# Rapport TP1

# Thibault Cart & Rami Albadri

# Introduction

Le problème de ce TP est qu’on ne sait pas quels sont les facteurs les plus importants qui déterminent si une personne gagne plus ou moins de 50K$

# Analyse préliminaire

Il y a 15 variables La variable cible est « income » car c’est ce qu’on recherche et elle est qualitative car 2 possibilités, soit plus que 50k ou moins que 50k

| **Attribute** | **Type** | **Justification** |
| --- | --- | --- |
| age | Quantitative | Valeur numérique continue (âge), adaptée aux calculs statistiques. |
| workclass | Qualitatif | |  | | --- | |  |  |  | | --- | | Catégories d’emploi (p.ex., privé, fonction publique d'État, etc) | |
| fnlwgt | Quantitative | |  | | --- | |  |  |  | | --- | | Score de pondération démographique. | |
| education | Qualitatif | Niveaux d’éducation non ordonnés (p.ex., licence, etc) |
| education-num | Quantitative | Version numérique ordonnée du niveau d’éducation. |
| marital-status | Qualitatif | Catégories d’état civil (marié, divorcé, etc.). |
| occupation | Qualitatif | Type de profession (vente, support technique, etc.). |
| relationship | Qualitatif | |  | | --- | |  |  |  | | --- | | Rôle au sein du foyer (époux, hors famille, etc.). | |
| race | Qualitatif | Groupe ethnique (blanc, noir, etc.). |
| sex | Qualitatif | Sexe biologique (homme, femme). |
| capital-gain | Quantitative | Gains en capital en valeur numérique. |
| capital-loss | Quantitative | Pertes en capital en valeur numérique. |
| hours-per-week | Quantitative | Nombre d’heures travaillées par semaine. |
| native-country | Qualitatif | Country of origin (United-States, India, etc.). |
| income | Qualitatif | |  | | --- | |  |   Classe cible (binaire : <=50k ou >50k |

# Certaines variables doivent-elles être supprimées ?

# Y a-t-il des données manquantes ?

# Variable cible et sa distribution