

PROJET RÉALISÉ PAR L'ÉQUIPE 7

QUELLES SONT LES VILLES LES PLUS SÛRES DE
FRANCE ?

Équipe: KAMMOUN Mohamed ali, MAKHLOUFI Khalil, VITOFFODJI
Adjimon, ROUIS Rania.



Département MIASHS, UFR 6 Informatique, Mathématique et Statistique
Université Paul Valéry, Montpellier 3

Mai 2023

SOU MIS COMME CONTRIBUTION PARTIELLE
POUR LE COURS SCIENCE DES DONNÉES 2 ET BASES DE DONNÉES

Déclaration de non plagiat

Nous déclarons que ce rapport est le fruit de notre seul travail, à part lorsque cela est indiqué explicitement.

Nous acceptons que la personne évaluant ce rapport puisse, pour les besoins de cette évaluation:

- la reproduire et en fournir une copie à un autre membre de l'université; et/ou,
- en communiquer une copie à un service en ligne de détection de plagiat (qui pourra en retenir une copie pour les besoins d'évaluation future).

Nous certifions que nous avons lu et compris les règles ci-dessus.

En signant cette déclaration, nous acceptons ce qui précède.

Signature: _____ Date: _____

Signature: _____ Date: _____

Signature: _____ Date: _____

Signature: _____ Date: _____

Remerciements

Nos plus sincères remerciements vont à notre encadrant pédagogique pour les conseils avisés sur notre travail.

Nous remercions aussi Michel Deudon pour nous avoir accompagné tout au long de ce projet et pour ces interventions, lorsque nous avons rencontrés des obstacles, pour nous avoir empêcher à prendre un retard considérable.

07/05/2023.

Résumé

Les résultats de notre étude nous ont permis de conclure que les villes avec une population moins importante présentent des taux de criminalité et des indices significativement moins élevés que les villes avec une grande population. Ainsi, **2018** la **première** ville la plus sûre de France est le **Cannet** avec un niveau de faits de délinquance **622** pour une population d'environ **41 447 habitants**, en **deuxième** position on retrouve la ville de **Montrouge** avec un niveau de faits de délinquance de **755** pour une population de **48 734 habitants**, la ville de **Vincennes** se hisse à la **troisième** position avec un niveau de faits de délinquance de **784** avec une population de **794 habitants**. Les villes **Saint-Louis, Saint-Germain-en-Laye, Meudon, Issy-les-Moulineaux, Caluire-et-Cuire, Puteaux** et **Massy**, viennent compléter le **top 10 des villes les plus sûres de France en 2018**. À la même période, la ville de **Paris** arrive en **première** dans le classement des **villes les moins sûres de France** avec un niveau de faits de délinquance de **63 918** pour une population d'environ **2 175 601 habitants**, en **deuxième** position on retrouve la ville de **Marseille** avec **31 098** de faits de délinquance enregistrés pour une population d'environ **868 277 habitants**, la ville de **Toulouse** termine le **top 3** avec un nombre de faits de délinquance **16 164** enregistrés en 2018 pour une population d'environ **486 828 habitants**. Les villes **lyon, Nice, Bordeaux, Lille, Montpellier, Strasbourg, Saint-Denis**, complètement respectivement le **top 10 des villes les moins sûres de France en 2018**. Cette insécurité est principalement influencée par la composition de la population et aussi comme nous l'avons montré dans cette étude, avec les villes ayant un grand nombre d'étudiants étant plus dangereuses. De plus, la répartition des entreprises sur le territoire français impacte également les différences de taux de criminalité entre les villes, avec celle abritant un plus grand nombre d'entreprises ayant des taux plus élevés. Les grandes villes métropolitaines, qui attirent une population plus importante en raison des opportunités d'emploi et d'autres facteurs, sont les plus menacées.

Table des matières

Chapitre 1	Contexte, définition et intérêt du projet	2
1.1	Enoncé du problème	2
1.2	Intérêt de l'étude	2
1.3	Responsabilité et composition de l'équipe	3
1.4	Objectif, hypothèse et l'imites de l'étude	3
1.4.1	Objectif de l'étude	3
1.4.2	Hypothèse de l'étude	3
Chapitre 2	Base de données	4
2.1	Nature, Source des données et description des tables	4
2.1.1	Nature et source	4
2.1.2	Description des tables	5
2.1.3	Vision générale des données	6
2.2	Modélisation des données	6
2.2.1	Modélisation conceptuelle des données	6
2.2.2	Modélisation organisationnelle des données (MOD)	7
2.3	SQL	8
2.3.1	Prétraitements	8
2.3.2	Requêtes en langage naturel et en SQL	8
Chapitre 3	Scicence de données	12
3.1	Logicieles	12
3.2	Description des Données	12
3.3	Identification des variables de l'étude.	12
3.4	Outils et modèles d'analyse	13
3.5	Méthodes descriptives	13
3.6	Analyse descriptive des Données	13
3.6.1	Resumé graphique du nombre de faits de déliquance.	13
3.6.2	Resumé graphique du nombre de faits de déliquance.	13
3.6.3	Resumé graphique du nombre d'étudiants.	14
3.6.4	Resumé graphique du nombre total d'entreprises.	14
3.6.5	Matrice de corrélation de Spearman entre les différentes variables	15
3.7	Limtes de l'étude	16
3.8	Difficultés rencontrés	16
3.8.1	Sur les données disponibles	16
3.8.2	Sur la taille des fichiers et l'import base de données.	16
3.8.3	Sur la modélisation	16

3.8.4	Sur les graphiques	16
3.9	Perspective	17
	Bibliographie	18
	Annexes	19
3.10	Requêtes en Lanquages SQL	19
3.11	ANNEXE SDD Scicences des Données	21
3.12	Tables {-} Résumé des tables (ANNEXE SDD1)	21

