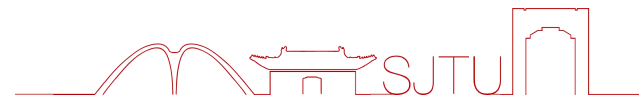




上海交通大学  
SHANGHAI JIAO TONG UNIVERSITY



# 基于萌娘百科wiki中明日方舟 相关数据的分析结果展示

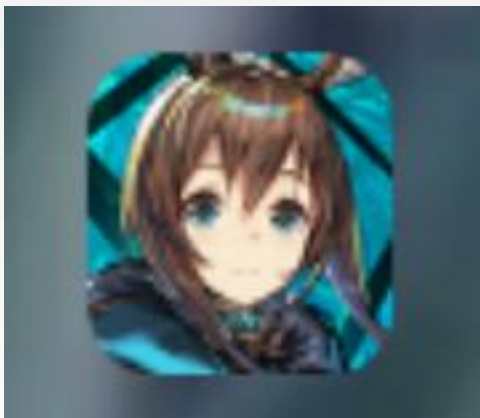
汇报人姓名：袁嘉靖，  
王桺赞，吴舒文

2024年1月11日

饮水思源 · 爱国荣校



# 简要介绍



明日方舟是一款由鹰角网络开发的二次元塔防游戏，游戏于2019年5月1日，到现在已经有4年半的时间了。游戏内实装的干员数量也已经由开服的几十个到现在的三百多个。游戏里的职业也由八个大职业拓展到现在的十几个分支职业。

|  |   |
|--|---|
|  <b>Vanguard</b><br>先锋<br>对部署有更佳效果的职业。  |  <b>Sniper</b><br>狙击<br>擅长远程物理攻击的输出职业。     |
|  <b>Guard</b><br>近卫<br>近战攻击为主的输出职业。     |  <b>Caster</b><br>术师<br>擅长远程法术攻击的输出职业。     |
|  <b>Defender</b><br>重装<br>高生命值与防御力的职业。  |  <b>Medic</b><br>医疗<br>提供治疗回复生命的职业。        |
|  <b>Specialist</b><br>特种<br>攻击手段多样的职业。 |  <b>Supporter</b><br>辅助<br>提供多样增益减益效果的职业。 |

八大职业分支





1

数据爬取

2

数据整理

3

数据展示及分析

01

# 数据爬取



我们使用requests库对萌娘百科 (<https://zh.moegirl.org.cn/Mainpage>) 中明日方舟词条下的部分数据进行爬取，存入json文件供后续处理

```
1 import requests
2 from lxml import etree
3 import json
4 import time
5
6 header = {'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/119.0.0.0 Safari/537.36 Edg/119.0.0.0'}
7
8 def get_data(url, header):
9     time.sleep(1)
10    response = requests.get(url, headers=header)
11    response.encoding = 'utf-8'
12    return response
13
14
15 originalUrl = 'https://zh.moegirl.org.cn/%E6%98%8E%E6%97%A5%E6%96%B9%E8%88%9F/%E5%B9%B2%E5%91%98%E5%9B%BE%E9%89%B4'
16
17 originalHTML = etree.HTML(get_data(originalUrl, header).text)
18 nextUrls = originalHTML.xpath('//div[@class="akI1l"]/span[@class="akI1lLayer7"]/a/@href')
19
20 final_data = []
```





# 数据爬取



萌 明日方舟/干员图鉴 - 萌娘百科 万 x +

< > C 家 zh.moegirl.org.cn/明日方舟/干员图鉴

萌娘百科

商务合作

力检测器要

搜索萌娘百科 (按"/快速搜索)



## 目录

- 1 干员筛选
- 2 按稀有度索引
  - 2.1 六星
  - 2.2 五星
  - 2.3 四星
  - 2.4 三星
  - 2.5 二星
  - 2.6 一星
- 3 按安装日期索引

上传文件 | 特殊页面 | 可打印版本 | 固定链接 | 页面信息 | 键入页面 | 相关更改

DevTools is now available in Chinese!

Always match Chrome's language Switch DevTools to Chinese Don't show again

Elements Console Sources Network >> 41 6 2

```
<div class="akI1l" data-operator-name="温蒂" style="cursor: pointer;">...</div>
<div class="akI1l" data-operator-name="歌蕾蒂娅" style="cursor: pointer;">...</div>
<div class="akI1l" data-operator-name="水月" style="cursor: pointer;">...</div>
<div class="akI1l" data-operator-name="老鲤" style="cursor: pointer;">...</div>
<div class="akI1l" data-operator-name="归溟幽灵鲨" style="cursor: pointer;">...</div>
<div class="akI1l" data-operator-name="多萝西" style="cursor: pointer;">...</div>
<div class="akI1l" data-operator-name="缄默德克萨斯" style="cursor: pointer;">...</div>
<div class="akI1l" data-operator-name="麒麟R夜刀" style="cursor: pointer;">...</div>
<div class="akI1l" data-operator-name="琳琅诗怀雅" style="cursor: pointer;">...</div>
<h3>...</h3>
```

Styles Computed Layout Event Listeners DOM Breakpoints Properties Accessibility

Filter :hov .cls +

```
element.style {
}
```

```
.mw-parser-output .akI1l a {
  color: #FFF!important;
}
```

```
a {
  --color: var(--theme-link-color);
  color: var(--color);
  text-decoration: none;
}
```

```
html, body, div, span, applet, iframe, h1, h2, h3, h4, h5, h6, p,
blockquote, pre, a, abbr, acronym, address, big, cite, code, del, dfn, em, img, ins, kbd, q,
s, samp, small, strike, strong, sub, sup, tt, var, b, u, i, center, dl, dt, dd, menu, ol,
ul, li, fieldset, form, label, legend, article, aside, canvas, details, embed, figure,
figcaption, footer, header, hgroup, main, nav, output, ruby, section, summary, time, mark,
audio, video {
```

在干员图鉴词条下爬取了所有已经录入的干员名字

”





初始 精英二 闪耀阶梯/I-宣传策略 表情差分

生命上限 725/1705  
攻击 178/528 (+90) (+24) (+27)  
防御 61/172 (+24)  
法术抗性 0  
再部署 慢 70秒 (-4秒)  
部署费用 12/14 (-2)  
阻挡数 1  
攻击速度 快 1.0秒 (+4)

\*\*\*\*\*

Archetto

空弦

画师: alchemaniac

配音: 水濑祈 (日语)

木雅瑞林 (汉语)



E0 E1

远程位

作者: alchemaniac [1]

### 基本资料

代号 空弦  
Archetto  
本名 席德佳  
别号 弦宝、三弦姐、弦子姐、鹰头猫<sup>[2]</sup>  
性别 女  
发色 金发  
瞳色 左蓝右红  
身高 152cm  
生日 11月16日  
种族 黎博利  
职业 狙击  
专精 射击 (弓)、军事教育、酿酒、烘焙  
萌点 金发、修女、异色瞳、披风、长靴、弓、猫耳  
宣传策略: 偶像  
出生地区 拉特兰  
角色EP 《Till the Bell Tolls》

对每一个干员的页面爬取如图所示的红色的框取部分的数据

上传文件



02

## 数据处理





我们把之前得到的json文件用python进行处理，删除不需要的数据，按分析需求调整数据格式，转存至csv

```
1
2 import numpy as np
3
4 import pandas as pd
5 import matplotlib.pyplot as plt
6
7 import re
8
9 from collections import Counter
10
11
12 df = pd.read_json('data.json')
13
```

```
df['hp1'] = [int(i.split('/')[0]) for i in df['生命上限']]
df['hp2'] = [int(i.split('/')[1]) for i in df['生命上限']]
```

```
df.to_csv('ArkModified1.csv')
```



