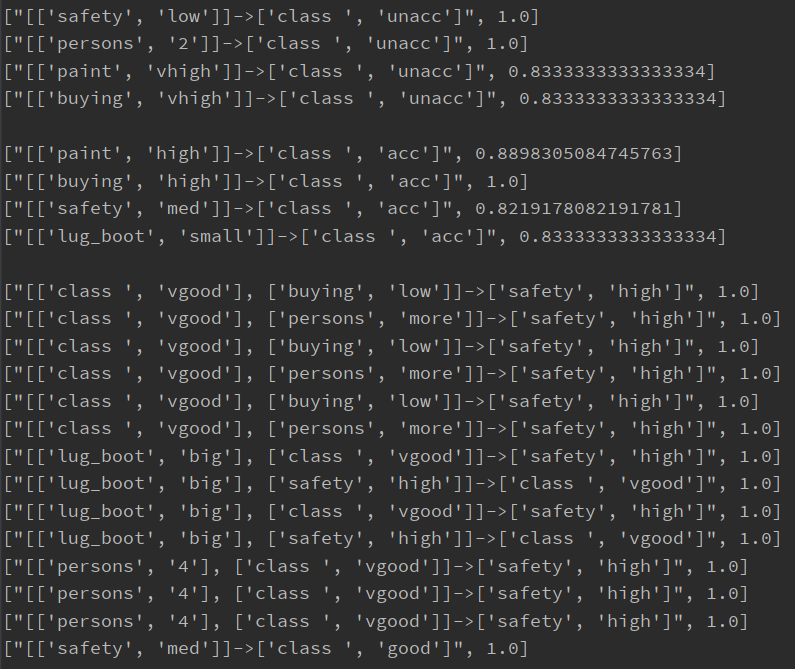
1. 首先按照定义写一个apriori算法，这里的情况与一般有些不同，项不是单独构成的，而是一个“列名”与一个“属性”组合在一起，所以需要略微调整。
2. 用apriori算法对数据集进行关联分析，发现由于unacc与acc占比较多，因此设置支持度时弱太小则运行时间过长，弱太大则不会有较少两类车出现
3. 暂时没有想到太好的解决方法，于是分别对最多、次多、最少两类 进行关联分析，得到一些关联结论如下，前面是关联规则，后面是置信度。



1. 随后进行聚类分析。我的选择是将除最后一列外的信息聚类，最后一列用来检验。