

DEVELOPPEUSE DATA

2020/2021



L'impact de la Vidéoprotection sur le choix Immo à **PARIS**



Michelle Saypraseut

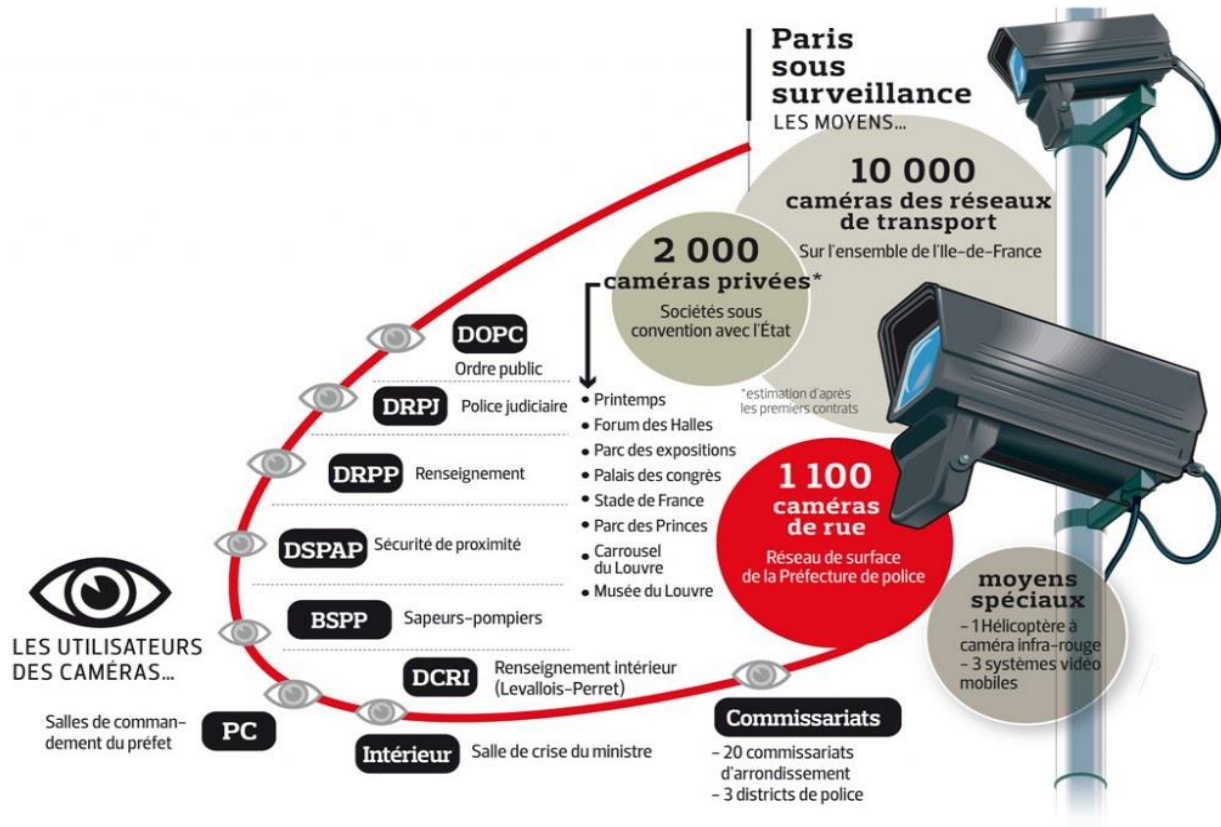


TABLE DES MATIERES

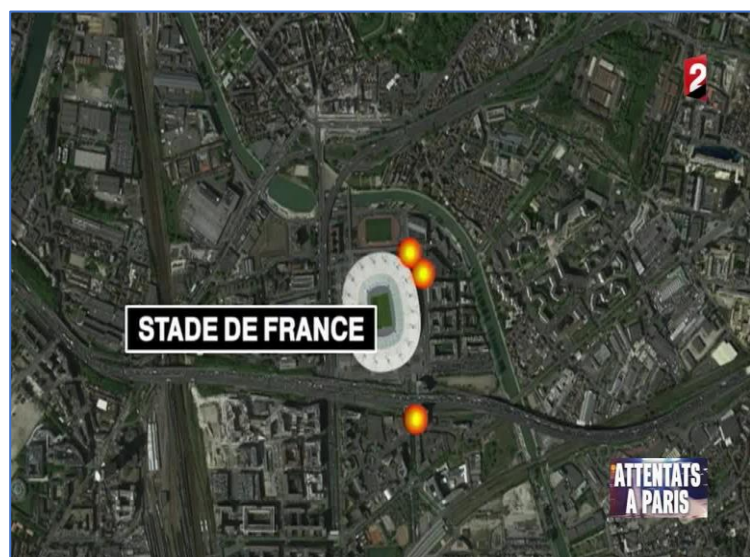
1. PRESENTATION DU PROJET	3
2. CAHIER DES CHARGES	5
2.1. CONTEXTE	5
2.2. STRATEGIE	5
2.3. OBJECTIF	6
3. GESTION DE PROJET & STACK TECHNIQUE	6
4. LES DONNEES	7
4.1. COLLECTE DES DONNEES	7
4.2. NETTOYAGE DES DONNEES	7
4.3. DESCRIPTION DES DONNEES	9
5. MODELISATION DE LA BASE DE DONNEE	11
5.1. MODELE CONCEPTUEL DES DONNEES (MCD)	12
5.2. MODELE LOGIQUE DES DONNEES (MLD)	13
5.3. MODELE PHYSIQUE DES DONNEES (MPD)	14
6. CREATION DE LA BASE DE DONNEES	15
6.1. CREATION DES DIFFERENTES TABLES	15
6.2. CONNEXION A LA BASE DE DONNEE	17
6.3. IMPORTATION DES DONNEES	17
7. SECURISATION & SAUVEGARDE	18
8. REQUETAGE & OPTIMISATION	19
9. ANALYSE & DATA VISUALISATION	20
10. CONCLUSION	35
11. REMERCIEMENTS	36
12. RESSOURCES	36

1. PRESENTATION DU PROJET

LA VIDEOPROTECTION, CET OUTIL PRIMORDIAL DONT ON SE MÉFIE



La vidéoprotection fait partie de nos vies depuis déjà plusieurs années, conscient de son impact sur la résolution des tristement célèbres affaires de terrorisme, à l'instar des attentats du Stade de France survenus le 13 Novembre 2015 ou encore de l'arrestation des frères Kouachi après les attentats de Charlie Hebdo.



Elle avait alors permis l'identification de la direction prise par les terroristes pour fuir.

Par ailleurs, elle a aussi pu prouver son efficacité dans des affaires de grand banditisme tels que dans l'arrestation des braqueurs de la bijouterie CHAUMET Place Vendôme (« le braqueur à trottinette »), son parcours a pu être refait entièrement grâce aux enregistrements des caméras de vidéoprotection.



D'autres affaires d'ordre politique ont éclaté grâce à la vidéoprotection comme l'affaire Benalla, par exemple...



Pourtant, nous en savons peu sur son usage réel au quotidien. De fait, nous connaissons mieux son utilité pour résoudre les affaires de terrorisme, de grand banditisme ou encore pour combattre les infractions de la route. Mais qu'en est-il de l'impact sur nos vies en tant que simple citoyen ?

Le monde tend à changer et aux vues des avancées technologiques telles que la reconnaissance faciale, thermique etc., il ne me semble pas invraisemblable de penser qu'un arrondissement citadin doté d'une bonne infrastructure de vidéoprotection pourrait devenir un gage de sécurité dans le futur...

Ainsi, le but de cette étude est de proposer l'usage de la vidéoprotection comme arme contre l'insécurité, ce qui, finalement, influencera nos choix en termes d'immobilier dans la capitale parisienne.

2.1. CONTEXTE

- **Enjeux :** *Mettre en exergue l'impact de la vidéoprotection sur la sécurité des quartiers Parisiens, à travers les chiffres relatifs aux délits recensés par la Police Nationale et pouvoir par conséquent, aider tout un chacun à choisir son lieu d'habitation en fonction de sa situation personnelle .*
- **Domaine d'application :** Paris intra-muros

2.2. STRATEGIE

Afin d'en savoir davantage, nous allons répondre à différentes questions , pour se faire nous allons rendre visuel un certain nombre de donnée :

FOCUS sur les DÉLITS recensés par la Police Nationale

- 1. Quels sont les délits traités dans cette analyse?
- 2. Quel est le top 10 des types de délits les plus fréquents?
- 3. Quel est le nombre de délits par arrondissement ?
- 4. Quel est le nombre de délits par année ?
- 5. Quel est le nombre de cambriolage par arrondissement ?

FOCUS sur la VIDEOPROTECTION

- 6. Quels sont les arrondissements les mieux dotés en caméra ?
- 7. Quelle est la fréquence d'installation de la vidéoprotection ?
- 8. Quel est le nombre de caméra par arrondissement en fonction de la superficie?
- 9. Quel est le nombre d'habitant par camera ?
- 10. Qui protège réellement la vidéo protection ?

FOCUS sur l'IMMO

- 11. Quels sont les arrondissements où les transactions sont les plus fréquentes?
- 12. Quel est le top 5 des types de biens les plus vendus?
- 13. Quel est la valeur foncière moyenne par arrondissement?
- 14. Quel est la proportion de transaction concernant les maisons à Paris?

GENERAL

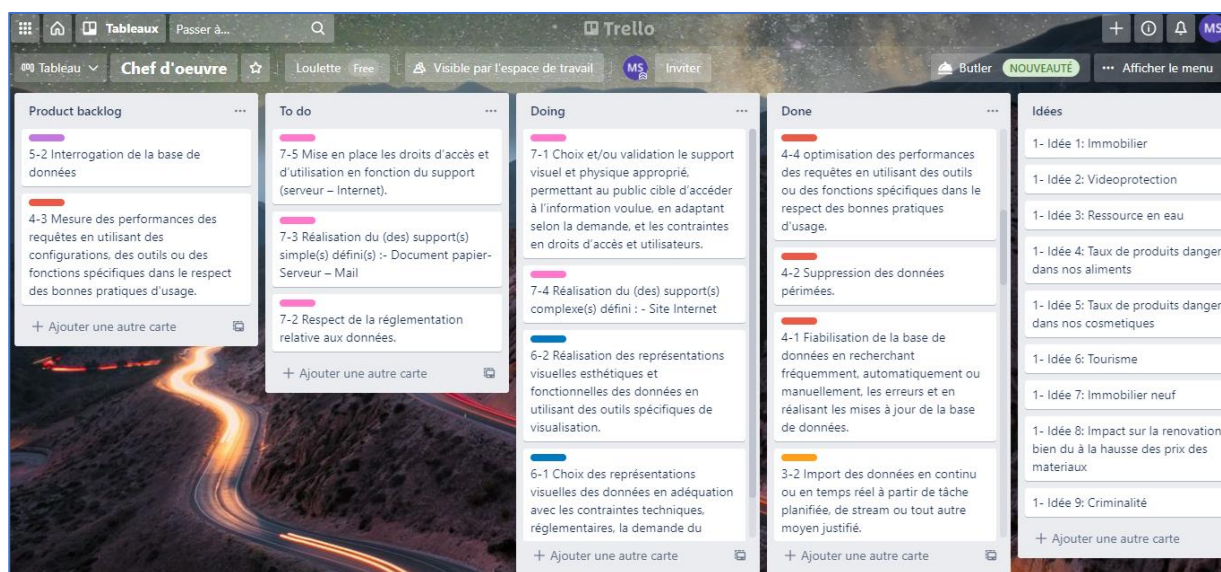
- Quels arrondissement semble offrir le meilleur compromis ?

2.3. OBJECTIF FINAL










A l'issu de cette analyse, il s'agira de proposer aux femmes et hommes, un outil d'aide à la décision pour le choix d'un futur achat dans les différents arrondissements parisiens.

3. GESTION DE PROJET & STACK TECHNIQUE

Afin d'organiser les différentes taches de ce projet, nous avons utiliser l'outil Trello, qui a pour particularité sa praticité en fonctionnant selon les méthodes agiles, ainsi pour arriver au bout de ce projet, la méthode kanban semblait être un choix judicieux.



<https://trello.com/b/f7aqVLmn/chef-doeuvre>

Langages/Outils	Utilisations
Python 	Langage orienté objet pour traiter la Data et développer des applications
SQL 	Langage de programmation permettant d'interroger les données
Jupyter Notebook 	Application web pour exécuter du code : sert aux chargement, exploration, traitement et nettoyage des données
MySQL 	SGBDR : Système de Gestion de Bases de Données Relationnelles, qui stocke des données de façon organisée et cohérente
MySQL Workbench 8.0 	IDE SQL pour la conception de la base de donnée, la visualisation et administration de la base de données
Power BI 	Visualisation des données (Data Visualisation)
Trello 	Gestion de projet, permet de visualiser les tâches selon les méthodes agiles.
Notion 	Gestion de projet, permet de centraliser son travail.
Github 	Service web d'hébergement et de gestion de développement de logiciels.

4. LES DONNEES

4.1. COLLECTE DES DONNEES

Les données constituant la database proviennent des sources suivantes :

- <https://www.data.gouv.fr/>
- <https://www.insee.fr/>

Afin de constituer cette base de donnée (BDD), 36 fichiers CSV ont été nécessaires :



EMPLACEMENT D'IMPLANTATION DE CAMERAS DE "VIDEOPROTECTION"

- **Format** : CSV
- **Contenu** : Emplacement de 707 caméras géo-tagger (sur 1424 en cours)
- **Détail des colonnes** : 7 colonnes
- <https://www.data.gouv.fr/fr/datasets/emplacements-dimplantation-de-cameras-de-videoprotection-bo-ville-de-paris-du-01-02-2019/>



DEMANDE DE VALEURS FONCIERES

- **Format** : CSV
- **Contenu** : Le présent jeu de données « Demandes de valeurs foncières », publié et produit par la direction générale des finances publiques, permet de connaître les transactions immobilières intervenues au cours des cinq dernières années sur le territoire métropolitain et les DOM-TOM, **à l'exception de l'Alsace, de la Moselle et de Mayotte**. Les données sont issues des actes notariés et des informations cadastrales de 2016 à 2020 (5 fichiers CSV)
- **Détail des colonnes** : 43 colonnes
- <https://www.data.gouv.fr/fr/datasets/demandes-de-valeurs-foncieres/>



CRIMES ET DELITS ENREGISTRES PAR LES SERVICES DE GENDARMERIE ET DE POLICE DEPUIS 2012

- **Format** : XLSX
- **Contenu** : Ces données constituent une actualisation des volumes Criminalité et délinquance constatées en France édités par la Direction centrale de la police judiciaire, publiés jusqu'en 2013 à la Documentation française, et disponibles sur son site [internet](#). Elles contiennent des informations, de caractère administratif, sur l'activité judiciaire des services de gendarmerie et de police, y compris celles des DOM-TOM, depuis 2012 à 2020 (9 fichiers XLSX)
- **Détail des colonnes** : 803 colonnes
- <https://www.data.gouv.fr/fr/datasets/crimes-et-delits-enregistres-par-les-services-de-gendarmerie-et-de-police-depuis-2012/>

- **Format** : CSV
- **Contenu** : Délimitation des arrondissements : le décret impérial du 1er novembre 1859 fixe les dénominations des vingt arrondissements de la ville de Paris. De Paris 1 à Paris 20 (20 fichiers CSV)
- **Détail des colonnes** : 10 colonnes
- <https://www.data.gouv.fr/fr/datasets/arrondissements-1/>

- **Format** : XLSX
- **Contenu** : Population par sexe et âge regroupé
- **Détail des colonnes** : 4 colonnes
- <https://www.insee.fr/fr/statistiques/5397443?sommaire=5397467&geo=COM-75101>

4.2. NETTOYAGE DES DONNEES

Avant de procéder à l'insertion des données dans la base, une grosse partie du projet est consacré au nettoyage de la donnée.

