

POLICE REPORT



Case No : 20202024

Date : 19.04.2024

Prepared by : Lysa HALLI, Michelle TOBING, Massyle DEKKAR.

Incident : étude de la criminalité dans la ville de Los Angeles.

Table des matières

La base de données :	3
Sources :	3
Description de la base de données	3
Prétraitement des données	4
Analyse des données :	5
Visualisation des données	5
Analyse statistique :	7
Analyse descriptive	8
Conclusion :	11
Table des illustrations :	12

Introduction :

La criminalité est un phénomène qui affecte les sociétés à travers le monde entier. Elle se définit comme l'ensemble des actes illégaux commis par un individu ou un groupe de personne, ça inclut des infractions mineures allant jusqu'à des crimes graves. Il constitue une préoccupation majeure au vu de ses lourdes conséquences qui touchent les individus mais induits également des coûts élevés pour les gouvernements ainsi que des dommages pour la sécurité, la stabilité et la prospérité d'une société. Des débats ont lieu depuis des décennies pour comprendre ce qui pousse un individu à commettre un acte criminel, mais les causes sont multiples allant de facteurs individuels, sociaux, économiques à culturels et politiques.

Afin de mettre en place des mesures pour réduire le taux de criminalité, les gouvernements, les forces de l'ordre doivent être dans la capacité d'identifier, de prévenir la criminalité sous toutes ses formes. C'est dans ce contexte que le Data Analyst joue un rôle important. En effet, l'analyse de données permet d'identifier les types de crimes, les zones à risque, les caractéristiques des victimes ainsi que les facteurs qui contribuent à ce phénomène. Et ainsi aider à comprendre les tendances de la criminalité et à utiliser les approches adéquates et efficaces afin d'y lutter.

C'est en prenant conscience de l'importance accrue du rôle d'un data analyst dans ce domaine qu'on s'est penché sur l'analyse d'une base de données recensant les crimes commis dans la ville de Los Angeles aux USA de 2020 à 2024.

La base de données :

Sources :

La première étape de notre analyse consiste à vérifier la fiabilité des sources de nos données. Les informations que nous utilisons proviennent de Data Lacity, un portail mis en place par la ville de Los Angeles. Ce site permet la transmission de données collectées par chaque département municipal, facilitant ainsi l'analyse publique et la visualisation de ces informations. Ces données nous ont été fournies par le Los Angeles Police Department (LAPD), ce qui en garantit la fiabilité.

Description de la base de données

Notre base de données répertorie 925 720 crimes enregistrés de 2020 à ce jour et comprend 28 colonnes. Ces colonnes fournissent des informations cruciales sur chaque crime, telles que la description de l'acte, la date et l'heure, le lieu, l'âge de la victime, son ethnie, etc. Nous avons également constaté de nombreuses valeurs manquantes au sein de la base.

Prétraitement des données

L'étape de prétraitement est essentielle pour assurer la qualité et la cohérence des données avant l'analyse. Nous avons adopté plusieurs mesures pour préparer notre ensemble de données :

- Gestion des valeurs manquantes :

Face à de nombreuses valeurs manquantes, nous avons procédé à l'imputation de nouvelles valeurs. Pour les "Mocodes", qui sont les codes décrivant le mode opératoire du criminel, nous avons attribué la valeur "1501 other MO" lorsque aucun code spécifique ne pouvait être appliqué. Cette approche a été répliquée pour les catégories d'armes, ainsi que pour le sexe et l'ethnie de la victime, en conformité avec les normes légales.

- Élimination des données non pertinentes :

Nous avons supprimé des colonnes et des lignes qui étaient superflues pour focaliser notre analyse sur les données les plus pertinentes. Toutes les lignes où la colonne « Premis desc », qui détaille le lieu précis du crime, était vide ont été éliminées. Nous avons également retiré les colonnes "crm_cd" de 1 à 4, jugées peu utiles, car une autre colonne fournissait déjà une description adéquate des crimes.

