【项目05】 多场景下的算法构建

课程数据：

某公司A,B产品在2018年1,2,3月的销量数据，数据格式为xlsx

作业要求：

1、批量读取数据，并输出以下信息

（1）数据量

（2）数据字段columns

（3）输出每个文件分别有多少缺失值

要求：

① 创建独立函数，从读取数据到以上输出要求

② 运行代码多次调用创建函数，对数据进行批量处理

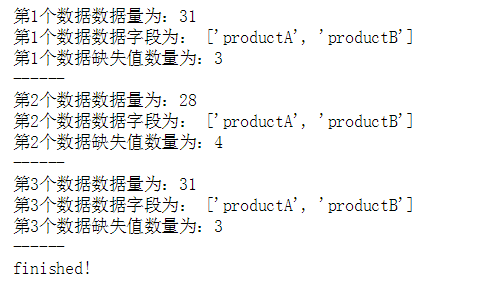
提示：

① 将课程excel数据放入单独文件夹(建议英文路径)，函数中通过input来输入excel数据所在文件夹路径，再通过遍历来读取文件

② pd.read\_excel()中用“index\_col”参数，将第一列变为index

③ os.walk(path) → 返回路径中的信息及文件，结果为一个生成器

④ 定位缺失值位置：data[data.isnull().values == True]



2、批量读取数据，用均值填充缺失值数据，并完成以下计算及图表制作

（1）读取数据并用均值填充缺失值；对“日期”字段进行时间序列处理，转换成日period ，最后输出三个Dataframe文件data1,data2,data3

（2）分别计算data1，data2，data3中A,B产品的月总销量，并绘制多系列柱状图，存储在对应的图片文件夹路径

（3）分别计算A产品在每个月中哪一天超过了月度80%的销量，输出日期

要求：

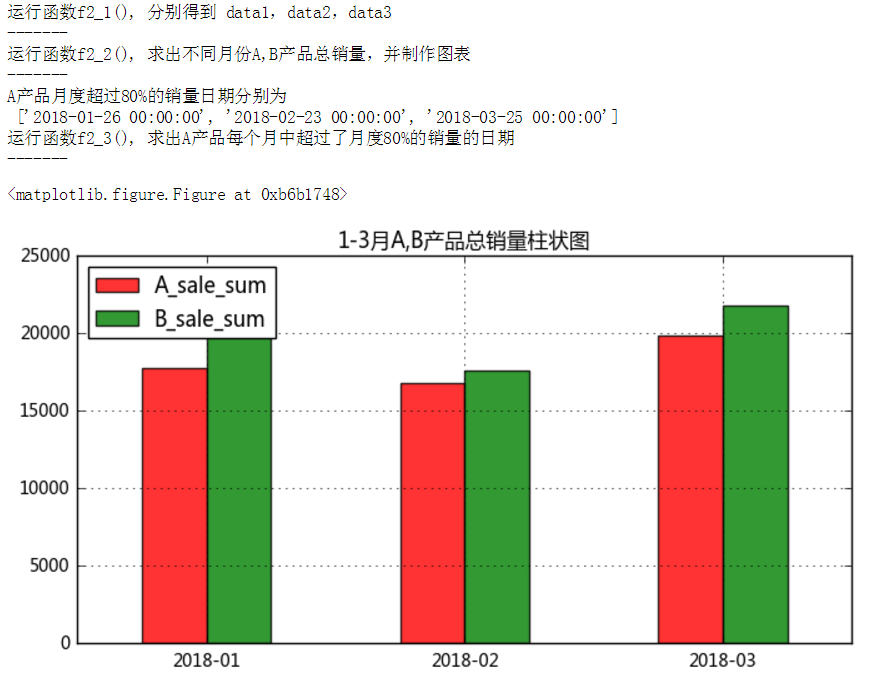
① 分别创建四个函数，对应完成上述需求

1. 数据文件夹路径和图片存储路径不要相同

提示：

1. 时间序列转换方法提示：pd.to\_period

② 时间戳Timestampe转化成字符串用str()方法



3、读取数据并合并，做散点图观察A,B产品销量，并做回归，预测当A销量为1200时，B产品销量值

（1）读取数据删除缺失值；对“日期”字段进行时间序列处理，转换成日period ，合并三个月数据，输出data；

（2）针对A产品销量和B产品销量数据做回归分析，制作散点图并存储，并预测当A销量为1200时，B产品销量值

要求：

① 分别创建两个函数，对应完成上述需求

② 数据文件夹路径和图片存储路径不要相同

提示：

1. 用pd.dripna方法去掉缺失值，注意inplace参数

