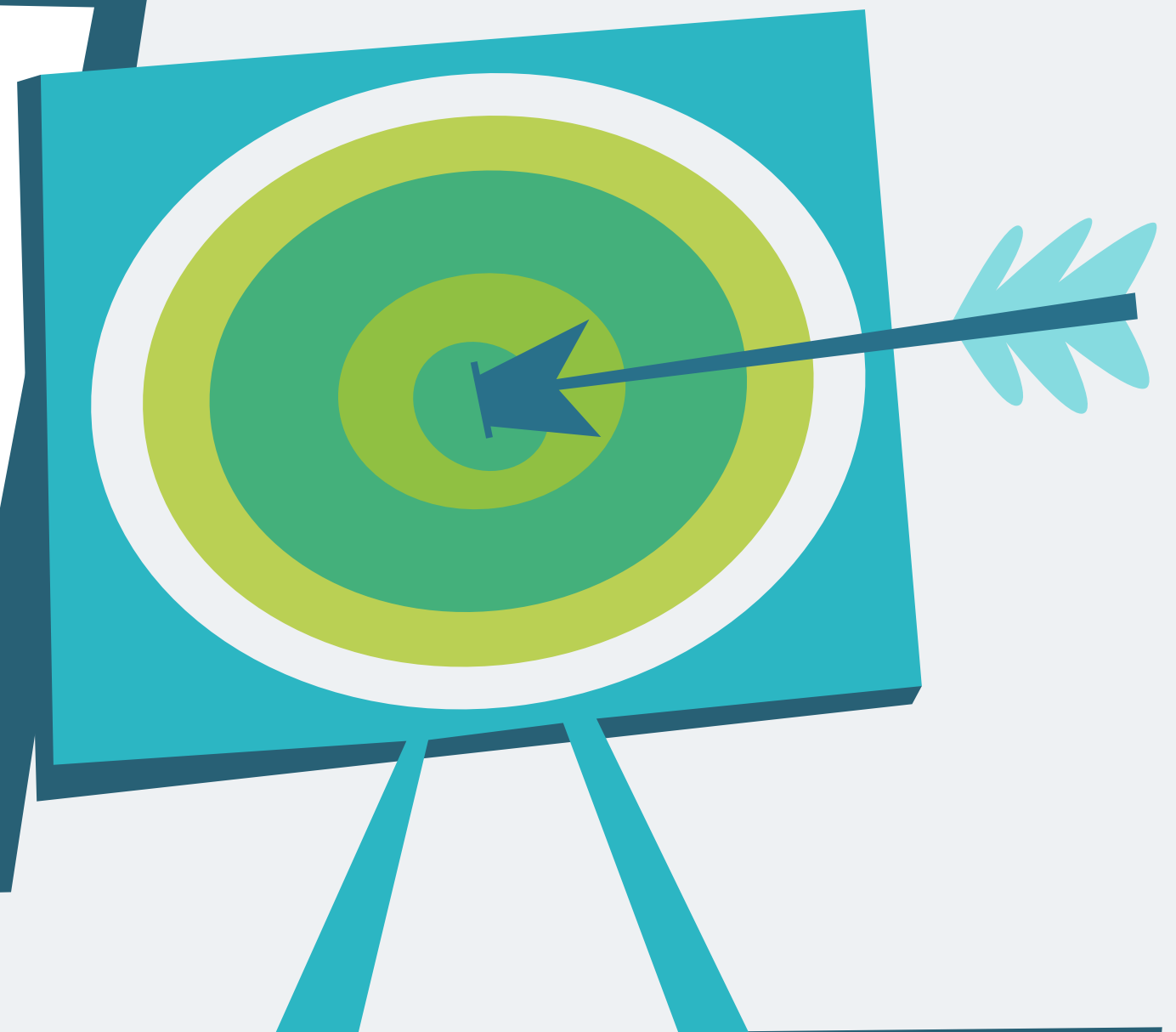


Le cycle de vie de la science des données est un processus structuré qui guide les praticiens à travers différentes étapes pour obtenir des informations efficaces sur les données. Les étapes clés comprennent :



## COMPRÉHENSION DE L'ENTREPRISE

Définir les objectifs du projet avec les parties prenantes pour aligner les solutions avec les objectifs.



## EXPLORATION DE DONNÉES:

Recueillir des données pertinentes, en tenant compte de la qualité, de la pertinence et de l'accessibilité.

