

# *RAPPORT*

## **IVG Scope**

AZZOUZI Hiba  
BENBETKA Rania  
BENGUENNA Célia  
LAVOGEZ Ethan  
NABE-NANA Hilarion

**Master 1 Data Science en Santé  
2025-2026**

## ABSTRACT

Despite the legalization of abortion in France and the availability of open institutional data, questions remain regarding territorial disparities, demographic patterns, healthcare accessibility, and the evolution of abortion methods. This project develops an interactive dashboard to centralize and visualize abortion-related data, enabling demographic and territorial analyses.

Data are collected through an automated web scraping process using the Opendatasoft API, extracting datasets from the French Directorate for Research, Studies, Evaluation and Statistics (DREES) and complementary sources available on [data.gouv.fr](https://data.gouv.fr). The methodology ensures reproducibility and structured integration of cleaned and harmonized data.

The dashboard highlights long-term national trends, the rise of medical abortion, and persistent territorial disparities. By combining indicators of abortion rates and healthcare supply, it contributes to understanding potential gaps between legal rights and effective accessibility.

**Keywords:** abortion, France, public health, reproductive health, territorial disparities, health data science, web scraping, data visualization, open data.

## RÉSUMÉ

Malgré la légalisation de l'Interruption Volontaire de Grossesse en France et la disponibilité de données institutionnelles en open data, des interrogations persistent concernant les disparités territoriales, les évolutions par âge et l'accessibilité aux soins. Ce projet propose la conception d'un dashboard interactif destiné à centraliser et visualiser les données relatives au recours à l'IVG afin de faciliter l'analyse démographique et territoriale du phénomène.

Les données sont collectées via un processus automatisé de web scraping fondé sur l'API publique Opendatasoft, à partir des jeux diffusés par la Direction de la recherche, des études, de l'évaluation et des statistiques (DREES) et de sources complémentaires issues du portail [data.gouv.fr](https://data.gouv.fr). La méthodologie assure la reproductibilité et la maîtrise de la chaîne de traitement, de l'extraction à l'intégration des données dans l'outil de visualisation.

Le dashboard met en évidence les tendances nationales depuis 1990, la progression de l'IVG médicamenteuse et les disparités départementales persistantes. En croisant indicateurs de recours et éléments d'offre de soins, il contribue à éclairer l'écart possible entre droit légal et accessibilité réelle.

**Mots-clés :** IVG, France, santé publique, santé sexuelle et reproductive, disparités territoriales, data science en santé, web scraping, visualisation de données, open data.

## **Sommaire**

1. Introduction
2. Contexte et état des lieux
3. Problématique
4. Objectifs du projet
5. Sources des données
6. Méthodologie
7. Description du dashboard
8. Résultats attendus
9. Utilité du projet
10. Limites
11. Perspectives
12. Conclusion
13. Bibliographie

## **1. Introduction**

L'interruption volontaire de grossesse (IVG) constitue un enjeu majeur de santé publique en France, au croisement des dimensions médicales, sociales et politiques. Légale depuis 1975, elle fait l'objet d'un suivi statistique régulier visant à évaluer les besoins et l'efficacité des dispositifs de prise en charge.

Cependant, bien que les données officielles soient nombreuses, elles demeurent souvent dispersées et peu accessibles sans compétences techniques spécifiques. Leur exploitation peut ainsi s'avérer complexe pour un public non spécialiste.

Dans ce contexte, le développement d'un dashboard interactif dédié aux IVG permet de centraliser les données ouvertes, d'en faciliter l'exploration visuelle et d'améliorer la compréhension des dynamiques nationales et territoriales. Ce projet présente la conception de cet outil, les sources mobilisées, la méthodologie adoptée ainsi que les principaux résultats et limites.

## **2. Contexte et état des lieux**

Chaque année, environ 200 000 IVG sont pratiquées en France. Ce chiffre relativement stable masque toutefois d'importantes variations selon l'âge, la situation socio-économique, le territoire ou l'évolution des politiques de santé reproductive.

L'accès à l'IVG constitue un indicateur clé des droits sexuels et reproductifs, mais aussi de l'organisation du système de soins. Des disparités territoriales persistent, notamment en raison de la fermeture de certaines structures pratiquant l'IVG, du manque de professionnels formés ou des contraintes d'accès aux services hospitaliers.

