



AMIGOS

Rapport

DATA CLEANING

2023



Contents

Informations sur le travail	01
introduction	02
Définition et Le but du nettoyage de base de données.	03
Les avantages du nettoyage de données	04
Outils de nettoyage de base de données	05
Explication de l'outil excel.	06
Explication de l'outil excel.	07
Relationships	08
conclusion	09

Contents

Ce rapport parle de l'importance et du but du nettoyage d'une base de données



Informations sur le travail



Date d'initiation:19--12-2022

Date de livraison:19-01-2023



Type travail collective.



Nom du créateur

keroum ahmed faycal

khettaf el mendil tayeb

tordjmane mohammed

1.introduction

Le nettoyage de base de données est un processus qui permet de nettoyer les données de toute erreur ou imperfection, y compris les données erronées ou vides et redondantes, et de vérifier que les données correspondent aux critères définis et forment une collection de données de bonne qualité et fiables.

Data Warehouse est un entrepôt de données conçu spécifiquement qui stocke des données de haute qualité et fiables conformes aux critères définis pour les fins statistiques et analytiques. Lorsque de nouvelles données sont ajoutées à l'entrepôt, ces données sont nettoyées pour s'assurer qu'elles forment une collection de données de bonne qualité et fiables.

Ainsi, le nettoyage de base de données peut être une partie du processus fondamental d'entrée de données correctes dans le Data Warehouse. Cela conduit à une amélioration des performances du système et à une garantie que les données stockées forment une collection de données de bonne qualité et fiables.

2-Définition et Le but du nettoyage de base de données.

2.1-Définition:

Le nettoyage de base de données est un processus visant à améliorer la qualité des données et à corriger les erreurs ou les imperfections de légitimité des données. Cela comprend la vérification de l'exactitude des données, la suppression des données erronées ou en double, la suppression des données vides et la vérification que les données peuvent être collectées et analysées de manière correcte. Le résultat final du processus de nettoyage est une base de données organisée et fiable.

2.2-Le but :

Le but de nettoyer une base de données avant de la stocker dans un data warehouse est de garantir la qualité et la fiabilité des données utilisées pour prendre des décisions commerciales importantes. Le nettoyage des données consiste à corriger les erreurs, à remplacer les valeurs manquantes, à supprimer les doublons et à normaliser les données pour réduire la redondance. En nettoyant les données, les entreprises peuvent être assurées que les informations stockées dans le data warehouse sont fiables et précises, ce qui améliore la précision des analyses et des décisions prises en conséquence. De plus, le nettoyage des données peut améliorer les performances du système en réduisant la quantité de données à traiter et en simplifiant les processus d'analyse.

<https://datascientest.com/data-cleaning-2>

3-Les avantages du nettoyage de données.

Le nettoyage de données offre de nombreux avantages. L'un des principaux avantages est d' permettre une meilleure prise de décision basée sur des données de qualité supérieure.

En raison de l'importance croissante des données dans tous les secteurs, la qualité des données a un impact positif sur toutes les activités impliquant des données.

Pour tirer le meilleur parti de cette pratique, le nettoyage de données doit être considéré comme un effort de l'ensemble de l'entreprise. Il peut fluidifier les opérations de l'entreprise et accroître la productivité en éliminant les données incorrectes qui peuvent prendre beaucoup de temps.

Les ventes peuvent augmenter si les équipes marketing ont accès aux meilleures données. La combinaison de ces différents avantages internes et externes peut conduire à une rentabilité accrue.

4-Étapes de nettoyage de la base de données

Il existe généralement plusieurs étapes pour le nettoyage de données :

1. Collecte de données : La première étape consiste à collecter les données de différentes sources telles que les bases de données, les fichiers, les formulaires, etc.
2. Inspection des données : La deuxième étape consiste à inspecter les données collectées pour détecter les erreurs, les incohérences, les données manquantes ou les données dupliquées.
3. Correction des erreurs : La troisième étape consiste à corriger les erreurs détectées en utilisant des algorithmes ou en modifiant manuellement les données.
4. Normalisation des données : La quatrième étape consiste à normaliser les données pour les formater de manière cohérente et standardisée.
5. Stockage des données nettoyées : La dernière étape consiste à stocker les données nettoyées dans une base de données centralisée pour une utilisation ultérieure.

Outils de nettoyage de base de données les plus populaires :

1. OpenRefine : un outil open source pour le nettoyage et l'analyse des données.
2. Trifacta Wrangler : une solution d'analyse de données pour le nettoyage et la transformation des données.
3. Talend Data Cleansing : un composant de l'environnement de développement Talend open source pour le nettoyage des données.
4. Informatica Data Quality : une solution d'analyse de données spécialisée dans le nettoyage des données.
5. Nettoyage de données Pythonic Avec pandas et NumPy
6. Microsoft Power Query : une outil dans Microsoft Excel pour le nettoyage et l'analyse des données.

L'outil excel est utilisé dans notre projet

Excel est un puissant et complet logiciel d'analyse de données utilisé largement pour la gestion et l'analyse des données. Lorsque vous utilisez Excel pour nettoyer une base de données, l'utilisateur peut facilement analyser les données et éliminer les erreurs.