- Catégorisation des crimes :

Nous avons classé les crimes en catégories selon les normes de l'Uniform Crime Reporting (UCR), comme les homicides, les viols ou les vols, chacun associé à un code criminel spécifique. Cette structuration rend notre base de données plus organisée et plus accessible

Il est bon d'expliquer certaines de ces catégories :

- BTFV désigne les cambriolages qui ont lieu dans des véhicules mais qui n'implique pas le vol du véhicule en lui-même (vol de biens dans le véhicule, dégradation du véhicule)
- MVT (ou GTA) représente les vols de véhicules.
- Vol personnelle représente les vols qui touche directement la victime (pickpocket (vol à la tire), vol à l'arraché, vol sur une personne ivre ou inconsciente)
- Agression simple désigne généralement une attaque physique contre une autre personne où l'intention est de causer des dommages ou des blessures mais sans l'usage d'armes et sans causer de blessures graves.
- Agression aggravée est comme une agression simple mais avec usage d'arme et peut donc causer des blessures graves.
- Le cambriolage consiste à entrer illégalement dans un bâtiment ou un espace clos (résidence, entreprise, entrepôt, etc.) avec l'intention de commettre un crime à l'intérieur. Il y a forcément usage de la force car le criminel doit briser une porte ou une fenêtre ou quoique ce soit qui s'en apparente pour pouvoir accéder à l'espace clos. Le vol lui désigne le vol simple, comme dit définit dans la loi comme l'action de prendre illégalement les biens d'autrui sans l'usage de la force.

Mais nous avons remarqué que nous pouvions encore généraliser les types de crimes. Nous pouvons créer des catégories qui regroupent les catégories précédentes selon des similarités plus globales. Les catégories précédentes permettent de généraliser les codes de crimes qui sont souvent trop spécifiques pour une analyse réellement pertinente, ce qui nous a valu de faire une catégorisation de ces codes. Mais nous pouvons voir que certaines catégories possèdent des similarités entre elles, notamment des similitudes plus générales telles que les crimes d'agression (assaults) et les crimes impliquant des vols (larceny). Nous allons donc effectuer une catégorisation des types précédents. L'objectif sera d'analyser des tendances plus générales.

Analyse des données :

Visualisation des données

Avant de commencer une analyse approfondie de la base de données, il est important de visualiser les données afin de révéler des tendances ou des schémas qui se répètent. Pour cela on a tracé plusieurs diagrammes qu'on va vous présenter.

- Nombres annuels de crimes :

Ce diagramme en barre illustre le nombre de crimes commis dans la ville chaque année de 2020 à 2024.

On remarque une nette augmentation depuis 2020 avec un pic notable en 2022. Le nombre de crimes, plus bas en 2020 par rapport aux autres années, peut s'expliquer par la crise COVID et les confinements mis en place à cette période. Le pic en 2022 correspond justement à la fin de la crise.

En 2024, on note une chute significative cela est dû à un manque de données car l'année n'est pas encore finie.

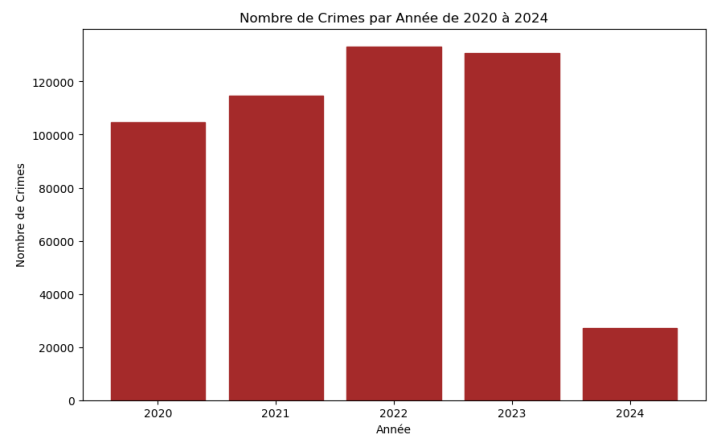


Figure 1: Nombre de crimes par années

- Tendance des crimes par mois :

Pour visualiser cette tendance temporelle plus en détails on a tracé un diagramme linéaire qui représente l'évolution du nombre de crimes par mois pendant ces quatre années. Chaque point sur le graphique correspond à un mois.