Par ailleurs, l'évolution des méthodes utilisées (médicamenteuse ou chirurgicale), l'âge moyen des patientes ou encore les délais de prise en charge font l'objet d'un suivi régulier par les autorités sanitaires. Ces informations sont essentielles pour orienter les politiques publiques et garantir un accès équitable aux soins.

Cependant, les données disponibles sont majoritairement présentées sous forme de rapports ou de tableaux statiques, rendant difficile une exploration interactive ou comparative. Les outils permettant de visualiser ces informations de manière intuitive restent encore peu nombreux.

Ainsi, la création d'un tableau de bord dédié vise à combler ce manque en proposant une synthèse visuelle et dynamique des données relatives aux IVG.

### 3. Problématique

Malgré la légalisation de l'IVG en France et la disponibilité de données ouvertes, des questions importantes subsistent concernant la répartition territoriale, l'évolution par âge, l'accès aux soins et l'utilisation des différentes méthodes d'IVG. Les disparités régionales et locales, les zones sous-dotées en professionnels de santé et les variations méthodologiques ne sont pas facilement visualisables à partir des rapports existants. La problématique de ce projet consiste donc à déterminer comment centraliser et visualiser ces données dans un tableau de bord interactif, afin de faciliter la compréhension des tendances démographiques, territoriales et méthodologiques, tout en mettant en lumière les inégalités d'accès aux soins et en soutenant la sensibilisation et l'analyse en santé publique.

### 4. Objectifs du projet

L'objectif principal du projet est de concevoir une interface homme-machine (IHM) dédiée à l'exploration des données sur l'IVG en France. Cette interface s'appuie sur des méthodes de web scraping afin de collecter, centraliser et exploiter des données issues de sources institutionnelles fiables.

L'outil permet d'analyser les données selon différents critères, de mettre en évidence les tendances temporelles ainsi que les disparités territoriales, et de proposer des visualisations claires facilitant la compréhension du phénomène. Il constitue à la fois un support pédagogique accessible à un large public et une application concrète de compétences en data science, en collecte automatisée de données et en visualisation interactive.

### 5. Sources et nature des données

Les données mobilisées dans le cadre de ce projet proviennent exclusivement de **sources officielles françaises en open data**, assurant ainsi leur fiabilité, leur traçabilité et leur mise à jour régulière. La principale source d'information relative à l'IVG en France est la Direction de la recherche, des études, de l'évaluation et des statistiques (DREES), organisme de référence en matière de production et de diffusion de statistiques sanitaires et sociales.

La DREES publie des données détaillées portant notamment sur le nombre annuel d'IVG, la répartition des interventions selon l'âge des patientes, les méthodes employées (médicamenteuses ou instrumentales), ainsi que la distribution territoriale des IVG aux niveaux régional et départemental. Ces informations sont accessibles via

les publications « *Études et Résultats* », la base de données Data DREES et la plateforme nationale open data [data.gouv.fr](https://data.gouv.fr). Des données complémentaires ont également été mobilisées à partir du portail national open data [data.gouv.fr](https://data.gouv.fr), notamment via le jeu de données dédié à l'IVG, afin d'enrichir et de consolider les analyses réalisées.

L'ensemble des données exploitées est agrégé et anonymisé, excluant toute information individuelle identifiable. Cette approche garantit le respect strict des principes de confidentialité et de protection des données personnelles, tout en permettant une analyse statistique approfondie, rigoureuse et pertinente du phénomène étudié.

## **6. Méthodologie**

Dans le cadre de la réalisation de ce dashboard, nous avons choisi de collecter les données directement depuis les widgets interactifs mis à disposition par la Direction de la recherche, des études, de l'évaluation et des statistiques (DREES), accessible via la plateforme [data.drees.solidarites-sante.gouv.fr](https://data.drees.solidarites-sante.gouv.fr). L'objectif était de mettre en place un processus automatisé de web scraping s'appuyant sur l'API publique Opendatasoft afin d'extraire les données de manière dynamique et reproductible. Cette démarche visait à garantir une maîtrise complète de la chaîne de traitement, depuis l'extraction des données brutes jusqu'à leur intégration dans le dashboard.

