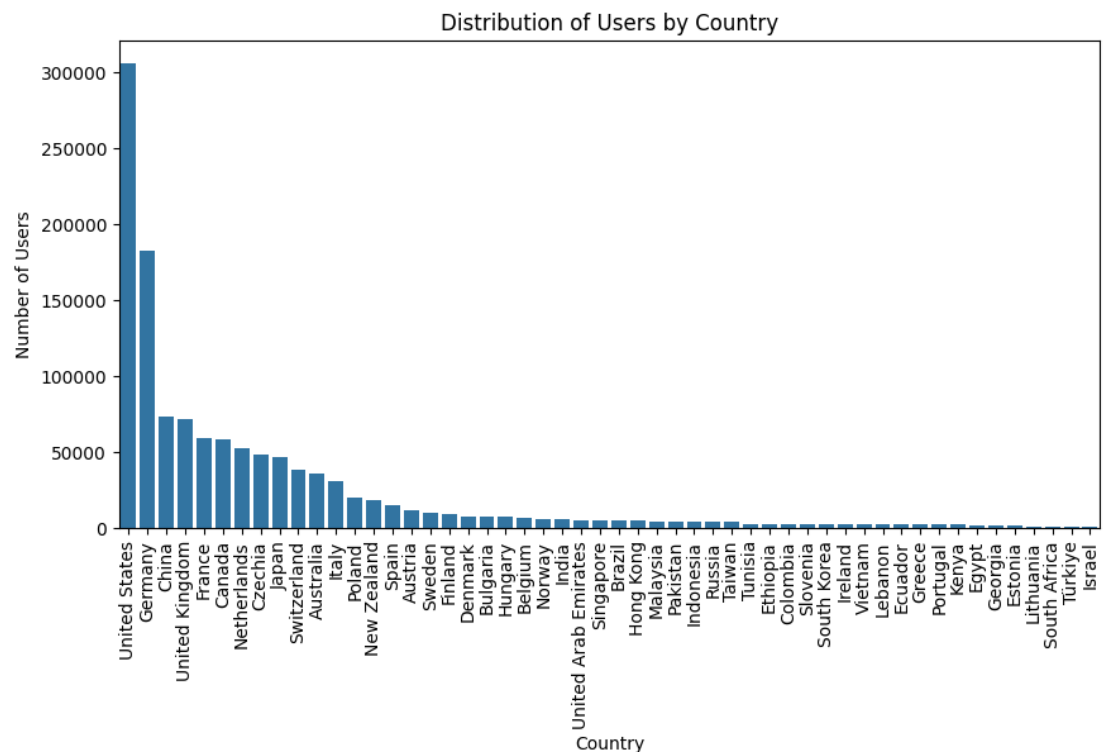


1. 国家和地区分布：

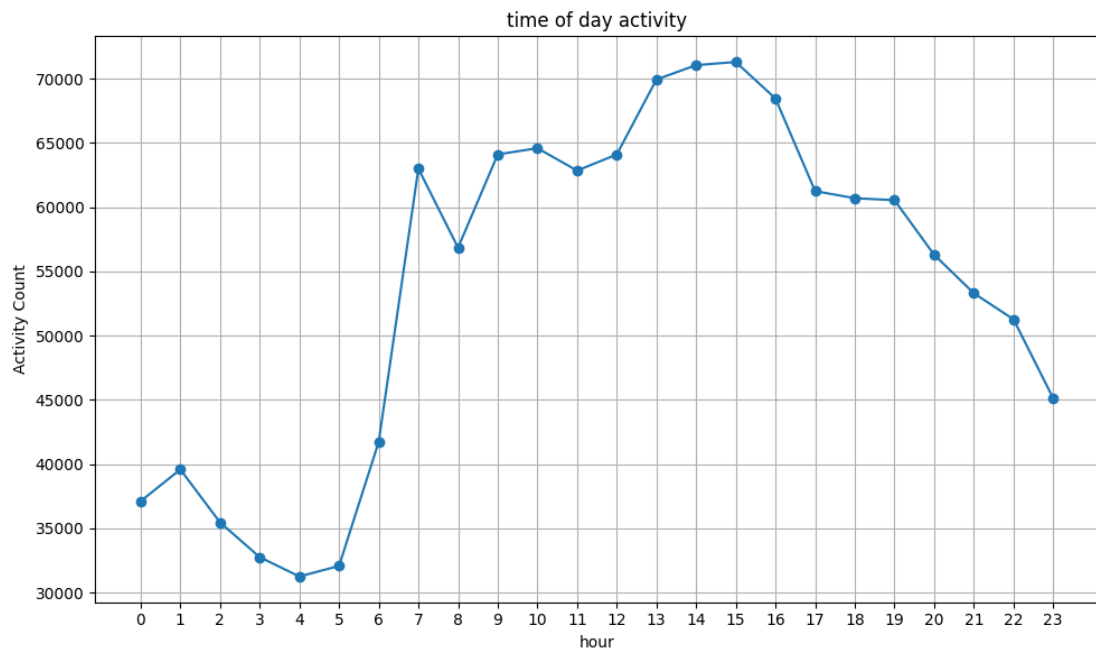
统计用户所属国家并作出柱状图：



主要开发者基地（国家）有：美国、德国、中国、英国、法国、加拿大等。这表明用户主要集中在少数几个国家。这些国家在开源项目和技术社区中具有较高的参与度。

主要开发者基地（城市）有：布拉格、加利福尼亚州帕洛阿尔托、纽约、旧金山、巴黎等。（某些数据的城市属性也为国家，因此并无计入结果）。可以看出科技发达的地方用户基数也大。

