**Intro to Data Science Final Project**

Objective: To identify profiles of households who own their real estate with mortgage or loan

Method: Building a predictive model which discriminates between  
 - Households that own their house with mortgage or loans (including equity loans)  
 - Households that own their homes free and without any loans

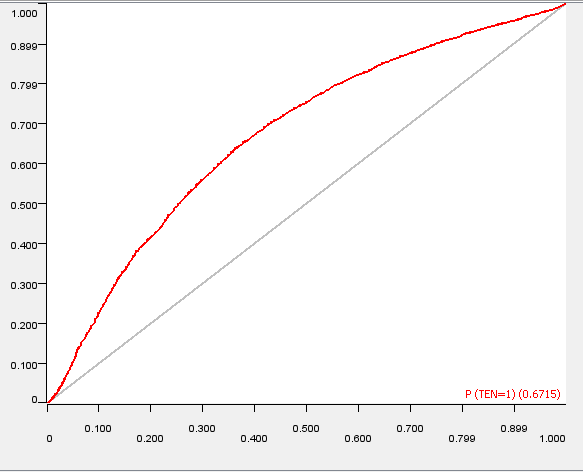
Data Notes: Output should be trained for Variable “TEN”: either   
1: “Owned with mortgage or loan”  
 or  
2: “Owned free and clear”

**Potential Relevant Criterion:**

Division: Area of the US where census was taken (0-9)  
Region: Region code of US where census was taken (1-4, 9 for Puerto Rico)  
ST: State Code (01-72, skips some)  
NP: Number of persons associated with this housing record  
ACR: Lot size (1-3, acreage based on number)  
BDSP: Number of Bedrooms (00-99)  
RMSP: Number of Rooms (00-99)  
VALP: Property Value (0000001 – 9999999)  
VEH: Vehicles (0-6)  
FES: Family type and employment status (1-8, Married, unmarried)  
FINCP: Family Income (-0059999 – 99999999)  
FPARC: Family presence and age of related children (1-4, children types)  
HHL: Household Language (1-5, language codes)  
HHT: Household / family type (1-7, code type of family)  
HINCP: Household Income (-0059999 – 99999999)  
LNGI: Limited English Speaking Household (1, at least one person speaks very well, 2, no English)  
MV: When moved into house or apartment (1-7, binned time slots)  
NPF: Number of persons in family (02-20)  
PARTNER: Unmarried partner household codes (0-4)  
SVAL: Specified owner unit (0,1, home by acreage)

Workflow Steps –   
1. Read Census CSV file  
2. Apply column filter to only filter columns that we want to analyze / are relevant to study

Naïve Solution: Income

  
ROC Curve for Naïve solution