

Projet 2

Contexte

Laplace Immo est un réseau national d'agences immobilières en pleine transformation digitale. Le directeur général, convaincu de l'importance des données pour améliorer la compétitivité, a initié le projet stratégique **DATAImmo** pour exploiter les données publiques et internes afin de mieux comprendre les dynamiques du marché immobilier et accompagner les clients dans leurs décisions.

Clara Daucourt, la CTO de l'entreprise, pilote le projet en étroite collaboration avec le service R&D dirigé par *Cécile Morel*, qui se concentre sur l'innovation dans l'analyse et la visualisation des données.

En tant que data analyst, vous jouez un rôle clé dans la conception, l'implémentation et l'interprétation des analyses. Vous devez notamment modifier la base de données permettant de collecter les transactions immobilières et foncières en France. Vous utiliserez ensuite cette base pour analyser le marché et aider les différentes agences régionales à mieux accompagner leurs clients. Votre travail servira non seulement à créer des outils décisionnels internes mais aussi à prouver la valeur ajoutée des solutions technologiques comme KNIME et Streamlit.

Le projet est structuré en plusieurs étapes, avec des livrables attendus à chaque phase, validés par Clara et Cécile.

Étape 1

À votre arrivée ce jeudi matin, vous avez reçu un e-mail de la part de Clara, qui donne plus de détails sur ce qui est attendu de votre part.

Objet : Réunion de validation – Modification de la base de données.

De : Clara

À : Moi

Hello,

Afin d'avancer sur le projet DATAImmo, je prévois une première réunion pour valider la modification de la base de données.

Tu trouveras en pièce jointe un fichier zip avec les données suivantes :

- des données extraites du site open data des Demandes de valeurs foncières (DVF) ;
- des données de l'INSEE avec les résultats des recensements de la population ;
- des données de data.gouv sur les régions, avec le référentiel géographique français, communes, unités urbaines, aires urbaines, départements, académies, régions.

D'ici la réunion, j'aurais besoin de 2 choses de ta part :

Il faut que tu prépares le dictionnaire des données en respectant le Template en PJ pour répertorier et décrire les données importantes à stocker car nous avons omis de le faire. Il faut le remplir pour les trois fichiers de données.

Ensuite, peux-tu modifier le schéma relationnel en pièce-jointe pour qu'il prenne en compte les nouvelles données région et population ? Ça nous permettra de bien visualiser les différentes entités, associations et cardinalités de la base de données. Enfin, il faudra que tu me présentes ce nouveau schéma relationnel normalisé (Il doit suivre la norme 3NF) de la base de données qui donnera lieu à la création des nouvelles tables.

On validera tout ça ensemble lors de notre réunion de validation, pour que tu puisses avancer sur la création de la base.

Je reste à ta disposition en cas de besoin.

Bien à toi,

Clara Daucourt, CTO

Pièces jointes (disponible sur le drive) :

- Template du dictionnaire des données
- Les données
- Le schéma relationnel à recréer et modifier

Étape 2

Deux semaines plus tard

Une fois cette première partie de votre travail effectuée, vous présentez votre travail à Clara qui valide votre schéma relationnel ainsi que le dictionnaire des données.

À ce stade du projet, vous pouvez valider avec les profs que votre conception est correcte avant d'aller plus loin.

Félicitations ! Elle vous envoie un e-mail à la suite de cette réunion.

Objet: Validation du schéma relationnel et suite du projet

De : Clara

À : Moi

Hello,

Félicitations pour cette première étape de conception ! Comme on en a parlé en réunion ce matin, on est bons pour partir sur la modification de la base de données. Il faut maintenant que tu implémentes les tables dans la base de données en respectant ce qu'on s'est dit ce matin (cf. compte rendu de réunion en pièce jointe).

On a échangé en interne, et on pense qu'un outil comme MySQL est pertinent pour ce type d'implémentation. Si tu es plus à l'aise, tu peux essayer d'implémenter un outil comme SQLite ou PostgreSQL.

Une fois que tu auras fait ça, tu pourras faire les requêtes pour extraire les données dont nous avons besoin. Pour ça, envoie-moi un document PDF avec tous les résultats des requêtes ainsi que les requêtes associées. Utilise des alias pour que ça soit plus lisible. Tu trouveras toutes les demandes auxquelles il faudra répondre dans le CR de réunion.

Une fois que tu auras terminé, on regardera tout ça ensemble pour voir ce que tu as réussi à tirer comme données.

Bon courage !

Clara Daucourt

CTO

Pièce jointe :

- Compte-rendu de réunion

Étape 3

1 mois plus tard

Cécile la responsable RD de l'entreprise explore de nouvelles pratiques professionnelles pour l'interprétation, la visualisation et l'analyse des données. Elle vous demande de l'aide :

Objet : Nouvelles pistes à explorer

De : Cécile

À : Moi

Bonjour,

J'aurais besoin de ton aide pour explorer de nouvelles pistes pour traiter nos données. Dans les prochains investissement de l'entreprise, nous nous posons la question suivante :

Faut-il acheter un logiciel d'analyse de données comme Tableau ou Power BI ?

