

数据挖掘实用案例分析

香水销售分析

一、前言介绍

- 本案的分析思路，从某电商网站上抓取到的香水产品销量数据，分析香水销售的影响因素，为香水销售商判定采购计划以及用户选择香水提供依据。

二、内容结果

- 香水销售数据预处理
- 香水销售数据统计分析
- 影响香水销量的因素分析
- 香水适用场所关联分析
- 香水聚类分析
- 香水营销建议

三、数据预处理

- 从某电商网站抓取1009条香水产品销售数据，包含了香水产品的商品名称、产品毛重、商品产地、包装、香调、净含量、分类、适用性别、适用场所、价格和评价数。
- Python编程处理 “评价” 和 “适用场所” 字段：
 - “评价” 字段的数据包含混合的中文和数字，末尾有一个 “+” 号，将其转为数值形式。即将类似 “1.9万+” 格式的 “评价” 字段的值转换为 “19000” ；
 - “适用场所” 分解为 “旅行” 、 “其他” 、 “约会” 、 “情趣” 等8个字段，其类型是0和1；
 - 将 “商品产地” 统一为 “中国” 。

三、数据预处理

- Python预处理完成示意图

1	商品名称	商品产地	包装	香调	净含量	分类	性别	适用场所	价格	评价	旅行	其它	约会	情趣	商务	日常	party聚会	运动
2	冰希黎695600860	中国	Q版香水	花果香调	1ml-15ml	浓香水EDP	女	日常, 约会, 情趣, 商务, party聚会, 旅行	10	19000	1	0	1	1	1	1	1	0
3	冰希黎695600860	中国	Q版香水	混合香调	1ml-15ml	浓香水EDP	女	日常, 约会, party聚会, 运动, 旅行	10	19000	1	0	1	0	0	1	1	1
4	(免邮)上海老臣		独立装	花果香调	31ml-100ml	浓香水EDP	女	日常, 约会, 商务, party聚会, 旅行	18	90	1	0	1	0	1	1	1	0
5	法颂浪漫梦境女士		Q版香水	花果香调	1ml-15ml	固体香水/香膏	女	日常, 约会, 商务, party聚会, 运动, 旅行	22	30	1	0	1	0	1	1	1	1
6	雅芳香水	中国	独立装	东方香调	1ml-15ml	淡香水EDT	女	日常, 旅行	24	100	1	0	1	0	1	1	1	1
7	雅芳香水	中国	独立装	东方香调	1ml-15ml	淡香水EDT	女	日常, 约会, 商务, 旅行	24	100	1	0	0	0	0	0	1	0
8	雅芳香水	中国	独立装	东方香调	1ml-15ml	淡香水EDT	女	日常, 约会, 商务, party聚会, 运动, 旅行	24	39000	1	0	0	0	0	1	0	0
9	雅芳香水	中国	独立装	东方香调	1ml-15ml	淡香水EDT	女	party聚会, 旅行	24	39000	1	0	1	0	1	1	0	0
10	美顿香水女士桂花		独立装	花果香调	31ml-100ml	淡香水EDT	女	日常, 约会, 旅行	25	50	1	0	1	0	0	1	0	0
11	冰希黎695600860	中国	独立装	混合香调	其它	固体香水/香膏	女	日常, 约会, 商务, party聚会, 运动, 旅行	25	19000	1	0	1	0	1	1	1	1
12	冰希黎695600860	中国	独立装	混合香调	其它	固体香水/香膏	女	日常, 约会, 情趣, 商务, party聚会, 旅行	25	19000	1	0	1	1	1	1	1	0
13	艾诗止汗香体走珠		独立装	花果香调	31ml-100ml	香体走珠	女	日常, 约会, 运动, 旅行	26	300	1	0	1	0	0	1	0	1
14	艾诗止汗香体走珠		独立装	花果香调	31ml-100ml	香体走珠	女	日常, 约会, 运动, 旅行	26	300	1	0	1	0	0	1	0	1
15	阿迪达斯女士净肤	中国		海洋香调	31ml-100ml	香体走珠	女		27	2700	0	1	0	0	0	1	0	1
16	阿迪达斯香体液	中国	独立装	混合香调	31ml-100ml	淡香水EDT	女	日常, 运动, 其它	27	2900	1	1	1	1	1	1	1	1
17	阿迪达斯香体液	中国		混合香调	31ml-100ml	香体走珠	女	日常, 约会, 情趣, 商务, party聚会, 运动, 旅行, 其它	27	5900	0	0	0	0	0	0	0	0
18	妮维雅走珠止汗露		独立装	花果香调	31ml-100ml	香体走珠	女	日常, 约会, party聚会, 运动	29	1300	0	0	1	0	0	1	1	1
19	Chanel香奈儿邂逅		独立装	花果香调	31ml-100ml	淡香水EDT	女	约会, 情趣, 商务, party聚会	29	40	1	0	1	0	0	1	0	0
20	美顿香水女士桂花	中国	独立装	花果香调	31ml-100ml	淡香水EDT	女	日常, 约会, 旅行	29	300	0	0	0	0	0	1	0	0

三、数据预处理

- 对香水产品的价格和评价数进行离散化处理。
 - 将价格等间距地分为6个等级，记为低、较低、中等、较高、高、非常高。
 - 将评价数等间距地分为7个等级，记为非常低、低、较低、中等、较高、高、非常高。
- 将价格和评价数离散化后的变量记为“价格等级”和“销量等级”，使用导出节点进行处理。

公式:

```
1 if (价格 <= 100) then '低'
2   else if (价格 <= 300) then '较低'
3     else if (价格 <= 500) then '中等'
4       else if (价格 <= 700) then '较高'
5         else if (价格 <= 1000) then '高'
6           else '非常高'
7         endif
8       endif
9     endif
10   endif
11 endif
```

公式:

```
1 if (评价 <= 100) then '非常低'
2   else if (评价 <= 500) then '低'
3     else if (评价 <= 1000) then '较低'
4       else if (评价 <= 2000) then '中等'
5         else if (评价 <= 5000) then '较高'
6           else if (评价 <= 10000) then '高'
7             else '非常高'
8           endif
9         endif
10       endif
11     endif
12   endif
13 endif
14
```