Pour se faire, j'ai importé les fichiers CSV sur Jupyter Notebook et utilisé la librairie Pandas.

Cette opération est cruciale pour éliminer les données erronées et rectifier les éventuelles erreurs.

Voici les différentes étapes de nettoyage effectuées :

- Supprimer les données superflues (doublon, NaN) et les valeurs aberrantes,
- Homogénéisation des informations : vérifier que le même nom (respect de la casse) et le même nombre de commune étaient présent dans toutes les tables car les deux fichiers contenaient des noms de commune différentes,
- Ajouter les valeurs manquantes : remplacer les NaN par des 0 ou autres quand cela est possible,
- Adapter les données à une structure standard :
 - Harmoniser le nom des colonnes
 - Normalisation des types de colonne : Transformer les « FLOAT » en « INT »
 - Enlever les « , », les « ; », les « « » et autres espaces inutiles,
- Réorganiser les colonnes,
- Concaténation des fichiers similaires après l'insertion d'une colonne « année » ou « code-postal »

- Sauvegarder une copie puis exporter les données en CSV.

Après leur nettoyage, les données doivent être validées, c'est-à-dire que l'on doit déterminer si des erreurs se sont produites dans le processus de préparation des données jusqu'à ce point et les corriger le cas échéant.

On obtient à la fin 9 fichiers de données prêts à être injectés dans les tables.

4.3. DESCRIPTION DES DONNEES DE MES DIFFERENTES TABLES

Table « immo_type »

code_type_local	type_local
0	2
1	4



Données	Type	Description
code_type_local	INT	Index
type_local	VARCHAR	Type de local

Table « annee »

code_annee	an
0	0
1	1



Données	Type	Description
code_annee	INT	Index de l'année
an	INT	Année

Table « commune »

code_postal	commune	nom_commune	superficie	lati	longi
0	75005	PARIS 05	Panthéon	2.539375e+06	48.844443
1	75013	PARIS 13	Gobelins	7.149311e+06	48.828388



Données	Type	Description
code_postal	INT	Index et code postal
commune	VARCHAR	Numéro d'arrondissement
nom_commune	VARCHAR	Nom de l'arrondissement
superficie	FLOAT	Superficie en m²
lati	FLOAT	Latitude (Coordonnées GPS)
longi	FLOAT	Longitude (Coordonnées GPS)

Table « delits »

code_delits		type_delits
0	2	Homicides pour voler et à l'occasion de vols
1	3	Homicides pour d'autres motifs



Données	Type	Description
code_delits	INT	Index du délit
type_delits	VARCHAR	Intitulé du délit

Table « population »

code_population		age	hommes	femmes	ensemble	code_postal
0	0	Moins de 3 ans	193	184	377	75001
1	1	3 à 5 ans	161	158	319	75001



Données	Type	Description
code_population	INT	Index de la population
age	INT	Ages des individus
hommes	INT	Sexe masculin
femmes	INT	Sexe féminin
ensemble	INT	Femmes et hommes confondus
code_postal	INT	Code postal de l'arrondissement

Table « immo_biens »

code_biens	n_voie	type_voie	voie	surface_terrain	surface_m2	nb_pieces	code_postal	code_type_local	
0	0	14	RUE	DE LARCADE	0	24	1	75008	2

Données	Type	Description
code_biens	INT	Index des biens
n_voie	INT	Numéro de voie
type_voie	VARCHAR	Type de la voie
voie	VARCHAR	Nom de la voie
surface_terrain	INT	Surface du terrain
surface_m²	INT	Surface du bien en m²
nb_pieces	INT	Nombre de pièce
code_postal	INT	Code postal de l'arrondissement
code_type_local	INT	Code du type de biens



Table « vidéoprotection »

	code_cam	implantation	precision_implantation	date_implantation	lat	lon	code_annee	code_postal
0	1051	Rue Saint-Denis rue de Rivoli	croisement	08/07/16	48.858750	2.347502	5	75001
1	1055	Rue Montmartre rue de Turbigo	croisement	10/03/17	48.863243	2.346124	6	75001



Données	Type	Description
code_cam	INT	Index des caméras
implantation	VARCHAR	Adresse des caméras
precision_implantation	VARCHAR	Précisions de l'adresse(croisement)
date_implantation	VARCHAR	Date d'installation
lat	FLOAT	Latitude (Coordonnées GPS)
lon	FLOAT	Longitude (Coordonnées GPS)
code_annee	INT	Code de l'année
code_postal	INT	Code postal de l'arrondissement

Table « immo_vente »

code_biens	code_annee	date_vente	nature_vente	valeur_fonciere
0	0	9 02/01/2020	Vente	268800.0
1	1	9 06/01/2020	Vente	878378.0



Données	Type	Description
code_biens	INT	Code des biens
code_annee	INT	Code de l'année
date_vente	VARCHAR	Date de vente
nature_vente	VARCHAR	Nature de la vente
valeur_fonciere	FLOAT	Valeur du bien en Euros

Table « indicateur_delits »



code_annee	code_postal	code_delits	code_indicateur	nb_delits
0	9	75002	2	0
1	9	75003	2	1

Données	Type	Description
code_annee	INT	Code de l'année
code_postal	INT	Code postal de l'arrondissement
code_delits	INT	Date de vente
code_indicateur	INT	Code de l'indicateur
nb_delits	INT	Nombre de délits

5. MODELISATION DE LA BASE DE DONNEE

LE MODÈLE RELATIONNEL

Mon choix d'une base de donnée relationnelle est motivé par plusieurs facteurs :

- Les données sont stockées d'une manière structurées
- L'intégrité des données n'est pas menacée

LA MODELISATION

Afin de modéliser ma base de donnée, l'outil Looping a été utilisé, cet outil permet de concevoir mon Modèle de Conception des Données (MCD), mon Modèle Logique des Données (MLD) ainsi que le Modèle Physique des Données (MPD).

Par la suite il est possible de générer un script qui me permettra de créer mes différentes tables en fonction de cette modélisation.

5.1. MODELE CONCEPTUEL DES DONNEES (MCD)

ENTITES

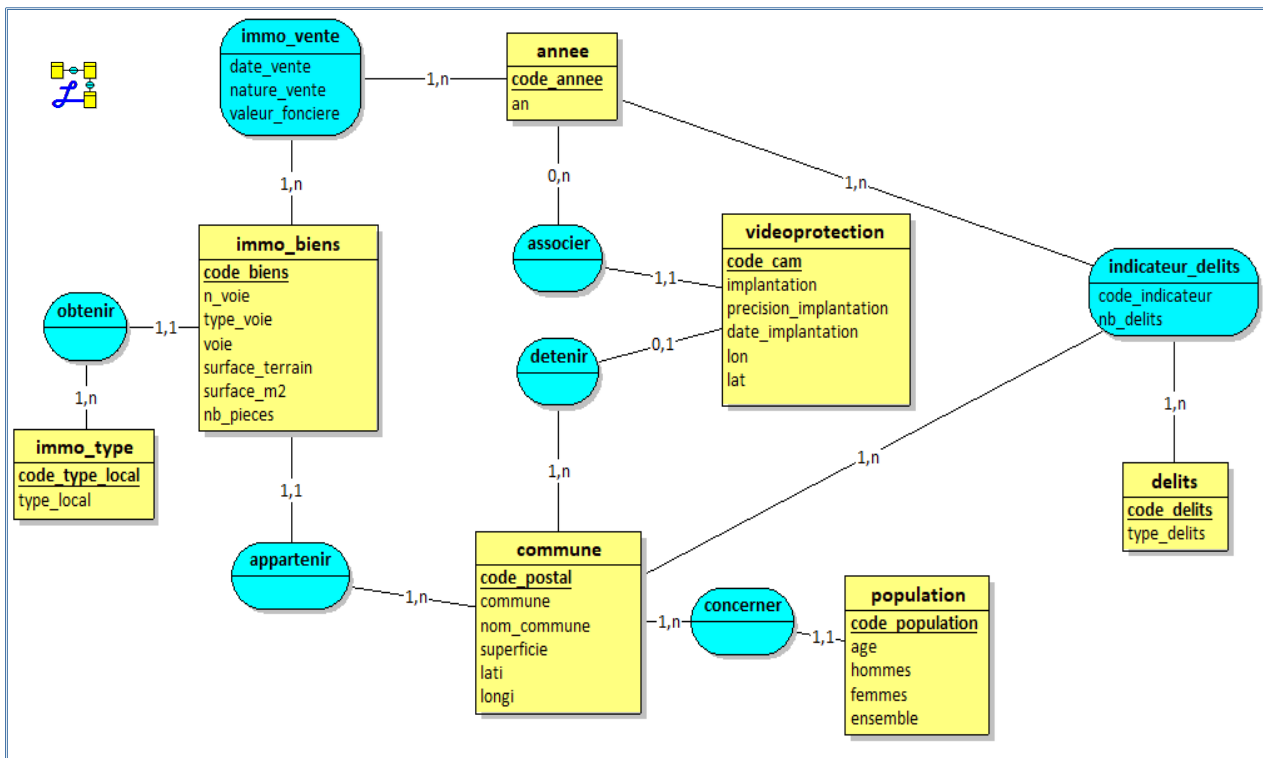
- Chaque **carrés jaunes** représentent une **entité**, ces 7 entités ont toutes un identifiant soulignée en première ligne ce qui permet d'identifier la table de manière unique. Elle ne peuvent être de valeur NULL.
- Ces entités deviendront des « tables » à la prochaine étape (MLD) et les identifiants, des clés primaires (Primary Key).

ASSOCIATIONS

- Chaque **bulle bleue** est une **association**, ces 7 associations lient les entités entre elles et peuvent devenir des « tables » ou créer une clé étrangère (Foreign Key) pour faire une liaison entre 2 tables à la prochaine étape (MLD).
- Dans cette modélisation, on dénombre 2 associations porteuses dont la clé primaire est composée par plusieurs clés étrangères, soit les clés primaires des différentes tables liées à l'association porteuse en question :
 - « immo_vente »
 - « indicateur_delits »

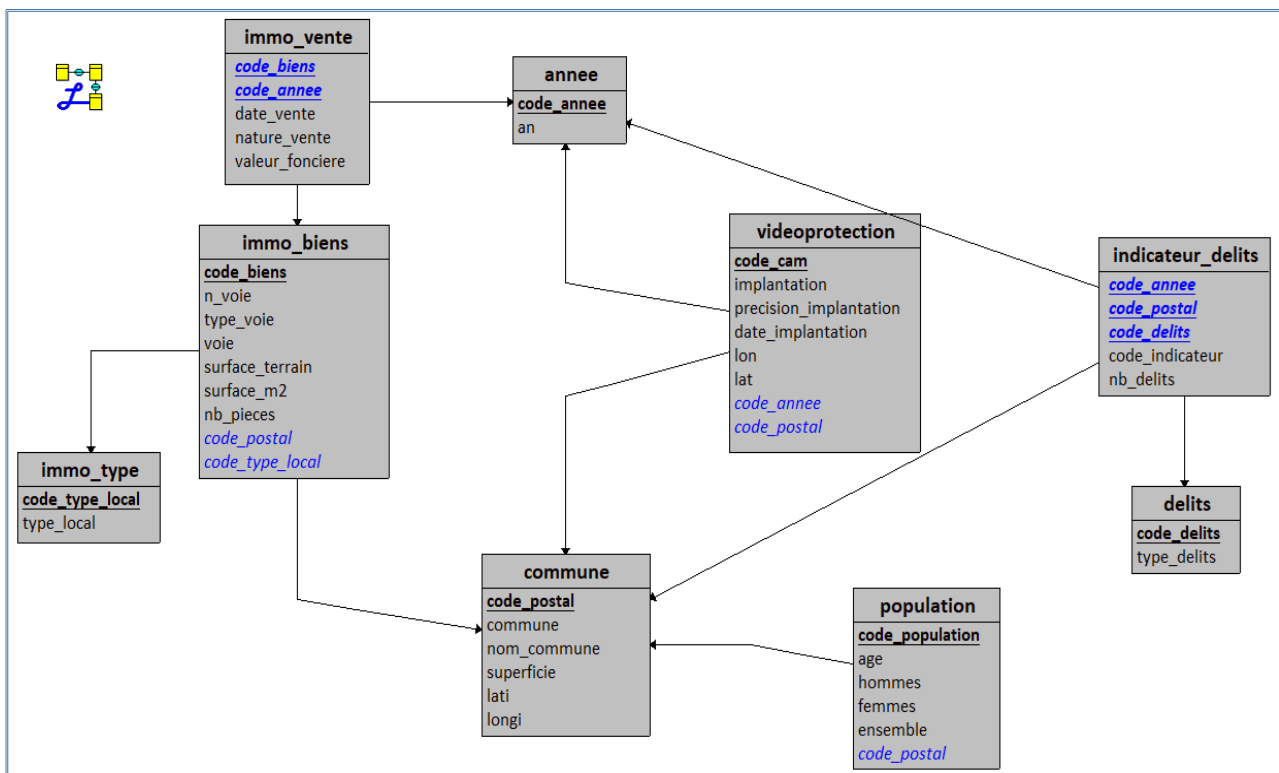
CARDINALITES

- Chaque **cardinalité** lie les associations aux entités et varient en fonction des différentes relations entre chaque table :
Entre l'entité « vidéoprotection » et « commune »
 - 0,1 : la vidéoprotection peut être détenu dans 0 ou 1 commune
 - 1,n : la commune peut détenir 1 ou plusieurs vidéoprotection



