

- 1、结合数据解读 4 月 4 日课堂程序；
- 2、将下述三个文件：

openData_Order201608_afterClear.csv、

openData_Order201609_afterClear.csv、

openData_Order201610_afterClear.csv

的数据导入数据库(数据库名称为：saleData.db)，其中 DAY_CODE 是交易日期编码，数据库中同样有该属性（或名字段、表头、列名称）、同时建立 YEAR、MONTH 属性，将 DAY_CODE 中的年份和月份数据放入对应属性；数据库中建立 AREA_NAME 属性，其数据来源于

SALEORG_CODE，其采集到数据库的编码格式为“QX”+SALEORG_CODE 对应值的最后三位，一共 5 位组成；CUST_CODE 代表客户编码；

ITEM_CODE 代表规格编码，其中前六位代表品牌，基于该数据建立

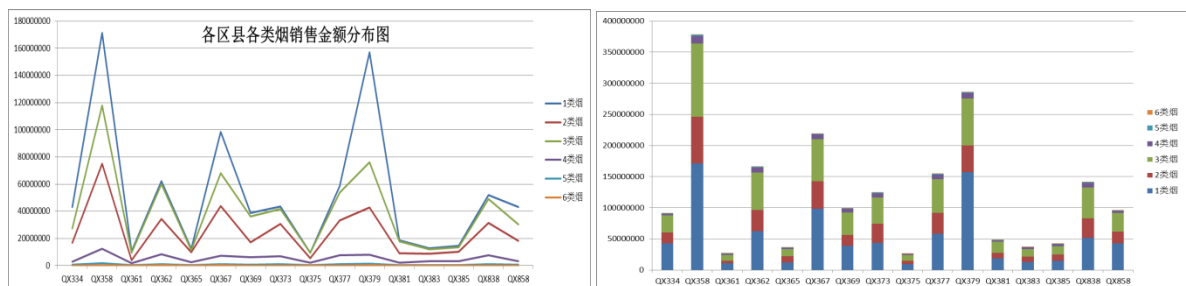
BRAND_CODE，数据来自 ITEM_CODE 的前六位；CUST_TYPE 代表客户

类别；KIND 代表规格对应的档次(价格高档次编号低，1 代表最高档)；

PRICE_TRADE 代表批发价；PURCH_QTY 代表采购数量；PURCH_TAX_AMT 代表采购金额(或者销售金额)。最终数据中有 DAY_CODE、YEAR、MONTH、

AREA_NAME、CUST_CODE、ITEM_CODE、BRAND_CODE、CUST_CODE、CUST_TYPE、KIND、PRICE_TRADE、PURCH_QTY、PURCH_TAX_AMT 等属性；

- 3、利用 saleData.db 中的数据，输出每个区县每类烟的销售金额，输出文件名为 areaKind.txt，每个数据之间的分隔符为英文逗号。在 Excel 中，



打开该文件，形成如上图所示的折线统计图。再次利用该数据，形成柱状堆积图；

- 4、利用该数据库中的数据，统计每个月商户数量、每个商户平均采购金额，操作过程第 3 题，并利用 Excel 图形化；
- 5、利用该数据库中的数据，统计出每个月销售规格数量、平均每个规格销售金额、平均每个金额的单价，操作过程第 3 题，并利用 Excel 图形

化；

- 6、 利用数据库中的数据，将商户销售金额从高到低排名，计算出每个商户销售金额占总销售的占比，计算出累积销售金额达到 80%时有多少个商户，商户占比是多少？利用 Excel 图形化；
- 7、 利用数据库中的数据，计算出每个商户销售金额的平均金额、标准差等，利用 Excel 图形化；
- 8、 利用数据库中的数据，将商户销售金额从高到低排名，根据数据分为 5 等份，计算出最高 20%销售总金额与其他各等份比值，并利用 Excel 图形化；
- 9、 根据数据，进一步发挥；
- 10、 作业完成完成 2018-04-16 18:00 前。