03

## 数据展示及分析



# 数据展示及分析——总体数值分布



**Vanguard**  
**先锋**

对部署有更佳效果的职业。



**Sniper**  
**狙击**

擅长远程物理攻击的输出职业。



**Guard**  
**近卫**

近战攻击为主的输出职业。



**Caster**  
**术师**

擅长远程法术攻击的输出职业。



**Defender**  
**重装**

高生命值与防御力的职业。



**Medic**  
**医疗**

提供治疗回复生命的职业。



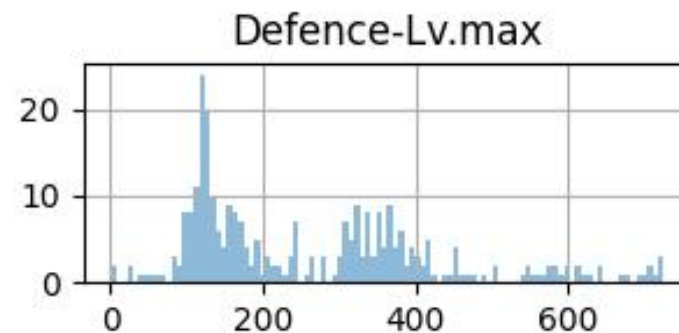
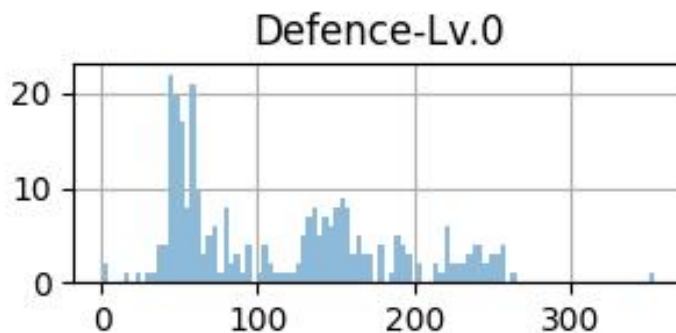
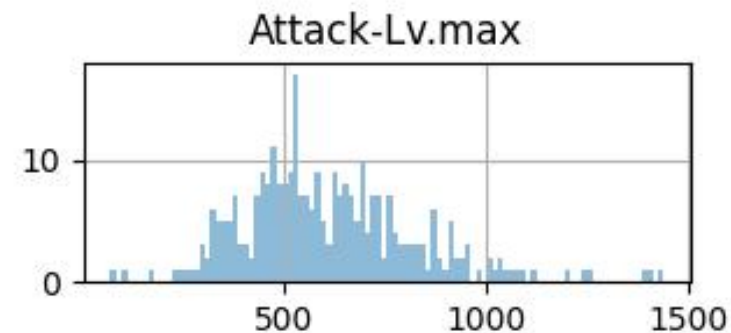
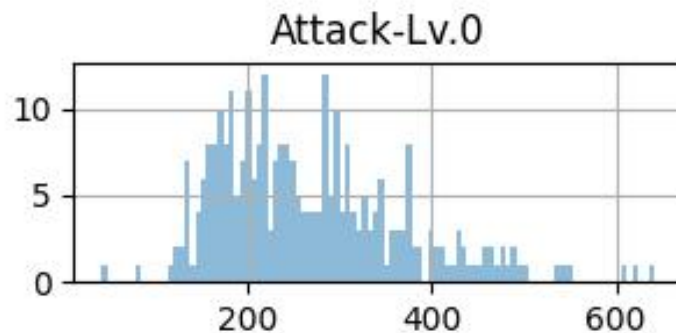
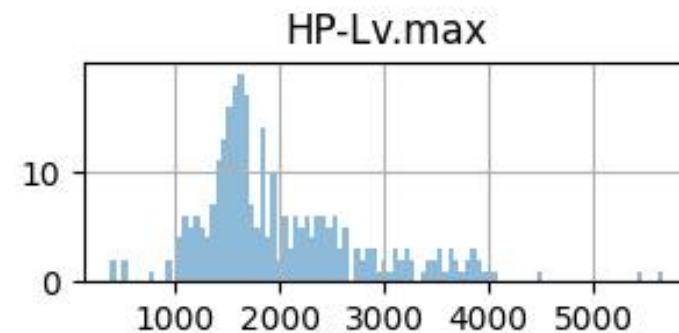
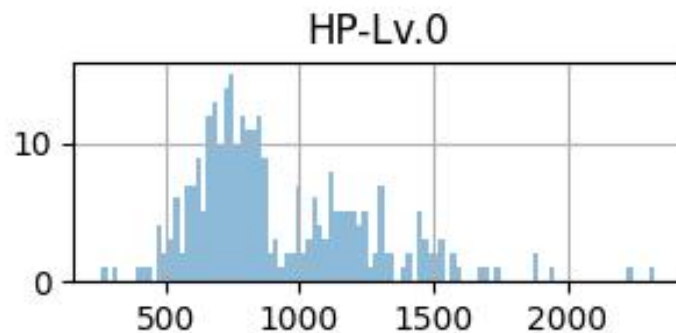
**Specialist**  
**特种**

攻击手段多样的职业。



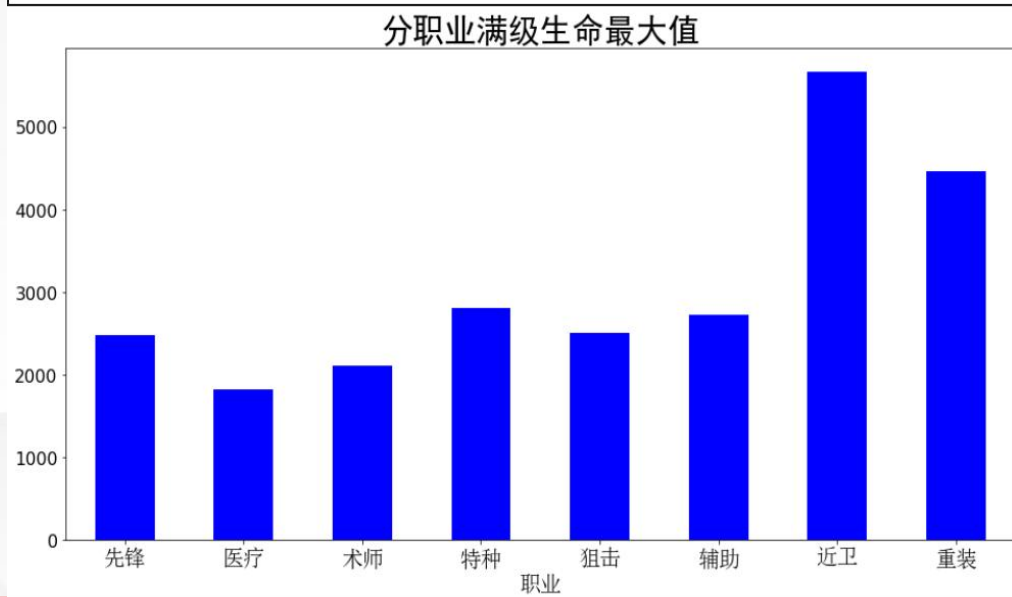
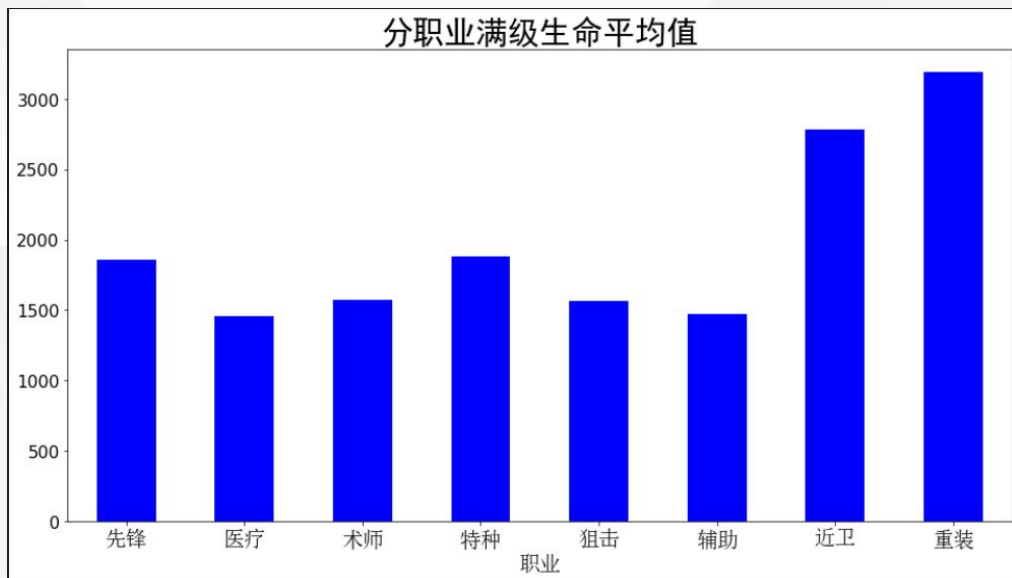
**Supporter**  
**辅助**

