

# Mastère 2

## **BRIEF PROJET** DATA SCIENCE : *IMMOBILIER FRANCE*

*ANNEE*  
*2019 – 2020*



## Présentation du projet

Ce projet consiste à étudier les transactions immobilières intervenues au cours des cinq dernières années. Ces données sont disponibles sur le site data gouv :

<https://www.data.gouv.fr/fr/datasets/demandes-de-valeurs-foncieres/#>

Les fichiers de 2015 à 2019 sont à concaténer en un seul. Ces données correspondent aux transactions immobilières sur le territoire métropolitain et les DOM-TOM, à l'exception de l'Alsace-Moselle et de Mayotte. Les données contenues sont issues des actes notariés et des informations cadastrales.

Ce projet a pour but de mettre en application vos acquis techniques mais aussi managériaux. Voici les deux problématiques que vous devez traiter :

### *1. Problématique Business :*

Il s'agira de créer un **dash board** permettant de synthétiser l'ensemble des informations pertinentes sur les logements immobiliers en France. Ce dashboard servira d'outil d'aide à la décision d'achat et permettra de répondre aux questions suivantes :

- Où investir ?
- Quel type de bien acheter ? (Donner des exemples)
- Quels types de bien sont à éviter ? (Donner des exemples)

Ce dashboard peut être réalisé sur Excel, Flask ou encore Dash. La solution Excel sera privilégiée. Il constituera, par ailleurs, le support de présentation de la soutenance du 7 septembre. Ce Dashboard comprendra, entre autres, les éléments des problématiques métiers ci-dessous.

### *2. Problématique métier :*

- Répartition des différents types de biens (Pie chart – barplot)
- Distribution du nombre de biens vendus en fonction du nombre de pièces
- Evolution du nombre de ventes en fonction des années
- Répartition géographique du nombre de biens vendus sur la France métropolitaine (Librairies Folium)

## Livrables

Les livrables attendus pour chaque étudiant sont :

- Modèle Physique de données des tables
- Notebook d'analyse de données en Python
- Dashboard Excel/Flask/Dash au choix (lien vers le Dashboard s'il s'agit d'une application)

On valorisera particulièrement la capacité de l'élève à aller au-delà des analyses proposées et à expliciter son raisonnement en vue de répondre aux problématiques business.

On rappellera enfin, que le chargement des données se fera via MySQL et que l'interrogation des tables se fera en python sur des données issues de tables SQL (*Cf projet data Science de cette année*).

## Modalités de rendu

### *A. Rendu écrit*

Le rendu écrit doit se faire sur **Classroom**, dans le devoir intitulé « *Projet Data Science – Mastère 2* », de préférence dans un *dossier zip* avec tous vos documents,

**au plus tard le 06 Septembre 2020.**

### *B. Soutenance orale*

La soutenance orale se tiendra sur la semaine du 7 au 11 Septembre, et aura lieu en visioconférence via Google Meet.

**Durée : 10 minutes**

- ⇒ 5 minutes sur la partie business
- ⇒ 5 minutes sur la partie technique

Cette soutenance consistera à présenter de manière vivante et intéressante vos travaux.

## **Références/liens utiles :**

- sqlalchemy : <https://docs.sqlalchemy.org/en/13/orm/tutorial.html>
- MCD : <https://web.maths.unsw.edu.au/~lafaye/CCM/merise/mcd.htm>
- Folium : <https://python-visualization.github.io/folium/>
- Dash : <https://dash.plotly.com/>
- Flask : <https://dev.to/admindashboards/flask-dashboard-atlantis-dark-open-source-admin-panel-with-dark-design-4l6>