批发商的客户类型分析（客户画像）（聚类）：

一、目标：

对批发商的客户类型进行聚类，以便批发商更好的组织物流和服务。

二、数据集: Wholesale customers data.csv

【属性】

8个变量，前6个代表客户每年在某一类商品上的消费金额，CHANNEL代表客户的销售渠道（0-餐饮， 1零售）， REGION代表客户的区域位置。

1) FRESH: annual spending on fresh products (Continuous);

2) MILK: annual spending on milk products (Continuous);

3) GROCERY: annual spending on grocery products (Continuous);

4) FROZEN: annual spending on frozen products (Continuous)  
5) DETERGENTS\_PAPER: annual spending on detergents and paper products (Continuous)

6) DELICATESSEN: annual spending on and delicatessen products (Continuous);

7) CHANNEL: customers Channel(包括：Horeca (Hotel/Restaurant/Cafe) or Retail channel (Nominal)

8) REGION: customers Region (Lisnon, Oporto or Other) (Nominal)

三、要求（部分要求如果难以实现，允许不完成，只要把数据挖掘过程描述清楚即可）：

1. 应用合适的数据探索方法，更好地了解数据。
2. 数据预处理：
   1. 检查数据集是否存在缺失值和异常值，如果存在，则消除它们（任意方法均可）。
   2. 根据选用的模型，自行决定是否需要归一化数据
   3. 是否存在标称属性或者序数属性需要给予处理
   4. 少量属性与聚类结果关联不大，删除该属性，如REGION
   5. 其他预处理步骤。
3. 选用合适的聚类模型，划分训练数据和测试数据，构建聚类模型。
   1. 解释所选模型的依据
   2. 解释模型的参数选择方法和最优参数结果。
   3. 选择合适的性能评价指标评价模型的性能。
   4. 绘制有关训练结果或训练过程的图。
4. 对聚类结果中的每一簇进行解释，合理推测他们属于哪种类型的批发客户（例如，大型超市？小杂货店，咖啡馆，餐饮店….）
5. 完成“xxx数据挖掘分析报告”。
6. 准备5分钟左右的ppt报告，在第18周随堂讲述本设计的主要工作。