

Définir les traitements (attributaires et spatiaux)

Dans un projet de données, **les traitements** sont les opérations appliquées aux données pour les analyser, les transformer ou en extraire des informations utiles.

Deux types de traitements principaux :

✓ **Traitements attributaires** → concerne toutes les données de tous les jeux de données, dès lors qu'on travaille sur des valeurs tabulaires (attributs des entités).

- **S'appliquent à toutes les variables non spatiales** (nombres, catégories, dates...).

✓ **Traitements spatiaux** → Basés sur les **données géographiques** il inclut aussi des **analyses spatiales** (distance, localisation, distances, intersections...).

- **S'applique uniquement aux données contenant une information géographique** (coordonnées, limites d'un territoire...).
-

Traitements attributaires (Analyse tabulaire des données)

Objectif : Nettoyer et enrichir les données avant analyse.

- **Tri des données** → Suppression des variables non essentielles au projet (ex. : variables techniques inutilisées).
- **Jointure des données** → Liaison des stations Vélib' avec les données démographiques via la clé `id_arrondissement` (arrondissement commun aux deux jeux de données).
- **Calculs statistiques** :
 - **Densité de population** = `Population / Superficie`.
 - **Ratio de stations par habitant** = `Nombre de stations / Population`.
- **Analyse comparative** → Comparaison du nombre de stations avec la population et la densité pour identifier les écarts d'équipement entre arrondissements.

Traitements spatiaux (Analyse géographique des données)

Objectif : Visualiser et analyser la distribution des stations Vélib' sur le territoire parisien.

- **Géolocalisation des stations** → Affichage des stations sur une carte via un Système d'Information Géographique (QGIS, GeoPandas, Leaflet...).
 - **Analyse spatiale** :
 - **Densité des stations par arrondissement** = Nombre de stations / Superficie de l'arrondissement.
 - **Corrélation entre densité de stations et densité de population** → Identification des zones sous-équipées en fonction de la demande potentielle.
-

Structure générale recommandée pour une réponse complète en entreprise

1- Préparation des données *(Optionnel selon le besoin)*

- **Tri des données** → Sélection des variables utiles.
- **Nettoyage des données** → Suppression des valeurs aberrantes ou manquantes.

2- Enrichissement des données *(Optionnel si plusieurs sources)*

- **Jointure des données** → Fusion avec d'autres jeux de données pertinents (ex. données démographiques).

3- Traitements analytiques (attributaires)

- **Calculs statistiques** → Exemples : ratios, moyennes, tendances.
- **Comparaison et analyse** → Identifier des corrélations ou tendances.

4- Traitements géographiques (spatiaux) (si applicable)

- **Géolocalisation et visualisation cartographique** → Affichage des données sur une carte.
- **Analyse spatiale avancée** → Calcul de distances, détection de zones sous-équipées...