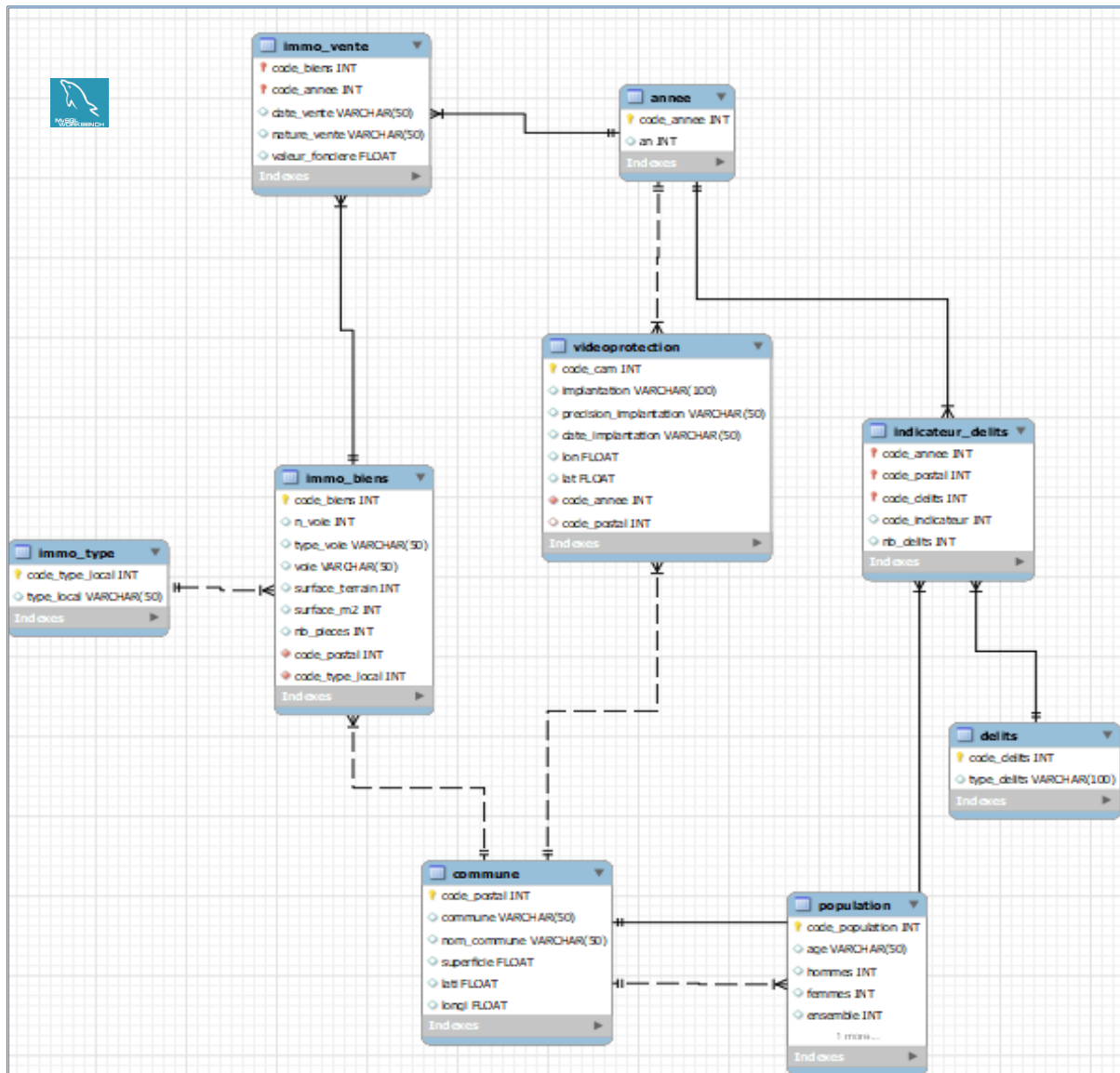
5.2. MODELE LOGIQUE DES DONNEES (MLD)

A ce stade de la modélisation, les entités deviennent des tables et les identifiants, des clés primaires (Primary Key). Les associations en fonction des différentes relations définies en amont deviennent des tables à leur tour. Sinon elle deviennent une clé étrangère (Foreign Key) dans les tables qui doivent être liées.



5.3. MODELE PHYSIQUE DES DONNEES (MPD)

Les différents types de données (INT, VARCHAR, FLOAT...) sont visibles à cette étape. Ils ont été définie lors de la conception du MCD.



6. CREATION DE LA BASE DE DONNEES

La base de données a été créée via MySQL Workbench en utilisant ce script :

```
drop database if exists immo;  
create database immo;  
use immo;
```

6.1. CREATION DES DIFFERENTES TABLES

Chaque table de la base de donnée a été créée en utilisant ce script généré automatiquement par l'outil Looping, le script a nécessité un nettoyage en amont et quelques modifications comme le changement de type de certaines colonnes.

```
5 • CREATE TABLE immo_type(  
6     code_type_local INT,  
7     type_local VARCHAR(50),  
8     PRIMARY KEY(code_type_local)  
9 );  
10  
11 • CREATE TABLE annee(  
12     code_annee INT,  
13     an INT,  
14     PRIMARY KEY(code_annee)  
15 );  
16  
17 • CREATE TABLE commune(  
18     code_postal INT,  
19     commune VARCHAR(50),  
20     nom_commune VARCHAR(50),  
21     superficie FLOAT,  
22     lati FLOAT,  
23     longi FLOAT,  
24     PRIMARY KEY(code_postal)  
25 );
```

```
27 • CREATE TABLE delits(  
28     code_delits INT,  
29     type_delits VARCHAR(100),  
30     PRIMARY KEY(code_delits)  
31 );  
32  
33 • CREATE TABLE population(  
34     code_population INT,  
35     age VARCHAR(50),  
36     hommes INT,  
37     femmes INT,  
38     ensemble INT,  
39     code_postal INT NOT NULL,  
40     PRIMARY KEY(code_population)  
41 );
```

```
43 • CREATE TABLE immo_biens(  
44     code_biens INT,  
45     n_voie INT,  
46     type_voie VARCHAR(50),  
47     voie VARCHAR(50),  
48     surface_terrain INT,  
49     surface_m2 INT,  
50     nb_pieces INT,  
51     code_postal INT NOT NULL,  
52     code_type_local INT NOT NULL,  
53     PRIMARY KEY(code_biens)  
54 );  
55  
56 • CREATE TABLE videoprotection(  
57     code_cam INT,  
58     implantation VARCHAR(100),  
59     precision_implantation VARCHAR(50),  
60     date_implantation VARCHAR(50),  
61     lon FLOAT,  
62     lat FLOAT,  
63     code_annee INT NOT NULL,  
64     code_postal INT,  
65     PRIMARY KEY(code_cam)  
66 );
```

```
68 • CREATE TABLE immo_vente(  
69     code_biens INT,  
70     code_annee INT,  
71     date_vente VARCHAR(50),  
72     nature_vente VARCHAR(50),  
73     valeur_fonciere FLOAT,  
74     PRIMARY KEY(code_biens, code_annee)  
75 );  
76  
77 • CREATE TABLE indicateur_delits(  
78     code_annee INT,  
79     code_postal INT,  
80     code_delits INT,  
81     code_indicateur INT,  
82     nb_delits INT,  
83     PRIMARY KEY(code_annee, code_postal, code_delits)  
84 );
```

..... CLES ETRANGERES

Les clés étrangères ont été insérées après la création des tables, grâce à la commande SQL Alter table qui permet de modifier une table existante et ce, afin d'éviter les erreurs au moment de l'insertion des données.

```
87 • ALTER TABLE population add FOREIGN KEY(code_postal) REFERENCES commune(code_postal);
88 • ALTER TABLE immo_biens add FOREIGN KEY(code_postal) REFERENCES commune(code_postal);
89 • ALTER TABLE immo_biens add FOREIGN KEY(code_type_local) REFERENCES immo_type(code_type_local);
90 • ALTER TABLE videoprotection add FOREIGN KEY(code_annee) REFERENCES annee(code_annee);
91 • ALTER TABLE videoprotection add FOREIGN KEY(code_postal) REFERENCES commune(code_postal);
92 • ALTER TABLE immo_vente add FOREIGN KEY(code_biens) REFERENCES immo_biens(code_biens);
93 • ALTER TABLE immo_vente add FOREIGN KEY(code_annee) REFERENCES annee(code_annee);
94 • ALTER TABLE indicateur_delits add FOREIGN KEY(code_annee) REFERENCES annee(code_annee);
95 • ALTER TABLE indicateur_delits add FOREIGN KEY(code_delits) REFERENCES delits(code_delits);
96 • ALTER TABLE indicateur_delits add FOREIGN KEY(code_postal) REFERENCES commune(code_postal);
```

6.2. CONNEXION A LA BASE DE DONNEE

La base de donnée est maintenant créée, ainsi que les tables encore vide pour le moment,. L'étape suivante consiste à créer une connexion entre la base de donnée et mon interface de travail, jupyter notebook en utilisant les librairies Pymysql et SQLAlchemy préalablement importés grâce aux commandes suivantes :

```
1 import pandas as pd
2 import pymysql
3 from sqlalchemy import create_engine
```

Connection à la BDD

```
1 my_conn = create_engine('mysql+pymysql://root:nounou@localhost/immobilier')
2 if my_conn:
3     print("Connection Successful!")
4 else:
5     print("Connection Failed!")
```

Connection Successful!

6.3. IMPORTATION DES DONNEES

L'insertion des données est réalisée via jupyter notebook grâce à la connexion préalablement établie et ce, en utilisant la méthode de lecture de pandas `read_csv` qui permet d'obtenir un Dataframe et par conséquent, de visualiser les données du fichier CSV. Puis en insérant ce Dataframe via la méthode `to_sql`.

Table Videoprotection

Lecture du CSV

```
1 videoprotection = pd.read_csv("CSV_table/videoprotection.csv", sep = ",")
2 videoprotection.head()
```

	code_cam	implantation	precision_implantation	date_implantation	lat	lon	code_annee	code_postal
0	1051	Rue Saint-Denis rue de Rivoli	croisement	08/07/16	48.858750	2.347502	5	75001
1	1055	Rue Montmartre rue de Turbigo	croisement	10/03/17	48.863243	2.346124	6	75001
2	1056	Rue Montorgueil rue de Turbigo	croisement	10/03/17	48.863273	2.346291	6	75001
3	1057	Quai François Mitterrand rue de l'Amiral de C...	croisement	08/06/18	48.858990	2.339848	7	75001
4	2006	Boulevard des Capucines rue Scribe	croisement	Début 2019	48.870250	2.329943	10	75002

Insertion des données dans la table

```
1 videoprotection.to_sql('videoprotection',con = my_conn,if_exists="append", index = False)
```

Nous procédons de la même manière pour toutes les tables de notre base de donnée.

Aperçu du nombre de table dans la Database

```
1 requete = ''' show tables'''
2 req = pd.read_sql(requete, my_conn)
3 req
4
```

Tables_in_immo

0	annee
1	commune
2	delits
3	immo_biens
4	immo_type
5	immo_vente
6	indicateur_delits
7	population
8	videoprotection

7. SECURISATION & SAUVEGARDE

SAUVEGARDE DE LA BDD (BackUp)

Nous allons maintenant effectuer une sauvegarde (BackUp) de la base de donnée avec les commandes suivantes :

```
Administrateur : Invite de commandes
Microsoft Windows [version 10.0.19042.1110]
(c) Microsoft Corporation. Tous droits réservés.

C:\windows\system32>mysqldump -u root -p immo >backup_immo.sql
Enter password: *****

C:\windows\system32>mysql -u root -p immo < backup_immo.sql
Enter password: *****

C:\windows\system32>
```

SECURISATION DE LA BDD

Nous allons maintenant effectuer la sécurisation de la base de donnée, pour se faire nous allons créer et attribuer des droits aux utilisateurs avec les commandes suivantes :

- Grace à l'Administrateur «root» et via son mot de passe créé par défaut, nous allons créer un utilisateur «support» .
- Nous allons lui attribuer un mot de passe et des droits «privilèges» (lecture « select») avec la commande « Grant select » et « flush privilèges » et uniquement sur la BDD IMMO.

Création de l'utilisateur « support » :

```
mysql> create user 'support'@'localhost';
Query OK, 0 rows affected (0.01 sec)

mysql> select user,host from mysql.user;
+-----+-----+
| user          | host          |
+-----+-----+
| mysql.infoschema | localhost    |
| mysql.session   | localhost    |
| mysql.sys       | localhost    |
| root           | localhost    |
| support        | localhost    |
+-----+-----+
5 rows in set (0.00 sec)
```

Attribution d'un mot de passe à « support » :

```
mysql> ALTER USER 'support'@'localhost' IDENTIFIED BY 'nounou';
Query OK, 0 rows affected (0.02 sec)
```

Accorder des privilèges à « support » et visualisation de ses privilèges:

```
mysql> grant select on immo.* to "support"@"localhost";
Query OK, 0 rows affected (0.02 sec)

mysql> flush privileges;
Query OK, 0 rows affected (0.01 sec)

mysql> show grants for "support"@"localhost";
+-----+-----+
| Grants for support@localhost |
+-----+-----+
| GRANT USAGE ON *.* TO `support`@`localhost` |
| GRANT SELECT ON `immo`.* TO `support`@`localhost` |
+-----+-----+
2 rows in set (0.00 sec)
```


8. REQUETAGE & OPTIMISATION

L'heure est à la frugalité numérique et si nous avons la possibilité d'économiser de l'énergie utilisé inutilement et avoir de meilleure performance autant en être conscient.

En effet, une base de donnée bien construite permet de meilleure performance tout comme l'optimisation d'une requête SQL va améliorer les temps de réponse et réduire le temps de calcul du processeur.

OPTIMISATION DE LA BASE DE DONNEE

- Eviter les redondances, et stocker dans une nouvelle table ces données textuelles, et les lier par une jointure (par exemple) :

La table « année » ci-après

annee
code_annee
an

se trouvait sous forme de donnée dans plusieurs tables, elle a été extraite pour en faire une seule table indépendante.

- L'index est utile pour accélérer l'exécution d'une requête SQL qui lit des données et ainsi améliorer les performances d'une application utilisant une base de données.

OPTIMISATION D'UNE REQUETE SQL

Par exemple :

- **Requête SQL d'origine :**

```
SELECT code_cam ,implantation
FROM videoprotection
WHERE implantation LIKE 'AVENUE%' AND 'RUE%';
```

- **Requête optimisée :**

```
SELECT code_cam ,implantation
FROM videoprotection
WHERE implantation LIKE ALL('AVENUE%' , 'RUE%');
```

9. ANALYSE & VISUALISATION

Nous passons maintenant à l'analyse de la donnée.

Pour cela, nous allons, à l'aide de requête SQL, rendre visible la donnée et ainsi répondre aux différentes questions définies lors de l'étape 2 du cahier des charges.

