**1.2 - Affichage et Analyse du Mapping**

GET crimes\_chicago\_2012\_2017/\_mapping

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.Une image contenant texte, capture d’écran, menu, Police

Le contenu généré par l’IA peut être incorrect.

**Champs mal typés :**

* « location » créé automatiquement en combinant « Latitude » et « Longitude » remplace donc « Location » qui est en plus du mauvais type : keyword.
* « Date » et « Updated On » mauvais type : keyword au lieu de date format « MM/dd/yyyy HH:mm:ss a »
* « Arrest » et « Domestic » mauvais type : keyword au lieu de boolean
* « Year » de type long pourrait être integer

**2.1 - Requêtes avec Filtres**

**Rechercher tous les documents contenant un mot spécifique :**

GET chicago\_crimes\_custom/\_search

{

"query": {

"match": {

"Description": "THEFT"

}

}

}

Une image contenant texte, menu, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

**Filtrer les résultats selon une plage de valeurs :**

GET chicago\_crimes\_custom/\_search

{

"query": {

"range": {

"Year": {

"gte": 2016,

"lte": 2017

}

}

}

}

Une image contenant texte, menu, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

**Appliquer un filtre booléen combinant plusieurs conditions ( ‘should’,‘must’) :**

GET chicago\_crimes\_custom/\_search

{

"query": {

"bool": {

"must": [

{ "match": { "Primary Type": "THEFT" }},

{ "term": { "Arrest": false }}

]

}

}

}

Une image contenant texte, menu, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

**2.2 - Agrégations**

**Calculer le nombre total de documents :**

GET chicago\_crimes\_custom/\_count

Une image contenant texte, capture d’écran, Police, nombre

Le contenu généré par l’IA peut être incorrect.

**Effectuer une agrégation par catégories :**

GET chicago\_crimes\_custom/\_search

{

"size": 0,

"aggs": {

"crimes\_par\_type": {

"terms": {

"field": "Primary Type",

"size": 10

}

}

}

}

Une image contenant texte, capture d’écran, menu, Police

Le contenu généré par l’IA peut être incorrect.

**Trouver la valeur moyenne, minimum et maximum d’un champ numérique :**

GET chicago\_crimes\_custom/\_search

{

"size": 0,

"aggs": {

"annee\_stats": {

"stats": {

"field": "Year"

}

}

}

}

Une image contenant texte, Police, capture d’écran, nombre

Le contenu généré par l’IA peut être incorrect.

**2.3 - Recherche Full-Text et Boosting**

**Effectuer une recherche `match` sur un champ textuel :**

GET chicago\_crimes\_custom/\_search

{

"query": {

"match": {

"Description": "AUTOMOBILE"

}

}

}

Une image contenant texte, menu, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

**Comparer une requête ‘match’ et ‘match\_phrase’ :**

GET chicago\_crimes\_custom/\_search

{

"query": {

"match": {

"Description": "AGGRAVATED HANDGUN"

}

}

}

Une image contenant texte, menu, capture d’écran, document

Le contenu généré par l’IA peut être incorrect.

GET chicago\_crimes\_custom/\_search

{

"query": {

"match\_phrase": {

"Description": "AGGRAVATED : HANDGUN"

}

}

}

Une image contenant texte, capture d’écran, menu, Police

Le contenu généré par l’IA peut être incorrect.

**Booster certains champs :**

GET chicago\_crimes\_custom/\_search

{

"query": {

"multi\_match": {

"query": "battery",

"fields": [

"Primary Type^3",

"Description^1"

]

}

}

}

Une image contenant texte, menu, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

**3.1 - Tester les Analyseurs Existant :**

**Test de l’analyseur ‘standard’ :**

GET \_analyze

{

"analyzer": "standard",

"text": "AGGRAVATED BATTERY OF A POLICE OFFICER"

}

Une image contenant texte, capture d’écran, Police, menu

Le contenu généré par l’IA peut être incorrect.