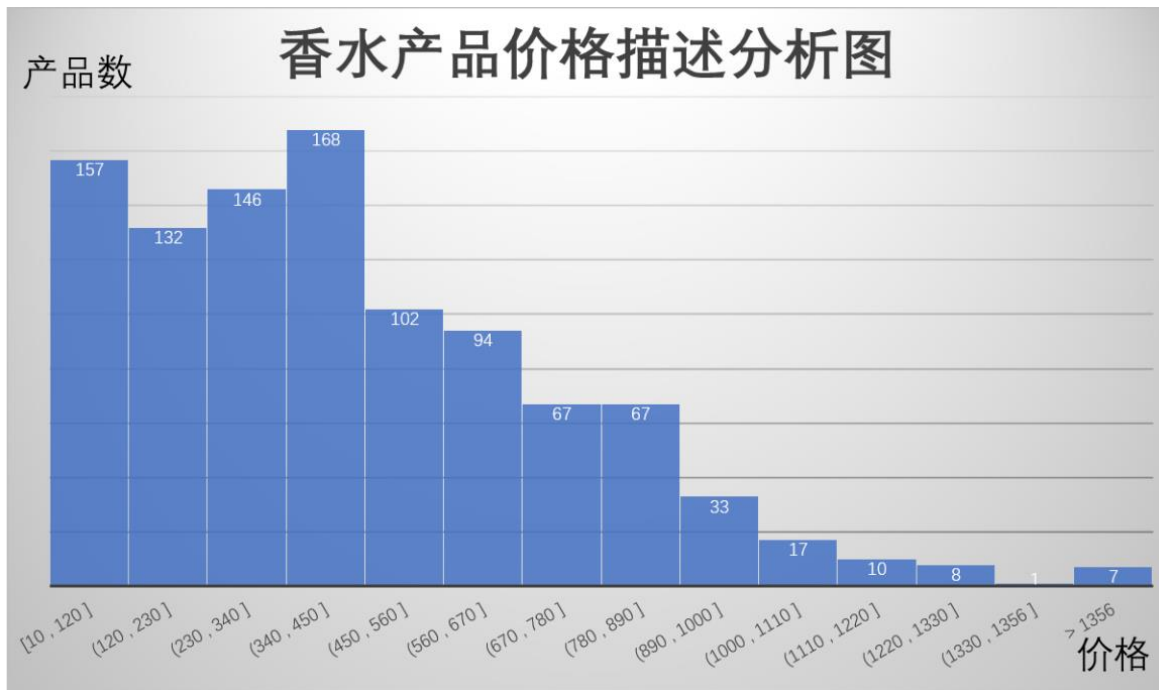
三、数据预处理

- 对香水产品的适用场合进行数量统计，得到新字段“适用场合数量”。
- 下图为最终处理得到的香水产品数据。

商品名称	商品产地	包装	香调	净含量	分类	性别	适用场所	价格	评价	旅行	其它	约会	情趣	商务	日常	party...	运动	适用场合数量	价格等级	销量等级
1 冰希黎6...	中国	Q版香水	花果香调	1ml-15ml	浓香水EDP	女	日常, 约会...	9.9...	190...	1.0...	0.0...	1.0...	1.0...	1.0...	1.0...	1.000	0.0...	6.000	低	非常高
2 冰希黎6...	中国	Q版香水	混合香调	1ml-15ml	浓香水EDP	女	日常, 约会...	9.9...	190...	1.0...	0.0...	1.0...	0.0...	0.0...	1.0...	1.000	1.0...	5.000	低	非常高
3 (免邮...		独立装	花果香调	31ml-100ml	浓香水EDP	女	日常, 约会...	18...	90.0...	1.0...	0.0...	1.0...	0.0...	1.0...	1.0...	1.000	0.0...	5.000	低	非常低
4 法颂良...		Q版香水	花果香调	1ml-15ml	固体香水/...	女	日常, 约会...	22...	30.0...	1.0...	0.0...	1.0...	0.0...	1.0...	1.0...	1.000	1.0...	6.000	低	非常低
5 雅芳香水	中国	独立装	东方香调	1ml-15ml	淡香水EDT	女	日常, 旅行	23...	100...	1.0...	0.0...	1.0...	0.0...	1.0...	1.0...	1.000	1.0...	6.000	低	非常低
6 雅芳香水	中国	独立装	东方香调	1ml-15ml	淡香水EDT	女	日常, 约会...	23...	100...	1.0...	0.0...	0.0...	0.0...	0.0...	0.0...	1.000	0.0...	2.000	低	非常低
7 雅芳香水	中国	独立装	东方香调	1ml-15ml	淡香水EDT	女	日常, 约会...	23...	390...	1.0...	0.0...	0.0...	0.0...	0.0...	0.0...	0.000	0.0...	2.000	低	非常高
8 雅芳香水	中国	独立装	东方香调	1ml-15ml	淡香水EDT	女	party聚会...	23...	390...	1.0...	0.0...	1.0...	0.0...	1.0...	1.0...	0.000	0.0...	4.000	低	非常高
9 美顿香...		独立装	花果香调	31ml-100ml	淡香水EDT	女	日常, 约会...	25...	50.0...	1.0...	0.0...	1.0...	0.0...	0.0...	1.0...	0.000	0.0...	3.000	低	非常低
10 冰希黎6...	中国	独立装	混合香调	其它	固体香水/...	女	日常, 约会...	25...	190...	1.0...	0.0...	1.0...	0.0...	1.0...	1.0...	1.000	1.0...	6.000	低	非常高
11 冰希黎6...	中国	独立装	混合香调	其它	固体香水/...	女	日常, 约会...	25...	190...	1.0...	0.0...	1.0...	1.0...	1.0...	1.0...	1.000	0.0...	6.000	低	非常高
12 艾诗止...		独立装	花果香调	31ml-100ml	香体走珠	女	日常, 约会...	25...	300...	1.0...	0.0...	1.0...	0.0...	0.0...	1.0...	0.000	1.0...	4.000	低	低
13 艾诗止...		独立装	花果香调	31ml-100ml	香体走珠	女	日常, 约会...	25...	300...	1.0...	0.0...	1.0...	0.0...	0.0...	1.0...	0.000	1.0...	4.000	低	低
14 阿迪达...	中国		海洋香调	31ml-100ml	香体走珠	女		26...	270...	0.0...	1.0...	0.0...	0.0...	0.0...	1.0...	0.000	1.0...	3.000	低	较高
15 阿迪达...	中国	独立装	混合香调	31ml-100ml	淡香水EDT	女	日常, 运动...	26...	290...	1.0...	1.0...	1.0...	1.0...	1.0...	1.0...	1.000	1.0...	8.000	低	较高
16 阿迪达...	中国		混合香调	31ml-100ml	香体走珠	女	日常, 约会...	26...	590...	0.0...	0.0...	0.0...	0.0...	0.0...	0.0...	0.000	0.0...	0.000	低	高
17 妮维雅...		独立装	花果香调	31ml-100ml	香体走珠	女	日常, 约会...	28...	130...	0.0...	0.0...	1.0...	0.0...	0.0...	1.0...	1.000	1.0...	4.000	低	中等
18 Chanel...		独立装	花果香调	31ml-100ml	淡香水EDT	女	约会, 情趣...	29...	40.0...	1.0...	0.0...	1.0...	0.0...	0.0...	1.0...	0.000	0.0...	3.000	低	非常低
19 美顿香...	中国	独立装	花果香调	31ml-100ml	淡香水EDT	女	日常, 约会...	29...	300...	0.0...	0.0...	0.0...	0.0...	0.0...	1.0...	0.000	0.0...	1.000	低	低
20 美顿香...	中国	独立装	花果香调	31ml-100ml	淡香水EDT	女	日常, 约会...	29...	300...	0.0...	0.0...	1.0...	1.0...	1.0...	0.0...	1.000	0.0...	4.000	低	低
21 美顿香...	中国	独立装	花果香调	31ml-100ml	淡香水EDT	女	日常, 约会...	29...	300...	1.0...	0.0...	1.0...	0.0...	0.0...	1.0...	0.000	0.0...	3.000	低	低

