数据清洗， 是整个数据分析过程中不可缺少的一个环节，其结果质量直接关系到模型效果和最终结论。在实际操作中，数据清洗通常会占据分析过程的50%—80%的时间。国外有些学术机构会专门研究如何做数据清洗，相关的书籍也不少。(Data cleaning)