Introduction

Il est désormais bien délicat, la nuit tombée, de faire la différence entre la place du Commerce et la cour des miracles. Parcourez quelques mètres à pieds et vous vous verrez proposer toutes sortes de stupéfiants, comme s'il était évident que le badaud qui s'aventure seul sur le pavé ne pouvait être qu'un client en quête de drogue. La place du Commerce n'a décidément jamais aussi bien porté son nom. Il flotte dans les nuits nantaises le souvenir d'une insouciance aujourd'hui disparue."affirme les habitants de Nantes dans un reportage mené par Le Point. Plongé au centre-ville de Nantes, sixième ville de France minée par l'insécurité, c'est certainement l'un des sujets majeurs dans notre contexte actuel. La sûreté des villes est l'un des principaux enjeux pour les politiciens pour mettre en avant leurs programmes comme c'était le cas dans les élections présidentielles en 2022. "La France a peur" affirme le journaliste Roger Gicquel dans le journal télévisé de TF1 le 18 février 1976. Depuis quelques années, on assiste à une flambée des crimes et des violences, les médias en témoignent, du danger qui règne sur les villes françaises. Régulièrement, les chiffres et les faits divers sont portés en étendard pour expliquer le sentiment d'insécurité des habitants. Violences sexuelles, vols avec arme, coups et blessures contre des personnes à la rue... Misère et insécurité règnent dans les métropoles françaises que les politiques attribuent généralement à des populations marginalisées, à des hordes de migrants clandestins. D'autres facteurs menacent la sûreté des villes : l'attrait des loyers, le coût de la vie expliquent le profil contrasté des régions et des départements français. Davantage de jeunes, de population actifs dans les plus grandes unités urbaines, les gens migrent pour se rapprocher des offres d'emplois ou des villes étudiantes. La dynamique des villes, ainsi que la densité, semblent propices à l'insécurité, mais pas seule. Pour ce faire, nous disposons de plusieurs jeux de données prenant source du site internet data.gouv. Nous nous demanderons dans ce contexte:

Comment la démographie des villes peut-elle expliquer la différence des politiques de sécurité mises en place par les autorités publiques dans le territoire français ?

CHAPITRE 1

Contexte, définition et intérêt du projet

1.1 Enoncé du problème

La sécurité dans nos différentes villes constitue un facteur très déterminant dans la lutte contre l'insécurité dans nos différentes régions. Le Parisien a réalisé un classement sur la base des données du Services Statistique Ministériel de la Sécurité Intérieure (SSMSI), ainsi les régions ayant les villes les plus sûres sont : le Maine-et-Loire (note de 16,22/20). En deuxième position, on retrouve Chemillé-en-Anjou dans le même département. Montaigu-Vendée (Vendée), près de Nantes, complète le podium. Selon Renée Zauberman, directrice de recherche au CNRS et responsable de l'OSC, ce résultat lié, en partie, à la densité de population dans ces villes. (Publié le 12/11/2022 sur SudOuest.fr). « Traditionnellement, là où il y a des populations plus denses, il y a plus d'infractions de manière générale. explique-t-elle auprès du journal. (Publié le 13/11/2022 sur Le Point). Les enquêtes Cadre de vie et sécurité permettent d'évaluer les taux de plainte. Ainsi par exemple, selon ces enquêtes en moyenne sur la période 2011-2018 seules 12% des victimes de violences sexuelles hors ménage portent plainte, contre 74% pour les victimes de cambriolages. Selon une analyse de l'INSEE, la délinquance est souvent analysée à partir des taux pour 1 000 habitants ou logements. Or pour des petites communes, une faible variation de la délinquance ou de la population induit une grande volatilité de ce taux. (INSEE, 2018) Si l'on s'accorde sur le fait que les variations de la densité de la population induit une grande volatilité du taux de délinquance, il est alors digne d'intérêt de se poser des questions sur le classement des villes les plus sûres de France, et les facteurs qui peuvent influencer ce classement. Cela suscite en nous les interrogations suivantes : quelles sont les villes de plus de 40 000 habitants les plus sûres de France ? Comment la densité peut influencer le taux de délinquance ? Autant de questionnements qui suscitent l'intérêt de notre thème "Analyse sur les villes de plus de 40 000 habitants les plus sûres de France"

1.2 Intérêt de l'étude

L'importance que joue la sécurité dans le bien être de la population d'une ville n'est plus à démontrer. Certes, elle est considérée comme moteur de toute croissance économique ce qui s'explique par les différentes actions mises en oeuvre par les dirigeants du pays pour assurer la sécurité des biens et services dans chaque

ville, bien qu'il soit un fait réel, ce moteur affiche certains résultats mitigés par rapport à l'insécurité grandissante observée dans certaines villes. Les résultats de cette étude permettront sur le plan théorique ou scientifique de pouvoir fournir aux différents acteurs du pouvoir public les différents classements des villes les plus sûres de la France. L'identification de ses villes, permettra d'apporter des informations pertinentes, complémentaires à celles déjà fournies par les études précédentes, aux autorités du tutelles et aux décideurs publiques pour une bonne gestion du secteur de la sécurité . Ceci va leur permettre de tirer les conséquences sur leurs politiques en France et d'orienter de façon objective leurs politiques pour tenter de reprendre le monopole sur l'insécurité dans certaines villes et de réduire dans ces villes l'insécurité observée qui est considéré comme un handicap aux développements de ces dernières.

1.3 Responsabilité et composition de l'équipe

Pour donner à chacun d'entre nous un pouvoir de décision et d'organisation équitable dans le projet, nous avons décidé conjointement de nommer des responsabilités, en plus de la répartition des tâches de travail. Ces présentes responsabilités ne substituent guère le recueil des avis des membres de l'équipe.

- **KAMMOUN Mohamed ali** :Étudiant 22014374 Participation dans la rédaction du premier rapport et dans la création des requêtes et des graphiques
- **MAKHOUI KHALIL** : Étudiant n°22104011 Responsable de la modélisation des requetes SQL.
- **ROUIS Rania**:Étudiant 22014518 Responsable de l'introduction du sujet.Participation dans la création de requêtes sql,et dans l'élaboration de graphes dans la partie descriptive.Relecture du sujet et vérification.
- **VITTOFODJI Adjimon Jérôme** : Étudiant n°22211133 Responsable de la planification des séances,de la rédaction des documents dans Rmarkdown et de l'élaboration des graphiques dans la partie descriptive.

1.4 Objectif, hypothèse et l'imites de l'étude

1.4.1 Objectif de l'étude

L'objectif général de cette étude est de déterminer les villes les plus sûr de France. De façon spécifique il s'agit de déterminer parmi les villes de plus de 40 000 habitants, lesquelles sont les plus sûres de France.

1.4.2 Hypothèse de l'étude

Pour atteindre ces objectifs, une hypothèse est formulée :

- Hypothèse 1: Une ville avec une population élevée est plus dangereuse.
- Hypothèse 2: Les villes estudiantines sont les plus dangereuses.
- Hypothèse 3: Le nombre d'entreprises dans une ville n'a aucune influence sur le nombre de faits de délinquance enregistrés.

