

FILIÈRE : CYCLE D'INGÉNIEUR GEOINFORMATION

Analyse Spatiale des Infrastructures Cellulaires

Détection de la Fracture Numérique par Data Science

RAPPORT DE MINI-PROJET

Réalisé par :
Mohammed BELFELAH

Encadré par :
Pr.Ouafae EL BOUHADI

Table des matières

1	Introduction Générale	3
1.1	Contexte du projet	3
1.2	Présentation des Données (Source OpenCellID)	3
1.2.1	Origine des données	3
1.2.2	Volume et Filtrage	3
1.3	Problématiques et Objectifs	4
1.3.1	Problématique 1 : Répartition Spatiale des Antennes	4
1.3.2	Problématique 2 : Influence de la Technologie	4
1.3.3	Problématique 3 : Logique de Couverture et Portée	4
1.4	Zone d'Étude : Tanger-Tétouan	4
2	Méthodologie et Pipeline de Traitement	5
2.1	Architecture Technique	5
2.2	Pré-traitement et Nettoyage des Données	5
2.2.1	Filtrage Spatial (Bounding Box)	5
2.2.2	Contrôle Qualité (Data Cleaning)	6
2.3	Modélisation Spatiale (Algorithme DBSCAN)	6
2.3.1	Choix de l'algorithme	6
2.3.2	Implémentation en Python	6
3	Analyse Spatiale et Répartition (Problématique 1)	7
3.1	Distribution Géographique Globale	7
3.2	Résultats du Clustering (Algorithme DBSCAN)	8
3.2.1	Identification des Cœurs Urbains (Clusters)	8
3.2.2	Analyse du "Bruit" (Antennes Isolées)	8
3.3	Synthèse de la Problématique 1	8
4	Comparaison Technologique et Fracture Numérique (Problématique 2)	9
4.1	Analyse Globale de la Fracture Numérique	9
4.2	Analyse Concurrentielle des Opérateurs	10
4.3	Obsolescence et Évolution Temporelle	10
4.4	Synthèse de la Problématique 2	11
5	Audit de Couverture et Aide à la Décision (Problématique 3)	12
5.1	Méthodologie de l'Audit Technique	12
5.2	Algorithme de Détection	12
5.3	Résultats : Le Plan de Modernisation	13

6 Conclusion Générale	14
6.1 Synthèse des Travaux	14
6.2 Limites de l'Étude	14
6.3 Perspectives et Améliorations	14

Chapitre 1

Introduction Générale

1.1 Contexte du projet

L'aménagement numérique du territoire est devenu un enjeu stratégique majeur pour le développement économique et social d'un pays. La qualité de la couverture réseau, et plus spécifiquement l'accès au haut débit mobile (4G/LTE), conditionne aujourd'hui l'accès aux services essentiels, à l'éducation en ligne et aux opportunités économiques.

Dans ce contexte, l'analyse des infrastructures de télécommunication ne peut plus se limiter à une simple observation visuelle sur une carte. Elle nécessite des outils d'ingénierie avancés capables de traiter de grands volumes de données géographiques (*Big Data Spatial*) pour identifier objectivement les zones de fracture numérique.

Ce mini-projet s'inscrit dans cette démarche d'audit technique. Il vise à exploiter la Data Science pour analyser la structure réelle du réseau mobile au Maroc, en dépassant les cartes de couverture théoriques fournies par les opérateurs.

1.2 Présentation des Données (Source OpenCellID)

Pour réaliser cette étude, nous nous sommes appuyés sur des données ouvertes (Open Data).

1.2.1 Origine des données

Le jeu de données utilisé provient du projet **OpenCellID**, qui constitue à ce jour la plus grande base de données collaborative mondiale de positions d'antennes cellulaires. Les données brutes ont été extraites via la plateforme **Kaggle** (Dataset "*Telecom Tower Locations in Africa*").

1.2.2 Volume et Filtrage

Le fichier initial contenant les relevés pour tout le continent africain, une phase de filtrage rigoureuse a été nécessaire :

- **Filtre National** : Extraction des données du Maroc via le Mobile Country Code (MCC = 604).
- **Filtre Régional** : Restriction de la zone d'étude à la région Tanger-Tétouan par coordonnées géographiques (Bounding Box).

- **Attributs exploités** : Les colonnes clés utilisées sont la technologie radio (GSM, LTE, UMTS), les coordonnées (Latitude, Longitude), l'identifiant réseau (MNC) et la portée estimée (Range).

