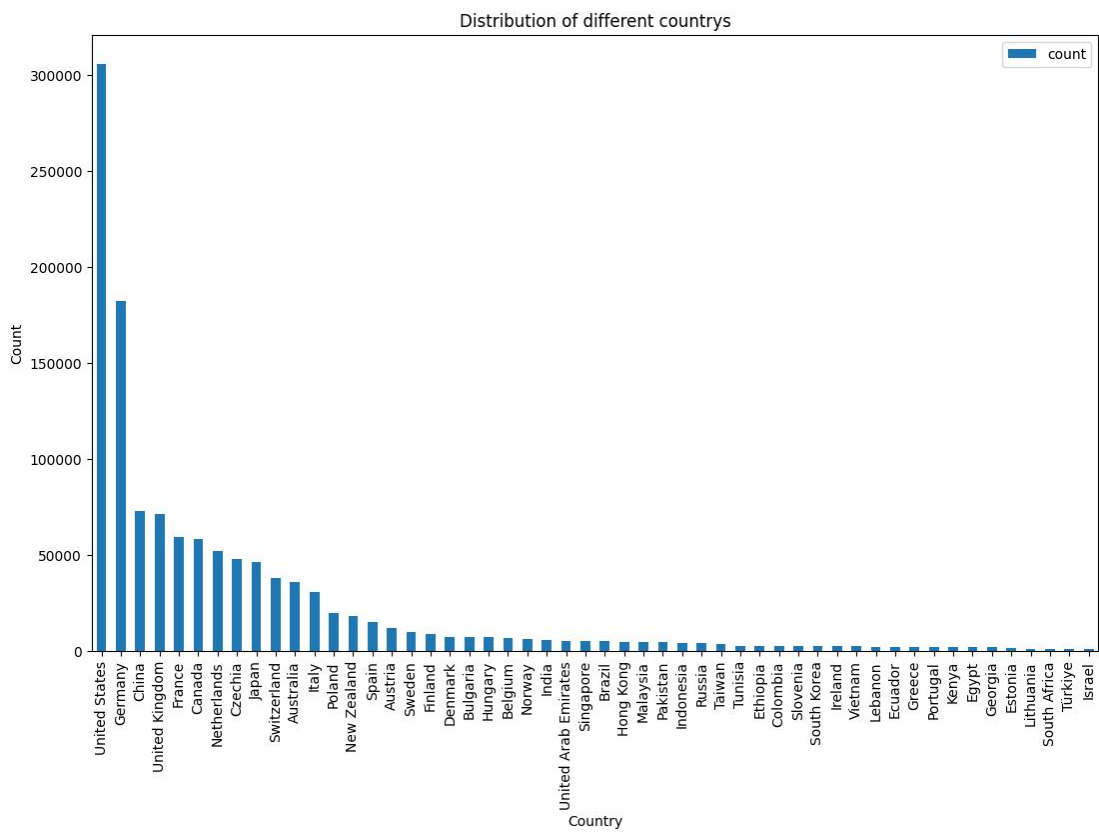


利用 GPT 大型模型工具完成数据洞察

1.人口统计分析

1.1 国家和地区分布：统计用户所在国家和地区的分布，识别主要的开发者集中地。



美国的数据数量在图中是最高的，接近 300000。这表明在该统计数据中，与美国相关的数据量占据了显著的优势。这种高数量可能反映了美国在该统计所涉及领域中的重要性或活跃度。

