

学生社团注册数据分析

一 分析字段

学号：长度为 41 的字符串，一种字符串对应一个学生，没用。

所在院系：47 种院系名称，还有一个“-”需要清理掉。

学期：第一学期、第二学期

学年：2021-2022、2022-2023、2023-2024

社团名称：120 种

状态：signin：学生提交了社团申请，处于未通过的状态。（4 条）太少了，忽略不计。

signOut：学生退出社团。（1834 条）

effective：学生被确认为正式的社团成员。（22506 条）大部分的图条件都是当状态为 effective。

社团指导单位：39 种院系名称

社团类型：7 种

社团星级：6 种

社团校区：中山北路校区、闵行校区

二 数据清洗和预处理

1 删除学号字段：经检验没有相同的学号有着大量数据，排除反复操作的情况。

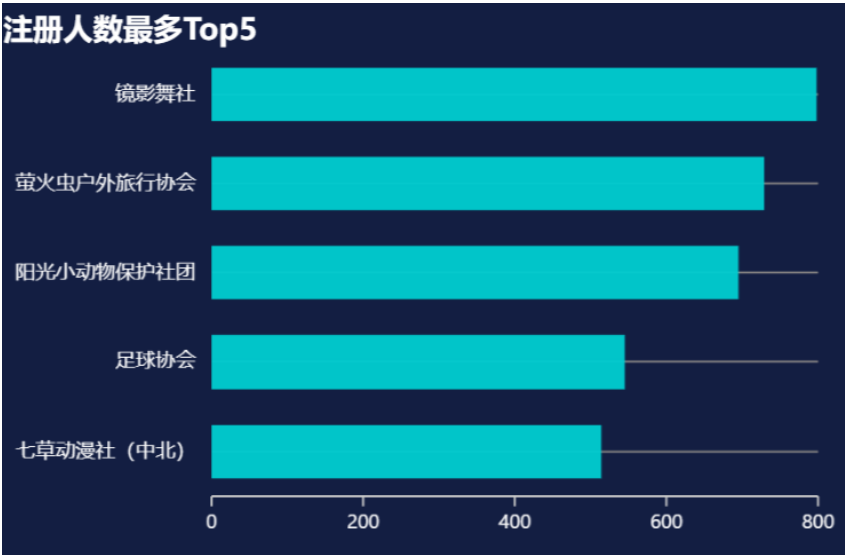
2 清理所在院系字段中的“-”

3 清理状态字段中的"signin"