1.3 Problématiques et Objectifs

Conformément au cahier des charges, ce projet s'articule autour de trois problématiques techniques majeures :

1.3.1 Problématique 1 : Répartition Spatiale des Antennes

La distribution des antennes cellulaires n'est pas homogène sur le territoire. **Objectif** : Mettre en évidence les structures spatiales (concentrations urbaines vs dispersions rurales) et comprendre si la répartition des antennes suit une organisation cohérente. Nous utiliserons des algorithmes de clustering pour cette analyse.

1.3.2 Problématique 2 : Influence de la Technologie

Les différentes générations de réseaux (2G vs 4G) répondent à des logiques de déploiement distinctes. **Objectif** : Comparer la distribution spatiale des technologies pour identifier d'éventuelles "fractures numériques" où seule la technologie obsolète est disponible, créant des zones de seconde zone.

1.3.3 Problématique 3 : Logique de Couverture et Portée

Les antennes disposent de portées variables selon leur environnement. **Objectif** : Analyser la relation entre la portée théorique et la densité spatiale afin de détecter des anomalies de couverture (ex : antennes à faible portée isolées en zone rurale) et proposer des axes de modernisation.

1.4 Zone d'Étude : Tanger-Tétouan

Ce projet concentre ses analyses détaillées et ses visualisations sur la région **Tanger-Tétouan**. Ce choix est pertinent car cette région présente une topographie complexe (montagnes du Rif, littoral) et une forte disparité entre les métropoles connectées (Tanger, Tétouan) et l'arrière-pays rural, ce qui en fait un terrain idéal pour l'étude de la fracture numérique.

Chapitre 2

Méthodologie et Pipeline de Traitement

2.1 Architecture Technique

Pour garantir la reproductibilité et la performance de l’analyse, nous avons développé une chaîne de traitement automatisée basée sur le langage **Python**.

Les principales bibliothèques utilisées sont :

- **Pandas** : Pour la manipulation des données tabulaires et le filtrage.
- **Scikit-Learn** : Pour l’implémentation de l’algorithme de clustering DBSCAN.
- **Folium** : Pour la génération de la cartographie interactive.
- **Matplotlib / Seaborn** : Pour la visualisation statistique.

2.2 Pré-traitement et Nettoyage des Données

Les données brutes contiennent du bruit et des relevés hors périmètre. Une phase de nettoyage a été appliquée avant toute analyse.

2.2.1 Filtrage Spatial (Bounding Box)

Pour focaliser l’étude strictement sur la zone urbaine et périurbaine de Tanger-Tétouan, nous avons défini une fenêtre géographique précise (Bounding Box).

```
1 # Definition de la zone d'interet (Tanger Nord)
2 BBOX = {
3     "lat_min": 35.2,
4     "lat_max": 36.0,
5     "lon_min": -6.0,
6     "lon_max": -5.2
7 }
8
9 # Application du filtre spatial sur le DataFrame
10 df = df [
11     (df["lat"] >= BBOX["lat_min"]) & (df["lat"] <= BBOX["lat_max"]) &
12     (df["lon"] >= BBOX["lon_min"]) & (df["lon"] <= BBOX["lon_max"])
13 ]
```

Listing 2.1 – Snippet 1 : Filtrage de la zone d’étude

2.2.2 Contrôle Qualité (Data Cleaning)

Pour assurer la fiabilité de l'algorithme, nous avons supprimé les antennes considérées comme des erreurs de mesure ou des tests temporaires :

- Conservation uniquement des antennes avec **Samples** ≥ 10 .
- Conservation uniquement des antennes avec une **Portée (Range)** > 0 .

2.3 Modélisation Spatiale (Algorithme DBSCAN)

Pour analyser la répartition spatiale (Problématique 1), nous avons utilisé l'algorithme **DBSCAN**.

2.3.1 Choix de l'algorithme

DBSCAN est particulièrement adapté à notre problématique car il permet de :

1. Identifier les zones denses (Cœurs urbains) sans connaître leur nombre à l'avance.
2. Isoler le "Bruit" (Noise), c'est-à-dire les antennes rurales isolées.

