特征

- age: 一个整数，表示被调查者的**年龄**。
- workclass: 一个类别变量表示被调查者的通常**劳动类型**，允许的值有 {Private, Self-emp-not-inc, Self-emp-inc, Federal-gov, Local-gov, State-gov, Without-pay, Never-worked}
- education_level: 一个类别变量表示**教育程度**，允许的值有 {Bachelors, Some-college, 11th, HS-grad, Prof-school, Assoc-acdm, Assoc-voc, 9th, 7th-8th, 12th, Masters, 1st-4th, 10th, Doctorate, 5th-6th, Preschool}
- education-num: 一个整数表示在学校**学习了多少年**
- marital-status: 一个**婚姻类别变量**，允许的值有 {Married-civ-spouse, Divorced, Never-married, Separated, Widowed, Married-spouse-absent, Married-AF-spouse}
- occupation: 一个类别变量表示一般的**职业领域**，允许的值有 {Tech-support, Craft-repair, Other-service, Sales, Exec-managerial, Prof-specialty, Handlers-cleaners, Machine-op-inspct, Adm-clerical, Farming-fishing, Transport-moving, Priv-house-serv, Protective-serv, Armed-Forces}
- relationship: 一个类别变量表示**（自身所在）家庭情况**，允许的值有 {Wife, Own-child, Husband, Not-in-family, Other-relative, Unmarried}
- race: 一个类别变量表示**人种**，允许的值有 {White, Asian-Pac-Islander, Amer-Indian-Eskimo, Other, Black}
- sex: 一个类别变量表示**性别**，允许的值有 {Female, Male}
- capital-gain: 连续值（**资本收益**）。
- capital-loss: 连续值（**资本损失**）。
- hours-per-week: 连续值（**每周工作时间**）。
- native-country: 一个类别变量表示**原始的国家（祖国）**，允许的值有 {United-States, Cambodia, England, Puerto-Rico, Canada, Germany, Outlying-US(Guam-USVI-etc), India, Japan, Greece, South, China, Cuba, Iran, Honduras, Philippines, Italy, Poland, Jamaica, Vietnam, Mexico, Portugal, Ireland, France, Dominican-Republic, Laos, Ecuador, Taiwan, Haiti, Columbia, Hungary, Guatemala, Nicaragua, Scotland, Thailand, Yugoslavia, El-Salvador, Trinadad&Tobago, Peru, Hong, Holand-Netherlands}

目标变量

- income: 一个**类别变量**，表示收入属于那个类别，允许的值有 {<=50K, >50K} - **这貌似是一个分类的问题。**