

US Police Shooting

Data Lab

Thème

Le Département de la Justice Américaine vous demande de les aider à comprendre les problèmes qu'il rencontre actuellement avec les services de police du pays.



Contexte

Suite aux récents évènements, la police américaine se trouve accusée de réaliser des discriminations lors de ses différentes interventions.

En effet, la tendance des fusillades mortelles par la police aux États-Unis semble ne faire qu'augmenter, avec en moyenne 1 000 fusillades mortelles par la police. De plus, ce taux semble beaucoup plus élevé pour les Noirs américains que celui de toute autre ethnie.

Le FBI a annoncé son intention de revoir la façon dont il suit les affrontements mortels avec la police.

Source: <https://www.kaggle.com/mrmorj/data-police-shootings>

Objectif

Mise en place d'une solution d'analyse de données

1. Construction d'un Data Warehouse suivant un modèle multi-dimensionnel.
2. Mise en place d'un ETL qui permettra d'ajouter à l'avenir les nouvelles données au fur et mesure.
3. Exploiter le nouveau modèle et ses données afin d'élaborer des rapports permettant de répondre aux questions qui sont posées dans le slide suivant (voire d'autres). Ces rapports peuvent être réalisés à l'aide d'outils tels que Power BI, Qlik Sense, Tableau, etc.

Requirements

1. Aider à comprendre la situation globale des fusillades mortelles réalisées par la police aux États-Unis.
2. Aider à comprendre s'il y a une discrimination d'une certaine ethnie parmi d'autres.
3. Aider à comprendre s'il y a des différences suivant les différentes régions des États-Unis.
4. Aider à comprendre s'il y a une évolution dans le temps.
5. Aider à comprendre si la police américaine réalise des bavures policières.

Sources de données

Une base de donnée de staging comprenant :

- **FatalShooting**
- La liste de toutes les fusillades mortelles par la police aux États-Unis entre 2015 et 2022 avec entre autres la date, le nom, le genre de la victime, son ethnie, son âge, la ville, si la victime était armé, si la victime a essayé de s'enfuir, etc...
- Nombre de lignes : 7 505.
- Source : <https://www.kaggle.com/mrmorj/data-police-shootings>

Sources de données

Mise à disposition d'une base de staging contenant déjà d'autres données

- **USCities**
 - Des informations sur les villes aux États-Unis.
- **Ethnicity**
 - Des informations sur la répartition ethniques aux États-Unis.

Ces sources des données peuvent être consolidées afin, par exemple, de comparer la vitesse de propagation des différentes épidémies.

Sources de données supplémentaires

Il s'agit d'un projet pour lequel une partie des données est fournie, toutefois seule une des sources de données doit être obligatoirement utilisée et ce, afin que tous les stagiaires partent d'un socle commun.

Les stagiaires sont absolument libres de rajouter d'autres sources de données et de les explorer au gré de leurs envies.

Délivrables

1. Backup de votre Data Warehouse.
2. Fichier de reporting Power BI, Qlik Sense, Tableau, etc.
3. Slides de présentation expliquant votre démarche et vos résultats.
 - Comment vous avez choisi de traiter le problème.
 - Quelles sont les données que vous avez choisies ?
 - Quels sont les constats que vous avez pu mettre en évidence.
 - Quelles sont les conclusions que vous pouvez apporter au sujet.