**DEADLINE : Vendredi 02 Décembre**

Après plusieurs années passées dans la grande chaine de librairie concurrente d’Amazon, vous décidez de changer d’entreprise et de relever de nouveaux défis. Votre choix se porte sur une grande entreprise spécialisée dans l’immobilier qui possède également un grand réseau d’agences en France. Avant de partir, vous marquez le coup avec un superbe pot de départ : vos collègues, la super ambiance et les infrastructures magnifiques vont vous manquer !!

A peine arrivé et une fois le tour de votre nouvelle entreprise effectuée, votre directrice technique Sophie vous positionne sur un nouveau Projet. Il s’agit d’un projet central car l’entreprise souhaite devancer la concurrence en créant une base de données immobilières pour faire des études de marché, implémenter un modèle ou des modèles de prédiction de prix immobilier et détecter de futures positions stratégiques pour leurs clients en se basant sur les données récoltées, ect…Voici donc les milestone des missions confiées par votre directrice technique Sophie lors de votre première réunion :



**Milestone 1 ( Etape 1 )  : Conception de la base de données**

1. Préparer le dictionnaire des données en respectant le Template du dictionnaire de données pour répertorier et décrire les données importantes à stocker. Les données extraites du site open data des [Demandes de valeurs foncières (DVF)](https://www.data.gouv.fr/fr/datasets/5c4ae55a634f4117716d5656/) sont également fournies
2. Ensuite, proposer un modèle conceptuel des données. Ça permettra de bien visualiser les différentes entités, associations et cardinalités de la base de données. Pour le formalisme, il faut le faire en UML

Présenter le schéma relationnel normalisé en 3NF de la base de données

1. vous avez reçu quelques heures plus-tard un e-mail de la part de Sophie, qui donne plus de détails sur ce qui est attendu de votre part :

*‘Bonjour,*

*Afin d’avancer sur le projet, 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 supplémentaires suivantes :*

* *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.*

1. *Ensuite, peux-tu* ***mettre à jour le dictionnaire de données,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. Pour le formalisme, nous avons choisi la méthode UML. Enfin, il faudra que tu me présentes ce nouveau schéma relationnel de la base de données qui donnera lieu à la création des nouvelles tables.’*

Les cardinalités doivent être présentes et cohérentes.

**Les Data pour le Milestone 1** Voici les fichiers à votre disposition

* *Template de données*
* *Le fichier zippé Donnée Immo contenant :*

*\*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.*

* *Schéma Relationnel*

**Milestone 2( Etape 2 ) : Implémentation de la base de donnée**

Une fois le modèle conceptuel de données, nouveau schéma relationnel modifié ainsi que le dictionnaire de données ont été validés, toute l’équipe vous félicite et se réunit pour définir le plan d’action d’implémentation de la base et le besoin en analyse de données dont vous serez le fer de lance.

#### Contraintes de la création de la base de données

\*Aucune technologie n’est imposée pour la création de la base, de même une solution open source est préférée (comme MySQL ou PostgreSQL ou SQlite)

\*Excel et/ou Power Query pourront être utilisés pour préparer les données et les faire correspondre au schéma relationnel.

\*Dans un premier temps, nous utiliserons seulement les données du premier semestre 2020. Cela servira de POC (Proof of Concept). Si la direction générale valide la solution, nous généraliserons plus tard avec les données des années précédentes

#### Requêtes SQL à effectuer pour les besoins d’analyse de données

À la demande du DG, Sophie vous demande d’extraire suivantes via des requêtes SQL sur les données :

1. Nombre total d’appartements vendus au 1er semestre 2020

2. Proportion des ventes d’appartements par le nombre de pièces.

3. Proportion des ventes d’appartements par le nombre de pièces

4. Liste des 10 départements où le prix du mètre carré est le plus élevé

5. Prix moyen du mètre carré d’une maison en Île-de-France

6. Liste des 10 appartements les plus chers avec le département et le nombre de mètres carrés

7. Taux d’évolution du nombre de ventes entre le premier et le second trimestre de 2020

8. Liste des communes où le nombre de ventes a augmenté d'au moins 20% entre le premier et le second trimestre de 2020

9. Liste des communes ayant eu au moins 50 ventes au 1er trimestre

10. Différence en pourcentage du prix au mètre carré entre un appartement de 2 pièces et un appartement de 3 pièces

11. Les moyennes de valeurs foncières pour le top 3 des communes des départements 6, 13, 33, 59 et 69

**MODALITES ET CRITERES DE PERFORMANCE**

-Le livrable final est un dépôt Github. Il devra contenir

* Dictionnaire des données
* Modèle conceptuel des données
* Schéma relationnel normalisé et schéma relationnel modifié avec
* Base de données opérationnelle avec les données du 1er semestre 2020 (captures d’écran directement du système de gestion de base de données des tables construites)
* Document avec les requêtes et les résultats ( dans un fichier PDF) et le fichier de code .sql associés à part