

# Projet M2 2022-23

## Modèles pour la régression

Mettre vos noms ICI

### Les données

Le jeu de données concerne la vente de 5891 logements et de 29 variables explicatives décrivant (presque) tous les aspects de ce logement et de son environnement en Corée du Sud dans la ville de Daegu, sur une période de 10 ans.

#### But de l'étude

L'objectif est de prédire le prix de vente et d'obtenir le modèle qui aura le plus petit **RMSE**. Vous devez contruire un (ou 2 modèles) qui permet de calculer le prix de vente du logement ( $Y=\text{SalePrice}$ ) en fonction de toutes ou une partie des variables (regresseurs/facteurs) contenues dans le jeu de données. Il y a 2 jeux de données :

- Le “train” avec 4189 observations, à partir duquel vous devez contruire votre modèle final
- Le “test” avec 1702 observations, que vous ne devez pas utiliser pour contruire votre modèle.

```
train=read.csv2("train.csv",header=TRUE,sep=",")
test=read.csv2("test.csv",header=TRUE,sep=",")
```

```
set.seed(2022)
```