

# 数据分析及实践 \_\_Assignment4

Xiaoma

2023 年 5 月 15 日

## 1 实验要求

利用 **Assignment3** 得到的与 **REPEAT** 列最相关的 5 个特征作为特征集，根据关联规则算法，挖掘这 5 个特征和 **REPEAT** 构成的频繁项集和关联规则。

注意事项：

- 数据集中与 **REPEAT** 列直接相关的三个特征在本次实验中仅允许合为一个特征使用 '**ST127Q01TA,ST127Q02TA,ST127Q03TA**'
- 需要挖掘的频繁项集的最小支持度阈值为 0.6，最小置信度阈值为 0.15
- 建议使用的关联规则算法：**Apriori** 或相关改进算法（自行实现）
- 选择的特征可以是数据集里有的，也可以是自己在实验三中构建的特征

## 2 实验原理

### 2.1 $3\sigma$ 准则

$3\sigma$  准则是一个简单的推论，几乎所有的值都在平均值正负三个标准差的范围内，也就是在实验上可以将 99.7% 的机率视为“几乎一定”。如果数据不服从正态分布，也可以用远离平均值的多少倍标准差来描述（这就使该原理可以适用于不同的业务场景，只是需要根据经验来确定  $k, \sigma$  中的  $k$  值，这个  $k$  值就可以认为是阈值）。

$$P(|x - \mu| > 3\sigma) \leq 0.003$$

其中， $\mu$  是样本的平均值， $\sigma$  是样本的标准差。这个公式表示，对于一个正态分布的样本，它的绝对偏差超过 3 倍标准差的概率小于等于 0.003，也就是说，99.7

### 2.2 提升度

提升度是一种关联规则中的度量，用于衡量两个事件之间的相关性。提升度越大，表示两个事件之间的相关性越强。提升度的数学原理可以用下面这个公式表示：

$$lift(A, B) = \frac{P(A, B)}{P(A)P(B)}$$

其中，A 和 B 是两个事件， $P(A)$  和  $P(B)$  分别是 A 和 B 的支持度， $P(A, B)$  是 A 和 B 同时出现的概率。<sup>3</sup>

### 2.3 置信度

置信度是指在统计学中描述概率的另一种方式。例如，使用 0.95 的置信水平构建置信区间，则有信心估计值在置信区间指定的上限和下限值之间的次数为 100 次中的 95 次。

## 2.4 Apriori 算法

Apriori 算法是经典的挖掘频繁项集和关联规则的数据挖掘算法。A priori 在拉丁语中指“来自以前”。当定义问题时，通常会使用先验知识或者假设，这被称作“一个先验”（a priori）。

Apriori 算法的数学原理如下：首先，找出所有频繁项集，然后由频繁项集产生强关联规则。Apriori 算法的核心是利用了频繁项集的性质，即频繁项集的所有非空子集也一定是频繁的。

以下是 Apriori 算法的伪代码：

---

**Algorithm 1:** Apriori 算法

---

**Data:** 数据集

**Result:** 频繁项集

initialization;

$k = 1$ ;

$F_k = \{i | i \in I, \sigma(\{i\}) \geq N \times \text{minsup}\}$ ;

**repeat**

$k = k + 1$ ;

$C_k = \text{apriori-gen}(F_{k-1})$  **for**  $\text{transaction} \in T$  **do**

$C_t = \text{subset}(C_k, t)$ ;

**for**  $c \in C_t$  **do**

$\sigma(c) = \sigma(c) + 1$

**end**

**end**

$F_k = \{c | c \in C_k, \sigma(c) \geq N \times \text{minsup}\}$ ;

**until**  $F_k = \emptyset$ ;

---

## 3 实验环境

VSCode + Python3.9.13

## 4 实验步骤

### 4.1 数据转换

参考 Assignment3 的特征相关性，首先抽取实验所需要的 5 个特征以及 REPEAT 列，'ST127Q02TA,ST001D01T,GRADE,ISCEDL,OCOD2,REPEAT'。根据实验要求，只抽取 REPEAT 列值为 1 的数据，丢弃所有缺失数据。

#### 4.1.1 数据去噪

使用  $3\sigma$  准则进行去噪。

#### 4.1.2 数据格式转换

为了将原格式数据转换为 Apriori 算法的输入数据，将挑选的 5 个特征的所有不同的取值均作为不同类别的事件，构建一个列为所有不同事件的矩阵  $\mathbf{R}$ ，若  $\mathbf{R}[i][j]$  非空，则说明在第  $i$  次记录中，事件  $j$  发生了。

### 4.2 寻找频繁项集和关联规则

使用 Apriori 算法寻找频繁项集和关联规则，由于 OCOD2 中有许多只出现过一次的值，考虑到计算时间复杂度，分别删去出现频率小于 10，小于 50 的值然后计算频繁项集。

#### 4.2.1 Fre > 10

---

k=1 的频繁项集

---

ISCEDL\_2.0

REPEAT

ST127Q02TA\_2.0

ST001D01T\_9.0

GRADE\_-1.0

---

---

k=2 的频繁项集

---

ISCEDL_2.0	REPEAT
ISCEDL_2.0	ST127Q02TA_2.0
REPEAT	ST127Q02TA_2.0
ISCEDL_2.0	ST001D01T_9.0
GRADE_-1.0	ISCEDL_2.0
ST001D01T_9.0	REPEAT
GRADE_-1.0	REPEAT
GRADE_-1.0	ST001D01T_9.0