FOCUS sur les DÉLITS recensés par la Police Nationale

1. Quels sont les délits traités dans cette analyse?

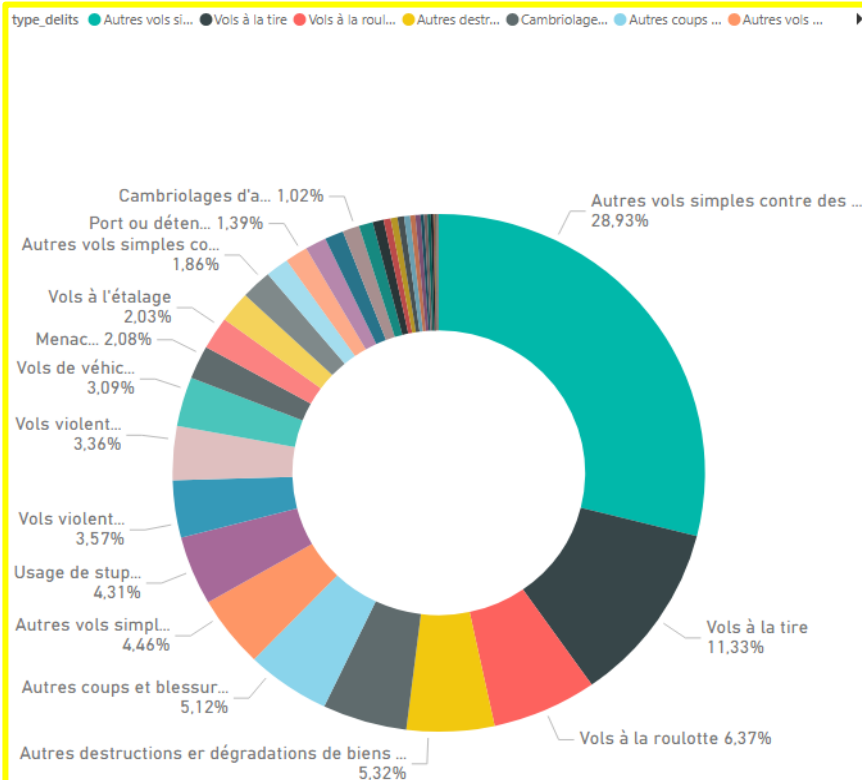
```
1 query_1 = '''SELECT DISTINCT type_delits AS 'Type de délit' , sum(nb_delits) AS 'Nb de délits'
2 FROM indicateur_delits
3 JOIN delits ON indicateur_delits.code_delits = delits.code_delits
4 GROUP BY type_delits
5 ORDER BY sum(nb_delits) DESC
6 ; '''
7
8 req_1 = pd.read_sql(query_1, my_conn)
9 req_1
```



Les **délits** traités dans cette analyse concernent :

- l'habitation ;
- l'environnement autour de l'habitation ;
- tout biens personnels susceptibles d'être entreposés proche du lieu d'habitation .

	Type de délit	Nb de délits
0	Autres vols simples contre des particuliers da...	487937.0
1	Vols à la tire	190997.0
2	Vols à la roulotte	107420.0
3	Autres destructions er dégradations de biens p...	89793.0
4	Cambriolages de locaux d'habitations principales	86491.0
5	Autres coups et blessures volontaires criminel...	86404.0
6	Autres vols simples contre des particuliers da...	75229.0
7	Usage de stupéfiants	72626.0
8	Vols violents sans arme contre des femmes sur ...	60195.0
9	Vols violents sans arme contre d'autres victimes	56711.0
10	Vols de véhicules motorisés à 2 roues	52116.0
11	Menaces ou chantages dans un autre but	35099.0
12	Vols à l'étalage	34259.0
13	Destructions et dégradations de véhicules privés	32710.0
14	Autres vols simples contre des établissements ...	31401.0
15	Camb de locaux industriels commerciaux ou fin...	23833.0
16	Port ou détention armes prohibées	23358.0
17	Vols d'automobiles	21667.0
18	Vols d'accessoires sur véhicules à moteur imma...	19270.0
19	Cambriolages d'autres lieux	17160.0
20	Usage revente de stupéfiants	14414.0
21	Menaces ou chantages pour extorsion de fonds	11145.0
22	Atteintes sexuelles	7341.0
23	Autres destructions er dégradations de biens p...	6954.0
24	Trafic et revente sans usage de stupéfiants	6759.0
25	Harcèlements sexuels et autres agressions sexu...	6130.0
26	Autres vols avec armes blanches	5304.0
27	Violations de domicile	5274.0
28	Incendies volontaires de biens privés	3481.0
29	Vols avec entrée par ruse en tous lieux	3480.0
30	Cambriolages de résidences secondaires	3386.0
31	Viols sur des majeurs	2566.0
32	Sequestrations	1168.0
33	Vols violents sans arme contre des particulier...	912.0
34	Harcèlements sexuels et autres agressions sexu...	866.0
35	Tentatives homicides pour d'autres motifs	688.0
36	Vols violents sans arme contre des établisseme...	616.0
37	Vols avec armes blanches contre des établisse...	395.0
38	Vols avec armes blanches contre des particule...	273.0
39	Viols sur des mineurs	260.0
40	Homicides pour d'autres motifs	105.0
41	Atteintes à l'environnement	74.0
42	Vols de véhicules de transport avec frêt	52.0
43	Tentatives d'homicides pour voler et à l'occas...	49.0
44	Vols à main armée contre des particuliers à le...	41.0
45	Coups et blessures volontaires suivis de mort	30.0
46	Prises d'otages à l'occasion de vols	11.0
47	Vols à main armée contre des entreprises de tr...	4.0
48	Homicides pour voler et à l'occasion de vols	3.0



2. Quels est le top 10 des types de délits les plus fréquents?

```

1 query_2 = '''SELECT type_delits AS 'Type de délits', SUM(nb_delits) AS 'Nb de délits'
2 FROM delits
3 JOIN indicateur_delits ON delits.code_delits = indicateur_delits.code_delits
4 GROUP BY type_delits
5 ORDER BY SUM(nb_delits) DESC
6 LIMIT 10;'''
7
8 req_2 = pd.read_sql(query_2, my_conn)
9 req_2

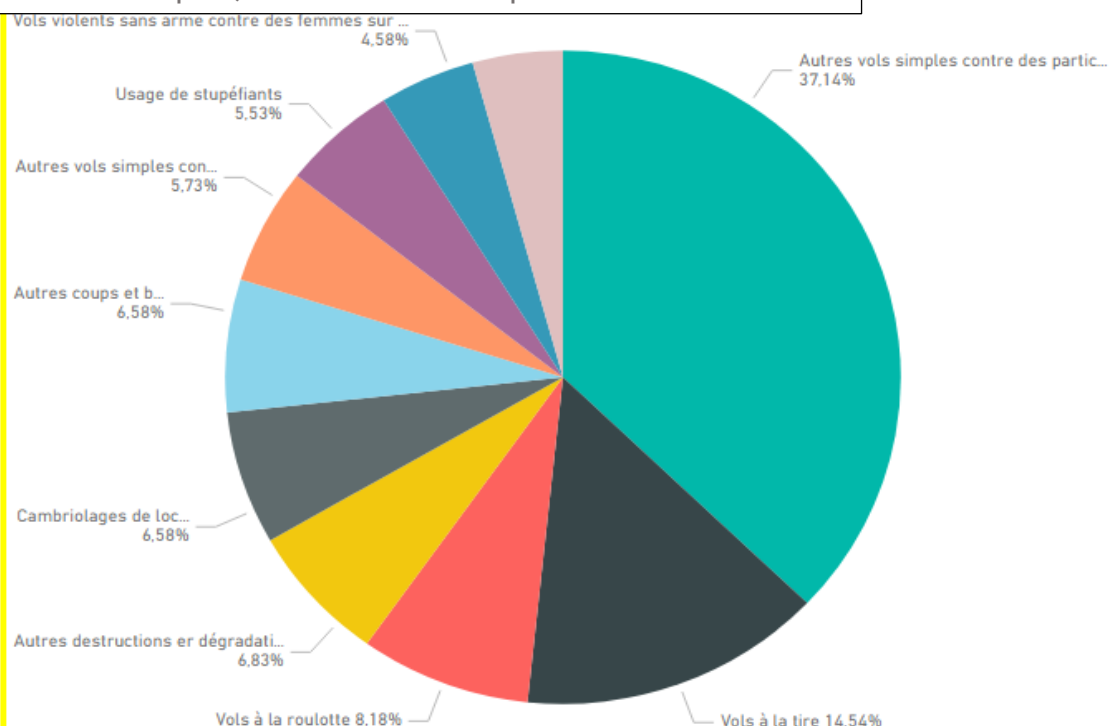
```



	Type de délits	Nb de délits
0	Autres vols simples contre des particuliers da...	487937.0
1	Vols à la tire	190997.0
2	Vols à la roulotte	107420.0
3	Autres destructions er dégradations de biens p...	89793.0
4	Cambriolages de locaux d'habitations principales	86491.0
5	Autres coups et blessures volontaires criminel...	86404.0
6	Autres vols simples contre des particuliers da...	75229.0
7	Usage de stupéfiants	72626.0
8	Vols violents sans arme contre des femmes sur ...	60195.0
9	Vols violents sans arme contre d'autres victimes	56711.0

Nous constatons que les délit les plus fréquents du Top 10 des délits sont les « autres vols simples contre des particuliers dans des lieux publics » avec 487937 vols recensés soit 37% des délits suivi des vols à la tire avec 190997 cas.

Les vols violents sans arme sont les moins fréquents avec 56711 vols, soit moins de 5% des délits suivi de ceux infligés aux femmes qui subissent 3484 vols en plus, soit 60195 vols depuis 2012.



3. Quel est le nombre de délits par arrondissement ?

```

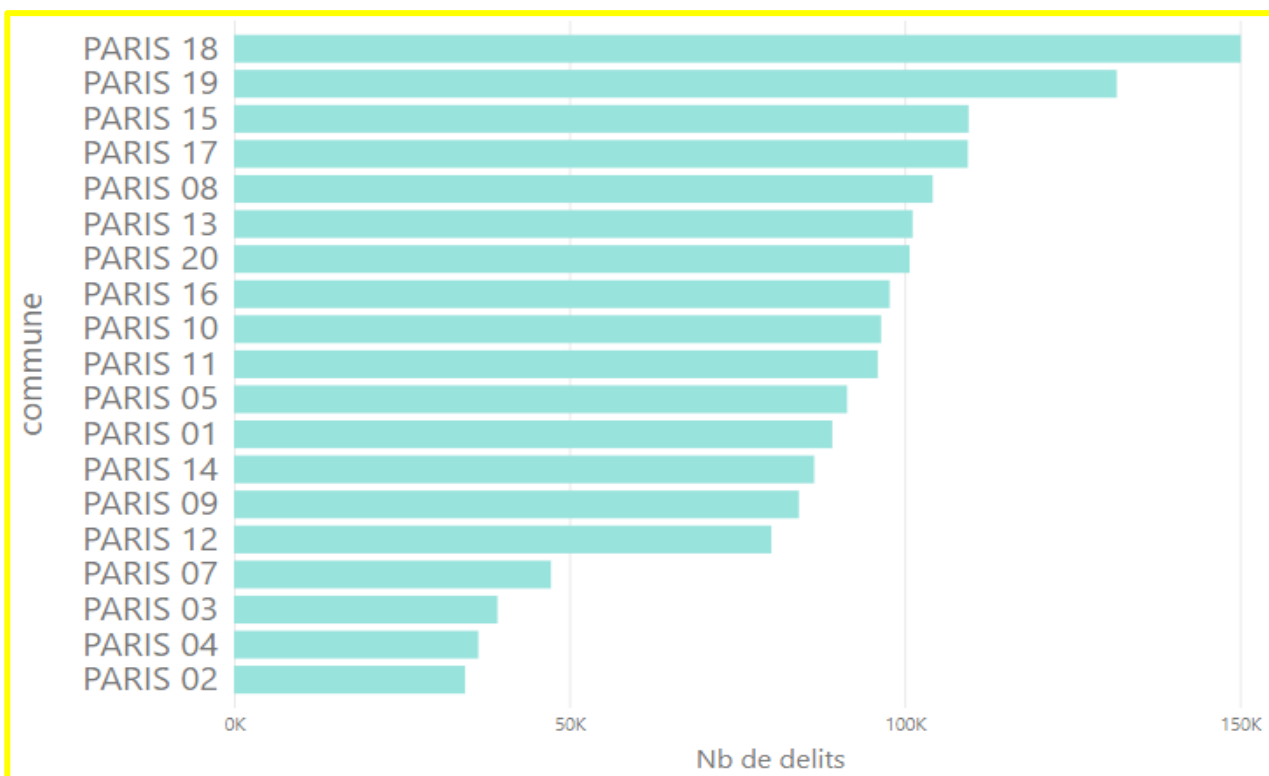
1 query_3 = '''SELECT DISTINCT commune.code_postal AS 'Code postal',
2 nom_commune as 'Quartier', SUM(nb_delits) AS 'Nb de délits'
3 FROM indicateur_delits
4 JOIN commune ON indicateur_delits.code_postal = commune.code_postal
5 GROUP BY commune.code_postal
6 ORDER BY SUM(nb_delits) DESC
7 ; '''
8
9 req_3 = pd.read_sql(query_3, my_conn)
10 req_3

```

	Code post	Quartier	Nb de délits
0	75018	Buttes-Montmartre	150153.0
1	75019	Buttes-Chaumont	131669.0
2	75015	Vaugirard	109565.0
3	75017	Batignolles-Monceau	109456.0
4	75008	Élysée	104213.0
5	75013	Gobelins	101218.0
6	75020	Ménilmontant	100773.0
7	75016	Passy	97779.0
8	75010	Entrepôt	96515.0
9	75011	Popincourt	96017.0
10	75005	Panthéon	91447.0
11	75001	Louvre	89230.0
12	75014	Observatoire	86535.0
13	75009	Opéra	84245.0
14	75012	Reuilly	80143.0
15	75007	Palais-Bourbon	47291.0
16	75003	Temple	39296.0
17	75004	Hôtel-de-Ville	36442.0
18	75002	Bourse	34470.0



Les **délits** les plus nombreux sont recensés dans le quartier des Buttes-Montmartre (18^{ème}) avec 150153 délits tout confondus et dans celui des Buttes-Chaumont(19^{ème}) avec 131669 délits. Alors que les moins nombreux sont dans le quartier Bourse (2^{ème}) avec 34470 et celui de l'hôtel-de-ville (4^{ème}), avec ses 36442 délits.