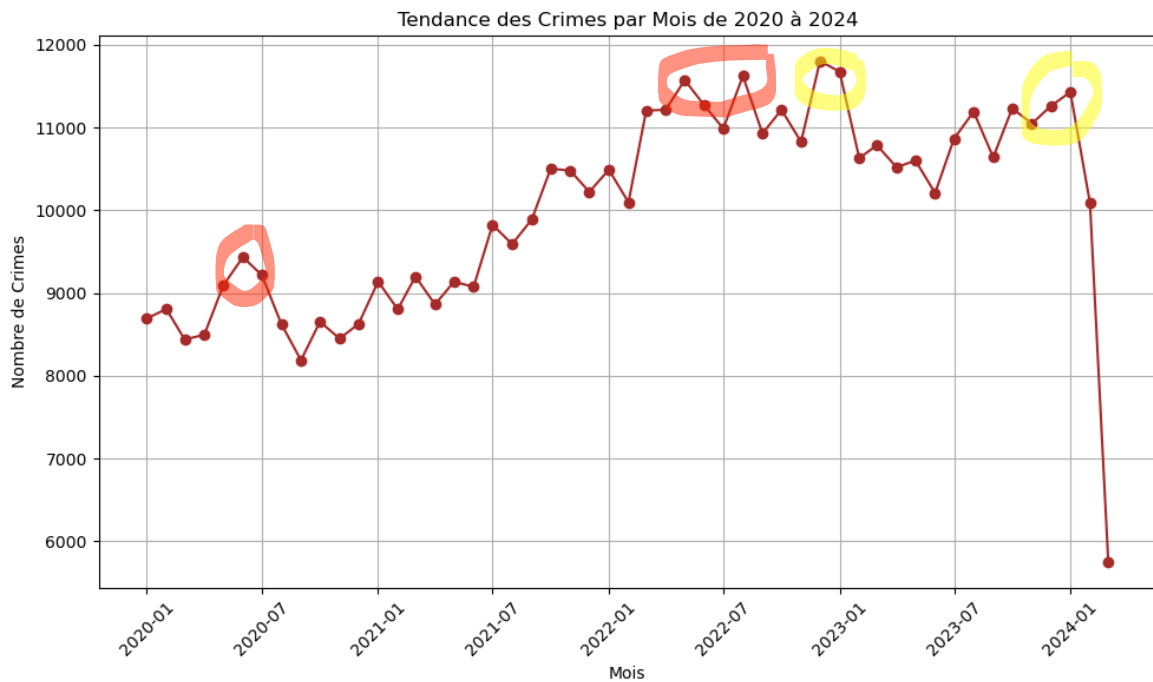


Figure 2: évolution des crimes par mois

Cette figure confirme la tendance observée plus haut quant à l'évolution positive du nombre de crimes dans la ville, on remarque ici que cette évolution s'est faite progressivement.

On peut également relever une certaine saisonnalité avec des pics récurrents. Si l'on observe attentivement on peut remarquer que les pics présents coïncident avec des moments spécifiques de l'année. En effet, ils s'observent soit en **fin d'année** (entre novembre et début janvier), on peut supposer que c'est durant les fêtes de Noël et du nouvel an, soit durant **la haute saison** (entre mai et juillet) autrement dit pendant les vacances d'été.