提供多样增益减益效果的职业。





# 数据展示及分析——分职业对比分析



对职业进行分类后计算属性的平均值。

我们可以发现生命值较多的职业是近卫，重装这些近战干员，而这些职业主要的作用也是提供防护，近战输出的功能。

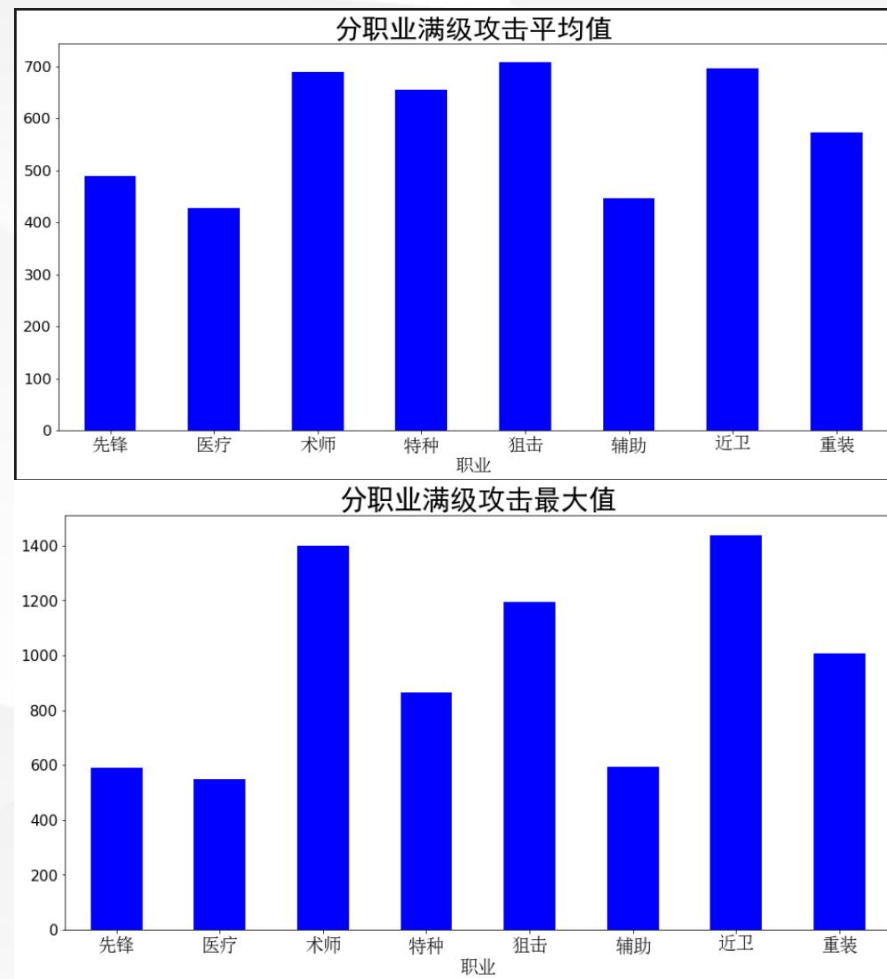
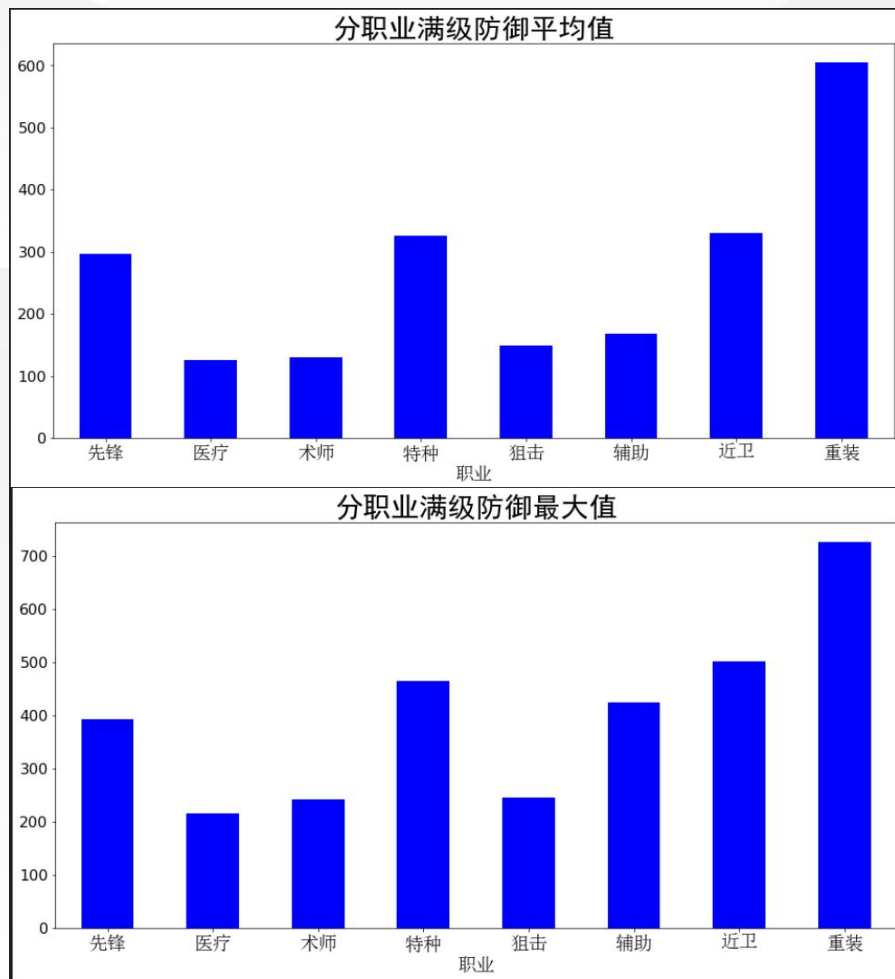
而医疗，术师，狙击，辅助等远程职业的生命值则明显更低，这说明我们需要更合理的选择摆放位置来让这些干员尽量少的受到伤害





