1. **Description de la base**

L'ensemble de données HMEQ rapporte les caractéristiques et les informations sur les impayés pour 5 960 prêts sur valeur domiciliaire. Un prêt sur valeur domiciliaire est un prêt dans lequel le débiteur utilise la valeur nette de son logement comme garantie sous-jacente.

L'ensemble de données présente les caractéristiques suivantes :

* BAD :
* 1 = demandeur en défaut de paiement ou gravement en souffrance ;
* 0 = prêt payé par le demandeur
* LOAN (Numérique) : Montant de la demande de prêt
* MORTDUE (Numérique) : Montant dû sur l'hypothèque existante
* VALUE (Numérique) : Valeur de la propriété actuelle
* REASON :
* DebtCon = consolidation de la dette
* HomeImp = amélioration de l'habitat
* JOB : Catégories professionnelles
* YOJ : Nombres d’années dans l'emploi actuel
* DEROG : Nombre de signalements dérogatoires majeurs
* DELINQ : Nombre de lignes de crédit en souffrance
* CLAGE : Âge de la ligne de crédit la plus ancienne en mois
* NINQ : Nombre de demandes de crédit récentes
* CLNO : Nombre de lignes de crédit
* DEBTINC : Ratio dette/revenu

1. **Que faut-il faire ?**

Proposer un modèle expliquant la variable « BAD » par les autres variables disponibles. J’ai fait le choix cette année de ne pas donner de consignes particulières. 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.

1. **Livrables attendus**

Un rapport .doc synthétique (25 pages max hors annexe) et les programmes informatiques (il n’est pas nécessaire de les mettre en annexe).