

ACTIVIDAD EL ALGORITMO DE ÁRBOL DE DECISIÓN

- a) Capturing patterns from data is called **Ajustar o entrenar al modelo**
- b) The data used to fit the model is called **Datos de entrenamiento**
- c) After the model has been fit, you can apply it to new data to **Predecir**
- d) You predict the price of any house by tracing through **El árbol de medición**
- e) The point at the bottom where we make a prediction is called **Hoja (a leaf)**
- f) Explica lo que representa cada fila de la siguiente tabla (revisa la sección 2)

	Rooms	Bathroom	Landsize	Lattitude	Longitude
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:** muestra cuantas filas tienen valores no faltantes.
 - **Mean:** es la media
 - **Std:** es la desviación estándar, la cual mide que tanto se expanden los valores numéricamente.
 - **Min:** Es valor mínimo, el más pequeño
 - **25%:** Son los números mayores a un cuarto pero menores al 75%
 - **50% y 70%:** Se definen de forma análoga
 - **Max:** el número mayor.
- g) The column we want to predict, which is called **El objetivo de predicción**
 - h) The columns that are inputted into our model (and later used to make predictions) are called **Características (features)**
 - i) The steps to building and using a model are
 - **Definir:** ¿Qué tipo de modelo será? ¿Un árbol de decisión? ¿Algún otro tipo de modelo? Otros parámetros del tipo de modelo se especifican también.
 - **Ajuste:** Captura patrones de los datos proporcionados.
 - **Predecir**
 - **Evaluar:** Determinar qué tan precisas son las mediciones del modelo

- j) Observa cómo es la predicción de los precios tomando en cuenta cinco registros.
NOTA: En realidad no son las primeras cinco en orden de aparición puesto que se excluyeron los registros que tienen celdas vacías en el archivo melb_data.csv
¿cuál sería el MAE para los datos predichos? **[1035000. 1465000. 1600000. 1876000. 1636000.]**
- k) ¿Cómo se define el Error Medio Absoluto (MAE Mean Absolute Error)? **Métricas para resumir la calidad del modelo. Con la métrica MAE, 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? **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? **262000.0**