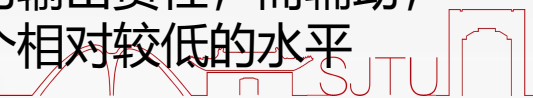


# 数据展示及分析——分职业对比分析



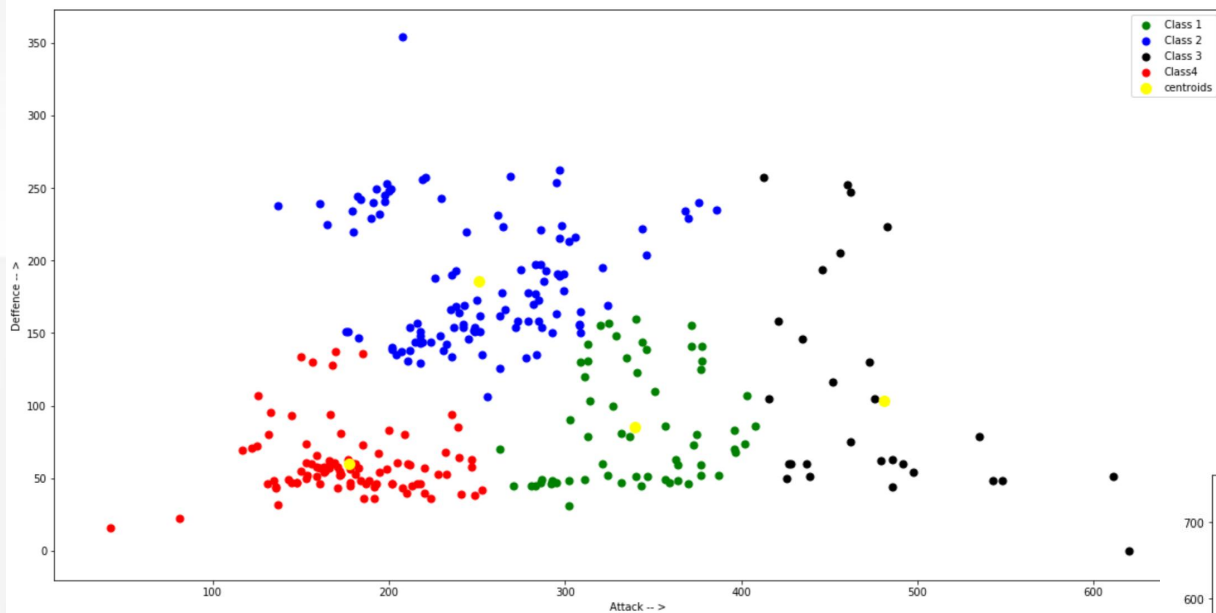
防御的数据中重装职业作为保护的职业自然遥遥领先

术师，狙击和近卫这些偏向于输出的干员职业都有较高的攻击力值，他们也承担着主要的输出责任，而辅助，医疗等职业的攻击力也自然在一个相对较低的水平



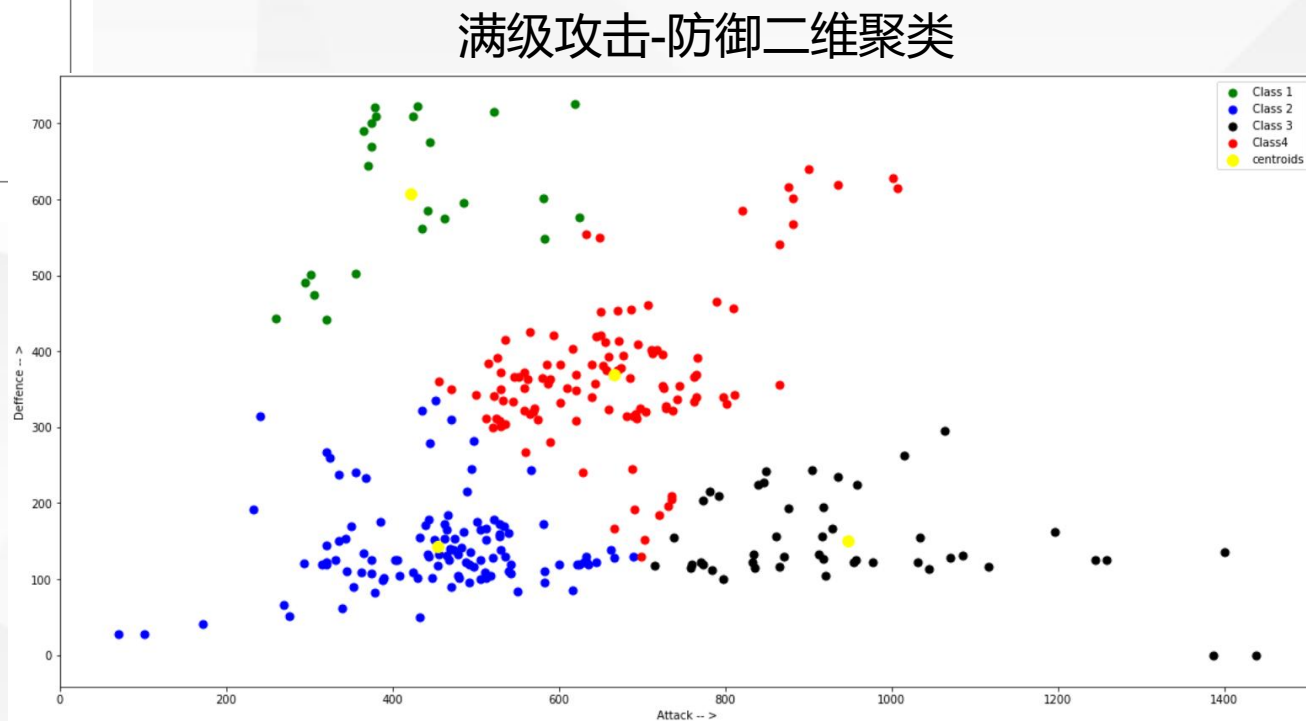