---

---

k=3 的频繁项集

---

ISCEDL_2.0	REPEAT	ST127Q02TA_2.0
ISCEDL_2.0	ST001D01T_9.0	REPEAT
GRADE_-1.0	ISCEDL_2.0	REPEAT
GRADE_-1.0	ISCEDL_2.0	ST001D01T_9.0
GRADE_-1.0	ST001D01T_9.0	REPEAT

---

---

k=4 的频繁项集

---

GRADE_-1.0	ST001D01T_9.0	ISCEDL_2.0	REPEAT
------------	---------------	------------	--------

---

分别以置信度  $conf$  和置信度  $conf +$  提升度  $Lift = \frac{P(Y|X)}{P(Y)} > 1$  计算关联规则，结果见附录。

#### 4.2.2 Fre > 50

k=1 的频繁项集	
ISCEDL_2.0	
REPEAT	
ST127Q02TA_2.0	
ST001D01T_9.0	
GRADE_-1.0	
k=2 的频繁项集	
ISCEDL_2.0	REPEAT
ISCEDL_2.0	ST127Q02TA_2.0
REPEAT	ST127Q02TA_2.0
ISCEDL_2.0	ST001D01T_9.0
GRADE_-1.0	ISCEDL_2.0
ST001D01T_9.0	REPEAT
GRADE_-1.0	REPEAT
GRADE_-1.0	ST001D01T_9.0

---

k=3 的频繁项集

---

ISCEDL_2.0	REPEAT	ST127Q02TA_2.0
ISCEDL_2.0	ST001D01T_9.0	REPEAT
GRADE_-1.0	ISCEDL_2.0	REPEAT
GRADE_-1.0	ISCEDL_2.0	ST001D01T_9.0
GRADE_-1.0	ST001D01T_9.0	REPEAT

---

---

k=4 的频繁项集

---

GRADE_-1.0	ST001D01T_9.0	ISCEDL_2.0	REPEAT
------------	---------------	------------	--------

---

分别以置信度  $conf$  和置信度  $conf +$  提升度  $Lift = \frac{P(Y|X)}{P(Y)} > 1$  计算关联规则，结果见附录。

## 5 总结

每个特征的大致含义如下：

- REPEAT: 是否留级过
- GRADE: 所在年级是否滞后与同龄人
- ST127Q02TA: 留级的次数
- ST001D01T: 学生年级
- ISCEDL: 学生学段
- OCOD2: 父亲的职业

通过观察频繁项集与关联规则，发现

1. 清洗后的数据中复读的学生均为初中生。
2. 九年级的学生与复读具有强相关性
3. 九年级的学生大部分只留级过一次。
4. 去除低频项的阈值大小与结果几乎无关，父亲职业相对于其他特征对复读影响较小。

根据分析结果以及现实生活中对上述特征的经验，推测实验结果是合理的。

### 5.1 探索性分析（附录 6.3）

- 探索对于国家与上述特征的关联规则，除了满足上述规则以外，发现西班牙的学生与留级，并且留级两次强相关。



- 只考虑巴拿马与上述特征的关联规则，除了满足上述规则以外，发现巴拿马的学生在年级低于同龄人一年时留级过两次，推测留级可能是超前学习导致的。
- 只考虑墨西哥与上述特征的关联规则，基本满足上述规则
- 值考虑西班牙与上述特征的关联规则，除了满足上述规则以外，发现西班牙的学生在年级低于同龄人一年时留级过两次，推测留级可能是超前学习导致的。

## 6 附录

### 6.1 Fre > 10 的数据关联规则

#### 6.1.1 只考虑 conf 的关联规则

ISCEDL\_2.0-> REPEAT  
together support: 1.0 , conf: 1.0 , lift: 1.0  
REPEAT-> ISCEDL\_2.0  
together support: 1.0 , conf: 1.0 , lift: 1.0  
ISCEDL\_2.0-> ST127Q02TA\_2.0  
together support: 0.752 , conf: 0.752 , lift: 1.0  
ST127Q02TA\_2.0-> ISCEDL\_2.0  
together support: 0.752 , conf: 1.0 , lift: 1.0  
REPEAT-> ST127Q02TA\_2.0  
together support: 0.752 , conf: 0.752 , lift: 1.0  
ST127Q02TA\_2.0-> REPEAT  
together support: 0.752 , conf: 1.0 , lift: 1.0  
ST001D01T\_9.0-> ISCEDL\_2.0  
together support: 0.776 , conf: 1.0 , lift: 1.0  
ISCEDL\_2.0-> ST001D01T\_9.0  
together support: 0.776 , conf: 0.776 , lift: 1.0  
ISCEDL\_2.0-> GRADE\_-1.0  
together support: 0.776 , conf: 0.776 , lift: 1.0  
GRADE\_-1.0-> ISCEDL\_2.0  
together support: 0.776 , conf: 1.0 , lift: 1.0  
ST001D01T\_9.0-> REPEAT  
together support: 0.776 , conf: 1.0 , lift: 1.0  
REPEAT-> ST001D01T\_9.0  
together support: 0.776 , conf: 0.776 , lift: 1.0  
GRADE\_-1.0-> REPEAT  
together support: 0.776 , conf: 1.0 , lift: 1.0

