# Machine Learning

## Introducción al problema

Se va a estudiar la concesión de hipotecas entre 1997 y 1998 en Boston de acuerdo a una serie de variables. Los datos provienen de la reserva federal de Boston, y pueden ser encontrados en https://vincentarelbundock.github.io/Rdatasets/doc/Ecdat/Hdma.html

La variable objetivo es “deny”: es una variable categórica binaria, Si el valor es “Yes” se rechaza la hipoteca, en cambio, cuando el valor es “no” se concede la hipoteca. EL objetivo del problema es construir un modelo clasificador binario de la variable deny en función de unas variables independientes.

Hay 12 variables independientes, de ellas 6 son continuas y 6 son categóricas.

Las variables continuas son las siguientes:

* Dir: el ratio entre deuda y el salario
* hir: ratio de los gastos de la casa y el salario
* lvr: ratio entre tamaño de la deuda y el valor de la propiedad
* ccs: puntuación de credito del consumidor del 1 al 6 (cuanto menor sea la puntuación mejor puntuación es)
* mcs: puntuación de crédito de la hipoteca del 1 al 4 (cuanto menor sea la puntuación mejor puntuación es)
* uria: la tasa de desempleo en el sector del solicitante en Massachusetts de 1989

Las variables categóricas son las siguientes:

* Pbcr: ¿ El solicitante tiene malos registros crediticios públicos? Las categorías de esta variable son “Yes” o “no”.
* dmi: ¿ Se le ha denegado el seguro de hipoteca al solicitante? Las categorías de esta variable son “Yes” o “no”.
* self: ¿Es el solicitante autónomo? Las categorías de esta variable son “Yes” o “no”.
* single: ¿Es el solicitante soltero? Las categorías de esta variable son “Yes” o “no”.
* condominium: ¿La casa es un condominio? Las categorías de esta variable son “Yes” o “no”.
* black ¿El solicitante es negro? Las categorías de esta variable son “Yes” o “no”.

## Exploración de variables

## Selección de variables