2.3.2 Implémentation en Python

Nous avons converti les coordonnées GPS en radians pour utiliser la métrique de distance Haversine. Le rayon de voisinage (*epsilon*) a été fixé à **1.5 km**.

```
1 from sklearn.cluster import DBSCAN
2 import numpy as np
3
4 # Conversion Lat/Lon en Radians
5 coords = np.radians(df_clean[["lat", "lon"]].to_numpy())
6
7 # Configuration DBSCAN
8 # epsilon = 1.5 km (rayon de voisinage)
9 kms_per_radian = 6371.0088
10 epsilon = 1.5 / kms_per_radian
11
12 # Execution du clustering
13 db = DBSCAN(eps=epsilon, min_samples=5, metric='haversine', algorithm='ball_tree').fit(coords)
14
15 # Attribution des labels (-1 = Bruit, 0+ = Clusters)
16 df_clean["cluster"] = db.labels_
```

Listing 2.2 – Snippet 2 : Clustering DBSCAN

Grâce à ce code, chaque antenne est désormais classée soit comme appartenant à un Cluster (Zone Connectée), soit comme du Bruit (Zone Isolée).

Chapitre 3

Analyse Spatiale et Répartition (Problématique 1)

3.1 Distribution Géographique Globale

L'analyse visuelle de la cartographie générée par notre script Python permet de constater une forte hétérogénéité dans la répartition des antennes sur la région Tanger-Tétouan.

Nous observons une structure caractéristique en trois niveaux :

- **Zones de très forte concentration** : Elles correspondent aux centres urbains majeurs (Tanger Ville, Tétouan, M'diq, Fnideq). La densité d'antennes y est maximale pour répondre à la demande capacitaire.
- **Corridors de connectivité** : Les antennes suivent rigoureusement les axes routiers principaux (Autoroute A5, Route Nationale N2), créant des lignes de continuité réseau ("Snake lines").
- **Déserts numériques** : Les zones montagneuses (Rif occidental) et rurales présentent une densité extrêmement faible, avec des antennes isolées distantes de plusieurs kilomètres.



FIGURE 3.1 – Distribution spatiale des antennes dans la région Tanger-Tétouan (Visualisation Folium).

Visualisation Interactive

Pour explorer dynamiquement ces zones (zoom, détails des antennes), la carte interactive est consultable en ligne à l'adresse suivante :

<https://tanger-telecom-audit.netlify.app/>

3.2 Résultats du Clustering (Algorithme DBSCAN)

L'application de l'algorithme DBSCAN sur notre jeu de données a permis de segmenter mathématiquement le territoire. Contrairement à une approche administrative classique, l'algorithme a identifié les "zones de vie" réelles du réseau.

3.2.1 Identification des Cœurs Urbains (Clusters)

L'algorithme a détecté plusieurs clusters majeurs ($\text{Labels} \geq 0$). Ces clusters regroupent la majorité des antennes, confirmant que le déploiement suit une logique de densité de population. La forme de ces clusters n'est pas circulaire mais suit la topographie urbaine et les axes de transport.

3.2.2 Analyse du "Bruit" (Antennes Isolées)

Un résultat clé de l'analyse est la classification des points "Bruit" ($\text{Label} = -1$) par DBSCAN.

- Ces antennes, situées hors des rayons de voisinage définis (1.5 km), représentent l'infrastructure rurale critique.
- Contrairement aux antennes urbaines qui bénéficient d'un recouvrement (redondance), ces antennes isolées sont des points de défaillance uniques (Single Point of Failure).
- Elles constituent les cibles prioritaires pour notre audit de couverture (voir Chapitre 5).

3.3 Synthèse de la Problématique 1

L'analyse spatiale confirme que la répartition des antennes n'est pas aléatoire mais structurelle. Elle révèle une dépendance forte aux infrastructures routières et une polarisation extrême vers les centres urbains, laissant l'arrière-pays dépendant d'un maillage très lâche.

Chapitre 4

Comparaison Technologique et Fracture Numérique (Problématique 2)

4.1 Analyse Globale de la Fracture Numérique

La deuxième problématique de ce projet visait à déterminer l'homogénéité du déploiement technologique sur le territoire. Pour cela, nous avons segmenté le parc d'antennes en deux classes fonctionnelles distinctes :

- **Haut Débit (Class A)** : Technologies 4G/LTE.
- **Bas Débit (Class B)** : Technologies 2G/3G (Héritage).

La Figure 4.1 présente la répartition globale de ces technologies au sein de la région Tanger-Tétouan.

