

肥料登记数据分析

目录

一、 行业背景与分析目标.....	2
1、 农业背景.....	2
2、 分析目标.....	2
二、 数据说明.....	2
三、 任务--数据的分析及建模	4
3.1、 任务 1—数据的预处理	4
3.1.1、 任务 1.1 —附件 1 的产品通用名称的规范化处理	4
3.1.2、 任务 1.2 —计算各肥料产品的总无机养分百分比	4
3.2、 任务 2—肥料产品的数据分析	4
3.2.1、 任务 2.1—对复混肥料按照总无机养分百分比分组、绘制直方图，并列出登 记数量前 3 组.....	4
3.2.2、 任务 2.2 —对有机肥料按照总无机养分百分比和有机质百分比进行分组，并 记录登记数量前 3 的分组.....	6
3.2.3、 任务 2.3 —对复混肥料按照氮、磷、钾养分的百分比进行聚类，并绘制三维 散点图.....	6
3.3、 任务 3 —肥料产品的多维度对比分析	6
3.3.1、 任务 3.1 对复混肥料中的各组按年份分析产品登记数量的变化趋势并将结果 可视化.....	6
3.3.2、 任务 3.2—对 2021 年 9 月 30 日仍有效的有机肥料产品，分析广西和湖 北在上述组别的分布差异.....	6
3.3.3、 任务 3.3—从附件 3 中提取登记数量大于 10 的肥料企业，并给出这些企业所 用到的原料集合.....	6
3.4、 任务 4—肥料产品的多维度对比分析	6
3.4.1、 任务 4.1—重复建 4 中提取氮、磷、钾养分和有机质的百分比，及肥料含氯 程度.....	6

一、行业背景与分析目标

1、农业背景

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。肥料是农业的动力，严格把握好肥料登记及质量关，加强肥料生产、销售市场管理，是保证农业生产安全、农民增收、农村稳定的重要措施之一。我国现有十三多亿人口，粮食是人们生存的基础。

俗话说：民以食为天，食以安为先，所以粮食的生产和安全是人们最关注的话题，而粮食的生产和安全主要取决于土壤肥料。现如今的国家经济正处在快速发展阶段，但与此同时出现了人口数量增多，农业资源不断被消耗和土地质量恶化等问题，要想这些问题得到有效解决，就应做到以下几点：(1)重复利用土地资源，增强对资源再利用的认识。(2)提高土地质量，高度保护生态环境和土壤资源。(3)在保证农民收入的同时也要注重生态资源效益，促进农业可持续发展，让农村经济和生态环境形成良好的循环。肥料的规范使用，与我们息息相关。

2、分析目标

1、对肥料登记数据进行预处理。使数据变得规范化与大部分数据可以相一致，让数据显而易见，让后期的任务处理减轻阻碍。

2、根据养分的百分比对肥料产品进行细分。进行多维度划分，阶段性的分析与展现。生产的肥料中无机盐含量与买家所真正需要的化肥无机盐含量，可以进行选择对比。正如现实生活中，肥料用的越多而产量越低，比如：磷肥过多时叶上又会出现小焦斑，系磷酸钙沉淀所致；磷过多还会阻碍植物对硅的吸收。植物易招致水稻感病。所以，了解化肥无机盐含量是很重要的。许多农民百思不得其解。这就是对化肥其中无机盐含量不了解，不能针对自己土地选择合适的肥料。本方案分析当中会给出具体的数据表格，进行分析解答。

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。使偏离实际的生产的化肥浮出水面。考虑所生产出的化肥，是否能在市场上发行。

4、对非结构化数据进行结构化处理。非结构化数据是数据结构不规则或不完整，没有预定义的数据模型，不方便用数据库二维逻辑表来表现的数据。所以我们将其结构化处理，把数据进行可视化展现。

二、数据说明

由题意可知：附件 1-附件 4 的数据收集自农业部门官方网站，部分数据细节与实际有差别，仅供比赛使用。附件 1 为安徽肥料登记数据，附件 2 为广西、湖北肥料登记数据。这两个附件中表的主要字段有企业名称、产品通用名称、

