



链滴

【释疑贴】社区帖子与用户的数据统计（新手必看）

作者：[soulfight](#)

原文链接：<https://ld246.com/article/1566621201255>

来源网站：[链滴](#)

许可协议：[署名-相同方式共享 4.0 国际 \(CC BY-SA 4.0\)](#)

<p>一开始对社区的导航方式无法适应，不知道从哪里开始查看感兴趣的贴子（开始没有发现登录后以直接使用搜索功能以为只能用 Google 的站内搜索），所以想把贴子标题爬下来快速浏览，最初的法是从最新贴子里找动态加载分页信息的 url，但是发现不好计算出总页数也无法区分动态加载的 json 数据是不是含有重复的混淆数据，试了好几次没有结果。</p>

<p>后来发现主页的底部导航有数据统计的页面链接，没想到社区所有注册用户信息是可以公开访问，于是先爬了下七百多页的用户列表页提取出了 50207 位用户的用户名和个人主页链接，接着逐个取用户主页上的数据，最后爬取公开贴子列表用户（有 68 位用户选择不公开贴子列表）的贴子列表取到 16124 条贴子的数据。最近将爬取到的数据做成了 3 个 Excel 表，部分截图如下，有兴趣的小伙伴可以查看隐藏的打赏内容。</p>

<p>

</p>

<p>注册用户数最多的 10 个城市（基于公开城市位置的用户统计）：</p>

排行	城市	人数
TOP.1	北京	4270
TOP.2	上海	2637
TOP.3	深圳	2224
TOP.4	广州	1870
TOP.5	杭州	1768
TOP.6	成都	1310

```
</tr>
<tr>
<td>TOP.7</td>
<td>武汉</td>
<td>969</td>
</tr>
<tr>
<td>TOP.7</td>
<td>南京</td>
<td>928</td>
</tr>
<tr>
<td>TOP.8</td>
<td>重庆</td>
<td>675</td>
</tr>
<tr>
<td>TOP.8</td>
<td>西安</td>
<td>613</td>
</tr>
<tr>
<td>TOP.9</td>
<td>苏州</td>
<td>596</td>
</tr>
<tr>
<td>TOP.9</td>
<td>长沙</td>
<td>542</td>
</tr>
<tr>
<td>TOP.9</td>
<td>郑州</td>
<td>526</td>
</tr>
<tr>
<td>TOP.10</td>
<td>济南</td>
<td>434</td>
</tr>
</tbody>
</table>
<p><strong>【黑客】成员与【画家】成员的数量:</strong></p>
<table>
<thead>
<tr>
<th>身份</th>
<th>人数</th>
<th>占比</th>
</tr>
</thead>
<tbody>
<tr>
```

```

<td>黑客</td>
<td>46424</td>
<td>92%</td>
</tr>
<tr>
<td>画家</td>
<td>3783</td>
<td>8%</td>
</tr>
</tbody>
</table>
<p><strong>活跃用户数:</strong></p>
<table>
<thead>
<tr>
<th>条件</th>
<th>人数</th>
<th>占比</th>
</tr>
</thead>
<tbody>
<tr>
<td>发贴数不为 0</td>
<td>2730</td>
<td>5.4%</td>
</tr>
<tr>
<td>发贴数不少于 10</td>
<td>396</td>
<td>0.8%</td>
</tr>
<tr>
<td>回贴数不为 0</td>
<td>4747</td>
<td>9.5%</td>
</tr>
</tbody>
</table>
<p><strong>贴子数据统计:</strong></p>
<table>
<thead>
<tr>
<th>过滤条件</th>
<th>贴子数量</th>
</tr>
</thead>
<tbody>
<tr>
<td>非管理员节点贴子浏览数不少于 500</td>
<td>1742</td>
</tr>
<tr>
<td>非管理员节点贴子回贴数不为 0</td>
<td>7227</td>