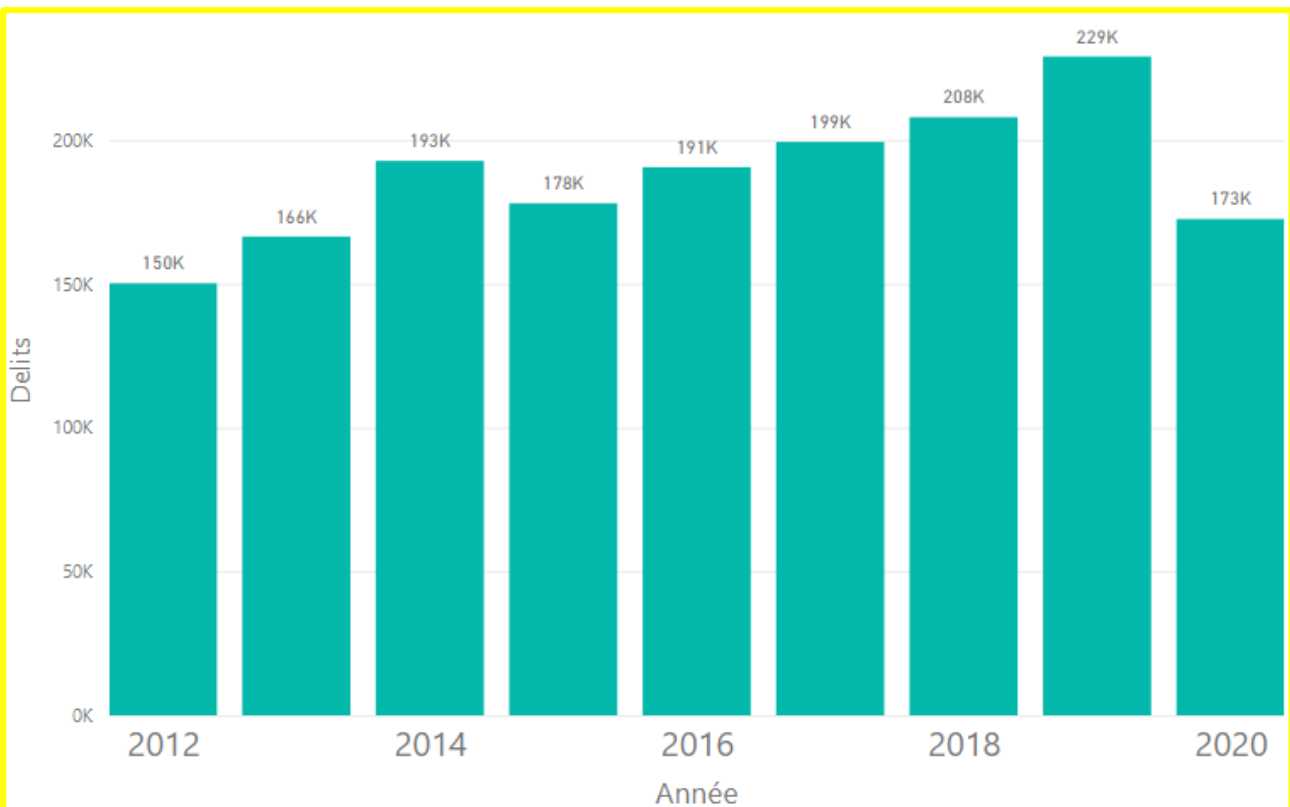
4. Quel est le nombre de délits par année ?

```
1 query_4 = '''SELECT an AS Année, sum(nb_delits) AS 'Nb de délits'
2 FROM indicateur_delits
3 JOIN annee ON indicateur_delits.code_annee = annee.code_annee
4 GROUP BY an
5 ; '''
6
7 req_4 = pd.read_sql(query_4, my_conn)
8 req_4
```

	Année	Nb de délits
0	2012	150143.0
1	2013	166400.0
2	2014	192719.0
3	2015	177956.0
4	2016	190546.0
5	2017	199290.0
6	2018	207895.0
7	2019	229008.0
8	2020	172500.0



Sur ce graph barre, nous observons une augmentation constante depuis 2012 (entre 4 et 10% d'augmentation) avec un pic de délinquance en 2014 (16%) jusqu'en 2019 puis une diminution flagrante en 2020 sûrement lié au covid 19, passant de 229008 délits à 172500 soit 56508 délits en moins.



5. Quel est le nombre de cambriolage par arrondissement ?

```

1 query_5 = '''SELECT DISTINCT commune.code_postal AS 'Code postal',
2 nom_commune AS 'Quartier', sum(nb_delits) AS 'Nb de délits'
3 FROM indicateur_delits
4 JOIN commune ON indicateur_delits.code_postal = commune.code_postal
5 JOIN delits ON indicateur_delits.code_delits = delits.code_delits
6 WHERE delits.code_delits LIKE 27
7 GROUP BY commune.code_postal
8 ORDER BY sum(nb_delits) DESC
9 ; '''
10
11 req_5 = pd.read_sql(query_5, my_conn)
12 req_5

```

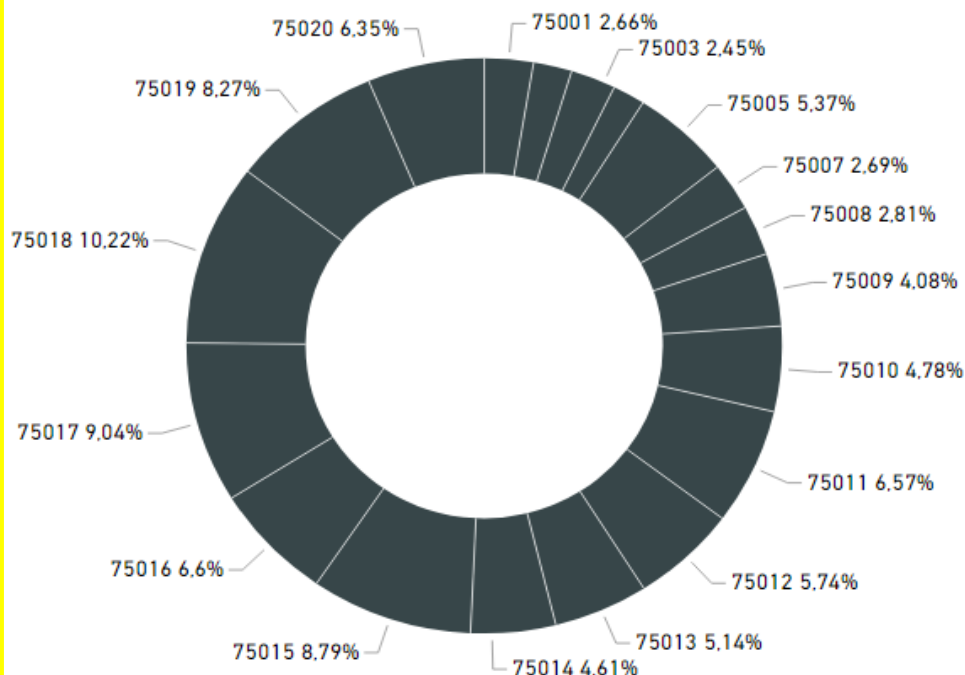


Le quartier qui subit le plus de cambriolage est le 18^{ème} et le 17^{ème} avec 8837 et 7816 cambriolages.

On recense en 1520 dans le 4^{ème} arrondissement soit 7317 cambriolages en moins et 1807 dans le 2^{ème}.

	Code postal	Quartier	Nb de délits
0	75018	Buttes-Montmartre	8837.0
1	75017	Batignolles-Monceau	7816.0
2	75015	Vaugirard	7601.0
3	75019	Buttes-Chaumont	7154.0
4	75016	Passy	5712.0
5	75011	Popincourt	5681.0
6	75020	Ménilmontant	5489.0
7	75012	Reuilly	4965.0
8	75005	Panthéon	4641.0
9	75013	Gobelins	4449.0
10	75010	Entrepôt	4135.0
11	75014	Observatoire	3984.0
12	75009	Opéra	3527.0
13	75008	Élysée	2430.0
14	75007	Palais-Bourbon	2328.0
15	75001	Louvre	2297.0
16	75003	Temple	2118.0
17	75002	Bourse	1807.0
18	75004	Hôtel-de-Ville	1520.0

Nombre de cambriolage par arrondissement



FOCUS sur la VIDÉOPROTECTION

6. Quels sont les arrondissements les mieux dotés en caméra ?

```

1 query_6 = '''SELECT commune.code_postal AS 'Code postal',
2 nom_commune AS 'Quartier', COUNT(code_cam) AS 'Caméra'
3 FROM videoprotection
4 JOIN commune ON videoprotection.code_postal = commune.code_postal
5 GROUP BY commune.code_postal
6 ORDER BY COUNT(code_cam) DESC
7 ; '''
8
9 req_6 = pd.read_sql(query_6, m
10 req_6

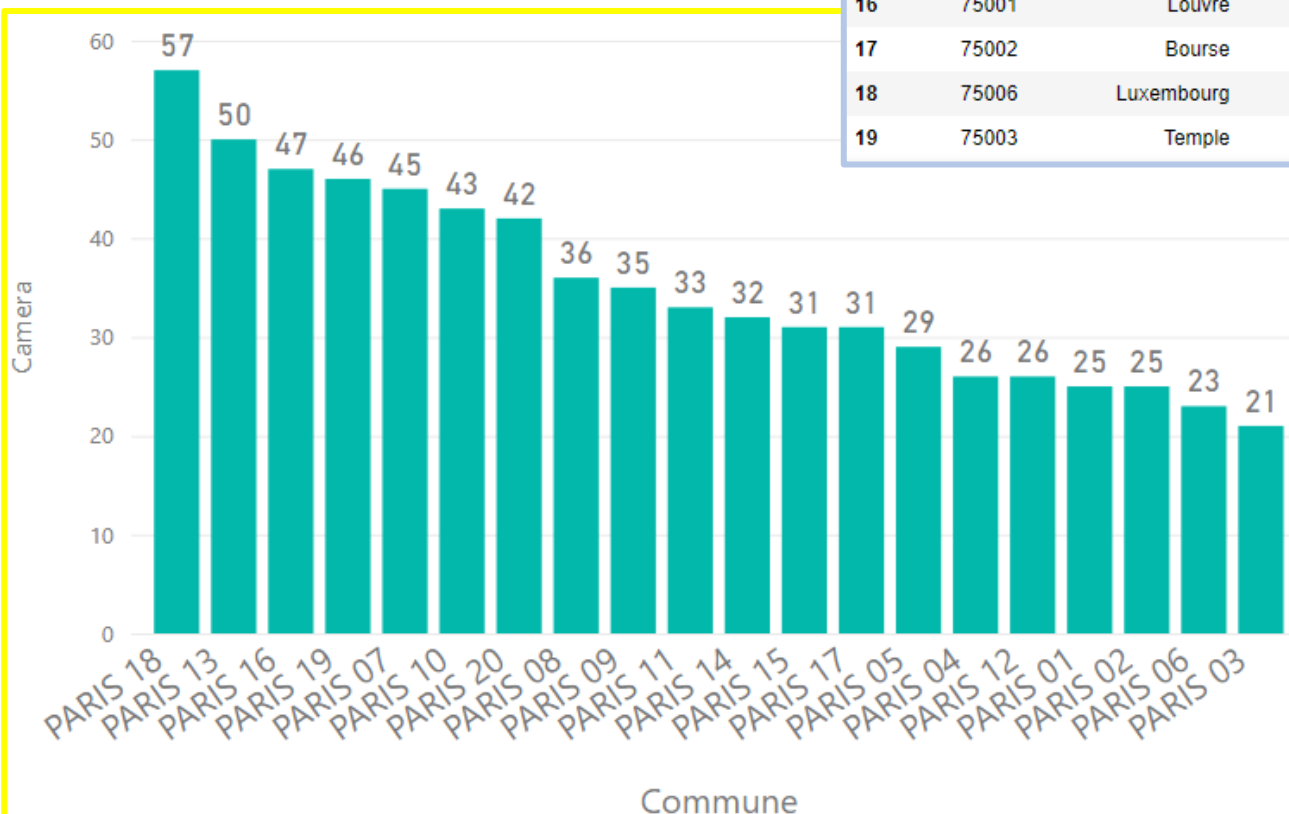
```

	Code postal	Quartier	Caméra
0	75018	Buttes-Montmartre	57
1	75013	Gobelins	50
2	75016	Passy	47
3	75019	Buttes-Chaumont	46
4	75007	Palais-Bourbon	45
5	75010	Entrepôt	43
6	75020	Ménilmontant	42
7	75008	Élysée	36
8	75009	Opéra	35
9	75011	Popincourt	33
10	75014	Observatoire	32



On constate 57 caméras dans le 18^{ème} arrondissement, c'est donc le plus doté en caméra, et 50 dans le 13^{ème} contre 21 et 23 caméras dans le 3^{ème} et le 6^{ème} arrondissement.

11	75015	Vaugirard	31
12	75017	Batignolles-Monceau	31
13	75005	Panthéon	29
14	75004	Hôtel-de-Ville	26
15	75012	Reuilly	26
16	75001	Louvre	25
17	75002	Bourse	25
18	75006	Luxembourg	23
19	75003	Temple	21



7. Quelle est la fréquence d'installation de la videoprotection ?

```

1 query_7 = '''SELECT an AS 'Année', COUNT(code_cam) AS 'Caméras installées'
2 FROM videoprotection
3 JOIN annee ON annee.code_annee = videoprotection.code_annee
4 GROUP BY an
5 ORDER BY COUNT(code_cam) DESC
6 ; '''
7
8 req_7 = pd.read_sql(query_7, my_conn)

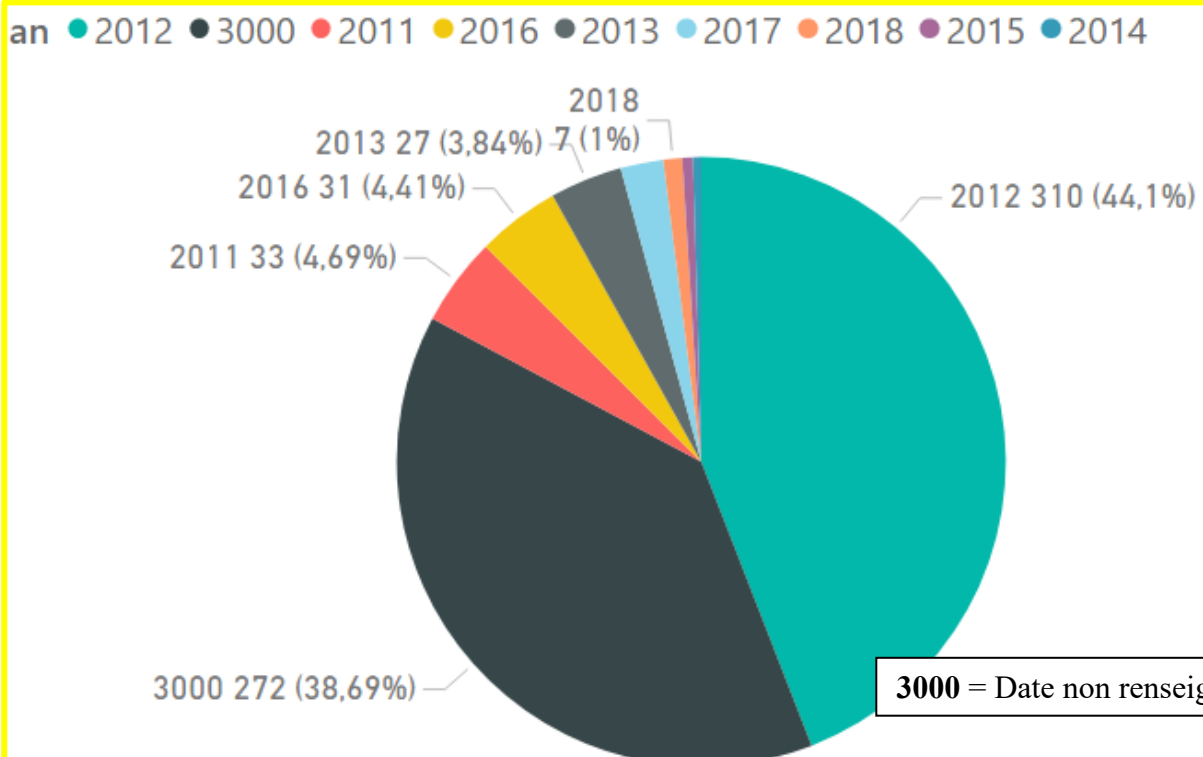
```



L'installation des caméras a surtout été faite en 2012, soit l'année suivant le lancement du projet des «1100 caméras» avec 310 caméras installées, mais il faut prendre en compte les 272 dates non renseignées qui apparaissent à « l'année 3000 ».

