|  |  |  |
| --- | --- | --- |
| univlyon2_logo201806-standard | **Master BI&A**  **Projet Tutoré - DG**  **Fadila Bentayeb**  **fadila.bentayeb@univ-lyon2.fr** | logo-ICOM_Coul-300x72 |

Projet

**Projet Tutoré – Gouvernance des Données**

**Groupes de 4 étudiants+**

**Contexte**

L’entreprise fictive **DataShop** souhaite instaurer un cadre de **gouvernance des données** pour améliorer la qualité, la sécurité et la valorisation de ses données clients, produits et commandes. Les données présentent actuellement des **doublons, incohérences, valeurs manquantes et anomalies**, ce qui complique leur exploitation fiable pour la prise de décision.

**Objectifs pédagogiques**

À la fin du projet, vous serez capables de :

1. Identifier et mesurer les problèmes de **qualité des données**.
2. Mettre en place un **cadre de gouvernance adapté** (documentation, rôles, règles).
3. Construire un **dictionnaire de données** et un **glossaire métier**.
4. Définir des **politiques de gouvernance** (qualité, sécurité, RGPD).
5. Créer un **tableau de bord interactif** de suivi qualité.
6. 33Expérimenter un **outil de data catalog open-source**.

**Ressource fournie**

* Le fichier Excel **projet\_data\_gouvernance\_large.xlsx** (Clients, Produits, Commandes).

**Étapes et répartition du temps**

**Étape 1 : Analyse initiale des données (6–8h)**

* Exploration des tables.
* Identification des anomalies (doublons, emails invalides, valeurs manquantes).
* Définition de 5 à 8 **indicateurs de qualité** (complétude, cohérence, unicité…).

**Étape 2 : Documentation & gouvernance (10–12h)**

* Construction d’un **dictionnaire de données** (noms, types, règles de validation).
* Élaboration d’un **glossaire métier**.
* Définition des **rôles et responsabilités** (Data Owner, Steward, Custodian…).

6

**Étape 3 : Gouvernance opérationnelle (10–12h)**

* Proposition de **règles de correction / nettoyage** (formatage des emails, gestion des doublons).
* Rédaction d’un **mini-cadre de gouvernance** (politiques d’accès, confidentialité, RGPD).
* Simulation d’un processus de validation / workflow.

**Étape 4 : Restitution et pilotage (10–12h)**

* Conception d’un **tableau de bord interactif** (Power BI, Tableau, Dash…) affichant :
  + taux de doublons,
  + taux de complétude,
  + validité des emails,
  + répartition des statuts commandes, etc.
* Rédaction d’un **rapport final (15–20 pages)**.
* **Présentation orale (12 min)** avec recommandations business.+

**Étape 5 (Optionnelle si temps dispo) : Exploration d’outil de Data Catalog (4–6h)**

* Tester un outil open-source (Amundsen, Apache Atlas, Talend Data Catalog).
* Simuler l’intégration du dictionnaire/glossaire.

**Livrables attendus**

1. **Rapport final** (15–20 pages) : analyse, dictionnaire, glossaire, cadre de gouvernance.
2. **Tableau de bord interactif** (fichier + captures d’écran).
3. **Présentation orale (12 min)** par groupe.
4. **Prototype de data catalog**.

**Évaluation**

* Analyse et indicateurs qualité : **20%**
* Dictionnaire + glossaire : **20%**
* Cadre de gouvernance proposé : **20%**
* Tableau de bord : **20%**
* Rapport + soutenance : **20%**

**Prototype possible de Data Governance que les étudiants peuvent produire :**

1. **Mini-data catalog** : table Excel ou outil simple listant toutes les tables et champs, avec métadonnées, Data Owner, règles de qualité.
2. **Tableau de bord qualité des données** : Power BI / Tableau / Excel avec indicateurs de qualité.
3. **Workflow de validation** : diagramme montrant comment les données sont contrôlées et corrigées avant exploitation.
4. **Documentation gouvernance** : dictionnaire, glossaire et règles de gouvernance centralisés dans un document unique.