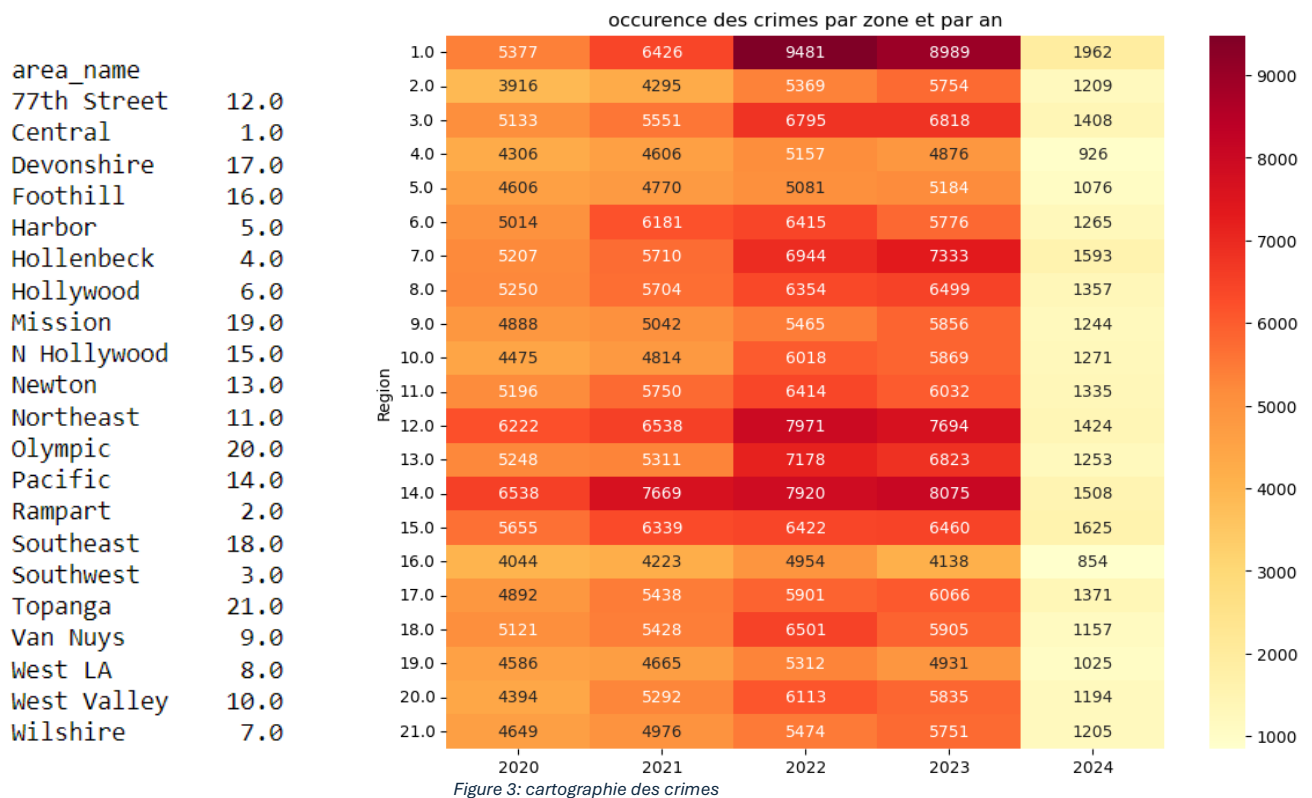
L'augmentation de la criminalité durant l'été peut s'expliquer par le fait que les gens ont tendance à passer plus de temps à l'extérieur, ce qui peut augmenter les opportunités pour les crimes tels que les vols, cambriolages ou les agressions. En fin d'année, durant les fêtes, les crimes ont tendance à augmenter pour plusieurs raisons, on peut présumer la consommation d'alcool ou de drogues qui conduits à des comportements violents, les déplacements des locaux qui constituent également une opportunité pour les vols.

- Cartographie des crimes :

Pour visualiser les crimes par zones et par an, on a décidé de créer une heatmap (carte de chaleur). Les données sont présentées sous forme de tableau avec des couleurs différentes. Les zones chaudes (valeurs plus élevées) sont de couleurs plus foncées, et les couleurs plus claires (jaune) représentent les valeurs plus faibles.

Cette figure va permettre d'identifier directement les zones chaudes où la criminalité est plus importante ce qui va permettre d'allouer les ressources et de mettre en place des mesures de manière plus efficace.

Sur la heatmap (figure de droite) les zones sont identifiées par des chiffres allant de 1 à 21. Pour une meilleure lisibilité, on a affiché à côté à quelles zone (quartier) correspond chaque chiffre.



On voit que dans chaque quartier la criminalité a augmenté au fil des années (les couleurs deviennent plus foncées). Par exemple à Rampart on passe de 3916 crimes en 2020 à 5754 crimes en 2023 soit une augmentation de 46.94% en seulement 3 ans.

On observe également une très grande variabilité entre les quartiers. Certains sont plus touchés par le phénomène que d'autres. Par exemple à Central et Pacific, où on enregistre le plus grand nombre de crimes (notamment en 2022 à Central). Cependant à Pacific l'augmentation de la criminalité s'est faite progressivement, alors qu'à Central en 2 ans (2020-2022) il y a eu une augmentation spectaculaire de 76,32% ce qui peut nous laisser penser à des changements dans le quartier ou des facteurs particuliers qu'il faut étudier.

En revanche, il y a également des quartiers plus sûrs comme Foothill, où le taux de criminalité a évolué de manière beaucoup moins significative et est même en baisse en 2023.

Analyse statistique :

Pour l'analyse statistique la seule variable exploitable et pertinente à étudier est vict_age qui répertorie l'âge des victimes.

	time_occ	AREA	rpt_dist_no	vict_age	premis_cd	weapon_used_cd	LAT	LON	Year	Month
count	925161.000000	925161.000000	925161.000000	925161.000000	925161.000000	925161.000000	925161.000000	925161.000000	925161.000000	925161.000000
mean	1337.673727	10.705659	1117.003261	29.566135	306.549140	452.902912	33.990424	-118.064710	2021.730794	6.290639
std	652.570756	6.101801	610.198071	21.862786	217.445993	97.367256	1.687170	5.848586	1.215159	3.506723
min	1.000000	1.000000	101.000000	-4.000000	101.000000	101.000000	0.000000	-118.667600	2020.000000	1.000000
25%	900.000000	6.000000	615.000000	0.000000	101.000000	400.000000	34.014500	-118.429900	2021.000000	3.000000
50%	1417.000000	11.000000	1141.000000	31.000000	203.000000	500.000000	34.058700	-118.322200	2022.000000	6.000000
75%	1900.000000	16.000000	1617.000000	45.000000	501.000000	500.000000	34.163900	-118.273900	2023.000000	9.000000
max	2359.000000	21.000000	2199.000000	120.000000	971.000000	516.000000	34.334300	0.000000	2024.000000	12.000000

