

一、摘要

肥料是农业生产中一种重要的生产资料，其生产销售必须遵循《肥料登记管理办法》，依法在农业行政管理部门进行登记。各省、自治区、直辖市人民政府农业行政主管部门主要负责本行政区域内销售的肥料登记工作。肥料材料成分是否符合标准直接影响到农业食品的安全，因此，利用数据分析技术对肥料产品的成分进行研究具有重大意义。

针对任务一，对所给数据集进行缺失值、重复值等方面的处理。

针对任务二、先对各种肥料产品进行分析，根据每个产品所在的分组

针对任务三，首先根据肥料产品发证日期的记录，将肥料中各组别不同年份产品登记数量的变化趋势进行可视化处理。

针对任务四

先对所给数据集进行规范化处理，从肥料产品技术指标中提取出氮、磷、钾和有机质的百分比，肥料含氯的程度相应数据。

目录

一、摘要.....	1
二、问题分析.....	2
三、任务.....	2
任务 1.1.....	3
任务 1.2.....	3
任务 2 肥料产品的数据分析.....	3
任务 2.1.....	3
任务 2.2.....	4
任务 2.3.....	4
.....	5
任务 3 肥料产品的多维度对比分析.....	6
任务 3.1.....	6
任务 3.2.....	6
Jaccard 相似度的 python 实现;	6

任务 4 肥料产品的多维度对比分析.....	7
任务 4.1	7
任务 4.2	7

二．问题分析

1. 对肥料登记数据进行预处理。
2. 根据养分的百分比对肥料产品进行细分和可视化处理。
3. 从省份、日期、生产商、肥料构成等维度对肥料登记数据进行对比分析。
4. 对非结构化数据进行结构化处理。

三．任务

任务 1 数据的预处理

任务 1.1 附件 1 的产品通用名称存在不规范的情况。请按照复混肥料（掺混肥料归入这一类）、有机-无机复混肥料、有机肥料和床土调酸剂这 4 种类别对附件 1 进行规范化处理。

序号	企业名称	产品通用名称	产品形态	总氮百分比	P2O5百分比	K2O百分比	含氯情况	有机质百分比	正式登记证号	发证日期	有效期	
0	1	安徽中元化肥股份有限公司	复混肥料	颗粒	0.13	0.17	0.20	低氮	0.0	皖农肥(2016)准字4255号	2016-01-08	2021-01
1	2	安徽中元化肥股份有限公司	复混肥料	颗粒	0.13	0.17	0.20	中氮	0.0	皖农肥(2016)准字4256号	2016-01-08	2021-01
2	3	安徽中元化肥股份有限公司	复混肥料	颗粒	0.20	0.15	0.16	低氮	0.0	皖农肥(2016)准字4257号	2016-01-08	2021-01
3	4	安徽中元化肥股份有限公司	复混肥料	颗粒	0.26	0.13	0.12	低氮	0.0	皖农肥(2016)准字4258号	2016-01-08	2021-01
4	5	安徽中元化肥股份有限公司	复混肥料	颗粒	0.26	0.13	0.12	中氮	0.0	皖农肥(2016)准字4259号	2016-01-08	2021-01

