

基于Python的 数据分析与可视化

第八讲 Pandas

刘子瑜 副研究馆员

liuziyu@cau.edu.cn 中国农业大学图书馆情报研究中心

<https://github.com/CAU-LiuZiyu/Data-Analysis-with-Python>

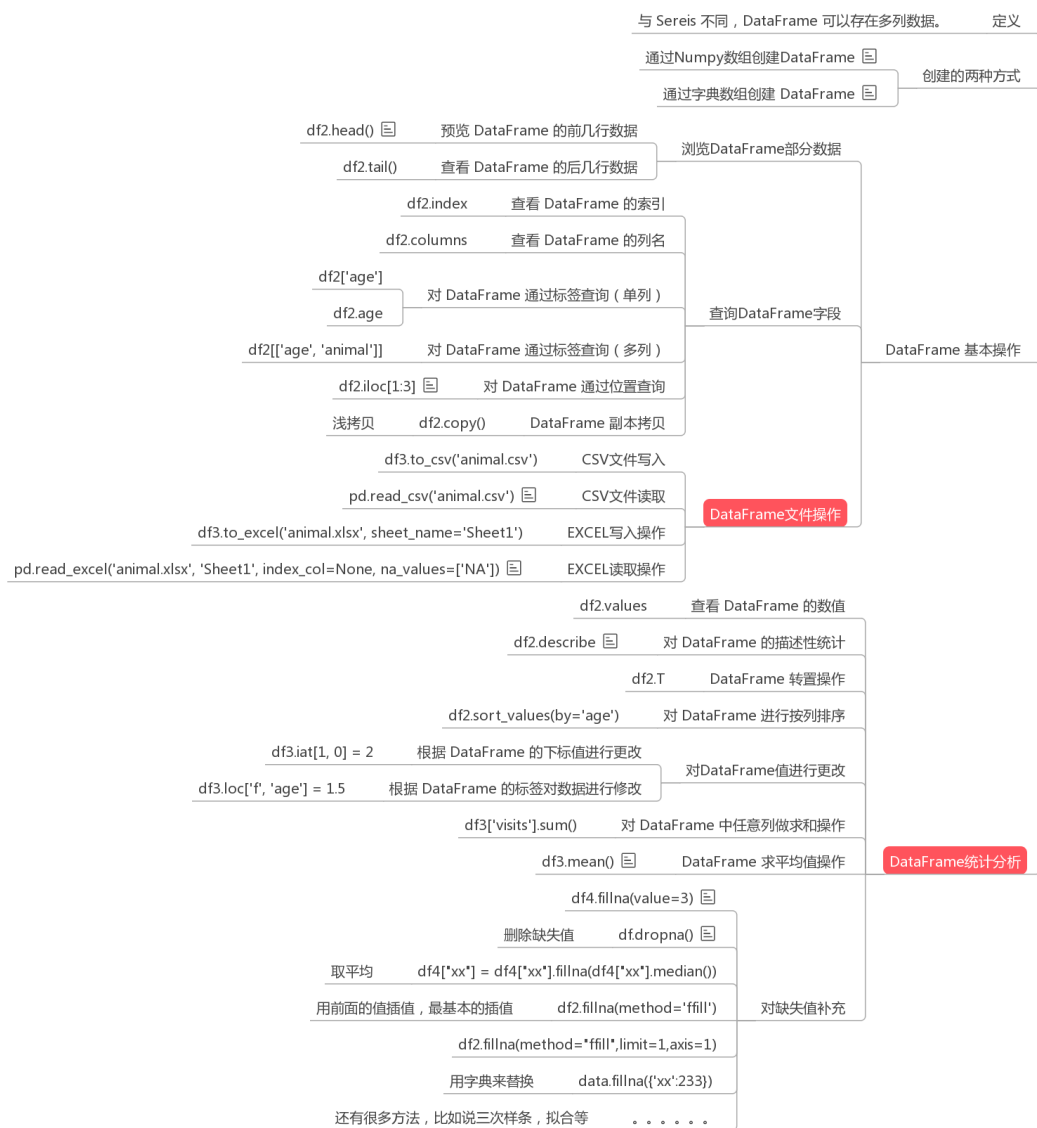
第三次作业（5.10提交）

探索Kaggle、天池、kesci、datafountain等网站。

完成天池AI学习的“红酒数据集分析”，形成一份简单的数据分析报告。
要求内容包括：

- （1）主要分析的问题是什么？
- （2）数据基本情况如何？
- （3）数据分析的结果能否回答或确认研究问题？

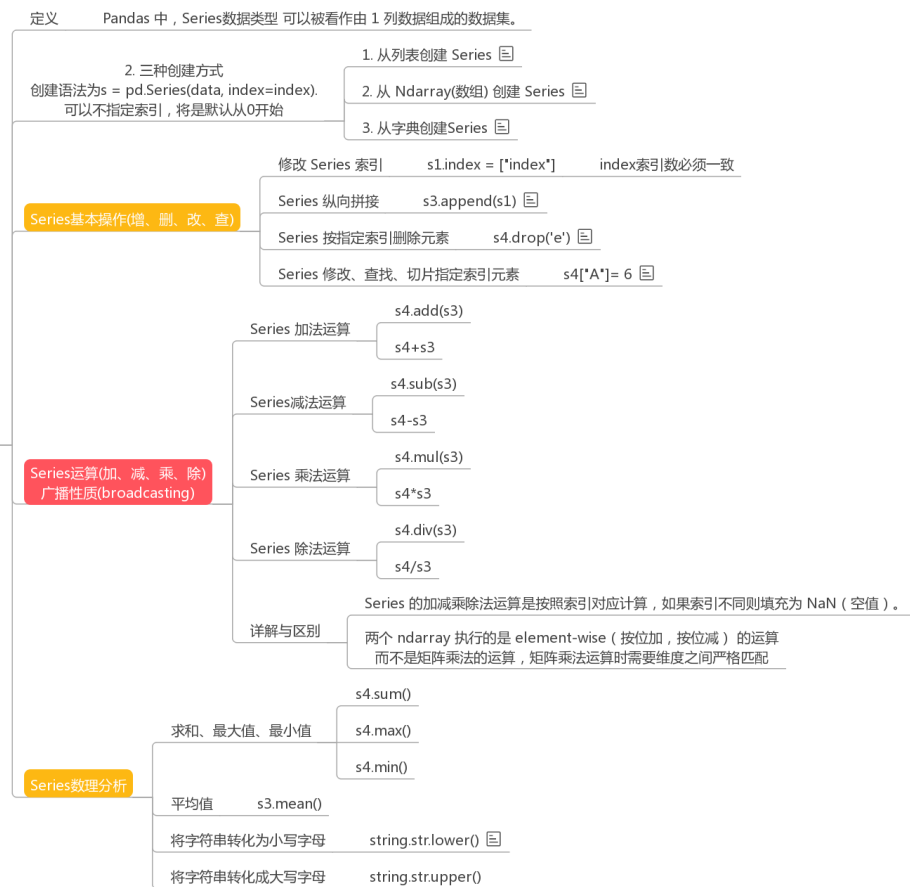
报告请控制在2000字以内（含代码、图表），正文和代码以ipynb文件提交。



DataFrame

pandas总结

Series



pandas总结

Series

定义 Pandas 中，Series数据类型 可以被看作由 1 列数据组成的数据集。

2. 三种创建方式
创建语法为 `s = pd.Series(data, index=index)`.
可以不指定索引，将是默认从0开始

- 1. 从列表创建 Series
- 2. 从 Narray(数组) 创建 Series
- 3. 从字典创建Series

Series基本操作(增、删、改、查)

- 修改 Series 索引 `s1.index = ['index*']` index索引数必须一致
- Series 纵向拼接 `s3.append(s1)`
- Series 按指定索引删除元素 `s4.drop('e')`
- Series 修改、查找、切片指定索引元素 `s4['A']= 6`

Series运算(加、减、乘、除) 广播性质(broadcasting)

- Series 加法运算
 - `s4.add(s3)`
 - `s4+s3`
- Series减法运算
 - `s4.sub(s3)`
 - `s4-s3`
- Series 乘法运算
 - `s4.mul(s3)`
 - `s4*s3`
- Series 除法运算
 - `s4.div(s3)`
 - `s4/s3`

Series 的加减乘除法运算是按照索引对应计算，如果索引不同则填充为 NaN（空值）。
两个 ndarray 执行的是 element-wise（按位加，按位减）的运算
而不是矩阵乘法的运算，矩阵乘法运算时需要维度之间严格匹配

Series数理分析

- 求和、最大值、最小值
 - `s4.sum()`
 - `s4.max()`
 - `s4.min()`
- 平均值 `s3.mean()`
- 将字符串转化为小写字母 `string.str.lower()`
- 将字符串转化成大写字母 `string.str.upper()`

