# 累产每日排名数据处理流程

## 1.原始数据描述

原始数据记录了各气井在一段时间内的生产数据，即单口井对多个生产日期。但部分气井投产后的一段时间内累计产气量为0，如表1所示。此类数据无法对气井的生产情况提供有效信息，属于冗余数据记录，故需要予以剔除。

**表1 无效数据示例**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 井号 | 日期 | 生产时间 | 历年累计产气量 | 甜点分区 | 生产天数 |
| 威202H10-5 | 2017/9/1 | 24 | 0 | 柠檬 | 1 |
| 威202H10-5 | 2017/9/2 | 24 | 0 | 柠檬 | 2 |
| 威202H10-5 | 2017/9/3 | 24 | 0 | 柠檬 | 3 |
| 威202H10-5 | 2017/9/4 | 24 | 0 | 柠檬 | 4 |

此外，为分析各气井的产气能力，需要对242口气井的产量进行横向比较。故以生产天数为单位，统计242口气井的每日产量并进行排名，以便评估各气井的相对生产能力。

## 2数据清洗操作流程

1. **剔除无效数据**

首先，对原始数据中历年累计产气量为0的数据条目进行剔除，保留历年累计产气量大于0的有效数据，并将清洗后的数据保存。

1. **生产天数重编号**

依次提取242口气井的历史生产数据，对每一口气井的历史生产数据，按照“日期“变量进行升序排列，以最早的“日期“记录作为首次生产日，重新编辑气井各个日期对应的生产天数。以表1中威202H10-5气井为例，重新编辑后的数据如表2所示。

**表2 生产天数重排数据示例**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 井号 | 日期 | 生产时间 | 历年累计产气量 | 甜点分区 | 生产天数 |
| 威202H10-5 | 2017/9/14 | 24 | 0.5732 | 柠檬 | 1 |
| 威202H10-5 | 2017/9/15 | 24 | 1.253 | 柠檬 | 2 |
| 威202H10-5 | 2017/9/16 | 24 | 2.1471 | 柠檬 | 3 |
| 威202H10-5 | 2017/9/17 | 24 | 3.3581 | 柠檬 | 4 |

1. **累计产气量排名**

为评估各气井在相同生产天数内的产气能力，对242口气井在每日的累计产气量进行排名，并增加一列新变量“排名”，变量取值范围为1-242。以威202H10-5气井为例，增加排名变量后的数据如表3所示。

**表3 每日产气量排名数据示例**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 井号 | 日期 | 生产时间 | 历年累计产气量 | 甜点分区 | 生产天数 | 排名 |
| 威202H10-5 | 2017/9/14 | 24 | 0.5732 | 柠檬 | 1 | 218 |
| 威202H10-5 | 2017/9/15 | 24 | 1.253 | 柠檬 | 2 | 219 |
| 威202H10-5 | 2017/9/16 | 24 | 2.1471 | 柠檬 | 3 | 217 |
| 威202H10-5 | 2017/9/17 | 24 | 3.3581 | 柠檬 | 4 | 218 |

1. **以不同生产周期划分数据**

对前3年历史生产数据，以90天（季度）为跨度，记录242口气井在不同时间范围内的生产数据，并增加排名变量。由于大部分气井的生产天数小于4年，从第4年开始，以年为单位进行划分。最终得到第1年各气井生产情况（1-4季度）、第2年各气井生产情况（1-4季度）、第3年各气井生产情况（1-4季度）、第4年各气井生产情况（全年）、第5年各气井生产情况（全年）、第6年各气井生产情况（全年）。

**表4 清洗数据集汇总**

|  |  |
| --- | --- |
| 数据集 | 生产天数范围 |
| 第1年1季度各气井生产情况 | 1-90 |
| 第1年2季度各气井生产情况 | 1-180 |
| 第1年3季度各气井生产情况 | 1-270 |
| 第1年4季度各气井生产情况 | 1-365 |
| 第2年1季度各气井生产情况 | 1-455 |
| 第2年2季度各气井生产情况 | 1-545 |
| 第2年3季度各气井生产情况 | 1-635 |
| 第2年4季度各气井生产情况 | 1-730 |
| 第3年1季度各气井生产情况 | 1-820 |
| 第3年2季度各气井生产情况 | 1-910 |
| 第3年3季度各气井生产情况 | 1-1000 |
| 第3年4季度各气井生产情况 | 1-1095 |
| 第4年各气井生产情况 | 1-1460 |
| 第5年各气井生产情况 | 1-1825 |
| 第6年各气井生产情况 | 1-2190 |