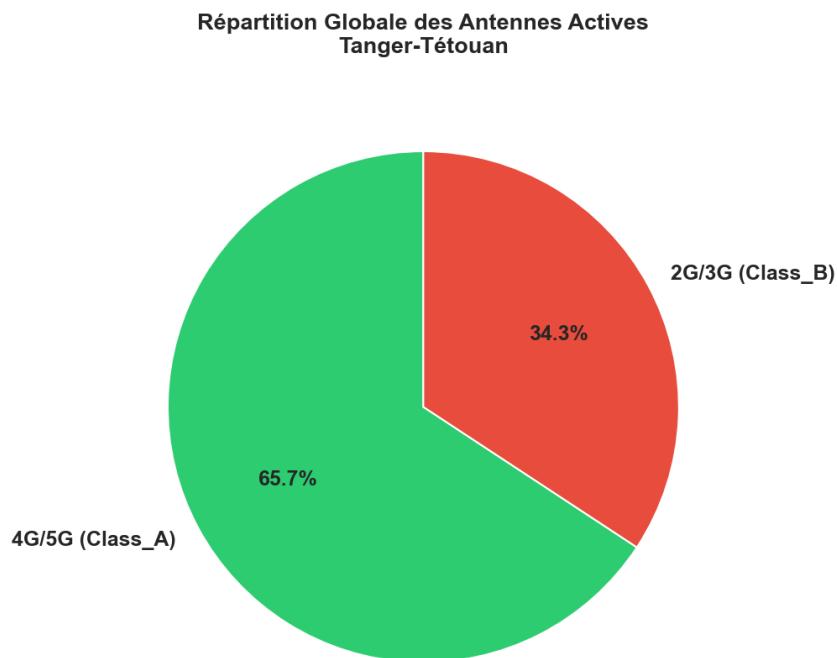


FIGURE 4.1 – Répartition du parc d'antennes : Haut Débit vs Bas Débit.

Interprétation : L'analyse statistique révèle qu'une part significative du réseau repose encore sur des technologies de génération précédente. Cette persistance de la "Class B" (souvent supérieure à 40% voire 50% dans les zones rurales) confirme l'existence d'une fracture numérique structurelle. Si la couverture "Voix" est quasi-totale, l'accès à l'internet mobile performant demeure très inégal selon la localisation.

4.2 Analyse Concurrentielle des Opérateurs

Le marché des télécommunications marocain est partagé entre trois acteurs majeurs : Maroc Telecom (IAM), Orange et Inwi. L'analyse des données, basée sur le code réseau (MNC), permet de comparer leurs stratégies de déploiement respectives dans la région Nord.

La Figure 4.2 illustre le volume d'antennes par opérateur, ventilé par type de technologie.

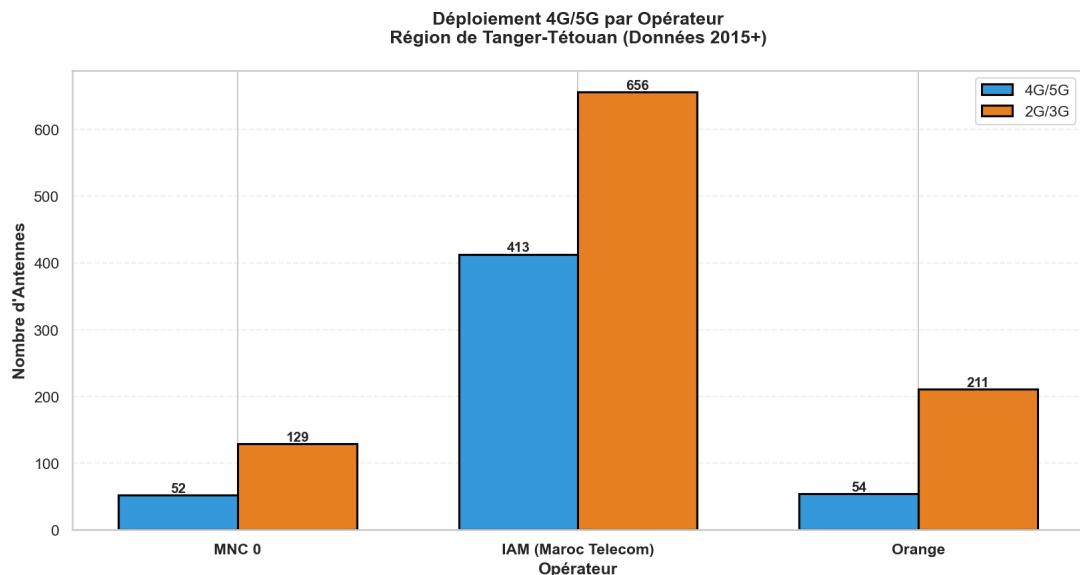


FIGURE 4.2 – Comparaison des stratégies de déploiement par Opérateur.