Le script développé interroge directement l'API associée à chaque widget. À partir de l'URL du widget, l'identifiant du jeu de données est extrait automatiquement, ce qui permet ensuite d'effectuer des requêtes vers l'end-point de l'API pour récupérer les métadonnées et les enregistrements. Une première requête permet d'obtenir la structure du dataset, notamment la liste des variables, leur ordre et leurs libellés officiels. Cette étape assure la cohérence entre les données extraites et la structuration institutionnelle définie par la DREES. Le script utilisé pour cette extraction est disponible dans le fichier .

Les données sont ensuite téléchargées de manière exhaustive grâce à un mécanisme de pagination. L'API limitant le nombre d'observations retournées par requête, les enregistrements sont récupérés par lots successifs jusqu'à atteindre le nombre total d'observations disponibles. Cette méthode garantit l'extraction complète du jeu de données, indépendamment de son volume. Les résultats sont ensuite convertis en Data Frame via la bibliothèque pandas afin de faciliter les traitements ultérieurs.

Un travail de data management a été mené après l'extraction afin de préparer les données à leur intégration dans l'outil de visualisation. Une attention particulière a été portée aux variables géographiques, dont la structure peut être complexe dans l'API,

notamment sous forme d'objets JSON ou de dictionnaires contenant des coordonnées. Un traitement spécifique a permis d'harmoniser ces formats en extrayant uniquement les informations pertinentes et en les convertissant dans une forme exploitable et cohérente. Les valeurs manquantes ont été normalisées afin d'éviter les incohérences lors de l'importation dans le dashboard. Les colonnes ont ensuite été renommées selon les libellés officiels fournis dans les métadonnées et réorganisées conformément à l'ordre institutionnel des variables. Un tri a également été appliqué selon les dimensions structurantes, notamment l'année et la zone géographique, afin de garantir une lisibilité optimale des fichiers produits.

Les jeux de données ont finalement été exportés dans un format standardisé, compatible avec les outils d'analyse et de visualisation utilisés pour le dashboard. L'encodage UTF-8 avec BOM et l'utilisation d'un séparateur adapté assurent une bonne interopérabilité, notamment avec les environnements bureautiques français. L'ensemble du processus repose sur un script unique, permettant une automatisation complète et une mise à jour facilitée des données en cas d'évolution des jeux de données sources.

Cette méthodologie garantit ainsi la traçabilité, la reproductibilité et la robustesse du projet. Elle permet d'assurer la fiabilité des données mobilisées dans le dashboard tout en conservant une capacité d'actualisation rapide et maîtrisée.

## 7. Description du dashboard

Le dashboard « **IVG en France — Droit légal, accessibilité réelle** » propose une analyse territoriale et démographique du recours à l'IVG sur la période 2016–2024, avec des séries historiques remontant à 1990 pour certains indicateurs. Il vise à décrire les dynamiques nationales et à mettre en évidence les disparités territoriales afin d'éclairer la question de l'accessibilité réelle aux soins.

L'interface repose sur un bandeau de filtres commun permettant de sélectionner l'année et le périmètre géographique (national, métropole, DROM-COM ou département), offrant une exploration multi-échelles. Le dashboard est organisé en quatre onglets thématiques : « Le constat », « La fracture », « Les patientes » et « L'offre de soins », correspondant à des approches globale, territoriale, sociodémographique et organisationnelle du phénomène

