

Présentation du dataset

Titre du projet : Regroupement des maisons de l'Ames (ville de l'Iowa)

Groupe 6 : Adam MATAOUI, Samuel METIN, Jonathan MOHAMEDALY, Valentin PHOU, Sarangan UTHAYAN

Le dataset que nous avons choisi est disponible sur Kaggle via le lien [suivant](#).

Le dataset *Ames Iowa Housing Data* contient des informations sur diverses propriétés vendues à Ames, dans l'Iowa, entre 2006 et 2010. Il rassemble les données de vente de plus de 2 900 maisons, (79 variables au total). Ces informations peuvent être regroupées en quatre catégories:

- **Caractéristiques de base de la maison**
- **Détails intérieurs**
- **Détails extérieurs**
- **Informations sur la vente de la propriété**

1/ Caractéristiques de base

On retrouve dans cette catégorie l'ensemble des informations essentielles sur la maison que ce soit des informations sur le voisinage et ses spécificités, mais aussi sur les matériaux utilisés pour sa construction.

Le dataset comporte aussi des informations plus spécifiques au logement telles que sa surface habitable (**Lot Area** en pieds carrés), le nombre de chambres disponible (**Bedroom AbvGr**) mais aussi le type de logement (**Bldg Type**, **House Style**, ...).

2/ Détails intérieurs

Les détails intérieurs contiennent toutes les informations concernant l'intérieur de la maison qui sont trop spécifiques pour être dans les caractéristiques de base. Les informations concernant l'intérieur du garage se situent dans la partie Détails extérieurs.

Parmi les informations dans cette partie on retrouve la qualité de la cuisine et des salles de bains (**Kitchen Qual**, **Full Bath**), le types de sol, mais aussi des informations moins importantes tels que le nombre de cheminées ainsi que leurs qualités (**Fireplaces**, **Fireplace Qu**).

3/ Détails extérieurs

Les caractéristiques extérieures englobent l'ensemble des informations relatives à l'extérieur de la propriété que ce soit des informations sur le garage, l'aménagement du terrain, le type de barrières utilisés mais aussi sur les matériaux utilisés.

On retrouve par exemple la capacité du garage (**Garage Cars**), la composition du toit (**Roof Matl**) ou des informations sur la piscine (**Pool Area**).

4/ Informations sur la vente de la propriété

Les informations sur la vente incluent la date de la vente (**Yr Sold**, **Mo Sold**) et les conditions de vente mais aussi le prix de vente (**SalePrice**).

5/ Étapes d'Analyse et Préparation des Données du Dataset

Dans la suite, nous allons nous concentrer sur le nettoyage et le pré-traitement des données du dataset. Cela comprend la correction des types de données et la transformation des variables catégorielles dans un format adéquat.

Une fois les étapes de nettoyage et de pré-traitement effectuées, nous utiliserons un algorithme de clustering pour regrouper les maisons en différentes catégories. Cela nous permettra de proposer, pour un budget donné, le type de maison optimal pour un client, en tenant compte de ses préférences et besoins.