《大数据分析方法》课程实验报告

学号： 2017202105

姓名： 林子豪

专业： 计算机科学与技术

班级： 2017级3班

队名：

**实验一：关联规则算法实验**

**实验目的：**在掌握关联规则算法的原理的基础上，能够应用关联规则算法解决实际问题。

**实验内容：**根据实验数据，采用Apriori等关联规则发现算法，给出相关关联规则。

**实验要求：**给出数据预处理过程、关联规则发现算法及发现关联规则，并对关联规则结果进行分析说明。

**实验题目：**蔬菜价格相关性分析

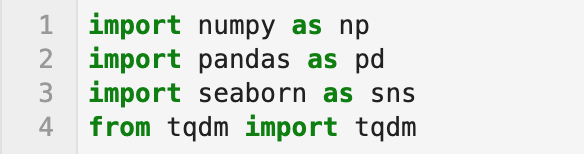
蔬菜的价格会受季节、天气等多方面因素的影响，但许多会出现同涨或者同跌等现象，请根据给出的蔬菜价格数据，采用关联规则发现算法，发现哪些蔬菜之间具有同涨、同跌或者涨跌不同步的现象。

数据格式如下：

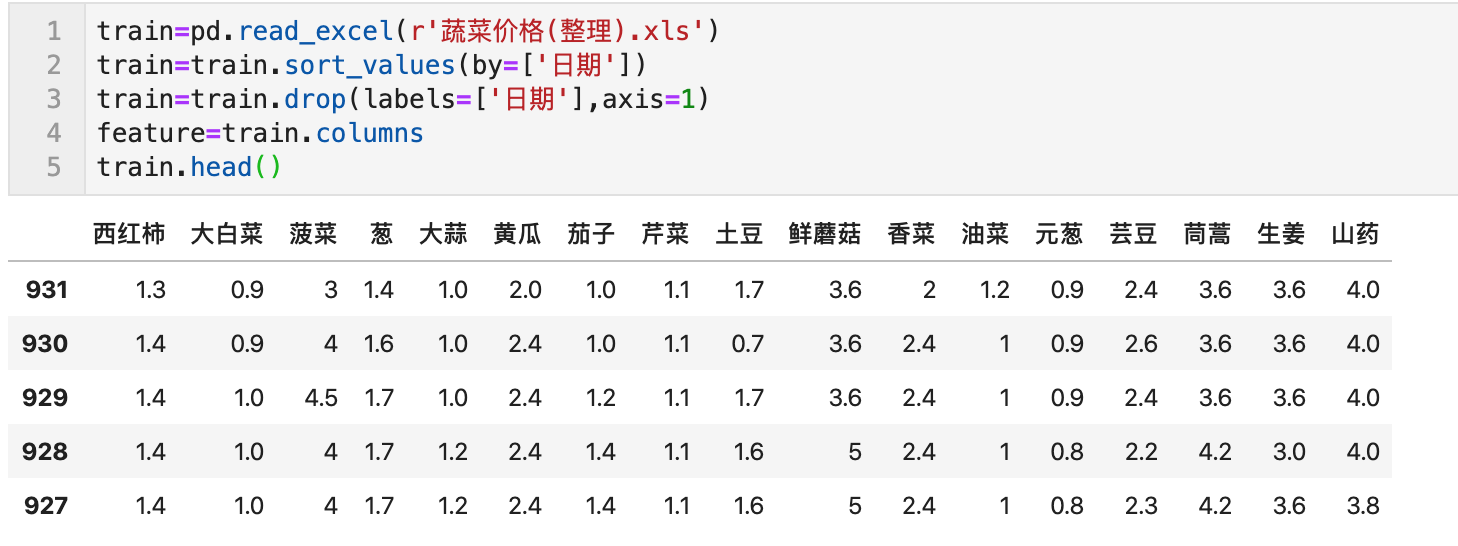
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 日期 | 蔬菜名 | 价格 | 肉食禽蛋 | 批发价格 |
| 2009/6/11 | 大白菜 | 0.7 | 精猪肚 | 28 |
| 2009/6/11 | 青萝卜 | 1.2 | 猪口条 | 28 |
| 2009/6/11 | 胡萝卜 | 2 | 猪心 | 24 |
| 2009/6/11 | 水萝卜 |  | 猪肝 | 7 |
| 2009/6/11 | 葱 | 2.5 | 猪大肠 | 22 |
| 2009/6/11 | 芽葱 |  | 猪蹄 | 15 |
| 2009/6/11 | 元葱 | 0.8 | 猪大排 | 20 |
| 2009/6/11 | 生姜 | 3 | 猪肋排 | 25 |
| 2009/6/11 | 大蒜 | 3 | 精猪肉 | 17 |
| 2009/6/11 | 土豆 | 1.3 | 精牛肉 | 30 |
| 2009/6/11 | 菠菜 | 1.4 | 白条肉 | 16 |
| 2009/6/11 | 韭菜 | 1 | 羊肉片 | 32 |
| 2009/6/11 | 芹菜 | 1.5 | 西装鸡 | 12 |
| 2009/6/11 | 油菜 | 1 | 鸡大腿 | 12 |
| 2009/6/11 | 苔菜 |  | 鸡翅 | 18 |
| 2009/6/11 | 香菜 | 2 | 鸡爪 | 16 |
| 2009/6/11 | 大头菜 | 0.6 | 鸡胸肉 | 14 |
| 2009/6/11 | 芸豆 | 1.2 | 翅根 | 14 |
| 2009/6/11 | 豆角 | 2 | 翅中 | 26.2 |
| 2009/6/11 | 茄子 | 1.2 |  |  |
| 2009/6/11 | 青椒 | 1.7 | 鸡心 | 11 |
| 2009/6/11 | 西红柿 | 1.4 | 松花蛋 |  |
| 2009/6/11 | 黄瓜 | 1.1 | 咸鸭蛋 |  |
| 2009/6/11 | 茭瓜 | 0.7 | 鸡蛋 | 6.6 |
| 2009/6/11 | 冬瓜 | 1.6 |  |  |

