摘要

肥料是农业生产中十分重要的必需品，对肥料登记进行合理管理和严格把控，能够有效保障农作物的产量和质量，随着《肥料登记管理办法》的颁布，各地区依法在农业行政管理部门进行登记，肥料的生产销售也必须严格遵循《肥料登记管理办法》，各省、自治区、直辖市逐渐开始重视肥料管理。肥料使用量的多少不仅关系到农民辛苦耕作的收益，而且关乎人们的饮食安全，因此，应维护肥料产业的健康发展，加强肥料农业肥料的管理，确保农作物健康、科学生长并实现高产优质的目标，促进我国农业不断向前发展。本文利用数据分析技术对肥料登记数据进行预处理，从多个维度对肥料登记进行对比分析。

ABSTRACT

Fertilizer is a very important necessity in agricultural production. Reasonable management and strict control of fertilizer registration can effectively guarantee the yield and quality of crops. With the promulgation of the "Management Measures for Fertilizer Registration", various regions shall carry out the management in the agricultural administrative department in accordance with the law. Registration, fertilizer production and sales must also strictly follow the "Management Measures for Fertilizer Registration". Provinces, autonomous regions, and municipalities have gradually begun to attach importance to fertilizer management. The amount of fertilizer used is not only related to the income of farmers' hard work, but also related to people's diet safety. Therefore, the healthy development of the fertilizer industry should be maintained, and the management of fertilizers should be strengthened to ensure the healthy and scientific growth of crops and achieve the goal of high yield and high quality. , To promote the continuous development of my country's agriculture. This paper uses data analysis technology to preprocess fertilizer registration data, and compares and analyzes fertilizer registration from multiple dimensions.

For task 1, the data in Annex 1 needs to be preprocessed. First, use the unique() function and regular expressions in Pandas to standardize the naming operation of the "Generic Product Name" field in Annex 1, according to compound fertilizers, organic-inorganic compound fertilizers, organic fertilizers and bed soil acid regulators. The 4

categories are standardized in Annex 1. Secondly, the percentage of total inorganic nutrients in each fertilizer product in Annex 1 is calculated, and the result is kept to 3 decimal places.

For task 2, it is necessary to analyze the data of fertilizer products. First, filter out the compound fertilizer products and organic fertilizer products from Annex 2, divide the compound fertilizer products into 10 groups equidistantly according to the percentage of total inorganic nutrients, and draw the histogram of the registered quantity of the products. For organic fertilizer products Divide them into 10 groups equidistantly according to the percentage of inorganic nutrients and organic matter. The top 3 groups with the largest number of registrations and the corresponding product registration numbers are listed from the largest to the smallest. Secondly, use the DBSCAN algorithm in the clustering algorithm to divide these The products are divided into 4 categories; finally, the three-dimensional scatter plot and scatter plot matrix of fertilizer products are drawn, and the characteristics of each cluster are obtained according to the radar chart analysis of the clustering results.

For task 3 and task 4, a multi-dimensional comparative analysis of fertilizer products is required. First, task 3 compares the change trend of the number of registered products in each group of compound fertilizers in different years according to the year analysis, extracts effective products, and extracts effective products. Among the products, the Guangxi and Hubei products registered in the top 5 groups were selected, and the distribution differences between the two were analyzed. Secondly, extract fertilizer companies whose product registration number is greater than 10 from Appendix 3, and calculate the Jackard similarity coefficient matrix between companies based on the raw materials used by each company; finally, the design algorithm in this paper is extracted from the technical indicators in Appendix 4 Extract the percentages of nitrogen, phosphorus, potassium nutrients and organic matter, as well as the degree of chlorine in fertilizers, and extract the names and percentages of various raw materials from Appendix 4 Raw Materials and Percentages.