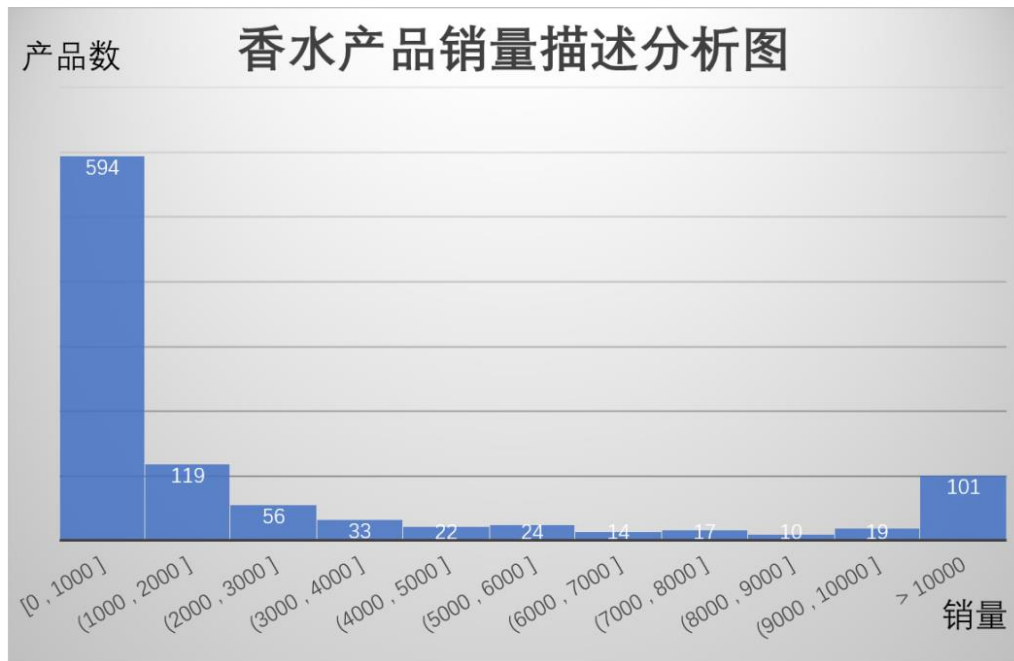
四、销售数据统计分析

- 香水产品价格描述分析图



四、销售数据统计分析

- 香水产品销量描述分析图



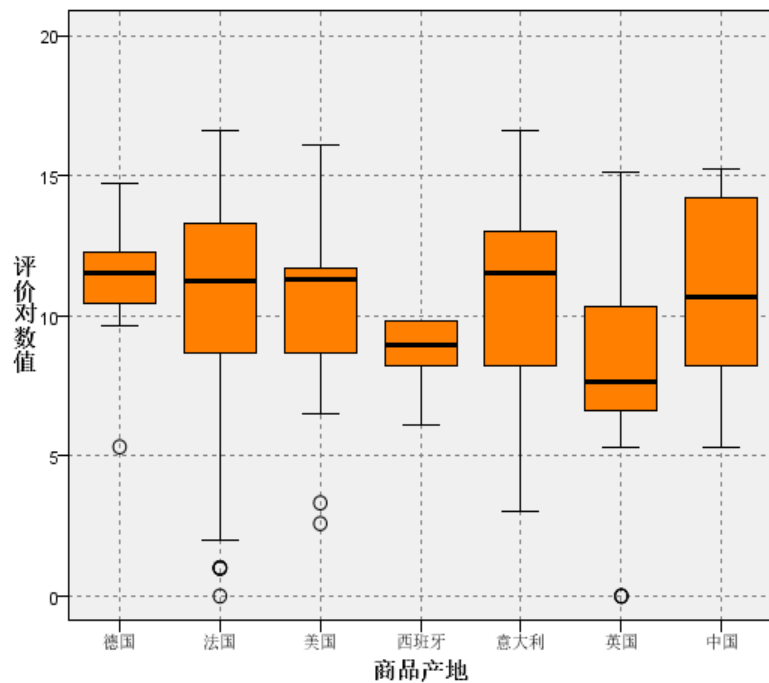
四、销售数据统计分析

- 香水产品产地分布图



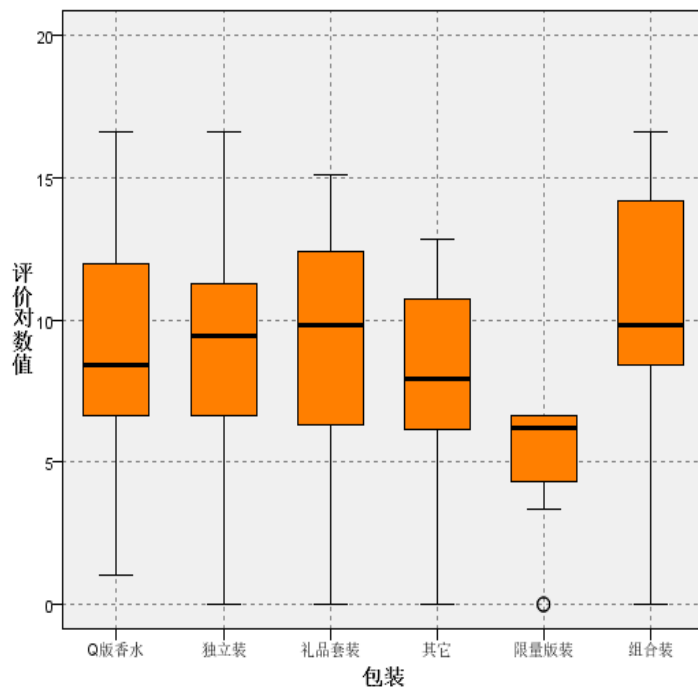
四、销售数据统计分析

- 各产地香水销量箱型图



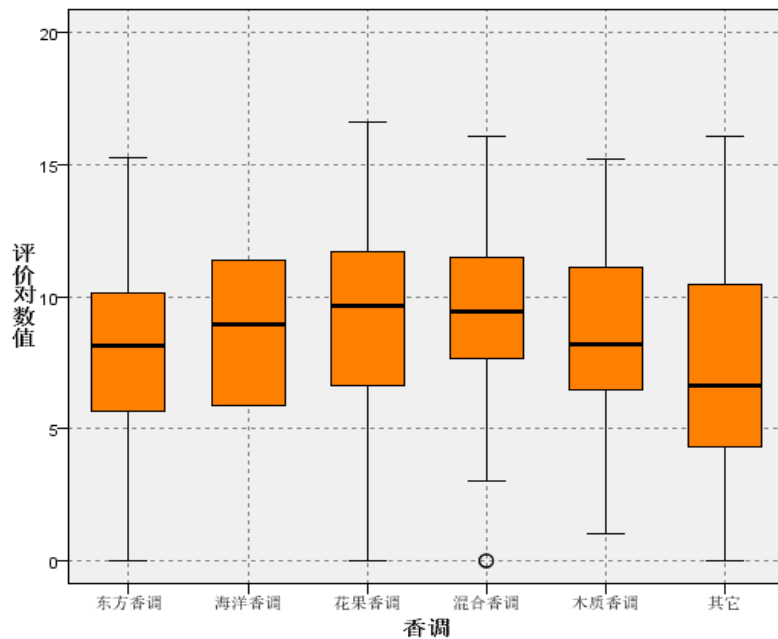
四、销售数据统计分析

