**Exercice : Prédire le prix de location d’un appartement à Dakar**

**Contexte :**

Vous travaillez pour une startup immobilière à Dakar qui souhaite prédire le **prix de location mensuel** d’un appartement. On vous confie un jeu de données contenant des informations sur des appartements loués récemment.

Votre mission : entraîner un **modèle de régression linéaire** pour prédire le **prix mensuel** à partir des caractéristiques disponibles.

| **Variable** | **Description** |
| --- | --- |

|  |  |
| --- | --- |
| surface\_m2 | Surface de l'appartement en m² |

|  |  |
| --- | --- |
| nombre\_chambres | Nombre de chambres |

|  |  |
| --- | --- |
| quartier | Quartier de Dakar (Plateau, Yoff, Almadies...) |

|  |  |
| --- | --- |
| etage | Étage où se situe l'appartement |

|  |  |
| --- | --- |
| ascenseur | Oui/Non |

|  |  |
| --- | --- |
| meublé | Oui/Non |

|  |  |
| --- | --- |
| prix\_mensuel | Loyer en FCFA (variable cible) |

### ****Tâches à réaliser**** :

#### Chargement des données

* Importer les librairies nécessaires.
* Charger le fichier appart\_dakar.csv.
* Afficher les 5 premières lignes.

#### Exploration des données

* Afficher les statistiques descriptives.
* Vérifier les valeurs manquantes.
* Nettoyer les données si nécessaire (valeurs manquantes ou aberrantes).

#### Visualisation

* Créer un nuage de points entre surface\_m2 et prix\_mensuel.
* Créer une heatmap de corrélation pour les variables numériques.

#### Encodage

* Convertir les colonnes catégorielles (quartier, ascenseur, meublé) avec pd.get\_dummies() ou LabelEncoder.

#### Régression Linéaire Simple

* Prédire prix\_mensuel à partir de surface\_m2 uniquement.
* Évaluer le modèle avec **MAE, RMSE et R²**.
* Tracer la droite de régression sur un nuage de points.

#### Régression Linéaire Multiple

* Utiliser toutes les variables pertinentes pour prédire prix\_mensuel.
* Évaluer les performances avec **MAE, RMSE et R²**.
* Comparer avec la régression simple.

#### . Conclusion

* Quel modèle est le plus performant ?
* Quelles sont les variables les plus influentes ?
* Quelles améliorations pourriez-vous proposer ?