目录

一、问题分析与目标

1、对肥料登记数据进行预处理，规范附件 1 中“产品通用名称”字段，同时计算附件 1 中总无机养分百分比，根据养分的百分比对肥料产品进行细分。

2、筛选出复混肥料产品、有机肥料产品，将所有复混肥料按照总无机养分百分比的取值等距分为 10 组，并且按照总无机养分百分比和有机质百分比对有机肥料产品分别等距分为 10 组，同时绘制复混肥料产品登记数量直方图以及有机肥料产品的分布热力图，按照登记数量从大到小列出登记数量最大的前 3 个分组及相应的产品登记数量。

3、使用聚类算法将附件 2 筛选出的复混肥料产品分为 4 类，依据聚类标签绘制肥料产品的三维散点图和散点矩阵图，并且根据聚类结果的雷达图分析每个聚类的特征。

4、提取文件“result2_1.xlsx”中发证日期的年份，分析比较复混肥料中各组别不同年份产品登记数量的变化趋势，并对进行可视化处理。

5、从文件“result3_2.xlsx”中提取 2021 年 9 月 30 日仍有效的有机肥料产品，并且筛选出广西和湖北产品登记数量在前 5 的组别，分析这两个省份上述组别的差异。

6、从附件 3 中提取产品登记数量大于 10 的肥料企业，统计这些企业用到的原料集合，并以各企业用到的原料为特征，计算企业之间的杰卡德相似系数矩阵。

7、设计算法或处理流程，从附件 4“技术指标”这一字段中提取出氮、磷、钾养分和有机质的百分比，以及肥料含氯的程度。

8、设计算法或处理流程，从附件 4 原料与百分比中提取各种原料的名称及其百分比，并且分析数据给出处理思路及过程。

二、任务 1

1、任务 1.1 数据规范化处理

在本次数据分析过程中，需要规范附件 1 中的字段“产品通用名称”，将其按照复混肥料（掺混肥料也归为此类）、有机-无机复混肥料、有机肥料和床土调酸剂这 4 种类别对该字段进行规范化处理。