REPEAT-> GRADE\_-1.0  
together support: 0.776 , conf: 0.776 , lift: 1.0  
ST001D01T\_9.0-> GRADE\_-1.0  
together support: 0.776 , conf: 1.0 , lift: 1.289  
GRADE\_-1.0-> ST001D01T\_9.0  
together support: 0.776 , conf: 1.0 , lift: 1.289  
ISCEDL\_2.0-> REPEAT,ST127Q02TA\_2.0  
together support: 0.752 , conf: 0.752 , lift: 1.0  
REPEAT-> ISCEDL\_2.0, ST127Q02TA\_2.0  
together support: 0.752 , conf: 0.752 , lift: 1.0  
ST127Q02TA\_2.0-> ISCEDL\_2.0, REPEAT  
together support: 0.752 , conf: 1.0 , lift: 1.0  
ST001D01T\_9.0-> ISCEDL\_2.0,REPEAT  
together support: 0.776 , conf: 1.0 , lift: 1.0  
ISCEDL\_2.0-> ST001D01T\_9.0,REPEAT  
together support: 0.776 , conf: 0.776 , lift: 1.0  
REPEAT-> ST001D01T\_9.0,ISCEDL\_2.0  
together support: 0.776 , conf: 0.776 , lift: 1.0  
GRADE\_-1.0-> ISCEDL\_2.0,REPEAT  
together support: 0.776 , conf: 1.0 , lift: 1.0  
ISCEDL\_2.0-> GRADE\_-1.0,REPEAT  
together support: 0.776 , conf: 0.776 , lift: 1.0  
REPEAT-> ISCEDL\_2.0,GRADE\_-1.0  
together support: 0.776 , conf: 0.776 , lift: 1.0  
ST001D01T\_9.0-> ISCEDL\_2.0,GRADE\_-1.0  
together support: 0.776 , conf: 1.0 , lift: 1.289  
ISCEDL\_2.0-> ST001D01T\_9.0,GRADE\_-1.0  
together support: 0.776 , conf: 0.776 , lift: 1.0  
GRADE\_-1.0-> ST001D01T\_9.0,ISCEDL\_2.0  
together support: 0.776 , conf: 1.0 , lift: 1.289  
GRADE\_-1.0-> ST001D01T\_9.0,REPEAT

together support: 0.776 , conf: 1.0 , lift: 1.289  
ST001D01T\_9.0-> GRADE\_-1.0,REPEAT  
together support: 0.776 , conf: 1.0 , lift: 1.289  
REPEAT-> ST001D01T\_9.0,GRADE\_-1.0  
together support: 0.776 , conf: 0.776 , lift: 1.0  
ISCEDL\_2.0-> REPEAT,ST001D01T\_9.0,GRADE\_-1.0  
together support: 0.776 , conf: 0.776 , lift: 1.0  
REPEAT-> ST001D01T\_9.0,ISCEDL\_2.0,GRADE\_-1.0  
together support: 0.776 , conf: 0.776 , lift: 1.0  
GRADE\_-1.0-> ST001D01T\_9.0,ISCEDL\_2.0,REPEAT  
together support: 0.776 , conf: 1.0 , lift: 1.289  
ST001D01T\_9.0-> REPEAT,ISCEDL\_2.0,GRADE\_-1.0  
together support: 0.776 , conf: 1.0 , lift: 1.289

### 6.1.2 同时考虑 conf 和 Lift 的关联规则

GRADE\_-1.0 -> ST001D01T\_9.0  
together support: 0.776 , conf: 1.0 , lift: 1.289  
ST001D01T\_9.0 -> GRADE\_-1.0  
together support: 0.776 , conf: 1.0 , lift: 1.289  
GRADE\_-1.0 -> ST001D01T\_9.0,ISCEDL\_2.0  
together support: 0.776 , conf: 1.0 , lift: 1.289  
ST001D01T\_9.0 -> GRADE\_-1.0,ISCEDL\_2.0  
together support: 0.776 , conf: 1.0 , lift: 1.289  
GRADE\_-1.0 -> REPEAT,ST001D01T\_9.0  
together support: 0.776 , conf: 1.0 , lift: 1.289  
ST001D01T\_9.0 -> REPEAT,GRADE\_-1.0  
together support: 0.776 , conf: 1.0 , lift: 1.289  
GRADE\_-1.0 -> REPEAT,ST001D01T\_9.0,ISCEDL\_2.0  
together support: 0.776 , conf: 1.0 , lift: 1.289  
ST001D01T\_9.0 -> REPEAT,GRADE\_-1.0,ISCEDL\_2.0