```

```
</tr>
<tr>
<td>非管理员节点贴子回贴数不为 0 或浏览数不少于 500</td>
<td>7776</td>
</tr>
<tr>
<td>非管理员节点贴子回贴数不少于 20 且浏览数不少于 500</td>
<td>235</td>
</tr>
</tbody>
</table>
```

<p>2019 年 8 月份各时间段访问黑客派的在线人数平均值（按数据统计里显示的在线用数减去所有数据里面最低的一个数值 116）：</p>

- 在线人数最多的时刻：8 月 19 日 17:05 (733)
在线人数最少的时刻：8 月 25 日 05:50 (116)

```
</ul>
<table>
<thead>
<tr>
<th>时间段</th>
<th>访问网站在线人数</th>
</tr>
</thead>
<tbody>
<tr>
<td>6~7 点</td>
<td>53</td>
</tr>
<tr>
<td>7~8 点</td>
<td>61</td>
</tr>
<tr>
<td>8~9 点</td>
<td>91</td>
</tr>
<tr>
<td>9~10 点</td>
<td>190</td>
</tr>
<tr>
<td>10~11 点</td>
<td>280</td>
</tr>
<tr>
<td>11~12 点</td>
<td>336</td>
</tr>
<tr>
<td>12~13 点</td>
<td>328</td>
</tr>
<tr>
```

```
<td>13~14 点</td>
<td>317</td>
</tr>
<tr>
<td>14~15 点</td>
<td>344</td>
</tr>
<tr>
<td>15~16 点</td>
<td>372</td>
</tr>
<tr>
<td>16~17 点</td>
<td>394</td>
</tr>
<tr>
<td>17~18 点</td>
<td>385</td>
</tr>
<tr>
<td>18~19 点</td>
<td>305</td>
</tr>
<tr>
<td>19~20 点</td>
<td>248</td>
</tr>
<tr>
<td>20~21 点</td>
<td>230</td>
</tr>
<tr>
<td>21~22 点</td>
<td>223</td>
</tr>
<tr>
<td>22~23 点</td>
<td>212</td>
</tr>
<tr>
<td>23~0 点</td>
<td>187</td>
</tr>
<tr>
<td>0~1 点</td>
<td>142</td>
</tr>
<tr>
<td>1~2 点</td>
<td>100</td>
</tr>
<tr>
<td>2~3 点</td>
<td>77</td>
```

```
</tr>
<tr>
<td>3~4 点</td>
<td>66</td>
</tr>
<tr>
<td>4~5 点</td>
<td>59</td>
</tr>
<tr>
<td>5~6 点</td>
<td>53</td>
</tr>
</tbody>
</table>
<p><strong>新增用户登录之后在黑客派停留总时长的统计:</strong></p>
<table>
<thead>
<tr>
<th>在线总时长</th>
<th>人数</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 分钟 +</td>
<td>11295</td>
</tr>
<tr>
<td>1 小时 +</td>
<td>6569</td>
</tr>
<tr>
<td>10 小时 +</td>
<td>2161</td>
</tr>
<tr>
<td>20 小时 +</td>
<td>1349</td>
</tr>
<tr>
<td>50 小时 +</td>
<td>670</td>
</tr>
<tr>
<td>100 小时 +</td>
<td>372</td>
</tr>
<tr>
<td>200 小时 +</td>
<td>196</td>
</tr>
<tr>
<td>720 小时 +</td>
```

```
<td>48</td>
</tr>
<tr>
<td>任意时长</td>
<td>19250</td>
</tr>
</tbody>
</table>
<p><strong>追加 3 个统计表格: </strong></p>
<table>
<thead>
<tr>
<th>年份</th>
<th>年度发贴数</th>
<th>年度新注册用户数</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>55</td>
<td>79</td>
</tr>
<tr>
<td>2013</td>
<td>498</td>
<td>380</td>
</tr>
<tr>
<td>2014</td>
<td>158</td>
<td>186</td>
</tr>
<tr>
<td>2015</td>
<td>953</td>
<td>1569</td>
</tr>
<tr>
<td>2016</td>
<td>3066</td>
<td>3591</td>
</tr>
<tr>
<td>2017</td>
<td>3505</td>
<td>5141</td>
</tr>
<tr>
<td>2018</td>
<td>4681</td>
<td>26953</td>
</tr>
<tr>
```



```
<td>2019.8</td>
<td>3207</td>
<td>12308</td>
</tr>
</tbody>
</table>
<table>
<thead>
<tr>
<th>月份</th>
<th>发帖分布数</th>
<th>新注册用户分布数</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 月</td>
<td>1317</td>
<td>5936</td>
</tr>
<tr>
<td>2 月</td>
<td>976</td>
<td>2678</td>
</tr>
<tr>
<td>3 月</td>
<td>1566</td>
<td>2884</td>
</tr>
<tr>
<td>4 月</td>
<td>1455</td>
<td>2732</td>
</tr>
<tr>
<td>5 月</td>
<td>1262</td>
<td>8574</td>
</tr>
<tr>
<td>6 月</td>
<td>1253</td>
<td>3987</td>
</tr>
<tr>
<td>7 月</td>
<td>1396</td>
<td>4814</td>
</tr>
<tr>
<td>8 月</td>
<td>1673</td>
<td>5239</td>
```

```

</tr>
<tr>
<td>9 月</td>
<td>1311</td>
<td>3575</td>
</tr>
<tr>
<td>10 月</td>
<td>1126</td>
<td>3326</td>
</tr>
<tr>
<td>11 月</td>
<td>1418</td>
<td>3362</td>
</tr>
<tr>
<td>12 月</td>
<td>1370</td>
<td>3100</td>
</tr>
</tbody>
</table>
<table>
<thead>
<tr>
<th>时间段</th>
<th>发帖总数</th>
<th>收到的回贴总数</th>
</tr>
</thead>
<tbody>
<tr>
<td>0~1 点</td>
<td>419</td>
<td>1978</td>
</tr>
<tr>
<td>1~2 点</td>
<td>209</td>
<td>891</td>
</tr>
<tr>
<td>2~3 点</td>
<td>66</td>
<td>290</td>
</tr>
<tr>
<td>3~4 点</td>
<td>32</td>
<td>120</td>
</tr>
<tr>
<td>4~5 点</td>