任务 1.2 计算附件 1 中各肥料产品的氮、磷、钾养分百分比之和，称为总无机养分百分比。

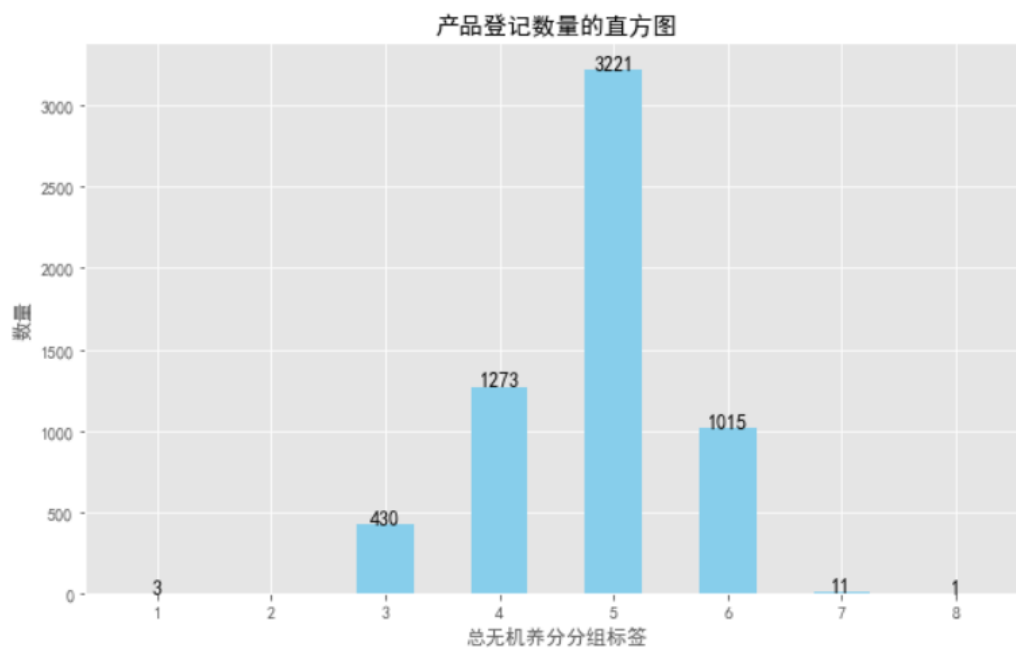
利用 python 中 math 函数通过如下公式计算得出总无机养分百分比

`['总无机养分百分比'] = ['总氮百分比'] + ['P2O5 百分比'] + ['K2O 百分比']`

任务 2 肥料产品的数据分析

任务 2.1 从附件 2 中筛选出复混肥料的产品，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。根据每个产品所在的分组，为其打上分组标

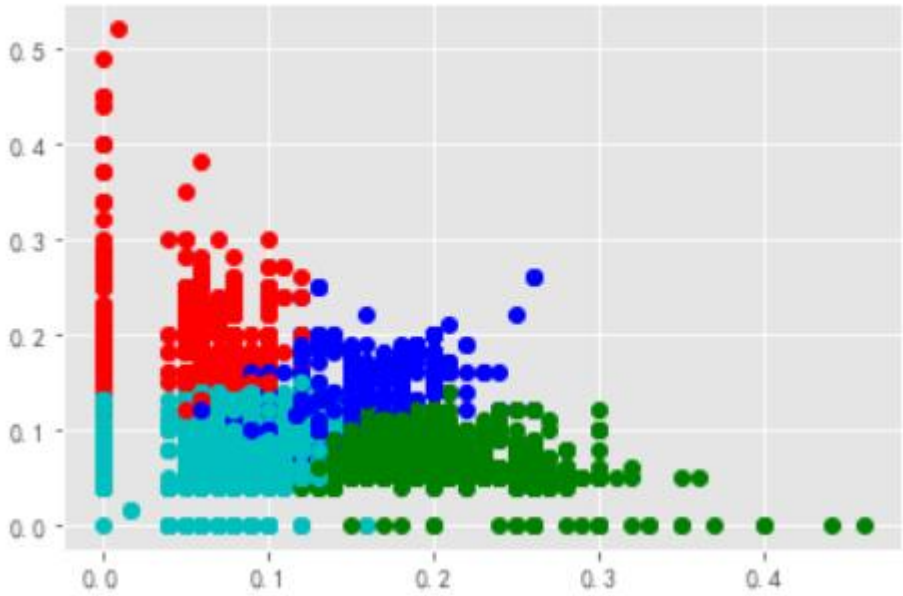
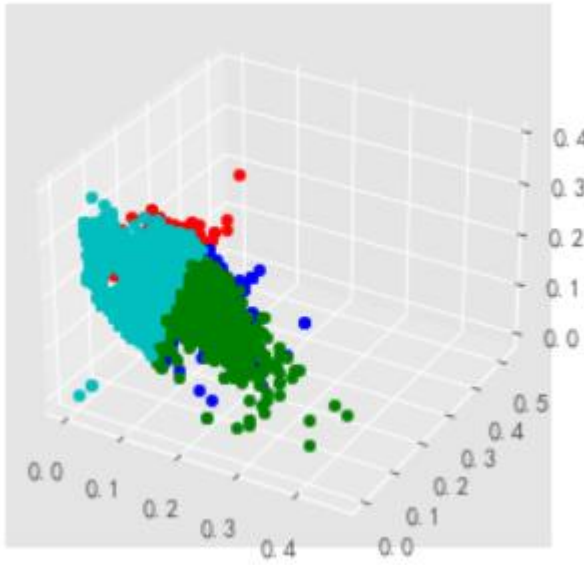
签（标签用 1~10 表示）。分析复混肥料产品的分布特点，在报告中绘制产品登记数量的直方图，并按登记数量从大到小列出登记数量最大的前 3 个分组及相应的产品登记数量