CHAPITRE 2

Base de données

2.1 Nature, Source des données et description des tables

2.1.1 *Nature et source*

Les données utilisées dans la présente étude proviennent d’Internet en licence libre et exploitable. En effet, ne nous somme par les auteurs de toutes ses données, ainsi donc pour permettre aux autres utilisateurs à consultants ou utiliser ses données dans leurs différents travaux, nous mettrons dans la bibliographie avec plus de précision les différentes sources de ses données.

Ces données proviennent de :

- Source de données 1 : Afin de favoriser l’ouverture des données sur la délinquance et l’insécurité, le Service statistique ministériel de la sécurité intérieure (SSMSI) met à disposition deux bases de données annuelles sur les principaux indicateurs des crimes et délits enregistrés par la police et la gendarmerie nationales, depuis 2016: l’une à l’échelle communale et l’autre à l’échelle départementale, toutes deux selon le lieu de commission.

<https://www.data.gouv.fr/fr/datasets/bases-statistiques-communale-et-departementale-de-la-delinquance-enregist>
Cette dernière est constituée de 13 colonnes et 1 000 000 lignes.

- Source de donnée 2 : MDB-INSSE. Elle comprend 99 colonnes et 36 678 lignes dont aucune valeur manquante.

<https://www.data.gouv.fr/fr/datasets/data-insee-sur-les-communes/>

Par la suite, nous avons créé et/ou complété les jeux de données, afin de référencer les différentes informations que nous n’avons pas trouvé. Cela permettra de lier toutes nos données et afficher des informations complémentaires dont nous jugeons importantes pour notre projet.

Ainsi, à partir de ces différentes bases de données, nous avons constitué notre jeu de données avec les colonnes cohérent avec le thème de notre projet d’étude. Les villes ayant une population de plus de 40 000 habitants ont été retenus dans l’étude notre projet. Cet dernier nous permis réduire non seulement le nombre de

lignes dans notre jeu de données, de ne plus avoir des valeurs manquantes dans notre base mais aussi de prendre en considération les villes qui ont une grande densité de la population.

Par la suite, nous avons ajouté à la base 1 la colonne nom de chaque ville à l'aide de la base 2 en fonction du CODEGEO de chaque ville.

Les principales variables de notre jeu de données sont :

- le code officiel géographique de la ville (**CODGEO_2022**);
- le libellé de la ville (**LIBGEO**);
- l'année (2018) pendant laquelle la délinquance a été enregistrée (**annee**);
- l'indicateur des crimes et délits (**classe**);
- l'unité de compte associé à cet indicateur (**unité.de.compte**);
- le nombre de faits de délinquance enregistré (**faits**) établis en commission de commune.
- la population municipale issu du recensement de la commune (**POP**) pour l'année précisée.

A cela s'ajoute:

- le nombre d'étudiants dans chaque ville (**Nb_Etudiants**);
- le nombre d'entreprise dans les différents secteurs services, commerce, construction, industrie;
- le nombre de la population active en âge de travailler, le nombre de salariés et le nombre de la population qui sont accès à l'éducation, à santé et les actions sociales.

(**Nb_Actifs,Nb_Actifs_Salarié,Nb_Education_santé_action_sociale**)

2.1.2 Description des tables

- **La table Commune:** dans cette table, nous avons le code officiel (Codegeo) qui est en Varchar unique pour identifier chaque commune. Elle représente notre clé primaire. Ensuite les autres attributs qu'on retrouve dans cette table sont: le nom de chaque ville (Libgeo), la population enregistré au cours de l'année 2018 (Pop),le nombre de la population actives(NbActif),le nombre d'étudiant dans cette commune (NbEtudiants) et le nombre de personne qui ont accès à l'éducation, la santé sociale(NbEducationSantesSociale). [*Source de données 1*]
- **La table Délinquant:** dans cette table, l'attribut identifiant délinquants (Id_délinquants) de type Integer est la clé primaire, l'unité de compte (unité.de.compte) qui est l'unité associé à l'indicateurs des crimes et délits est un attribut de cette table.Elle est de type Varchar. [*source de données 1*]
- **La table Accidents:** dans cette table, l'attribut identifiant accidents (Id_Accidents) de type Integer est la clé primaire, la classe de type Varchar et faits sont les attributs de cette table.[*Source de données 1*]
- **La table Entreprise:** dans cette table, le type de secteur est la clé primaire de cette table.Cette variable est de type Integer. Ensuite le NbEntrepriseServices,NbEntrepriseConstruction,NbEntrepriseCommerce et NbEntrepriseIndustrie sont des attributs de cette table. [*Source de données 2*]

- **La table Résider:** dans cette table, l'Id_Délinquants et Codegeo2022 sont les clés primaires de cette table.[*crée par les auteurs*]
- **La table Avoir-lieu:** Cette représente la ville dans laquelle l'accidents a eu lieu.[*crée par les auteurs*]
- **Causer:** dans cette table, l'Id_Délinquants et l'Id_Accidents sont les clés primaires de cette table. [*crée par les auteurs*]
- **Localiser:** Cette représente la ville dans laquelle l'entreprise est située. [*crée par les auteurs*]

2.1.3 Vision générale des données

Table 2.1: Description des types d'attribut.

Identifiants	Variables quantitative	Variables qualitative
4- Unique	Discretes : 15	Norminal: 3

Table	Lignes	Type	Interclassement	Taille	Perte
ACCIDENTS	180	InnoDB	utf8_general_ci	96,0 kio	–
Causer	180	InnoDB	utf8_general_ci	16,0 kio	–
COMMUNES	181	InnoDB	utf8_general_ci	16,0 kio	–
DÉLINQUANTS	180	InnoDB	utf8_general_ci	16,0 kio	–
ENTREPRISES	186	InnoDB	utf8_general_ci	32,0 kio	–
Résider	180	InnoDB	utf8_general_ci	16,0 kio	–
6 tables	1 087	InnoDB	utf8_general_ci	192,0 kio	0 o

Figure 2.1: Vision de la structure de la base de données.

2.2 Modélisation des données

2.2.1 Modélisation conceptuelle des données

Après avoir défini correctement nos bases de données, pour avoir un raisonnement plus concret et raisonnable tout au long de l'analyse, nous nous sommes amenés à créer un MCD (modèle conceptuel de données) avec l'aide du site [<https://www.mocodo.net/>] telle que celle visible sur la Figure 2.2 ci-dessous :

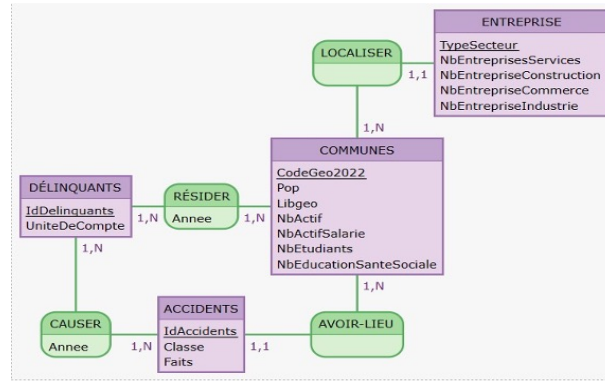


Figure 2.2: Modèle Conceptuelle de Donnée (MCD).