## Onglet « Le constat »



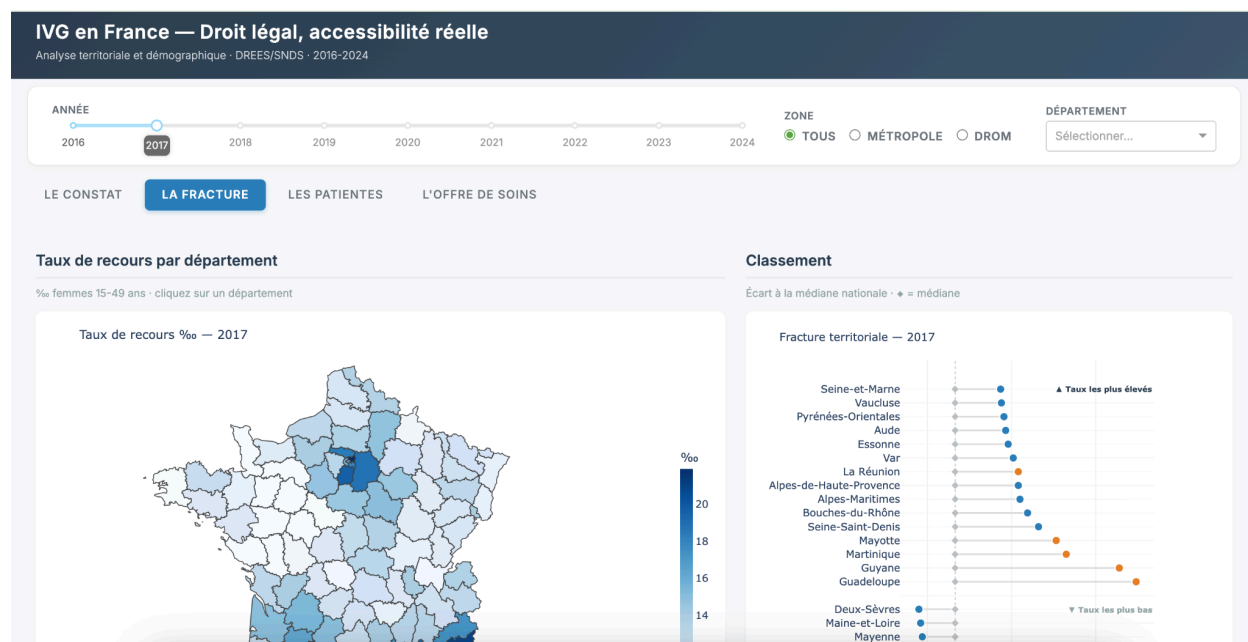
Cet onglet propose une vision synthétique et historique du recours à l'IVG en France. Il s'ouvre sur des indicateurs clés présentés sous forme de tuiles : nombre total d'IVG, taux pour 1 000 femmes de 15 à 49 ans, indicateur conjoncturel d'avortement (ICA) et part des IVG médicamenteuses, permettant d'appréhender à la fois le volume, l'intensité du recours et l'évolution des pratiques.

Un premier graphique retrace l'évolution du nombre d'IVG depuis 1990, mettant en évidence des fluctuations modérées puis une hausse après 2020, tout en contextualisant les données (période COVID, distinction valeurs brutes/hors reprises). Un second graphique présente l'évolution des modalités de prise en charge depuis 2016 et souligne la progression des IVG médicamenteuses, révélatrice d'une transformation structurelle des pratiques.

Enfin, un graphique combinant le taux pour 1 000 femmes et l'ICA permet de comparer mesure relative et indicateur synthétique du recours au cours de la vie reproductive, avec un signalement explicite de la rupture méthodologique de 2020 invitant à la prudence dans les comparaisons temporelles.