正式登记证编号、发证日期、有效时间、产品形态、营养成分百分比、含氯情况等。其中产品通用名称实际上是 肥料产品的类型，需要在省级农业行政主管部门登记的肥料有复混肥料（包括掺 混肥料）、有机-无机复混肥料、有机肥料和床土调酸剂这 4 类。肥料的营养成分百分比指标，通常标记出属于无机成分的氮、磷、钾的含量，以及有机质的含量。我国规定，氮肥成分以总氮的质量来计算含量，磷肥成分按磷元素的量折算成五氧化二磷（P2O5）的质量来计算含量，钾肥成分按钾元素的量折算成氧化钾（K2O）的质量来计算含量。注意，肥料正式登记证有效期为 5 年，可以续期，会出现有效期距发证日期大于 5 年的情况。附件 3 给出了某省登记肥料的产品配方，相比附件 1 和附件 2 增加了关于肥料原料的信息。附件 4 给出了某省肥料登记数据中营养成分及原料构成的原始数据。字段技术指标以字符串的形式给出了肥料的营养成分的百分比。例如某复混肥料的技术指标字段取值为“N+P2O5+K2O≥20%（7-10-3） 有机质≥20% 含氯”，表示肥料中氮磷钾三大元素的总养分含量不小于 20%；“（7-10-3）”指的是氮磷钾的配比，氮含量为 7%，磷肥成分（折算为 P2O5）含量为 10%，钾肥成分（折算为 K2O）含量为 3%；“有机质≥20%”表示肥料中有机质的含量不小于 20%；“含氯”表示肥料中含有氯元素。有机肥料由于不含无机养分或含量较少，有些产品只在技术指标中标明“总养分≥…%”，没有给出氮、磷、钾 3 大元素的具体含量。字段原料与百分比以字符串的形式给出了糖蜜酒精废液（占 25%），发酵菌种（占 1%），木糠（占 25%），滤泥（占 49%），肥料的原料构成及质量百分比，例如某有机肥料的原料与百分比字段取值为“表明了该有机肥料由蜜糖酒精废液、发酵菌种、木糠和滤泥四种原料构成，质量百分比分别是 25%、1%、25%及 49%。

三、任务--数据的分析及建模

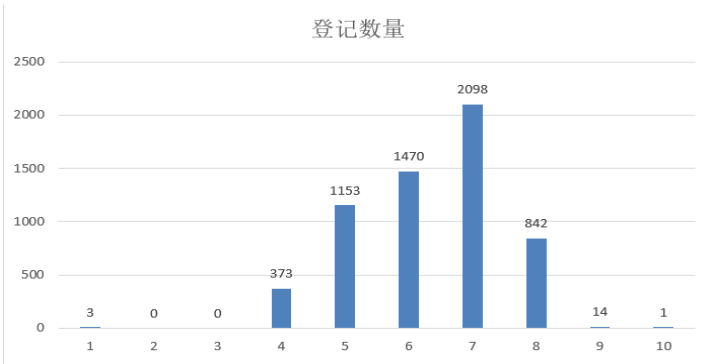
3.1、任务 1—数据的预处理

3.1.1、任务 1.1 一附件 1 的产品通用名称的规范化处理

3.1.2、任务 1.2 一计算各肥料产品的总无机养分百分比

3.2、任务 2—肥料产品的数据分析

3.2.1、任务 2.1—对复混肥料按照总无机养分百分比分组、绘制直方图，并列出生成数量前 3 组



针对于生产出来的复混肥料产品登记情况，在 4-8 组比较集中（无机盐百分比含量而言在 21.6%-57.6%相对集中）出现峰值 2098，其中植物最需要的三种无机盐含量比较均衡，现实生活中比较适用于大多数土地。

针对于前三的数据量如下图所示：

排名	一	二	三
分组标签	7	6	5
产品登记数量	2098	1470	1153

组数	登记数量
7	2098
6	1470
5	1153

其它组，基于总体数据来说占比不足 1%。对于直方图来说，其含量大多在于第 9 组，其中含 P（磷）的非超高，对于缺少 P 的土地可以考虑这家企业的化肥。

植物的生活需要多种无机盐。其中以含氮的、含磷的、含钾的无机盐需要量最大，不同的植物需要的无机盐的含量和种类都不同，同一植物的不同时期对无机盐的需要量也不同，所以对植物要做到合理施肥。但，像第 1 组，无机盐含量不足 5%。因此也要谨防生活中一些不良商家的违心买卖！

3.2.2、任务 2.2 一对有机肥料按照总无机养分百分比和有机质百分比进行分组，并记录登记数量前 3 的分组

3.2.3、任务 2.3 一对复混肥料按照氮、磷、钾养分的百分比进行聚类，并绘制三维散点图

3.3、任务 3 一肥料产品的多维度对比分析

3.3.1、任务 3.1 对复混肥料中的各组按年份分析产品登记数量的变化趋势并将结果可视化

3.3.2、任务 3.2一对 2021 年 9 月 30 日仍有效的有机肥料产品，分析广西和湖北在上述组别的分布差异

3.3.3、任务 3.3—从附件 3 中提取登记数量大于 10 的肥料企业，并给出这些企业所用到的原料集合

3.4、任务 4—肥料产品的多维度对比分析

3.4.1、任务 4.1—重复建 4 中提取氮、磷、钾养分和有机质的百分比，及肥料含氯程度

1	序号	原料与占比	
2	1	尿素	15%
3		高岭土	15.50%
4		硫酸铵	28.16%
5		磷酸一铵	16.34%
6		硫酸钾	25%
7	2	尿素	15%
8		高岭土	30.23%
9		氯化铵	28%
10		磷酸一铵	12.25%
11		氯化钾	14.52%
12	3	木薯渣（干基）	84.90%
13		菌种	0.10%
14		黄豆渣	15%
15	4	尿素	15%
16		高岭土	20%
17		粉状磷酸一铵	40%
18		氯化钾	25%