# 数据展示及分析——Kmeans聚类



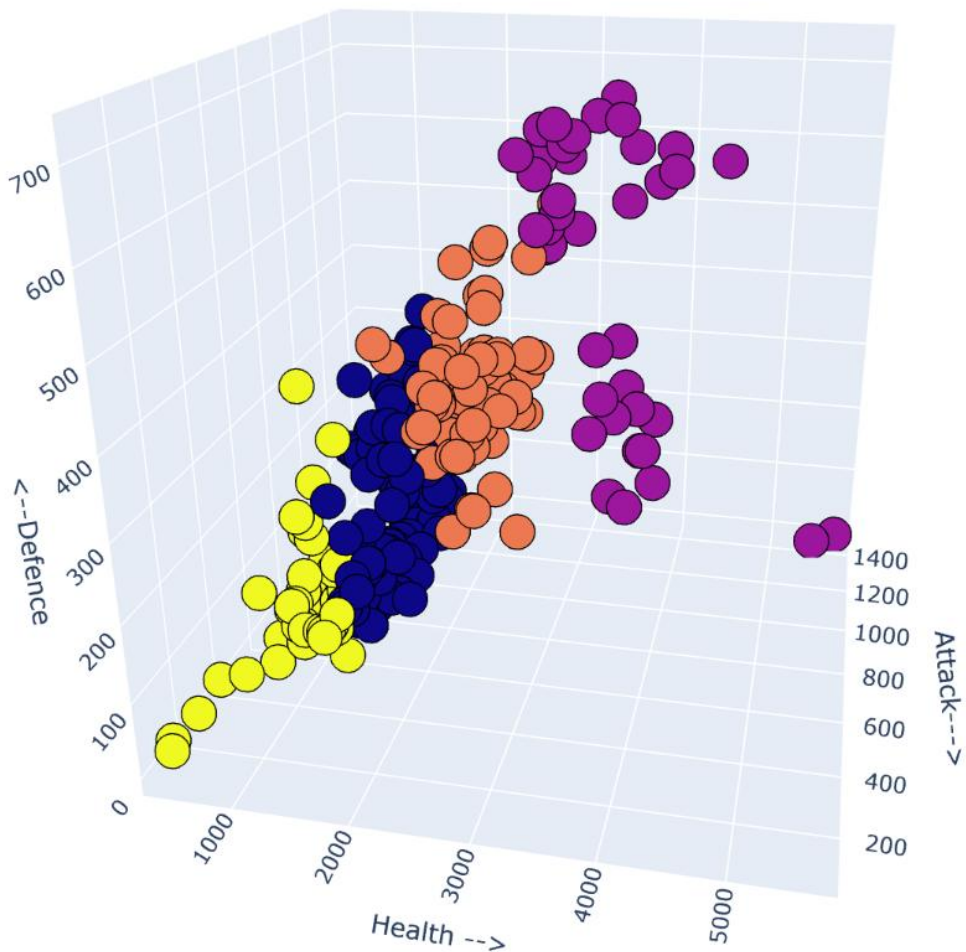
零级攻击-防御二维聚类

通过WCSS算法选择K值，完成聚类算法

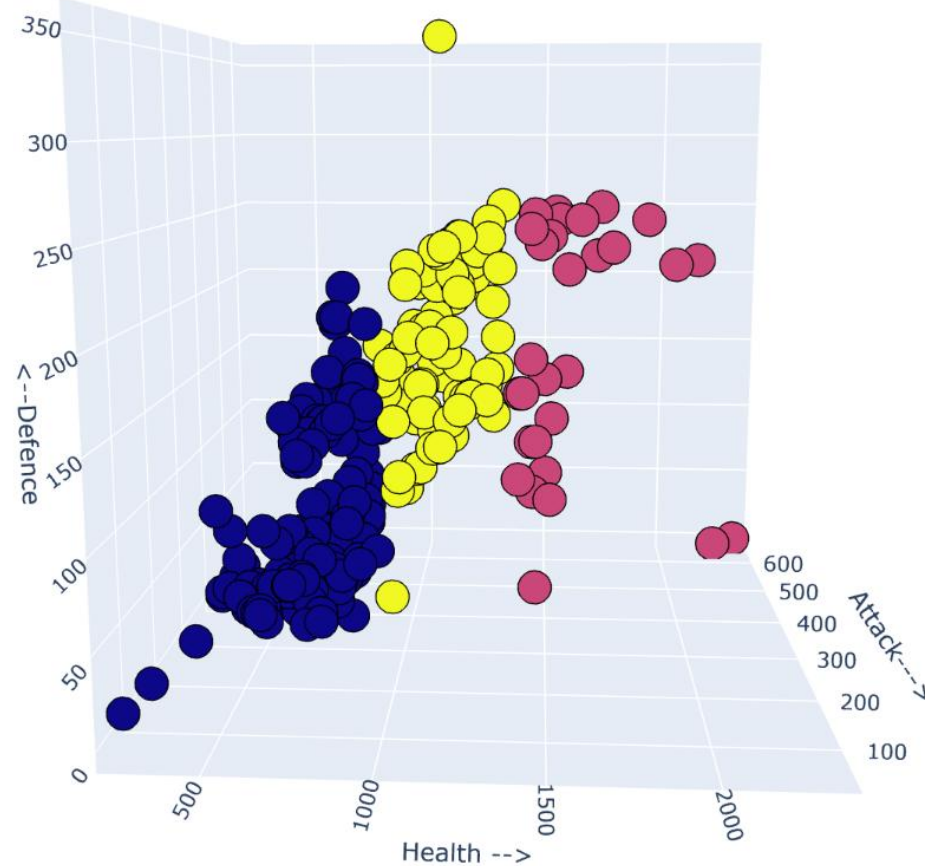




# 数据展示及分析——三维Kmeans聚类



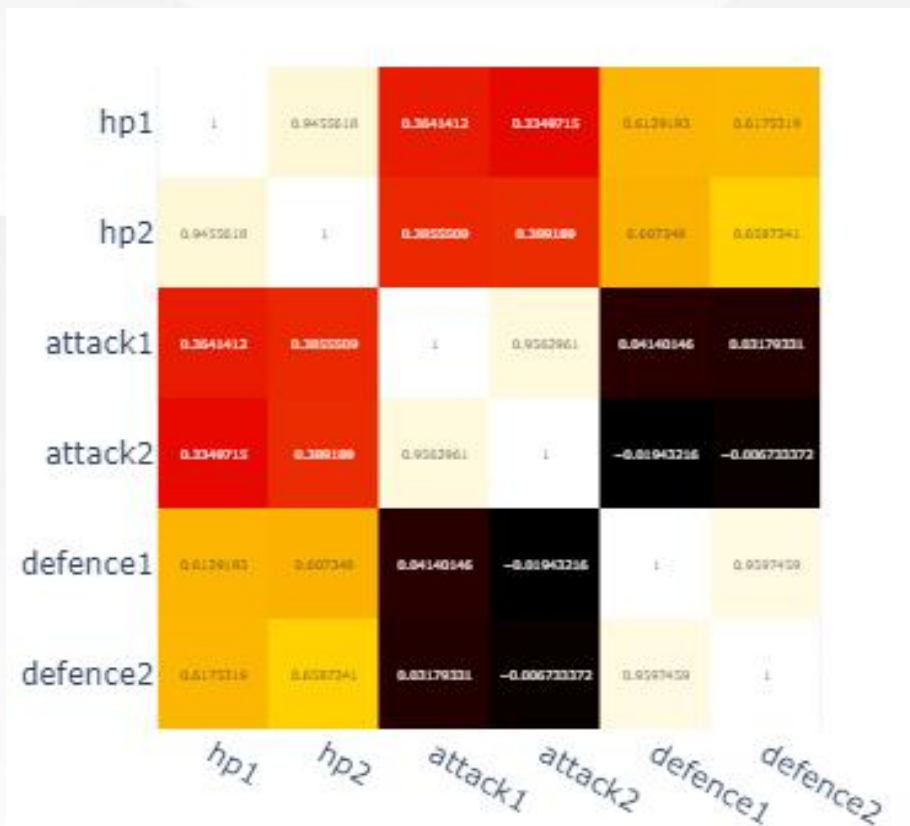
满级攻击-防御-生命值三维聚类