2、任务 1.2

此任务要求计算各肥料产品的氮、磷、钾养分百分比之和，称为总无机养分百分比。

三、任务 2

1、任务 2.1

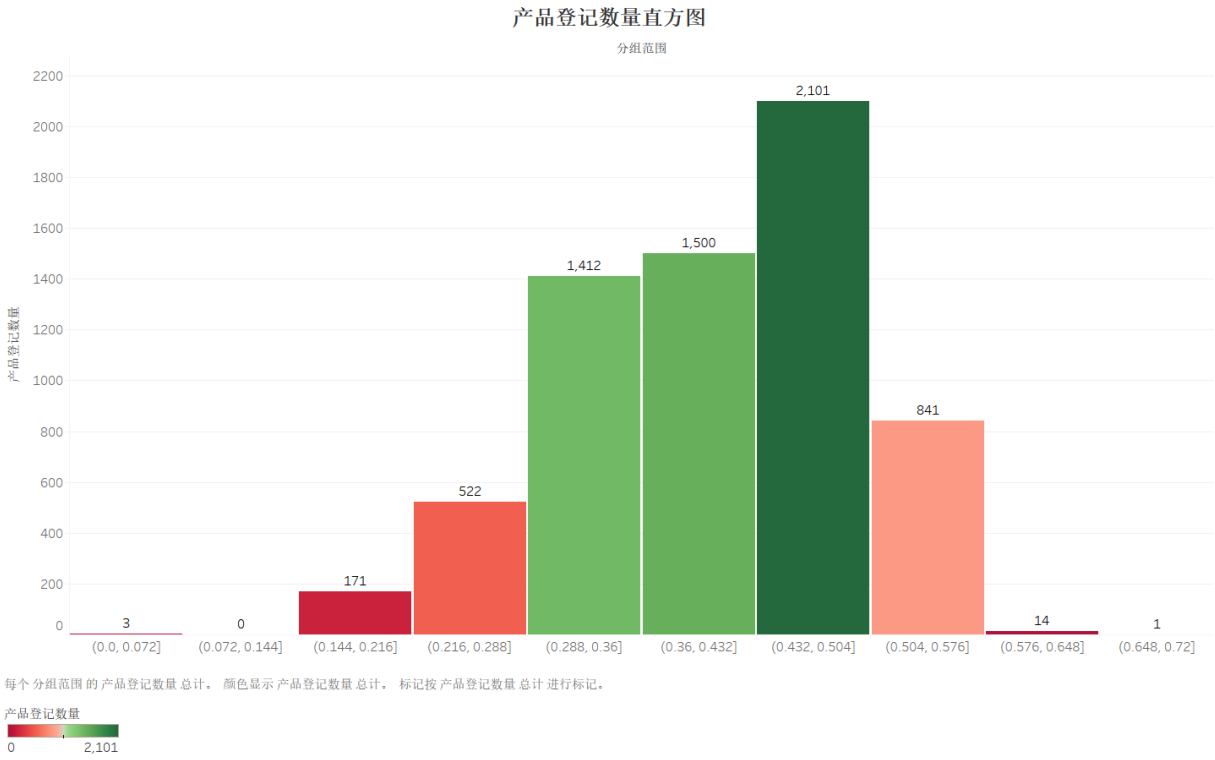

图 5 产品登记数量直方图

图 5 根据前文给出的复混肥料产品无机养分百分比的取值区间和产品登记数量绘制出的直方图，从图 5 亦可以直观的看出无机养分百分比的分布集中在(0.288，0.504]这个区间。

按照登记数量降序排列，本位依次列出登记数量最大的前 3 个分组及相应的产品登记数量，如表 2 所示。

表 2 复混肥料产品登记数量前 3 的分组

排名	一	二	三
分组标签	7	6	5
产品登记数量	2101	1500	1412

2. 任务 2.2

此任务要求分别按照总无机养分百分比和有机质百分比的范围对有机肥料产品等距分 10 组。

图 6 核心代码示意图 5

然后，本文根据分组情况绘制有机肥料产品的热力分布图，如图 7 所示，其中横轴表示总无机养分组，纵轴表示有机质分组。