J'ai besoin de ton aide pour explorer cette pise. Pour cela, je t'invite à prendre le fichier Excel d'origine

Est-ce que tu peux avec l'aide des quatre types d'analyses possibles :

- **L'analyse descriptive** : répond à la question « que s'est-il passé ? »
- **L'analyse diagnostique** : donne des réponses à la question « pourquoi cet événement a-t-il eu lieu ? »
- **L'analyse prédictive** : cherche à prévoir ce qui est susceptible de se produire
- **L'analyse prescriptive** : répond à l'interrogation « quelle action devrait-on mettre en place ? »

Est-ce que tu peux nous produire des rapports, documents, tableaux, schémas et graphiques afin de souligner les corrélations, les relations, les tendances, ... relevées à partir des données.

Essaye d'utiliser les techniques de datavisualisation pour rendre les données statiques compréhensibles à tous, y compris aux non spécialistes

Le nombre de graphique / tableaux de bord n'est pas important, le plus important est de répondre à la question : quel est l'apport de ce dispositifs en terme d'analyse

Étape 4

Exploration de nouvelles pistes

Cécile est très satisfaite de votre travail et vous sollicite à nouveau pour mener de nouvelles recherches et expérimentations. A votre arrivée ce vendredi, vous trouverez l'email suivant de Cécile.

Objet : Exploration de nouvelles pistes avec KNIME et Streamlit

De : Cécile Morel, Responsable R&D

À : [Votre nom]

Bonjour,

Le projet DATAImmo avance bien, et je souhaiterais explorer de nouvelles pistes d'analyse des données. Voici ce que j'aimerais que tu fasses :

- **KNIME** : Construis un pipeline avec KNIME pour automatiser l'intégration, la transformation et l'analyse des données. Fais-nous des propositions sur les analyses prédictives possibles (modélisation des prix immobiliers).

- **Streamlit** : Développe un prototype d'application pour visualiser les données et rendre les résultats accessibles aux parties prenantes. Pense à intégrer des fonctionnalités interactives pour simuler des prix immobiliers.

Ces outils pourraient devenir des solutions pérennes pour nos équipes, alors concentre-toi sur leur potentiel et leur apport en termes de productivité et d'accessibilité.

Pour t'aider, j'ai fait appel à un consultant data qui a produit le document joint (knime-streamlit) avec des pistes possibles à traiter

Merci de préparer des livrables détaillés pour que nous puissions les présenter au Directeur Général.

À bientôt,

Cécile Morel, Responsable R&D

Livrables

1. Étape 1 :

- a. Dictionnaire des données.
- b. Schéma relationnel normalisé modifié.

2. Étape 2 :

- a. Fichier zip avec captures d'écran des 4 tables de la base de données opérationnelle, qui incluent le nombre de lignes chargées dans la base :
 - Bien
 - Vente
 - Commune
 - Région
- b. Document avec les requêtes et les résultats (PDF).

3. Étape 3 :

- a. Ficher tableau ou Power BI

4. Étape 4 :

- a. Fichier Knime
- b. Projet Streamlit

Pour faciliter votre passage devant le jury, déposez sur le drive (dossier Rendu), dans un dossier zip nommé « **Titre_du_projet_num_groupe** », tous les livrables du projet comme suit : Num_Groupe_n° du livrable_nom du livrable_date de démarrage du projet. Cela donnera :

- Num_Groupe_1_dictionnaire_de donnees_mmaaaa
- Num_Groupe_2_schéma_mmaaaa
- Num_Groupe_3_bdd_mmaaaa
- Num_Groupe_4_requetes_mmaaaa
- ...

Soutenance

Pendant la soutenance, le jury jouera le rôle de Clara, la CTO, à qui vous présentez votre travail. Vous pouvez créer un support de présentation mais ce n'est pas obligatoire :

- **Présentation (15 minutes)**
 - Vous expliquerez la démarche de la création de la base de données.
 - Vous présenterez la base de données avec les données importées directement dans l'outil SGBD en live (en réalisant une démonstration, sans vous appuyer sur des captures d'écran).
 - Vous répondrez aux analyses demandées par le Directeur Général.
 - Vous exposerez votre choix de l'outil de datavisualisation
 - Vous présenterez vos réalisations Knime et Streamlit
- **Discussion (10 minutes)**
 - Le jury, jouant le rôle de Clara, vous challengera sur vos choix.
 - À la fin de la soutenance, le jury arrêtera de jouer le rôle de Clara pour vous permettre de débriefer ensemble.

Votre présentation devrait durer 15 minutes (+/- 5 minutes). Puisque le respect des durées des présentations est important en milieu professionnel, **les présentations en dessous de 10 minutes ou au-dessus de 20 minutes peuvent être refusées.**