

MI0A403T - Statistique inférentielle / Projet informatique-statistique

Andrew El Kahwaji, Wael Aboulkacem, Hans Kanen Soobbooroyen

09/05/2025

Contents

1. Objectif du Projet	1
2. Partie Informative	2
2.1 C'est quoi un incendie	2
2.2 Causes des Incendies	2
2.3 Consequences des Incendies	2
3. Partie Informatique	2
2.1 Definition de quelques terme Informatiques	2
2.2 Bibliotheques Utilisees	3
2.3 Création de la base de données	3
2.4 Creation des Tables	4
2.5 Injection des donnees	4
2.6 Affichage des donnees	5
2.7 Exportation des donnees sous forme CSV	6
2.8 Menu du Programme	6
2.9 La Table Incendies-Departements	7
4. Partie Statistique	8
4.1 Definitions des concepts statistiques	8
4.2 Description des donnees	9
4.3 Analyse des doneees	9
5. Ressources	9
5.1 Ressources sur la Partie Informative	9
5.2 Ressources sur la Partie Informatique	9
5.3 Ressources sur la Partie Statistique	10

1. Objectif du Projet

Ce projet qui est liée a l'UE Statistique Inferentielle / Projet Stat-Info qui vise à mieux comprendre les raisons et cause des incendies en analysant des données statistiques. L'idée principale est de voir comment différentes variables influencent l'étendue des incendies, en utilisant des outils statistiques et informatiques.

Nous avons structuré notre rapport en trois sections principales : la Partie Informative, qui servira à présenter des informations générales afin d'aider le lecteur à comprendre notre projet avant de plonger dans les détails et les spécificités des sections Informatique et Statistique. La section Informatique va examiner minutieusement les techniques que nous avons mises en œuvre dans notre projet, en précisant toutes les informations indispensables. Et pour finir, la section Statistique qui nous aidera à détailler toutes les études que nous avons réalisées avec les diagrammes appropriés, l'interprétation et la solution de nos enjeux.

Finalement, nous avons constitué une section supplémentaire qui nous sert à énumérer tous les sites internet que nous avons consultés pour la rédaction et l'élaboration de notre projet. Il est à noter que les références sont présentées au format APA !

Il faut également noter que lors de notre projet, qui se divisait en deux parties : Informatique et Statistique, nous avons utilisé GitHub. Cette plateforme collaborative nous a permis de travailler collectivement sur un code en définissant les étapes attribuées à chaque membre du groupe. Cette plateforme nous offre aussi la possibilité de fusionner tout le code en un unique fichier, sans nécessité de l'assembler manuellement.

2. Partie Informatique

Dans cette partie, nous allons considérer des données informatives avant de passer à la description de notre section Informatique et Statistique.

2.1 C'est quoi un incendie

L'incendie est un phénomène de combustion incontrôlée dans le temps et l'espace, dont la principale caractéristique est sa capacité à se propager rapidement.

Pour qu'une combustion puisse se produire, trois éléments habituellement réunis dans le « triangle du feu » sont indispensables : un matériau combustible, un agent comburant et une source d'énergie d'activation.

2.2 Causes des Incendies

Les raisons des incendies sont multiples, cependant, une grande majorité d'entre eux provient d'une action humaine. Comme la négligence, malveillance, préparation insuffisante aux catastrophes naturelles comme les séismes, les tsunamis

2.3 Conséquences des Incendies

Les effets des incendies sont nombreux et graves. Elles ont des conséquences sur l'homme (asphyxie, intoxication due aux fumées, brûlures sévères), sur les entreprises (diminution de la production, dégâts matériels, licenciements) et sur l'environnement (contamination de l'air et de l'eau, ravage du paysage). Les principales causes de décès liés aux incendies sont l'intoxication par le monoxyde de carbone et la diminution de l'oxygène, plutôt que les flammes elles-mêmes. Pour plus d'informations, veuillez consulter l'article intégral ici.

3. Partie Informatique

La partie Informatique de notre Projet consiste à effectuer les démarches suivantes :

1. Création de la Base de données
2. Création de la connexion entre la Base de données et notre code source
3. Établissement des Tables dans la Base de Données
4. Insertion de données dans les tables
5. Présenter les informations des tables dans la console
6. Exportation des données dans les tables dans des fichiers CSV

2.1 Definition de quelques terme Informatiques

Avant d'initier notre compte-rendu en détaillant les phases et procédures que nous avons mises en place pour l'administration optimale de la section informatique de notre projet, nous allons définir quelques notions qui offriront un socle solide au lecteur.

