

|  |
| --- |
| Université Savoie Mont Blanc |
| Suivi de projet |
|  |

|  |
| --- |
|  |

**Année Universitaire 2024 – 2025**

**Réalisé Par :**

* BEPTEMPS Quentin
* EL ABLAK Sammy
* PERROUX Valentin
* TASDEMIR Sefer
* TINASTEPE FEYZA

**Encadré Par :**

* M. V. Couturier
* M. C. Lin-Kwong-Chon



**Sommaire**

[Semaine 11 2](#_Toc195088110)

[Semaine 12 3](#_Toc195088111)

[Semaine 13 4](#_Toc195088112)

[Semaine 14 5](#_Toc195088113)

[Semaine 15 6](#_Toc195088114)

Semaine 11

**Objectifs principaux de la semaine :**

1. Prendre connaissance de la base existante (fichier CSV de 326,6 Mo issu d’Open Food Facts).
2. Analyser la structure actuelle et identifier les premiers points d’amélioration (qualité, performance, sécurité).
3. Mettre en place l’environnement de travail (PostgreSQL, Azure si nécessaire, Power BI, Knime).
4. Organiser l’équipe et détailler les rôles de chacun selon les tâches à venir.

**Déroulé détaillé :**

* **Analyse de la base existante (OLTP initial).**
  + Nous avons exploré le fichier CSV fourni par Open Food Facts afin de comprendre les champs disponibles (informations nutritionnelles, ingrédients, labels, pays de vente, etc.).
  + Nous avons observé l’hétérogénéité des données (valeurs manquantes, doublons résiduels, différences de format).
* **Organisation du projet.**
  + Répartition des rôles dans l’équipe (référent base de données PostgreSQL, référent Azure, référent Power BI, référent IA/Knime, référent gestion globale du projet)
* **Installation et configuration des environnements.**
  + Vérification de PostgreSQL en local
  + Préparation d’un compte de test sur Azure (base PostgreSQL hébergée) pour prévoir le futur déploiement.
* **Tâches complémentaires & décisions :**
  + Rédaction d’un petit document interne pour lister toutes les colonnes du CSV et décider lesquelles sont critiques et lesquelles sont optionnelles.
  + Préparation d’un plan de test pour l’importation des données (extraits de 10 000 lignes, par exemple) avant de procéder à l’intégration complète.

**Bilan et points d’amélioration à la fin de la Semaine 11 :**

* Nous avons une vision claire du contenu de la base de données existante.
* L’équipe est structurée, les outils sont installés.
* Les premiers constats sur la qualité (valeurs manquantes, doublons, colonnes obsolètes) sont répertoriés.
* Nous avons formalisé la liste des actions à mener pour la suite (conception du modèle relationnel OLTP).

Semaine 12

**Objectifs principaux de la semaine :**

1. Concevoir le modèle relationnel OLTP optimisé à partir du CSV (limiter la redondance, améliorer la structure).
2. Débuter l’implémentation (script SQL pour PostgreSQL).
3. Planifier l’insertion des données et simuler les volumes de ventes (loi normale).

**Déroulé détaillé :**

* **Conception du modèle relationnel OLTP.**
  + Après avoir listé les attributs du CSV, nous avons identifié les entités principales (ex. : Produit, Catégorie, Marque, Distributeur, etc.).
  + Nous avons conçu un **MCD** puis un **MLD** correspondant.
  + Nous avons tenu compte des nouveaux besoins d’OFF :
    - Conserver les informations nutritionnelles clés
* **Optimisations retenues :**
  + Utilisation de types de données appropriés
  + Mise en place de **clés étrangères** pour relier la table des produits à celle des ventes (relation 1-n).
  + Préparation de futurs **index** sur les champs fréquemment interrogés (par exemple la clé produit, le distributeur, et la période de vente).
* **Élaboration du script SQL.**
  + Rédaction des CREATE TABLE en veillant à appliquer les bonnes contraintes
  + Premiers tests de création de la base en local sur PostgreSQL.
* **Simulations des chiffres de vente (loi normale).**
  + Choix de la bibliothèque Python ou de Knime pour générer aléatoirement ~100 données de ventes par produit sur trois ans.
  + Sélection d’une **loi normale** adaptée (moyenne et écart-type variant selon la popularité du produit).
* **Points de coordination :**
  + Nous avons revu la dépendance entre la conception OLTP et la future base OLAP : assurer qu’il sera facile d’extraire les ventes vers une table de faits OLAP.

**Bilan de la Semaine 12 :**

* Modèle relationnel OLTP finalisé et validé en interne.
* Script SQL en version bêta pour la création des tables.
* Début de la génération des ventes simulées.

Semaine 13

**Objectifs principaux de la semaine :**

1. Déployer la base OLTP finale sur un serveur PostgreSQL (local et/ou Azure).
2. Vérifier la qualité et la performance via des tests d’insertion et de requêtes.
3. Démarrer les travaux de **reporting (**Power BI) et de **workflows** (Knime).