排名	一	二	三
分组标签	5	4	6
产品登记数量	3221	1273	1015

任务 2.2 从附件 2 中筛选出有机肥料的产品，将产品按照总无机养分百分比和有机质百分比分别等距分为 10 组，并为每个产品打上分组标签 (1,1), (1,2), ..., (10,10)，并根据分组情况绘制有机肥料产品的分布热力图，分析有机肥料产品的分布特点。

任务 2.3 从附件 2 中筛选出复混肥料的产品，按照氮、磷、钾养分的百分比，使用聚类算法将这些产品分为 4 类。根据聚类结果为每个产品打上聚类标签（标签用 1~4 表示），，根据聚类标签绘制肥料产品的三维散点图和散点图矩阵，分析每个聚类的特征。



序号	企业名称	产品通用名称	产品形态	总氮百分比	P2O5百分比	K2O百分比	含氮情况	有机质百分比	正式登记证号	发证日期	有效期	产品商品名称	适用作物	总无机养分百分比	聚类标签
0	4493.0 湖北庄园肥业股份有限公司	复混肥料	粒状	0.00	0.16	0.09	中氮	0.0	鄂农肥(2017)准字2960号	2017-01	2022-01	NaN	NaN	0.25	1.0
1	4494.0 湖北庄园肥业股份有限公司	复混肥料	粒状	0.00	0.10	0.18	无氮	0.0	鄂农肥(2017)准字2961号	2017-01	2022-01	NaN	NaN	0.28	1.0
2	254.0 广西林茂肥料有限责任公司	复混肥料	粒状	0.00	0.05	0.20	无氮	0.0	桂农肥(2015)准字2345号	2015-12-22	2020-12	爱高健	NaN	0.25	1.0
3	5.0 湖北澳特尔化工有限公司	复混肥料	粒状	0.15	0.15	0.15	无氮	0.0	鄂农肥(2009)准字00079号	2014-10-25	2019-10	NaN	NaN	0.45	1.0
4	255.0 广西林茂肥料有限责任公司	复混肥料	粒状	0.00	0.05	0.20	中氮	0.0	桂农肥(2015)准字2346号	2015-12-22	2020-12	爱高健	NaN	0.25	1.0
...

任务 3 肥料产品的多维度对比分析

任务 3.1 从文件“result2_1.xlsx”中提取发证日期中的年份，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。

任务 3.2 从文件“result2_2.xlsx”中提取 2021 年 9 月 30 日仍有效的有机肥料产品。从有效产品中分别筛选出广西和湖北（根据正式登记证号区分）产品登记数量在前 5 的组别，分析两个省份上述组别的分布差异。

序号	企业名称	产品通用名称	产品形态	总氮百分比	P2O5百分比	K2O百分比	含氮情况	有机质百分比	正式登记证号	发证日期	有效期	产品商品名称	适用作物	总无机养分百分比	分组标签	年	月
7	恒发新天地生物科技(大悟)有限公司	有机肥料	粉状	0.0200	0.0200	0.0200	无氮	0.45	鄂农肥(2011)准字0873号	2016-11-18	2021-11-01	NaN	NaN	0.0600	[1, 5]	2021	11
110	湖北三宁化工股份有限公司	有机肥料	粒状	0.0167	0.0167	0.0167	无氮	0.45	鄂农肥(2016)准字2047号	2016-10-10	2021-10-01	NaN	NaN	0.0501	[1, 5]	2021	10
111	荆州市丰禾源生物科技有限公司	有机肥料	粒状	0.0167	0.0167	0.0167	无氮	0.45	鄂农肥(2016)准字2048号	2016-10-10	2021-10-01	NaN	NaN	0.0501	[1, 5]	2021	10
112	湖北吉顺磷化有限公司	有机肥料	粉状	0.0167	0.0167	0.0167	无氮	0.45	鄂农肥(2016)准字2083号	2016-10-20	2021-10-01	NaN	NaN	0.0501	[1, 5]	2021	10
113	湖北吉顺磷化有限公司	有机肥料	粒状	0.0167	0.0167	0.0167	无氮	0.45	鄂农肥(2016)准字2084号	2016-10-20	2021-10-01	NaN	NaN	0.0501	[1, 5]	2021	10

任务 3.3 从附件 3 中提取产品登记数量大于 10 的肥料企业，给出这些企业所用到的原料集合（发酵菌剂除外）。以各企业用到的原料作为特征，计算企业之间的杰卡德相似系数矩阵。