德国（Germany）的数据数量大约在 150000 左右，位居第二。虽然远低于美国，但仍然是一个相当高的数量，显示出德国在该领域也有一定的影响力。

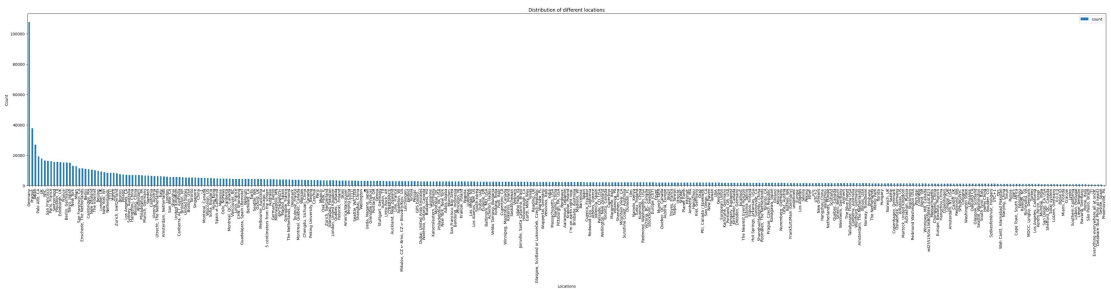
中国（China）、加拿大（Canada）、荷兰（Netherlands）、

澳大利亚（Australia）、新西兰（New Zealand）等国家的数据数量依次递减。这些国家的数据量处于中等水平，表明它们在该统计领域有一定的参与度，但不如美国 and 德国突出。

一些国家的数据数量相对较少。例如，印度（India）、印度尼西亚（Indonesia）、俄罗斯（Russia）等国家的数据量明显低于上述国家。这可能暗示这些国家在该统计所涉及的特定领域中的参与度较低，或者该统计的数据来源存在一定的偏向性。

埃塞俄比亚（Ethiopia）、南非（South Africa）、以色列（Israel）等国家的数据数量非常少，几乎接近 0。这表明这些国家在该统计所涉及的领域中参与极少，或者数据收集过程中这些国家的数据较难获取。