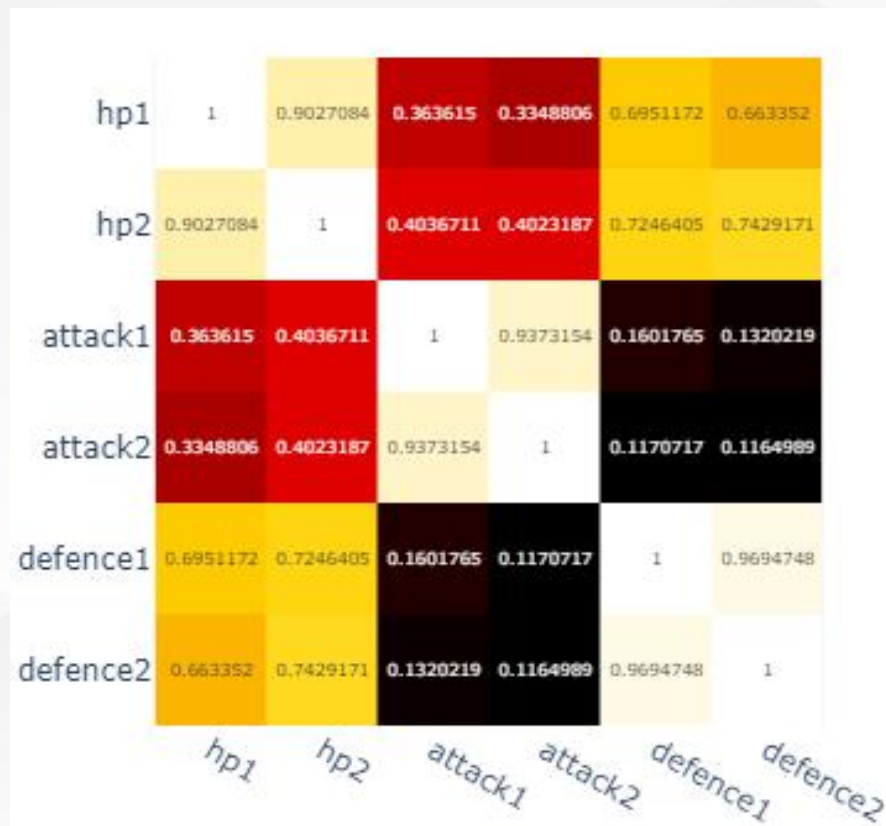
零级攻击-防御-生命值三维聚类



# 数据展示及分析——相关性分析



全部角色各属性Pearson相关系数



全部角色各属性Spearman相关系数

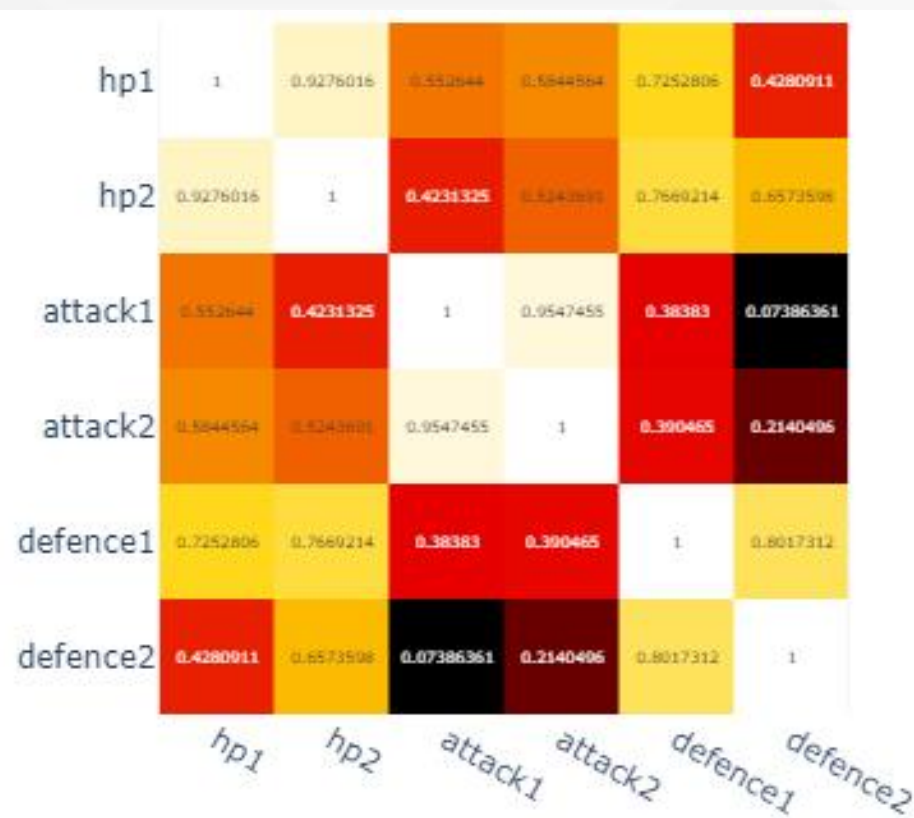
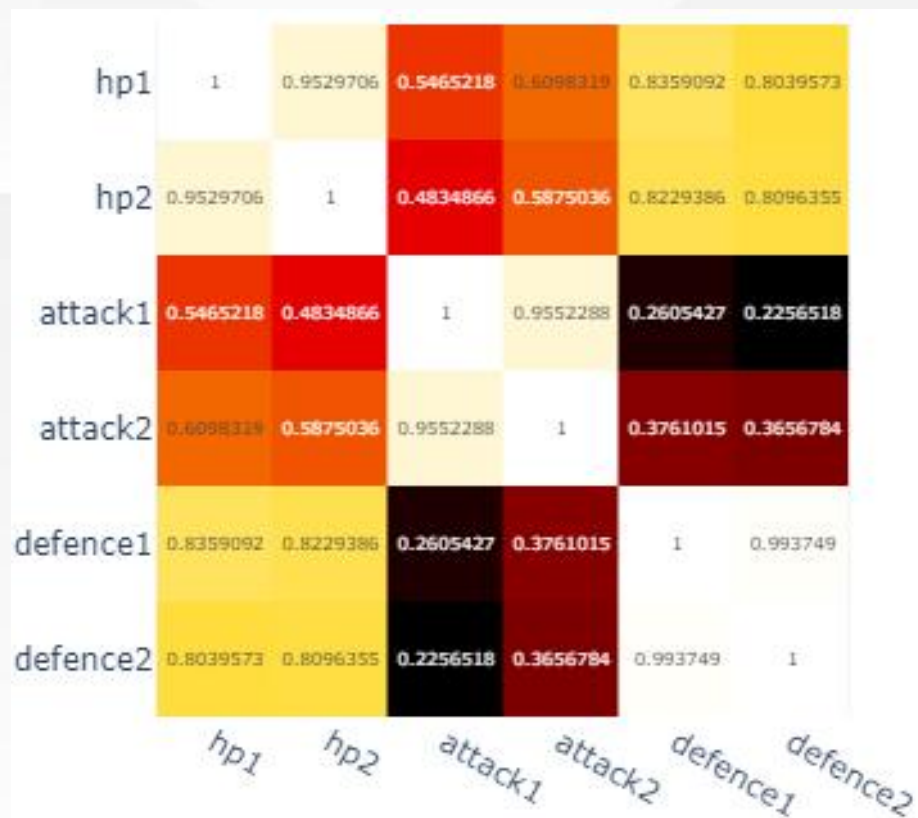
总体特征：对应的零级属性与满级属性高度相关