## Onglet « La fracture »

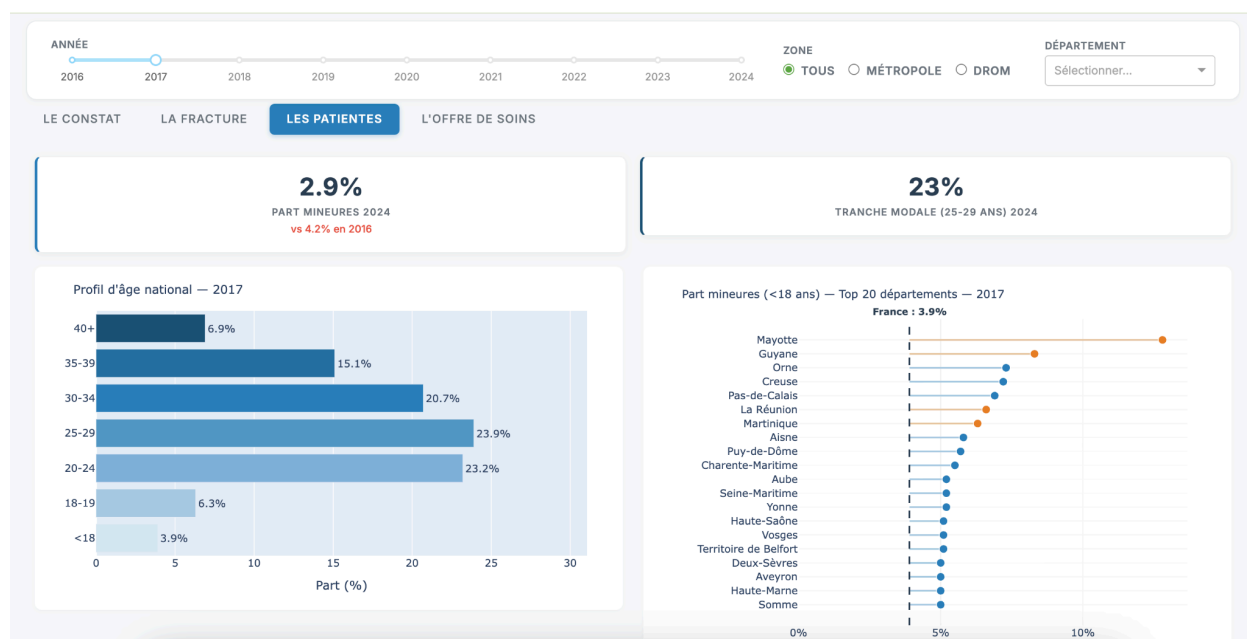


L'onglet « La fracture » met en lumière les inégalités territoriales de recours à l'IVG. Une carte choroplèthe des taux départementaux permet de visualiser immédiatement les disparités géographiques, notamment entre certains départements métropolitains et ultramarins.

Un classement des 15 départements aux taux les plus élevés et les plus faibles, accompagné d'une médiane nationale de référence, complète cette lecture. La combinaison de la carte et du classement renforce l'analyse en associant visualisation intuitive et hiérarchisation explicite.

L'interactivité, qui permet d'explorer le profil d'un département, inscrit l'onglet dans une démarche exploratoire et souligne que le taux de recours reflète non seulement l'accessibilité, mais aussi des déterminants démographiques et sociaux.

