# 实验9\_任务2\_关联分析

# 问题：

任务**2** 关联分析（对应⽂件 **car.data.csv**）

任务描述：给定⼆⼿评估情况的数据，第⼀⾏为各类属性名称，最后⼀列为⼆⼿的评估情况分类，以⽐例分为如下4类：

unacc 1210 70.023

acc 384 22.222

good 69 3.993

v-good 65 3.762

各属性的可能取值如下（并不全是数值）：

buying： v-high, high, med, low

maint： v-high, high, med, low

doors： 2, 3, 4, 5more

persons： 2, 4, more

lug\_boot： small, med, big

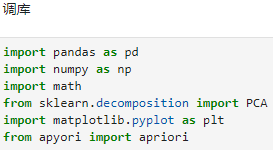
safety： low, med, high

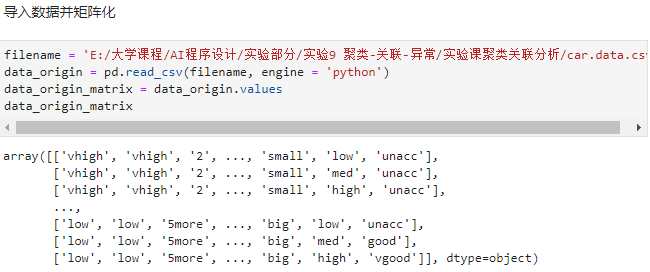
1. 请对数据进⾏关联分析，计算哪些属性的哪些值对于每个分类结果有较强的关联（请⾃定置信度和⽀持的阈值）。

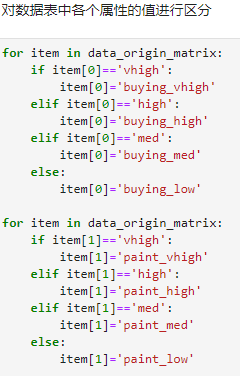
2. 尝试⽤前述聚类⽅法结果对该数据进⾏聚类，尝试分析聚类结果与关联分析的联系。

