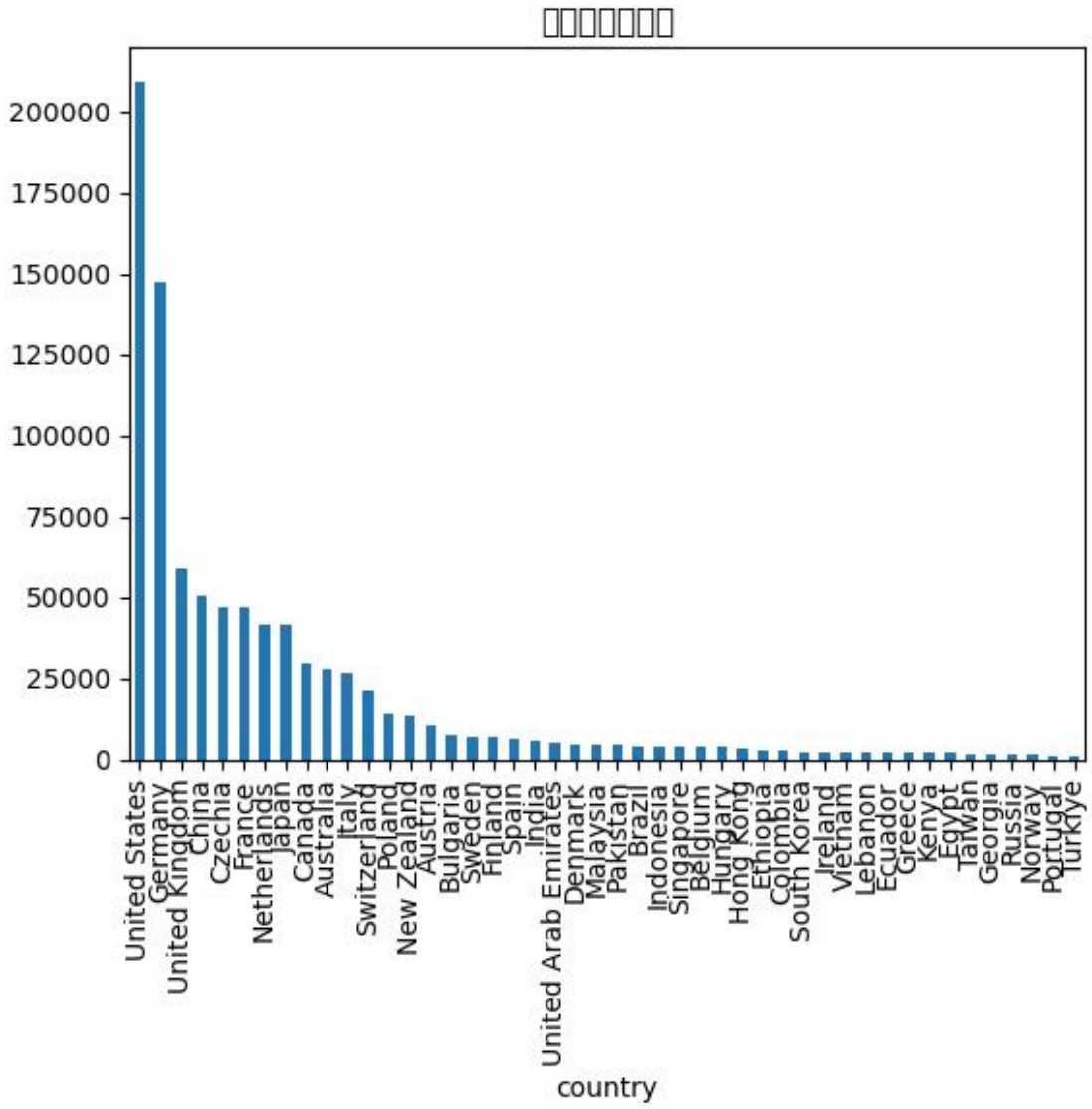
数据洞察分析

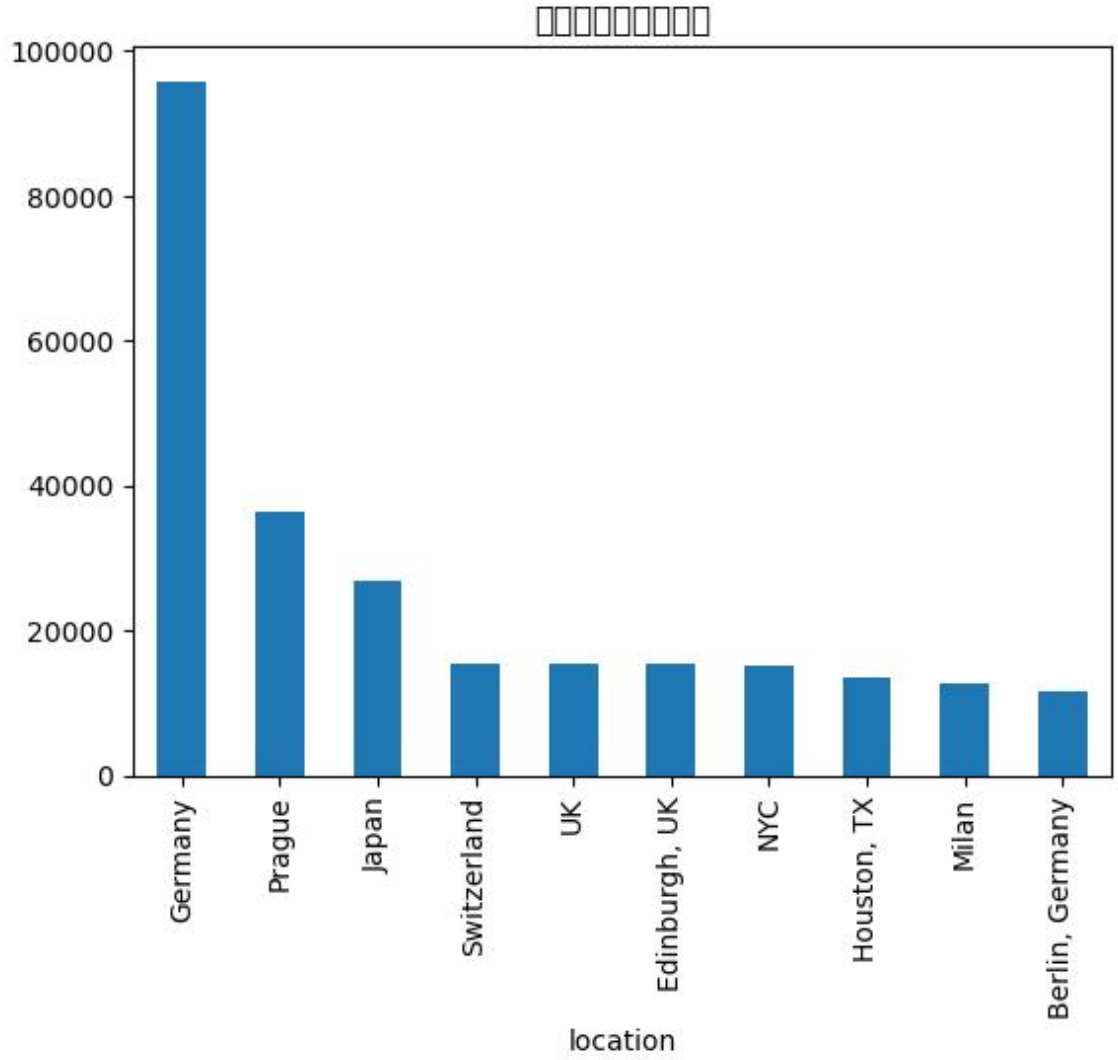
1.国家分布：



1.头部国家贡献显著：图中显示 United States（美国）的用户提交数量遥遥领先，是其他国家的几倍以上。United Kingdom（英国）、Canada（加拿大）、Germany（德国）等也贡献了大量提交。这些国家通常是技术发展较为成熟的地区，表明它们对全球技术生态的核心贡献。

2.长尾效应：除了前几个主要国家，其他国家的提交量明显下降，呈现典型的长尾分布。这表明技术生态的贡献者虽然高度集中在少数国家，但也有来自全球许多国家的开发者贡献。

