**Computing Problem**

Use the test data set HDBresale\_subset24.csv in ELEARN TestA. Data are HDB 4-room flats resale statistics from Jan 2021 to Sep 2023. Features are Height (3 implies 1 to 3 storeys, 6 implies 4 to 6 storeys,…, 24 implies 22 to 24 storeys), Remaining\_lease in years, categorical nominal variable of location (C, N, E, S, W implies central, northern, eastern, southern, western districts respectively), and categorical flat models (Improved type, Model\_A type, New\_Generation type, Premium type, and Simplified type). The binary classification is Rel\_value (relative value: whether S$ resale price per square metre is > 6,500 for Y=1, or £ 6,500 for Y=0). There are 29,669 transactions.

Perform MinMaxScaler preprocessing of the features (including label Y) before the Train, Test sets splitting. Then use: "from sklearn.model\_selection import train\_test\_split \ Train, Test = train\_test\_split(df, test\_size=0.25, random\_state=0)".

Apply GradientBoostingClassifier to predict Y using "GradientBoostingClassifier(n\_estimators=500, random\_state=1, max\_features="sqrt", learning\_rate=0.01, max\_depth=None)".

Also apply RandomForestClassifier to predict Y using "RandomForestClassifier(n\_estimators=500, random\_state=1, max\_features="sqrt", max\_depth=None)".