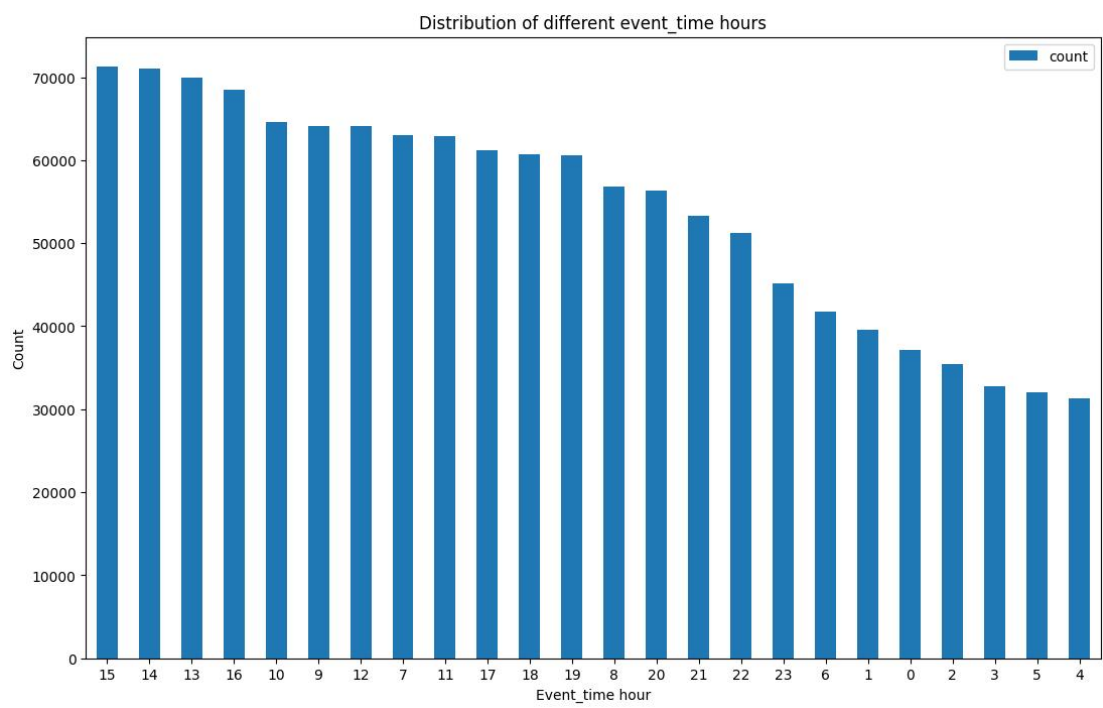
1.2 城市级别分布：分析主要城市的开发者密度，发现技术热点区域。



高数量地点。在图的左侧，可以看到有几个地点的数据数量较高。例如，靠近纵轴起点的一些地点，其柱状条明显较高，表明这些地点的数据量较大。这可能意味着这些地点在该统计所涉及的领域中具有重要地位或较高的活跃度。

低数量地点。随着横轴从左向右移动，大部分地点的柱状条高度逐渐降低，表明这些地点的数据数量较少。特别是在图的右侧，大多数地点的数据量都非常低，接近 0。这可能暗示这些地点在该统计所涉及的特定领域中的参与度较低，或者数据收集过程中这些地点的数据较难获取。

1.3 时区分布:了解用户的时区分布，分析不同地区用户的协作时间模式。



在图的左侧，可以看到 15 点（3 PM）的数据数量最高，接近 70000。这表明在该统计中，事件在 15 点发生的频率最高。14 点（2 PM）和 16 点（4 PM）的数据数量也非常高，仅次于 15 点。这表明在下午 2 点到 4 点之间是事件发生的高峰期。

在图的右侧，4 点（4 AM）的数据数量最低，接近 30000。

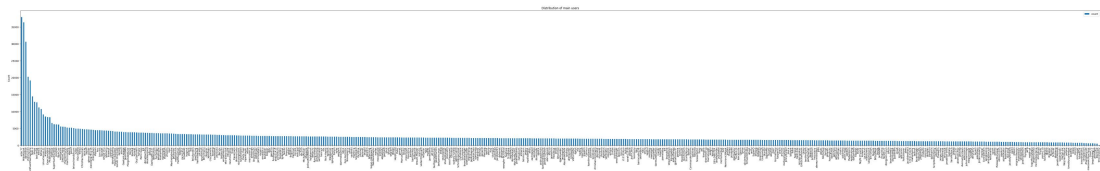
这表明在凌晨 4 点事件发生的频率最低。0 点（12 AM）、1 点（1 AM）、2 点（2 AM）和 3 点（3 AM）的数据数量也相对较低，表明凌晨时段是事件发生的低谷期。

