueur (clé étrangère)
- poor_rate_of_the_country : Taux de pauvreté du pays (*données inventées/générées*)
- crime_rate : Taux de criminalité local (*donnée inventé*)
- legal_gun_ownership : Possession légale d'armes (oui/non) (*données inventées/générées*)
- country : Pays concerné (*données inventées/générées*)
- city : Ville principale d'activité (*données inventées/générées*)

- Table investigations

- investigation_id (PK) : Identifiant de l'enquête (clé primaire)
- killer_id (FK) : Lien vers le tueur (clé étrangère)
- start_year : Année de début de l'enquête (*données inventées/générées*)
- end_year : Année de fin de l'enquête (*données inventées/générées*)
- arrest_year : Année de l'arrestation (*données inventées/générées*)
- conviction_year : Année de la condamnation (*données inventées/générées*)
- number_of_victims_confirmed : Victimes confirmées dans l'enquête (*données inventées/générées*)
- number_of_victims_claimed : Victimes revendiquées (par le tueur ou soupçons) (*données inventées/générées*)

Modélisation des Relations :

La structure repose sur 3 types de relations :

Relation	Type	Explication
serial_killers → methods	1:n (Un-à-plusieurs)	Un tueur peut avoir plusieurs méthodes.
serial_killers → locations	1:1 (Un-à-un)	Chaque tueur est associé à une zone géographique principale.
serial_killers → investigations	1:1 (Un-à-un)	Chaque tueur a un seul dossier d'enquête.

Permet notamment :

- Éviter la redondance : Séparer methods permet d'éviter de répéter "Strangulation" 10 fois.
- Flexibilité : Ajouter une nouvelle table ou nouvelle informations (ex: "Empoisonnement" dans la table méthode sans modifier la table principale).
- Performance : Requêtes simplifiées par des relations claires (ex: JOIN methods ON...)

Relations entre les tables :

La base de données repose sur 4 tables :

- serial_killers (table centrale)
- methods (méthodes de meurtre)
- locations (contexte géographique)
- investigations (suivi judiciaire)

Table serial_killers (table principale)

- Contient les informations de base sur chaque tueur en série
- Clé primaire : killer_id (entier unique pour chaque tueur)
- Relations :
 - 1:n avec methods (un tueur peut avoir plusieurs méthodes)
 - 1:1 avec locations (chaque tueur a une localisation principale)
 - 1:1 avec investigations (chaque tueur a un dossier d'enquête)

Table methods

- Recense les méthodes de meurtre employées
- Clé primaire composite : (method_id, killer_id)
- Clé étrangère : killer_id référence serial_killers(killer_id)
- Justification : Permet de lister plusieurs méthodes pour un même tueur
- Exemple : Un tueur peut utiliser à la fois strangulation et arme blanche

Table locations

- Donne le contexte géographique des crimes
- Clé primaire composite : (location_id, killer_id)
- Clé étrangère : killer_id référence serial_killers(killer_id)
- Particularité : Relation 1:1 (un tueur = une localisation principale)
- Contient des données socio-économiques (taux de pauvreté, criminalité...)

Table investigations

- Suivi complet des enquêtes judiciaires
- Clé primaire : investigation_id (entier unique)
- Clé étrangère : killer_id référence serial_killers(killer_id)
- Particularité : Relation 1:1 (une enquête par tueur)
- Contient la chronologie judiciaire (arrestation, condamnation...)

Justification de l'utilisation de clé primaire composite :

Une clé primaire composite a été utilisée pour la table methods car :

- Un tueur peut avoir plusieurs méthodes (ex: couteau + strangulation)
- Une méthode peut être utilisée par plusieurs tueurs
- => La combinaison (method_id + killer_id) est unique

- Une clé primaire composite a été utilisée pour la table locations car :
- Chaque tueur à une seule localisation principale
- Plusieurs tueurs peuvent avoir des localisations similaires
- => killer_id doit être unique mais location_id peut se répéter

Avantage :

- Évite les tables de liaison inutiles tout garantissant :
- Pas de doublons
- Relations claires entre tables
- Meilleures performances

Perspectives d'évolution

- Ajout d'une table victims (genre, âge, profession, etc...)
- Séparer la colonne "name" en 2 colonnes "nom" et "prénom"
- Extension à d'autres types de criminels (antérieur au 19ème siècle par exemple)