

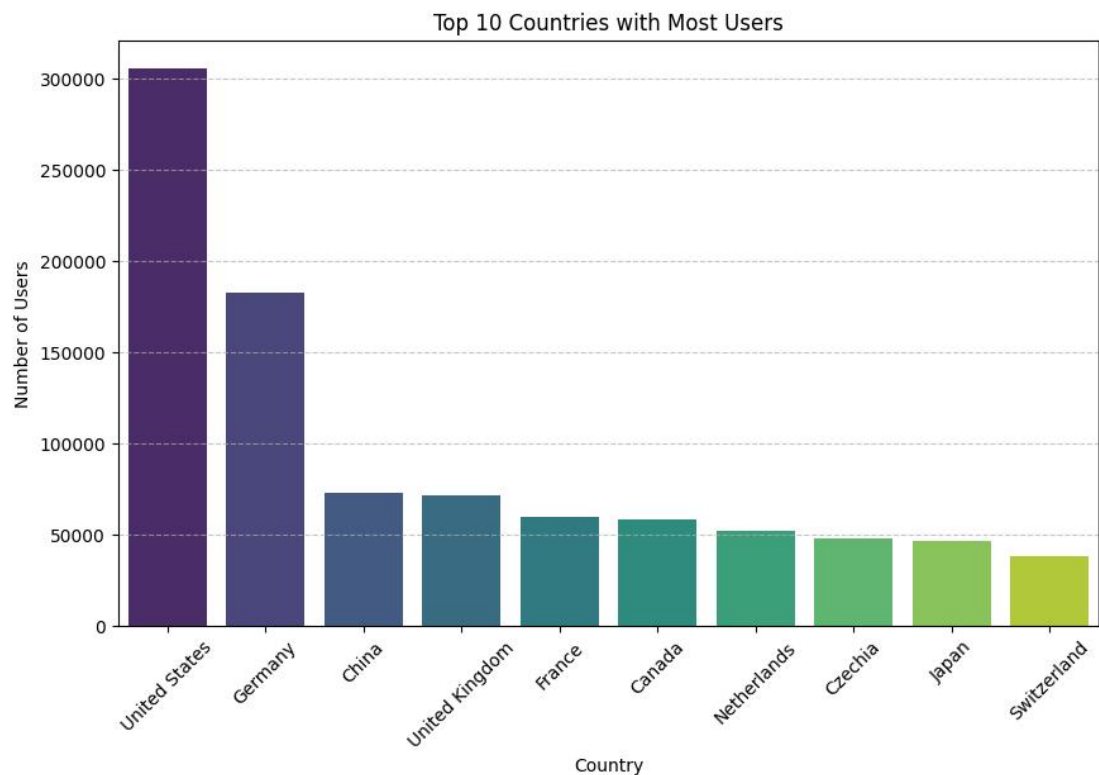
github 用户数据数据洞察报告

杨依婷 10223360435

1. 人口统计分析

1.1 国家和地区分布

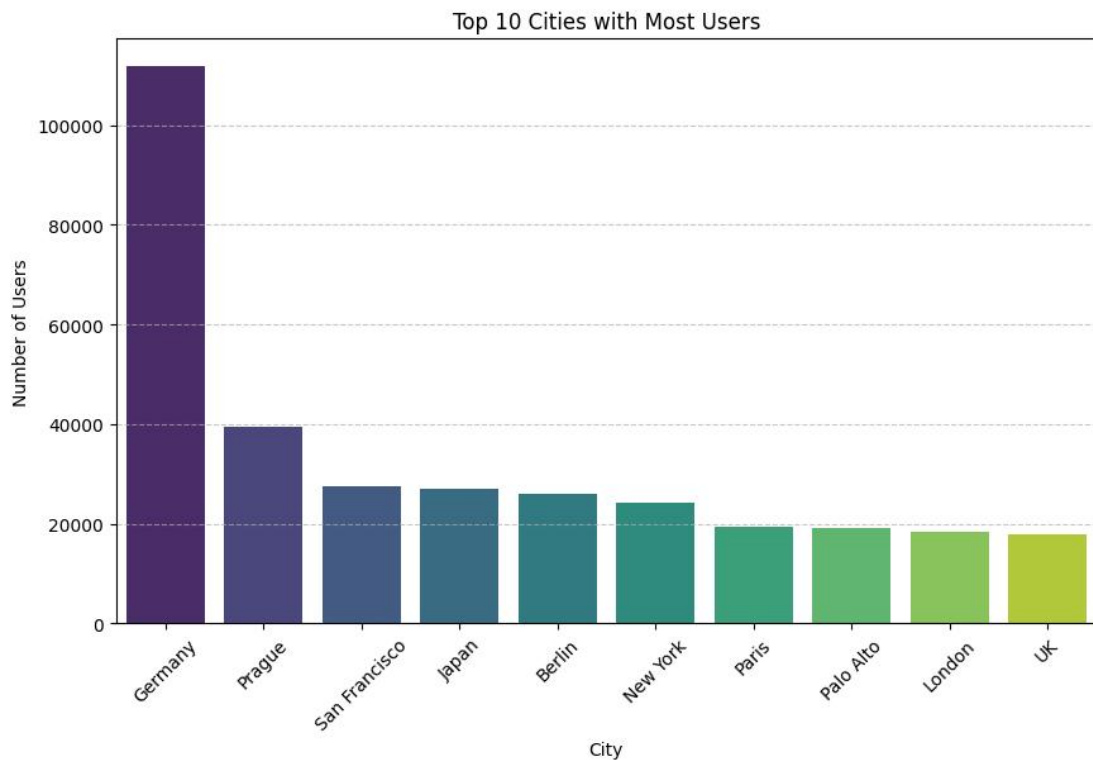
通过分析用户所在国家的分布，我们可以识别出全球主要的开发者集中地。以下是用户数量最多的前 10 个国家



