

Projet : Pipeline ETL & Dashboard de Comptes Clients

Objectif général

Mettre en place un processus ETL complet (Extraction – Transformation – Loading) à partir de fichiers Excel, puis produire un dashboard (PDF, ou Excel, ou Tableau, lookerstudio) présentant différents indicateurs sur les comptes clients de la microfinance.

Les étudiants doivent :

1. **Lire** les données sources (Excel)
 2. **Transformer** les données pour les rendre exploitables
 3. **Charger** les données finales dans un Data Warehouse (SQLite, PostgreSQL, ou autre...)
 4. **Construire un dashboard** inspiré de *dashboardCompte.PNG*
 5. **Proposer éventuellement des indicateurs ou graphiques supplémentaires**
-

1. Description des données

Les données se trouvent dans un fichier Excel (dataset_account.xlsx) contenant notamment :

Feuille 1 : data source

Colonnes disponibles :

- **Code** : Identifiant client
 - **noCompte** : Numéro de compte
 - **AvailableBalance** : Solde disponible du client
 - **OpeningDate** : Date d'ouverture du compte
 - **AccountStatus** : Statut du compte (utiliser uniquement les comptes "Actif")
 - **Report_date_to** : Date de situation du compte
 - **Gestionnaire** : Gestionnaire de compte
-

Composition du numéro de compte

Le champ noCompte est structuré comme suit :

Segment	Signification	Exemple
00115	Code banque	00115
00001	Code agence	00001
9...	Produit (1er digit) + numéro de compte	9000031 6601
39	Clé RIB	39

Exemple complet : **00115 00001 90000316601 39** (9 : code produit ; 00001 : code agence)

Feuille 2 : Product Mapping

Contient la correspondance entre :

- **Code produit** (ex : 9)
- **Nom du produit** (ex : “Dépôt à Terme”)

Feuille 3 : Branch Mapping

Contient la correspondance entre :

- **Code agence** (ex : 00001)
- **Nom de l’agence** (ex : “Champagne-Ardenne”)

. Travail demandé (Pipeline ETL)

Étape 1 – Extraction

Importer les données depuis :

- la feuille data source
- la feuille Product Mapping
- la feuille Branch Mapping

Étape 2 – Transformations

Appliquer les opérations suivantes :

1. **Décomposer le numéro de compte (noCompte)**

- Extraire :
 - Code banque
 - Code agence
 - Code produit (1er digit de la 3e section)
 - Numéro de compte proprement dit
 - Rapprocher les codes avec :
 - Mapping des produits
 - Mapping des agences
 - ➔ Ici vous avez 2 options :
 - Faire les « join » dans l'ETL (plus rapide côté rendu vers le dashboard).
- Join avec panda dataframe:

```
In: Df_merged = d.merge (df_left, df_right, on="ici_le_PK_de_jointure", how='left')
#how -> (left, right, inner, outer) # **mais attention!
```

- Faire les “join “ dans le Data warehouse (plus souples pour les rajouts ou modification de nom d'agence, ou utilisation par d'autre « fact ») **

**** Souvenez-vous : la bonne pratique pour mettre en place les tables dans les data warehouse c'est d'avoir des tables des dimensions (Agence, produit, gestionnaire etc....) et des tables des faits (exemple : les situations de chaque compte à date avec les clés primaires des dimensions)**

2. **Filtrer uniquement les comptes actifs**
Garder seulement AccountStatus = "Actif".
3. **Formatage des dates, montants, etc.**
4. **Création d'un dataset final**
Harmonisé, enrichi, prêt pour la visualisation et agrégation.

Étape 3 – Agrégation

Préparer les indicateurs nécessaires au dashboard, par exemple :

- Répartition des comptes par agence
- Solde total disponible par agence

- Solde moyen par produit
- Top N gestionnaires par solde total
- Total de comptes actifs

Ou tous autres indicateurs que vous jugez import

Étape 4 – Chargement

Charger les données finales dans un Data Warehouse :

- SQL Server
- PostgreSQL
- MySQL
- SQLite
- ou autre base relationnelle

L'objectif ici c'est de représenter un Data warehouse et non d'en avoir un vraie (donc à vous d'en choisir)

Assurez-vous de fournir l'accès à la base

(connection string / privilèges).

3. Dashboard à produire

Créer un dashboard inspiré du fichier **dashboardCompte.PNG**.

Vous pouvez utiliser :

- Power BI
- Tableau
- Metabase
- Grafana
- Dash / Streamlit
- ou un autre outil BI de votre choix (même MS Excel)

Le dashboard doit présenter :

- Au moins **3 indicateurs clés**
- Au moins **4 graphiques**

Liberté accordée :

- Vous n'êtes pas obligés de reproduire exactement les mêmes graphiques.
 - Vous pouvez ajouter des visuels que vous jugez pertinents.
 - Design et lisibilité comptent plus que la technologie ou plateforme choisie.
-

4. Livrables

Chaque étudiant doit fournir :

1. **Code ETL** (Python, SQL)
2. **Schéma du pipeline ETL**
3. **Dashboard final**
4. **Courte documentation** expliquant :
 - Le pipeline ETL
 - Les transformations choisies
 - Les indicateurs du dashboard

Format des livrables : à envoyer par mail à razaonja@gmail.com en respectant les points ci-après

- Nom de fichier : ASJA -L3_ETL_PROJECT-MatriculeXXXX-2025_2026.zip
- Object du mail : ASJA -L3 ETL PROJECT MatriculeXXXX 2025-2026

Ex :

- Filename : ASJA-L2-ETL_PROJECT-Matricule5984-2025_20263.zip
 - Mail Object : ASJA -L3 ETL PROJECT Matricule5984 2025-2026
-

Objectif pédagogique

À la fin du projet, les compétences que vous devriez acquérir sont:

- Lire et transformer un dataset réel
- Manipuler des Jointures
- Concevoir un pipeline ETL complet
- Documenter un processus analytique

- Produire un dashboard professionnel