

# MF004 – semestrální projekt

Marek Bryša, 323771

Brno, 8. ledna 2012

## 1 Data

### 1.1 Úvod

Zdrojem dat je soubor `MortgageDefaulters.xls`. Obsahuje údaje o 15153 klientech banky, kteří mají hypoteční úvěr. Cílem je identifikovat klienty s vysokou pravděpodobností, že nebudou splácet. K dispozici jsou tyto proměnné:

**Bo\_Age** Věk klienta

**Ln\_Orig** Výše půjčky v USD

**Orig\_LTV\_Ratio\_Pct** Poměr půjčky a nákupní ceny domu

**Credit\_score** Klientovo credit score

**First\_home** První klientův nákup dům (Y/N)

**Tot\_mthly\_debt\_exp** Klientův celkový měsíční výdaj na půjčku

**Tot\_mthly\_incm** Klientův celkový měsíční příjem

**orig\_apprd\_val\_amt** Odhad ceny domu v době žádosti

**pur\_prc\_amt** Kupní cena domu

**DTI\_ratio** Poměr nákladu na dluh a příjmu klienta – měsíčně

**StatusCurrent** Stav půjčky

**OUTCOME** Binary binární verze stavu: 0=v pořádku, 1=default

**StateUS** Stát USA, kde se dům nachází

**Median.state.inc** Střední příjem domácnosti v daném státě (2002-2004)

**UPB>Appraisal** Je půjčka vyšší než odhad? 0=ne, 1=ano

## 1.2 Popisné statistiky

V následující tabulce jsou uvedeny popisné statistiky intervalových proměnných:

	Mean	Std.Err.	Median	Std.Dev.	Min	Max
Bo_Age	36.79	0.08	37	10.03	18	99
Ln_Orig	153467.57	555.42	141500	68370.61	19600	599000
Orig_LTV_Ratio_Pct	93.08	0.07	95	8.85	20	111
Credit_score	687.67	0.51	688	62.90	440	999
Tot_mthly_debt_exp	1745.46	8.85	1578	1089.20	0	17225
Tot_mthly_incm	5024.71	23.98	4632	2952.16	500	65000
orig_apprd_val_amt	170661.44	664.31	154000	81775.07	0	870000
pur_prc_amt	164681.56	647.62	148650	79719.84	20000	870000
DTI Ratio	0.37	0.00	0	0.18	0	3
Median.state.inc	44945.07	44.12	43988	5431.51	32589	57352

Je zřejmé, že data obsahují některé podezřelé hodnoty. Bude nutné provést transformaci.

## 1.3 Transformace dat