



pandas入门教程

库 数据科学

处理Dataframe数据
对应字典 numpy对应list

读取数据excel,

pd.read_csv("文件路径")

希望反斜杠不转义

pd.read_excel("文件路径", sheet_name="sheet1")

存储为csv

type()

DataFrame

常用基本函数

titanic.head() 前8行

titanic.tail() 后10行

titanic.dtypes 数据类型

titanic.info() 更具体的信息

DataFrame数据选择

ages = titanic["age"]

其数据类型是series

DataFrame是基于Pandas的库

to titanic["age"].shape 大小

选择两行

age-sex = titanic[["age", "sex"]]

数据筛选

above_age = titanic[titanic["age"] > 35]

class_23 = titanic[titanic["Pclass"].isin([2, 3])]

选出年龄不为缺失值的信息

age_no_na = titanic[titanic["Age"].notna()]

判断是否为空

只选出35以上乘客的姓名

above_35 = titanic[titanic["Age"] > 35, "Name"]

对第10行到第25行和第3列到5列

titanic.iloc[9:25, 2:5]

创建新列

改变列名

统计各种信息