# 数据展示及分析——分职业相关性分析

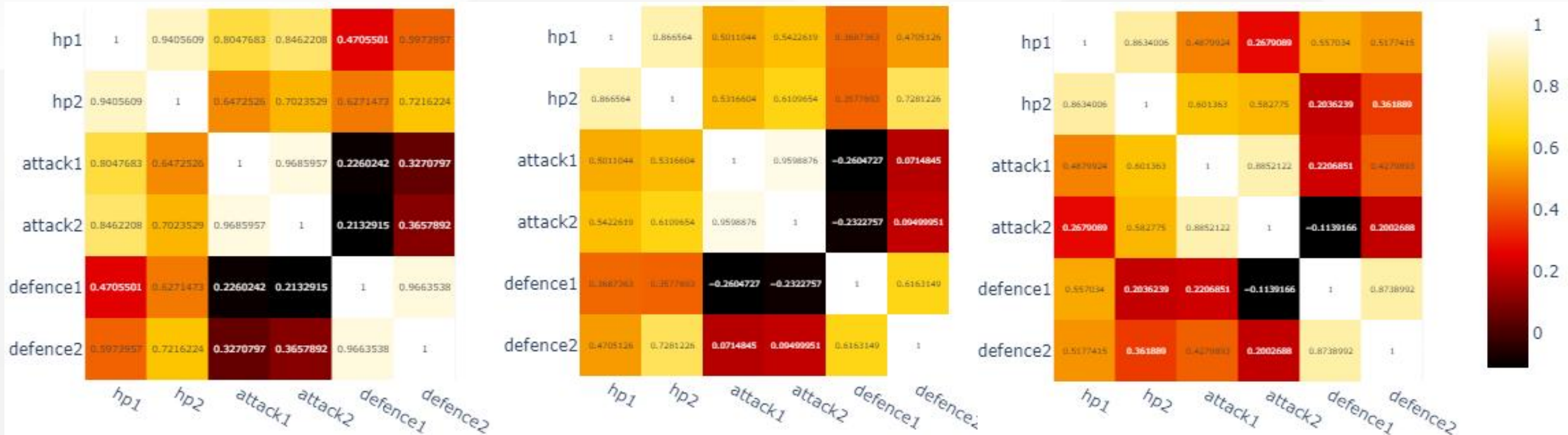


第一类：防御与生命的相关性较高





# 数据展示及分析——分职业相关性分析

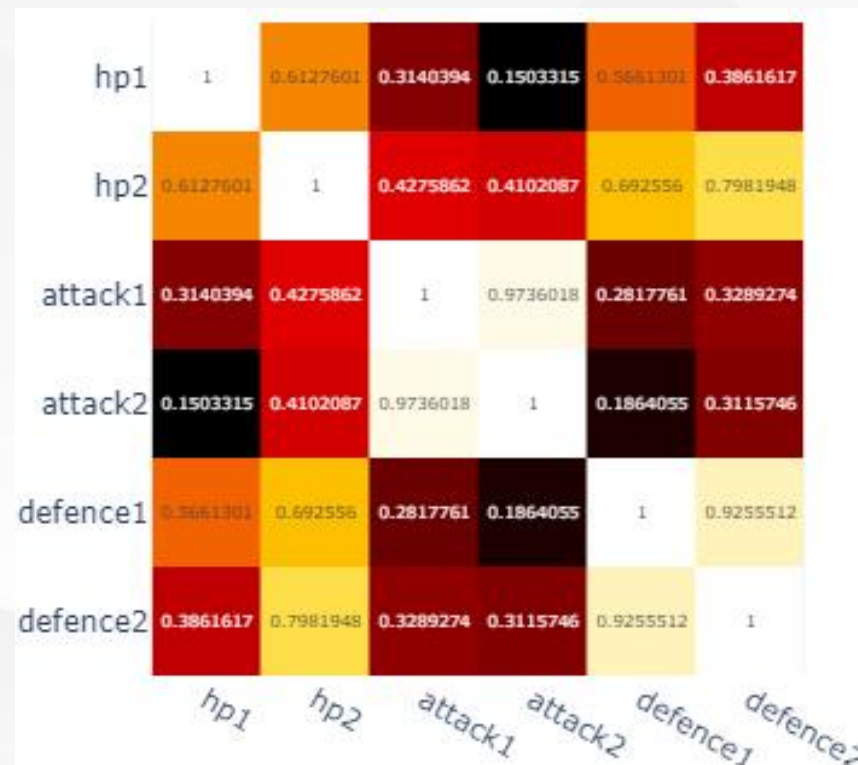
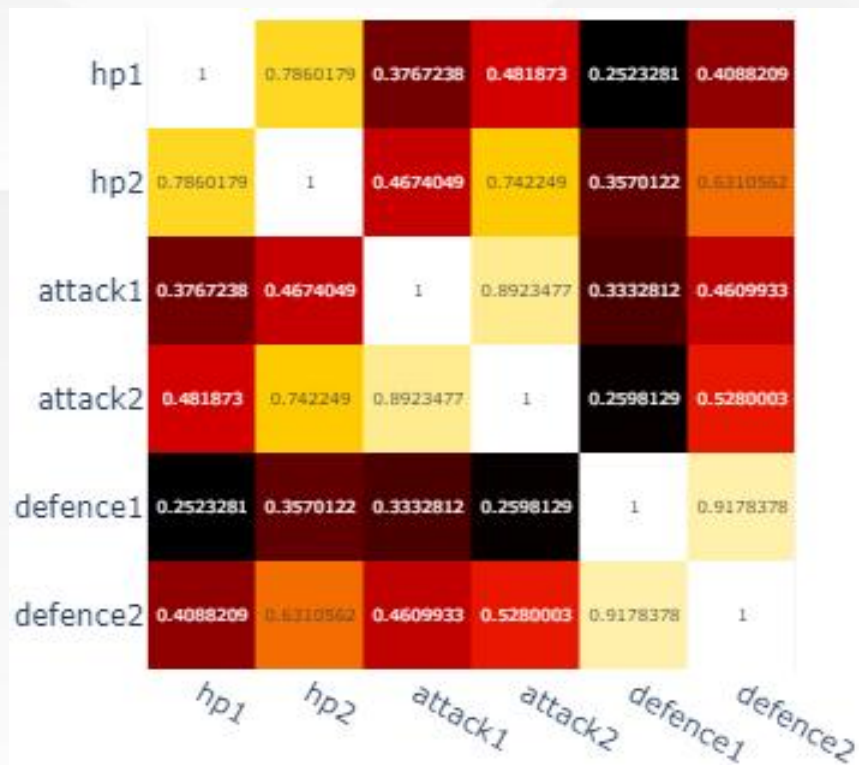