together support: 0.776 , conf: 1.0 , lift: 1.289

## 6.2 Fre > 50 的数据关联规则

### 6.2.1 只考虑 conf 的关联规则

ISCEDL\_2.0 -> REPEAT

together support: 1.0 , conf: 1.0

REPEAT -> ISCEDL\_2.0

together support: 1.0 , conf: 1.0

ISCEDL\_2.0 -> ST127Q02TA\_2.0

together support: 0.752 , conf: 0.752

ST127Q02TA\_2.0 -> ISCEDL\_2.0

together support: 0.752 , conf: 1.0

REPEAT -> ST127Q02TA\_2.0

together support: 0.752 , conf: 0.752

ST127Q02TA\_2.0 -> REPEAT

together support: 0.752 , conf: 1.0

ISCEDL\_2.0 -> ST001D01T\_9.0

together support: 0.776 , conf: 0.776

ST001D01T\_9.0 -> ISCEDL\_2.0

together support: 0.776 , conf: 1.0

GRADE\_-1.0 -> ISCEDL\_2.0

together support: 0.776 , conf: 1.0

ISCEDL\_2.0 -> GRADE\_-1.0

together support: 0.776 , conf: 0.776

ST001D01T\_9.0 -> REPEAT

together support: 0.776 , conf: 1.0

REPEAT -> ST001D01T\_9.0

together support: 0.776 , conf: 0.776

GRADE\_-1.0 -> REPEAT

together support: 0.776 , conf: 1.0  
REPEAT -> GRADE\_-1.0  
together support: 0.776 , conf: 0.776  
GRADE\_-1.0 -> ST001D01T\_9.0  
together support: 0.776 , conf: 1.0  
ST001D01T\_9.0 -> GRADE\_-1.0  
together support: 0.776 , conf: 1.0  
ISCEDL\_2.0 -> REPEAT,ST127Q02TA\_2.0  
together support: 0.752 , conf: 0.752  
REPEAT -> ISCEDL\_2.0,ST127Q02TA\_2.0  
together support: 0.752 , conf: 0.752  
ST127Q02TA\_2.0 -> ISCEDL\_2.0,REPEAT  
together support: 0.752 , conf: 1.0  
ISCEDL\_2.0 -> ST001D01T\_9.0,REPEAT  
together support: 0.776 , conf: 0.776  
ST001D01T\_9.0 -> ISCEDL\_2.0,REPEAT  
together support: 0.776 , conf: 1.0  
REPEAT -> ISCEDL\_2.0,ST001D01T\_9.0  
together support: 0.776 , conf: 0.776  
GRADE\_-1.0 -> ISCEDL\_2.0,REPEAT  
together support: 0.776 , conf: 1.0  
ISCEDL\_2.0 -> GRADE\_-1.0,REPEAT  
together support: 0.776 , conf: 0.776  
REPEAT -> GRADE\_-1.0,ISCEDL\_2.0  
together support: 0.776 , conf: 0.776  
GRADE\_-1.0 -> ISCEDL\_2.0,ST001D01T\_9.0  
together support: 0.776 , conf: 1.0  
ISCEDL\_2.0 -> GRADE\_-1.0,ST001D01T\_9.0  
together support: 0.776 , conf: 0.776  
ST001D01T\_9.0 -> GRADE\_-1.0,ISCEDL\_2.0  
together support: 0.776 , conf: 1.0

GRADE\_-1.0 -> ST001D01T\_9.0,REPEAT  
together support: 0.776 , conf: 1.0  
ST001D01T\_9.0 -> GRADE\_-1.0,REPEAT  
together support: 0.776 , conf: 1.0  
REPEAT -> GRADE\_-1.0,ST001D01T\_9.0  
together support: 0.776 , conf: 0.776  
GRADE\_-1.0 -> ISCEDL\_2.0,ST001D01T\_9.0,REPEAT  
together support: 0.776 , conf: 1.0  
ST001D01T\_9.0 -> GRADE\_-1.0,ISCEDL\_2.0,REPEAT  
together support: 0.776 , conf: 1.0  
ISCEDL\_2.0 -> GRADE\_-1.0,ST001D01T\_9.0,REPEAT  
together support: 0.776 , conf: 0.776  
REPEAT -> GRADE\_-1.0,ISCEDL\_2.0,ST001D01T\_9.0  
together support: 0.776 , conf: 0.776

### 6.2.2 同时考虑 conf 和 Lift 的关联规则

GRADE\_-1.0 -> ST001D01T\_9.0  
together support: 0.776 , conf: 1.0 , lift: 1.289  
ST001D01T\_9.0 -> GRADE\_-1.0  
together support: 0.776 , conf: 1.0 , lift: 1.289  
GRADE\_-1.0 -> ST001D01T\_9.0,ISCEDL\_2.0  
together support: 0.776 , conf: 1.0 , lift: 1.289  
ST001D01T\_9.0 -> GRADE\_-1.0,ISCEDL\_2.0  
together support: 0.776 , conf: 1.0 , lift: 1.289  
GRADE\_-1.0 -> REPEAT,ST001D01T\_9.0  
together support: 0.776 , conf: 1.0 , lift: 1.289  
ST001D01T\_9.0 -> REPEAT,GRADE\_-1.0  
together support: 0.776 , conf: 1.0 , lift: 1.289  
GRADE\_-1.0 -> REPEAT,ST001D01T\_9.0,ISCEDL\_2.0  
together support: 0.776 , conf: 1.0 , lift: 1.289