## NETTOYAGE DES DONNÉES:

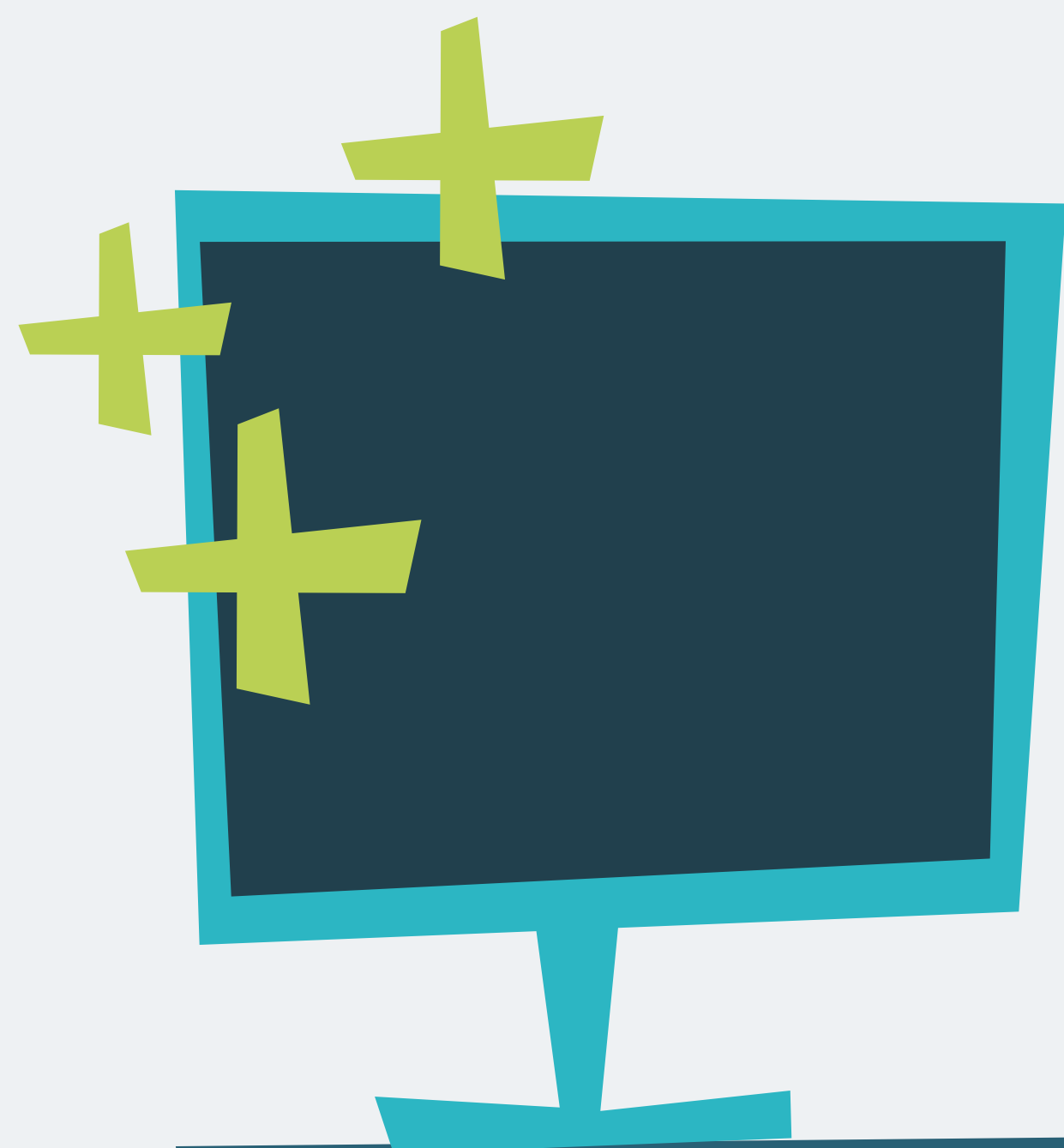
Gérez les incohérences, les erreurs et les valeurs manquantes pour une analyse précise.



## EXPLORATION DES DONNÉES:

Analyser les données nettoyées pour les tendances et les idées en utilisant des statistiques et des visuels.





## **INGÉNIERIE DES FONCTIONNALITÉS:**

Améliorer les performances du modèle en sélectionnant ou en transformant les fonctionnalités.

## **MODÉLISATION PRÉDICTIVE:**

Développer et affiner les modèles d'apprentissage automatique pour des prédictions précises.