2.2.2 Modélisation organisationnelle des données (MOD)

À partir du Modèle Conceptuel des Données, nous avons créé le Modèle Organisationnel des données, grâce à l'outil Concepteur de MAMP - PhpMyAdmin. Lien vers l'image plus grande.

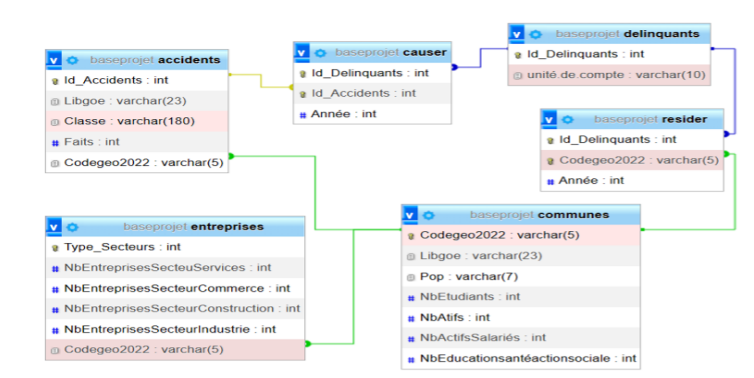


Figure 2.3: Modèle Organisationnel de Données (MCO).

La version manuscrite du MOD est :

- **ACCIDENTS** (IdAccidents, Classe, Faits, Codegeo2022);
- **causer CAUSER** (annee, IdAccidents, IdDelinquants);
- **COMMUNES** (CodeGeo2022, pop, LibGeo, NbActif, NbActifSalarie, NbEtudiants, NbEduactionSanteSociale);
- **DÉLINQUANTS** (IdDelinquants, UniteDeCompte);
- **ENTREPRISE** (TypeSecteur, NbEntreprisesServices, NbEntreprisesConstruction, NbEntreprisesCommerce, NbEntreprisesIndustrie, Codegeo2022);
- **résider** (Annee, CodeGeo2022, IdDelinquants) ;

2.3 SQL

2.3.1 Prétraitements

Afin d'importer nos données, nous avons commencé par les réunir dans nos tables des différents classeurs Excel, en faisant correspondre nos données via nos clés primaires (cf. Modélisation des données) et en croisant nos sources, avec les traitements mentionnés ci-après.

Après l'import dans SQL nous avons modifié les types de nos variables pour qu'elles aient le bon type. Ensuite, nous avons rajouté les clés primaires et étrangères qui nous ont permis de créer les liens entre nos tables, et de générer le MOD.

- Source 1 :
 - Nous avons filtré les données en fonction de l'année 2018 et en retenant uniquement les villes dont la population est supérieure ou égale à 40 000 habitants.
 - Suppression des colonnes (tauxpourmille) qui représente le nombre de faits pour mille habitants ou logements dans le cas des cambriolages, le nombre de logement issu du recensement de la commune (LOG) pour l'année précisée par une variable millésime (millLOG).
 - La colonne **Faits** a été modifiée. Cette dernière est obtenue en faisant la somme des faits enregistrés pendant l'année 2018 seulement.
- Source 2: Nous disposons de plus de 20 colonnes. Ainsi nous avons retenu uniquement 8 colonnes. Celles qu'ils pourront nous permettre d'atteindre notre objectif de recherche.

2.3.2 Requêtes en langage naturel et en SQL

Sur la table *COMMUNE*

- Savoir les principales informations sur une ville précise

```
SELECT COMMUNES.Codegeo2022, COMMUNES.Pop, COMMUNES.Libgoe
from COMMUNES
where COMMUNES.Libgoe = "Marseille";
```

- Connaître les villes les plus peuplées à partir d'un certain nombre. (Voir ANNEXE 1-3)

```
SELECT COMMUNES.Codegeo2022, COMMUNES.Libgoe, COMMUNES.Pop
from COMMUNES
where COMMUNES.Pop >= "200000"
```

- Ajout d'une colonne d'une variable qualitative remplie selon la pop de la ville.
(Voir ANNEXE 1-4)

```
ALTER TABLE communes
ADD COLUMN taille VARCHAR(10);
UPDATE communes SET taille= "petite" WHERE communes.Pop < 100000;
UPDATE communes SET taille= "moyenne" WHERE communes.Pop
BETWEEN 100000 AND 200000;
UPDATE communes SET taille= "grande" WHERE communes.Pop > 200000 ;
```

Sur la table COMMUNES ET ACCIDENTS

- Savoir combien de faits de délinquance ont été enregistrés dans une ville au choix.

```
SELECT COMMUNES.Codegeo2022, COMMUNES.Libgoe, ACCIDENTS.Faits
from COMMUNES, ACCIDENTS
Where COMMUNES.Codegeo2022 = ACCIDENTS.Codegeo2022
And COMMUNES.Libgoe = "Paris"
```

- Dans quelle ville a lieu le plus d'accidents.(Voir ANNEXE 1-5)

```
SELECT COMMUNES.Codegeo2022, COMMUNES.Libgoe, ACCIDENTS.Faits, COMMUNES.Pop
from COMMUNES, ACCIDENTS
where ACCIDENTS.Codegeo2022 = COMMUNES.Codegeo2022
and ACCIDENTS.Faits = ( SELECT Max(ACCIDENTS.Faits) FROM ACCIDENTS)
```

- Afficher le nombre de ville dont le nombre de Faits est supérieur à une valeur saisi.

```
SELECT count(ACCIDENTS.IdACCIDENTS)
from ACCIDENTS
where ACCIDENTS.Faits>"1525"
```

- Connaître le nombre d'étudiant et de faits de délinquance d'une ville choisi aléatoirement. (Voir ANNEXE 1-6)

```
SELECT COMMUNES.Libgoe,COMMUNES.NbEtudiants,ACCIDENTS.Faits
from COMMUNES, ACCIDENTS
where ACCIDENTS.Codegeo2022 = COMMUNES.Codegeo2022
AND ACCIDENTS.IdAccidents =
(SELECT ACCIDENTS.IdAccidents FROM ACCIDENTS ORDER BY RAND() LIMIT 1)
```

