

## E 题 基于无监督学习的客户细分研究

### 注意事项

1. 限制拿到数据集后 168 小时内完成。
2. 禁止将数据集分享给他人，造成的法律纠纷由分享者承担。
3. 限制使用 Python 语言编写分析算法。

无监督学习的主要应用场景之一就是市场细分。虽然在我们开发市场时获取的数据都没有标记，但是将市场细分成不同类型至关重要，这样人们就可以关注各自的市场类型了。市场细分对广告投放、库存管理、配送策略的实施、大众传媒等市场行为都非常有用。

下面我们将与一个零售商和他的客户打交道，采用 <https://archive.ics.uci.edu/ml/datasets/Wholesale+customers> 的数据进行分析。数据表里包含了不同类型商品的销售数据，目标是找到数据集群，从而为客户提供最优的销售和分销策略。

请根据所采集的零售商和客户数据，运用数据挖掘的方法（如：无监督聚类的方法），完成以下任务：

- （1）建立无监督聚类模型，如均值漂移聚类模型。
- （2）将 Milk 和 Groceries 两个特征进行聚类，并将聚类结果可视化，以获取直观的输出。
- （3）评价聚类算法的聚类效果。（提示：可以采用轮廓系数）
- （4）对客户类别进行细分，并为零售商提供建议，为不同的客户提供销售和分销策略。

### Attribute Information:

1) FRESH: annual spending (m.u.) on fresh products (Continuous);  
2) MILK: annual spending (m.u.) on milk products (Continuous);  
3) GROCERY: annual spending (m.u.) on grocery products (Continuous);  
4) FROZEN: annual spending (m.u.) on frozen products (Continuous)  
5) DETERGENTS\_PAPER: annual spending (m.u.) on detergents and paper products (Continuous)  
6) DELICATESSEN: annual spending (m.u.) on delicatessen products (Continuous);  
7) CHANNEL: customers' Channel - Horeca (Hotel/Restaurant/Café) or Retail channel (Nominal)  
8) REGION: customers' Region - Lisbon, Oporto or Other (Nominal)

### Descriptive Statistics:

(Minimum, Maximum, Mean, Std. Deviation)

FRESH ( 3, 112151, 12000.30, 12647.329)  
MILK (55, 73498, 5796.27, 7380.377)  
GROCERY (3, 92780, 7951.28, 9503.163)  
FROZEN (25, 60869, 3071.93, 4854.673)  
DETERGENTS\_PAPER (3, 40827, 2881.49, 4767.854)  
DELICATESSEN (3, 47943, 1524.87, 2820.106)

### REGION Frequency

Lisbon 77  
Oporto 47  
Other Region 316  
Total 440

### CHANNEL Frequency

Horeca 298  
Retail 142  
Total 440