ST001D01T\_9.0 -> REPEAT,GRADE\_-1.0,ISCEDL\_2.0

together support: 0.776 , conf: 1.0 , lift: 1.289

## 6.3 考虑 CNTRYID 的关联规则

### 6.3.1 全部 CNTRYID-同时考虑 conf 和 Lift 的关联规则

ST001D01T\_9.0 -> GRADE\_-1.0

together support: 0.776 , conf: 1.0 , lift: 1.289

GRADE\_-1.0 -> ST001D01T\_9.0

together support: 0.776 , conf: 1.0 , lift: 1.289

ST127Q02TA\_2.0 -> CNTRYID\_724.0

together support: 0.67 , conf: 0.891 , lift: 1.04

CNTRYID\_724.0 -> ST127Q02TA\_2.0

together support: 0.67 , conf: 0.782 , lift: 1.04

ST001D01T\_9.0 -> CNTRYID\_724.0

together support: 0.666 , conf: 0.858 , lift: 1.002

CNTRYID\_724.0 -> ST001D01T\_9.0

together support: 0.666 , conf: 0.777 , lift: 1.002

GRADE\_-1.0 -> CNTRYID\_724.0

together support: 0.666 , conf: 0.858 , lift: 1.002

CNTRYID\_724.0 -> GRADE\_-1.0

together support: 0.666 , conf: 0.777 , lift: 1.002

ST001D01T\_9.0 -> ISCEDL\_2.0,GRADE\_-1.0

together support: 0.776 , conf: 1.0 , lift: 1.289

GRADE\_-1.0 -> ST001D01T\_9.0,ISCEDL\_2.0

together support: 0.776 , conf: 1.0 , lift: 1.289

ST001D01T\_9.0 -> REPEAT,GRADE\_-1.0

together support: 0.776 , conf: 1.0 , lift: 1.289

GRADE\_-1.0 -> ST001D01T\_9.0,REPEAT

together support: 0.776 , conf: 1.0 , lift: 1.289



ST127Q02TA\_2.0 -> ISCEDL\_2.0,CNTRYID\_724.0  
together support: 0.67 , conf: 0.891 , lift: 1.04  
CNTRYID\_724.0 -> ISCEDL\_2.0,ST127Q02TA\_2.0  
together support: 0.67 , conf: 0.782 , lift: 1.04  
ST127Q02TA\_2.0 -> REPEAT,CNTRYID\_724.0  
together support: 0.67 , conf: 0.891 , lift: 1.04  
CNTRYID\_724.0 -> REPEAT,ST127Q02TA\_2.0  
together support: 0.67 , conf: 0.782 , lift: 1.04  
ST001D01T\_9.0 -> ISCEDL\_2.0,CNTRYID\_724.0  
together support: 0.666 , conf: 0.858 , lift: 1.002  
CNTRYID\_724.0 -> ST001D01T\_9.0,ISCEDL\_2.0  
together support: 0.666 , conf: 0.777 , lift: 1.002  
GRADE\_-1.0 -> ISCEDL\_2.0,CNTRYID\_724.0  
together support: 0.666 , conf: 0.858 , lift: 1.002  
CNTRYID\_724.0 -> ISCEDL\_2.0,GRADE\_-1.0  
together support: 0.666 , conf: 0.777 , lift: 1.002  
ST001D01T\_9.0 -> REPEAT,CNTRYID\_724.0  
together support: 0.666 , conf: 0.858 , lift: 1.002  
CNTRYID\_724.0 -> ST001D01T\_9.0,REPEAT  
together support: 0.666 , conf: 0.777 , lift: 1.002  
GRADE\_-1.0 -> REPEAT,CNTRYID\_724.0  
together support: 0.666 , conf: 0.858 , lift: 1.002  
CNTRYID\_724.0 -> REPEAT,GRADE\_-1.0  
together support: 0.666 , conf: 0.777 , lift: 1.002  
ST001D01T\_9.0 -> GRADE\_-1.0,CNTRYID\_724.0  
together support: 0.666 , conf: 0.858 , lift: 1.289  
GRADE\_-1.0 -> ST001D01T\_9.0,CNTRYID\_724.0  
together support: 0.666 , conf: 0.858 , lift: 1.289  
CNTRYID\_724.0 -> ST001D01T\_9.0,GRADE\_-1.0  
together support: 0.666 , conf: 0.777 , lift: 1.002  
ST001D01T\_9.0 -> ISCEDL\_2.0,REPEAT,GRADE\_-1.0