Sur la table DÉLINQUANTS

- Ajouter un nouveau délinquants à la liste. (Voir ANNEXE 1-7)

```
INSERT INTO delinquants (delinquants.IdDelinquants,  
delinquants.UnitéDeCompte)  
VALUES (183, "Autre coup et blessures volontaires" )
```

Sur les tables DELINQUANTS, COMMUNES, Resider

- Connaître le lieu où s'est dérouler une délinquance.

```
SELECT Resider.IdDelinquants, COMMUNES.Libgoe,COMMUNES.Codegeo2022  
FROM DELINQUANTS, COMMUNES, Resider  
WHERE DELINQUANTS.IdDelinquants = Resider.IdDelinquants  
AND Resider.Codegeo2022 = COMMUNES.Codegeo2022  
AND DELINQUANTS.IdDelinquants = "26"
```

Sur la table ENTREPRISE

Sur les tables COMMUNES, ACCIDENTS, ENTREPRISES

- Afficher le nombre total d'entreprise de la ville où il y a eu le moins de faits de délinquances.(Voir ANNEXE 1-8)

```
SELECT COMMUNES.Libgoe,ACCIDENTS.Faits,ENTREPRISES.NbEntreprisesSecteuServices  
FROM ENTREPRISES, COMMUNES, ACCIDENTS  
WHERE ENTREPRISES.Codegeo2022 = COMMUNES.Codegeo2022  
and COMMUNES.Codegeo2022 = ACCIDENTS.Codegeo2022  
and ACCIDENTS.Faits =( SELECT MIN(ACCIDENTS.Faits) FROM ACCIDENTS)
```

- Afficher les trois principaux facteurs (population, nombre d'étudiants, nombre d'entreprises) des 10 villes les plus ou moins peuplés.(Voir ANNEXE 1-9)


```

SELECT* FROM(SELECT COMMUNES.Codegeo2022,COMMUNES.Libgoe,ACCIDENTS.Faits,
                COMMUNES.Pop,COMMUNES.NbEtudiants,
                ENTREPRISES.NbEntreprisesSecteuServices +
                ENTREPRISES.NbEntreprisesSecteurCommerce +
                ENTREPRISES.NbEntreprisesSecteurConstruction+
                ENTREPRISES.NbEntreprisesSecteurIndustrie As
NbEntrepriseTotal
FROM ENTREPRISES, COMMUNES, ACCIDENTS
WHERE ENTREPRISES.Codegeo2022 = COMMUNES.Codegeo2022
and COMMUNES.Codegeo2022 = ACCIDENTS.Codegeo2022
Order by ACCIDENTS.Faits desc
Limit 10
) AS T1
UNION
SELECT * FROM(SELECT
COMMUNES.Codegeo2022,COMMUNES.Libgoe,ACCIDENTS.Faits,
                COMMUNES.Pop,COMMUNES.NbEtudiants,
                ENTREPRISES.NbEntreprisesSecteuServices +
                ENTREPRISES.NbEntreprisesSecteurCommerce +
                ENTREPRISES.NbEntreprisesSecteurConstruction+
                ENTREPRISES.NbEntreprisesSecteurIndustrie As
NbEntrepriseTotal
FROM ENTREPRISES, COMMUNES, ACCIDENTS
WHERE ENTREPRISES.Codegeo2022 = COMMUNES.Codegeo2022
and COMMUNES.Codegeo2022 = ACCIDENTS.Codegeo2022
Order by ACCIDENTS.Faits ASC
Limit 10
) AS T2;

```

CHAPITRE 3

Science de données

3.1 Logicieles

Dans la réalisation de notre projet, nous avons utilisé le logiciel excel(2016) pour constituer nos différentes tables. Ensuite nous avons importer ces tables sous format csv dans le logiciel sql PhpMyAdmin à l'aide de la version gratuite du logiciel MAMP pour certain et WAMP pour d'autres afin de relier à travers les clés primaires et étrangère, les différentes tables. Nous avons utiliser la version gratuite de MAMP (6) Par la suite nous avons exporter au format sql nos table sur le serveur filess.io afin d'interagir depuis le logiciel Rstudio avec nos différentes tables.

Pour la rédaction de notre projet, nous avons utiliser la version 28-1-2 du logiciel Rstudio sous Mac pro à l'aide de R Markdown pour générer directement notre document en format pdf. L'analyse descriptive a été faite aussi avec ce dernier.

3.2 Description des Données

Les données de notre étude sont typiquement annuelles couvrant les année 2014 à 2022. Ainsi, dans la réalisation de ce projet, seul les données couvrant l'année 2018 ont été retenu. Ainsi, à partir différentes données, nous avons constitué notre jeu de données avec les colonnes en adéquation avec le thème de notre projet d'étude. Les villes ayant une population de plus de 40 000 habitants ont été retenus dans l'étude notre projet. Cet dernier nous permis réduire non seulement le nombre d'individus dans notre jeu de données, de ne plus avoir des valeurs manquantes dans notre base mais aussi de prendre en considération les villes qui ont une grande densité de la population.

Afin d'avoir le nombre de faits de délinquance enregistrés (faits) établis en commune de commission pour l'année 2018, nous avons à l'aide d'un tableau croisé dynamique fait la sommes des (faits) de déliquescence enregistré dans chaque commune. A l'aide des requêtes en SQL, nous avons déterminer le nombre total d'entreprises de chaque ville Le

Nous distinguons 180 observations et 15 variables. (Voir Table en ANNEXE SDD1)

3.3 Identification des variables de l'étude.

Eu égard à la littérature et aux travaux empiriques effectués sur la sureté des villes, tout en tenant compte des problèmes liés à l'indisponibilité des données sur une longue période, nous retenons pour l'étude, les variables suivantes :

- **La variable endogène:** le code géographique de la commune.
- **Les variables exogènes:** le nombre total de faits de délinquance enregistrés établis en commune de commission d'une part, La population (Pop), Nombre Etudiants, et le nombre total d'entreprises d'autres part.

3.4 Outils et modèles d'analyse

Dans le cadre de notre étude, les méthodes employées sont descriptives. Les méthodes descriptives ont pour but d'appréhender l'évolution des variables d'intérêt et de valider ou de réfuter les hypothèses de recherche.

