

La documentation détaillée pour les fichiers Jupyter Notebook pour la gestion des tables dans la base de données Azure.

Voici un aperçu rapide de ce que chaque fichier fait :

1. **Création_BDD.ipynb** : Ce notebook est responsable de la création de la base de données. Il commence par charger les variables d'environnement nécessaires pour se connecter à la base de données. Ensuite, il établit une connexion à la base de données et crée plusieurs tables, y compris Produit, Caracteristiques, Dimensions et USER_API.
2. **Nettoyage_BDD.ipynb** : Ce notebook est utilisé pour nettoyer la base de données. Il effectue plusieurs tâches de nettoyage, notamment la mise à jour des URL de redirection, la suppression des doublons, la mise à jour des prix en fonction de l'URL, et la mise à jour de la colonne Marque
3. **CRUD.ipynb** : Ce notebook contient des opérations CRUD (Create, Read, Update, Delete) sur la base de données. Il crée de nouvelles entrées dans les tables Produit, Caracteristiques et Dimensions, lit les données de la base de données, met à jour le prix et le bruit dans les tables Produit et Caracteristiques, et supprime les produits et les caractéristiques en fonction du prix et du bruit.
4. **RGPD.ipynb** : Ce notebook est utilisé pour gérer les données des utilisateurs conformément au Règlement général sur la protection des données (RGPD). Il supprime les utilisateurs qui ne se sont pas connectés depuis 24 mois, les utilisateurs qui ont créé un compte mais ne se sont jamais connectés, et les utilisateurs qui souhaitent supprimer leur compte.

Pour chaque fichier, je vais maintenant détailler les dépendances, les commandes, les enchaînements logiques de l'algorithme, et les choix de nettoyage et d'homogénéisation des formats de données.

1. Création_BDD.ipynb

Dépendances

- os
- dotenv
- pyodbc

Commandes

- Charger les variables d'environnement
- Établir une connexion à la base de données
- Créer les tables **Produit**, **Caracteristiques**, **Dimensions** et **USER_API**

Enchaînements logiques de l'algorithme

1. Charger les variables d'environnement
2. Établir une connexion à la base de données
3. Créer les tables **Produit**, **Caracteristiques**, **Dimensions** et **USER_API**

Choix de nettoyage et d'homogénéisation des formats de données

- Il y a pas de nettoyage dans ce fichier il ne fait que crée des tables dans la basse de données Azure

2. Nettoyage_BDD.ipynb

Dépendances

- os
- dotenv
- pyodbc
- requests
- lxml

Commandes

- Mise à jour des URL de redirection
- Suppression des doublons
- Mise à jour des prix en fonction de l'URL
- Mise à jour de la colonne Marque

Enchaînements logiques de l'algorithme

1. Charger les variables d'environnement
2. Établir une connexion à la base de données
3. Effectuer les tâches de nettoyage

Choix de nettoyage et d'homogénéisation des formats de données

- Les URL de redirection sont mises à jour pour s'assurer que toutes les URL sont valides.
- Les doublons sont supprimés pour éviter la redondance des données.
- Les prix sont mis à jour en fonction de l'URL pour s'assurer que les prix sont à jour.
- La colonne Marque est mise à jour pour s'assurer que toutes les entrées ont une marque.

3. CRUD.ipynb

Dépendances

- os
- dotenv
- pyodbc

Commandes

- Création de nouvelles entrées dans les tables Produit, Caracteristiques et Dimensions
- Lecture des données de la base de données
- Mise à jour du prix et du bruit dans les tables Produit et Caracteristiques
- Suppression des produits et des caractéristiques en fonction du prix et du bruit

Enchaînements logiques de l'algorithme

1. Charger les variables d'environnement
2. Établir une connexion à la base de données
3. Effectuer les opérations CRUD

Choix de nettoyage et d'homogénéisation des formats de données

- Les nouvelles entrées sont créées avec des types de données spécifiques pour chaque colonne pour assurer l'homogénéité des données.
- Les prix et les bruits sont mis à jour pour s'assurer que les données sont à jour.

Les produits et les caractéristiques sont supprimés en fonction du prix et du bruit pour maintenir la pertinence des données. Ce fichier est crée pour tester plusieurs scenario possible pour tester le CRUD (Create, Read, Update, Delete) sur la base de données. Il y a du code python pour tester dans l'éditeur de code ou directement en SQL dans le portail Azure le code est doublé pour tester les 2 cas

4. RGPD.ipynb

Dépendances

- os
- dotenv
- pyodbc
- datetime

Commandes

- Suppression des utilisateurs qui ne se sont pas connectés depuis 24 mois
- Suppression des utilisateurs qui ont créé un compte mais ne se sont jamais connectés
- Suppression des utilisateurs qui souhaitent supprimer leur compte

Enchaînements logiques de l'algorithme

1. Charger les variables d'environnement
2. Établir une connexion à la base de données
3. Effectuer les tâches de suppression des utilisateurs

Choix de nettoyage et d'homogénéisation des formats de données

- Les utilisateurs qui ne se sont pas connectés depuis 24 mois sont supprimés pour respecter le RGPD.
- Les utilisateurs qui ont créé un compte mais ne se sont jamais connectés sont supprimés pour maintenir la pertinence des données.
- Les utilisateurs qui souhaitent supprimer leur compte sont supprimés pour respecter le RGPD.

Ce fichier a été créé pour exécuter avec rapidité des demandes d'utilisateur qui souhaite supprimer leur compte, mais aussi pour supprimer de la base de données Azure les personnes qui ne se sont pas connectés depuis 24 mois en adéquation avec les règles établies pour la gestion des utilisateurs de l'API