- 各包装香水销量箱型图



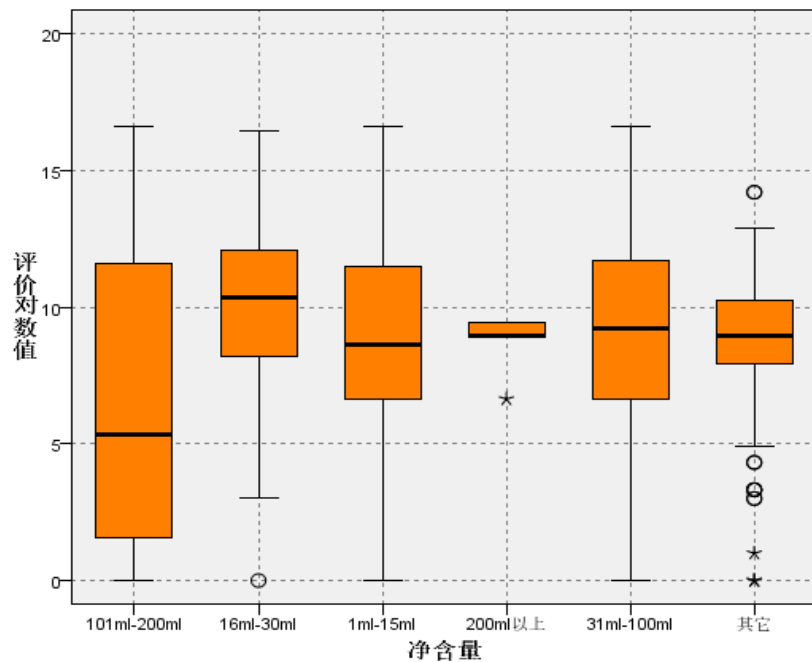
四、销售数据统计分析

- 不同香调香水销量箱型图



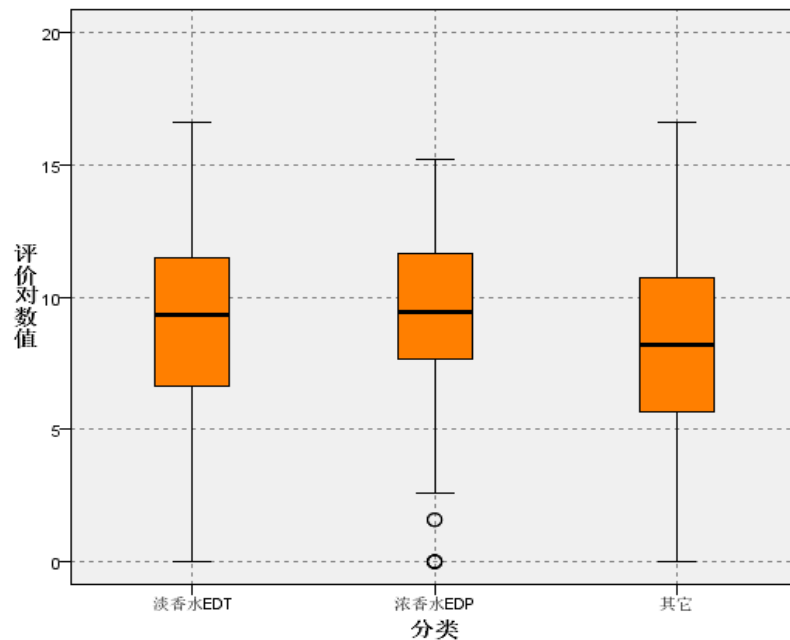
四、销售数据统计分析

- 不同净含量香水销量箱型图



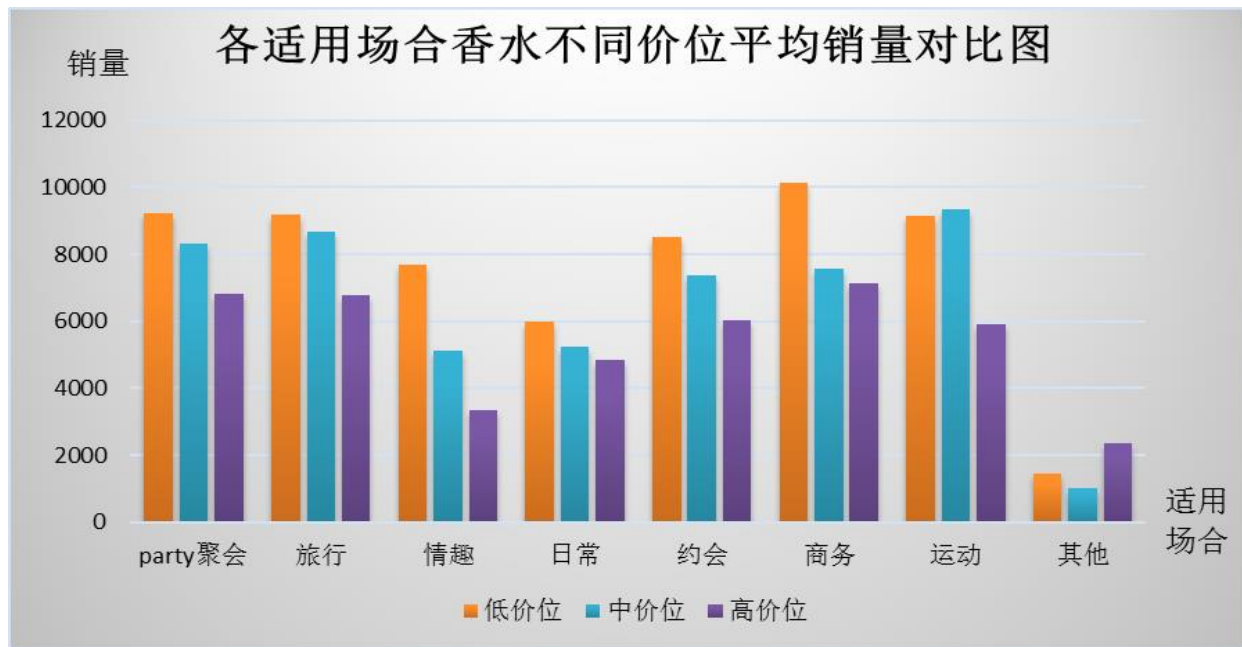
四、销售数据统计分析

- 各分类香水销量箱型图



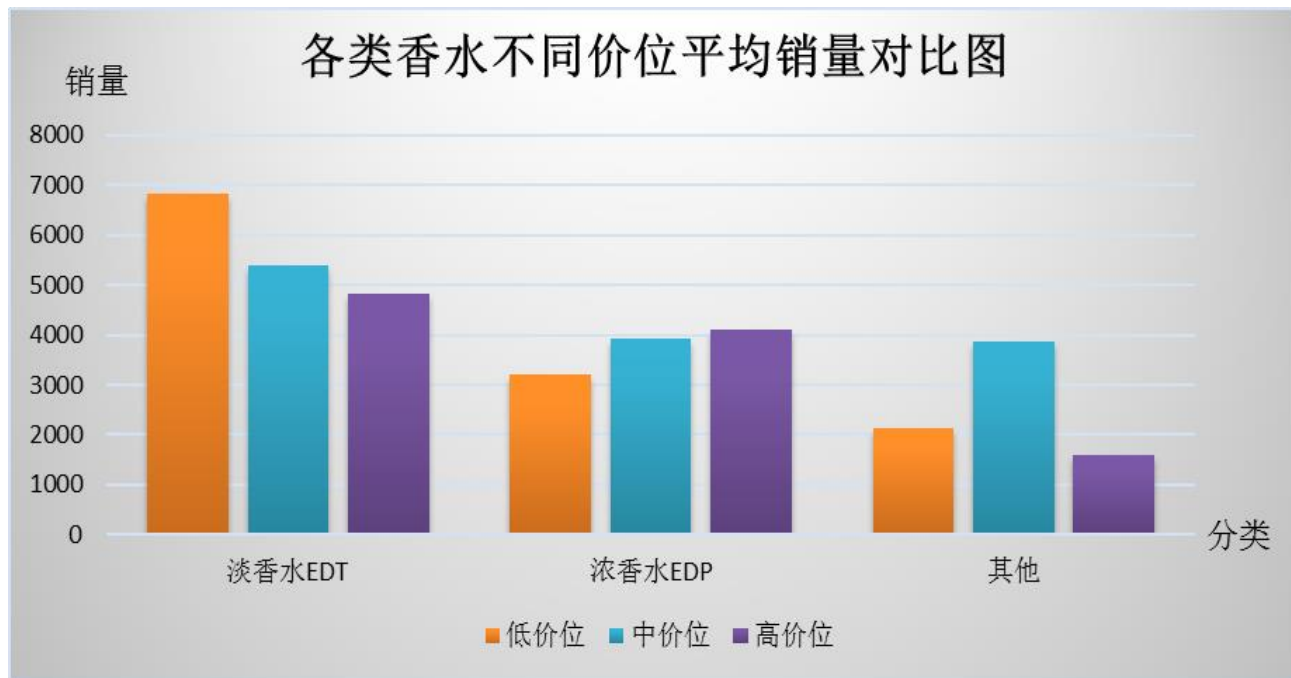
四、销售数据统计分析

