

Description du dictionnaire et autres éléments :

1.1) Dictionnaire des données

La première étape de la mise en place de la base de données a été l'élaboration minutieuse du dictionnaire des données. Ce document essentiel recense tous les champs inclus dans la base, accompagnés de leur signification, format, et contraintes éventuelles (ex. : unicité, valeurs par défaut, ou types de validation). Un champ de commentaire a également été ajouté pour garder une trace de la source originelle des données et des décisions de traitement appliquées, permettant ainsi une traçabilité complète. Cela vise à faciliter, au besoin, l'utilisation de ces données pour des analyses futures et leur intégration dans d'autres systèmes.

Les fichiers initiaux étant très volumineux et complexes, une sélection rigoureuse des champs pertinents s'est imposée. L'objectif était d'inclure uniquement les données stratégiques et utiles, afin d'optimiser la performance et de réduire la taille de la base, évitant ainsi des traitements superflus ou coûteux en termes de temps de calcul.

Les données sources proviennent principalement de data.gouv.fr, incluant des ensembles variés, tels que des données communes, des référentiels géographiques et le fichier des valeurs foncières. Afin de maximiser leur pertinence pour notre projet, nous avons effectué des transformations supplémentaires : nettoyage des données, unification des formats et application de règles de standardisation pour harmoniser les informations.

Parmi les champs retenus, nous avons intégré des informations clés répondant aux besoins de l'entreprise, comme des identifiants uniques pour faciliter les croisements de données, ainsi que des métadonnées précisant les périodes de validité et la fréquence de mise à jour des données.

En procédant ainsi, nous avons construit une base de données non seulement compacte, mais aussi évolutive et bien documentée, qui sera facilement extensible avec de nouvelles données ou adaptable à des futures analyses plus spécifiques, tout en assurant une excellente qualité de données pour l'ensemble des utilisateurs.

Tables bien :

La table bien est structurée principalement autour des caractéristiques des biens, à savoir le type de bien (maison ou appartement), la surface totale (Surface_Bâti), et le nombre de pièces. Chaque bien est identifié de manière unique par un ID, nommé ID_bien, généré via un système d'auto-incrémentation démarrant à 1 et s'étendant jusqu'au nombre total de biens disponibles

dans le répertoire de l'entreprise. Ce choix de numérotation permet d'attribuer un identifiant intrinsèque et stable à chaque bien, en conformité avec la méthodologie définie pour cet exercice.

La table bien constitue la table principale de notre base de données, servant de pivot pour la connexion avec d'autres tables. Elle est reliée, notamment, à la table commune, seconde table clé dans notre architecture, grâce à une clé étrangère Code_commune. Cette structure permet d'assurer une association cohérente entre chaque bien et la localisation géographique de sa commune, renforçant ainsi la dimension spatiale de notre analyse.

Par ailleurs, pour garantir une gestion efficace et optimiser les requêtes, nous avons implémenté des index sur les principaux champs de cette table, tels que ID_bien et Code_commune. Cela améliore la rapidité des requêtes de consultation et facilite l'exploration des données par localisation ou par typologie de bien.

Tables communes :

La table commune est principalement constituée du code de la commune, du code du département et du nom de la commune. Pour constituer la clé primaire de cette table, nous avons utilisé un système d'auto-incrémentation des valeurs d'ID pour chaque commune. Bien que cette méthode puisse sembler délicate dans un contexte d'importation, nous avons vérifié que les identifiants générés automatiquement ne causeront aucun problème lors de l'intégration de la base de données.

Cette table respecte une règle de gestion clé : à une commune peuvent être associés plusieurs biens, tandis qu'un bien ne peut être lié qu'à une seule commune. Par ailleurs, elle contient une clé étrangère, depart_ID, qui référence l'identifiant unique du département. Ainsi, chaque commune est rattachée à un département unique, tandis qu'un département peut regrouper plusieurs communes.