# VISUALISATION DES DONNÉES:

Utiliser des éléments visuels pour communiquer des constatations complexes aux intervenants, ce qui facilite la prise de décisions.



# QU'EN EST-IL DES DATA SCIENTISTS ? VOUS ÊTES-VOUS DÉJÀ POSÉ DES QUESTIONS SUR LEURS RÔLES?

Le rôle d'un Data Scientist peut être décrit comme celui d'un "Jack-of-All-Trade" en raison de la diversité des compétences requises pour exceller dans ce domaine.

Un Data Scientist porte plusieurs chapeaux, combinant l'expertise de divers domaines pour extraire des informations significatives des données.





Voici une exploration des différentes facettes qui font d'un data scientist un véritable polymathe:

# MATHÉMATICIEN

**RÔLE:** Les scientifiques des données utilisent des concepts et des techniques mathématiques pour créer des modèles et des algorithmes robustes. De l'algèbre linéaire pour les opérations matricielles au calcul pour l'optimisation, une base mathématique solide est cruciale pour comprendre et développer des modèles complexes d'apprentissage automatique.





# STATISTICIEN



**RÔLE:** Les connaissances statistiques sont fondamentales pour les Data Scientists en termes de tests d'hypothèses, de théorie des probabilités et de statistiques inférentielles. Les scientifiques utilisent des méthodes statistiques pour tirer des conclusions fiables à partir des données, faire des prédictions et évaluer la signification de leurs résultats.

# SCIENTIFIQUE

**RÔLE:** Les scientifiques des données suivent une approche scientifique dans leur travail. Ils forment des hypothèses, conçoivent des expériences (ou des analyses) et évaluent systématiquement les résultats. Cet état d'esprit scientifique assure la fiabilité et la reproductibilité de leurs résultats, contribuant à la rigueur globale du processus de Data Science.




# PROGRAMMEUR

**RÔLE:** La maîtrise de la programmation est un attribut clé d'un scientifique des données. Ils utilisent des langages de programmation tels que Python ou R pour manipuler les données, construire des modèles et créer des solutions évolutives. Les compétences de codage sont essentielles pour mettre en œuvre des algorithmes, automatiser des processus et intégrer des solutions analytiques dans des environnements de production.



Dans les domaines des STIM, l'expertise d'un scientifique des données compétent est inestimable pour exploiter les connaissances tirées des données afin de faire progresser la recherche, d'optimiser les processus et de prendre des décisions éclairées, contribuant ainsi au progrès scientifique et à l'innovation.

Quels sont les avantages d'utiliser des outils?



01001  
11010



01010100  
01011011  
01110110

## 1. AMÉLIORATION DU FLUX DE TRAVAIL

Jupyter Notebooks favorise un flux de travail intégré, combinant code, visuels et texte pour un processus d'analyse cohésif et itératif.



## 2. GAIN DE TEMPS

Des outils comme Apache Spark permettent le traitement parallèle de grands ensembles de données, ce qui réduit les temps de calcul. Les fonctionnalités d'automatisation et d'optimisation libèrent les data scientists des tâches manuelles, leur permettant de se concentrer sur le travail de haut niveau.