- 各适用场合香水不同价位平均销量对比图



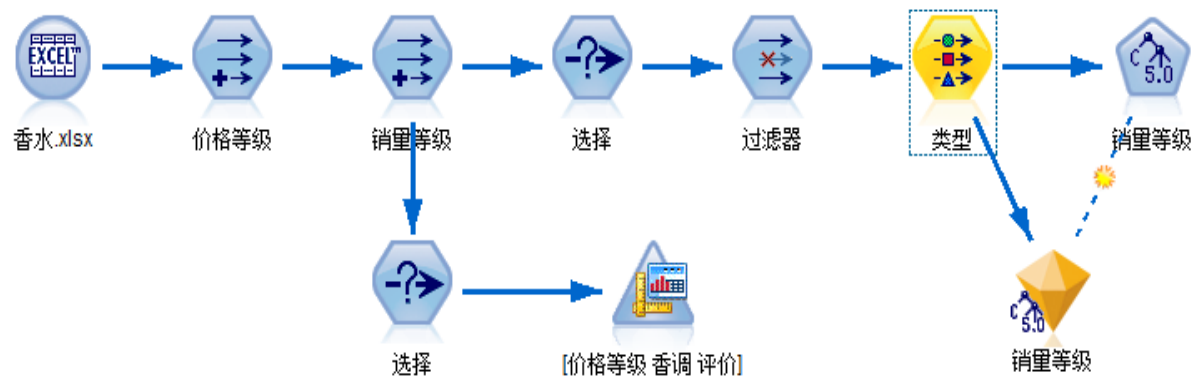
四、销售数据统计分析

- 各类香水不同价位平均销量对比图



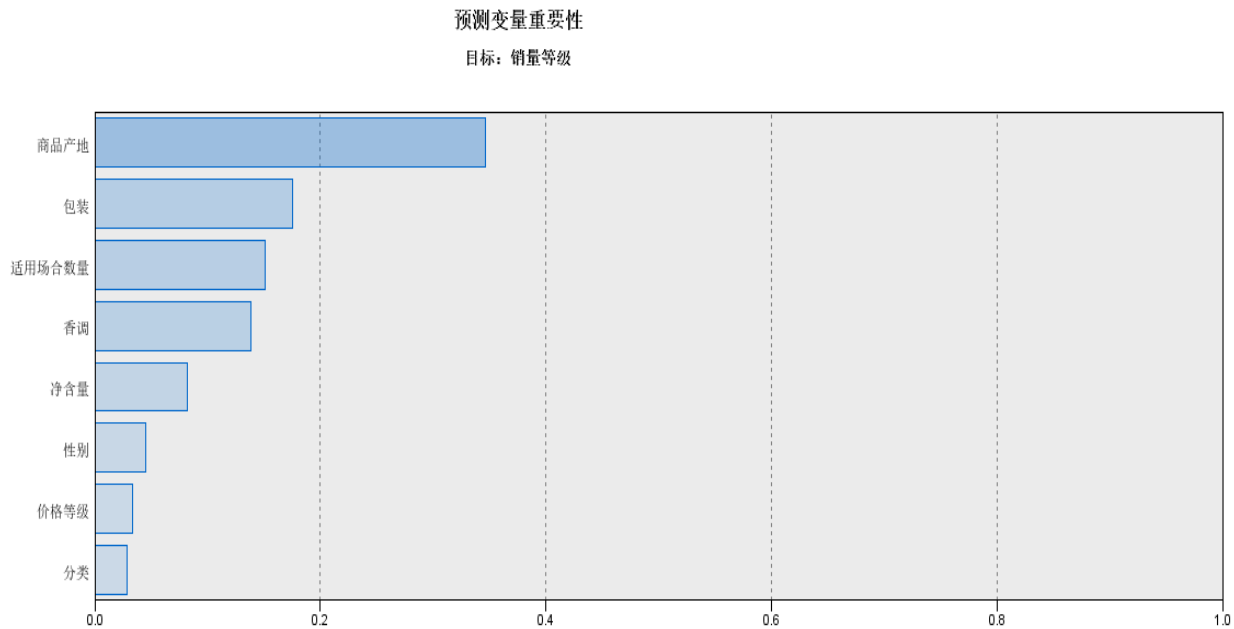
五、影响香水销量因素分析

- 使用C5.0决策树算法，挖掘影响香水产品销量等级的因素



五、影响香水销量因素分析

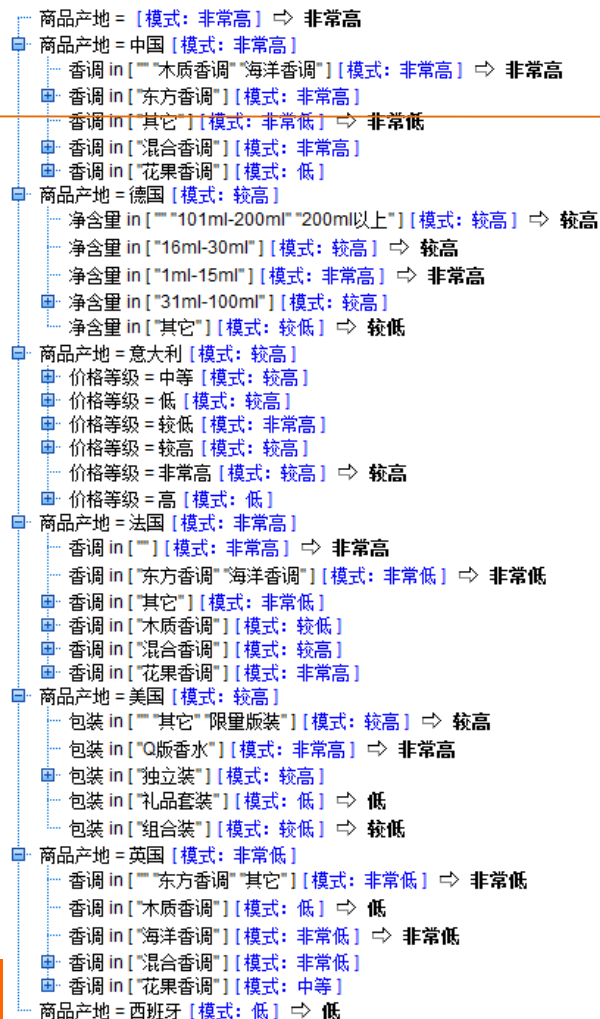
- 预测变量重要性



五、影响香水销量因素分析

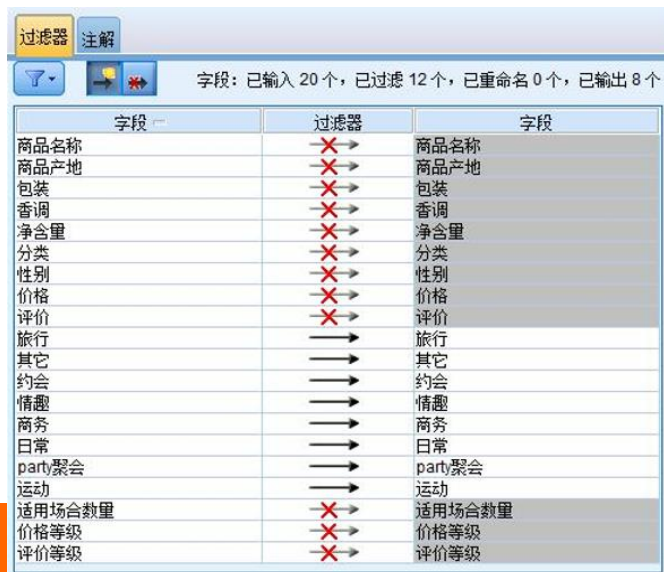