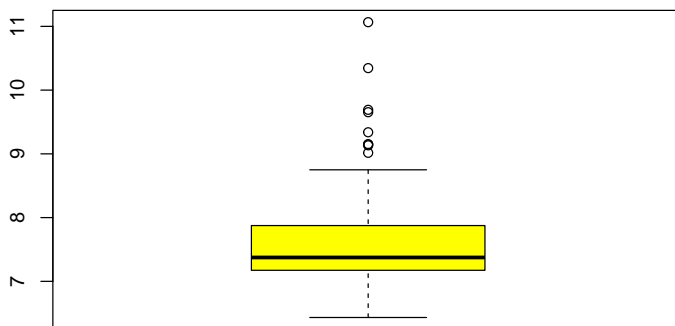
3.5 Méthodes descriptives

Compte tenu de la structure de nos données, nous procéderons à une analyse descriptive des différentes variables. Nous nous intéressons dans un premier temps à une représentation graphique des données de nos modèles à travers un boxplot. Par suite, nous analyserons la liaison entre les différentes variables exogènes avec la variable endogène à travers la matrice de corrélation avec la méthode de Pearson. Nous réaliserons ces analyses sous le logiciel Rstudio.

3.6 Analyse descriptive des Données

3.6.1 *Resumé graphique du nombre de faits de délinquance.*

L'analyse de ce graphique révèle que le nombre médian de faits de délinquance enregistrés est de 7,4%. Cette valeur est encadrée par le premier quartile et le troisième quartile respectivement 7,1% et 7,9%. Par ailleurs, on note quelques valeurs atypiques se situent au-delà de la frontière haute (7,9%).



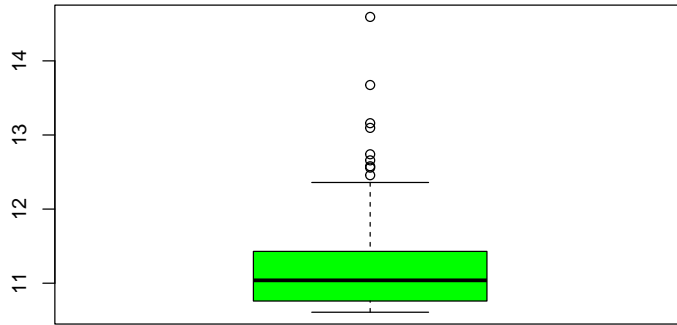


Figure 3.2: Boite à moustache de la densité de la Population

3.6.3 *Resumé graphique du nombre d'étudiants.*

L'analyse de ce graphique révèle que le nombre L'analyse de ce graphique révèle que le nombre médian d'élève enregistrés est de 8,3%. Cette valeur est encadrée par le premier quartile et le troisième quartile respectivement 8,1% et 9%. Par ailleurs, on note quelques valeurs atypiques se situent au-delà de la frontière hausse (9%) et une valeur atypique se situe en dessous de la frontière basse (8,3%).

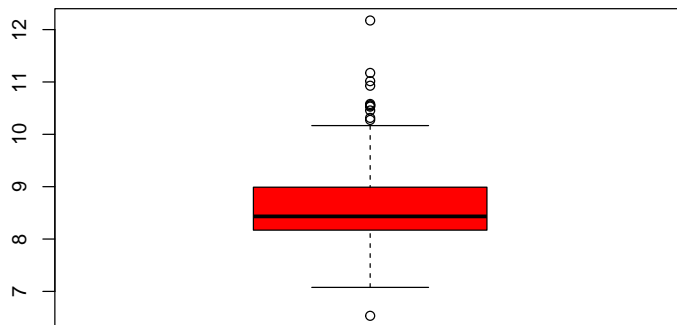


Figure 3.3: Boite à moustache du nombre d'étudiants

3.6.4 *Resumé graphique du nombre total d'entreprises.*

L'analyse de ce graphique révèle que le nombre médian d'entreprise total enregistrés est de 8,2%. Cette valeur est encadrée par le premier quartile et le troisième quartile respectivement 7,9% et 8,9%. Par ailleurs, on note quelques 04 valeurs atypiques se situent au-delà de la frontière hausse (7,9%) et et une valeur atypique se situe en dessous de la frontière basse (7,9%).

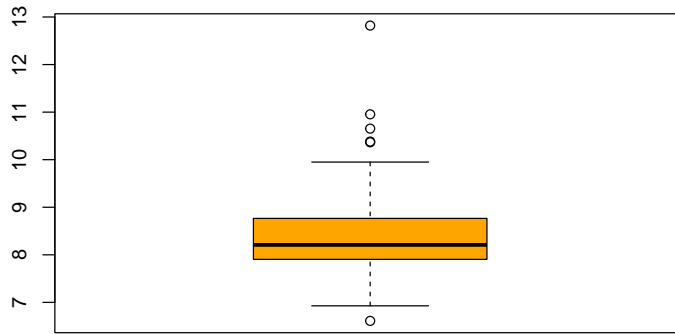
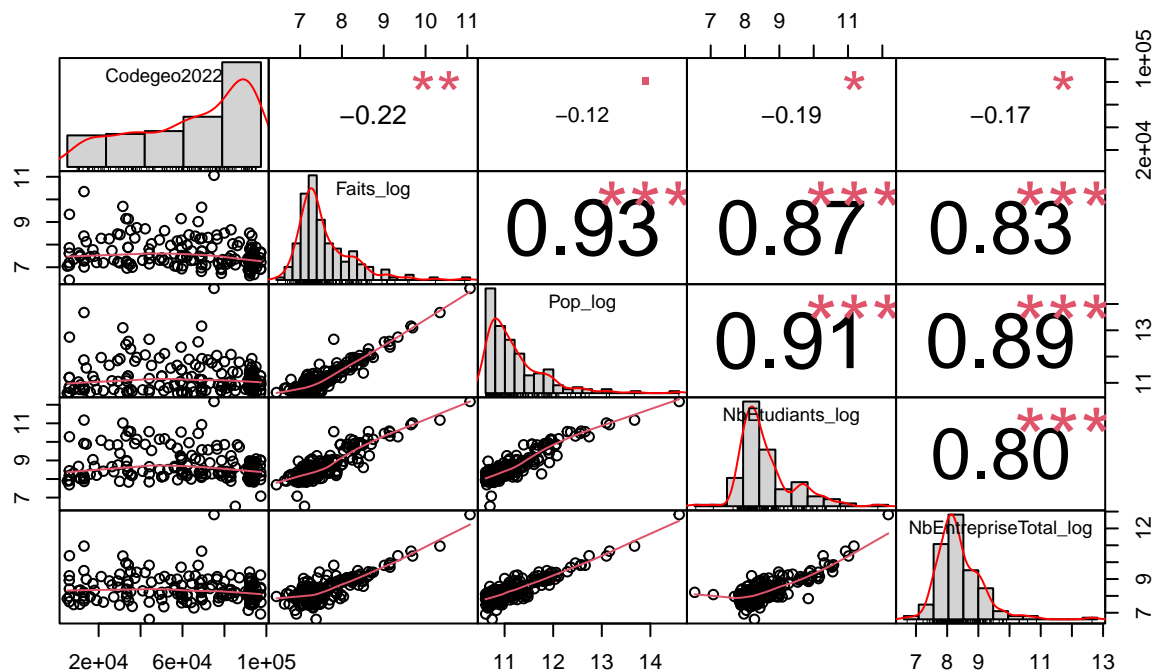