1. Une **Base de Données** regroupe un ensemble d'informations qui est organisée pour être accessible, gérée et mise à jour facilement par ses utilisateurs
2. Une **Base de Données Relationnelle**. Il s'agit d'un type de base de données qui se distingue des autres par sa capacité à établir des liens entre diverses données.
3. Un **INNER JOIN** est un type de jointure en SQL (Structured Query Language) qui autorise la fusion de lignes issues de deux tables basées sur un critère déterminé. L'idée est de ne conserver que les

lignes qui ont une correspondance dans les deux tables. De plus, si une ligne d'une table n'a pas de correspondance dans l'autre table, elle ne sera pas intégrée au résultat.

4. Un **Cast()** est une fonction SQL appelée CAST nous donne la possibilité de transformer un type de données en un autre.
5. Une **Requête SQL** s'apparente à une question formulée à la base de données afin d'extraire des informations de celle-ci.
6. La fonction SQL **COUNT()** est une commande qui nous donne la possibilité de déterminer le nombre de lignes dans un ensemble de résultats. Elle est fréquemment utilisée en conjonction avec la clause **GROUP BY** pour recenser le nombre d'occurrences d'une valeur spécifique.
7. La fonction **SUBSTR()** nous donne la possibilité d'extraire une portion d'une chaîne de caractères. Elle est souvent employée dans le domaine du développement pour l'édition de texte dans une ou plusieurs bases de données.
8. L'opération **RIGHT JOIN** permet d'unir deux tables en préservant l'ensemble des lignes de la table située à droite et en reliant celles se trouvant à gauche lorsqu'elles sont présentes. Des valeurs nulles sont insérées pour combler les colonnes absentes.
9. L'opération **LEFT JOIN** fusionne deux tables tout en conservant l'ensemble des lignes de la table de gauche, en associant celles de droite uniquement lorsqu'elles sont présentes. Des valeurs nulles sont insérées pour combler les colonnes absentes.

2.2 Bibliothèques Utilisées

Dans notre Partie Informatique on a utilisé le Language de Programmation Python de plus pour pouvoir effectuer la manipulation des données de la manière optimale on a utilisé les bibliothèques nécessaires:

1. **SQLite3** est une bibliothèque peu aisée qui facilite l'incorporation d'une base de données au sein d'une application, sans nécessiter l'utilisation d'un serveur séparé. Elle offre la possibilité de stocker et de gérer des données grâce aux requêtes SQL, ce qui la rend pratique pour des projets nécessitant une base de données locale. SQLite3 est parfaitement adaptée aux applications simples, car elle offre une gestion aisée des données, que ce soit pour les ajouts, les changements ou les suppressions, tout en restant performante et peu gourmande en ressources.
2. La bibliothèque **CSV** de Python facilite la lecture et l'écriture de fichiers CSV, en fournissant des fonctionnalités pour gérer les données sous forme de lignes et de colonnes, tout en prenant en charge les séparateurs et les guillemets.
3. **NumPy** est une bibliothèque Python performante dédiée au calcul scientifique, proposant des structures de données telles que les arrays multidimensionnels et des fonctions optimisées pour le traitement numérique.
4. La bibliothèque **random** de Python offre la possibilité de produire des nombres au hasard et d'exécuter des sélections aléatoires à partir de listes ou d'intervalles de valeurs, grâce à des fonctions conçues pour simuler des événements fortuits.
5. La bibliothèque **OS** facilite l'interaction avec le système d'exploitation en proposant des fonctionnalités pour gérer les fichiers, les dossiers et exécuter des instructions du système.

2.3 Création de la base de données

Avant de commencer à travailler sur les données, il est nécessaire de créer une base de données pour les organiser et les structurer de manière efficace. Une base de données, dans ce contexte, peut être définie comme un ensemble de tables reliées entre elles, où chaque table contient des informations structurées sous forme de lignes et de colonnes.