together support: 0.776 , conf: 1.0 , lift: 1.289  
GRADE\_-1.0 -> ST001D01T\_9.0,ISCEDL\_2.0,REPEAT  
together support: 0.776 , conf: 1.0 , lift: 1.289  
CNTRYID\_724.0 -> ISCEDL\_2.0,REPEAT,ST127Q02TA\_2.0  
together support: 0.67 , conf: 0.782 , lift: 1.04  
ST127Q02TA\_2.0 -> ISCEDL\_2.0,REPEAT,CNTRYID\_724.0  
together support: 0.67 , conf: 0.891 , lift: 1.04  
CNTRYID\_724.0 -> ST001D01T\_9.0,ISCEDL\_2.0,REPEAT  
together support: 0.666 , conf: 0.777 , lift: 1.002  
ST001D01T\_9.0 -> ISCEDL\_2.0,REPEAT,CNTRYID\_724.0  
together support: 0.666 , conf: 0.858 , lift: 1.002  
CNTRYID\_724.0 -> ISCEDL\_2.0,REPEAT,GRADE\_-1.0  
together support: 0.666 , conf: 0.777 , lift: 1.002  
GRADE\_-1.0 -> ISCEDL\_2.0,REPEAT,CNTRYID\_724.0  
together support: 0.666 , conf: 0.858 , lift: 1.002  
CNTRYID\_724.0 -> ST001D01T\_9.0,ISCEDL\_2.0,GRADE\_-1.0  
together support: 0.666 , conf: 0.777 , lift: 1.002  
ST001D01T\_9.0 -> ISCEDL\_2.0,GRADE\_-1.0,CNTRYID\_724.0  
together support: 0.666 , conf: 0.858 , lift: 1.289  
GRADE\_-1.0 -> ST001D01T\_9.0,ISCEDL\_2.0,CNTRYID\_724.0  
together support: 0.666 , conf: 0.858 , lift: 1.289  
CNTRYID\_724.0 -> ST001D01T\_9.0,REPEAT,GRADE\_-1.0  
together support: 0.666 , conf: 0.777 , lift: 1.002  
ST001D01T\_9.0 -> REPEAT,GRADE\_-1.0,CNTRYID\_724.0  
together support: 0.666 , conf: 0.858 , lift: 1.289  
GRADE\_-1.0 -> ST001D01T\_9.0,REPEAT,CNTRYID\_724.0  
together support: 0.666 , conf: 0.858 , lift: 1.289  
CNTRYID\_724.0 -> ST001D01T\_9.0,ISCEDL\_2.0,REPEAT,GRADE\_-1.0  
together support: 0.666 , conf: 0.777 , lift: 1.002  
ST001D01T\_9.0 -> ISCEDL\_2.0,REPEAT,GRADE\_-1.0,CNTRYID\_724.0  
together support: 0.666 , conf: 0.858 , lift: 1.289

GRADE\_-1.0 -> ST001D01T\_9.0,ISCEDL\_2.0,REPEAT,CNTRYID\_724.0  
together support: 0.666 , conf: 0.858 , lift: 1.289

### 6.3.2 CNTRYID591.0-同时考虑 conf 和 Lift 的关联规则

ST001D01T\_9.0 -> ST127Q02TA\_2.0  
together support: 0.612 , conf: 0.771 , lift: 1.086  
ST127Q02TA\_2.0 -> ST001D01T\_9.0  
together support: 0.612 , conf: 0.862 , lift: 1.086  
ST127Q02TA\_2.0 -> GRADE\_-1.0  
together support: 0.612 , conf: 0.862 , lift: 1.086  
GRADE\_-1.0 -> ST127Q02TA\_2.0  
together support: 0.612 , conf: 0.771 , lift: 1.086  
ST001D01T\_9.0 -> GRADE\_-1.0  
together support: 0.793 , conf: 1.0 , lift: 1.261  
GRADE\_-1.0 -> ST001D01T\_9.0  
together support: 0.793 , conf: 1.0 , lift: 1.261  
ST001D01T\_9.0 -> ST127Q02TA\_2.0,GRADE\_-1.0  
together support: 0.612 , conf: 0.771 , lift: 1.261  
ST127Q02TA\_2.0 -> ST001D01T\_9.0,GRADE\_-1.0  
together support: 0.612 , conf: 0.862 , lift: 1.086  
GRADE\_-1.0 -> ST001D01T\_9.0,ST127Q02TA\_2.0  
together support: 0.612 , conf: 0.771 , lift: 1.261  
ST001D01T\_9.0 -> ST127Q02TA\_2.0,ISCEDL\_2.0  
together support: 0.612 , conf: 0.771 , lift: 1.086  
ST127Q02TA\_2.0 -> ST001D01T\_9.0,ISCEDL\_2.0  
together support: 0.612 , conf: 0.862 , lift: 1.086  
ST001D01T\_9.0 -> ST127Q02TA\_2.0,REPEAT  
together support: 0.612 , conf: 0.771 , lift: 1.086  
ST127Q02TA\_2.0 -> ST001D01T\_9.0,REPEAT  
together support: 0.612 , conf: 0.862 , lift: 1.086

ST127Q02TA\_2.0 -> GRADE\_-1.0,ISCEDL\_2.0  
together support: 0.612 , conf: 0.862 , lift: 1.086  
GRADE\_-1.0 -> ST127Q02TA\_2.0,ISCEDL\_2.0  
together support: 0.612 , conf: 0.771 , lift: 1.086  
ST127Q02TA\_2.0 -> GRADE\_-1.0,REPEAT  
together support: 0.612 , conf: 0.862 , lift: 1.086  
GRADE\_-1.0 -> ST127Q02TA\_2.0,REPEAT  
together support: 0.612 , conf: 0.771 , lift: 1.086  
ST001D01T\_9.0 -> GRADE\_-1.0,ISCEDL\_2.0  
together support: 0.793 , conf: 1.0 , lift: 1.261  
GRADE\_-1.0 -> ST001D01T\_9.0,ISCEDL\_2.0  
together support: 0.793 , conf: 1.0 , lift: 1.261  
ST001D01T\_9.0 -> GRADE\_-1.0,REPEAT  
together support: 0.793 , conf: 1.0 , lift: 1.261  
GRADE\_-1.0 -> ST001D01T\_9.0,REPEAT  
together support: 0.793 , conf: 1.0 , lift: 1.261  
ST001D01T\_9.0 -> ST127Q02TA\_2.0,GRADE\_-1.0,ISCEDL\_2.0  
together support: 0.612 , conf: 0.771 , lift: 1.261  
ST127Q02TA\_2.0 -> ST001D01T\_9.0,GRADE\_-1.0,ISCEDL\_2.0  
together support: 0.612 , conf: 0.862 , lift: 1.086  
GRADE\_-1.0 -> ST001D01T\_9.0,ST127Q02TA\_2.0,ISCEDL\_2.0  
together support: 0.612 , conf: 0.771 , lift: 1.261  
ST001D01T\_9.0 -> ST127Q02TA\_2.0,REPEAT,GRADE\_-1.0  
together support: 0.612 , conf: 0.771 , lift: 1.261  
ST127Q02TA\_2.0 -> ST001D01T\_9.0,GRADE\_-1.0,REPEAT  
together support: 0.612 , conf: 0.862 , lift: 1.086  
GRADE\_-1.0 -> ST001D01T\_9.0,ST127Q02TA\_2.0,REPEAT  
together support: 0.612 , conf: 0.771 , lift: 1.261  
ST001D01T\_9.0 -> ST127Q02TA\_2.0,REPEAT,ISCEDL\_2.0  
together support: 0.612 , conf: 0.771 , lift: 1.086  
ST127Q02TA\_2.0 -> ST001D01T\_9.0,REPEAT,ISCEDL\_2.0