第二类：总体看相关性较低，但攻击与生命的相关性相对较高





# 数据展示及分析——分职业相关性分析

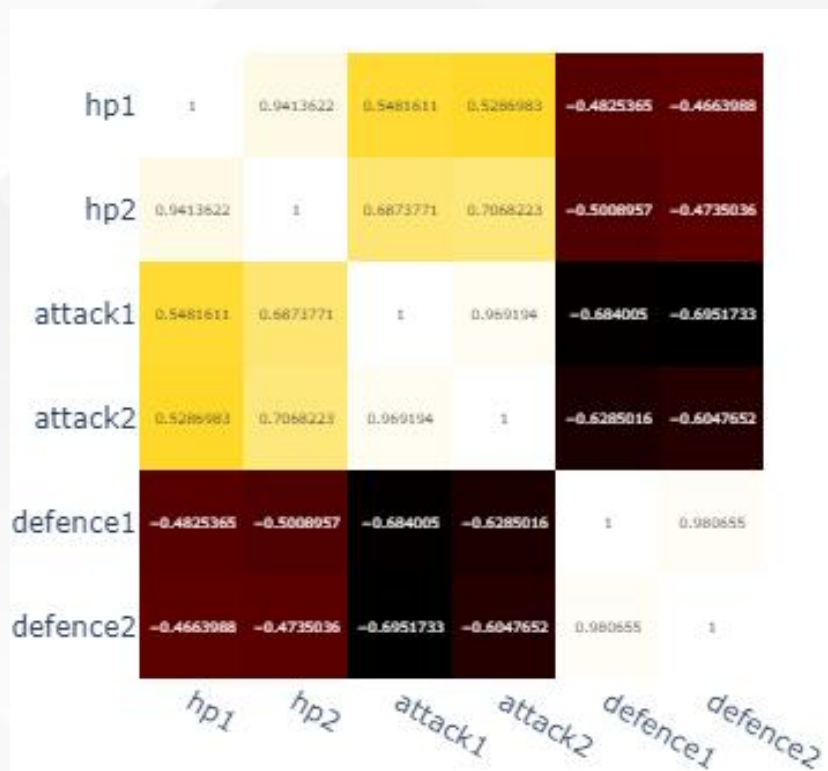


第三类：相关性较低





# 数据展示及分析——分职业相关性分析



第四类：只有近卫一个职业，攻击与生命呈正相关，攻击和防御呈负相关







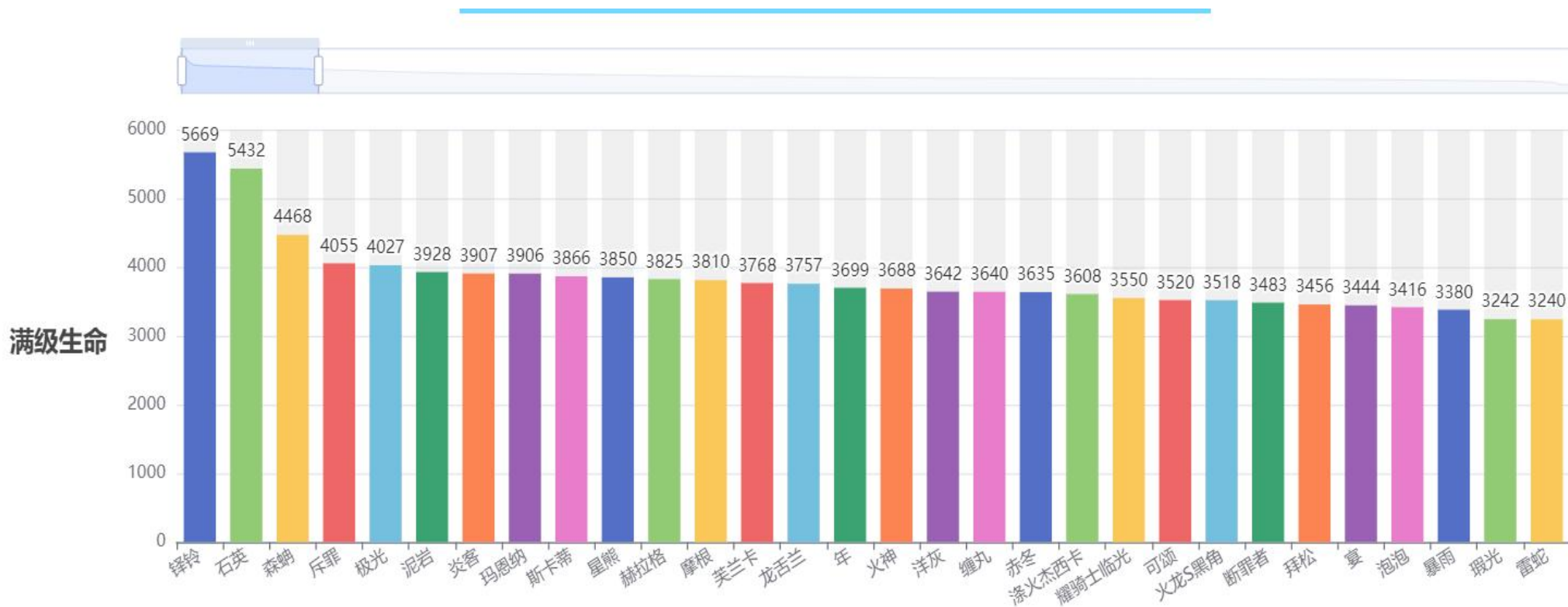
最后，我们用echarts实现了对数据的网页可视化展示

主要能够展示干员的生命值，攻击力，防御力，法术抗性，身高和生日，还有画师的‘劳模’程度。而且根据不同职业单独做了页面展示





# 数据展示及分析



用降序排列的直方图展示了所有干员的生命值等信息



# 数据展示及分析

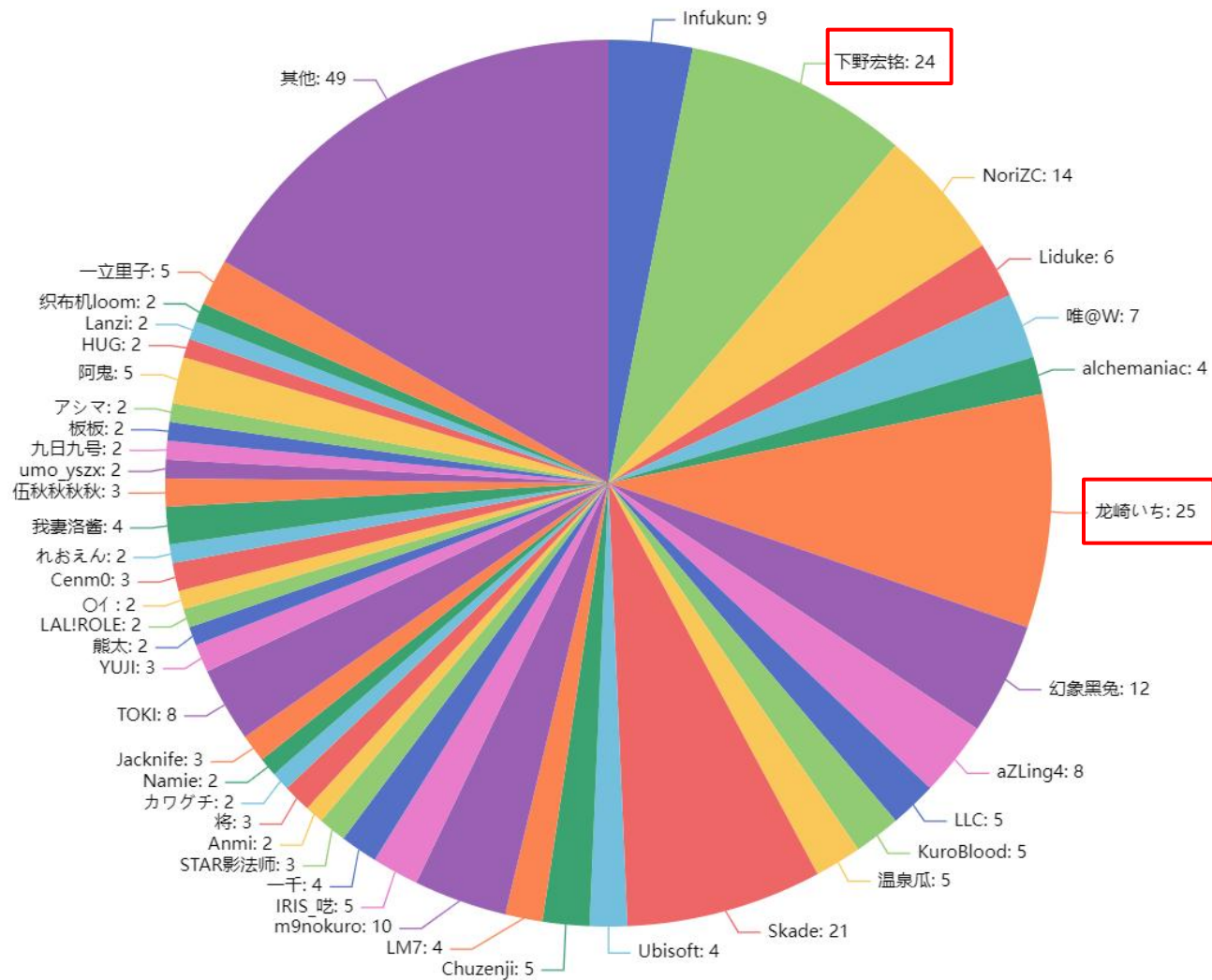
生日

2023

| 1月          | 2月       | 3月         | 4月          | 5月         | 6月         | 7月        | 8月            | 9月        | 10月     | 11月    | 12月     |
|-------------|----------|------------|-------------|------------|------------|-----------|---------------|-----------|---------|--------|---------|
| 1 15 22 29  | 5        | 5 12 19 26 | 2 23 30     | 14         | 25         | 2 9 16 30 | 13 20         | 10 17     | 8 15 29 |        | 3 24 31 |
| 2 23 30     | 13 20 27 | 6 13 20 27 | 3 1 8 15    | 5 12 26    | 10 17 31   |           | 4 25          | 2 9 16 30 | 6 20    | 11 25  |         |
| 3 17 31     | 14       | 7 14       | 4 11 2 9 30 | 6 27       | 8 22 29    | 12        | 3 10 17 24 31 | 14 21 28  | 5 12    |        |         |
| 4 18 1 8 15 | 15 22    | 5 3 10 31  | 7 21 28     | 12 26 2    | 23 30      |           | 18 25 1       | 22 29     | 13 20   |        |         |
| 5 19 2 16 2 | 20       | 4 18 1     | 22          | 6 20 27    | 3 10       | 7 21      |               | 2 9 16 30 | 21      |        |         |
| 13 20 27    | 3 10 17  | 7 5 19     | 30          | 7 11 18 25 | 1 15 22 29 | 6 3 10    | 1 8 15 22 29  |           |         |        |         |
| 7 4         | 25       | 15 22      | 6 13 20 27  | 24         | 1 15 22 29 | 19        | 9 30          | 7 28      | 11      | 2 9 23 |         |

图中标出的日期都是有的干员的生日

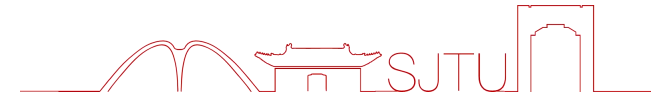
”







上海交通大学  
SHANGHAI JIAO TONG UNIVERSITY



# 结论



饮水思源 · 爱国荣校



# 感谢聆听

上海交通大学

SHANGHAI JIAO TONG UNIVERSITY

饮水思源 爱国荣校