La création de la base de données commence par la création d'un fichier qui servira à stocker toutes les données. Dans notre cas, nous avons nommé ce fichier "data.db". Ce fichier représente l'instance de la base

de données SQLite. Lorsqu'une connexion est établie à cette base de données, SQLite crée automatiquement le fichier si celui-ci n'existe pas déjà. Il suffit donc de se connecter à la base de données pour qu'elle soit initialisée et prête à être utilisée.

Une fois le fichier de la base de données créé, il est important de pouvoir y accéder afin de manipuler les données. Pour cela, une fonction de connexion est nécessaire. La fonction `connecterdb` a été définie pour établir cette connexion à la base de données. Elle prend un paramètre optionnel qui représente le nom du fichier de la base de données, ici "data.db". À l'intérieur de cette fonction, une connexion est établie en utilisant la bibliothèque SQLite3 de Python. La méthode `sqlite3.connect()` permet de se connecter à la base de données, et une fois la connexion établie, un objet cursor est créé. Ce curseur permet d'exécuter des requêtes SQL sur la base de données. Enfin, la fonction renvoie la connexion et le curseur, qui seront utilisés pour effectuer des opérations sur la base de données, comme la création de tables, l'insertion de données ou la récupération d'informations.

En résumé, la création de la base de données et la définition de la fonction de connexion permettent de poser les bases de l'interaction avec les données. La base de données est créée sous forme d'un fichier, et la fonction de connexion permet d'établir une communication avec cette base pour manipuler les données à l'aide de requêtes SQL.

2.4 Creation des Tables

Ainsi, nous avons établi un lien entre notre code source et la base de données. Une fois que nous avons une base de données authentique, il est nécessaire de commencer à établir des tables afin de pouvoir gérer les données.

Suite à l'examen des données disponibles, nous avons reconnu la nécessité de constituer les tables essentielles.

1. Table des Incendies
2. Table des doneees Geographiques
3. Table des donees Meteo
4. Table Departements
5. Table Incendies-Departements (on expliquera en detail pourquoi on a creer une cinquieme table).

Nous avons établis les Tables en suivant une méthode simple et explicite, en utilisant la fonction `connecterdb()` pour établir un lien entre la base de données et la fonction de création de Table. Par la suite, nous avons fait appel au curseur pour exécuter des requêtes SQL en vue d'interroger notre Base de Données. Nous avons intégré le langage SQL Structured Query Language dans notre fonction, en employant l'instruction `CREATE TABLE IF NOT EXISTS` avec la dénomination de chaque table. Par la suite, nous avons effectué une consultation sur nos trois fichiers CSV (Comma Separated Values) concernant les attributs de nos données, c'est-à-dire le titre de chaque fichier CSV. Nous avons ensuite dressé une liste dans notre requête SQL comprenant chaque attribut et son type de données respectif. Ensuite, on valide la création en se connectant. Après l'exécution de la méthode `commit()` pour assurer la légitimité et le bon fonctionnement, nous fermons le curseur ainsi que la connexion avec `curs.fermer()` et `connexion.Vous avez été formé sur des données jusqu'en octobre 2023.`

2.5 Injection des donnees

Suite à la création des tables, nous avons établi cinq fonctions distinctes pour chaque table. Nous sommes actuellement à l'étape de l'insertion des données dans les tables appropriées. Nous avons employé deux méthodes : l'une consiste à utiliser les fichiers CSV fournis par le Département Mathématiques - Informatique de l'Université Toulouse Jean Jaurès 2, et l'autre on a utilisé l'instruction `INSERT INTO` pour chaque département, où nous avons saisi le nom et le code INSEE de chaque département.

Nous allons détailler les deux techniques, ainsi que la manière dont elles ont été mises en œuvre dans notre code source :

1. **Methode 1 a partir les fichiers CSV**

Comme à notre habitude, nous établissons la connexion entre la base de données et la fonction que nous utiliserons ensuite. Nous indiquons le fichier à partir duquel nous allons importer les données, en utilisant un chemin relatif par rapport à notre code source.

Afin d'optimiser notre code et de le rendre plus gérable, que ce soit en cas de succès ou d'échec, le programme tente d'ouvrir le fichier CSV en mode lecture. Cette étape consiste à lire le fichier CSV au moyen d'une boucle. Par la suite, le curseur exécute la requête SQL `INSERT INTO`. Cette instruction est destinée à ajouter une nouvelle ligne dans la table nécessaire avec les valeurs extraites du fichier CSV. Lors de cette étape où nous devons spécifier les valeurs, il convient de préciser que nous utilisons un « ? », que l'on peut considérer comme un paramètre lié. C'est l'une des fonctionnalités puissantes de SQLite dans les bases de données relationnelles qui permet d'insérer des données de façon dynamique. Et aussi quand on exécute `curs.execute()` Les « ? » seront substitués par les valeurs dérivées du fichier CSV au fur et à mesure de notre boucle `for`.