**Test de l’analyseur ‘stop’ :**

GET \_analyze

{

"analyzer": "stop",

"text": "AGGRAVATED BATTERY OF A POLICE OFFICER"

}

Une image contenant texte, Police, menu, capture d’écran

Le contenu généré par l’IA peut être incorrect.

**Test de l’analyseur ‘keyword :**

GET \_analyze

{

"analyzer": "keyword",

"text": "AGGRAVATED BATTERY OF A POLICE OFFICER"

}

Une image contenant texte, capture d’écran, Police, nombre

Le contenu généré par l’IA peut être incorrect.

**3.2 - Création d’un Analyseur Personnalisé :**

**Créer un analyseur personnalisé adapté à votre dataset :**

PUT chicago\_crimes\_smart\_analyzer

{

"settings": {

"analysis": {

"analyzer": {

"description\_smart": {

"type": "custom",

"tokenizer": "standard",

"filter": [

"lowercase",

"stop",

"porter\_stem"

]

}

}

}

},

"mappings": {

"properties": {

"Description": {

"type": "text",

"analyzer": "description\_smart"

}

}

}

}

Une image contenant texte, Police, ligne, capture d’écran

Le contenu généré par l’IA peut être incorrect.

**Tester cet analyseur sur votre dataset :**

GET chicago\_crimes\_smart\_analyzer/\_analyze

{

"analyzer": "description\_smart",

"text": "AGGRAVATED BATTERIES ON A POLICE OFFICER"

}

Une image contenant texte, Police, document, ligne

Le contenu généré par l’IA peut être incorrect.

**Comparer les résultats avec ceux obtenus avant :**

L’analyseur standard conserve tous les mots, alors que l’analyseur personnalisé met en minuscule, enlève les mots inutiles (stopwords), et réduit les mots à leur racine. Cela permet d’avoir plus de résultats pertinents, même avec des variations grammaticales

**Justifier vos choix :**

J’ai choisi de créer un analyseur personnalisé pour le champ "Description" afin d’améliorer la pertinence des recherches. Contrairement à l’analyseur standard, celui-ci met les mots en minuscule, supprime les mots inutiles (comme "of", "a") et réduit les mots à leur racine.

Cela permet de retrouver plus facilement les différentes formes d’un même terme, comme "battery", "batteries" ou "battered", tout en filtrant les éléments peu pertinents. C’est particulièrement utile quand les descriptions sont souvent longues et répétitives.

**4.1 - Comprendre le Score de Pertinence**

Le champ \_score reflète à quel point un document correspond à la requête. Plus le terme est bien placé (exemple : dans un champ boosté comme "Primary Type"), plus le score est élevé.

**4.2 - Calcul du Rappel et de la Précision**

**Calculer la précision et le rappel avant et après optimisation des requêtes**

**Requête initiale :**

GET chicago\_crimes\_custom/\_search

{

"query": {

"match": {

"Description": "THEFT"

}

}

}

Une image contenant texte, capture d’écran, Police, nombre

Le contenu généré par l’IA peut être incorrect.

Cette requête retourne 1439 documents, seuls 777 sont pertinents (662 non pertinent)

Il y a véritablement 4879 documents pertinents existants (vérifié avec les filtres sur le documents CSV via Excel)

On a donc :

Précision = 777 / 1439 ≈ 0.54

Rappel = 777 / 4879 ≈ 0.16

**Requête optimisée :**

GET chicago\_crimes\_custom/\_search

{

"query": {

"multi\_match": {

"query": "THEFT",

"fields": ["Primary Type^3", "Description^1"]

}

}

}

Cette requête retourne 4904 documents, seuls 4879 sont pertinents (25 non pertinent)

Il y a véritablement 4879 documents pertinents existants.

On a donc :

Précision = 4879 / 4904 ≈ 0.99

Rappel = 4879 / 4879 = 1

**Proposer une amélioration pour optimiser ces valeurs**

Pour éviter les faux positifs (comme "identity theft" ou "attempted theft"), on pourrait affiner la requête avec match\_phrase ou encore appliquer un analyzer personnalisé avec suppression des stopwords, etc…