【参考答案】

1. 对数据集进行描述性统计

收入字段有缺失值13个（2分），可处理也可以不处理



大部分变量都存在异常值，暂不做处理（2分）



连续变量的描述性统计，所有连续字段均有一定程度的右偏（2分）



量纲不一致（1分）

对零售店客户进行细分可以参考RFM模型，所以可以选择与RFM三个维度相关的指标进行（言之有理即可,共3分，1个维度的说明给1分）

（2）该步骤包括以下处理：

构造新变量（3分）

**R：构造新变量购买距今天数**，购买距今天数=提取日期-最近购买日期

**M：构造花费总金额，**花费总金额=葡萄酒花费+水果花费+肉类花费+水产品花费+甜品花费

**F：构造总购买次数，**总购买次数=网站购买次数+实体店购买次数

基础数据如下：（3分）



选择用RFM来细分，接下来需要使用中位数对3个维度分别进行细分（每个维度4分，共12分）



也可以采用K-means对数据进行聚类分析，聚类分析之前需要进行数据形态转换、标准化、相关性探索等探索和预处理（每个步骤4分，共12分）

 (3) 根据上一步，对中位数处理后的数据进行客户细分，分成8类，通过数据透视表，每一类的客户特征和客户数如下表所示，也可以选择K-means聚类分析，选择聚成几类均可（K设置为几只要言之有理即可）（4分）



结果分析合理即得分（8类用户的结果分析共8分）

(4)对8类客户进行特征描述并给出营销建议

**特征描述：**（言之有理即得分，共5分）

**营销建议：**（言之有理即得分，共5分）