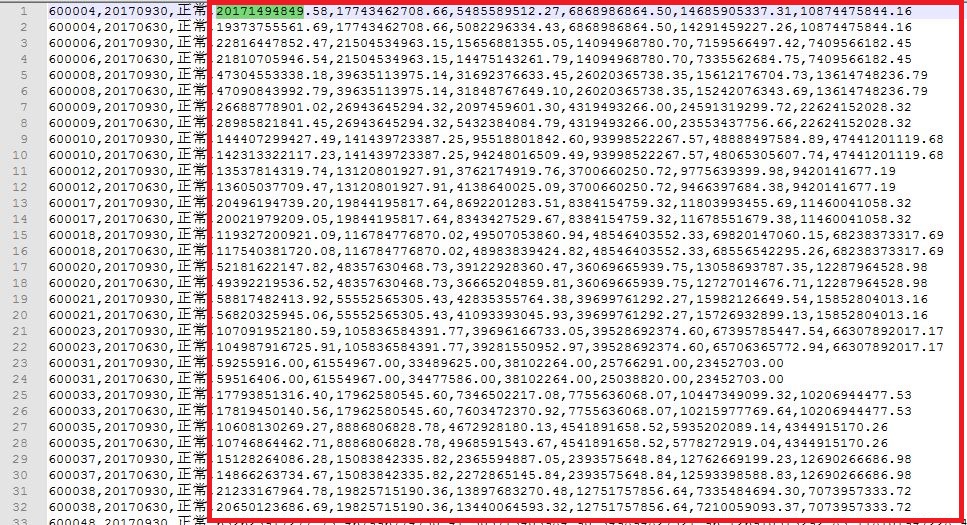
目前kmeans聚类算法中的K为人为设定为k=3

经过会计规则和专家规则处理后得到的部分数据如下：



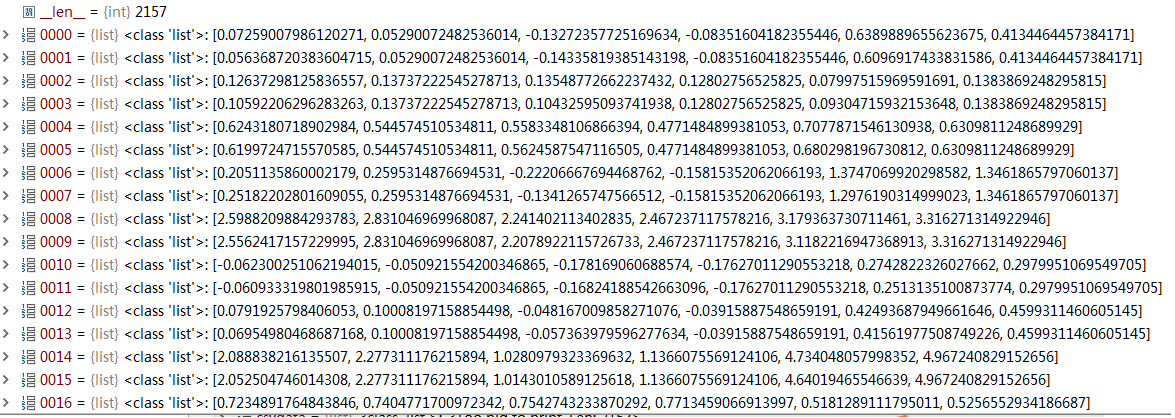
没有经过数据预处理聚类结果：

第一类数据占比：96.38386648122392%

第二类数据占比：0.3708854891052388%

第三类数据占比：3.2452480296708393%

预处理后的部分数据如下：



数据预处理后的分类结果（中心化和标准化预处理）

第一类数据占比：96.6156699119147%

第二类数据占比：0.3708854891052388%

第三类数据占比：3.013444598980065%

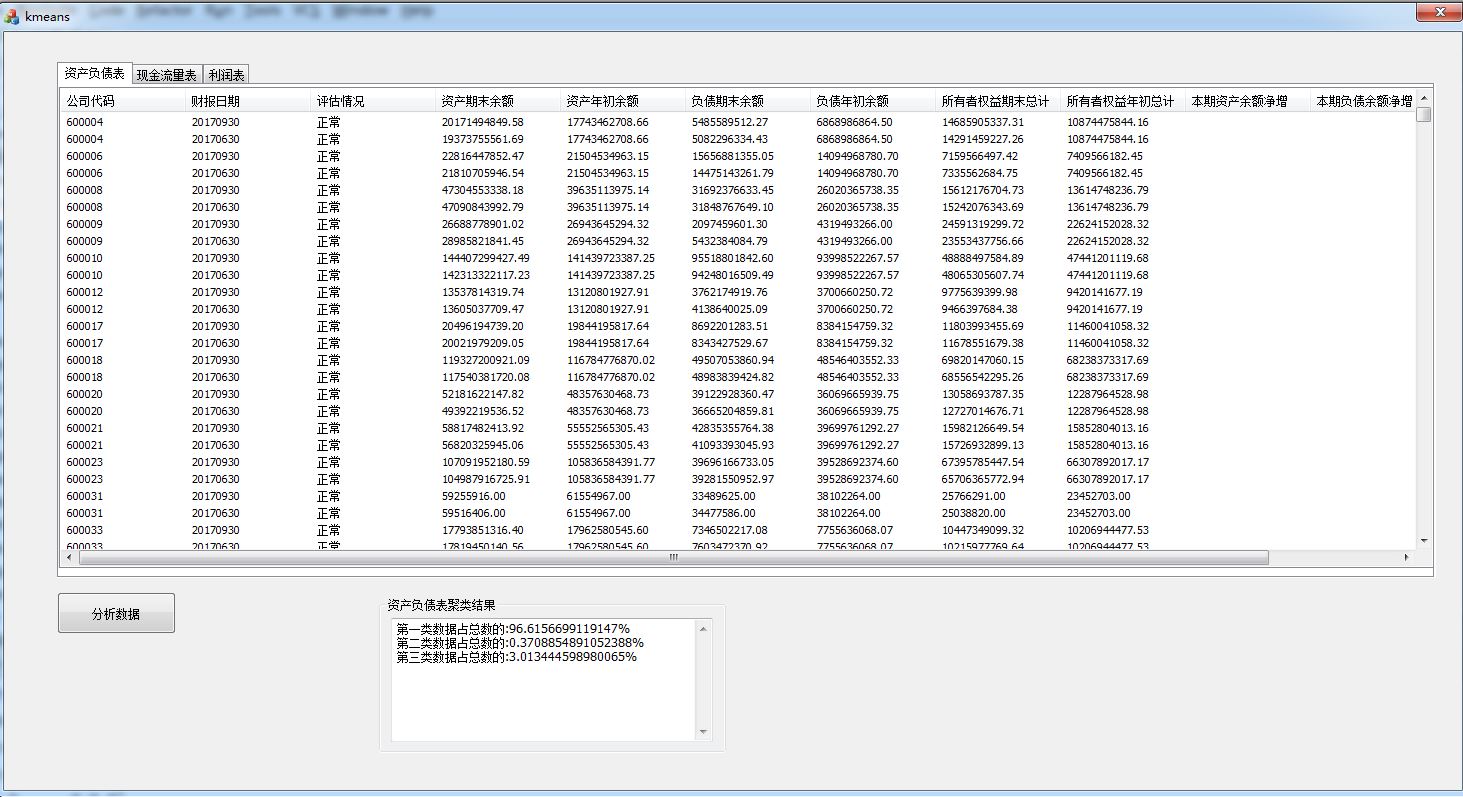
聚类出的结果大致一致，经过初步分析，原因如下：

数据预处理主要目的是在一些实际问题中，我们得到的样本数据都是多个维度的，即一个样本是用多个特征来表征的。

比如预测房价的问题中，影响房价y的因素有房子面积x1、卧室数量x2等，我们得到的样本数据就是(x1,x2)这样的一些样本点，这里x1、x2又被称为特征，很显然这些特征的量纲和数值的量级都是不一样的，由于kmeans聚类一般采用欧氏距离作为分类的依据，所以在预测房价时，如果直接使用原数据值，那么他们对房价的影响程度将是不一样的，而中心化和标准化处理，可以使得不同特征具有相同的尺度（scale）

下一步准备进行cost、vrc方差率和silh轮廓系数分析，求得最优的聚类类别数k，即自适应k

目前界面展示如下：



《资产负债表》