Figure 3.4: Boite à moustache du nombre total d'entreprises

3.6.5 Matrice de corrélation de Spearman entre les différentes variables

De l'analyse de cette matrice de corrélation, on note qu'il y a une liaison négative entre le faits de délinquance enregistrés et le nom de la ville. Tandis qu'on note une forte liaison positive entre le nombre de faits de délinquance, le taille de la population, la ville estudiantine et le nombres total d'entreprises dans cette ville. Autrement dit, toute augmentation de la densité de la population entraine une forte augmentation des faits de délinquance dans cette ville. Ce qui explique le fait que les villes les plus peuplées sont les plus dengereuse. De même toute une augmentation du nombre total d'entreprises ou du nombre d'élève dans une ville entraine une forte augmentation de faits de délinquance dans ces villes. La diminution des faits de délinquance dans ces villes sera effective à long terme.



3.7 Limites de l'étude

- Notre étude s'est basé dans un second temps uniquement sur une analyse descriptive à travers l'interprétation des graphes. Ainsi donc, nous nous sommes pas allés en profondeurs de l'étude, en faisant une analyse approfondie à travers la modélisation de nos modèles afin de vérifier nos hypothèses.
- Nous nous sommes pas assuré de la distribution des variables de cette étude, ni de vérifier à travers les test d'hypothèses de la validation de nos modèles.

3.8 Difficultés rencontrés

3.8.1 *Sur les données disponibles*

- Lors de nos recherches des données, nous n'avons pas réussi à trouver des données annuelles allant de la période 2014 à 2022 afin de faire une analyse en série temporelle ainsi donc nous nous sommes limités à faire nos analyses pour l'année 2018.
- La première étape était de créer une base à l'aide de données issues d'internet. Nous avons pris du temps pour avoir une base parfaite car des tables comportaient des données manquantes, le choix de garder ou supprimer des tables était un peu compliqué. De plus, nous avons dû aller chercher des données issues d'autres sources afin d'avoir une base complète.

3.8.2 *Sur la taille des fichiers et l'import base de données.*

- Certaines bases de données étaient trop volumineuses alors nous avons dû filtrer nos données afin de sélectionner uniquement les villes dont la taille de la population est supérieure à 40 000 habitants.
- Nous n'avons pas importé directement nos données sur le serveur en ligne, alors nous avons dans un premier temps, exporté depuis excel au format csv nos tables dans PhpMyAdmin, et après nous avons pu exporter sous format SQL nos différentes tables sur le serveur en ligne filess.io pour pouvoir interagir avec la base en ligne.
- Nous n'avons pas pu avoir tous accès à la base en ligne depuis Rstudio, nous rencontrons un problème de max_User.

3.8.3 *Sur la modélisation*

- Dans un premier temps, nous avons réalisé un MCD qui comportait des tables où l'on avait aucune donnée. Nous avons dû les supprimer et alors écrire un autre modèle. À travers ce deuxième modèle, nous avons des données pour toutes nos tables mais plus qu'un seul problème persistait encore, c'était le choix des clés primaires. Au moment où nous avons mis comme clé primaires le Code officiel géographique le MCD était alors correct.

3.8.4 *Sur les graphiques*

- Élaboration de plusieurs types de graphes mais qui ne faisaient peu ou même aucun lien avec nos hypothèses. D'autre part, certains graphes étaient difficiles à réaliser car les chunk ne reconnaissaient pas nos requêtes.

3.9 Perspective

D'après les résultats de notre étude, nous avons constaté que les villes avec une forte densité de population avaient tendance à présenter des taux et indices de criminalité plus élevés que les villes de petite taille. Cette tendance s'explique par le fait que les grandes villes offrent de nombreuses opportunités économiques, sociales et culturelles, ce qui attire un grand nombre d'individus, y compris des étudiants et des entreprises qui cherchent à s'y installer.

Cependant, ces mêmes facteurs peuvent contribuer à l'insécurité dans les grandes villes, augmentant ainsi les risques pour les habitants. Pour remédier à cette situation, les résultats de notre étude peuvent être utilisés pour informer les politiques et les programmes visant à réduire la criminalité dans les villes. Ces initiatives peuvent inclure le renforcement des effectifs de la police dans les zones les plus à risques, la création d'un réseau de caméras de surveillance connecté à des équipes opérationnelles, la revitalisation des quartiers, l'émission d'amendes pour les infractions et les délits, ainsi que des campagnes de sensibilisation pour lutter contre la violence.

CONCLUSION

Pour conclure, on pourrait dire que l'insécurité urbaine n'est pas un mythe mais une réalité qui crée une véritable pression sur les infrastructures et les services de nos villes françaises. Les forts taux de criminalité qu'on témoigne chaque jour sont largement impactés par la densité croissante de la population, ce qui constitue un véritable enjeu pour les autorités locales qui ne cessent de développer des programmes et des stratégies adaptées afin d'assurer la sécurité de leurs populations. Cependant, l'étude de l'insécurité des villes est une question complexe et multidimensionnelle qui mérite une étude encore plus approfondie pour mieux comprendre ce phénomène et améliorer davantage la sécurité des villes et la qualité de vie des habitants.

Bibliographie

- Sandra Bringay. *Module de cours création de base de donnée,SQL*.
- P.Lafaye de Micheaux,(Semestre 2 / L2 / 2022–2023).*Science des données 2*.
- Pierre Lafaye de Micheaux,Rémy Drouilhet et Benoit Liqueur,(2eme édition). *Le logiciel RMaîtriser le langage Effectuer des analyses (bio)statistiques*.
- Julien Barnier,(2023-04-02).*Introduction à R et au tidyverse*.
- Malo Ait Yahia et al.,(Version du 27 février 2023). *Rapport de Bases de données*.
- Jéhovahni G-B. M. SODJINOUE et Jérôme Adjimon. G. VITOFFODJI.(Juin 2019).*Étude de la demande d'essence dans les stations-services de la SONACOP de 2012 à 2017: approche économétrique*.

