Algoritmo de Árbol de Decisión (Preguntas)

a) Capturing patterns from data is called… fitting or training the model.

b) The data used to fit the model is called… the training data.

c) After the model has been fit, you can apply it to new data to… predict prices of additional homes.

d) You predict the price of any house by tracing through… the decision tree.

e) The point at the bottom where we make a prediction is called… a leaf.

f) Explica lo que representa cada fila de la siguiente tabla:

	Rooms	Bathroom	Landsize	Lattitude	Longtitude
count	60.000000	60.000000	60.000000	60.000000	60.000000
mean	2.716667	1.566667	251.133333	-37.777957	144.939105
std	0.783120	0.620734	244.073028	0.048900	0.054444
min	1.000000	1.000000	0.000000	-37.848100	144.867900
25%	2.000000	1.000000	123.000000	-37.808125	144.878975
50%	3.000000	1.500000	165.500000	-37.801550	144.952150
75%	3.000000	2.000000	266.750000	-37.723775	144.995400
max	6.000000	3.000000	1063.000000	-37.716400	145.000400

- **count:** The first number, the count, shows how many rows have non-missing values.
- **mean:** The second value is the mean, which is the average.
- **std:** std is the standard deviation, which measures how numerically spread out the values are.

To interpret the min, 25%, 50%, 75% and max values, imagine sorting each column from lowest to highest value.

- **min:** The first (smallest) value is the min.
- **25%:** If you go a quarter way through the list, you'll find a number that is bigger than 25% of the values and smaller than 75% of the values. That is the 25% value.
- **50% & 75%:** The 50th and 75th percentiles are defined analogously.

- **max:** the max is the largest number.

g) The column we want to predict, which is called… the prediction target.

h) The columns that are inputted into our model (and later used to make predictions) are called: "features."

i) The steps to building and using a model are:

- **Define:** What type of model will it be? A decision tree? Some other type of model? Some other parameters of the model type are specified too.
- **Fit:** Capture patterns from provided data. This is the heart of modeling.
- **Predict:** Just what it sounds like.
- **Evaluate:** Determine how accurate the model's predictions are.

j) Observa cómo es la predicción de los precios tomando en cuenta cinco registros, ¿cuál sería el MAE para los datos predichos?

R= 1035000, 1465000, 160000, 1876000, 1636000

k) ¿Cómo se define el Error Medio Absoluto (MAE Mean Absolute Error)?

R= Tomamos el valor absoluto de cada error. Esto convierte cada error en un número positivo. Luego tomamos el promedio de esos errores absolutos. Esta es nuestra medida de calidad del modelo.

l) ¿Cuál es el MAE de tu entrenamiento con cinco registros?

R= 666.6666666666666

m) Después de dividir el conjunto de datos en las variables de entrenamiento y validación (train_X, val_X, train_y, val_y), así como después de entrenar nuevamente, ¿cuál es el MAE que obtienes?

R= 257933.33333333334