资产期末余额 20171494849.58 13.68%

资产年初余额 17743462708.66

负债期末余额 5485589512.27 -20.14%

负债年初余额 6868986864.50

所有者权益期末总计 14685905337.31 35.05%

所有者权益年初总计 10874475844.16

本期资产余额净增

本期负债余额净增

本期所有者权益净增

《利润表》

本期营业总收入 1704827663.49

上期营业总收入 1549942047.12

年初至报告期末总收入 4918489181.87

上年初至报告期末总收入 4534592334.78

本期营业总成本 1185926100.84

上期营业总成本 1073507763.83

年初至报告期末营业总成本 3379648080.37

上年初至报告期末营业总成本 3144787863.07

本期营业利润 556628923.05

上期营业利润 469016014.27

年初至报告期末营业利润 1607031101.91

上年初至报告期末营业利润 1394991936.14

本期利润总额 526087723.53

上期利润总额 482429314.96

年初至报告期末利润总额 1577293396.34

上年初至报告期末利润总额 1420889587.25

本期营业净利润 393598476.59

上期营业净利润 361616150.82

年初至报告期末营业净利润 1180049882.34

上年初至报告期末营业净利润 1059693410.16

年初至报告期末扣非净利润

上年初至报告期末扣非净利润

本期扣非净利润 1198716151.01

上期扣非净利润 1047461452.40