**Déroulé détaillé :**

* **Déploiement de la base sur Azure et insertion des données.**
  + Nous avons configuré un serveur PostgreSQL sur Azure (option “Flexible Azure Database for PostgreSQL”).
  + Exécution des scripts de création de tables (OLTP).
  + Insertion du jeu de données principal (avec un échantillon pour tester avant de charger l’intégralité).
  + Vérification de la cohérence et de la rapidité (temps d’insertion, indexes, logs d’erreur).
* **Optimisation de la base PostgreSQL (en local puis sur Azure).**
  + Mise en place d’**index** sur les champs de jointure les plus utilisés.
  + Vérification du plan d’exécution de certaines requêtes critiques (ex. : recherche de produit par catégorie, agrégations sur les ventes).
* **Démarrage des workflows Knime.**
  + Création d’un premier workflow pour effectuer un **contrôle de qualité** des données (détection de valeurs aberrantes, valeurs manquantes).
  + Début de l’exploration des données (statistiques descriptives) pour préparer l’entraînement des futurs modèles IA.
* **Premiers rapports Power BI (en anglais).**
  + Connexion de Power BI au serveur Azure PostgreSQL.
  + Mise en place d’un **schéma en étoile** minimal (ou du moins, d’un schéma logique) pour faciliter la création des visuels.
  + Premier tableau de bord : répartition des produits par pays, distribution du Nutri-Score, volumes de vente simulés.
* **Points de coordination et problèmes résolus :**
  + Amélioration de la charte de nommage des tables pour uniformiser le modèle (harmoniser majuscules/minuscules, etc.).
  + Gestion des problèmes de ports pour se connecter au serveur Azure depuis l’IUT.

**Bilan de la Semaine 13 :**

* Base OLTP fonctionnelle à la fois en local et sur Azure.
* Données insérées sans perte majeure, l’équipe a corrigé quelques erreurs de type ou d’encodage.
* Début de la création des premiers rapports Power BI et des workflows Knime.

Semaine 14

**Objectifs principaux de la semaine :**

1. Finaliser les rapports Power BI pour la base OLTP.
2. Concevoir puis déployer le modèle OLAP (base en étoile) si nécessaire pour des analyses plus poussées.
3. Mettre en place et entraîner les modèles IA dans Knime.

**Déroulé détaillé :**

* **Création avancée des rapports Power BI (modèle OLTP).**
  + Élargissement des visualisations
  + Analyse de performance
* **Conception du modèle relationnel OLAP (étoile).**
  + Identification des **tables de faits et des tables de dimensions**
  + Mise en place d’une base dédiée (OLAP) sur PostgreSQL ou sur Azure.
  + Tests d’alimentation (ETL) depuis la base OLTP vers la base OLAP (scripts SQL + Knime).
  + Validation de la cohérence (intégrité référentielle, granularité des faits).
* **Rapports Power BI sur la base OLAP.**
  + Connexion de Power BI à la base OLAP.
  + Création de mesures spécifiques (par ex. : total des ventes, moyenne par segment de produits, analyses par année).
  + Comparaison de la fluidité et de la réactivité par rapport au modèle OLTP.
* **Mise en place des modèles IA sous Knime.**
  + **Nutri-Score automatique** :
    - Séparation du jeu de données
    - Entraînement de plusieurs algorithmes (Random Forest, Logistic Regression, etc.) pour prédire le Nutri-Score.
    - Évaluation des performances (précision, F1-score).
  + **Open Fact Food score** :
    - Test de méthodes de clustering ou de classification supervisée pour proposer une note en fonction des variables disponibles.
* **Problèmes rencontrés / résolus :**
  + Mise en place des règles de nettoyage (absence de valeurs de sodium/fibres) : comment intégrer ces données dans les IA ?
  + Utilisation de proxies ou de paramétrages réseau pour accéder à Azure depuis Knime.

**Bilan de la Semaine 14 :**

* Premiers rapports Power BI finalisés sur l’OLTP, et migration progressive vers OLAP.
* Base OLAP opérationnelle, ETL en place
* IA : prototypes fonctionnels pour le Nutri-Score et le OFF Score

Semaine 15

**Objectifs principaux de la semaine :**

1. Finaliser et optimiser les modèles IA (Nutri-Score et OFF Score).
2. Rédiger tous les livrables (document technique, scripts, rapports, etc.).
3. Préparer la soutenance finale (présentation en français et conclusion en anglais).

**Déroulé détaillé :**

* **Optimisation et validation des modèles IA.**
  + Ajustement des paramètres
  + Sélection du ou des meilleurs algorithmes selon la précision
  + Documentation claire du workflow Knime (nœuds, paramètres, variables d’entrée et sortie).
* **Rédaction des livrables.**
  + **Document technique** :
    - Modèle OLTP (MCD, MLD, script SQL).
    - Modèle OLAP (schéma en étoile, justification technique).
    - Détails sur l’insertion des données, la gestion des ventes simulées.
    - Démarche IA (données d’apprentissage, choix des modèles, performance).
    - Compte-rendu sur la sécurisation de la base, éventuelles failles corrigées.
  + **Rapports Power BI** : finalisation, mise en forme
  + **Workflows Knime** : remise de tous les fichiers .knwf
  + **Résumé (abstract) d’une demi-page en anglais** présentant la problématique et les principaux résultats.
* **Préparation de la soutenance (finale le 11 avril).**
  + Répartition du temps de parole entre les membres de l’équipe.
  + Création du diaporama avec un fil conducteur :
    - Architecture globale (OLTP / OLAP / ETL / IA),
    - Détails sur la base OLTP (modèle, intégration),
    - Rapports Power BI (démos),
    - IA (modèles, résultats),
    - Conclusion en anglais.
  + Répétition générale pour respecter le temps imparti

**Bilan final :**

* Les modèles IA sont prêts, avec des performances satisfaisantes et une consommation de ressources maîtrisée.
* Tous les livrables sont rédigés et relus (document technique, scripts, rapports, abstract en anglais).
* L’équipe est prête pour la soutenance du 11 avril, avec un support de présentation validé