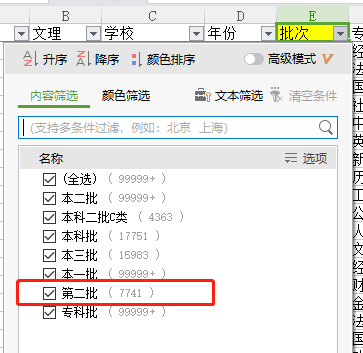
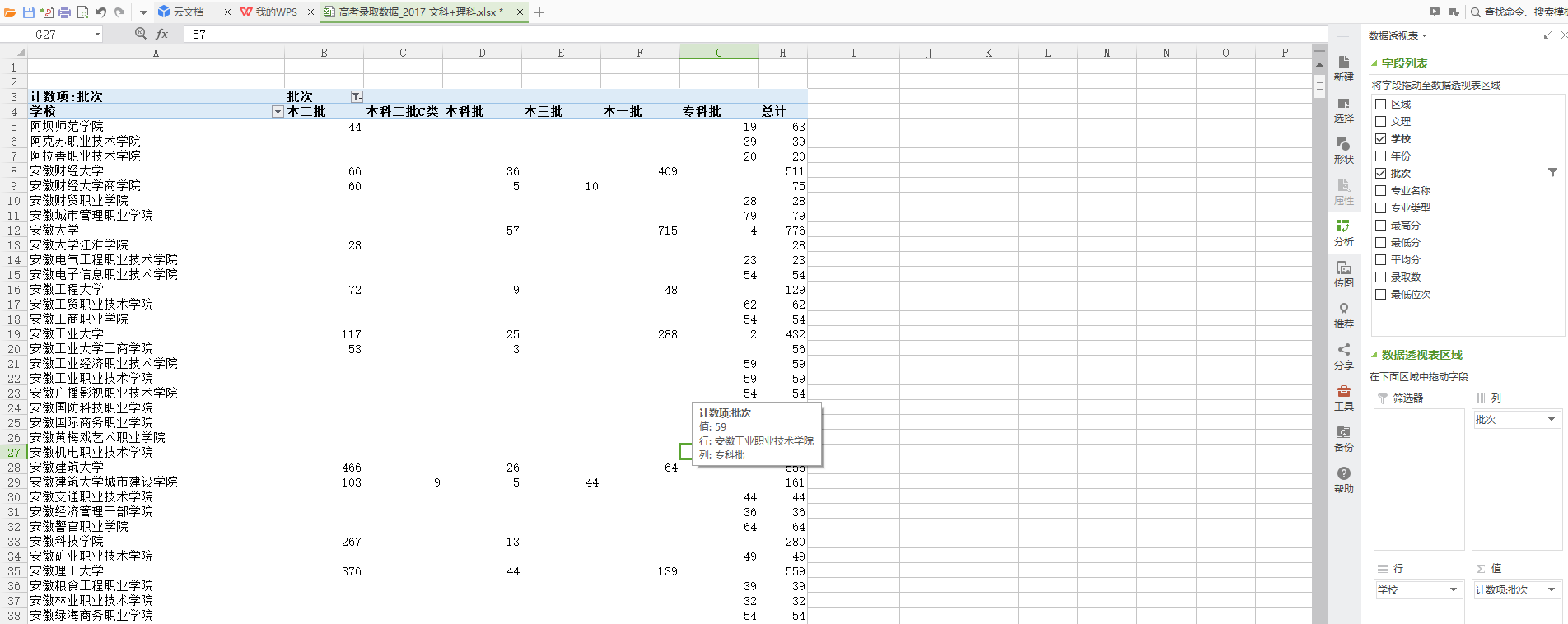
1.将文科和理科两个表格拼接为一张表，筛选“批次”字段，将其中的“第二批”改为“本二批”：

