

介绍 1-2 同级红酒群体发现的背景及解决方案

建议课时：10 分钟

一、 教学目标

1. 初步了解红酒关注群体透视表的维度
2. 了解该实战的应用方向

二、 背景介绍

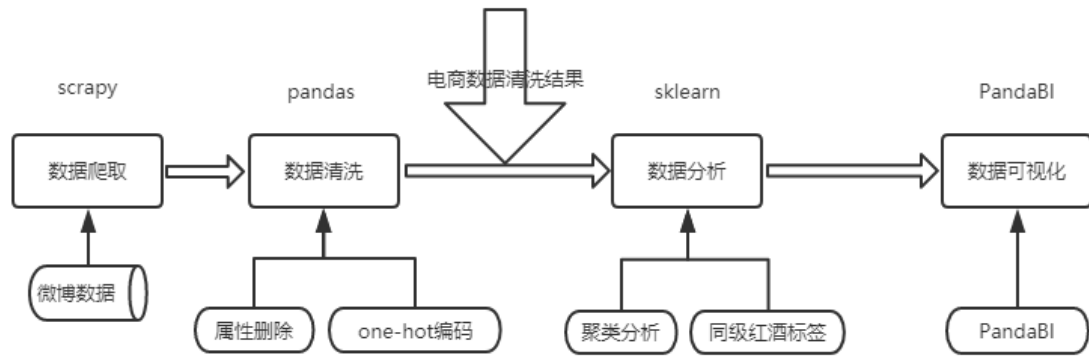
在辅助海关红酒定价决策中，我们可以得到电商数据清洗之后的红酒数据，利用聚类可以得到对红酒进行一个级别的划分，得到同类红酒。在此基础上，可应用的方向有：

- 对同级红酒的关注群体所在的城市、区域、年龄段等信息做透视表，用以辅助运营相关的决策。
- 开发新的红酒品牌，得到各级红酒的特性，辅助定价决策。
- 若有时间线的持续收集数据，可以观测某个话题或者促销之后的关注群体的变化，用以辅助运营品牌决策。
- 等等

本案例中涉及的应用方向是对同级红酒的关注群体的各个属性做透视表，用以辅助运营相关的决策。

三、 架构分析

主要解决的问题是同级红酒的定义，在此基础上做同级红酒群体的透视表。该问题可用聚类或者分类的方法解决，由于海关决策已讲解了分类的处理流程，此次采用聚类方法解决。完整解决方案如下：



具体解释如下：

步骤	处理内容	工具/第三方包
数据爬取	根据预先整理的红酒品牌列表，检索微博的实时话题，爬取话题人的相关信息	Scrapy
数据清洗	<ul style="list-style-type: none">● 数据清洗：清洗微博的数据，涉及的操作有缺失值处理，列计算，列删除等● 数据整合：把微博数据和电商数据整合	Pandas
数据分析	采用 kmeans 方法聚类	sklearn
数据可视化	展示各个维度的透视表	PandaBI