

数据清洗

- 数据重排
 - 缺失数据NaN，哨兵值
 - 参数：
 - axis:
 - how
 - thresh：要求保留至少n个非NaN值
 - inplace：修改原始数据
 - method
 - limit
 - 函数
 - fillna:默认返回新对象
 - dropna: 返回非空的记录，等价于data[data.notnull()]
 - isnull
 - notnull
- 数据转换
 - 参数：
 - keep
 - replace
 - index/columns 也适用这些函数,字典的作用是精确匹配
 - 函数
 - duplicated:重复值的bool索引，默认判断全部列
 - drop_duplicates:默认保留第一个出现的集合
 - map:
 - replace
 - rename
 - cut：分组，返回Categorical对象，连续的数据被离散化，拆分成面元（分组），默认面元定长
 - right = False: 左闭右开
 - labels：设置分组名称
 - precision：设置小数位数
 - qcut：通过样本分位数分组，每组数量不一定相同，但是大小基本相等
 - describe():统计汇总
 - random.permutation：随机重排序
 - take
 - sample:n个随机数，允许重复样本
 - get_dummies：分类列表转换，将值转换成列，构建指标矩阵
 - join：合并
 - 字符串操作

- split
- strip
- join
- in
- index : 找不到引发异常
- find : 找不到返回-1
- count :
- replace
- lower
- upper
- 正则表达式 : re模块

以上内容整理于 [幕布](#)