```

```
<td>27</td>
<td>30</td>
</tr>
<tr>
<td>5~6 点</td>
<td>13</td>
<td>32</td>
</tr>
<tr>
<td>6~7 点</td>
<td>40</td>
<td>162</td>
</tr>
<tr>
<td>7~8 点</td>
<td>74</td>
<td>245</td>
</tr>
<tr>
<td>8~9 点</td>
<td>306</td>
<td>1650</td>
</tr>
<tr>
<td>9~10 点</td>
<td>1160</td>
<td>5817</td>
</tr>
<tr>
<td>10~11 点</td>
<td>1419</td>
<td>7815</td>
</tr>
<tr>
<td>11~12 点</td>
<td>1335</td>
<td>6368</td>
</tr>
<tr>
<td>12~13 点</td>
<td>698</td>
<td>2939</td>
</tr>
<tr>
<td>13~14 点</td>
<td>820</td>
<td>4050</td>
</tr>
<tr>
<td>14~15 点</td>
<td>1200</td>
<td>4970</td>
</tr>
<tr>
```

15~16 点	1217	5022
16~17 点	1273	4550
17~18 点	1257	3920
18~19 点	793	2214
19~20 点	610	1597
20~21 点	719	2563
21~22 点	784	2362
22~23 点	906	2914
23~0 点	746	2729

分析发现，每天最佳发帖时间段是 10~11 点。

以上社区开放的数据只放在社区给大伙解疑。如果你还对会员升级系统感兴趣，请看：

[https://hacpai.com/article/1567678498082](https://ld246.com/forward?goto=https%3A%2F%2Fhacpai.com%2Farticle%2F156778498082)