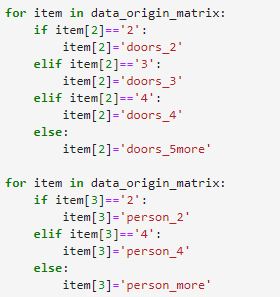
# 实验过程记录：

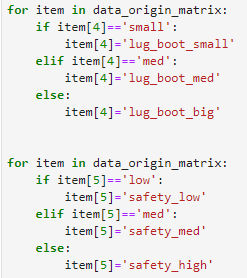


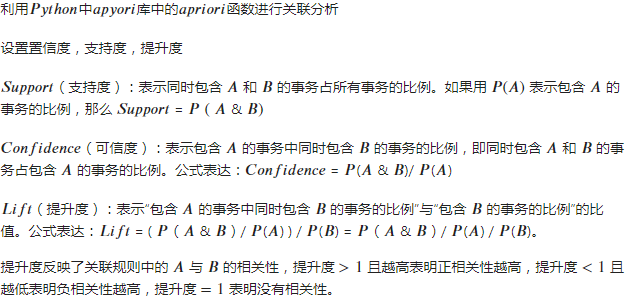




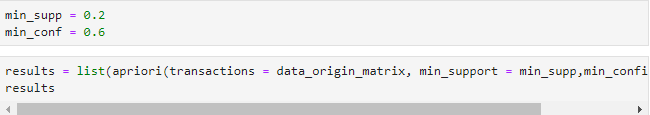




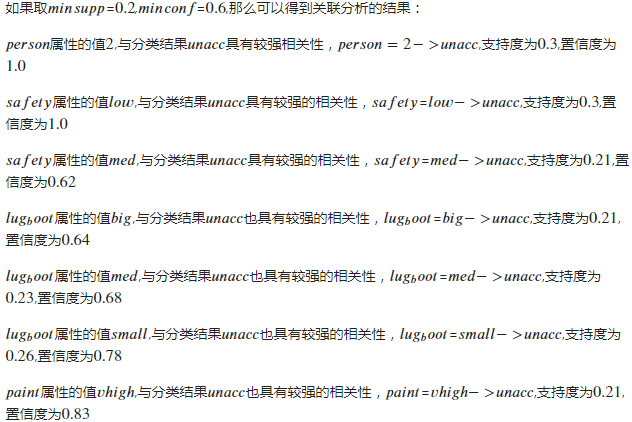




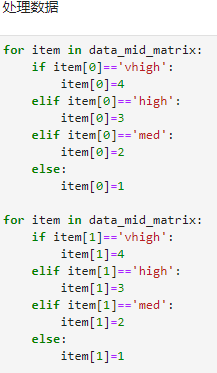


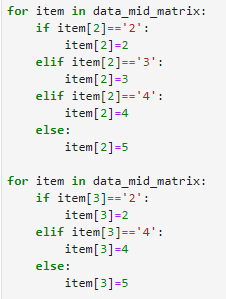


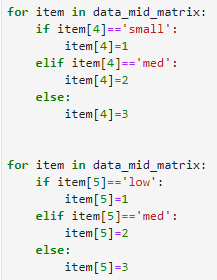


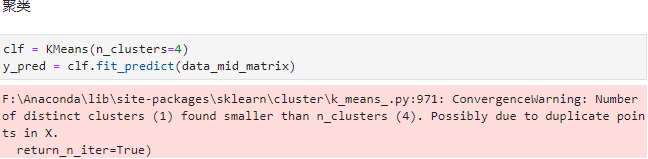


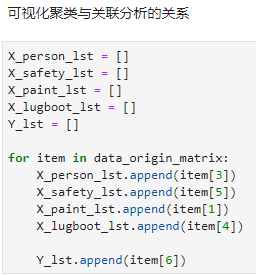


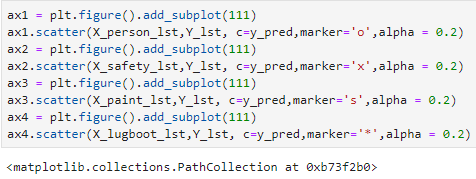


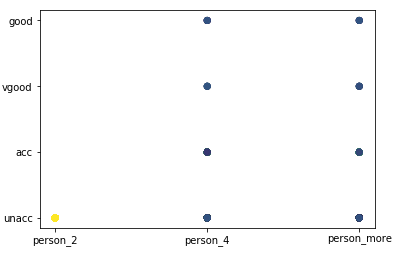


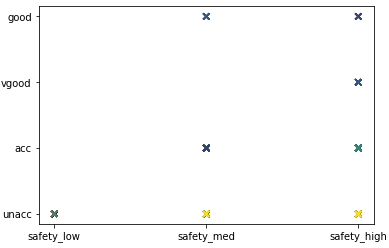


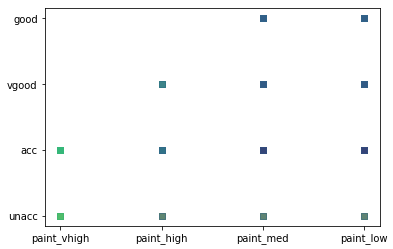


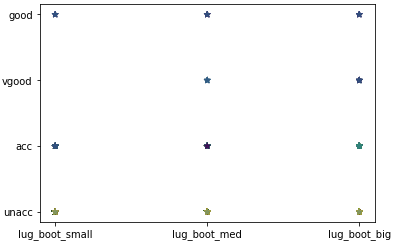


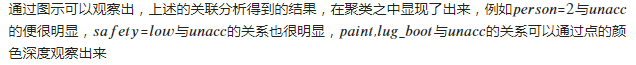












**问题1：**

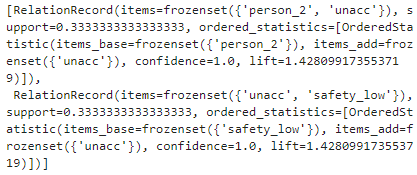
I．数据处理主要包括以下步骤：

1. 由于许多属性可以去到相同的值，为了避免相同的值在关联分析中互相影响，先把各个属性的值之前也附上属性的名称。

II.关联分析成果：

1）采用不同的支持度和置信度参数进行关联分析，观察结果。

a)取最小支持度为0.3，最小置信度为0.8，那么得到如下结果：



意味着：

person属性的值2,与分类结果unacc具有较强相关性，person=2->unacc,支持度为0.3,置信度为1.0.

safety属性的值low,与分类结果unacc也具有较强的相关性，safety=low->unac,支持度为0.3,置信度为1.0.