2.城市与地区分布：



1. 头部地区：Germany 的提交数量远超其他地区。这表明德国是一个重要的技术热点。

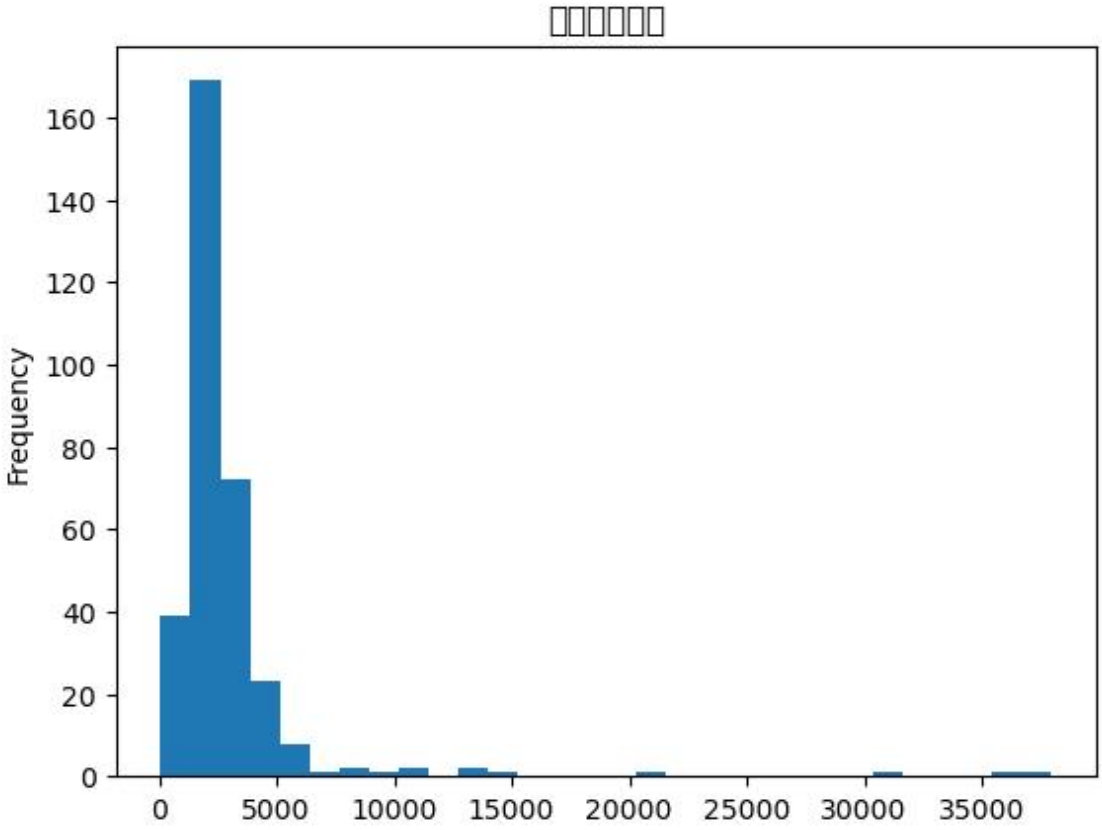
Prague 和 Japan 的贡献数量也较高，这些地方可能是技术发展活跃的地区。

Edinburgh, UK 和 NYC 等城市的贡献较多，表明这些城市在技术生态中的重要性。

2.长尾城市：除了头部几个城市，其余城市的提交量较低，但仍然有一定的技术贡献。

长尾部分表明技术热点城市之外的地区也有分散的开发者活动。

3. 识别大多数用户的提交行为



数据分布分析：

中位数为 2234：表示一半的用户提交次数在 2234 次以下，另一半用户提交次数在 2234 次以上。

25% 分位数为 1726.75，75% 分位数为 2986.25：

25% 的用户提交次数少于约 1727 次，75% 的用户提交次数少于约 2986 次。

数据的中间 50% 用户提交次数集中在 1727 到 2986 次 之间。

数据解读：

1.用户提交行为分布较集中：中间 50% 的用户提交次数范围（1727 到 2986）并不太宽，说明大多数用户的提交次数差异不大。低于 1727 的用户可能是低活跃用户，高于 2986 的用户可以划分为高活跃用户。