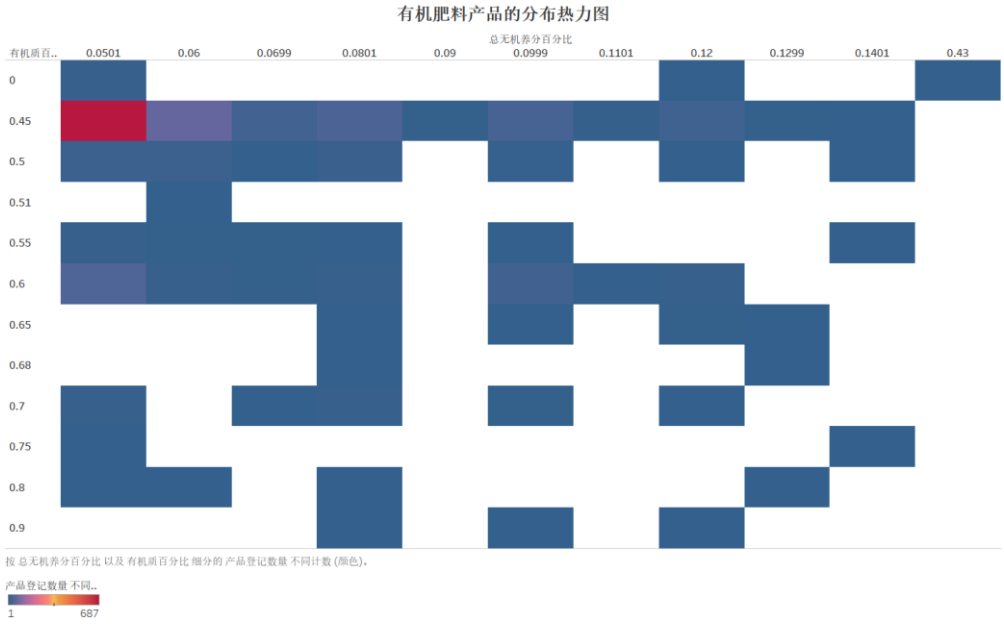

图 7 有机肥料产品的分布热力图

3、任务 2.3

3.1 DBSCAN 算法简介

DBSCAN 算法是一种基于密度的聚类算法，聚类时不需要预先指定簇的个数，并且最终的簇的个数不确定。

DBSCAN 算法将数据点分为 3 类：

（1） 核心点：在半径 Eps 内含有超过 MinPts 数目的点。

（2） 边界点：在半径 Eps 内点的数量小于 MinPts,但是落在核心点的领域内的点。

（3） 噪音点：既不是核心点也不是边界的点。

和传统的 K-Means 算法相比，DBSCAN 最大的不同就是不需要输入类别数 k，它最大的优势是可以发现任意形状的聚类簇，同时它可以在聚类的同时找出异常点。

由于本任务所给的数据集是稠密的，并且数据集不是凸的，那么使用 DBSCAN 算法来聚类具有很好的效果。

本文将 DBSCAN 算法中的 min_samples 的参数设置为 4，eps 的参数使用默认值 0.5，metric 使用默认值，依次将复混肥料分为 4 类。