Annexes

3.10 Requêtes en Lanquages SQL

- Annexe 1-1 : Connexion au serveur en lignes

```
#install.packages("RMySQL")
#install.packages("DBI")
library(DBI)
library(RMySQL)
bd <- DBI::dbConnect(RMySQL::MySQL(),
host = "15n.h.filess.io",
port = 3307,
username = "Suretedesvilles_everybody",
password = "0e81a5c6d256045ed145cf9addba62624ac145cc",
dbname = "Suretedesvilles_everybody",
unix.sock="/var/lib/mysql/mysql.sock"
)
```

- Annexe 1-2:Resumer des tables

```
SHOW tables;
```

Table 3.1: 6 records

Tables_in_Suretedesvilles_everybody
ACCIDENTS
COMMUNES
Causer
DELINQUANTS
ENTREPRISES
Resider

- Annexe 1-3 Connaître les villes peuplées à partir d'un certain nombre.

Table 3.2: Displaying records 1 - 10

Codegeo2022	Libgoe	Pop
10387	Troyes	61996
11069	Carcassonne	46513

Codegeo2022	Libgoe	Pop
11262	Narbonne	55375
13004	Arles	51031
13005	Aubagne	47208
13055	Marseille	868277
13056	Martigues	48420
13103	Salon-de-Provence	45400
16015	Angoulme	41711
17300	La Rochelle	76275

- Annexe 1-4 : Ajout d'une colonne d'une variable qualitative remplie selon la pop de la ville:

Aix-en-Provence	143097	moyenne
Arles	51031	petite
Aubagne	47208	petite
Marseille	868277	grande

- Annexe 1-5: Dans quel vie il y a eu le plus d'accidents.

Table 3.3: 1 records

Codegeo2022	Libgoe	Faits	Pop
75056	Paris	63918	2175601

- Annexe 1-6: Connaître le nombre d'étudiants et de faits de délinquance d'une ville choisi aléatoirement.

Table 3.4: 0 records

Libgoe	NbEtudiants	Faits
--------	-------------	-------

- Annexe 1-7: Ajouter un nouvel unité de délinquance à la liste

IdDelinquants	UniteDeCompte
183	Autre coup
180	victime
179	infraction

- Annexe 1-8 : Afficher le nombre total d'entreprise de la ville où il y a eu le moins de faits de délinquances.

Table 3.5: 1 records

Libgoe	Faits	NbEntrepriseTotal
Le Cannet	622	2880

- Annexe 1-9: Afficher les trois principaux facteurs (population, nombre d'étudiants, nombre d'entreprises) des 10 villes les plus ou moins peuplées.

Codegeo2022	Libgoe	Faits	Pop	NbEtudiants	NbEntrepriseTotal
75056	Paris	63918	2175601	193636	369297
13055	Marseille	31098	868277	71159	57164
31555	Toulouse	16164	486828	60792	31742
69123	Lyon	15574	518635	55711	42321
6088	Nice	11368	341032	29004	32244
33063	Bordeaux	9412	257068	30147	20941
59350	Lille	9410	233098	37365	16356
34172	Montpellier	9261	290053	39317	18797
67482	Strasbourg	8234	284677	34725	18231
97411	Saint-Denis	7759	150535	14086	10078
6030	Le Cannet	622	41471	2185	2880
92049	Montrouge	755	48734	2724	1844
94080	Vincennes	794	49635	2663	3410
97414	Saint-Louis	823	53589	3315	4447
78551	Saint-Germain-en-Laye	835	44750	3551	3283
92048	Meudon	844	45748	3312	1850
92040	Issy-les-Moulineaux	881	68260	3797	4061
69034	Caluire-et-Cuire	912	42847	3293	2755
92062	Puteaux	953	44837	2756	2701
91377	Massy	973	50632	3098	1270

3.11 ANNEXE SDD Scicences des Données

3.12 Tables {-} Résumé des tables (ANNEXE SDD1)

Variables	Descriptions	Sources	Méthodes de calculs et Observations
Codegeo	C'est le code officiel géographique de la commune	Base communale	Extraite de la base de la base communale
Année	C'est l'année pendant laquelle la délinquance a été enregistrée	Base communale	Extraite de la base de la commune

Indicateur des crimes et délits (classe)	c'est l'indicateur des crimes et délits au cours de l'année 2018	Base commune	Extraite de la base de la commune
Unité de compte (UnitéDe-Compte)	C'est l'unité de compte associée à cet indicateur	Base communale	Extraite de la base la commune
Le nombre de faits de délinquance enregistrés (faits)	C'est le nombre total de faits de délinquance enregistrés pendant l'année 2018	Auteurs	Elle est calculé en faisant la somme de tout les faits enregistrés au cours de l'année 2018
La population (Pop)	C'est la population municipale issue du recensement de la commune	Base communale	Extraite de la base de la commune
Nom de la commune (Libgeo)	C'est le libellé de la commune	INSEE	Extraite de la base de l'INSEE
Nombre Etudiants (NbEtudiants)	C'est le nombre d'étudiants de la commune	INSEE	Extraite de la base de l'INSEE
Nombre de la population Actifs (NbAtifs)	C'est le nombre de la poplation active	INSEE	Extraite de la base de l'INSEE
Nombre de la salarié actfs (NbActifs-Salariés)	C'est le nombre de salarié acrifts	INSEE	Extraite de la base de l'INSE.
Nombre Education Santnté action sociale (NbEducation-Santntéactionso-ciale)	C'est le nombre de la population qui a accès à l'éducation, la santé et qui bénéficie des actions sociales	INSEE	Extraite de la base de l'INSEE
Nombre entreprise service (NbEntrepriseSecteurServices)	C'est le nombre d'entreprise dans le domaine des services	INSEE	Extraite de la base de l'INSEE

Nombre entreprise commerce (NbEntreprises- SecteurCom- merce)	C'est le nombre d'entreprise dans le domaine des commerces	INSEE	Extraite de la base de l'INSEE
Nombre entreprise construction (NbEntreprises- SecteursCon- struction)	C'est le nombre d'entreprise dans le domaine de la construction	INSEE	Extraite de la base de l'INSEE
Nombre entreprise dans l'industrie (NbEntreprises- SecteurIndus- trie)	C'est le nombre d'entreprise dans le domaine de l'industrie	INSEE	Extraite de la base de l'INSEE

Table 3.6: Description et source des variables