**实验过程：**

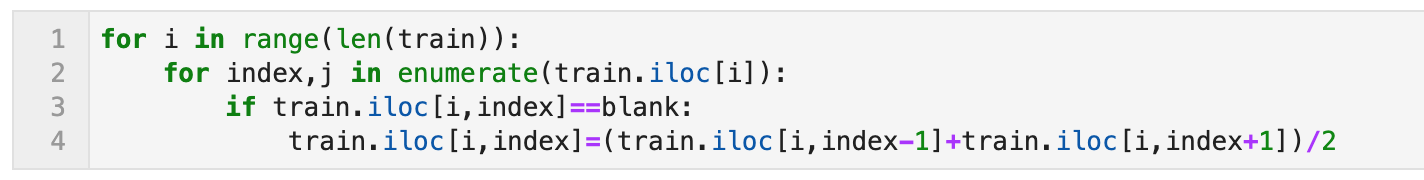
1.引入需要的包



2.读入数据并对日期排序

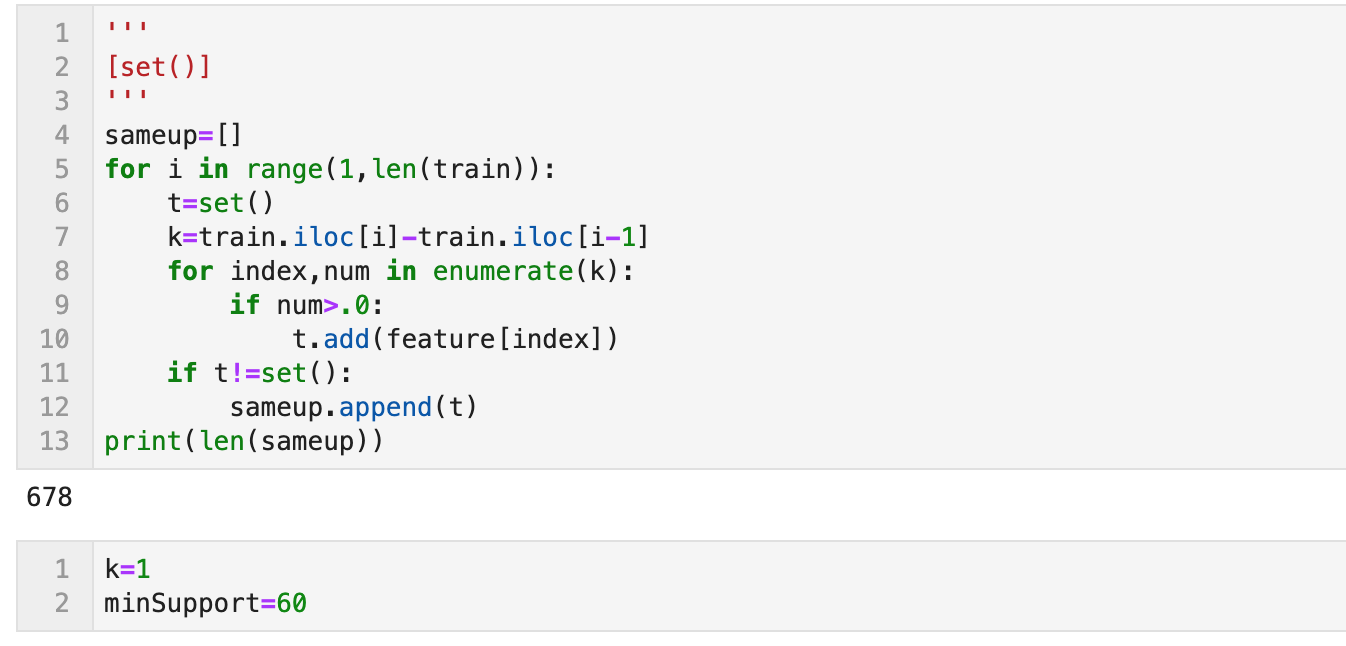


3.处理缺失值