b)取最小支持度为0.2，最小置信度为0.6，那么得到如下结果：



意味着：

person属性的值2,与分类结果unacc具有较强相关性，person=2->unacc,支持度为0.3,置信度为1.0

safety属性的值low,与分类结果unacc具有较强的相关性，safety=low->unacc,支持度为0.3,置信度为1.0

safety属性的值med,与分类结果unacc具有较强的相关性，safety=med->unacc,支持度为0.21,置信度为0.62

lug\_boot属性的值big,与分类结果unacc也具有较强的相关性，lug\_boot=big->unacc,支持度为0.21,置信度为0.64

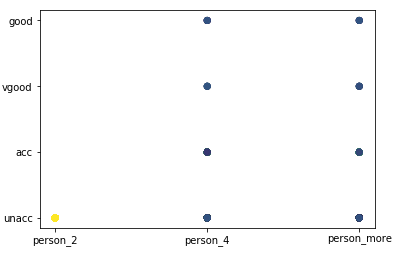
lug\_boot属性的值med,与分类结果unacc也具有较强的相关性，lug\_boot=med->unacc,支持度为0.23,置信度为0.68

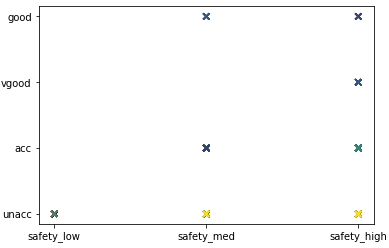
lug\_boot属性的值small,与分类结果unacc也具有较强的相关性，lug\_boot=small->unacc,支持度为0.26,置信度为0.78

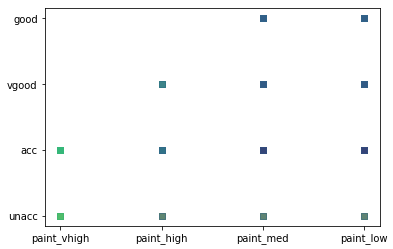
paint属性的值vhigh,与分类结果unacc也具有较强的相关性，paint=vhigh->unacc,支持度为0.21,置信度为0.83

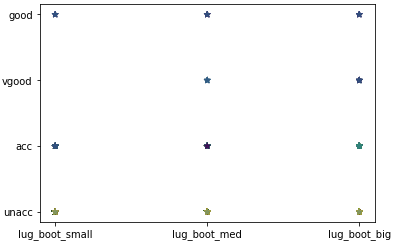
**问题2：**

I聚类成果

1. 可视化：



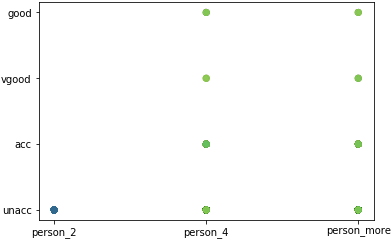


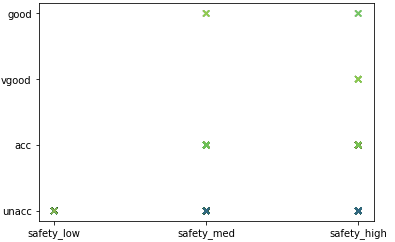


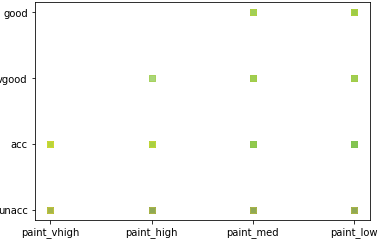
2）分析：

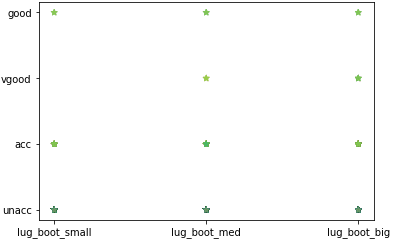
其中，通过不同的颜色并设置透明度，可以观察到数据点的聚集情况：

1. 在第一、二张图中，明显可以看到unacc这类属性与person=2和safety=low的关联明显，符合关联分析结果
2. 在第三、四张图中，通过把透明度改为0.1,在比较点的颜色深度，就可以比较该点处点的多少，可以发现：









lug\_boot对应的small，med，big在unacc这一栏的标记颜色均很深，符合关联分析结果

但是，paint的vhigh属性在unacc这一栏上体现的并不明显（颜色并不深），并没有体现出强关联

同时，safety的high属性在unacc上这一栏的点颜色也比较深，但在关联分析中，并没有反映出这一点。