



## Capítulo 2. Proyecto de aprendizaje automático de un extremo a otro

En este capítulo, trabajará en un proyecto de ejemplo de principio a fin, pretendiendo ser un científico de datos contratado recientemente en una empresa inmobiliaria. 1 Estos son los pasos principales por los que pasará:

Mira la imagen completa.

Obtenga los datos.

Descubra y visualice los datos para obtener información.

Prepare los datos para los algoritmos de Machine Learning.

Seleccione un modelo y entrénelo.

Afina tu modelo.

Presenta tu solución.

Inicie, supervise y mantenga su sistema.

### Trabajar con datos reales

Cuando esté aprendiendo sobre Machine Learning, es mejor experimentar con datos del mundo real, no con conjuntos de datos artificiales. Afortunadamente, hay miles de conjuntos de datos abiertos para elegir, que abarcan todo tipo de dominios. Aquí hay algunos lugares en los que puede buscar para obtener datos:

### Repositorios de datos abiertos populares

Repositorio de aprendizaje automático de UC Irvine

Conjuntos de datos de Kaggle

Conjuntos de datos de AWS de Amazon

Metaportales (enumeran repositorios de datos abiertos)

Portales de datos

OpenDataMonitor

Quandl

Otras páginas que enumeran muchos repositorios de datos abiertos populares

Lista de Wikipedia de conjuntos de datos de aprendizaje automático

Quora.com

El subreddit de conjuntos de datos

En este capítulo usaremos el conjunto de datos de precios de vivienda de California del repositorio 2 de StatLib (consulte la Figura 2-1 ). Este conjunto de datos se basa en datos del censo de California de 1990. No es exactamente reciente (una bonita casa en el Área de la Bahía todavía era asequible en ese momento), pero tiene muchas cualidades para el aprendizaje, por lo que pretendemos que son datos recientes. ...