一分析并读取给定的数据文件: 深圳市一卡通2018年8月31日-9月1日刷卡数据

导入pandas 库，通过pd.read\_csv 来导入数据

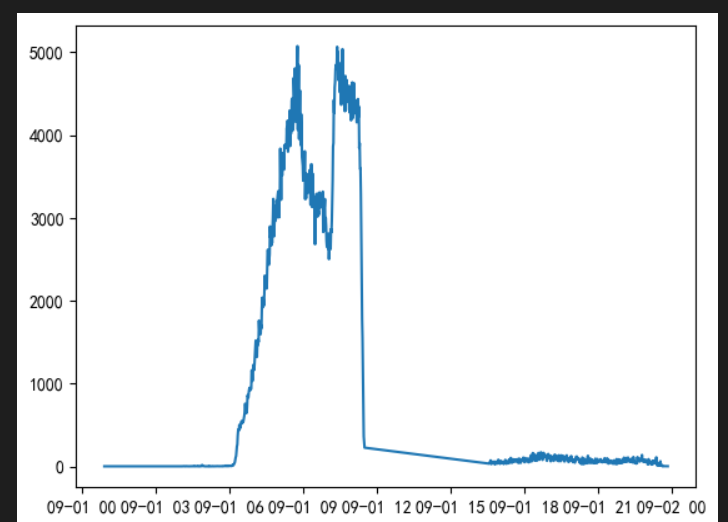
二按照一定规则清洗数据;

将导入的两个数据表，按列合并，并显示缺失值，发现有**car\_no station conn\_mark deal\_money equ\_no 这些属性列有缺失值，但缺失值很少，对这些列进行删除，后将deal\_date，进行apply函数调用lambda 函数，lambda x:'2018/9/1 '+x[9:]+':00'，将deal\_date列属性进行时间格式化，然后转化为datetime，然后将数据进行按这列进行sort\_values，并重新设置索引**



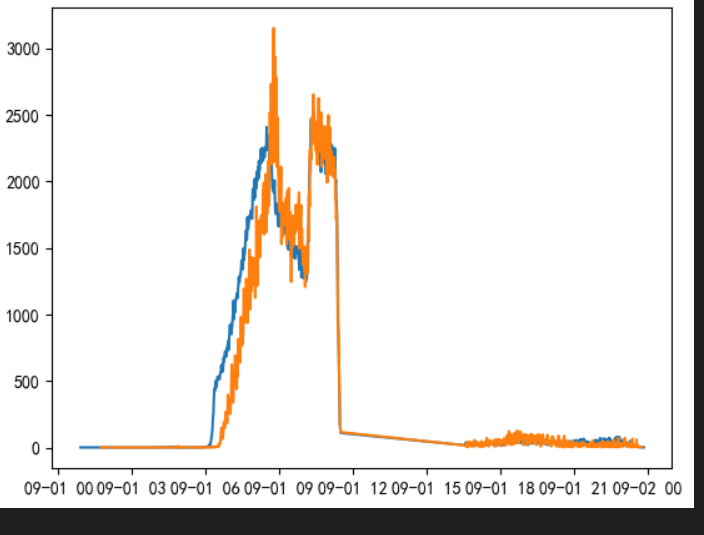
**三以时间为x轴，人数为y轴，勾勒出整体人流量曲线图**

**通过分析清洗后的数据，x轴为数据的索引，y轴为对deal\_date 进行value\_count，并按照时间从小到大排序的人员流量。导入matplotlib，绘制整体人流量曲线图**



**四以时间为x轴，人数为y轴，按照不同线路勾勒出曲线图，反映出线路之间差异**

**x 轴为清洗后的数据的索引，y轴为deal\_type进行value\_count,统计地铁出站和地铁入站的每个时间段的人数，并按照时间从小到大排序的人员流量，绘制出地铁出战和地铁入站的人数曲线图。**



**五以时间为x轴，人数为y轴，按照不同站点勾勒出曲线图，反映出站点之间差异**

**x 轴为清洗后的数据的索引，y轴为station进行value\_count,统计每一个不同站台的每个时间段的人数，并按照时间从小到大排序的人员流量，总共有170个站台，挑选十个站台，绘制人数曲线图**

