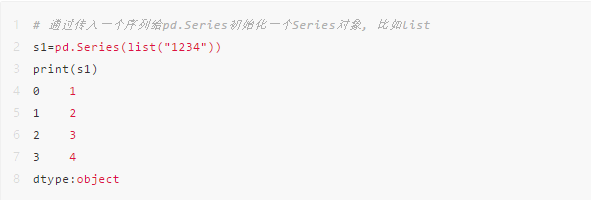
**Pandas-Learing**

Pandas是基于numpy的一种工具，提供了大量快速处理数据的函数和方法。

Pandas主要有两张数据对象：Series、DataFrame

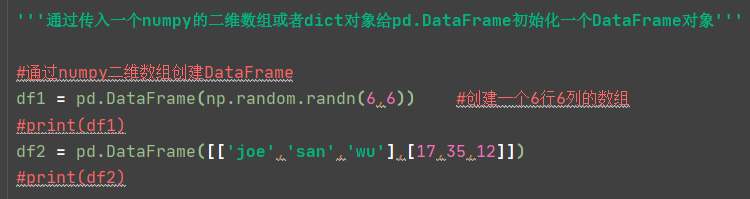
1. Series

Series是一种带有索引的序列对象

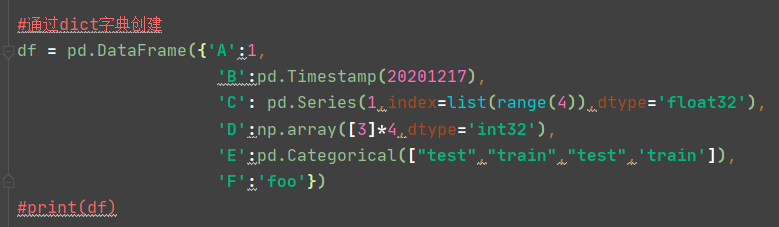


1. DataFrame：一个表格型的数据结构，包含有一组有序的列，每列可以是不同的值类型(数值、字符串、布尔型等)，DataFrame即有行索引也有列索引，可以被看做是由Series组成的字典。

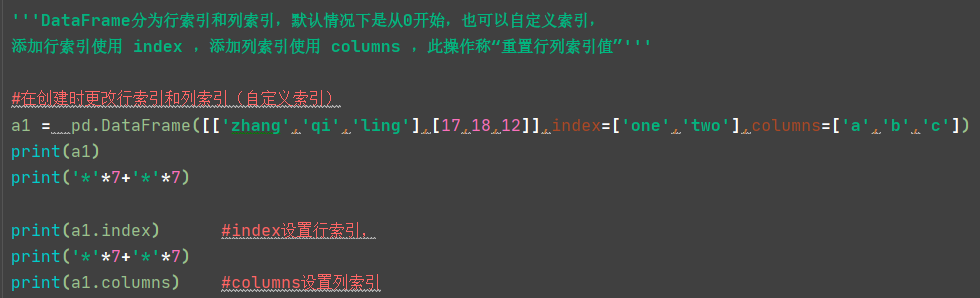
（1）通过二维数组创建



（2）通过字典创建

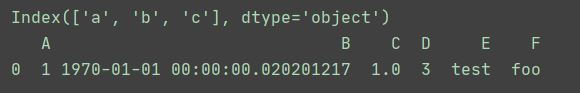


（3）DataFrame索引：

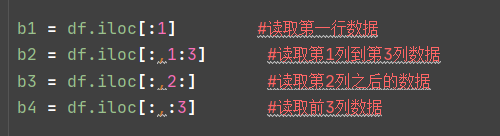


读取数据（.iloc方法可以通过列号索引数据）：





df.iloc[:1]读取第一列  
df.iloc[:,1:3]读取第1列到第3列  
df.iloc[:,2:]读取第2列之后的数据  
df.iloc[:,:3]读取前3列数据



1. Pandas常用函数

计算每列数据对象是数值的count, mean, std, min, max：

df.describe()

单独统计Open列的平均值：

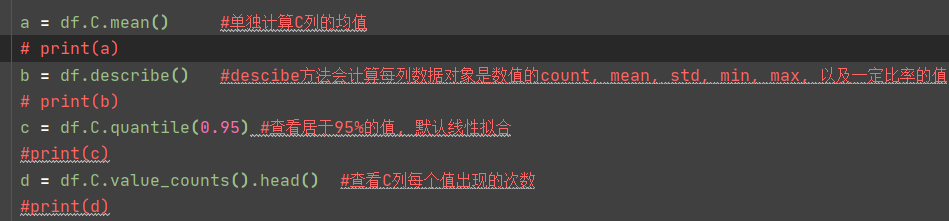
df.Open.mean()

查看居于95%的值, 默认线性拟合：

df.Open.quantile(0.95)

查看Open列每个值出现的次数：

df.Open.value\_counts().head()



1. pandas的绘图是使用matplotlib，如果想要画的更细致, 可以使用matplotplib