2.长尾效应的存在：从直方图可以看到，部分用户的提交次数远远高于中位数（比如 10,000+ 的极值）。这些极值会显著影响整体的均值，但对中位数影响较小。

因此，根据分位数，将用户划分为以下三类：

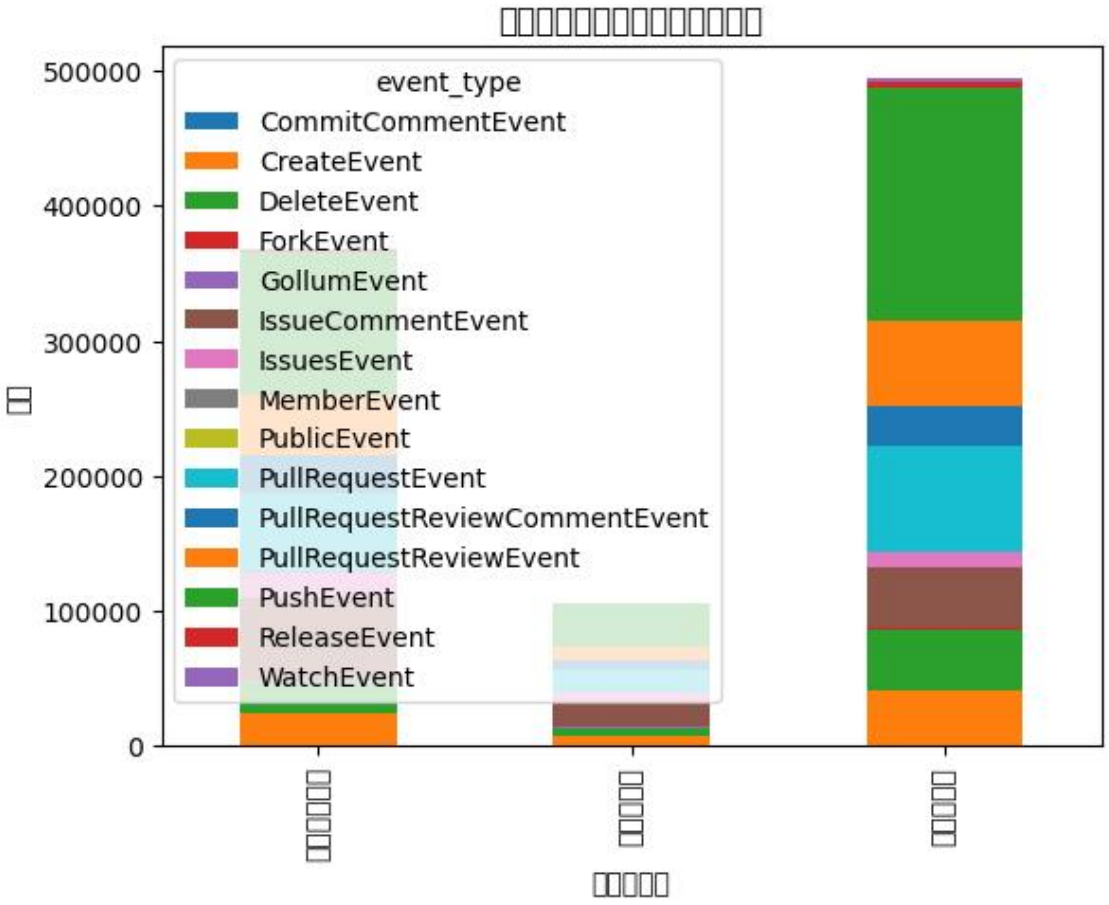
低活跃用户：提交次数 ⩽ 1727

中等活跃用户：提交次数在 1727 - 2986 之间

高活跃用户：提交次数 > 2986

3.影响力与提交行为的洞察分析

洞察 1：提交类型与用户行为



通过分析不同活跃度用户的提交类型分布，可以观察到：高活跃用户主要集中在代码相关活动（如 PushEvent、PullRequestEvent），表明他们更多参与项目核心开发。低活跃用户则更倾向于外围活动（如 ForkEvent、WatchEvent），说明他们可能是项目的观察者或初步贡献者。