从上面的数据来看，美国是全球开发者最集中的国家，紧随其后的是德国和中国。这表明这些国家可能有更多的技术和开发活动，可能也是开发者社群和技术会议的主要举办地。

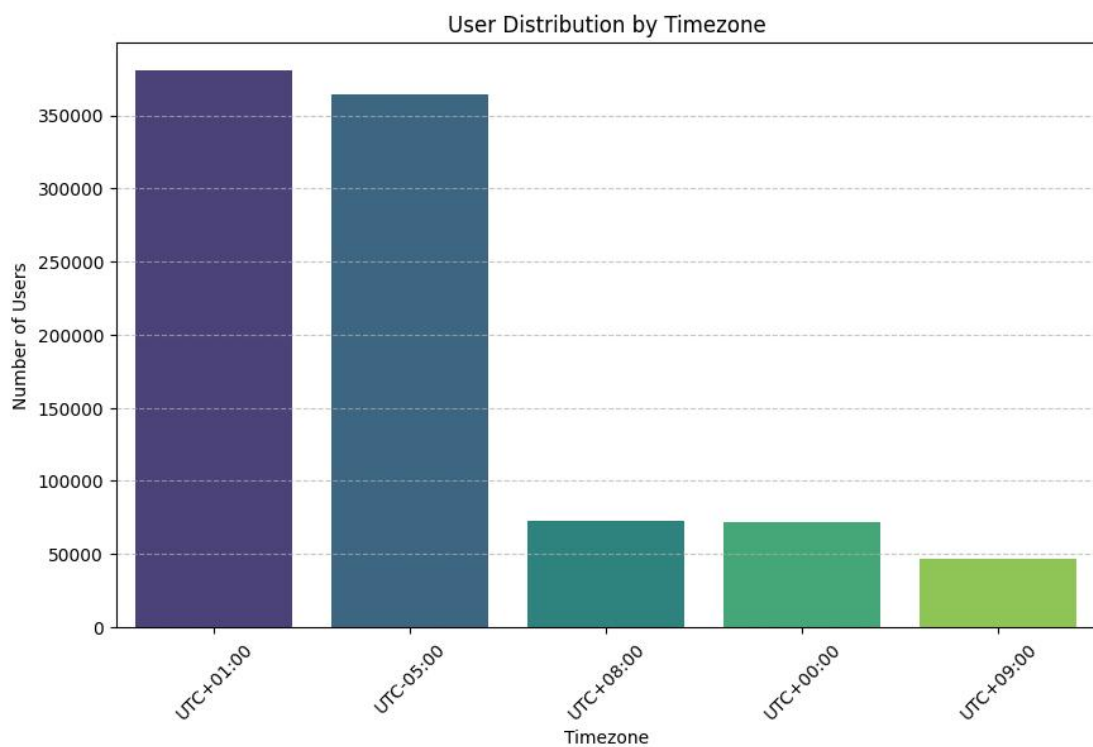
1.2 城市级别分布

接下来我们分析了开发者所在的城市分布，发现以下 10 个城市用户数量最多：



此处德国被识别城市是因为数据问题。从城市分布来看，**Prague**（捷克首都）和 **San Francisco**（美国硅谷的中心）开发者密度最高，显示出这些地方的技术活跃度较高，尤其是在开发和初创公司领域。