Enfin, pour garantir l'intégrité et l'efficacité des relations entre les tables, nous avons mis en place des contraintes de référence et des index, notamment sur les champs Code_commune et depart_ID. Cela optimise les performances des requêtes et assure la cohérence des données lors de la navigation entre les tables commune et bien, ainsi que dans les analyses basées sur des critères géographiques.

Table vente :

Concernant les ventes, cette table contient les informations suivantes : la date de la vente et le prix du bien. La date de la vente est enregistrée au format DATE, afin de conserver le format fourni dans le fichier transmis par l'entreprise et de faciliter le tri et les calculs basés sur des périodes.

La valeur de la mutation représente le montant total des ventes pour l'ensemble des biens, en fonction de leur date de transaction et de leurs caractéristiques. Le prix du bien est stocké au format décimal, permettant une précision accrue pour des valeurs comportant des décimales, ce qui évite les problèmes d'arrondi ou d'incompatibilité dans le SGBD pour des montants précis.

Cette table est reliée à la table bien via la clé étrangère ID_bien, qui identifie le bien concerné par chaque vente. Cela permet d'associer chaque transaction aux informations spécifiques du bien vendu et facilite la réalisation de requêtes pour générer des statistiques pertinentes pour l'entreprise, comme les tendances de vente, les prix moyens par type de bien ou par localisation, ou encore l'évolution des valeurs immobilières dans le temps.

Table département :

Cette table contient l'ID du département, le nom du département et l'ID de la région. Elle fournit des informations essentielles pour l'analyse géographique des départements, notamment en permettant de nommer chaque département et d'identifier sa région associée. Cela facilite la génération de statistiques au niveau des départements et des régions, répondant ainsi à des besoins d'analyse comme l'obtention du nombre de biens disponibles dans chaque département et chaque région.

En liant les informations de cette table aux autres tables de la base de données, il devient possible de réaliser des analyses plus approfondies, par exemple, sur la répartition des biens immobiliers, les tendances de prix par région ou département, et la densité des transactions immobilières dans différentes zones géographiques. Cette structure apporte également une flexibilité pour étendre les données géographiques et faciliter des études territoriales plus poussées.

Table région :

Cette table contient principalement l'identifiant unique de la région (ID_region) et le nom de la région. Pour les colonnes de cette table, nous avons choisi des types de données en chaînes de caractères afin d'assurer la conformité avec les standards d'importation et de mise à jour, réduisant ainsi les risques d'erreurs et respectant les contraintes de domaine.

Contrairement aux autres tables, cette table n'est pas directement liée à elles par des clés étrangères. Cependant, elle occupe une position fondamentale dans la structure de la base de données, car les autres tables en dépendent : il est nécessaire que cette table soit préalablement remplie pour que les données des autres tables, puissent être correctement importées et structurées. Ainsi, elle assure la cohérence des données régionales et permet de maintenir une hiérarchie géographique solide pour toutes les analyses futures.

1.2) Schéma relationnel et modèle logique des données

Notre schéma relationnel comprend principalement cinq tables, que nous avons décrites précédemment : bien, commune, département, région, et vente. Le modèle logique de données pour notre base se présente comme suit :

Table Bien :

Contient les informations spécifiques aux biens immobiliers (ID_bien, type de bien, surface totale, nombre de pièces) et est liée à la table commune via la clé étrangère Code_commune.

Table Commune :

Inclut les informations sur chaque commune (Code_commune, nom de la commune, ID de département) et est liée à la table département.

Table Département :

Regroupe les informations sur les départements (depart_ID, nom du département, ID de région) et assure la connexion avec la table région.

Table Région : Fournit les informations sur les régions (ID_region, nom de la région) et constitue la base géographique de la hiérarchie.

Table Vente :

Stocke les transactions de vente (ID_vente, date de la vente, prix du bien) et se lie à la table bien par l'ID_bien, permettant des analyses

Table Départements :

Définis par l'identification depart_ID, lesquels sont eux-mêmes reliés au département et aux communes.