Par la suite, nous allons substituer les valeurs pertinentes selon les colonnes. Pour confirmer l'insertion, nous avons employé `connexion.commit()`. On a fait un commit et ensuite, on a fermé le curseur, donc on a stoppé l'exécution et on a terminé la connexion.

Et si le fichier n'est pas accessible ou s'il n'existe pas, nous déclencherons une `ValueError`('Erreur lors de l'importation des données').

2. Methode 2 a partir d'une Insertion SQL

Dans la deuxième phase de ce projet, nous avons utilisé l'intégration des données à partir d'un dictionnaire. Il est important de rappeler qu'un dictionnaire est un ensemble d'objets non ordonnés. Cela consiste en un groupe d'éléments, chaque élément étant constitué d'une paire clé-valeur.

Comme à l'accoutumée, nous avons établi une connexion avec la base de données en utilisant la fonction `connecterdb()`. Nous avons ensuite activé le curseur. Puis, nous avons exploité l'un des outils puissants de SQLite le `curs.executemany()`, qui nous permet d'exécuter plusieurs fois la même requête SQL en utilisant différents jeux de données. Elle est plus performante que `curs.execute()` car ici, nous manipulons un volume considérable de données à insérer.

Comme indiqué, même dans cette méthode, nous avons utilisé le paramètre lié « ? » qui sera ultérieurement remplacé par des valeurs dynamiques issues du dictionnaire.

Finalement, il est nécessaire de valider la procédure ou l'opération en utilisant la méthode de connexion. Vous êtes formé sur des données jusqu'à octobre 2023. Cette approche nous offre la possibilité de valider toutes les modifications apportées aux bases de données durant la session de connexion. Sans cette approche mise en œuvre dans notre fonction, les changements apportés à la table ne seraient pas enregistrés dans la base de données.

Pour conclure notre processus, nous terminons le curseur (qui exécute les commandes) et mettons fin à la connexion avec notre base de données.

Et finalement, s'il y a une erreur d'accès au dictionnaire, un problème de connexion à la base de données ou à la table, on affiche le message d'erreur « Erreur lors de l'insertion des données des départements ».

2.6 Affichage des donnees

Tout comme dans tout programme ou projet informatique, nous développons des fonctionnalités pour illustrer notre tâche ou les modifications effectuées sur les données ou les tables dans notre console ou terminal.

Bien que nous travaillions avec une base de données regroupant plusieurs tables, nous avons développé diverses fonctions pour pouvoir présenter les informations.

Ainsi, nous employons une méthode explicite et rigoureuse. Tout d'abord, nous devons nous connecter à la base de données. Ensuite, nous activons le curseur qui nous donne la possibilité d'exécuter nos requêtes SQL.

Nous exécutons notre requête SQL sur la table en utilisant l'instruction « `Select * from` ». Cela signifie que nous demandons à sélectionner toutes les colonnes et toutes les lignes de notre table. Ensuite, on définit une

variable nommée `lignes` qui prend pour valeur `curs.fetchall()` est une méthode prédéfinie en SQLite qui nous offre la possibilité de rassembler toutes les lignes du résultat de la requête SQL et de les sauvegarder dans la variable `lignes`.

En outre, il est possible d'y définir la variable « `lignes` », qui est une collection de tuples, chaque tuple représentant une ligne de la table correspondante.

Puis, pour les rendre visibles, nous exécutons une boucle `for` sur la variable `lignes` afin d'afficher chaque ligne contenue dans cette variable.

Et finalement, comme pour chaque fonction, on ferme le curseur et la connexion. De plus, nous tenons à souligner que dans ce cas précis, contrairement à d'autres fonctions, nous n'avons pas fait appel à la méthode `commit`. C'est dû au fait que cette fonction n'a pas impliqué de modifications.