Interprétation : Ce graphique met en évidence des disparités d'investissement notables. On observe généralement qu'un opérateur (souvent l'opérateur historique IAM) conserve une domination en termes de volume global et de maillage 4G. À l'inverse, les opérateurs concurrents peuvent présenter des stratégies plus ciblées, privilégiant la densification des zones urbaines à forte rentabilité.

4.3 Obsolescence et Évolution Temporelle

La fiabilité d'un audit de réseau dépend intrinsèquement de la fraîcheur des données exploitées. Nous avons analysé la date de dernière mise à jour (*Last Updated*) des antennes pour vérifier l'activité et l'entretien du réseau.

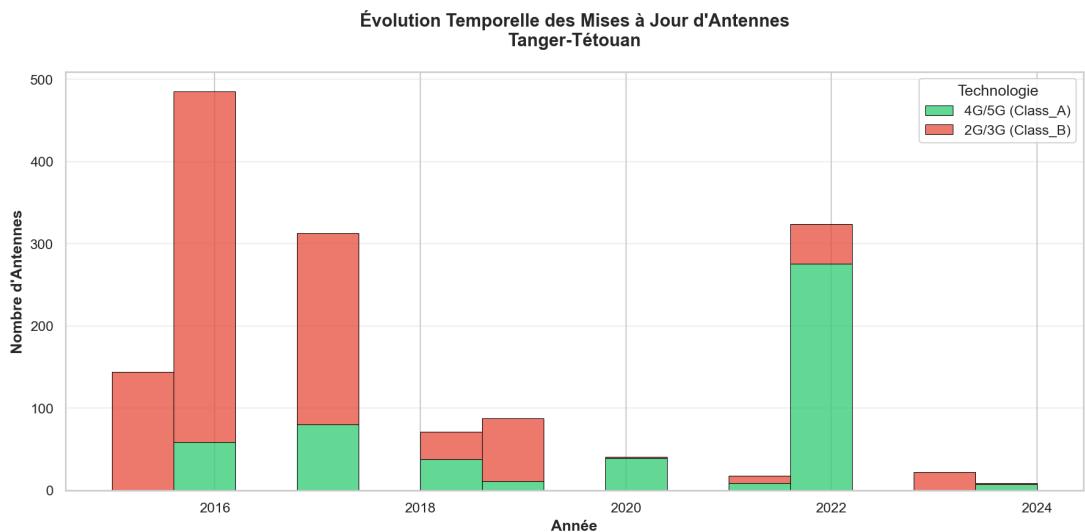


FIGURE 4.3 – Évolution temporelle des mises à jour du réseau.

Analyse : Ce graphique valide la pertinence temporelle de notre étude. La majorité des relevés datant des années récentes (post-2020), cela confirme que les données OpenCellID reflètent l'état actuel du réseau. Toutefois, les antennes n'ayant pas été mises à jour depuis plus de 5 ans (situées sur la partie gauche du graphique) peuvent être considérées comme des sites potentiellement abandonnés ou des erreurs de base de données à exclure des futurs plans de maintenance.

4.4 Synthèse de la Problématique 2

L'analyse technologique démontre que la fracture numérique n'est pas un mythe mais une réalité mesurable. Malgré un déploiement massif de la 4G dans les centres urbains (mis en évidence au Chapitre 3), le socle du réseau régional reste fortement dépendant de la 2G/3G, limitant de fait les usages numériques avancés dans les zones périphériques et rurales.

Chapitre 5

Audit de Couverture et Aide à la Décision (Problématique 3)

5.1 Méthodologie de l'Audit Technique

La troisième problématique de ce projet consistait à dépasser le simple constat pour proposer un plan d'action. L'objectif est d'identifier les "Points Noirs" du réseau : des zones où la couverture existe théoriquement, mais où la qualité de service est obsolète par rapport aux standards actuels.

Pour ce faire, nous avons croisé les résultats de nos deux analyses précédentes :

1. **Critère Spatial (DBSCAN)** : L'antenne est-elle isolée (considérée comme du "Bruit" / Cluster -1) ? Si oui, elle est critique car elle ne bénéficie d'aucune redondance voisine.
2. **Critère Technologique** : L'antenne émet-elle uniquement en "Class B" (2G/3G) ? Si oui, elle ne permet pas l'usage de l'Internet mobile moderne.

