

【Task5精选问题】

Q1: pivot方法可否看做excel的数据透视表操作？若不能两者有什么区别？

excel的数据透视表功能很强大，pd.pivot()只实现了一个功能——将两列数据变为二维表格。

Q2: 数据合并要是两个表数据量都很大的话，内存不够怎么办？

数据合并是在数据分析处理最消耗空间和时间的操作。

因此为减少消耗，在数据合并前分别对两个表格进行条件和属性选择，减少单个表格的数据。

Q3: join 和 merge的区别是什么？

```
pd.merge(df1, df2)
```

```
df1.join(df2)
```

二者都执行连接操作，本质等价。

merge()比join()多了df的索引控制，而这个可以通过删除列/增加列的方式实现等效。

Q4: 对于大型的数据 pandas的数据合并和层次话索引还可以很好操作吗？感觉一两个标签就已经很晕了

Pandas基于NumPy实现，便捷高效。

数据分析可没有一把梭哈的方法，你预先得规划数据处理过程，再进行编码。即思维运算->Python实现