together support: 0.612 , conf: 0.862 , lift: 1.086  
ST127Q02TA\_2.0 -> GRADE\_-1.0,REPEAT,ISCEDL\_2.0  
together support: 0.612 , conf: 0.862 , lift: 1.086  
GRADE\_-1.0 -> ST127Q02TA\_2.0,REPEAT,ISCEDL\_2.0  
together support: 0.612 , conf: 0.771 , lift: 1.086  
ST001D01T\_9.0 -> GRADE\_-1.0,REPEAT,ISCEDL\_2.0  
together support: 0.793 , conf: 1.0 , lift: 1.261  
GRADE\_-1.0 -> ST001D01T\_9.0,REPEAT,ISCEDL\_2.0  
together support: 0.793 , conf: 1.0 , lift: 1.261  
ST001D01T\_9.0 -> ST127Q02TA\_2.0,REPEAT,ISCEDL\_2.0,GRADE\_-1.0  
together support: 0.612 , conf: 0.771 , lift: 1.261  
ST127Q02TA\_2.0 -> ST001D01T\_9.0,GRADE\_-1.0,REPEAT,ISCEDL\_2.0  
together support: 0.612 , conf: 0.862 , lift: 1.086  
GRADE\_-1.0 -> ST001D01T\_9.0,ST127Q02TA\_2.0,REPEAT,ISCEDL\_2.0  
together support: 0.612 , conf: 0.771 , lift: 1.261

### 6.3.3 CNTRYID724.0-同时考虑 conf 和 Lift 的关联规则

ST127Q02TA\_2.0 -> ST001D01T\_9.0  
together support: 0.619 , conf: 0.792 , lift: 1.018  
ST001D01T\_9.0 -> ST127Q02TA\_2.0  
together support: 0.619 , conf: 0.797 , lift: 1.018  
ST127Q02TA\_2.0 -> GRADE\_-1.0  
together support: 0.619 , conf: 0.792 , lift: 1.018  
GRADE\_-1.0 -> ST127Q02TA\_2.0  
together support: 0.619 , conf: 0.797 , lift: 1.018  
ST001D01T\_9.0 -> GRADE\_-1.0  
together support: 0.777 , conf: 1.0 , lift: 1.286  
GRADE\_-1.0 -> ST001D01T\_9.0  
together support: 0.777 , conf: 1.0 , lift: 1.286  
ST127Q02TA\_2.0 -> ST001D01T\_9.0,GRADE\_-1.0

together support: 0.619 , conf: 0.792 , lift: 1.018  
ST001D01T\_9.0 -> ST127Q02TA\_2.0,GRADE\_-1.0  
together support: 0.619 , conf: 0.797 , lift: 1.286  
GRADE\_-1.0 -> ST127Q02TA\_2.0,ST001D01T\_9.0  
together support: 0.619 , conf: 0.797 , lift: 1.286  
ST127Q02TA\_2.0 -> ST001D01T\_9.0,ISCEDL\_2.0  
together support: 0.619 , conf: 0.792 , lift: 1.018  
ST001D01T\_9.0 -> ST127Q02TA\_2.0,ISCEDL\_2.0  
together support: 0.619 , conf: 0.797 , lift: 1.018  
ST127Q02TA\_2.0 -> ST001D01T\_9.0,REPEAT  
together support: 0.619 , conf: 0.792 , lift: 1.018  
ST001D01T\_9.0 -> ST127Q02TA\_2.0,REPEAT  
together support: 0.619 , conf: 0.797 , lift: 1.018  
ST127Q02TA\_2.0 -> GRADE\_-1.0,ISCEDL\_2.0  
together support: 0.619 , conf: 0.792 , lift: 1.018  
GRADE\_-1.0 -> ST127Q02TA\_2.0,ISCEDL\_2.0  
together support: 0.619 , conf: 0.797 , lift: 1.018  
ST127Q02TA\_2.0 -> GRADE\_-1.0,REPEAT  
together support: 0.619 , conf: 0.792 , lift: 1.018  
GRADE\_-1.0 -> ST127Q02TA\_2.0,REPEAT  
together support: 0.619 , conf: 0.797 , lift: 1.018  
ST001D01T\_9.0 -> GRADE\_-1.0,ISCEDL\_2.0  
together support: 0.777 , conf: 1.0 , lift: 1.286  
GRADE\_-1.0 -> ST001D01T\_9.0,ISCEDL\_2.0  
together support: 0.777 , conf: 1.0 , lift: 1.286  
ST001D01T\_9.0 -> GRADE\_-1.0,REPEAT  
together support: 0.777 , conf: 1.0 , lift: 1.286  
GRADE\_-1.0 -> ST001D01T\_9.0,REPEAT  
together support: 0.777 , conf: 1.0 , lift: 1.286  
ST001D01T\_9.0 -> ST127Q02TA\_2.0,GRADE\_-1.0,ISCEDL\_2.0  
together support: 0.619 , conf: 0.797 , lift: 1.286