Excel dispose de plusieurs fonctionnalités de nettoyage de bases de données, par exemple, l'utilisateur peut supprimer les lignes vides et les lignes en double, supprimer les mots répétés, corriger les erreurs de frappe et éliminer les entrées incorrectes.

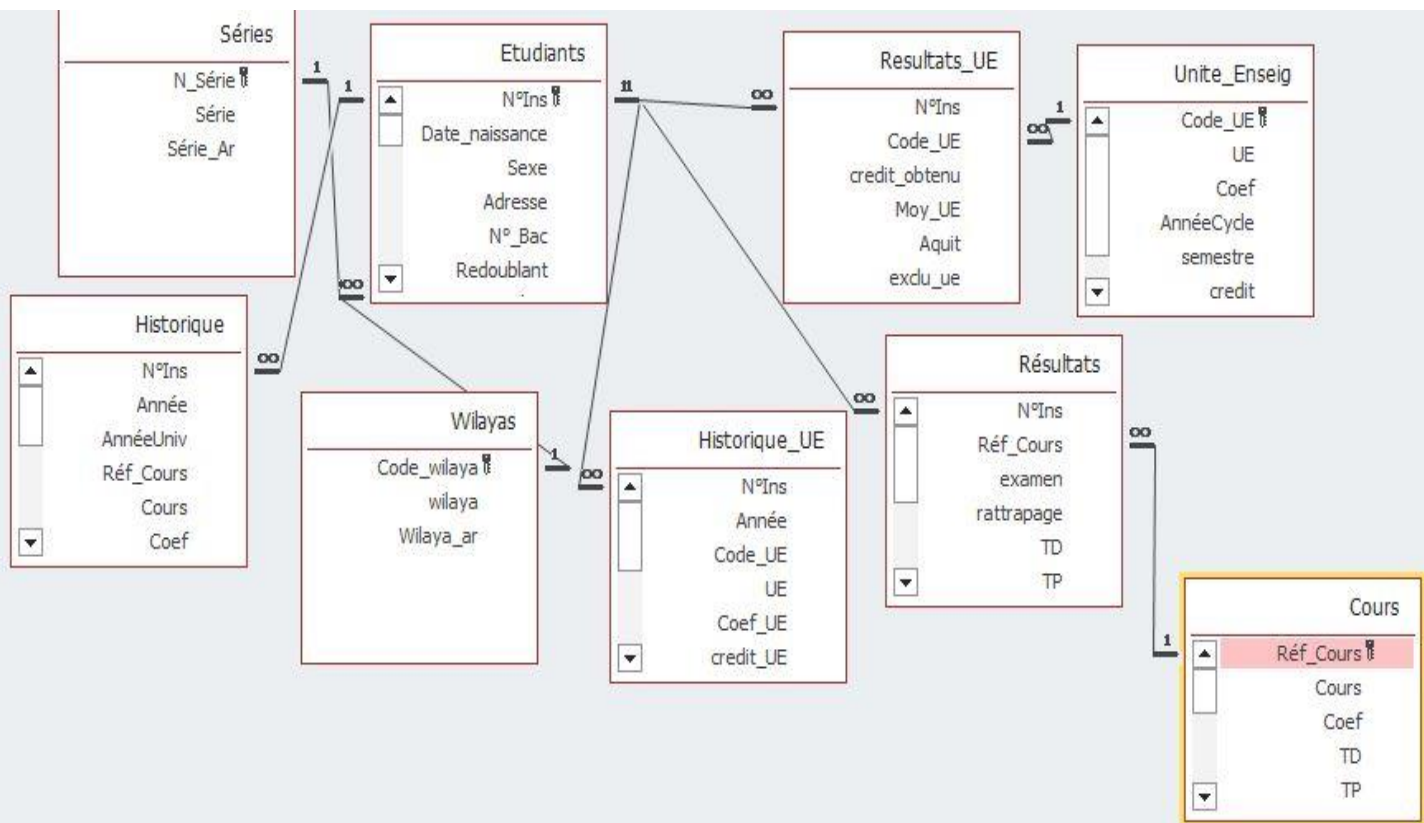
Pour obtenir de meilleurs résultats lors du nettoyage de la base de données à l'aide d'Excel, il est possible d'utiliser les outils pris en charge par Microsoft, tels que Microsoft Power Query, qui permet de nettoyer les données plus rapidement et plus efficacement.

En conclusion, Excel peut être utilisé comme un outil efficace et flexible pour nettoyer une base de données, mais l'utilisateur doit être qualifié et connaître les compétences nécessaires pour le faire correctement.



6-Relationships

L'image suivante représente les relations entre la table après le nettoyage de la base de données



7-conclusion

Le nettoyage des données dans un data warehouse est crucial pour garantir la qualité et la fiabilité des informations utilisées pour prendre des décisions commerciales. Cela peut inclure la suppression des doublons, la correction des erreurs, la standardisation des formats et l'enrichissement des données. Le nettoyage régulier des données dans un data warehouse peut améliorer les performances et la pertinence des analyses, ce qui peut conduire à des décisions plus informées et à un meilleur retour sur investissement.