

## Analyse de l'immobilier à Paris

**Cursus concerné :** Analytics Engineer

**Difficulté :** 9/10

**Description détaillée :**

Le gouvernement français et la ville de Paris publient régulièrement des données sur les biens immobiliers. Quelques exemples sont les prix de vente, les limites légales de prix de location, le risque d'inondation, la localisation des écoles maternelles et bien d'autres choses encore. L'objectif de ce projet est d'exploiter plusieurs de ces sources pour analyser le prix de l'immobilier à Paris.

Etape	Description	Objectif	Modules / Masterclass / Templates	Conditions de validation du projet
1	Exploration de données non structurées	<p>Télécharger le jeu de données des valeurs foncières déclarées pour l'année 2022 sur <a href="https://data.gouv.fr">data.gouv.fr</a>. Cet ensemble de données doit être nettoyé, à cet effet vous allez :</p> <ul style="list-style-type: none"><li>- Conserver uniquement les données de Paris</li><li>- Supprimer les colonnes contenant plus de 50 % de données manquantes.</li><li>- Après l'étape précédente, supprimer toutes les lignes contenant des données manquantes.</li></ul> <p>Passer par l'API de <a href="https://opendata.paris">opendata.paris</a> pour récupérer des données pertinentes pour le potentiel commercial d'un bien immobilier. Quelques exemples : <a href="#">secteurs de risques</a>, <a href="#">encadrement des loyers</a>, <a href="#">espaces libres à végétaliser</a>, etc.</p> <p>Cette étape est importante, vous devez comprendre les données que vous pouvez récupérer et faire un choix des routes à utiliser.</p>	103 - Introduction aux APIs de données	<p>Un fichier explicatif du traitement et des différentes données accessibles (doc / pdf)</p> <p>Un exemple de données collectées.</p>
2	Modélisation des données	Concevoir un schéma de base de données respectant la troisième forme normale (3NF) ou en <b>appliquant les principes de la dénormalisation</b> .	145 - Fondamentaux de l'intégration de données	Données stockées dans une base de données relationnelle

			142 - SQL	(PostgreSQL/MySQL )  <b>Diagramme UML</b>  Un fichier de requêtes SQL pour montrer que la base est fonctionnelle
<b>3</b>	Construction du pipeline ETL/ELT	<p>Créer un pipeline ETL/ELT pour remplir la base de données :</p> <ul style="list-style-type: none"> <li>- <b>Extraction</b> : charger les données à partir des sources d'origine</li> <li>- <b>Chargement</b> : dénormaliser (schéma en étoile) et charger les données dans le système cible (SQL Database)</li> <li>- <b>Transformation</b> : Fusion, Encodage,</li> </ul> <p>Exécuter le pipeline pour intégrer autant de données que possible à partir de l'ensemble de données d'origine</p>	Snowflake  BigQuery  Snaplogic  144 - ETL avec pyspark	Démo du DataWarehouse   Schéma en étoiles
<b>4</b>	Dashboard final	Création du Dashboard Power BI (ou autre) et préparation de la soutenance.	181 - Power BI  182 - Tableau  185 - Looker Studio	Soutenance  Rapport