**Plan de Travail pour l'Analyse de la Base de Données**

**1. Définition des Objectifs**

* Identifier les indicateurs clés à extraire.
* Déterminer les questions spécifiques que vous souhaitez répondre avec les régressions.
* Fixer des objectifs mesurables pour l'analyse.

**2. Exploration des Données (EDA)**

* Charger la base de données mheath.
* Analyser la structure des données (types de variables, dimensions).
* Visualiser les distributions des variables et les relations entre elles.
* Identifier les valeurs manquantes et les anomalies.

**3. Extraction des Indicateurs**

* Calculer des indicateurs pertinents (moyennes, médianes, écarts-types, etc.).
* Créer des visualisations pour illustrer ces indicateurs (histogrammes, boîtes à moustaches, etc.).
* Identifier les corrélations entre les variables pour sélectionner des candidats à la régression.

**4. Préparation des Données**

* Nettoyer les données : traiter les valeurs manquantes et les doublons.
* Normaliser ou standardiser les variables si nécessaire.
* Créer des variables dérivées si pertinent (par exemple, ratios, interactions).

**5. Sélection et Application des Modèles de Régression**

* **Régression Linéaire** : Analyser les relations linéaires entre les variables.
* **Régression Logistique** : Si vous travaillez avec des variables de classification.
* **Régression Polynomiale** : Pour capturer des relations non linéaires si nécessaire.
* Sélectionner les variables indépendantes en fonction des corrélations et des résultats de l'EDA.

**6. Évaluation des Modèles**

* Diviser les données en ensembles d'entraînement et de test.
* Évaluer la performance des modèles avec des métriques appropriées (MSE, RMSE, R², AUC pour la logistique).
* Utiliser la validation croisée pour garantir la robustesse des résultats.

**7. Interprétation des Résultats**

* Analyser les coefficients des modèles pour comprendre l'impact des variables.
* Visualiser les résultats avec des graphiques (prédictions vs valeurs réelles, résidus).

**8. Prise de Décisions**

* Basé sur les résultats, formuler des recommandations concrètes.
* Identifier les mesures à prendre pour optimiser les performances ou résoudre des problèmes.

**9. Documentation et Communication**

* Préparer un rapport ou une présentation des résultats clés, des méthodes et des recommandations.
* Utiliser des visualisations pour rendre les résultats accessibles aux parties prenantes.

**10. Suivi et Amélioration**

* Évaluer l'impact des mesures prises.
* Mettre à jour régulièrement le modèle avec de nouvelles données et affiner les analyses.

**1. Prédiction**

* **Anticipation** : Les modèles permettent de prédire des valeurs futures basées sur des données historiques, ce qui est essentiel dans des domaines comme la finance, la santé et le marketing.

**2. Compréhension des Relations**

* **Analyse des Corrélations** : Ils aident à comprendre comment différentes variables sont liées entre elles. Par exemple, dans une étude de marché, on peut analyser l'impact des prix sur les ventes.

**3. Identification des Facteurs Déterminants**

* **Causalité** : Les modèles de régression peuvent identifier quels facteurs ont le plus d'influence sur un résultat donné, aidant ainsi à cibler les actions à entreprendre.

**4. Optimisation des Ressources**

* **Prise de Décision** : En fournissant des insights basés sur des données, les modèles aident les entreprises à allouer efficacement leurs ressources pour maximiser le rendement.

**5. Évaluation de Scénarios**

* **Simulation** : Les modèles permettent de simuler différents scénarios pour évaluer les conséquences potentielles de décisions ou de changements dans les variables.

**6. Détection d'Anomalies**

* **Surveillance** : Les modèles peuvent identifier des valeurs aberrantes ou des comportements inattendus dans les données, ce qui est crucial pour la détection de fraudes ou de problèmes opérationnels.

**7. Amélioration Continue**

* **Feedback** : En analysant les résultats des modèles, les organisations peuvent ajuster leurs stratégies et améliorer leurs processus de manière continue.

**8. Communication des Résultats**

* **Visualisation** : Les modèles facilitent la création de visualisations claires qui aident à communiquer des résultats complexes de manière accessible aux parties prenantes.

**9. Prise de Mesures Proactives**

* **Prévention** : En anticipant des problèmes potentiels, les organisations peuvent agir de manière proactive pour atténuer les risques.

**Conclusion**

Utiliser des modèles d'analyse de données, notamment des modèles de régression, est essentiel pour tirer des insights significatifs à partir des données, améliorer la prise de décision et optimiser les performances dans divers domaines.

**Prédictions liées à la mère :**

**1°) Risque de complications pendant l'accouchement**

Prédiction des risques de complications telles que l'hémorragie post-partum, l'éclampsie ou la rupture utérine en fonction des antécédents médicaux de la mère, de son âge, de son IMC et des données de la consultation prénatale (CPN) [**table accouchement et cpn**]

**2°) Type d'accouchement**

Prédiction du type d'accouchement (voie basse ou césarienne) en fonction de divers facteurs tels que l'âge de la mère, ses antécédents de césariennes, la présentation du fœtus et les complications survenues pendant la grossesse.

**Prédictions liées au nouveau-né :**

**1°) Poids de naissance**

Prédiction du poids de naissance du bébé en fonction de facteurs tels que l'âge gestationnel, la santé de la mère et les données de la CPN.

**2°) Risque de morbidité néonatale**

Évaluation du risque de morbidité néonatale, comme l'admission en unité de soins intensifs néonatals (USIN), en fonction des caractéristiques de la mère et de l'accouchement.