L'année 2014 est celle qui contient le moins d'installation avec seulement 3 caméras installées.

	Année	Caméras installées
0	2012	310
1	3000	272
2	2011	33
3	2016	31
4	2013	27
5	2017	16
6	2018	7
7	2015	4
8	2014	3



8. Quel est le nombre de caméra par arrondissement en fonction de la superficie?

```

1 query_8 = '''SELECT commune.code_postal AS 'Arrondissement',
2 ROUND((superficie),0) AS 'Superficie en m²', COUNT(code_cam) AS 'Nb de cam'
3 FROM commune
4 JOIN videoprotection ON commune.code_postal = videoprotection.code_postal
5 GROUP BY commune.code_postal
6 ORDER BY superficie DESC;'''
7
8 req_8 = pd.read_sql(query_8, my_conn)
9 req_8

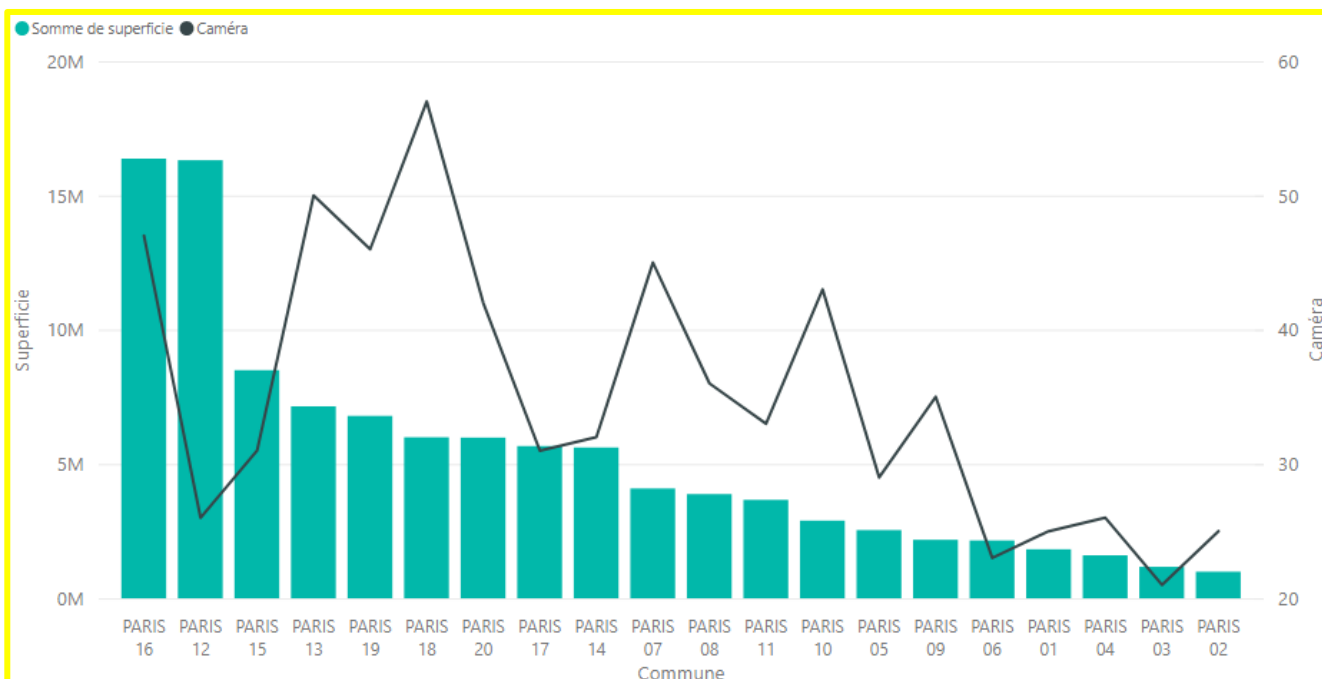
```

	Arrondissement	Superficie en m²	Nb de cam
0	75016	16372542.0	47
1	75012	16314783.0	26
2	75015	8494994.0	31
3	75013	7149311.0	50
4	75019	6792651.0	46
5	75018	5996052.0	57
6	75020	5983446.0	42
7	75017	5668834.0	31
8	75014	5614878.0	32
9	75007	4090057.0	45
10	75008	3880036.0	36
11	75011	3665442.0	33
12	75010	2891740.0	43
13	75005	2539374.0	29
14	75009	2178303.0	35
15	75006	2153096.0	23
16	75001	1824613.0	25
17	75004	1600586.0	26
18	75003	1170883.0	21
19	75002	991154.0	25



Nous constatons que l'arrondissement le plus vaste est le 16^{ème} avec 16372542 m² et 47 caméras. Il est suivi de près par le 12^{ème} avec 16314783 m² et 26 caméras, soit près de la moitié en terme de caméra.

Le 2^{ème} arrondissement est le plus petit avec 991154 m² et 25 caméras, soit autant doté en caméra que le 12^{ème} arrondissement.



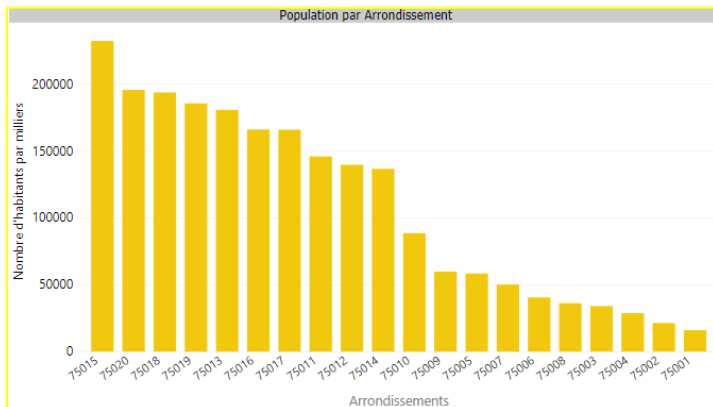
9. Quel est le nombre d'habitant par camera ?

query9_1 = Population par arrondissement

```
1 query9_1 = '''SELECT SUM(ensemble) AS 'population', population.code_postal
2 FROM population
3 GROUP BY population.code_postal
4 ORDER BY population ASC; '''
5
6 req9_1 = pd.read_sql(query9_1, my_conn)
7 req9_1
```

	population	code_postal
0	16093.0	75001
1	21421.0	75002
2	28762.0	75004
3	34021.0	75003
4	36222.0	75008
5	40525.0	75006
6	50197.0	75007
7	58379.0	75005
8	59835.0	75009
9	88558.0	75010
10	136597.0	75014

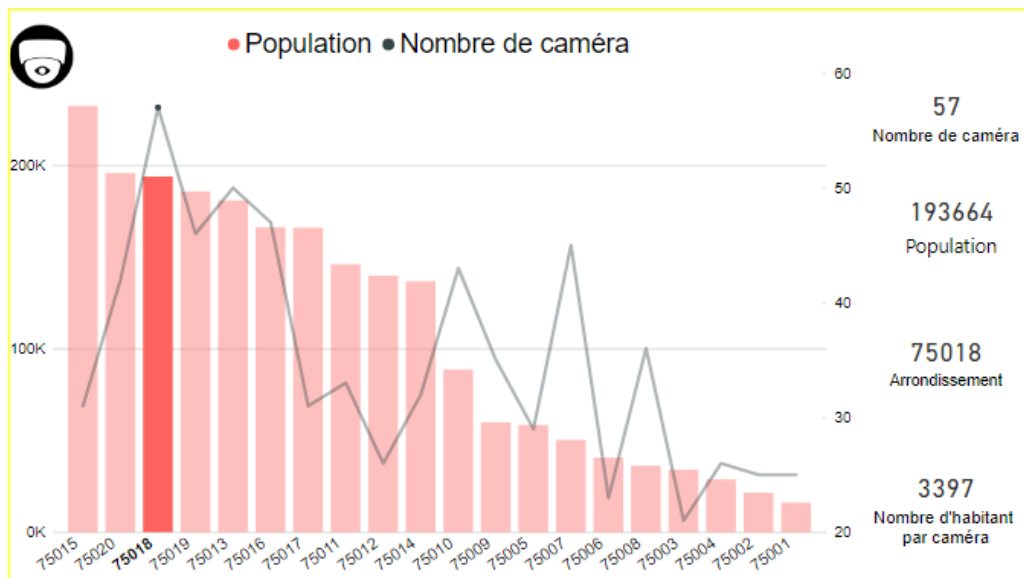
11	139664.0	75012
12	145903.0	75011
13	165860.0	75017
14	166013.0	75016
15	180632.0	75013
16	185514.0	75019
17	193664.0	75018
18	195599.0	75020
19	232144.0	75015



Le 15^{ème} arrondissement est le plus peuplé avec 232144 habitants pour 31 caméras alors que le plus petit arrondissement qui est le 1^{er} et est doté de 25 caméras pour seulement 16093 habitants. Cela s'explique par l'affluence touristique du 1^{er} arrondissement alors que le 15^{ème} est un quartier plus populaire

9_2 = MERGE avec la requete query9_1: Population par arrondissement et la requête query_6 : Nombre de caméra par arrondissement

```
1 req9_2 = req9_1.merge(req_6, how='inner', on='code_postal')
2 req9_2
```



	population	code_postal	Quartier	Caméra
0	16093.0	75001	Louvre	25
1	21421.0	75002	Bourse	25
2	28762.0	75004	Hôtel-de-Ville	26
3	34021.0	75003	Temple	21
4	36222.0	75008	Élysée	36
5	40525.0	75006	Luxembourg	23
6	50197.0	75007	Palais-Bourbon	45
7	58379.0	75005	Panthéon	29
8	59835.0	75009	Opéra	35
9	88558.0	75010	Entrepôt	43
10	136597.0	75014	Observatoire	32
11	139664.0	75012	Reuilly	26
12	145903.0	75011	Popincourt	33
13	165860.0	75017	Batignolles-Monceau	31
14	166013.0	75016	Passy	47
15	180632.0	75013	Gobelins	50
16	185514.0	75019	Buttes-Chaumont	46
17	193664.0	75018	Buttes-Montmartre	57
18	195599.0	75020	Ménilmontant	42
19	232144.0	75015	Vaugirard	31

10. Qui protège réellement la vidéo protection ?

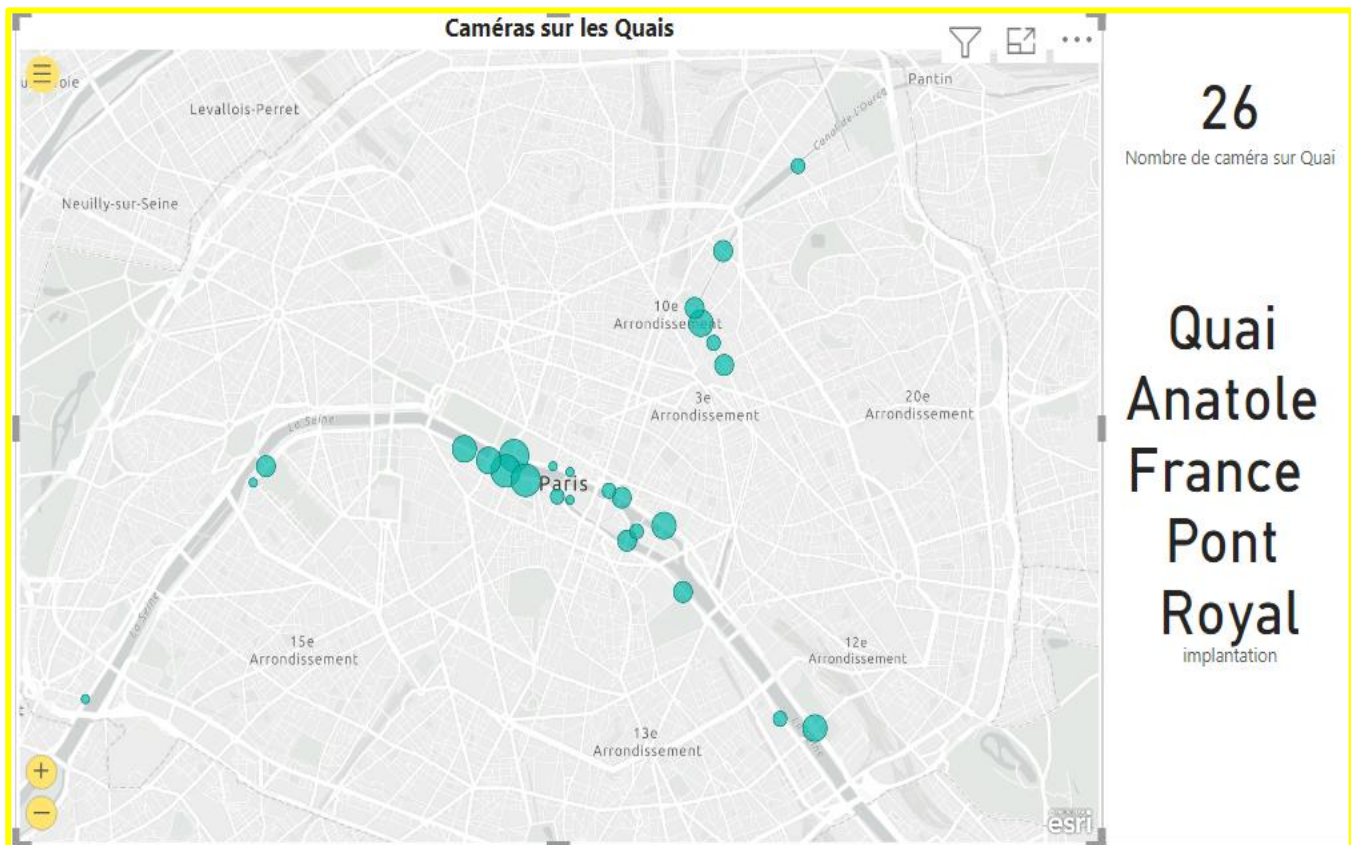
- Nombre de camera proche des quais

```
1 query_10 = '''SELECT code_cam ,implantation
2     FROM videoprotection
3     WHERE implantation LIKE ('quai%'); '''
4
5 req_10 = pd.read_sql(query_10, my_conn)
6 req_10
```