Tableau 1: analyse statistique

On peut voir que l'âge moyen des victimes est de 29.6 ans ce qui est relativement jeune.

Le médian indique l'âge qui sépare la moitié des victimes en deux groupes égaux. Dans ce cas, le médian est de 31 ans, ce qui signifie que la moitié des victimes ont moins de 31 ans et l'autre moitié ont plus de 31 ans.

L'écart type étant 21.9, signifie une grande variabilité dans les âges des victimes. Les incidents touchent généralement des victimes âgées de 7 à 51 ans.

Analyse descriptive

Dans cette section nous allons analyser plus en profondeur notre base de données.

Premièrement, on va s'intéresser au le type de crimes le plus fréquent. Tout d'abord on crée un tableau qui résume le nombre d'incidents pour chaque type de crime et leur proportion. On le visualise ensuite dans un histogramme, offrant une vision claire de la représentation proportionnelle de chaque type de crime.

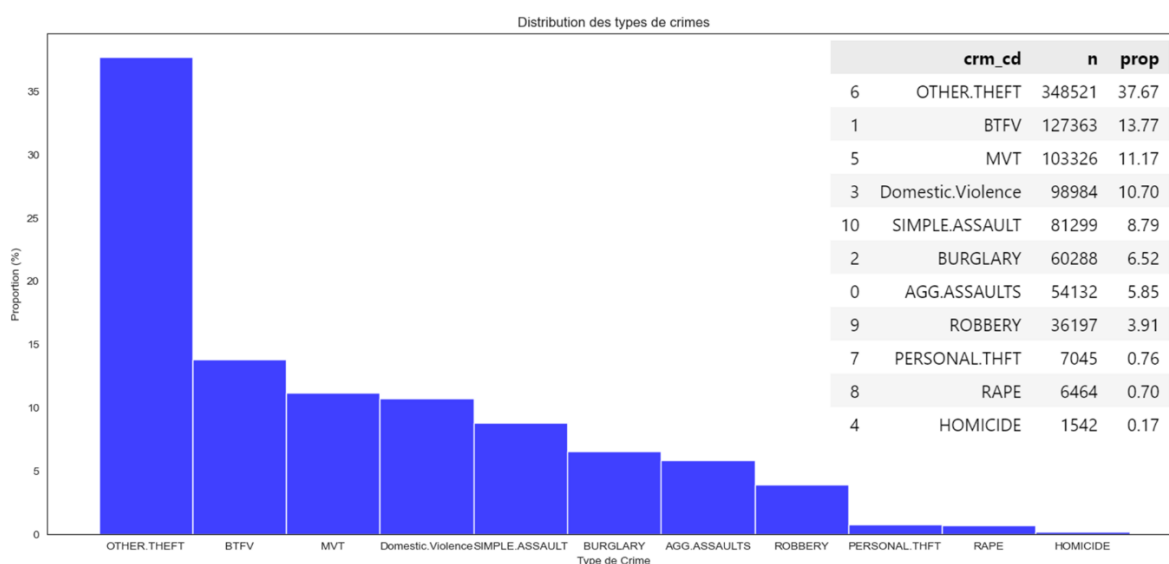


Figure 4: types de crimes

On peut voir que *Other Theft* prend la plus grande proportion avec 38%, cela signifie que les crimes de vol, *shoplifting*, vol de vélo etc., c'est-à-dire les crimes mineurs dominent les incidents criminels. Les autres types de crimes qui ont une proportion plus de 10% sont BTFV (*Burglary Theft from Vehicle*) ou cambriolage depuis un véhicule à 14%, MVT (*Motor Vehicle Theft*) ou vol de véhicules à 11%, et violence domestique à 10.7%.

On va catégoriser les types de crimes en *larceny* s'il s'agit de vol et de cambriolage ou en *assault* s'il s'agit d'agression, de violation et d'attaques physiques. Cette opération va nous permettre de déterminer quelle catégorie de crime est plus répandue.