决策树

- 中国和法国生产的香水最受消费者欢迎
- 在中国和法国生产的香水中，消费者更加看重的是香水的香调。整体销量最高的“花果香调”在国产香水中销量反而较低，整体销量较低的“木质香料”销量却非常高
- 德国、意大利和美国的香水整体销量较高；对德国香水，消费者更注重香水的净含量；对意大利香水，消费者更看重价格；对美国香水，消费者更看重包装
- 英国和西班牙的产品销量较低。对英国香水，消费者更看重香调



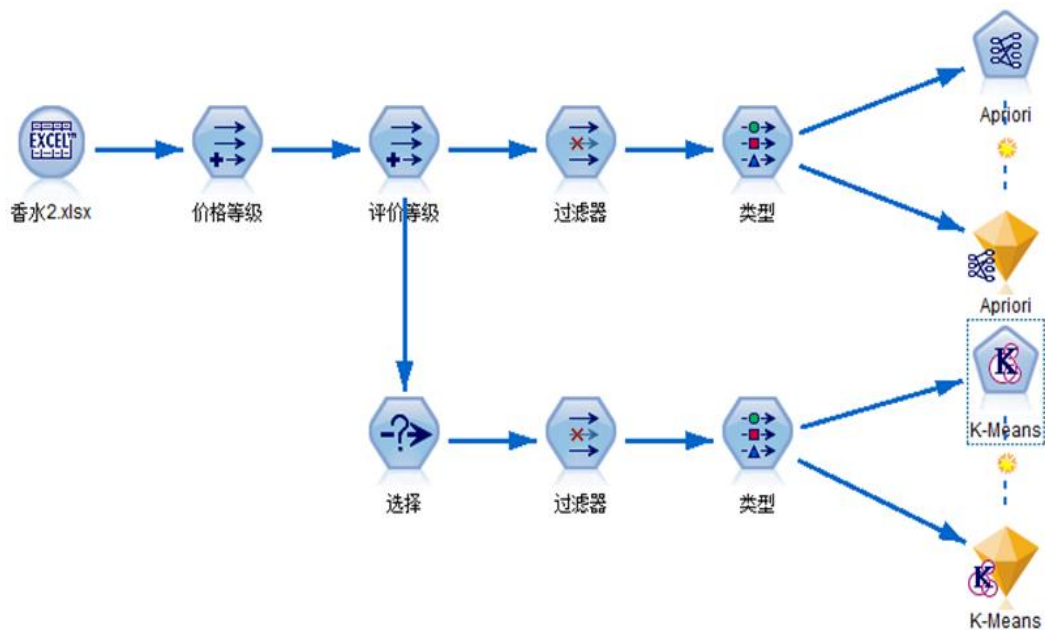
六、香水适用场景关联分析