从图中可以观察到一个明显的趋势：事件发生频率从下午的高峰期逐渐下降，经过傍晚和夜间，到凌晨达到最低点，然后在上午逐渐回升。

如果基于白天工作的假设，那么欧亚洲总量大，美洲单国家量大。

2.协作行为分析

2.1 提交频率：统计每个用户的提交次数，识别高活跃用户和低活跃用户。



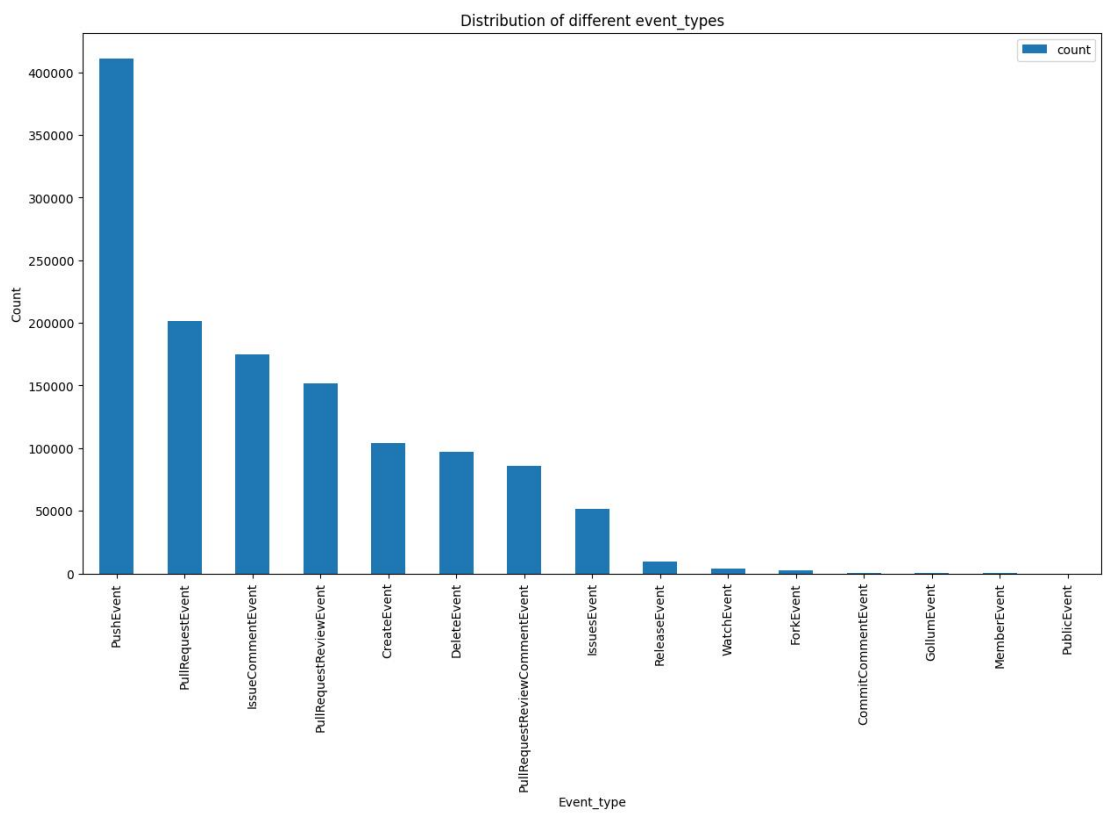
高数量用户。在图的左侧，可以看到有几个用户的数据数量较高。例如，靠近纵轴起点的一些用户，其柱状条明显较高，表明这些用户的数据量较大。这可能意味着这些用户在该统计所涉及的领域中具有重要地位或较高的活跃度。

低数量用户。随着横轴从左向右移动，大部分用户的柱状条高度逐渐降低，表明这些用户的数据数量较少。特别是在图的右侧，大多数用户的数据量都非常低，接近 0。这可能暗示这些用户在该统计所涉及的特定领域中的参与度较

低，或者数据收集过程中这些用户的数据较难获取。

3.其他维度有趣的洞察（至少 2 个）

3.1 日志数据事件类型统计



高数量事件类型。在图的左侧，可以看到“PushEvent”事件类型的数据数量最高，接近 400000。这表明在该统计中，“PushEvent”事件发生的频率远远高于其他事件类型。“PullRequestEvent”和“IssueCommentEvent”事件类型的数据数量也较高，分别在 200000 左右和 150000 左右。这些事件类型在该统计中也占有重要的比例。