3.2 本文方法及分析

本文使用聚类算法中的 DBSCAN 算法（具有噪声的基于密度的聚类方法）对筛选出来的复混肥料产品按照氮、磷、钾养分的百分比将产品分为 4 类，首先根据化学元素周期表，分别计算磷的相对原子质量在其对应的五氧化二磷相对分子质量的百分比，和钾的相对原子质量在氧化钾相对分子质量中所占的百分比，其中磷的相对原子质量为 30.97，钾的相对原子质量为 39，氧的相对原子质量为 16。

然后本文根据聚类结果为每个产品打上聚类标签，用 1~4 表示，最后，本文依照聚类标签绘制肥料产品的三维散点图和散点图矩阵，如图 9~10 所示。

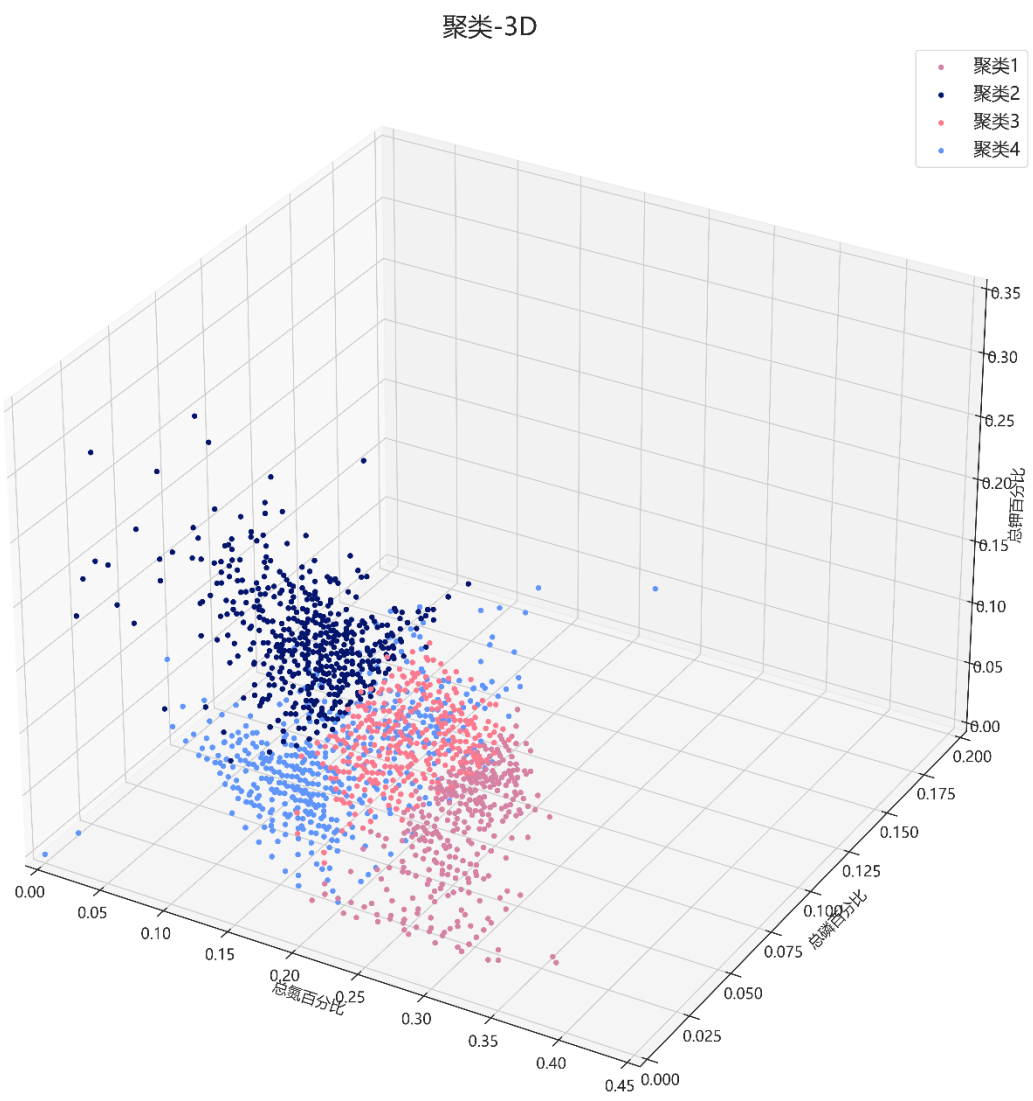

图 9 三维散点图

