**1. Compréhension et nettoyage des données**

**Compréhension du jeu de données**

* **Table principale : TIME**  
  Contient les heures réellement travaillées par consultant, avec les métadonnées de mission, taux horaires (Charge-Out Rate), surcharges (Adm. Surcharge (%)), clients, phases de projet, etc.
* **Structure temporelle**  
  Création de la colonne **“Début de semaine”** à partir de la Work Date, pour structurer les analyses en **base hebdomadaire**, indispensable pour les KPI et la planification.
* **Enrichissement du modèle**
  + **Fusion** avec Staffing pour croiser les **heures réelles vs heures planifiées**.
  + Ajout de la table Budget pour suivre les **coûts réels vs budget projeté**.
  + Intégration de SeuilsBench pour évaluer le **niveau critique du bench** en fonction du niveau hiérarchique.
* **Création manuelle et catégorisation des consultants**
  + Création d’une **table Consultant type** permettant de distinguer les **consultants internes** (coût basé sur une estimation salariale) et **externes** (coût estimé selon une pondération du Std. Price et d’un référentiel interne).
  + Cette catégorisation a permis de calculer **des marges précises**, différenciées selon les profils, et d’améliorer les analyses par typologie de ressources.

**Nettoyage et structuration des données**

* **Traitement des dates manquantes et erreurs de format**  
  Suppression des lignes invalides ou incomplètes, conversion robuste de Work Date en datetime et gestion des erreurs.
* **Suppression des doublons**
  + Appliquée dans les tables TIME et Staffing, notamment lors des jointures sur Employee Name, Semaine, No. Phase\_projet.
  + Cette étape a évité les **erreurs de surreprésentation d’heures** ou de coûts dans les agrégations.
* **Harmonisation des noms de variables**  
  Nettoyage et uniformisation des noms de colonnes (Personnel No., Phase\_projet, etc.) pour permettre des jointures solides entre les différentes tables (même s’il y avait initialement des **relations many-to-many** difficiles à gérer).

**Création des tables clés du modèle Power BI**

* **TableSemaines** :  
  Table virtuelle contenant l’agrégation par Employee Name et Début de semaine avec la variable y\_reel (heures réelles travaillées). Utilisée dans plusieurs mesures clés (Utilization Rate, RMSE…).
* **TableFinale\_Join** (Table centrale physique) :  
  Créée dans **Power Query** via fusion entre TableSemaines et Staffing, elle intègre :
  + y\_reel (heures réelles)
  + y\_prev (heures planifiées)
  + Toutes les colonnes contextuelles (Staff Level, Client Name, etc.) Cette table permet une **analyse complète, filtrable dynamiquement**, avec une **cohérence parfaite entre les sources**.
* **Staffing - 2024 / Staffing - 2025** :  
  Séparation de l’historique et de la prévision, utilisée pour tester différents scénarios de planification, et entraîner le modèle de prédiction sur le passé.