Jaccard 相似度的 python 实现;

定义

给定两个集合A,B，Jaccard 系数定义为A与B交集的大小与A与B并集的大小的比值，定义如下：

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

当集合A，B都为空时，J(A,B)定义为1。

杰卡德距离(Jaccard Distance) 是用来衡量两个集合差异性的一种指标，它是杰卡德相似系数的补集，被定义为 1 减去 Jaccard 相似系数。而杰卡德相似系数(Jaccard similarity coefficient)，也称杰卡德指数(Jaccard Index)，是用来衡量两个集合相似度的一种指标。

(1) 杰卡德相似系数

两个集合 A 和 B 的交集元素在 A，B 的并集中所占的比例，称为两个集合的杰卡德相似系数，用符号 $J(A,B)$ 表示。

杰卡德相似系数是衡量两个集合的相似度一种指标。

(2) 杰卡德距离

与杰卡德相似系数相反的概念是杰卡德距离(Jaccard distance)。杰卡德距离可用如下公式表示：

杰卡德距离用两个集合中不同元素占有所有元素的比例来衡量两个集合的区分度。

任务 4 肥料产品的多维度对比分析

任务 4.1 设计算法或处理流程，从附件 4 技术指标中提取出氮、磷、钾养分和有机质的百分比，以及肥料含氯的程度。

序号		产品通用名称	总氮百分比	P2O5百分比	K2O百分比	有机质百分比	含氯情况
0	1	复混肥料	0.140	0.080	0.130	0.00	无氯
1	2	复混肥料	0.150	0.060	0.090	0.00	中氯
2	3	有机肥料	0.333	0.333	0.333	0.45	无氯
3	4	复混肥料	0.100	0.180	0.150	0.00	低氯
4	5	有机肥料	0.333	0.333	0.333	0.45	无氯
...
195	196	有机-无机复混肥料	0.060	0.080	0.060	0.00	含氯
196	197	有机-无机复混肥料	0.090	0.070	0.130	0.20	含氯
197	198	有机肥料	0.333	0.333	0.333	0.20	无氯
198	199	有机肥料	0.333	0.333	0.333	0.45	无氯
199	200	有机-无机复混肥料	0.333	0.333	0.333	0.15	含氯

任务 4.2 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。

参考文献

- [1]邓少江.规范强基础 产业作支撑 联合促增收——记秭归县康源中药材种植专业合作社[J].中国农民合作社,2021(11):25-26.
- [2]刘华彬.发展品质农业 搭建共享平台——成都市温江区富农蔬菜专业合作社发展纪实[J].中国农民合作社,2021(11):48-50.
- [3]赖金平,姚锋先.动植物有机肥的应用研究现状及展望[J].现代园艺,2021,44(21):23-25.
- [4]陈建萍,杜宏山,周坤昌,余维胜.云南红豆杉药用原料林培育技术[J].特种经济动植物,2021,24(11):94-96.
- [5]赵黎昀,毛可馨. 化工企业“带资”驶入新赛道 磷酸铁锂大宗化进程加速[N]. 证券时报,2021-11-09(A04).
- [6]熊建军,黄采文,曾勇,纪雄辉,陈山,官迪,吴家梅.有机肥对水稻镉生物有效性的影响研究进展[J].湖南生态科学学报,2021,8(04):92-96.
- [7]宋晓,陈莉,李建芬,李夕军.增施微生物菌剂对设施土壤理化性质及微生物的影响[J].安徽农业科学,2021,49(21):169-171.
- [8]杨波,梁劲,黎冬梅.玉林大蒜富硒安全生产技术[J].长江蔬菜,2021(21):52-54.
- [9]刘为国. 2022 年化肥行情或将出现大幅下滑[N]. 河北农民报,2021-10-28(005).
- [10]石旭旭. 复合肥市场观望气氛仍浓厚[N]. 农资导报,2021-10-19(A03).
- [11]王艳春,周宏.有机肥在林业可持续发展中的重要性及应用探析[J].南方农业,2021,15(24):100-101.
- [12]石旭旭. 原料价格下跌复合肥成本松动[N]. 农资导报,2021-08-24(A03).
- [13]李若晨,吕健飞.武义县推进有机肥替代化肥工作的主要做法与建议[J].中国农技推广,2021,37(07):15-16+13.
- [14]汤菲菲. 尿素负荷微降库存续增[N]. 农资导报,2021-07-06(A03).
- [15]雷雅茹. 出口形势好 肥市价格高[N]. 河北农民报,2021-07-01(004).