## Onglet « Les patientes »

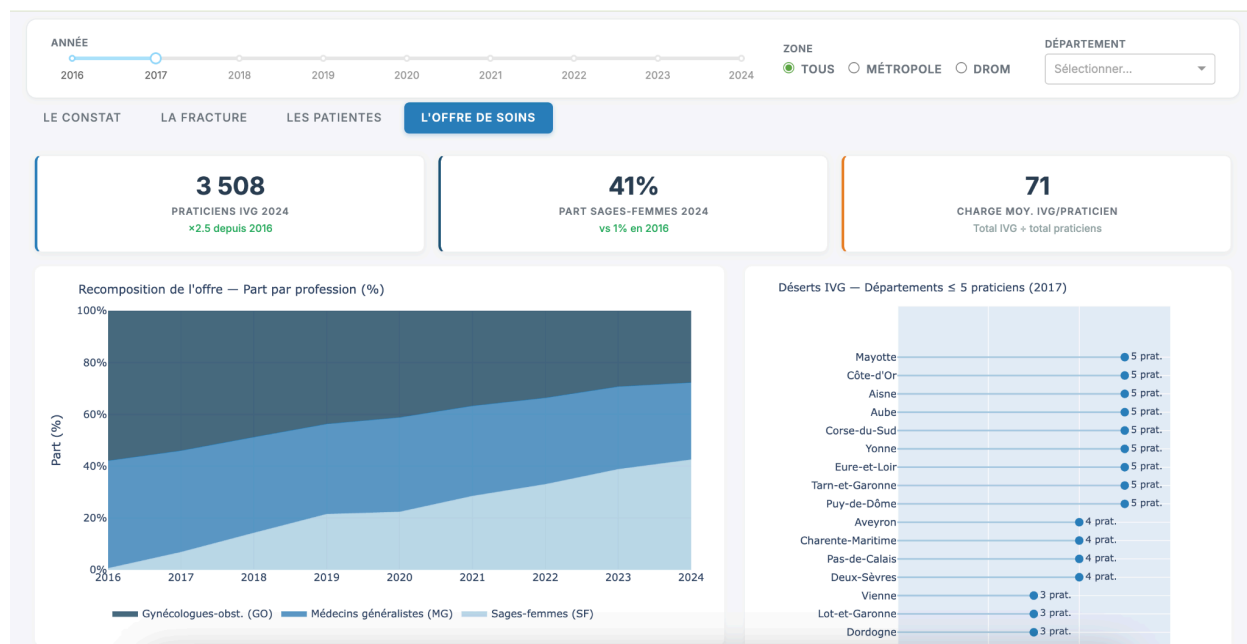


Cet onglet propose une analyse sociodémographique centrée sur la structure par âge des femmes ayant recours à l'IVG. Des indicateurs synthétiques mettent en avant la part des mineures et la tranche d'âge modale, confirmant la prédominance des 20–29 ans.

Le profil d'âge national, présenté sous forme de barres horizontales, facilite la comparaison entre classes d'âge et montre que le recours reste majoritairement concentré chez les 20–29 ans, tandis que la part des mineures demeure minoritaire et en diminution.

Un classement départemental selon la part de mineures, accompagné d'une référence nationale, permet d'identifier les territoires où cette proportion est plus élevée. Enfin, l'évolution depuis 2016 met en évidence une tendance décroissante, complétant l'analyse territoriale par une lecture temporelle des dynamiques démographiques.

## Onglet « L'offre de soins »



Ce dernier onglet analyse la dimension organisationnelle du recours à l'IVG à travers l'offre de professionnels impliqués. Des indicateurs clés présentent le nombre total de praticiens, la part des sages-femmes et la charge moyenne d'IVG par praticien, afin d'évaluer l'ampleur et la répartition de l'offre.

Un graphique met en évidence la recomposition progressive des rôles depuis 2016, marquée par la montée en puissance des sages-femmes et une évolution de l'équilibre entre professions.

Enfin, un classement des départements comptant cinq praticiens libéraux ou moins identifie des situations potentielles de fragilité territoriale, qualifiées de « déserts IVG », et permet de mettre en regard l'analyse de l'offre avec celle du recours, éclairant l'écart possible entre droit formel et accessibilité réelle.

## 8. Résultats attendus

Les résultats attendus de ce projet sont d'abord d'ordre analytique. Le dashboard doit permettre d'objectiver les dynamiques récentes du recours à l'IVG en France, en distinguant les évolutions nationales des disparités territoriales. Il s'agit notamment de mettre en évidence les tendances temporelles depuis 1990, la transformation

progressive des modalités de prise en charge (essor de l'IVG médicamenteuse), ainsi que les différences marquées entre départements.

Un second résultat attendu concerne la compréhension des profils des patientes. L'analyse de la structure par âge et de son évolution doit permettre d'identifier les groupes les plus concernés et d'observer les transformations démographiques du recours, en particulier la diminution relative de la part des mineures. L'articulation entre taux de recours et indicateur conjoncturel d'avortement doit également fournir une lecture plus complète de l'intensité du phénomène.

Enfin, le projet vise à produire une vision intégrée croisant demande et offre de soins. L'identification de territoires à faible densité de praticiens, combinée aux niveaux de recours observés, doit permettre d'explorer l'hypothèse d'un décalage possible entre droit légal et accessibilité réelle. Le résultat attendu n'est pas seulement descriptif, mais analytique : fournir un outil capable de soutenir une réflexion sur les inégalités territoriales en santé reproductive.

## **9. Utilité du projet**

Ce projet présente une utilité à la fois scientifique, opérationnelle et pédagogique. Sur le plan scientifique, il propose une synthèse visuelle et interactive de données publiques parfois dispersées, facilitant leur appropriation et leur interprétation. En structurant l'information autour de dimensions complémentaires — volume, taux, profils sociodémographiques et offre de soins — le dashboard favorise une lecture systémique du recours à l'IVG.