### 3. COHÉRENCE

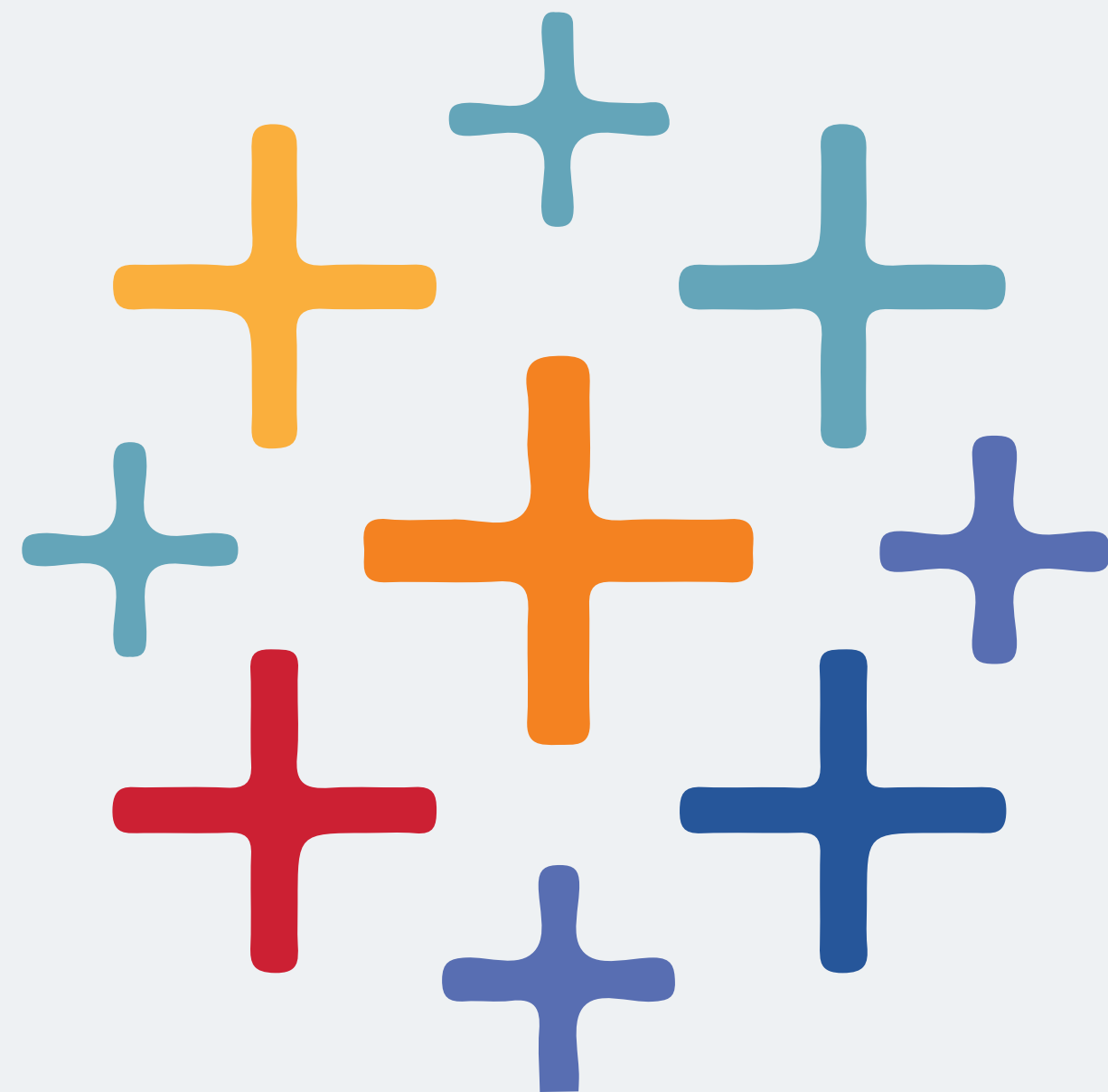
La cohérence dans l'analyse est vitale pour la fiabilité, et Power BI permet aux data scientists de visualiser les données, en se connectant à diverses sources de données pour des rapports interactifs et des tableaux de bord. Son intégration avec les produits Microsoft et son interface conviviale facilitent la narration des données et la prise de décision collaborative.



## VOICI D'AUTRES OUTILS CLÉS DE LA SCIENCE DES DONNÉES:

### EXCEL

Excel, un logiciel de tableur largement utilisé, est essentiel pour la manipulation et l'analyse de base des données. Son interface intuitive facilite l'exploration des ensembles de données, les calculs et les visualisations simples, ce qui en fait un atout précieux dans la boîte à outils d'un scientifique des données.