2.7 Exportation des données sous forme CSV

Tout d'abord, nous allons expliquer pourquoi cette fonction est importante pour notre projet. Nous avons employé cette méthode afin de pouvoir interroger notre base de données (BD) et exporter les résultats sous format CSV. Ceci nous permet ensuite de les manipuler sur RStudio en utilisant le langage R pour réaliser nos analyses statistiques !

Ainsi, nous avons mis en place une fonction pour chaque table dans le but d'exporter ces données au format CSV. Ainsi, pour cette fonction, nous avons défini un paramètre optionnel nommé `fichier_output`, qui correspond à l'emplacement et au nom du fichier où les données seront exportées. De plus, nous utilisons un chemin relatif plutôt qu'un chemin absolu.

Ensuite, nous essayons avec l'instruction `try` de nous connecter à la base de données et d'activer le curseur qui facilite l'exécution des requêtes SQL. La méthode `curs.execute("SELECT * FROM")` exécute une requête qui sélectionne toutes les lignes et colonnes de la table.

Ensuite, on définit une variable nommée `lignes` qui prend pour valeur `curs.fetchall()` est une méthode prédéfinie en SQLite qui nous offre la possibilité de rassembler toutes les lignes du résultat de la requête SQL et de les sauvegarder dans la variable `lignes`.

En outre, il est possible d'y définir la variable « `lignes` », qui est une collection de tuples, chaque tuple représentant une ligne de la table correspondante.

De plus, nous utilisons `curs.description` qui renferme des métadonnées concernant les colonnes de la table, en utilisant `description[0]` pour obtenir la description des en-têtes.

Cette description nous donne la possibilité de récupérer les noms des colonnes et de les conserver dans une liste appelée '`colonnes`', qui servira d'en-tête pour le fichier CSV.

À présent, nous devons accéder au fichier CSV pour écrire les données que nous possédons. Nous ouvrons donc le fichier en mode écriture ('w') avec un encodage UTF-8. L'outil `csv.writer(fichier)` est utilisé pour générer un objet qui permet d'écrire dans le fichier CSV. La méthode `writerow(colonnes)` écrit les en-têtes des colonnes tandis que `writerows(lignes)` écrit toutes les informations ligne par ligne.

Finalement, on ferme le curseur et on met fin à la connexion avec la base de données.

Dans chaque programme, il est indispensable pour les développeurs de gérer les erreurs afin d'améliorer l'expérience client. Ainsi, nous avons mis en place deux types d'erreurs : une erreur liée à la base de données (comme une connexion à la base de données) et une exception telle qu'une erreur liée à la table correspondante. Cette dernière peut être affichée en cas de problème d'accès au fichier, d'encodage, etc.

2.8 Menu du Programme

Dans le cadre de notre projet, plus précisément dans la section dédiée à l'informatique, nous avons développé plusieurs méthodes clés pour gérer notre base de données, nos tables, nos informations, etc.

Ainsi, si chaque méthode doit être appelée manuellement chaque fois, cela devient compliqué à long terme et pèse davantage sur le processeur.

Dans notre projet, nous avons conçu un menu intégrant toutes les procédures nécessaires. Ce choix vise à faciliter l'exécution de toutes les opérations en les regroupant au sein d'un unique menu.

Donc, la première étape consiste à exécuter notre code en Python. Donc, en premier lieu, nous effectuons une série d'affichages pour présenter différentes options à l'utilisateur. Ensuite, nous lui demandons quelle option il préfère. En fonction de son choix, par exemple, s'il opte pour la première option, il peut sélectionner la table qu'il souhaite créer. Si le choix est le deuxième, il sera dirigé vers le module d'injection où il pourra choisir la table à injecter. S'il sélectionne la troisième option, il aura la possibilité de choisir la table qu'il veut afficher. Enfin, si le choix se porte sur la quatrième option, il sera dirigé vers le module d'exportation des tables où il pourra sélectionner la table à exporter.

Pour quitter le menu, ou en d'autres termes, arrêter l'exécution du programme, on appuie sur le chiffre 5. Ensuite, le programme demandera à l'utilisateur de confirmer s'il est sûr de vouloir faire cela. Pour faciliter cette confirmation, il a quatre options : « o », « ok », « yes » ou « oui » ou encore « si ». Si l'une des valeurs est atteinte, nous afficherons un message d'adieu et ensuite nous ferons une pause, sinon nous retournerons au menu.