进一步的细节：

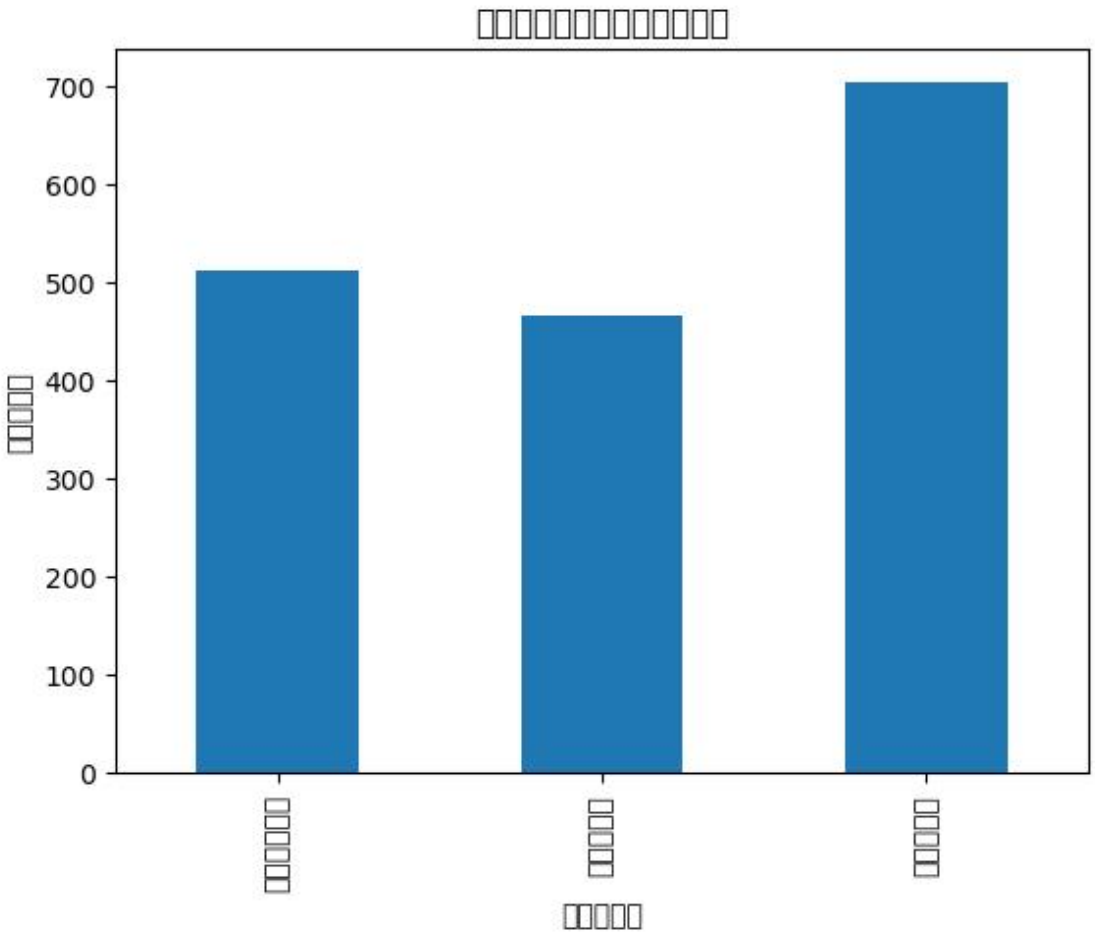
- 高活跃用户：显著偏向 PushEvent 和 PullRequestEvent，表明对代码提交和审核活动的主导性。

- 中等活跃用户：行为类型多样，可能兼具外围贡献和核心开发活动。

- 低活跃用户：以 WatchEvent、ForkEvent 为主，更多参与旁观或初步尝试的行为。

通过对不同活跃度的提交类型分布进行分析，可以帮助更好地理解用户的参与行为及其对项目的贡献模式。

洞察 2：影响力与提交频率的关系



从数据分析可以看出，提交频率与影响力之间存在一定的弱相关性（相关系数为 0.23）。这表明提交次数的增加可能会对用户的影响力产生一定的提升作用，但这种关系并不强。

从图表中可以看出：

- 高活跃用户：平均影响力显著高于中等和低活跃用户，说明频繁提交确实能提升影响力。

- 中等活跃和低活跃用户：影响力之间的差异较小，表明提交频率不足时影响力提升空间有限。

此外，影响力可能还受到其他因素的影响，例如提交的质量、活动类型（如 PullRequestEvent 的贡献度可能高于 WatchEvent）、以及用户参与的项目是否具有高人气等。