Revisa las **secciones 1 a 5** del siguiente link:

https://www.kaggle.com/learn/intro-to-machine-learning

Responde correctamente las siguientes preguntas guía:

1. La captura de patrones a partir de datos se llama:

**Ajuste o capacitación del modelo.**

1. Los datos utilizados para ajustar el modelo se denominan:

**Datos de entrenamiento.**

1. Después de que el modelo se haya ajustado, puede aplicarlo a los datos nuevos para después de que el modelo se haya ajustado, puede aplicarlo a:

**Nuevos datos para predecir los precios de viviendas adicionales**.

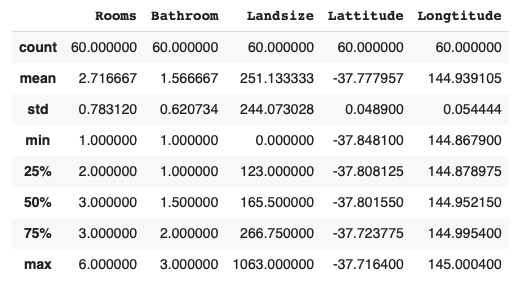
1. Usted predice el precio de cualquier casa rastreando:

**A través del árbol de decisiones.**

1. El punto en la parte inferior donde hacemos una predicción se llama:

**Hoja.**

f) Explica lo que representa cada fila de la siguiente tabla (revisa la sección 2).



**-Cuenta.**, muestra cuántas filas tienen valores no faltantes.

-**La media**, que es el promedio.

-**Desviación estándar**, que mide la extensión numérica de los valores.

**-Los valores mínimo, 25%, 50%, 75% y máximo**, imagine ordenar cada columna del valor más bajo al más alto. El primer valor (el más pequeño) es el mínimo. Si recorre un cuarto de camino en la lista, encontrará un número que es mayor que el 25% de los valores y menor que el 75% de los valores. Ese es el valor del 25% (pronunciado "percentil 25"). Los percentiles 50 y 75 se definen de forma análoga, y el máximo es el número más grande.

1. La columna que queremos predecir, que se llama:

**Target de predicción.**

h) Las columnas que se ingresan en nuestro modelo (y luego se usan para hacer predicciones) son llamados:

**Características.**

i) Los pasos para construir y usar un modelo son:

**Definir:** ¿qué tipo de modelo será? ¿Un árbol de decisiones? ¿Algún otro tipo de modelo? Algunos También se especifican otros parámetros del tipo de modelo.

**Ajuste:** captura patrones de los datos proporcionados. Este es el corazón del modelaje.

**Predecir:** justo como suena.

**Evaluar:** determine cuán precisas son las predicciones del modelo.

j) Observa cómo es la predicción de los precios tomando en cuenta cinco registros.

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*)?

**El error = valor actual - valor predicho.  
Toma el valor absoluto de cada error, esto convierte cada error en un número positivo.**

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?

**220433.33333333334**