Il me semble intéressant de constater le grand nombre de caméra installées près des quais, soit 26 au total, c'est dans certains arrondissements, le même nombre de caméra observé. Cela semble logique, vu le nombre de visiteur venant flâner sur les quais lors des beaux jours.

	code_cam	implantation
0	1041	Quai des Orfèvres rue de Harlay
1	1057	Quai François Mitterrand rue de l'Amiral de C...
2	14429	Quai Saint-Exupéry rue du Général Niox
3	16887	Quai de la Mégisserie Pont Neuf
4	17103	Quai Branly avenue de Suffren
5	25626	Quai de Jemmapes rue Alibert
6	27843	Quai d'Orléans rue Jean du Bellay
7	28130	Quai de la Marne rue de Crimée
8	29770	Quai de Gesvres rue Saint-Martin
9	30922	Quai François Mauriac passerelle Simone de Be...
10	31195	Quai de Conti pont Neuf
11	39992	Quai Branly pont d'Iéna
12	45873	Quai de Jemmapes boulevard Jules Ferry
13	48209	Quai de Valmy rue Louis Blanc
14	48238	Quai de l'Archevêché pont de l'Archevêché
15	51005	Quai Saint-Bernard rue Cuvier
16	51230	Quai de Valmy rue Lucien Sampaix rue Recollet
17	59063	Quai de l'Hôtel de Ville Pont d'Arcole
18	65346	Quai de Bercy pont de Tolbiac
19	66461	Quai des Célestins pont Marie
20	71064	Quai Anatole France Pont Royal
21	73562	Quai de Valmy rue de Lancry
22	78263	Quai Anatole France rue de la Légion d'Honneur
23	86313	Quai Malaquais rue Bonaparte
24	87406	Quai Voltaire pont du Carrousel
25	99000	Quai François Mitterrand Pont du Carrousel



FOCUS sur l'IMMO

11. Quels sont les arrondissements où les transactions sont les plus fréquentes?

```

1 query_11 = '''SELECT commune.code_postal AS 'Code postal',
2   nom_commune AS 'Quartier', COUNT(*) AS 'Nb de vente'
3 FROM immo_biens
4 JOIN commune ON immo_biens.code_postal = commune.code_postal
5 GROUP BY commune.code_postal
6 ORDER BY COUNT(*) DESC
7 ; '''
8
9 req_11 = pd.read_sql(query_11, my_conn)
10 req_11

```

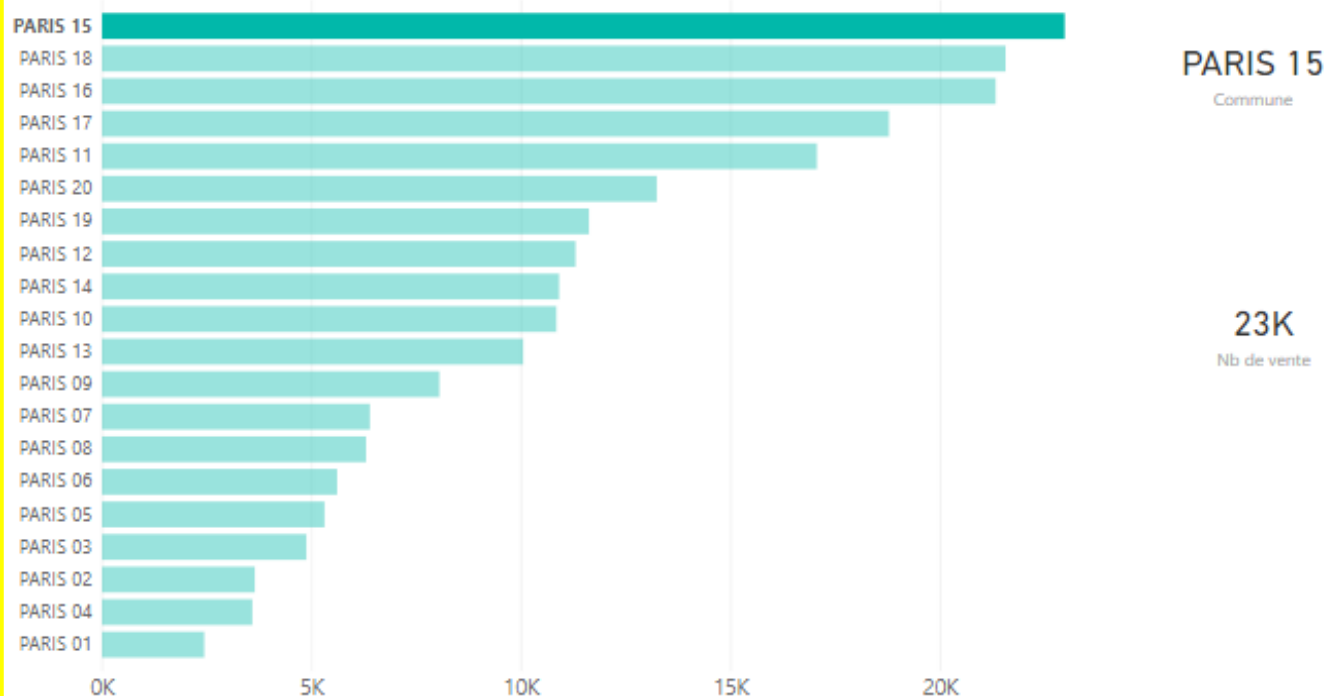
	Code postal	Quartier	Nb de vente
0	75015	Vaugirard	22976
1	75018	Buttes-Montmartre	21566
2	75016	Passy	21329
3	75017	Batignolles-Monceau	18783
4	75011	Popincourt	17063
5	75020	Ménilmontant	13243
6	75019	Buttes-Chaumont	11624
7	75012	Reuilly	11305
8	75014	Observatoire	10913
9	75010	Entrepôt	10847
10	75013	Gobelins	10055
11	75009	Opéra	8056
12	75007	Palais-Bourbon	6398
13	75008	Élysée	6307
14	75006	Luxembourg	5616
15	75005	Panthéon	5315
16	75003	Temple	4883
17	75002	Bourse	3652
18	75004	Hôtel-de-Ville	3592
19	75001	Louvre	2446



Le quartier Vaugirard (15ème) est celui qui subit le plus de transaction immobilière avec 22976 ventes entre 2016 et 2020, suivi par le 18ème et ses 21566 ventes.

Les transactions immobilières les moins nombreuses sont celle du 1^{er} arrondissement avec 2446 ventes et celle de l'Hôtel-de-ville (4ème) avec 3592 ventes.

Nb de vente par commune



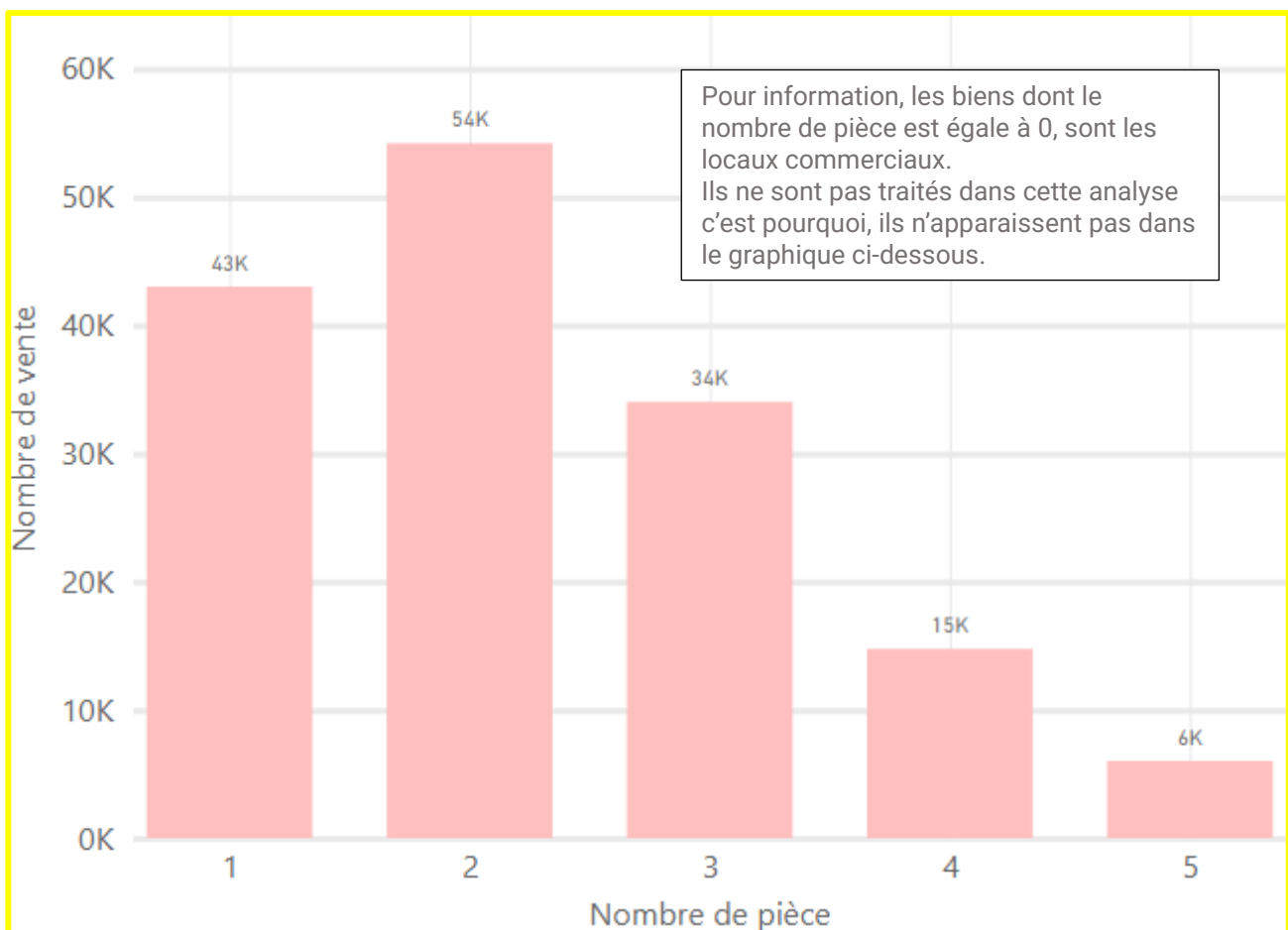
12. Quel est le top 5 des types de biens les plus vendus?

```
1 query_12 = '''SELECT nb_pieces AS 'Nb de pièces', COUNT(*) AS 'Nb de Ventes'
2 FROM immo_biens
3 GROUP BY nb_pieces
4 ORDER BY nb_pieces
5 LIMIT 6
6 ; '''
7
8 req_12 = pd.read_sql(query_12, my_conn)
9 req_12
```



Les biens les plus populaires sont les 2 pièces avec 54202 ventes, alors que les 5 pièces ne représentent que 6018 ventes.

	Nb de pièces	Nb de Ventes
0	0	60610
1	1	43016
2	2	54202
3	3	34021
4	4	14756
5	5	6018



13. Quel est la valeur foncière moyenne par arrondissement?

```

1 query_13 = '''SELECT code_postal AS 'Code postal', ROUND(AVG(valeur_fonciere),1) AS 'Prix moyen'
2 FROM immo_biens
3 NATURAL JOIN immo_vente
4 GROUP BY code_postal
5 ORDER BY ROUND(AVG(valeur_fonciere),1) DESC
6 ; '''
7
8 req_13 = pd.read_sql(query_13, my_conn)
9 req_13

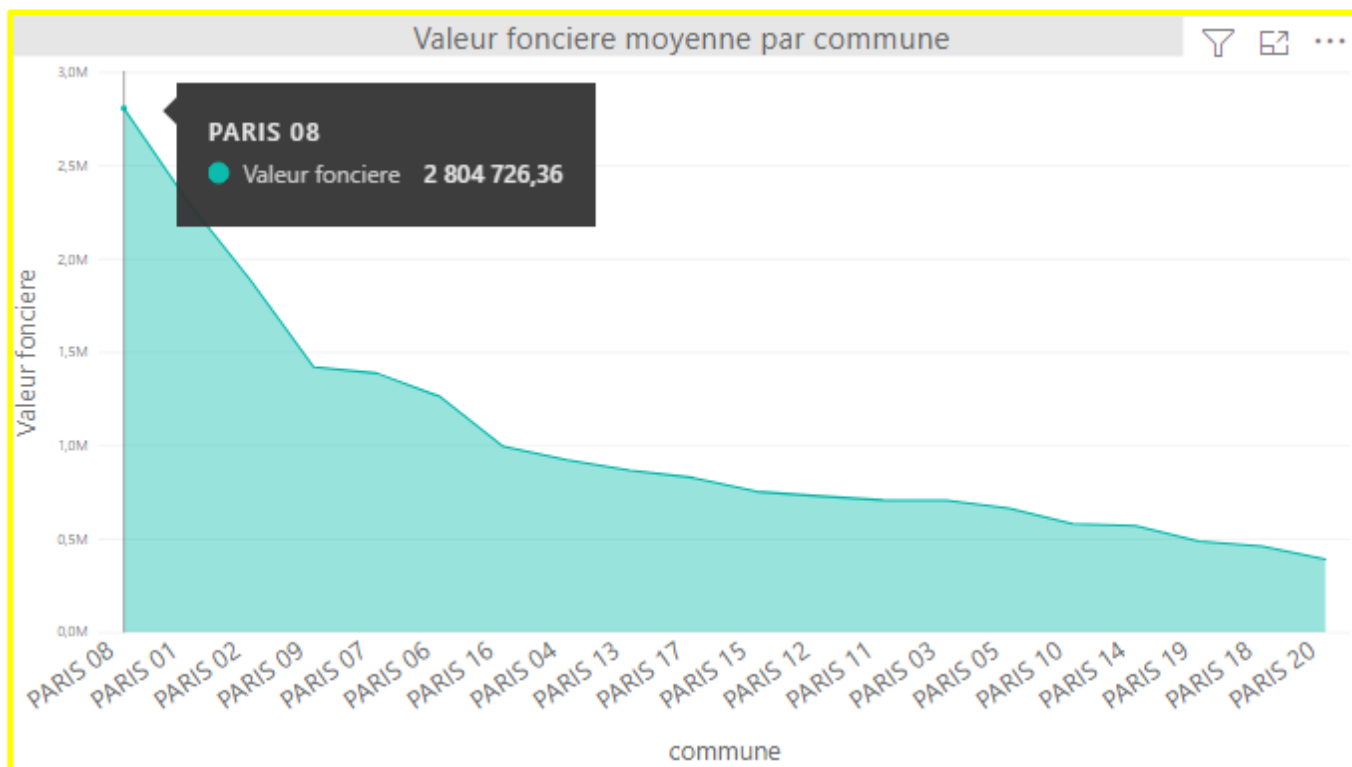
```

	Code postal	Prix moyen
0	75008	2804726.5
1	75001	2310044.6
2	75002	1884485.9
3	75009	1419366.8
4	75007	1386025.9
5	75006	1260813.1
6	75016	993139.9
7	75004	922144.8
8	75013	865229.4
9	75017	826761.6
10	75015	752260.5
11	75012	728257.3
12	75011	706386.7
13	75003	704253.6
14	75005	663404.1
15	75010	579255.0
16	75014	569584.7
17	75019	486263.4
18	75018	458664.0
19	75020	389741.7



La valeur foncière moyenne la plus importante est celle du 8^{ème} arrondissement avec 2 804 726.5 €, suivi du 1^{er} avec 2 310 044.6 €.