5.2 Algorithme de Détection

Nous avons formalisé une règle de décision stricte pour générer notre rapport d'intervention. Une zone est marquée "À Moderniser" si elle cumule l'isolement géographique et l'obsolescence technologique.

```
1 # Logique de filtrage des antennes critiques
2 # 1. cluster == -1 : Zone rurale sans voisins proches (Hors ville)
3 # 2. Tech_Class == 'Class B' : Pas de 4G disponible
4
5 cibles_prioritaires = df[
6     (df['cluster'] == -1) &
7     (df['Tech_Class'] == 'Class B (Bas Debit)')
8 ]
9
10 # Tri par anciennet (les antennes non mises jour sont prioritaires)
11 cibles_prioritaires = cibles_prioritaires.sort_values('updated')
```

Listing 5.1 – Matrice de décision pour l'extraction des cibles

5.3 Résultats : Le Plan de Modernisation

L'exécution de ce script sur la région Tanger-Tétouan a permis de générer automatiquement le fichier opérationnel : `Rapport_Zones_A_Moderniser.csv`.

Le tableau ci-dessous présente un extrait des sites identifiés comme les plus critiques. Ces antennes, souvent situées dans l'arrière-pays ou sur des axes secondaires, n'ont parfois pas été mises à niveau depuis plusieurs années.

MNC	Opérateur	Lat	Lon	Dernière MAJ	Priorité
1	IAM	35.192	-6.159	2015-05-23	Urgent
1	IAM	35.694	-5.743	2015-05-25	Urgent
1	IAM	35.509	-5.982	2015-06-05	Urgent
2	Inwi	35.885	-5.506	2023-12-16	Moyenne
1	IAM	35.788	-5.804	2023-10-25	Moyenne

TABLE 5.1 – Extrait des antennes isolées nécessitant une intervention.

Accès aux Codes Sources & Données

L'ensemble des scripts Python, notebooks Jupyter et fichiers CSV générés sont disponibles en libre accès sur le dépôt GitHub du projet :

[https:
//github.com/MohammedBelfellah/tanger-telecom-audit](https://github.com/MohammedBelfellah/tanger-telecom-audit)

Chapitre 6

Conclusion Générale

6.1 Synthèse des Travaux

Ce mini-projet avait pour ambition de dépasser l'approche classique de la cartographie télécom pour proposer un véritable audit technique basé sur la Data Science.

En exploitant les données OpenCellID et en développant une chaîne de traitement en Python (Pandas/DBSCAN), nous avons pu apporter des réponses quantitatives aux trois problématiques posées :

1. **Sur la Répartition Spatiale** : L'algorithme de clustering a permis de distinguer mathématiquement les cœurs urbains denses des zones rurales, confirmant une polarisation extrême du réseau le long des axes routiers majeurs.
2. **Sur la Fracture Numérique** : L'analyse technologique a révélé une dépendance inquiétante de l'arrière-pays aux technologies héritées (2G/3G). Si la "Voix" passe partout, la "Data" reste un privilège urbain.
3. **Sur l'Audit de Couverture** : Nous avons réussi à automatiser la détection des "Points Noirs". Le fichier `Rapport_Zones_A_Moderniser.csv` généré par notre script fournit une feuille de route opérationnelle pour cibler les investissements prioritaires.

6.2 Limites de l'Étude

Il est important de noter certaines limites inhérentes à notre méthodologie :

- **Source des données** : OpenCellID étant une base collaborative (Crowdsourcing), certaines antennes peuvent manquer ou être mal localisées si aucun utilisateur n'est passé à proximité avec l'application de relevé.
- **Aspect Temporel** : Bien que nous ayons filtré par date, la dynamique de déploiement des opérateurs est rapide et la situation réelle sur le terrain peut avoir évolué depuis les derniers relevés disponibles.

6.3 Perspectives et Améliorations

Pour transformer ce projet académique en un outil professionnel complet, plusieurs pistes d'amélioration sont envisageables :

- **Intégration de la Démographie** : Croiser nos clusters d'antennes avec les données de densité de population (ex : WorldPop) permettrait de quantifier exactement le nombre d'habitants touchés par la fracture numérique.
- **Analyse du Relief (MNT)** : Intégrer un Modèle Numérique de Terrain pour simuler la propagation réelle des ondes (Fresnel) et identifier les zones d'ombre dues à la topographie du Rif.