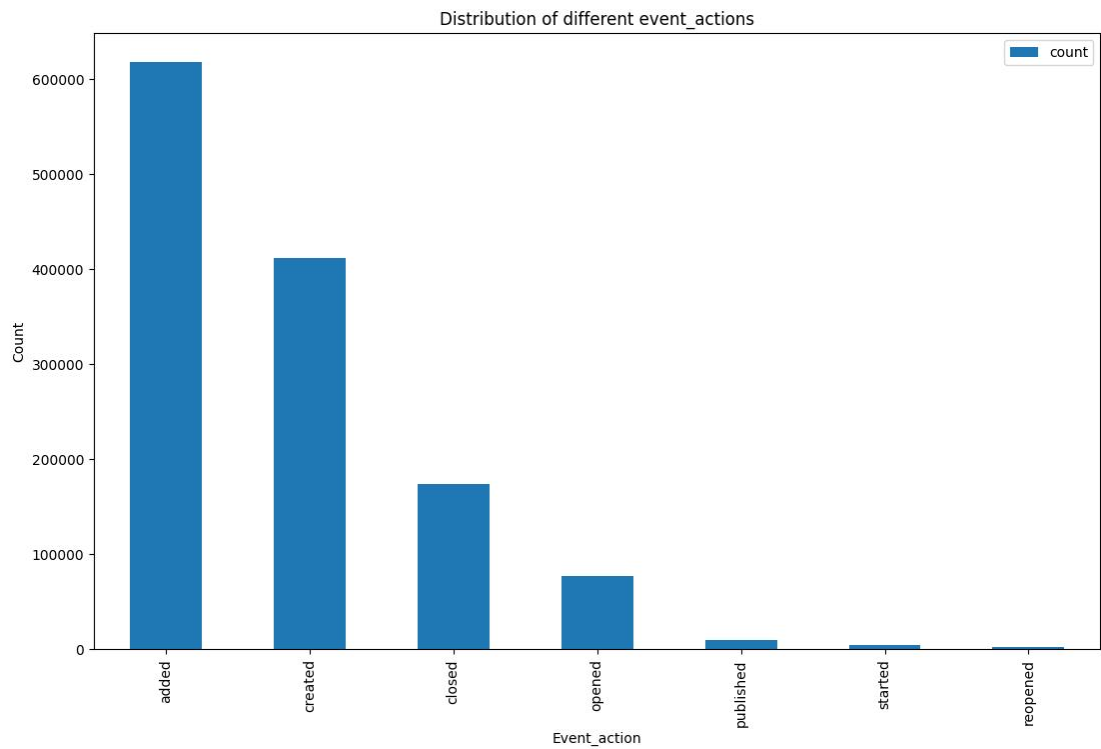
低数量事件类型。在图的右侧，“GollumEvent”、“MemberEvent”和“PublicEvent”等事件类型的数据数量非

常低，几乎接近 0。这表明这些事件类型在该统计所涉及的特定领域中的发生频率极低。

其他事件类型，如“CreateEvent”、“DeleteEvent”、“ForkEvent”等，其数据数量处于中等水平，介于 10000 到 100000 之间。这些事件类型在该统计中也有一定的出现频率，但不如“PushEvent”等主要事件类型频繁。

用户行为模式潜在意义。这是关于用户行为的数据，根据用户在某个应用上的操作类型分布，可以推测用户更倾向于进行“PushEvent”等主要操作，而较少进行“GollumEvent”等低频操作。这可以帮助应用开发者优化操作界面或功能。

3.2 日志数据的事件操作



在图的左侧，“added”事件动作的数据数量最高，接近

600000。这表明在该统计中，“added”事件动作发生的频率远远高于其他事件动作。这可能意味着在该统计所涉及的领域中，添加（added）操作是最常见的动作。

“created”事件动作的数据数量次之，大约在 400000 左右。这表明创建（created）操作也是一个较为常见的动作。中等数量事件动作。

“closed”事件动作的数据数量在 200000 左右，属于中等水平。这表明关闭（closed）操作在该统计所涉及的领域中也有一定的发生频率。

“opened”、“published”、“started”和“reopened”事件动作的数据数量非常低，几乎接近 0。这表明这些事件动作在该统计所涉及的特定领域中的发生频率极低。

这是关于用户在某个应用上的操作类型的数据，那么可以推测用户更倾向于进行“added”和“created”操作，而较少进行“opened”、“published”、“started”和“reopened”操作。这可以帮助应用开发者优化操作界面或功能，重点关注用户频繁使用的操作。

4. 总结

本文主要通过了 AI 工具进行了图像数据的综合分析。作者还借用了 pandas_profiling 库对整个 csv 文件进行了数据分析，也可以视作是某种形式的 AI 工具分析。