1.3) Description des requêtes

Requête 1 :

La requête 1 est composée de deux sous-requêtes, visant à obtenir le nombre de biens vendus par type dans chaque région pour la période du 1er janvier 2020 au 30 juin 2020. La première sous-requête fournit le nombre d'appartements vendus par région, tandis que la seconde donne le nombre de maisons vendues.

Requête 2 :

L'objectif de la requête 2 est de comparer le nombre de maisons vendues au nombre d'appartements. La première sous-requête fournit, en fonction du nombre de pièces des maisons, le ratio entre les maisons vendues et le nombre d'appartements pour les deux premiers trimestres de l'année 2020. La deuxième requête donne le ratio entre les appartements et les maisons pour les deux premiers trimestres de l'année 2020.

Requête 3 :

Cette requête compare le nombre de ventes de maisons et d'appartements pour les deux premiers trimestres de l'année 2020. La première sous-requête calcule le ratio entre le nombre de maisons vendues et celui des appartements, en fonction du nombre de pièces des maisons et de la région. La deuxième sous-requête, quant à elle, fournit le ratio inverse : celui des appartements par rapport aux maisons pour la même période et selon les mêmes critères.

Requête 4 :

L'objectif de cette requête SQL est d'analyser la valeur foncière moyenne par mètre carré pour un certain type de bien immobilier dans une région spécifique, pour une période donnée. Elle regroupe et structure les informations sur les biens vendus afin de donner une vue d'ensemble sur les prix moyens des biens en fonction des régions et des types de biens.

Requête 5 :

Cette requête permet de connaître le nombre de ventes du premier et du deuxième trimestre.

Requête 6 :

La requête 6 a consisté à calculer principalement le prix des moyens des appartements ayant plus de 4 pièces par personnes.

Principalement on note les résultats suivants :

L'Île-de-France est de loin la région la plus chère avec un prix moyen de 8 030,81 €/m², ce qui n'est pas surprenant étant donné qu'elle inclut Paris.

La Région PACA (Provence-Alpes-Côte d'Azur) se situe en troisième position avec 2 942,35 €/m², reflétant l'attractivité du Sud-Est de la France.

On observe un écart très important entre l'Île-de-France et les autres régions, le prix au m² y étant presque deux fois plus élevé que dans la deuxième région la plus chère.

Les régions où le prix moyens des appartements de plus de 4 personnes sont : La Martinique (564,22 €/m²), Bourgogne-Franche-Comté (1 068,93 €/m²), Grand Est (1 313,26 €/m²).

Requête 7 :

La requête 7 vise à comparer les prix moyens entre les appartements de 2 pièces et ceux de 3 pièces, ce qui nous montre une tendance intéressante :

Elle montre principalement que les petites surfaces sont généralement plus demandées en zone urbaine, les appartements de 2 pièces sont souvent prisés par les investisseurs pour la location et la demande est plus forte pour les petites surfaces, notamment de la part des étudiants et jeunes actifs. Ce sont ces raisons qui font que plus la surface est petite, plus le prix au mètre carré tend à être élevé.

Requête 8 :

La requête 8 vise à présenter les communes ayant enregistré plus de 50 ventes durant janvier à mars 2020. On remarque cependant :

Une forte concentration dans les communes parisiennes et la présence des villes importantes comme Marseille, Nice, Versailles et Sète. Les nombres varient significativement allant de plus de 50 à des centaines pour les communes de Paris.

Requête 9 :

L'objectif de cette requête SQL est d'extraire les 3 communes les plus chères par département pour plusieurs départements français sélectionnés, en se basant sur la valeur foncière moyenne des biens immobiliers vendus. Cette analyse permet de comparer la valeur foncière moyenne dans différentes communes au sein de chaque département, en mettant en évidence les communes ayant les prix moyens les plus élevés.

On remarque principalement que la commune 356 a une valeur foncière moyenne de 330000 EUR classée 1ère, suivie de la commune 336 avec une position de 314425 EUR et enfin en 3^{ème} position avec la commune 329 qui compte jusqu'à 314416,88 EUR.