

# 数据加载

---

- 输入输入分类
  - 本地文件
  - 数据库
  - 网络资源
- 数据输入
  - 参数
    - skiprows : 跳过行
    - sep/delimiter:指定分隔符
    - header : 用作列名的行号, 默认为0, 没有的话设置None
    - index\_col:指定行索引的列名或者编号
    - na\_values:指定缺失值映射, 赋值字典或者数组
    - nrows : 读取行数
    - chunksize:按块读取
  - read\_csv():默认分隔符','
  - read\_table():默认分隔符'\t'
  - read\_fwf: 读取定宽列格式
  - read\_clipboard:读取剪贴板数据
  - read\_excel: xls或者xlsx
  - read\_json: json数据
  - read\_hdf:HDF5格式, 适用一次写多次读
  - read\_html:依赖lxml、Beautiful Soup,默认解析<table>标签的数据k
  - read\_msgpack
  - read\_pickle: 二进制序列, pickle序列
  - read\_sas
  - read\_sql:使用sqlalchemy访问
  - read\_stata
  - read\_feather
  - ExcelFile模块
  - csv模块
    - 参数
      - delimiter: 分割字符
      - lineterminator:行结束符
      - quotechar : 引用符号
      - skipinitialspace:忽略分割符后的空白字符
    - reader
    - writer
- 数据输出
  - 参数

- sep:
- index:是否导出行标签，默认为True
- header：是否导出列标签，默认为True
- columns：导出的列名
- na\_rep:默认NaN值导出为空字符串，可以修改
- 函数：
  - to\_csv():
  - to\_json
  - to\_pickle
  - to\_excel

以上内容整理于 [幕布](#)