ST127Q02TA\_2.0 -> ST001D01T\_9.0,ISCEDL\_2.0,GRADE\_-1.0  
together support: 0.619 , conf: 0.792 , lift: 1.018  
GRADE\_-1.0 -> ST127Q02TA\_2.0,ST001D01T\_9.0,ISCEDL\_2.0  
together support: 0.619 , conf: 0.797 , lift: 1.286  
ST001D01T\_9.0 -> ST127Q02TA\_2.0,GRADE\_-1.0,REPEAT  
together support: 0.619 , conf: 0.797 , lift: 1.286  
ST127Q02TA\_2.0 -> ST001D01T\_9.0,GRADE\_-1.0,REPEAT  
together support: 0.619 , conf: 0.792 , lift: 1.018  
GRADE\_-1.0 -> ST127Q02TA\_2.0,ST001D01T\_9.0,REPEAT  
together support: 0.619 , conf: 0.797 , lift: 1.286  
ST001D01T\_9.0 -> ST127Q02TA\_2.0,ISCEDL\_2.0,REPEAT  
together support: 0.619 , conf: 0.797 , lift: 1.018  
ST127Q02TA\_2.0 -> ST001D01T\_9.0,ISCEDL\_2.0,REPEAT  
together support: 0.619 , conf: 0.792 , lift: 1.018  
ST127Q02TA\_2.0 -> REPEAT,GRADE\_-1.0,ISCEDL\_2.0  
together support: 0.619 , conf: 0.792 , lift: 1.018  
GRADE\_-1.0 -> ST127Q02TA\_2.0,ISCEDL\_2.0,REPEAT  
together support: 0.619 , conf: 0.797 , lift: 1.018  
ST001D01T\_9.0 -> GRADE\_-1.0,ISCEDL\_2.0,REPEAT  
together support: 0.777 , conf: 1.0 , lift: 1.286  
GRADE\_-1.0 -> ST001D01T\_9.0,ISCEDL\_2.0,REPEAT  
together support: 0.777 , conf: 1.0 , lift: 1.286  
ST127Q02TA\_2.0 -> ST001D01T\_9.0,GRADE\_-1.0,ISCEDL\_2.0,REPEAT  
together support: 0.619 , conf: 0.792 , lift: 1.018  
GRADE\_-1.0 -> ST127Q02TA\_2.0,ST001D01T\_9.0,ISCEDL\_2.0,REPEAT  
together support: 0.619 , conf: 0.797 , lift: 1.286  
ST001D01T\_9.0 -> ST127Q02TA\_2.0,GRADE\_-1.0,ISCEDL\_2.0,REPEAT  
together support: 0.619 , conf: 0.797 , lift: 1.286

#### 6.3.4 CNTRYID484.0-同时考虑 conf 和 Lift 的关联规则

ST001D01T\_9.0  $\rightarrow$  GRADE\_-1.0 together support: 0.857 , conf: 1.0 , lift: 1.167

GRADE\_-1.0  $\rightarrow$  ST001D01T\_9.0

together support: 0.857 , conf: 1.0 , lift: 1.167

ST001D01T\_9.0  $\rightarrow$  GRADE\_-1.0,ISCEDL\_2.0

together support: 0.857 , conf: 1.0 , lift: 1.167

GRADE\_-1.0  $\rightarrow$  ST001D01T\_9.0,ISCEDL\_2.0

together support: 0.857 , conf: 1.0 , lift: 1.167

ST001D01T\_9.0  $\rightarrow$  GRADE\_-1.0,REPEAT

together support: 0.857 , conf: 1.0 , lift: 1.167

GRADE\_-1.0  $\rightarrow$  ST001D01T\_9.0,REPEAT

together support: 0.857 , conf: 1.0 , lift: 1.167

ST001D01T\_9.0  $\rightarrow$  GRADE\_-1.0,REPEAT,ISCEDL\_2.0

together support: 0.857 , conf: 1.0 , lift: 1.167

GRADE\_-1.0  $\rightarrow$  ST001D01T\_9.0,REPEAT,ISCEDL\_2.0

together support: 0.857 , conf: 1.0 , lift: 1.167

#### 6.3.5 CNTRYID152.0-同时考虑 conf 和 Lift 的关联规则

数据量过少无法产生关联规则。

#### 6.3.6 CNTRYID214.0 -同时考虑 conf 和 Lift 的关联规则

数据量过少无法产生关联规则。