**Autres analyses et prédictions possibles :**

* **Durée du travail :** Prédiction de la durée des différentes phases du travail (dilatation, expulsion, délivrance) en fonction de divers facteurs maternels et obstétricaux.
* **Nécessité d'interventions médicales :** Prédiction de la nécessité d'interventions médicales telles que l'épisiotomie, la perfusion d'ocytocine ou l'analgésie péridurale en fonction des caractéristiques de la mère et de la progression du travail.
* **Impact des soins prénataux :** Évaluation de l'impact des soins prénataux sur les résultats de l'accouchement, comme le taux de césariennes ou les complications néonatales.
* **Analyse des tendances :** Analyse des tendances au fil du temps concernant les pratiques d'accouchement, les complications et les résultats pour identifier les domaines à améliorer.

**Techniques de modélisation possibles :**

* **Régression logistique :** Pour prédire les résultats binaires tels que le type d'accouchement (voie basse ou césarienne) ou la présence de complications.
* **Régression linéaire :** Pour prédire les variables continues telles que le poids de naissance ou la durée du travail.
* **Analyse de survie :** Pour analyser le temps jusqu'à l'accouchement spontané et identifier les facteurs qui influencent la durée de la grossesse.
* **Machine learning :** Utilisation d'algorithmes de machine learning tels que les arbres de décision, les forêts aléatoires ou les réseaux neuronaux pour améliorer la précision des prédictions et identifier des relations complexes entre les variables.

# **I°) Accouchement et Santé Maternelle**

## 1°) Prévalence des Complications

Utiliser des données de la table accouchement pour prédire les complications obstétricales en fonction de facteurs comme l'âge, l'état de santé de la mère et les soins reçus.

* **Régression logistique** pour prédire la présence de complications obstétricales (binaire : oui/non) en fonction de facteurs tels que l'âge, l'état de santé, et les soins reçus.

2°) Risque de Décès Maternel et Néonatal

Prédire les risques de décès avant ou après la sortie en analysant les variables pertinentes des tables cpon\_mere, cpon\_enfant, et infection\_neo\_natale.

* **Régression logistique** pour prédire le risque de décès avant ou après la sortie en fonction des caractéristiques de la mère et du nouveau-né.

# II°) Vaccination et Suivi de la Santé Infantile

## 1°) Taux de Vaccination

Prédire les taux de vaccination des enfants à partir des données des tables vaccination, enfant, et rapport\_vad\_pfe.

* **Régression linéaire** pour prédire le taux de vaccination en fonction de variables démographiques et de la disponibilité des services de santé.

## **2°) Efficacité des Campagnes de Vaccination**

Évaluer l'impact des différentes campagnes (campagne) sur les taux de vaccination et de maladies infantiles.

* **Analyse de variance (ANOVA)** pour comparer les taux de vaccination entre différentes campagnes.

# III°) Planification Familiale

**Demande et Adoption de la Planification Familiale**

Analyser les données dans cpon\_mere et evaluation\_adoption pour prédire l'acceptation des méthodes de planification familiale.

**Régression logistique** pour prédire l'adoption des méthodes de planification familiale en fonction des caractéristiques sociodémographiques et des services offerts.

# IV°) Suivi des Intrants Médicaux

**Disponibilité des Médicaments**

Prédire les ruptures de stocks ou la disponibilité des médicaments en utilisant les tables intrant\_commande, intrant\_reception, et suivi\_dispo\_medoc\_intrant.

* **Régression linéaire** pour prédire la disponibilité des médicaments en fonction des commandes et des réceptions.

**Efficacité des Chaînes d'Approvisionnement**

Évaluer l'efficacité de la distribution des intrants à travers les différentes localités (localite, district).

* **Régression multiple** pour évaluer les facteurs influençant l’efficacité de la distribution des intrants.

# V°) Évaluation des Services de Santé

* **Satisfaction et Qualité des Soins** : Prédire les niveaux de satisfaction des patients en utilisant les données de contact, alerte, et suivi\_activites\_communautaires.

**Régression ordinale** si la satisfaction est mesurée sur une échelle (par exemple : très insatisfait à très satisfait).

* **Impact des Interventions Communautaires** : Évaluer l'impact des programmes de sensibilisation (causerie\_groupe, suivi\_activites\_communautaires) sur la santé des populations.

**Régression linéaire** pour évaluer l'impact des programmes sur les résultats de santé, comme la réduction des maladies.

# VI°) Analyse Démographique

* **Caractéristiques des Ménages** : Utiliser les données de menage, femme, et enfant pour prédire des indicateurs de santé, comme la prévalence de maladies ou les besoins en soins.

**Régression linéaire** pour prédire des indicateurs de santé à partir des caractéristiques des ménages.

* **Migration et Mobilité** : Analyser les données de district\_user et espc\_agent pour comprendre les mouvements de population et leur impact sur l'accès aux soins.

**Régression logistique** pour prédire l'impact de la migration sur l'accès aux soins.

# VII°) Suivi des Activités Communautaires

* **Engagement Communautaire** : Prédire le niveau d'engagement des membres de la communauté dans les activités de santé en utilisant des données de suivi\_activites\_communautaires et suivi\_besoin\_medoc\_intrant.

**Régression logistique** pour prédire le niveau d'engagement en fonction de divers facteurs sociaux et économiques.

# VIII°) Surveillance des Maladies

* **Surveillance Épidémiologique** : Utiliser les tables surveillance, signe\_danger\_pec, et infection\_neo\_natale pour prédire des épidémies potentielles ou des pics de maladies.

**Modèles de régression de Poisson** pour modéliser le nombre d'événements (ex. : cas de maladie) en fonction du temps et d'autres facteurs.