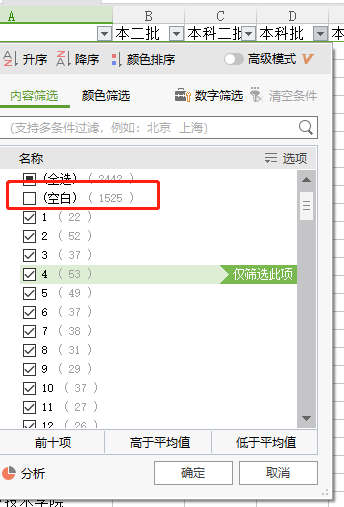


2.新建数据透视表，统计各学校的各批次录取人数：

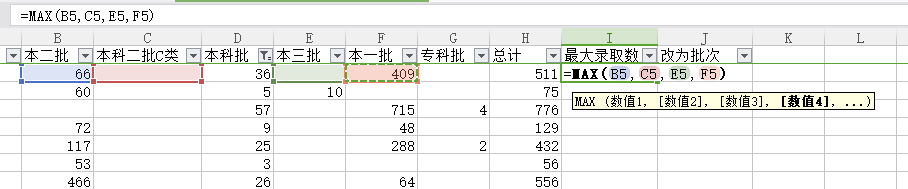


3.将透视表复制为数值，删除总计行。筛选字段“本科批”为非空的数据，并添加3个字段：“字段拼接”、“最大录取数”、“改为批次”：

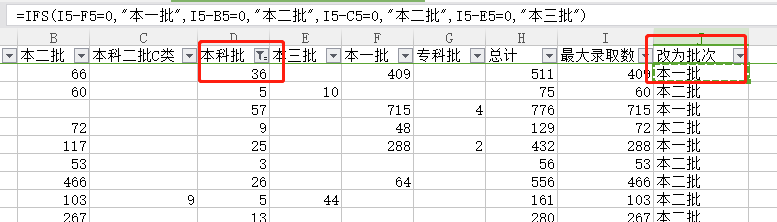




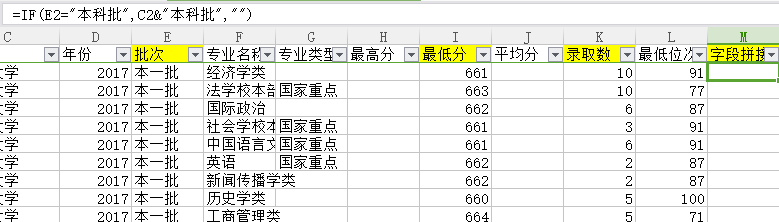
提取各个本科批次的最大值：

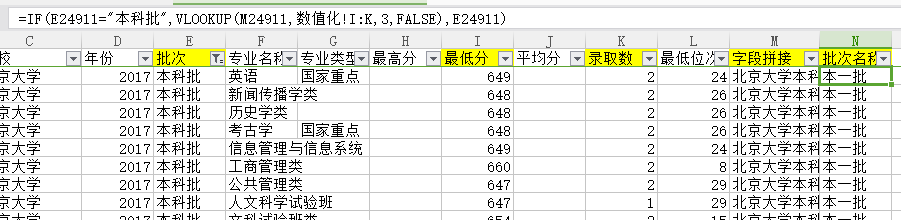


在“改为批次”列，将本科批有值的数据批次改为其所在的行的数据里其他各本科批次录取人数最大的那个批次名称：

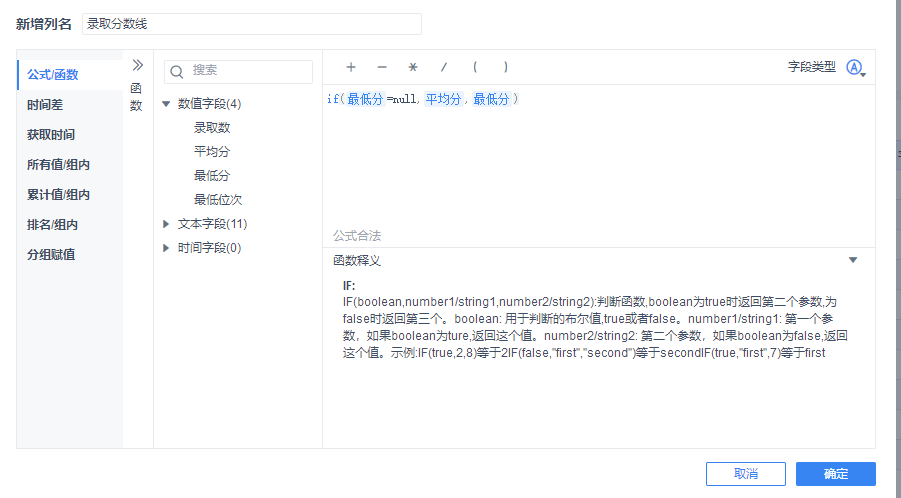


4.返回数据源表，新增两个字段：“数据拼接”和“批次名称”，并将字段“专业名称”中有后缀的处理下，前缀名称一致的统一命名，选中所有数据去重复项。





5.将以上处理好的数据源导入BI业务包，并创建自主数据集《高考录取数据分析》，由于最低分字段有缺失值，新增列“录取分数线”，当最低分非空时取值最低分，最低分为空值时则取值平均分：



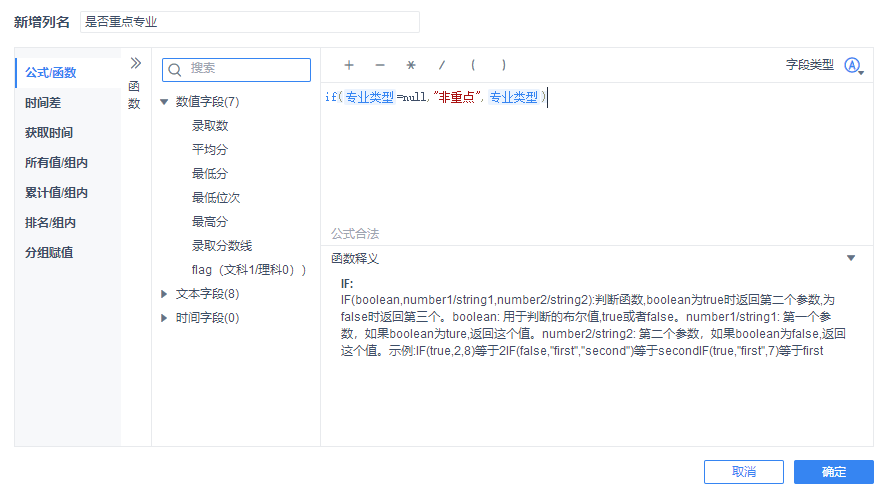
1. 过滤掉录取数为空值的数据：



1. 新增列：







1. 据此自助数据集建立仪表板“2017年高考录取数据分析”：