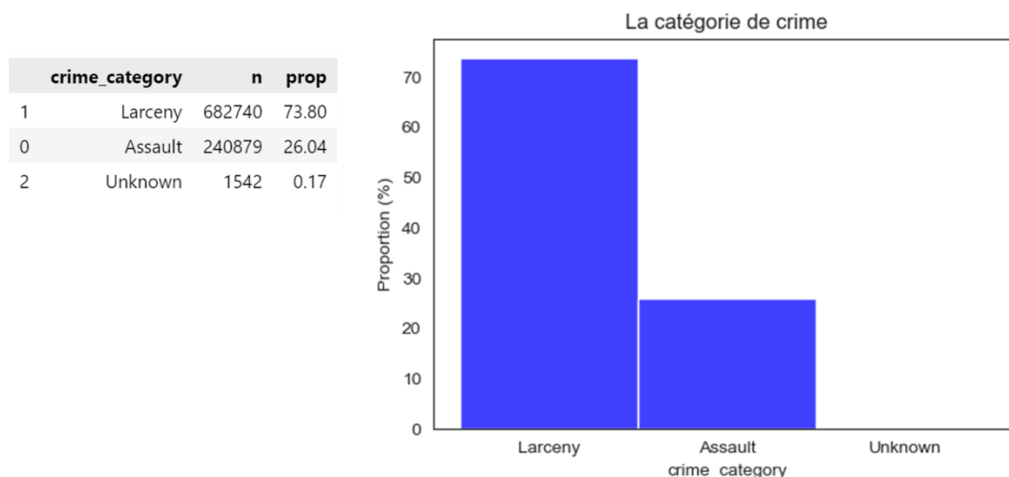


Figure 5: catégorie des crimes

L'histogramme nous montre que la majorité de crimes sont celles de catégorie *larceny*.

- Crimes par zone :

Ensuite, on va étudier les incidents criminels par zone pour démontrer quelle zone est plus touchée par ce phénomène et quelle catégorie de crime est plus répandue.

On peut voir que la zone avec le plus nombre de crimes est la zone Central, et le deuxième est la zone 77th Street. On observe également que pour toutes les zones la majorité de crimes sont *larceny*. Néanmoins on peut constater que la zone de 77th Street a la plus nombre d'*assault* par rapport aux autres zones, même plus que celles de la zone Central, on peut tenter d'expliquer ça par le fait que 77th Street fait partie du sud de Los Angeles, une région caractérisée par une forte densité de population ce qui la rend sujette à un taux de criminalité plus élevé en raison de la proximité et de la fréquence des interactions entre les personnes.

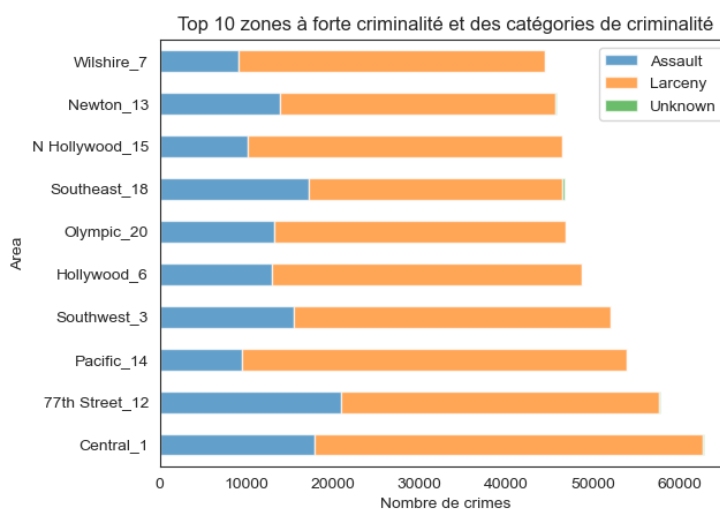


Figure 6: top 10 zones à forte criminalité

- Age des victimes :

Le pie chart nous permet de visualiser la tranche d'âge des victimes. On constate que plus de moitié de victimes sont adultes.

La deuxième tranche d'âge la plus important est « Unknown Age ». En effet, 25.3% des victimes, on ne connaît pas leur âge car non renseigné dans la base de données. Néanmoins, on a choisi de garder les lignes avec cette donnée manquantes car il y a environ 200k de données qui seront perdues si l'on décide de les supprimer. On peut ainsi perdre des données qui sont pertinentes pour d'autres analyses (type de crimes etc).