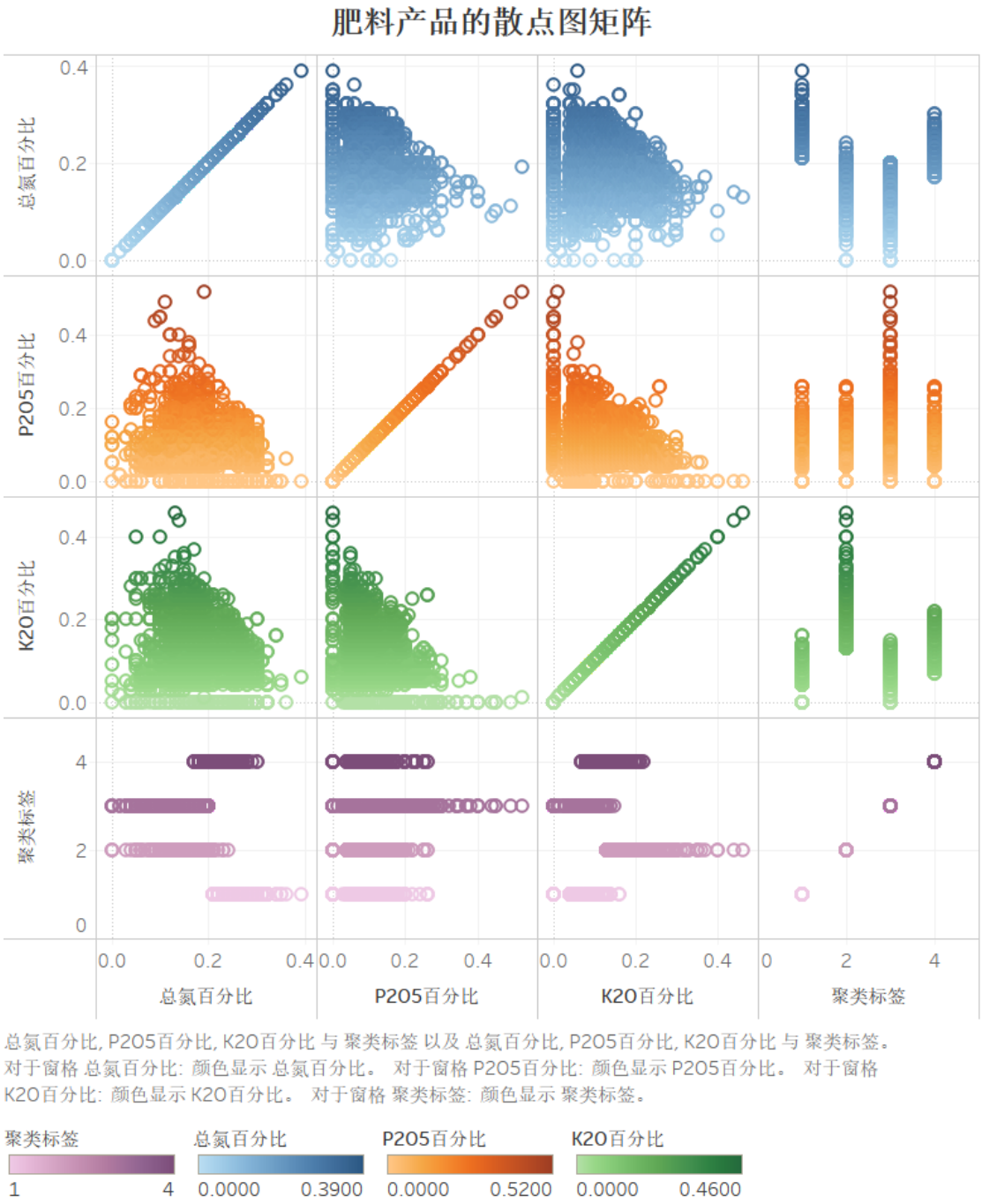
肥料产品的散点图矩阵
总氮百分比
P2O5百分比
K2O百分比
聚类标签
0.0
0.2
0.4
0
2
4
总氮百分比, P2O5百分比, K2O百分比 与 聚类标签 以及 总氮百分比, P2O5百分比, K2O百分比 与 聚类标签。
对于窗格 总氮百分比: 颜色显示 总氮百分比。 对于窗格 P2O5百分比: 颜色显示 P2O5百分比。 对于窗格 K2O百分比: 颜色显示 K2O百分比。 对于窗格 聚类标签: 颜色显示 聚类标签。
聚类标签
1
4
总氮百分比
0.0000
0.3900
P2O5百分比
0.0000
0.5200
K2O百分比
0.0000
0.4600