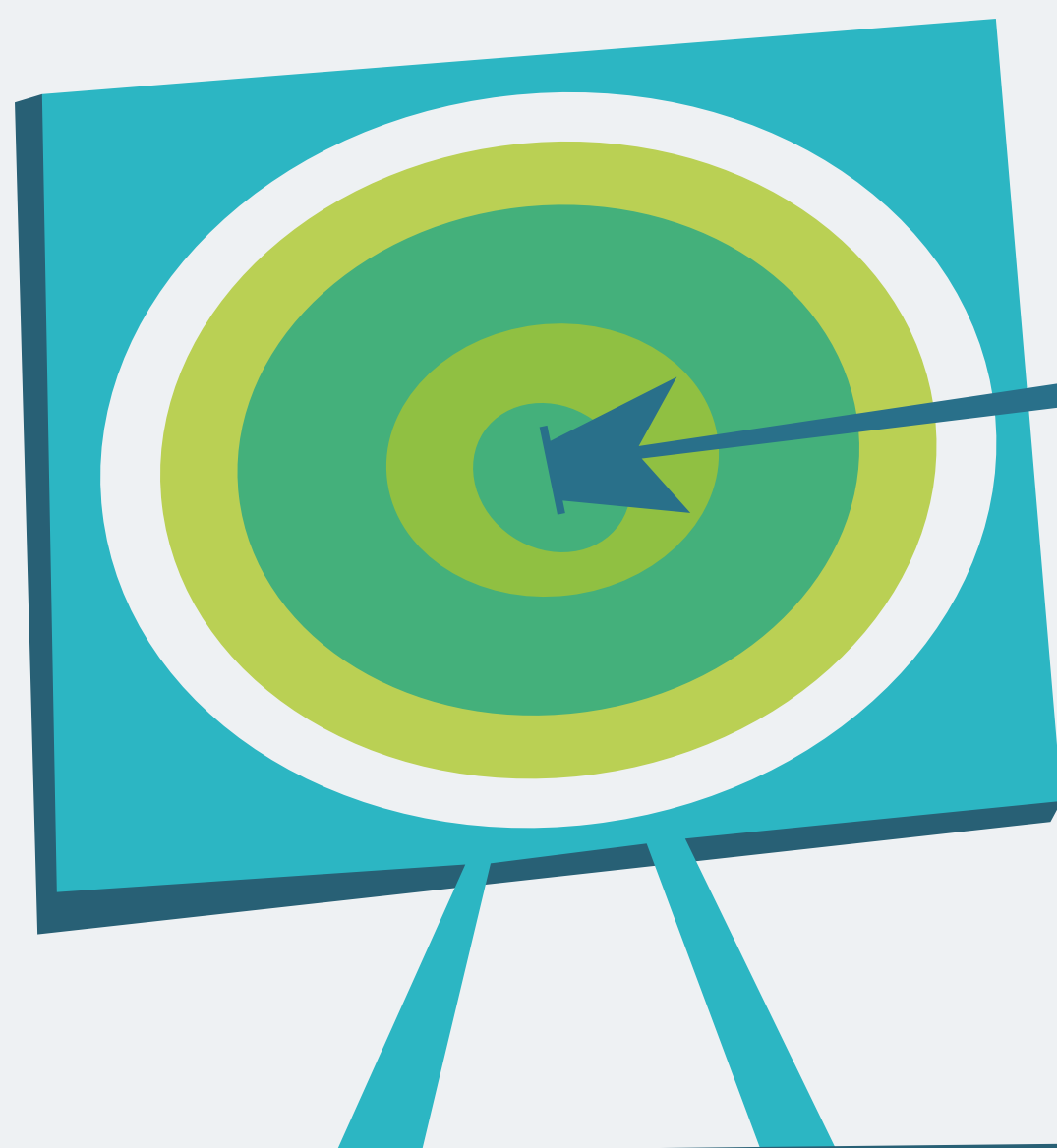
### TABLEAU

Tableau est un outil intuitif de visualisation des données qui permet aux Data Scientists de créer des tableaux de bord interactifs et perspicaces avec une interface glisser-déposer, rendant les données complexes facilement compréhensibles pour tous les niveaux techniques.



Dans divers domaines, notamment les affaires, les soins de santé et les finances, la science des données facilite la prise de décisions plus éclairées et stratégiques grâce à un processus systématique en cinq étapes :





## 1. DÉFINISSEZ VOS OBJECTIFS

Énoncer clairement les objectifs de la prise de décisions axée sur les données, de la définition des défis ou des possibilités en fonction des objectifs organisationnels.

## 2. DÉTERMINER LES SOURCES DE DONNÉES

Évaluer et sélectionner les sources de données, en tenant compte des ensembles de données internes, des sources externes ou d'un mélange. Évaluer la qualité, l'accessibilité et la pertinence des données.



## CRÉER OU ACHETER UNE SOLUTION DE TRAITEMENT DES DONNÉES

Développez ou sélectionnez une solution de traitement des données, en assurant l'efficacité, l'évolutivité, la sécurité des données et l'intégration avec diverses sources.



## EXTRAIRE LA VALEUR DES DONNÉES

Effectuez des analyses à l'aide de méthodes statistiques et d'apprentissage automatique pour obtenir des informations exploitables, alignées sur les objectifs définis.



# PROMOUVOIR UNE CULTURE AXÉE SUR LES DONNÉES DANS VOTRE ORGANISATION

Cultiver une culture organisationnelle valorisant la prise de décision axée sur les données. Sensibiliser, encourager la collaboration et améliorer la littératie des données.