2. 协作时间与一天活跃时间分布：



用户在每日 7-20 时相对活跃。5-7 时显著增加，20-24 时缓慢下降。这与大多数开发者的工作时间相吻合。

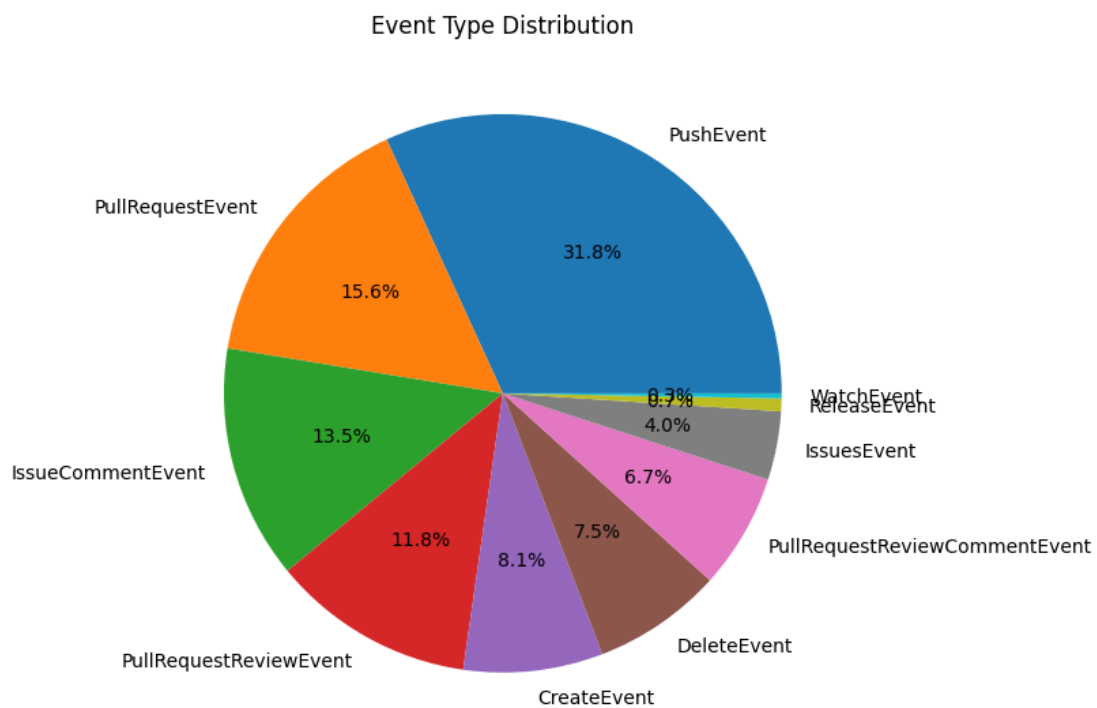
3. 识别高活跃与低活跃用户：

```

user_id
158862      37960
11146458    36400
28706372    30616
1580956     20300
40306929    19215
43724913    14505
9824526     12905
50149701    12704
195327      11218
663432      10764
Name: count, dtype: int64
user_id
4560679      750
999278        743
6379399       650
3162115       644
39504233       635