Sur le plan opérationnel, l'outil peut constituer un support d'aide à la décision pour les acteurs institutionnels ou territoriaux. L'identification rapide de disparités départementales ou de zones à faible densité de praticiens peut éclairer des politiques de répartition de l'offre ou de renforcement de l'accès aux soins. Bien que le dashboard ne permette pas d'établir des relations causales, il contribue à orienter l'attention vers des situations potentiellement problématiques.

Enfin, du point de vue pédagogique, le projet illustre la valeur ajoutée des approches de data science en santé publique. Il démontre comment des données administratives peuvent être transformées, structurées et mises en récit à travers des outils interactifs afin de soutenir l'analyse et la compréhension de phénomènes complexes

## **10. Limites**

Plusieurs limites doivent être soulignées. Les analyses reposent sur des données agrégées, ce qui expose à un risque de biais écologique : les disparités observées à l'échelle départementale ne permettent pas d'inférer des comportements individuels ni d'identifier précisément les déterminants du recours.

Une rupture méthodologique en 2020 impose également une interprétation prudente des comparaisons temporelles. De plus, les indicateurs disponibles ne couvrent pas pleinement l'accessibilité réelle (délais, mobilité interdépartementale, barrières sociales et culturelles).

L'analyse de l'offre de soins s'appuie principalement sur le nombre de praticiens déclarés, sans information détaillée sur leur activité effective, ce qui rend la notion de « désert IVG » indicative plutôt que exhaustive. Enfin, le dashboard adopte une approche essentiellement descriptive, sans modélisation statistique formelle des hypothèses explicatives.

## **11. Perspectives**

Plusieurs pistes d'amélioration peuvent être envisagées. L'analyse pourrait être enrichie par des méthodes quantitatives complémentaires, comme des modèles de régression ou des analyses spatiales, ainsi que par l'intégration d'indicateurs socio-économiques afin de mieux comprendre les variations territoriales du recours à l'IVG.

L'évaluation de l'accessibilité réelle pourrait également être approfondie grâce à des données sur les délais de prise en charge, les flux interrégionaux ou la distance aux structures pratiquant l'IVG, avec une approche multi-échelle plus fine.

Enfin, le dashboard pourrait évoluer vers un outil plus dynamique intégrant des comparaisons personnalisées, des analyses de tendance automatisées, des visualisations prédictives et une mise à jour entièrement automatisée via le pipeline de scraping.

## **12. Conclusion**

Ce projet s'inscrit à l'interface entre santé publique, analyse territoriale et data science. À travers la construction d'un dashboard interactif, il propose une lecture structurée du recours à l'IVG en France, articulant évolution temporelle, disparités géographiques, profils des patientes et organisation de l'offre de soins.

L'analyse met en évidence des transformations importantes des pratiques, notamment la progression de l'IVG médicamenteuse et la montée en puissance des sages-femmes, ainsi que des écarts territoriaux persistants. Elle souligne également que le droit légal à l'IVG ne garantit pas nécessairement une accessibilité homogène sur l'ensemble du territoire, ce qui justifie une lecture conjointe des indicateurs de recours et des éléments relatifs à l'offre de soins.

Ce travail a été réalisé dans le cadre du Master 1 Data Science en Santé, avec pour objectif de mettre en application concrète des compétences techniques acquises au cours de la formation. Le projet a notamment permis de mobiliser des méthodes de web scraping via API pour l'extraction automatisée de données publiques, ainsi que des principes d'interface homme-machine pour la conception d'un outil interactif structuré, lisible et orienté utilisateur. Il constitue ainsi une illustration appliquée de l'apport de la data science à l'analyse des enjeux de santé publique, en combinant acquisition de données, traitement, structuration et visualisation interactive au service d'une problématique sanitaire contemporaine.

### 13. Bibliographie

Institut national d'études démographiques. *Avortements*.

<https://www.ined.fr/fr/tout-savoir-population/chiffres/france/avortements-contraception/avortements/>

Gouvernement français. *IVG – Interruption volontaire de grossesse*. <https://ivg.gouv.fr/>

La plateforme des données publiques françaises - Jeux de données - IVG.

<https://www.data.gouv.fr/datasets/ivg>

Direction de la recherche, des études, de l'évaluation et des statistiques. *Données relatives à l'IVG*.

<https://data.drees.solidarites-sante.gouv.fr/explore/?q=IVG&sort=modified>