见代码 清洗.py

三 数据可视化

1

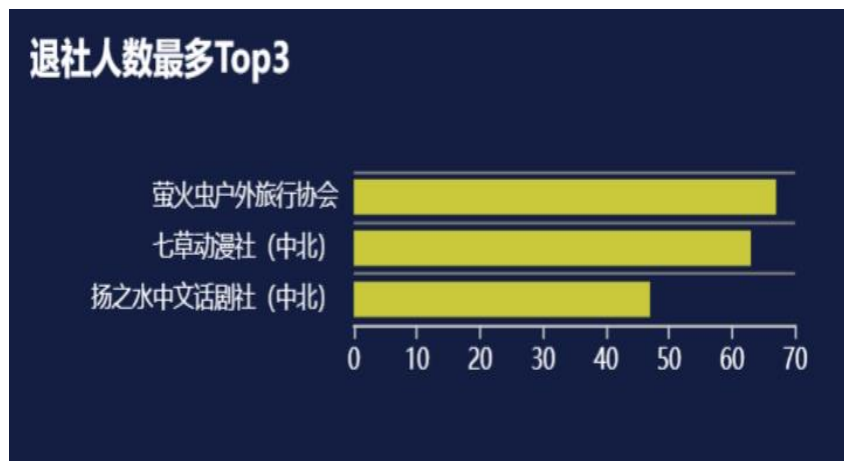


横向柱状图

状态为 effective

表明最受学生欢迎的五个社团 有具体数字

镜影舞社、萤火虫户外旅行协会、阳光小动物保护社团、足球协会、七草动漫社（中北）

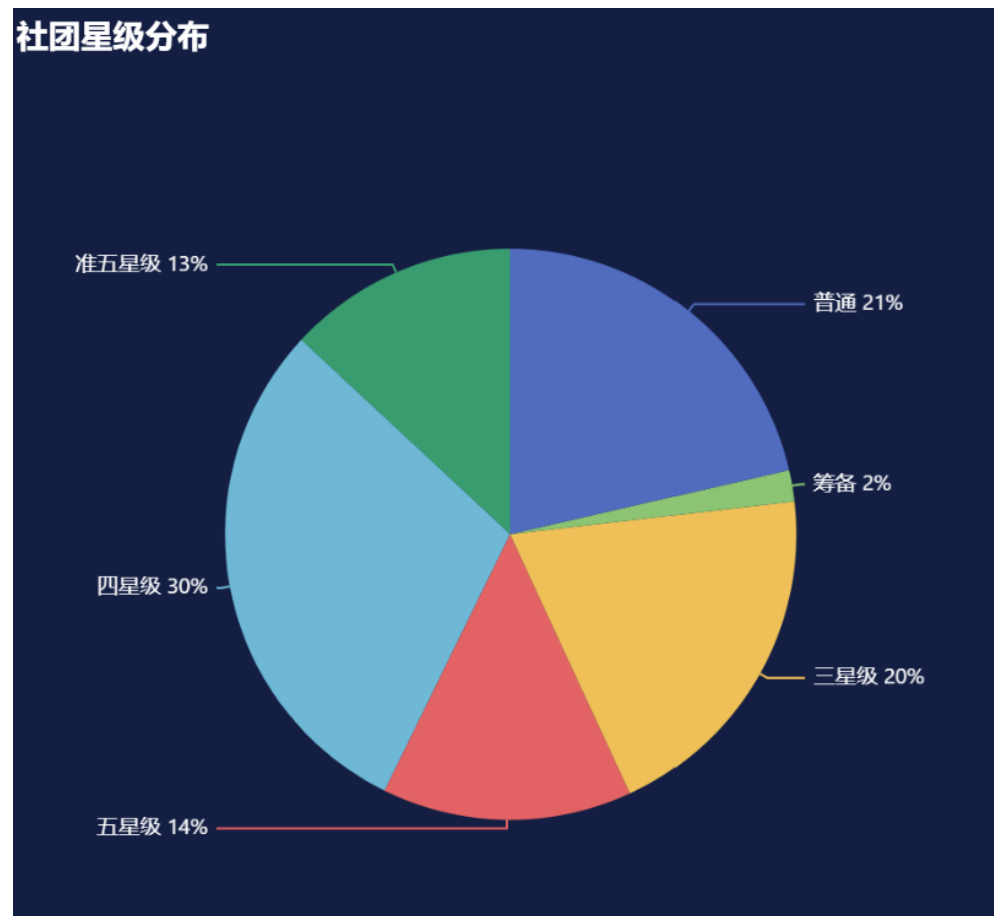


横向柱状图

状态为 signOut

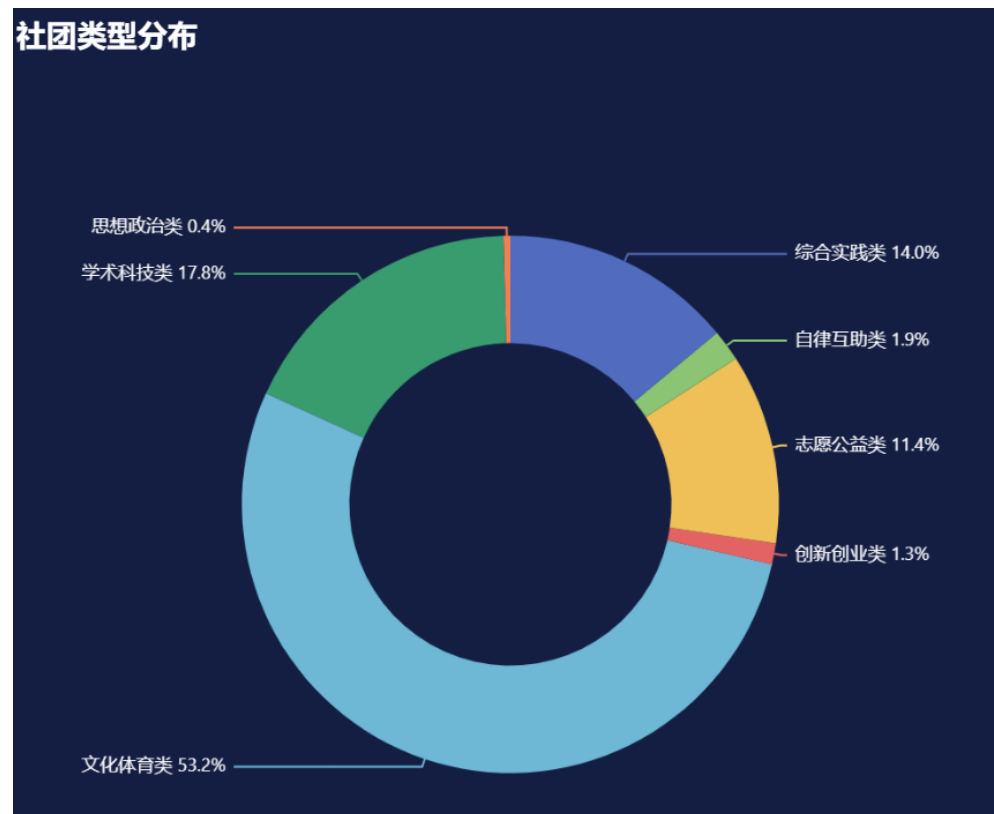
表明学生最想退出的三个社团 有具体数字

萤火虫户外旅行协会、七草动漫社（中北）、扬之水中文话剧社（中北）



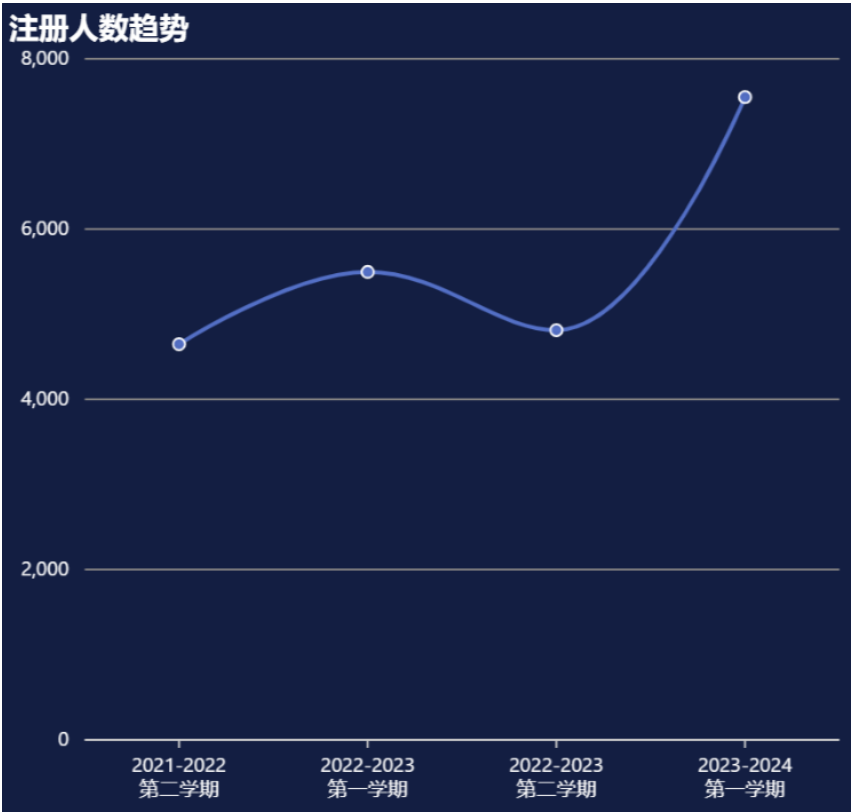
饼图

表明社团星级分布 有百分比和具体数字



环形图

表明社团类型分布 有百分比和具体数字



折线图

状态为 effective

