

## Travail Pratique #2 (17.5 points)

Ce travail compte pour 17.5% de votre note finale.

Date de remise : mardi 3 décembre 2024 avant 22h

### Directives :

- Même équipe que pour le travail pratique #1
- Pour répondre aux questions suivantes, vous aurez besoin du fichier de données qui se trouve sur Moodle dans le dossier *data\_projet.zip* et devez utiliser python.
- Expliquez clairement ce que vous avez tenté de réaliser à chaque étape/sous-étape à l'aide de commentaires dans le code. Pour inscrire un commentaire, vous pouvez simplement écrire une phrase avec un # devant ou englober un paragraphe de trois guillemets anglais, au début et à la fin du paragraphe ("").
- Vous devez remettre soit :
  - ❑ un code (.py) commenté en expliquant vos étapes aux différentes questions clairement en commentaire dans le code pour chaque question.
  - ❑ un code (.ipynb) incluant vos explications.
- Veuillez déposer votre projet sur moodle, au plus tard le 3 décembre 2024 avant 22h.
- Inscrire tous les prénoms, noms et matricules des participants de l'équipe à l'intérieur de chaque document.

Dans le cadre de ce travail, vous utiliserez le fichier suivant:

- MilwaukeePropertySalesData.csv

déposés sur Moodle dans le dossier *data\_projet.zip*.

Les données proviennent du site «Milwaukee Open Data». Ces données contiennent des informations concernant les ventes de propriétés résidentielles et commerciales dans la ville de Milwaukee aux États-Unis. Même si les données sont disponibles pour téléchargement sur leur site web, vous devez obligatoirement utiliser le fichier .csv fourni. Voici une description des variables du jeu de données:

- `_id` : Identifiant unique pour chaque entrée dans le jeu de données.
- `PropertyID` : Identifiant unique pour chaque propriété.
- `PropType` : Type de propriété (par exemple, résidentielle, commerciale, condominium).
- `taxkey` : Clé fiscale unique associée à la propriété.
- `Address` : Adresse physique de la propriété.
- `CondoProject` : Nom du projet de condominium, le cas échéant.
- `District` : District administratif où se trouve la propriété.
- `nbhd` : Code du quartier où se trouve la propriété.
- `Style` : Style architectural de la propriété.

- Extwall : Matériau principal des murs extérieurs.
- Stories : Nombre d'étages de la propriété.
- Year\_Built : Année de construction de la propriété.
- Rooms : Nombre total de pièces dans la propriété.
- FinishedSqft : Surface habitable en pieds carrés.
- Units : Nombre d'unités dans la propriété (pertinent pour les immeubles à plusieurs logements).
- Bdrms : Nombre de chambres à coucher.
- Fbath : Nombre de salles de bain complètes.
- Hbath : Nombre de demi-salles de bain.
- Lotsize : Taille du terrain en pieds carrés.
- Sale\_date : Date de la vente de la propriété.
- Sale\_price : Prix de vente de la propriété.

Les bases du langage Python sont abordées en classe, mais n'hésitez pas à explorer par vous-même, par exemple sur le web, pour vous aider à résoudre les questions.

Dans votre script Python, identifiez les différentes sections par l'utilisation de zones de commentaires. Seul Python doit être utilisé pour réaliser le projet.

1. (2 points) Y-a-t-il des valeurs extrêmes dans la variable **FinishedSqft**. Si oui, enlevez-les avec la méthode de votre choix.
2. (2 points) Un agent immobilier affirme que le prix de vente moyen des propriétés résidentielles dans le District 11 est supérieur à 250 000 \$. Utilisez le test d'hypothèse approprié pour déterminer s'il y a suffisamment de preuves pour soutenir cette affirmation à un alpha de 5 %.
3. (2 points) Un chercheur souhaite déterminer s'il existe une différence significative dans la superficie habitable (**FinishedSqft**) moyenne entre les propriétés résidentielles construites avant 1950 et celles construites en 1950 (**Year\_Built**) ou après. Effectuez un test d'hypothèse approprié à un alpha de 5 % pour répondre à cette question.
4. (2 points) Un analyste du marché immobilier souhaite déterminer s'il existe une relation significative entre le nombre de chambres (**Bdrms**) dans une propriété résidentielle et le fait que la propriété ait été vendue au-dessus ou en dessous du prix de vente médian. Regroupez les propriétés ayant 5 chambres ou plus dans la même catégorie de nombre de chambres. Effectuez le test d'hypothèse approprié avec un alpha de 5 % pour déterminer s'il existe une association entre le nombre de chambres et la catégorie de prix de la propriété (au-dessus ou en dessous de la médiane).
5. (2 points) Calculez le pourcentage de variation du prix moyen de vente d'un mois à l'autre et créez un graphique de ces variations par mois.
6. (1.5 points) Vérifiez si la série temporelle du nombre de propriétés vendues quotidiennement est stationnaire.
7. (6 points) Effectuez un regroupement avec K-means sur les ventes de propriétés immobilières du mois de juin. Expérimentez différents nombres de regroupement et utilisez la méthode du coude pour déterminer le nombre optimal de groupe. Discutez de vos choix de variables pour effectuer les regroupements. Décrivez chacun des regroupements obtenus à l'aide de graphiques pertinents.

Bon travail!