4.处理单项集，发现一共有678个单项集

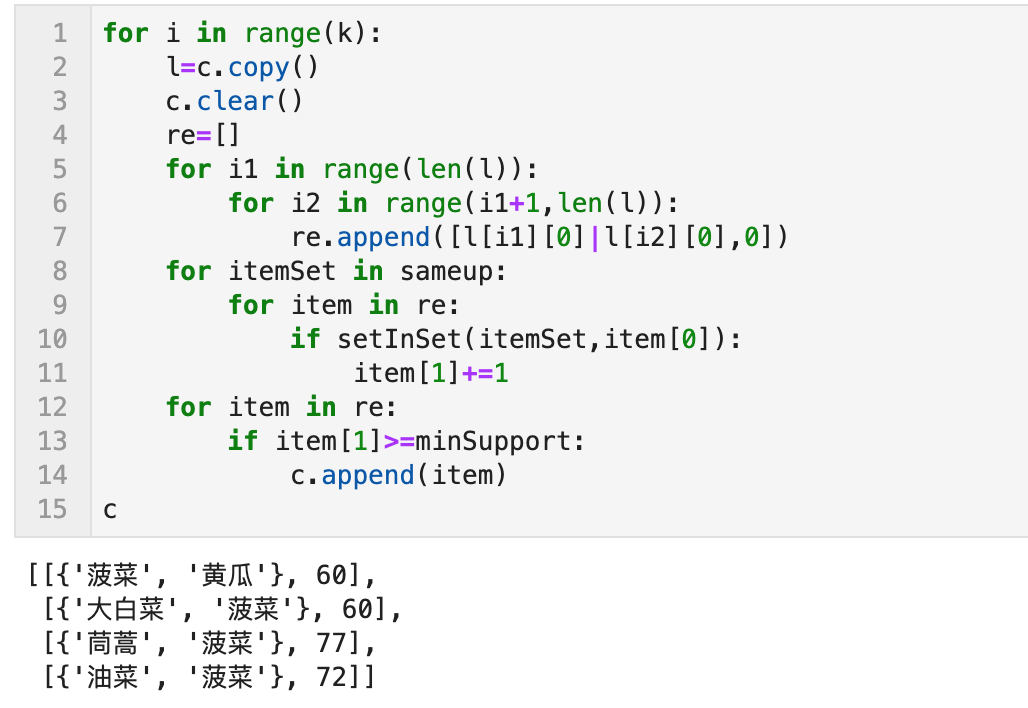
并且设置迭代次数和最小支持度



5.得到满足最小支持度的单项集



1. 在k次迭代后得到的频繁二项集和其出现次数



1. 