图 10 散点矩阵图

四、任务 3

1、任务 3.1

2、任务 3.2

在对数据进行处理之后，从文件“result2_2.xlsx”中提取 2021 年 9 月 30 日仍有效的肥料产品，并从有效产品中分别筛选出广西和湖北产品登记数量在前 5 的组别。

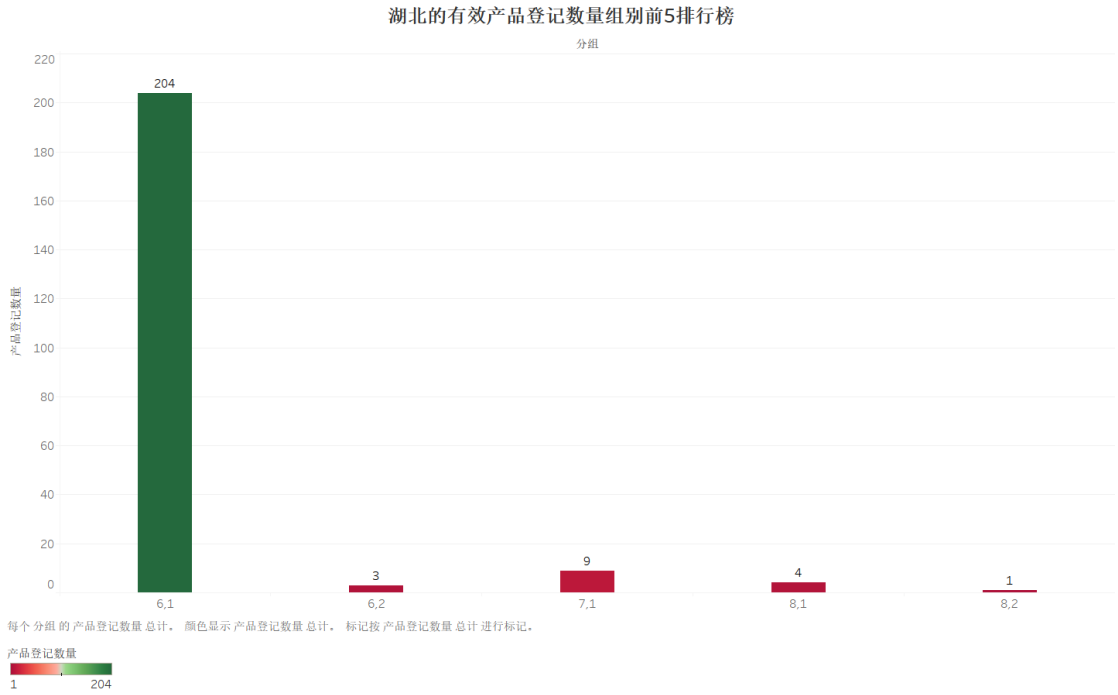

图 12 湖北有效产品登记数量组别前 5 排行榜

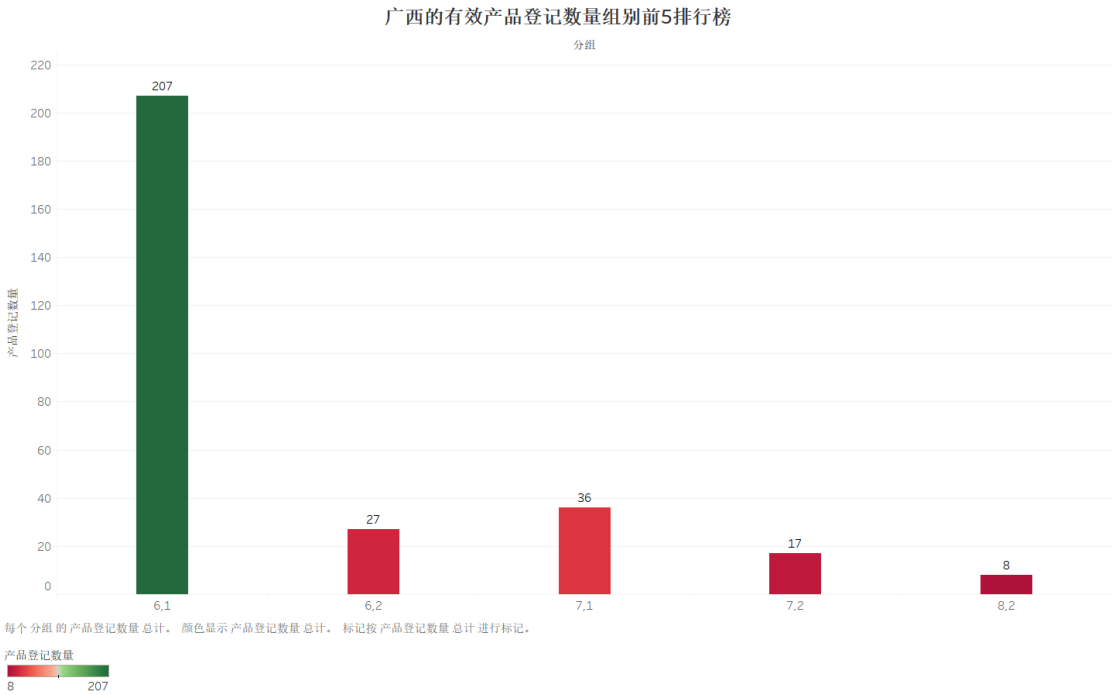

图 13 广西有效产品登记数量组别前 5 排行榜

3、任务 3.3

此任务需要从附件 3 中提取产品登记数量大于 10 的肥料企业，给出这 10 个企业所用到的原料集合（发酵剂除外）。

最后，我们根据杰卡德相似系数的定义制作出杰卡德相似系数矩阵。

	ID1	ID10	ID12	ID2	ID3	ID4	ID5	ID6	ID7	ID9
ID1	1.000000	0.200000	0.125000	0.166667	0.125000	0.222222	0.111111	0.285714	0.058824	0.125000
ID10	0.200000	1.000000	0.400000	0.214286	0.272727	0.285714	0.333333	0.555556	0.400000	0.555556
ID12	0.125000	0.400000	1.000000	0.307692	0.400000	0.285714	0.600000	0.400000	0.400000	0.272727
ID2	0.166667	0.214286	0.307692	1.000000	0.307692	0.312500	0.357143	0.214286	0.214286	0.062500
ID3	0.125000	0.272727	0.400000	0.307692	1.000000	0.500000	0.333333	0.272727	0.400000	0.166667
ID4	0.222222	0.285714	0.285714	0.312500	0.500000	1.000000	0.428571	0.285714	0.285714	0.285714
ID5	0.111111	0.333333	0.600000	0.357143	0.333333	0.428571	1.000000	0.333333	0.454545	0.333333
ID6	0.285714	0.555556	0.400000	0.214286	0.272727	0.285714	0.333333	1.000000	0.272727	0.272727
ID7	0.058824	0.400000	0.400000	0.214286	0.400000	0.285714	0.454545	0.272727	1.000000	0.272727
ID9	0.125000	0.555556	0.272727	0.062500	0.166667	0.285714	0.333333	0.272727	0.272727	1.000000