L'arrondissement le plus attractif financièrement est celui du 20^{ème}, avec des prix moyen avoisinant 389741.7 € et celui du 18^{ème}, 458664 € tout bien confondus. (Prix moyen basée sur les ventes de l'année 2016 à 2020)



14. Combien de transaction concernant les maisons à Paris?

```

1 query_14 = '''SELECT commune.code_postal AS 'Code postal', nom_commune AS 'Quartier',
2 ROUND(AVG(valeur_fonciere),1) AS 'Prix moyen', COUNT(type_local) AS 'Nb de vente'
3 FROM immo_biens
4 JOIN immo_type ON immo_type.code_type_local = immo_biens.code_type_local
5 JOIN immo_vente ON immo_biens.code_biens = immo_vente.code_biens
6 JOIN commune ON immo_biens.code_postal = commune.code_postal
7 WHERE immo_biens.code_type_local LIKE 1
8 GROUP BY commune.code_postal
9 ORDER BY COUNT(type_local) DESC ; '''
10
11 req_14 = pd.read_sql(query_14, my_conn)
12 req_14

```

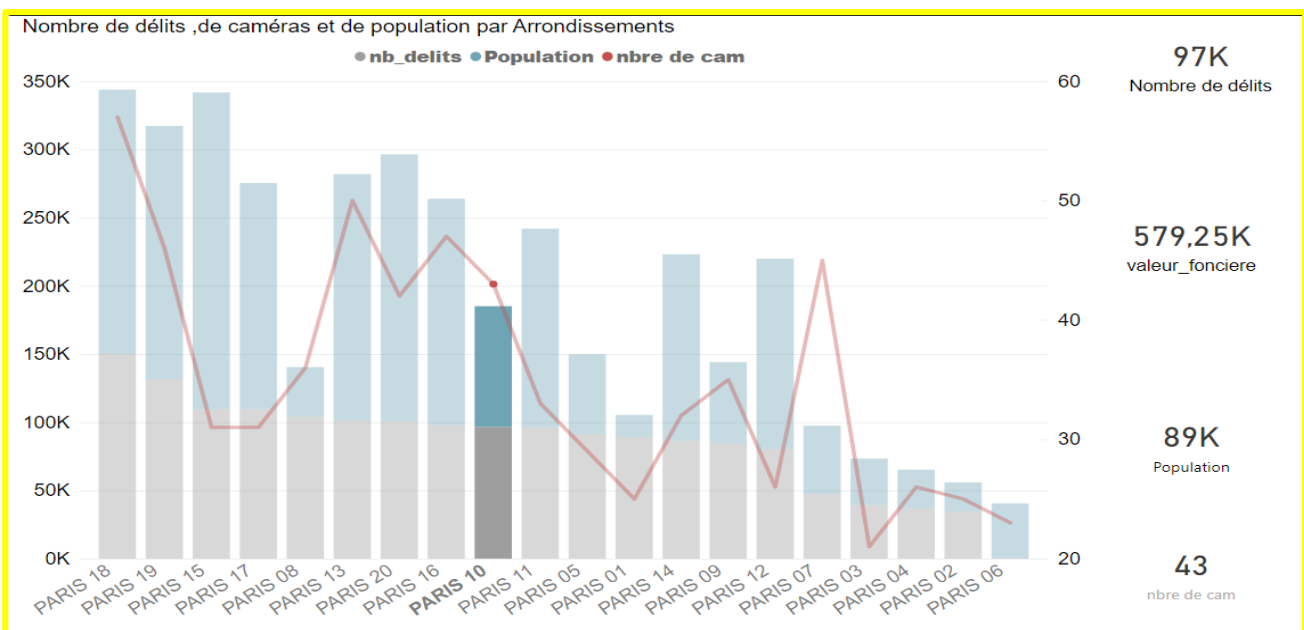
	Code postal	Quartier	Prix moyen	Nb de vente
0	75016	Passy	3958506.9	170
1	75019	Buttes-Chaumont	1330534.3	121
2	75020	Ménilmontant	1056520.6	110
3	75014	Observatoire	1458929.6	109
4	75013	Gobelins	1486092.1	72
5	75015	Vaugirard	1524076.8	60
6	75018	Buttes-Montmartre	1486413.2	51
7	75017	Batignolles-Monceau	2532799.2	39
8	75012	Reuilly	1297632.2	26
9	75007	Palais-Bourbon	16697133.7	23
10	75008	Élysée	7053292.0	10
11	75005	Panthéon	2781000.0	10
12	75006	Luxembourg	7264077.0	10
13	75011	Popincourt	1311066.7	6
14	75009	Opéra	3539679.2	6
15	75003	Temple	1500000.0	1
16	75010	Entrepôt	425500.0	1

Les maisons sont peu nombreuses intra-muros, on dénombre 170 ventes dans le quartier de Passy(16ème) avec un prix moyen de 3 958 507 € et 121 transactions aux Buttes-Chaumont (19ème) avec une moyenne estimée à 1 330 534 €

On dénombre 1 seule vente en 5 ans dans le 10ème et le 3ème arrondissement avec des prix moyens de 425500 et 1 500 000 €.

GENERAL

Quels arrondissement semble offrir le meilleur compromis ?



En prenant en compte différents critères, tels que le nombre de délits, d'habitants, le nombre de caméra et le prix moyen, le 10ème arrondissement semble être un bon compromis, attention à prendre en compte les critères personnels de chacun, d'une rue à l'autre, le quartier peut sembler bien différent.

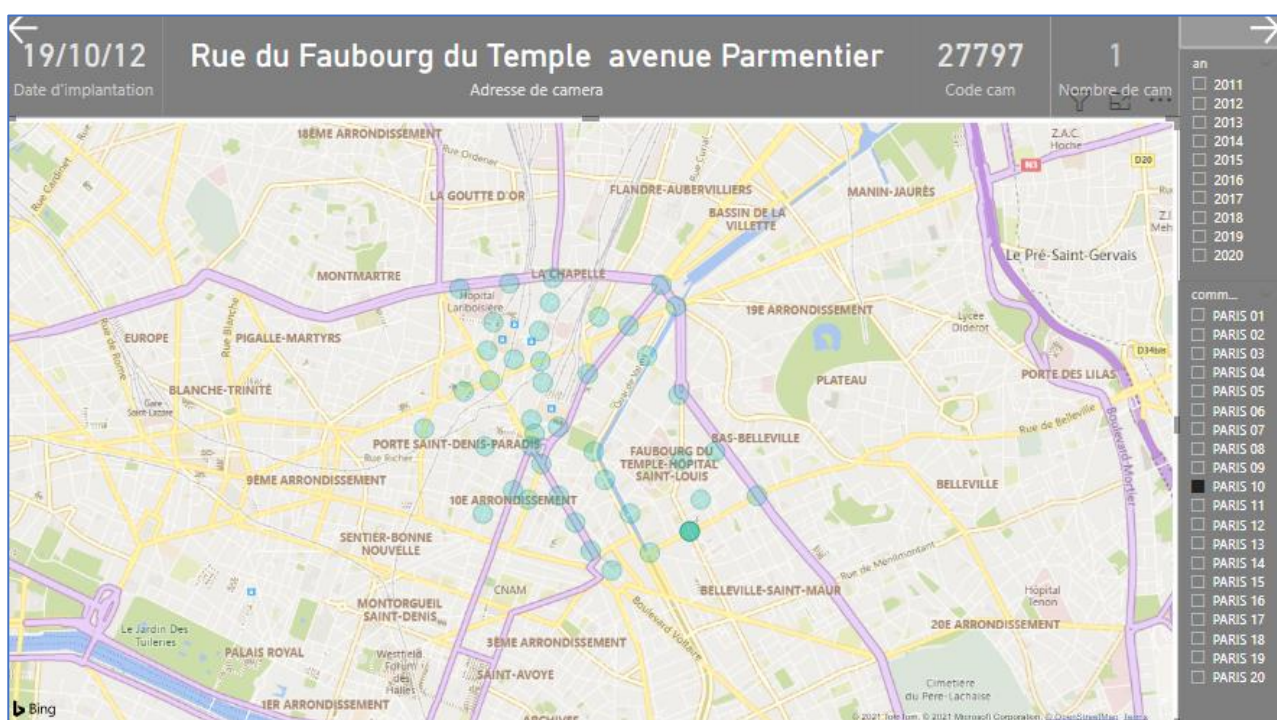
Retour sur l'**objectif final**, qui a vocation à guider un utilisateur lambda dans la quête d'un bien en fonction de ses propres critères. Au vue de la complexité de trouver un quartier idéal, il peut être intéressant de se servir d'un outil tel que celui-ci :

Voici un exemple type de recherche pour un profil : femme, d'une trentaine d'année dont le salaire est plutôt confortable, qui a tendance à se restaurer en extérieur et qui se déplace seule en transport en commun.

Sa recherche sera orientée en fonction de sa propre problématique alors qu'une famille cherchera peut être à éviter les quartiers gangrené par les stupéfiants par exemple...



Cette seconde interface tend plus à repérer les caméras installées autour de l'habitation cible, il faut savoir qu'il est possible, en cas de problème, d'indiquer au service de l'ordre quel caméra consulter si l'on a conscience de sa présence.



10. CONCLUSION

Grâce aux différentes analyses réalisées dans cette étude, nous pouvons apporter des informations déterminantes pour un individu lambda dans le choix d'un investissement immobilier, bien sur le coup de cœur aura toujours une très grande importance comparé à des données statistiques, mais il peut être intéressant de prendre les critères suivant en compte.

Nous savons que l'analyse porte sur les délits autour de l'habitation principale et son environnement, que les vols sur les lieux publics sont les plus courants alors que les vols violents sont les moins nombreux du top 10, et qu'ils touchent plus les femmes que les hommes. Malgré l'installation de la plupart des caméras en 2012, soit 1 an après le lancement du projet des « 1100 caméras », nous observons une augmentation constante de la délinquance depuis 2012 avec un pic en 2014 (16%), puis une diminution flagrante en 2020 sûrement lié au covid 19.

Certains arrondissement ressortent plus que d'autres, que ce soit positivement ou négativement. Prenons le quartier des Buttes-Montmartre (18^{ème}), il fait partie des arrondissement qui subissent le plus de cambriolage et de délit en général mais c'est aussi le mieux doté en caméra et surtout l'un des plus attractif financièrement, juste après le 20^{ème}, ce qui explique qu'il soit l'un des plus populaire en terme de transaction immobilière.

Le quartier des Buttes-Chaumont (19^{ème}), fait lui aussi parti des arrondissement les plus criminogène mais se démarque en étant l'un des mieux doté en bien de type maison, juste après le quartier de Passy (16^{ème}), qui est aussi le plus grand arrondissement de Paris en superficie.

Sans grande surprise, nous constatons qu'il y a très peu de transaction concernant les maisons dans la capitale, cela s'explique par la rareté de ce type de bien, en revanche les 2 pièces sont les biens les plus populaires.

Les arrondissement les moins criminogènes sont les quartiers de la Bourse (2^{ème}) et de l'Hôtel-de-Ville (4^{ème}), c'est aussi ceux qui subissent le moins de cambriolage, il y a très peu de transaction immobilière dans ces quartiers, cela s'explique par le fait qu'il fassent parti des plus petits arrondissement, tout comme le quartier du Louvre (1^{er}), un des plus chères, juste après le 8^{ème} et sa moyenne de prix estimé à 2 804 726 €...

Malgré toutes ces informations, il reste difficile de trouver un quartier parfait, en effet, chacun à ses propres critères, le prix, le calme ou encore l'environnement. Mais le 10^{ème} arrondissement semble offrir plusieurs avantages, ce n'est certes, pas le moins cher, donc pour un budget limité on se tournera sans doute vers des quartiers plus attractif...

Enfin, afin que chacun puisse trouver son bonheur et ce en fonction de ses critères personnelles, il peut être intéressant d'utiliser l'outil d'aide à la décision.

Concernant la vidéo protection, installée principalement sur des grands axes routiers, devant des monuments, des bâtiments publics, des lieux touristiques et de grand passage comme les quais, rappelons qu'elle a été prévue pour lutter principalement contre le terrorisme (voir même le grand banditisme), mais celle-ci n'a pas pour but pour le moment de traiter les délits de la vie courante.

Toutefois , les nouvelles technologies, en l'occurrence la vidéoprotection évolue sans cesse , et ce, grâce à l'intelligence artificiel comme la reconnaissance faciale , la détection de mouvement de foule suspecte, de plaque d'immatriculation, de chaleur et bien plus encore... Malheureusement ou heureusement en France, certaines lois ne permettent pas encore d'utiliser ces technologies. Mais qui sait ? peut-être un jour !

Houra , encore plus de donnée à analyser..... :)

11. REMERCIEMENTS

Un grand **MERCI**, à chacun de mes formateurs #Manel #Josselin #David #Nicolas.

Sans oublier Ana, et Alexia qui nous a donné notre chance et à l'équipe #Hackeuses.

Mais aussi à chacun de mes camarades pour l'entraide constante, que ce soit psychologique ou technique.

Et bien sûr, au jury de nous permettre de nous améliorer.

Et surtout...

Merci à **SIMPLON** <3

12. RESSOURCES



<https://pandas.pydata.org/>



<https://github.com/>



<https://stackoverflow.com/>



<https://www.pierre-giraud.com/>



<https://sql.sh/>