- 对源数据进行预处理，将适用场所分隔开，生成不同的字段，总共为8类；
- 将含有该类适用场所的值设置为1.0，否则设置为0.0；
- 在关联分析前滤除除适用场合外的所有本次分析不需要的字段，将所有适用场合的类型设置为任意。



六、香水适用场景关联分析

- 采用Apriori算法，将最低条件支持度设为55%，最小规则置信度设置为90%



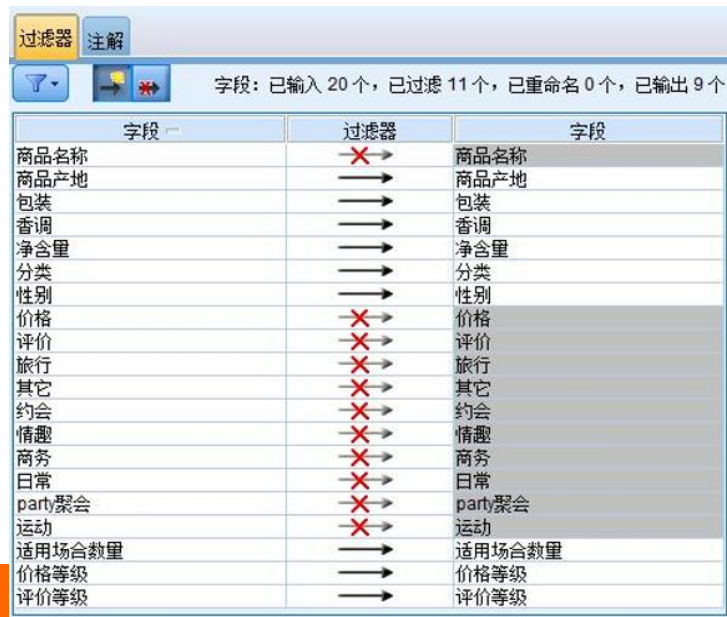
六、香水适用场景关联分析

- 大多数的适用场合之间关联性非常强;
- 日常、商务、party聚会、约会出现次数最多，也是相互关联性最强的场所;