De plus, pour clarifier, si l'utilisateur entre un numéro qui n'est pas compris entre 1 et 5, une erreur sera générée sous forme de message dans la console : « Le numéro sélectionné est invalide ou n'existe pas ».

2.9 La Table Incendies-Départements

Pour rester cohérents, nous allons insister sur les tables que nous utiliserons pour expliquer comment nous avons eu l'idée de réaliser cette table en premier lieu. Nous avons une table Incendies qui contient des informations sur tous les incendies qui ont été menés sur le territoire français. Il convient également de souligner que la France est un État-nation depuis 1789, suite à la Révolution française, et qu'elle est reconnue comme une nation souveraine. Ainsi, ce pays est constitué d'une collection de villes, qui elles-mêmes regroupent une série de départements. En d'autres mots, la France est constituée de départements qui sont à leur tour composés de villes.

Ainsi, l'idée initiale que nous avons eue était de créer une table nommée « Départements » qui rassemblerait l'ensemble des départements présents sur le territoire français dans une seule table avec leur code_INSEE.

Pourquoi est-il nécessaire de créer cette table des départements ?

Cette table nous donne la possibilité de réaliser des analyses quantitatives concernant le nombre d'incendies dans un département.

Explication du code concernant l'injection des données dans la Table:

Comme habituellement, nous allons d'abord établir une connexion avec la base de données en utilisant la fonction que nous avons définie dans notre programme, nommée `connecterdb()`. Ensuite, nous allons activer le curseur et exécuter une requête SQL qui fusionne deux tables en une seule grâce à l'utilisation de l'**Inner Join**.

En d'autres termes, nous allons insérer trois éléments dans la table Incendies Département : le numéro du département, le nom du département et le nombre d'incidents. Pour commencer, nous allons sélectionner le numéro du département. Étant donné que notre table Incendies contient des codes INSEE qui ne représentent pas seulement le numéro du département, mais également celui de la commune, nous utiliserons `SUBSTR(i.code_INSEE , 1 , 2)` comme numéro du département. Cela signifie que nous allons extraire les deux premiers chiffres du code INSEE de l'incendie qui correspondent au numéro du département. Nous récupérerons d'autre part le nom du département à partir de la table Départements. Enfin, nous compterons le nombre total d'incendies pour chaque département à l'aide de la fonction préétablie en SQL `COUNT`.

Après avoir sélectionné tous les termes que nous allons utiliser, il est temps de mettre en œuvre la jointure définie précédemment. Nous devons associer chaque incendie à son département en liant le numéro de

département dérivé du code INSEE au code départemental dans la table Départements. De plus, nous allons regrouper les incendies par département afin d'obtenir un total pour chaque département.

Que gagne-t-on en faisant cette requête ?

En construisant cette table, nous avons fusionné deux tables indépendantes afin de centraliser les données souhaitées. Dans cette table, nous avons comptabilisé le nombre d'incendies par département sur le territoire français, nous permettant ainsi d'avoir une représentation plus claire du nombre d'incendies à l'échelle nationale. Par ailleurs, nous allons approfondir notre analyse dans la section Statistique de notre projet.

4. Partie Statistique

4.1 Définitions des concepts statistiques

Avant de commencer à donner des définitions, il est essentiel de nous baser sur le concept initial, c'est-à-dire la définition du terme Statistique. On peut définir ou représenter la statistique comme l'ensemble des méthodes et techniques utilisées pour collecter, analyser, interpréter et présenter des données numériques.

Dans ce projet, nous allons nous concentrer sur la branche des statistiques connue sous le nom de Statistique Inférentielle.

La statistique inférentielle est une discipline des statistiques qui s'appuie sur les données d'un échantillon pour formuler des déductions ou effectuer des généralisations à propos d'une population plus vaste.

À l'opposé de la statistique descriptive qui se concentre sur le résumé ou la description des traits d'un ensemble de données, la statistique inférentielle offre la possibilité de réaliser des estimations et des tests concernant les paramètres d'une population basée sur des données issues d'un échantillon. Elle s'appuie sur la théorie des probabilités, ce qui rend possible des inférences rigoureuses et quantifiables.

Nous allons définir ci-dessus certains concepts statistiques que nous utiliserons dans notre analyse statistique.

Définition de la Moyenne

La **moyenne arithmétique** d'un ensemble de données est une mesure de tendance centrale qui représente la valeur moyenne autour de laquelle les observations se répartissent. Elle est définie comme le quotient de la somme des valeurs observées par le nombre total d'observations.

