



## Machine Learning predicción Airbnb

En el módulo de Machine Learning se realizó la predicción del precio de Airbnb. Tras la carga del dataset y la limpieza de los datos se dividió en Train/Test.

Con los datos de Train se revisaron las variables y se eliminaron aquellas que no eran relevantes para el estudio. El dataset contenía bastantes valores NA por lo que se estudió y aquellas variables que contaban con un porcentaje superior al 50% se eliminaron completamente. En el caso de variables con menor porcentaje se realizó la media con respecto al precio para rellenar aquellos campos que contenían NA. Para las variables que únicamente contenían porcentajes inferiores al 5% se eliminaron los datos NA.

Una vez que las variables estaban limpias aquellas que no eran int64 o float64, es decir, objetos se transformaron para que todas tuvieran el mismo formato y los modelos de entrenamiento pudieran trabajar con ellos.

En cuanto a los modelos se utilizaron:

\* **Regresión:**

- MSE train (grado 2): 842.1464
- MSE test (grado 2): 1229.2923
- RMSE train (grado 2): 29.0197
- RMSE test (grado 2): 35.0612

\* **Random Forest Regressor:**

- Best mean cross-validation score: 0.637
- Best parameters: max\_depth 13
- Train: 0.8591
- Test: 0.6105
- MSE train Random Forest: 323
- MSE test Random Forest: 1.18e+03
- RMSE train Random Forest): 18
- RMSE test Random Forest: 34.4

\* **Bagging Regressor:**

- Best mean cross-validation score: 0.637
- Best parameters: max\_depth 12
- Train: 0.8556
- Test: 0.6349
- MSE train Bagging: 323
- MSE test Bagging:: 1.18e+03
- RMSE train Random Forest): 18
- RMSE test Random Forest: 34.4