### 1.3 时区分布



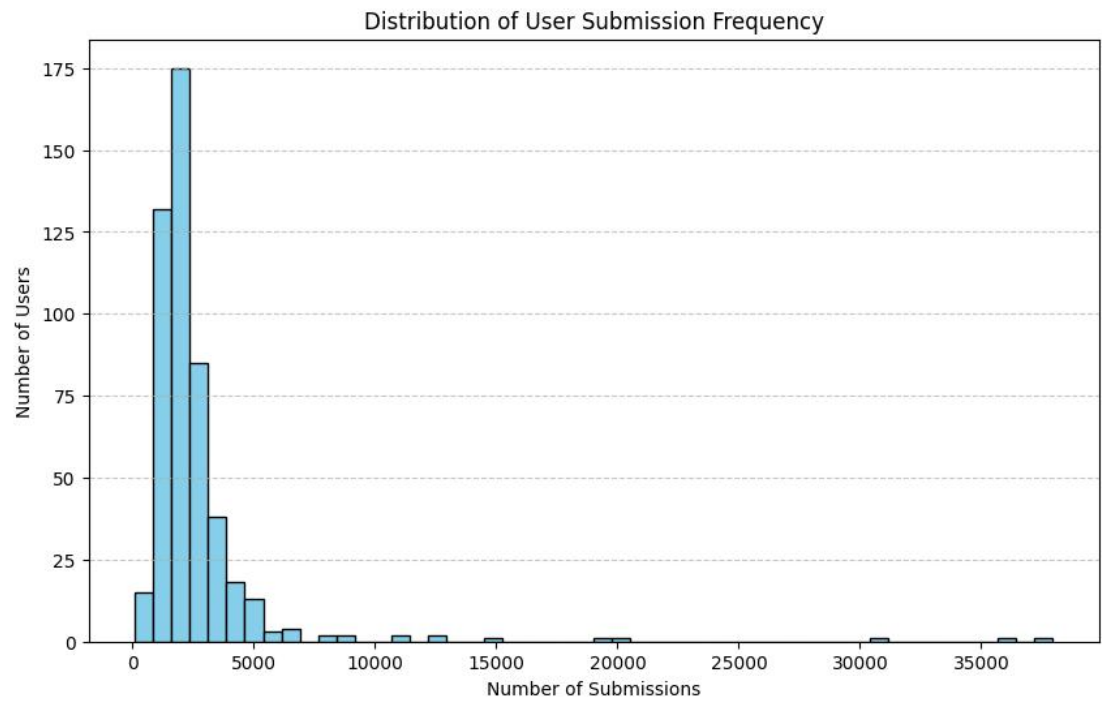
从时区分布来看，活跃用户主要集中在 UTC+01:00 和 UTC-05:00 时区，分别代表欧洲和美洲的用户。UTC+08:00 时区的用户较少，主要来自中国和东南亚。整体而言，欧洲和美洲的用户活跃度较高，而亚洲地区的用户相对较少。

2. 协作行为分析

2.1 提交频率分析

通过分析每个用户的提交频率，我们识别了用户的活跃度。以下是提交频率的描述性统计数据：

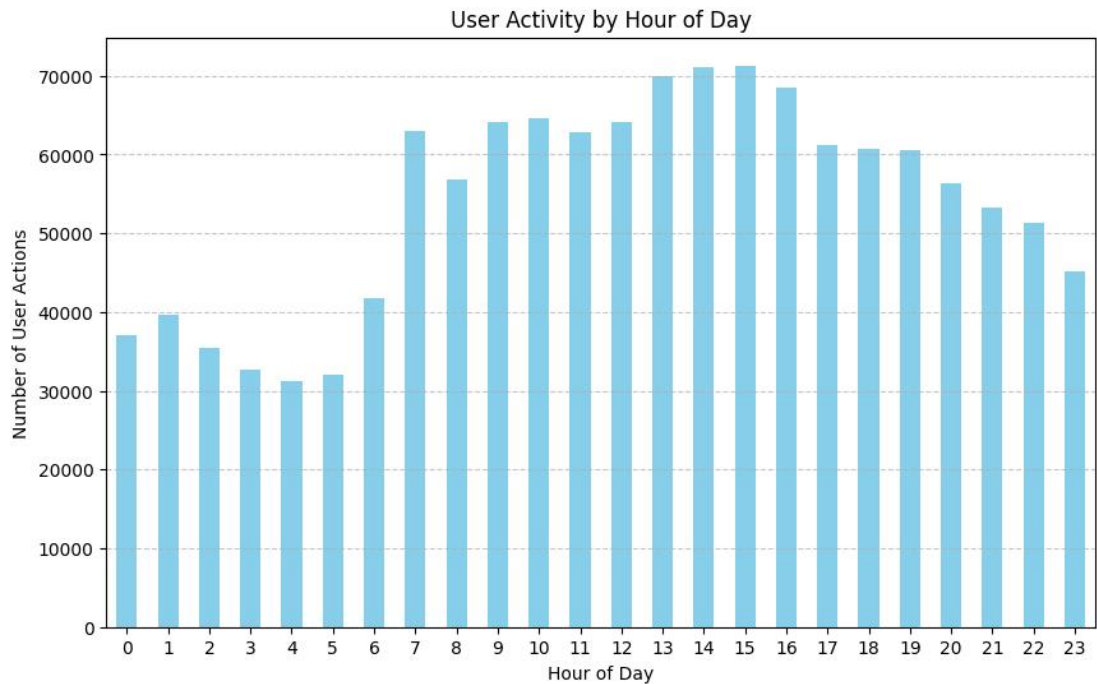
指标	数值
用户数	497
平均值	2605.18
标准差	3182.59
最小值	75
中位数	2026
最大值	37960



