

1. Description de la base

The data set HMEQ reports characteristics and delinquency information for 5,960 home equity loans. A home equity loan is a loan where the obligor uses the equity of his or her home as the underlying collateral. The data set has the following characteristics:

- BAD: 1 = applicant defaulted on loan or seriously delinquent; 0 = applicant paid loan
- LOAN: Amount of the loan request
- MORTDUE: Amount due on **existing** mortgage
- VALUE: Value of current property
- REASON: DebtCon = debt consolidation; HomImp = home improvement
- JOB: Occupational categories
- YOJ: Years at present job
- DEROG: Number of major derogatory reports
- DELINQ: Number of delinquent credit lines
- CLAGE: Age of oldest credit line in months
- NINQ: Number of recent credit inquiries
- CLNO: Number of credit lines
- DEBTINC: Debt-to-income ratio

Les données vous sont fournies sous format CSV.

2. Que faut-il faire ?

Proposer un modèle expliquant la variable « BAD » par les autres variables disponibles et éventuellement de nouvelles variables créées. L'objectif est de fournir un rapport **structuré** et convaincant. Il doit être riche en « Analyses statistiques » utiles au problème posé. La qualité de la rédaction (la syntaxe, les fautes de grammaire et d'orthographe, etc.) primera sur la quantité.

3. Livrables attendus

Un rapport .doc synthétique (25 pages max hors annexes) et les programmes informatiques. Les programmes informatiques doivent être commentés.

Le projet peut être réalisé sous les logiciels suivants : Python, R et/ou SAS.

Les travaux sont à envoyer sur : ibrahim.toure@univ-paris1.fr en mettant en Objet « Groupe n°XX – Scoring – IMSD ».

4. Date limite

Les projets sont à rendre pour le 22/01/2025