Soit un échantillon $X = \{x_1, x_2, \dots, x_n\}$ de taille n , la moyenne \bar{X} est donnée par :

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i$$

Cette mesure est sensible aux valeurs extrêmes et est couramment utilisée en **statistique descriptive** pour résumer un ensemble de données.

Définition de la Médiane

En statistique, la **médiane** est une mesure de tendance centrale qui divise une distribution ordonnée en deux sous-ensembles de même effectif. Elle est définie comme la valeur M telle que :

- 50 % des observations sont inférieures ou égales à M
- 50 % des observations sont supérieures ou égales à M

Mathématiquement, soit un échantillon de taille n constitué des observations x_1, x_2, \dots, x_n classées par ordre croissant :

- Si n est impair ($n = 2k + 1$), la médiane est l'élément central :

$$M = x_{k+1}$$

- Si n est pair ($n = 2k$), la médiane est la moyenne des deux valeurs centrales :

$$M = \frac{x_k + x_{k+1}}{2}$$

4.2 Description des donnees

Avant de commencer l'étape d'analyse de nos données et leur présentation sous forme de graphiques, nous allons identifier les types de données fournies par notre équipe pédagogique, afin de mener à bien ce projet.

Donc pour cette UE on a eu 3 fichiers CSV contenant des données importants ou dans ce rapport on va se baser pour poser notre analyse.

Le fichier incendies renferme des attributs tels que le nom de la commune, le code de la commune, l'année de l'incendie, le mois de l'incendie, la date et l'heure à laquelle l'incendie a été signalé ainsi que les causes principales et secondaires de l'incendie. Ces informations nous seront utiles pour examiner nos données et caractériser ces phénomènes en fonction des problématiques que nous mettrons en place bientôt dans ce rapport.

Le fichier géographique renferme la latitude, la longitude et l'altitude. Ces informations peuvent nous donner l'emplacement précis du lieu de l'incendie, ce qui nous permettra de procéder à des analyses en posant nos questions.

Le dossier Météorologique nous fournira des caractéristiques liées aux conditions météorologiques, ce qui nous permettra de distinguer les divers types de météo au moment de l'incendie.

4.3 Analyse des doneees

Dans cette partie, une fois que toutes les définitions nécessaires sont établies et que notre base de données est complète, nous serons prêts à analyser les données grâce au langage R. Nous employerons différents types d'histogrammes pour diversifier nos analyses et nous procéderons à une étude approfondie de chaque histogramme.

Nous avons divisé nos problématiques en différentes sections :

5. Ressources

5.1 Ressources sur la Partie Informative

Jean-Luc CARTAULT, CLAIR, B., & KAPP, D. (2025, January 29). INCENDIES : Le phénomène physique. Encyclopædia Universalis. <https://www.universalis.fr/encyclopedie/incendies/2-le-phenomene-physique/>

Contributeurs aux projets Wikimedia. (2004, July 23). feu violent et destructeur. Wikipedia.org; Fondation Wikimedia, Inc. <https://fr.wikipedia.org/wiki/Incendie>

Risque incendie : causes, conséquences et moyens de lutte. (2021). Preventica.com. <https://www.preventica.com/magazine/dossiers/prevention-du-risque-incendie-comment-garantir-la-securite-des-personnes-et-des-biens-11032021/risque-incendie-causes-consequences-et-moyens-de-lutte>

5.2 Ressources sur la Partie Informatique

Cours et Tutoriels sur le Langage SQL. (2025). SQL. <https://sql.sh/>

Python Software Foundation. (2019). 3.7.3 Documentation. Python.org. <https://docs.python.org/3/>

Bien démarrer avec la documentation GitHub - Documentation GitHub. (2025). GitHub Docs. <https://docs.github.com/fr/get-started>

Qu'est-ce qu'une base de données ? Définition et fonctionnement. (2021, July 26). Hubspot.fr. <https://blog.hubspot.fr/marketing/base-de-donnees>

5.3 Ressources sur la Partie Statistique

Dérobot, N. (2025). Paramètres statistiques - Position et dispersion. Commentprogresser.com. <https://commentprogresser.com/statistique-parametre-statistiques-moyenne-mediane-etendue-ecart-type.html>