

## 复旦大学本科生毕业论文（设计）检查情况记录

指导教师对学生论文指导过程记录		
时间	讨论内容(包括存在的问题和改进措施)	指导教师签名
2017.3.30	<ol style="list-style-type: none"> <li>1. 问题：发现数据清洗过程中，关系表主码缺失。改进措施：通过熟悉 SQL 语句逐个排查，找到主键。</li> <li>2. 问题：数据原子结构不支持布尔型关联规则。改进措施：学习数值型关联规则，进行后续操作。</li> <li>3. 问题：销量的数据日数据过于庞大，无法作为聚类的输入。改进措施：先算出 180 个 item 的日销量相关关系矩阵，以矩阵系数作为聚类模型输入。</li> <li>4. 问题：观测期间的价格变动无法察觉，同一种产品的价格方差较大，并有悖常理。改进措施：将观测期的每个 item 数据进行逐一排查。剔除异常值，对价格进行按照常理推测。或从数据外部获得商家调价信息，或以销量最大的价格作为商品的价格基准。</li> <li>5 问题：数据清理耗时严重，需要加快进度。</li> </ol>	