On peut néanmoins constater qu'après les adultes, les victimes sont les jeunes à 10.7%, les seniors à 5.7% et les enfants à 1.4%.

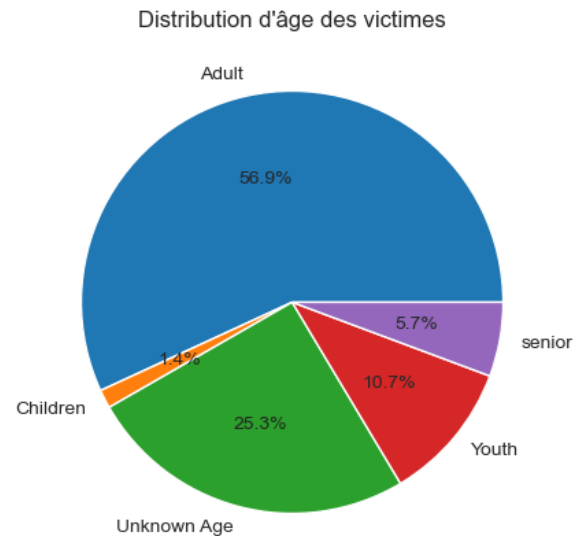


Figure 7: distribution de l'age des victimes

On visualise ensuite la répartition des âges des victimes dans les zones pour voir comment les profils d'âge des victimes peuvent varier selon les zones.

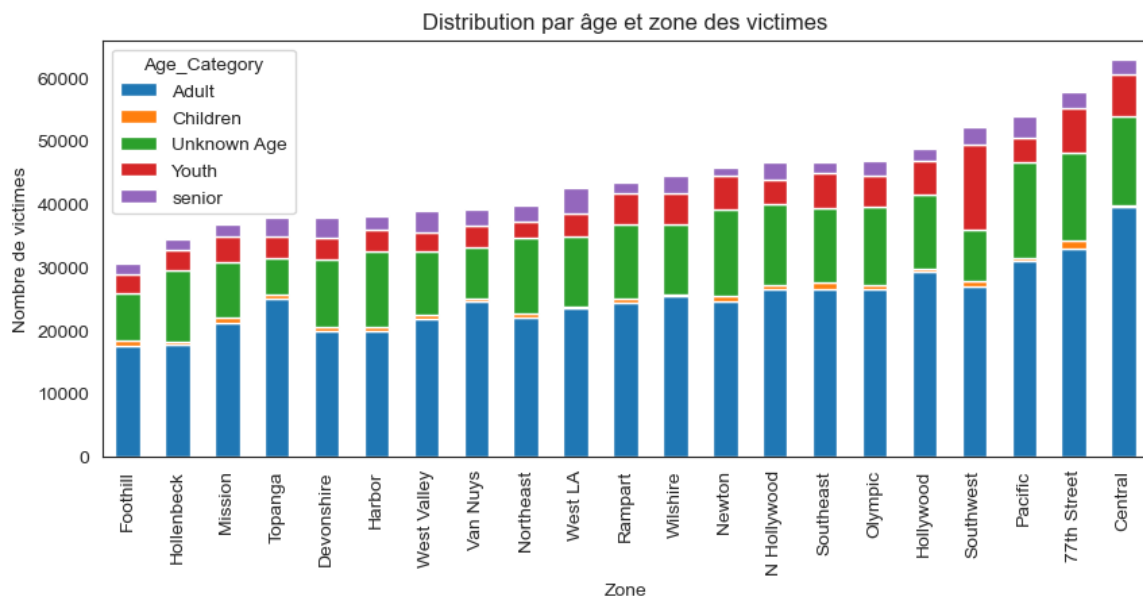


Figure 8: distribution par age et zone des victimes

On peut voir que tous les zones ont environ la même proportion d'âge, la zone Southwest où la proportion de victimes jeunes est beaucoup plus importante que dans les autres zones. Cela peut s'expliquer par le fait que cette zone a une population plutôt jeune. En effet, elle comporte des quartiers comme Playa Vista qui attirent des jeunes par leur proximité avec des entreprises de tech et son ambiance urbaine moderne. De plus, elle est proche d'universités réputées (comme l'Université de Californie du Sud).

- Origine des victimes :

Il est intéressant d'étudier l'origine ethnique des victimes. On peut constater que la plus grande proportion de 30% sont des victimes *Hispanic/Latin/Mexican*, le deuxième c'est inconnu et on le garde pour la même raison de celui de l'âge, suivi par les victimes *White* à 20%, les victimes *Black* à 14% et les autres descendance prennent une proportion inférieure à 10%.