图 17 杰卡德相似系数矩阵结果图表

五、任务 4

1、任务 4.1

通过观察“技术指标”字段，我们可以看该字段字符串存在不规范的现象，按照要求，如果技术指标中只给出总养分百分比（“≥”按照“=”处理）而无明细数据，则氮、磷、钾养分的百分比按照总百分比的 1/3 来计算，结果保留 3 位小数（例如 1.0%，即 0.010）。复混肥料属于无机肥料，它的有机质百分比设定为 0。含氯情况分为“无氯”、“低氯”、“中氯”和“高氯”4 种。如果肥料 产品的技术指标中没有给出含氯情况，则视为“无氯”；如果注明“含氯”，则视为“低氯”。

	序号	产品通用名称	技术指标	原料与占比
0	1	复混肥料	总养分总养分≥35%(14-8-13)	尿素 (占15%),高岭土 (占15.5%),硫酸铵 (占28.16%),磷酸一铵 (占16...
1	2	复混肥料	总养分总养分≥30%(15-6-9)中氯	尿素 (占15%),高岭土 (占30.23%),氯化铵 (占28%),磷酸一铵 (占12.2...
2	3	有机肥料	总养分≥5%有机质≥45%	木薯渣（干基） (占84.9%),菌种 (占0.1%),黄豆渣 (占15%)
3	4	复混肥料	总养分总养分≥43%(10-18-15)含氯(低氯)	尿素 (占15%),高岭土 (占20%),粉状磷酸一铵 (占40%),氯化钾 (占25%)
4	5	有机肥料	总养分总养分≥5.0%有机质≥45%	畜禽粪便 (占50%),菌种 (占2%),桐 麸 (占30%),滤 泥 (占18%)
5	6	有机-无机复混肥料	总养分总养分≥20%(10-4-6)有机质≥20%含氯	尿素 (占10%),氯化铵 (占22%),肥料级磷酸氢钙 (占20%),氯化钾 (占10%)...
6	7	有机肥料	总养分总养分≥5%有机质≥45%	甘蔗制糖滤泥 (占85%),复合发酵菌 (占0.05%),米糠 (占10%),酒精发酵浓缩液...
7	8	有机肥料	总养分≥6%有机质≥50%	滤泥 (占78.4%),发酵菌剂 (占0.1%),酒精废液 (占21.5%)
8	9	有机肥料	总养分≥7%有机质≥45%	滤泥 (占78.6%),发酵菌剂 (占0.1%),酒精废液 (占21.3%)
9	10	有机肥料	总养分≥7%有机质≥55%	滤泥 (占77.9%),发酵菌剂 (占0.1%),酒精废液 (占22%)
10	11	有机肥料	总养分≥9%有机质≥45%	滤泥 (占76.6%),发酵菌剂 (占0.1%),酒精废液 (占23.3%)
11	12	有机肥料	总养分≥10%有机质≥45%	滤泥 (占75.6%),发酵菌剂 (占0.1%),酒精废液 (占24.3%)

图 19 数据处理后图

2、任务 4.2

六、参考文献

[1]周海平.肥料市场存在的问题与对策[J].农业科技与信息,2017(20):22-23.

[2]田雪,郑亚玲.肥料产品市场监管问题探析[J].石河子科技,2018(06):14-15.

[3]陈蕾.复合肥:传统肥料销售遇 阻科学施肥理念提升[J].中国农资,2019(26):7.

[4]陈金凤.绵阳市肥料生产经营使用情况的调查研究[J].四川农业与农机,2019(06):49-50.

[5]何庆虎.加强农业肥料登记管理的措施探究[J].南方农业,2019,13(23):174-175.

[6]刘宁莉.山西省有机肥料生产管理制度现状分析及探讨[J].中国农技推广,2019,35(07):15-18.
[7]魏萌,李阳.《肥料登记管理办法》新政落地　农资江湖波澜再起,肥料登记何去何从?[J].中国农资,2017(47):3-4.
[8]翟玉健,余承智,高星.一种基于DBSCAN聚类的雷达点迹处理方法[J].舰船电子对抗,2021,44(05):58-61.