从这个分布来看，大多数用户的提交频率集中在较低的范围（75-3,000 次提交）。少数活跃用户有非常高的提交频率，说明有一部分用户参与项目的深度较高，可能是核心贡献者或维护者。

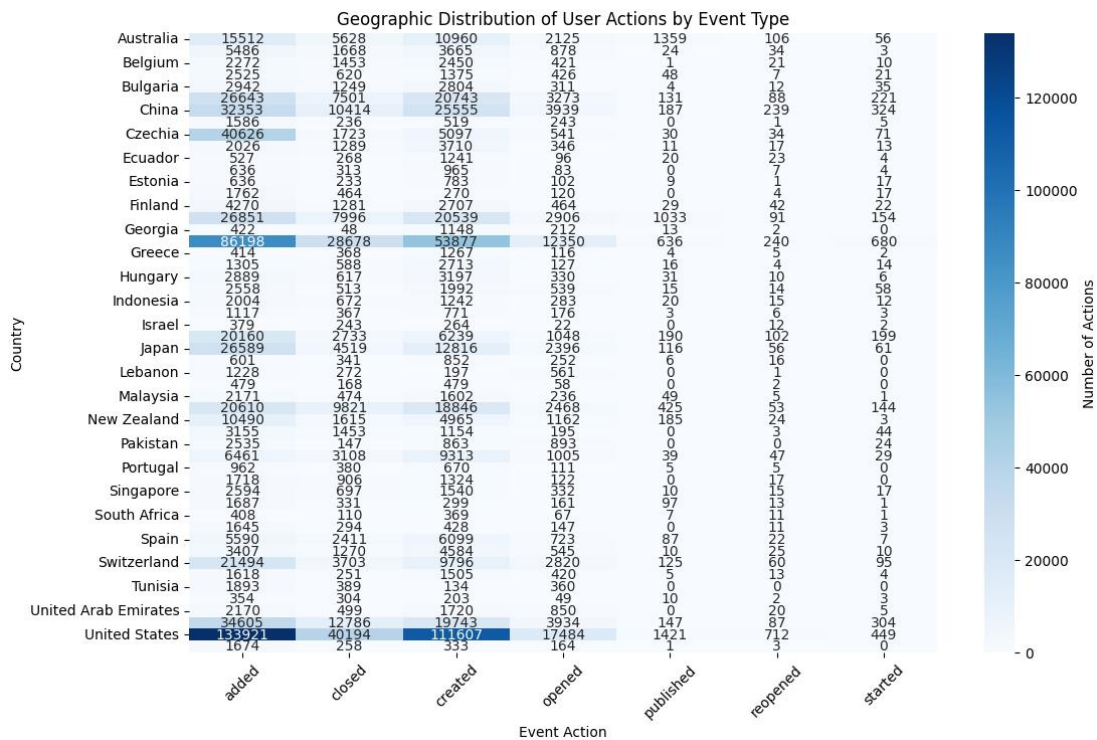
3. 其他维度有趣的洞察

3.1 用户活跃时间段分析



通过对用户提交行为时间的分析，我们可以识别出用户的活跃时间段。根据用户的活跃时间（按小时划分），我们发现开发者们在白天时段的活跃度较高，尤其是在上午 9 点到下午 6 点之间活跃频率最为集中。这符合全球开发者的工作时间模式，并且显示出开发者更倾向于在工作日的正常办公时间内进行活动。

### 3.2 地理位置与行为模式分析



结合用户的地理位置（城市和国家）与其行为类型，我们可以看到不同地区用户的行为差异。例如，美国的用户倾向于更多地创建和启动项目，而欧洲的用户可能更频繁地进行项目编辑

和关闭。这些差异可能反映了各地区在技术发展和项目管理的不同关注点，也可能影响全球协作时的工作模式。