Cela peut s'expliquer par plusieurs facteurs. Premièrement les groupes ethniques qui sont économiquement désavantagés peuvent être plus susceptibles de vivre dans des zones avec un taux de criminalité plus élevé. De plus, certains groupes ethniques peuvent être ciblés pour des crimes en raison de stéréotypes, de discrimination ou de racisme.

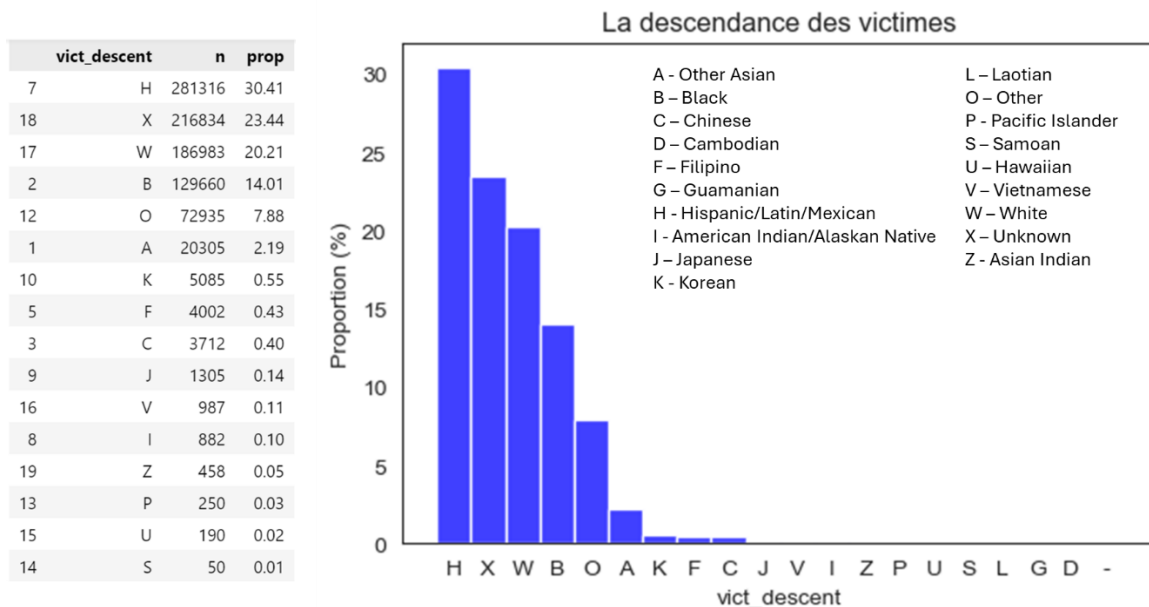


Figure 9: origine ethnique des victimes

Conclusion :

À travers cette étude approfondie des données criminelles de la ville de Los Angeles de 2020 à 2024, nous avons pu mettre en lumière plusieurs aspects cruciaux de la criminalité urbaine. Notre analyse, basée sur une base de données exhaustive fournie par le Los Angeles Police Department, a révélé des tendances et des schémas spécifiques. En effet, Nous avons constaté une augmentation notable des crimes entre 2020 et 2022, avec un pic durant la reprise post-COVID. L'analyse démographique a mis en évidence que les Hispaniques/Latins/Mexicains représentent la plus grande proportion des victimes, ce qui pourrait refléter à la fois la composition démographique de certaines zones et des facteurs socio-économiques plus larges. En outre, les jeunes adultes semblent être les plus affectés. Par ailleurs, la cartographie des crimes et l'analyse par zone ont permis de détecter les "points chauds" de criminalité, facilitant ainsi l'allocation plus ciblée des ressources policières et de prévention.

Cette analyse souligne également l'importance du rôle des data analysts dans la compréhension de la criminalité urbaine. Une telle étude pourrait guider les politiques publiques afin d'améliorer la sécurité et la qualité de vie dans les zones urbaines.

Table des illustrations :

Figure 1: Nombre de crimes par années	5
Figure 2: évolution des crimes par mois	6
Figure 3: cartographie des crimes.....	7
Figure 4: types de crimes.....	8
Figure 5: catégorie des crimes	9
Figure 6: top 10 zones à forte criminalité.....	9
Figure 7: distribution de l'age des victimes	10
Figure 8: distribution par age et zone des victimes	10
Figure 9: origine ethnique des victimes	11