表明四个学期社团注册人数趋势 有具体数字



词云

状态为 effective

表明学生所在院系分布 有具体数字



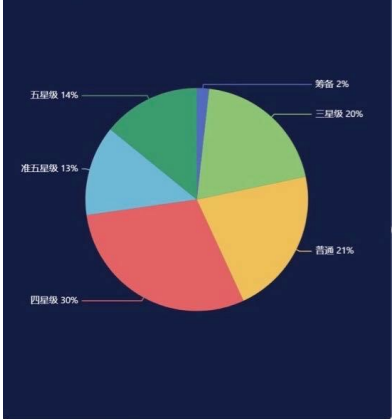
词云

状态为 effective

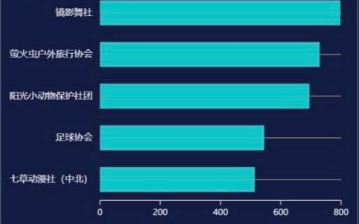
表明社团指导单位分布 有具体数字

学生社团注册数据分析报告

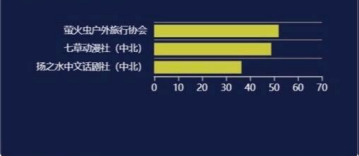
社团星级分布



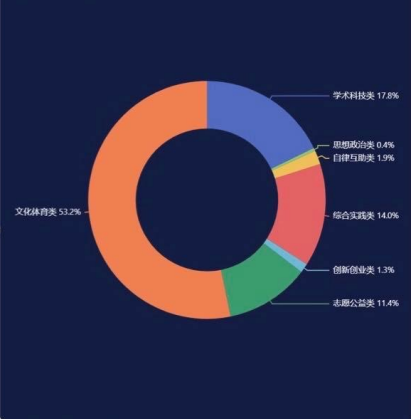
注册人数最多Top5



退社人数最多Top3



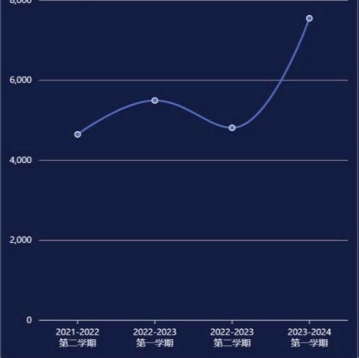
社團類型分布



学生所在院系分布



注册人数趋势



社团指导单位分布



注：数据是 dataease 平台上的 只能下载下来 1000 条