排序依据: 置信度百分比				12	的	12
后项	前项	支持度百分比	置信度百分比			
日常	商务	57.676	99.64			
日常	party聚会	56.432	98.897			
	约会					
日常	party聚会	60.373	98.625			
约会	party聚会	59.544	93.728			
	日常					
party聚会	商务	57.676	93.525			
party聚会	商务	57.469	93.502			
	日常					
约会	party聚会	60.373	93.471			
日常	约会	66.805	93.168			
约会	商务	57.676	91.727			
约会	商务	57.469	91.697			
	日常					
商务	party聚会	55.809	90.335			
	约会					
	日常					
商务	party聚会	59.544	90.244			
	日常					

七、香水产品聚类分析

- 将数据中的商品产地、包装、香调、净含量、分类、性别、适用场合数量作为输入字段进行聚类分析
- 使用过滤器节点过滤不需要的字段



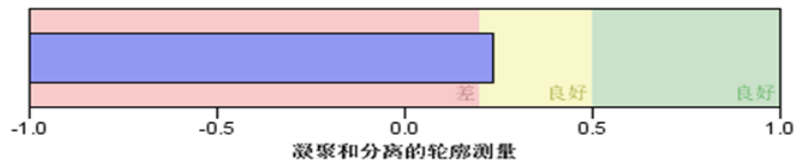
七、香水产品聚类分析

- 使用K-Means算法进行聚类，将聚类数设置为6
- 得到聚类模型概要和聚类质量

模型概要

算法	K-Means
输入	7
聚类	6

聚类质量



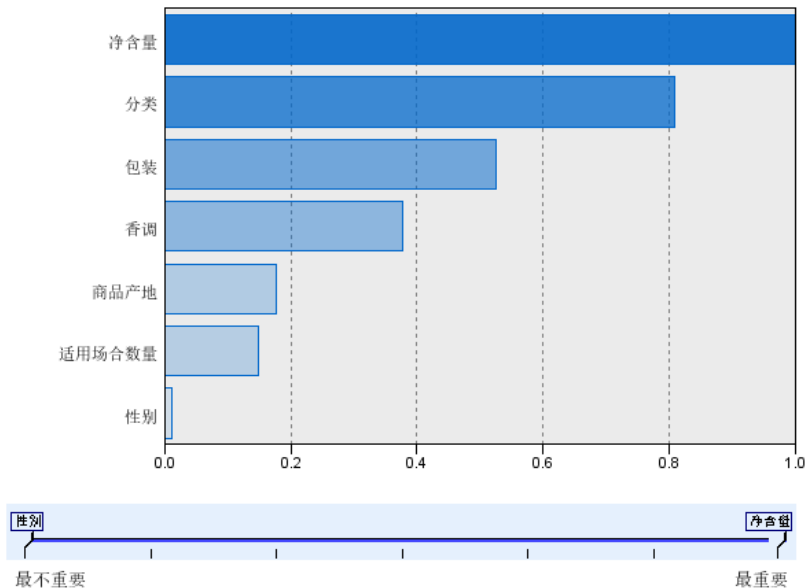
七、香水产品聚类分析

K-Means聚类结果

聚类

聚类 标签	输入（预测变量）重要性					
	1.0	0.8	0.6	0.4	0.2	0.0
描述	聚类-1	聚类-4	聚类-3	聚类-6	聚类-2	聚类-5
大小	24.5% (112)	23.0% (105)	22.1% (101)	11.6% (53)	9.6% (44)	9.2% (42)
输入	净含量 31ml-100ml (66.1%)	净含量 31ml-100ml (98.1%)	净含量 31ml-100ml (74.3%)	净含量 31ml-100ml (79.2%)	净含量 1ml-15ml (90.9%)	净含量 16ml-30ml (97.6%)
	分类 淡香水EDT (100.0%)	分类 浓香水EDP (100.0%)	分类 淡香水EDT (100.0%)	分类 淡香水EDT (98.1%)	分类 淡香水EDT (79.5%)	分类 浓香水EDP (95.2%)
	包装 独立装 (94.6%)	包装 独立装 (96.2%)	包装 独立装 (94.1%)	包装 独立装 (92.5%)	包装 Q版香水 (61.4%)	包装 独立装 (92.9%)
	香调 花果香调 (98.2%)	香调 花果香调 (62.9%)	香调 花果香调 (96.0%)	香调 混合香调 (73.6%)	香调 花果香调 (52.3%)	香调 花果香调 (73.8%)
	商品产地 意大利 (41.1%)	商品产地 法国 (61.0%)	商品产地 法国 (75.2%)	商品产地 法国 (50.9%)	商品产地 法国 (29.5%)	商品产地 法国 (57.1%)
	适用场合数量 1,000 (52.7%)	适用场合数量 1,000 (22.9%)	适用场合数量 6,000 (57.4%)	适用场合数量 1,000 (30.2%)	适用场合数量 1,000 (29.5%)	适用场合数量 1,000 (33.3%)
	性别 女 (96.4%)	性别 女 (99.0%)	性别 女 (94.1%)	性别 女 (92.5%)	性别 女 (90.9%)	性别 女 (100.0%)

预测变量重要性



八、香水产品营销建议

- 制定价格方面：将产品价格定位在大众消费品的水平，并保持正常利润空间。对淡香水EDT类产品，在不亏损的前提下适当降低价格；对浓香水EDT类产品，在调整空间内适当提高价格；
- 产品分类方面：香水产品的产地、香调、净含量等对销售有很大影响，需要综合考虑几种因素才能获得更多收入；
- 销售策略方面：消费者在购买香水产品时体现出了明显的价格敏感性，价格低的销量更好。组合装的香水销量好于其他包装。商家需要结合不同使用场合推出更多的香水组合和礼品装香水，以刺激消费。