62625502       621
1609022        599
1133777        582
3238291        485
814283         75
Name: count, dtype: int64

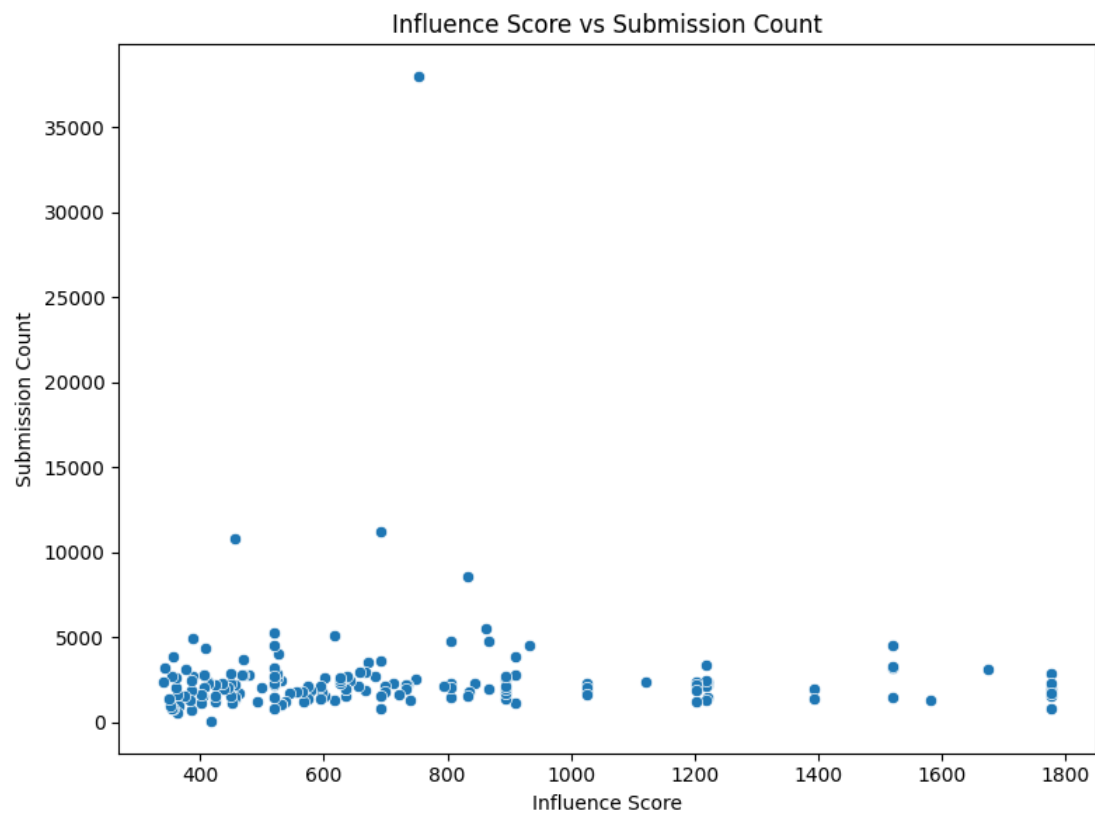
```

4. 分析不同类型的活动的分布情况，了解哪种类型的活动最常见，绘制饼图。



PUSH 占比最高，达到 31.8%。这表明代码推送是用户最频繁的操作之一，反映了 GitHub 作为一个代码托管平台的核心功能。前五名分别是 Push、PullRequest、IssueComment、PullRequestReview、Create。

5、分析用户影响力得分与提交次数的关系，绘制散点图。



影响力分数与提交数之间的相关性：0.009447。可见两者之间并无主要关联。这反映出仅仅提交是无法获得较高的影响力分数的，也许还要有浏览量、关注数量、和其它用户互动数量等等。