

GitHub500 名用户协作行为日志数据分析

一、人口统计分析

1.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。

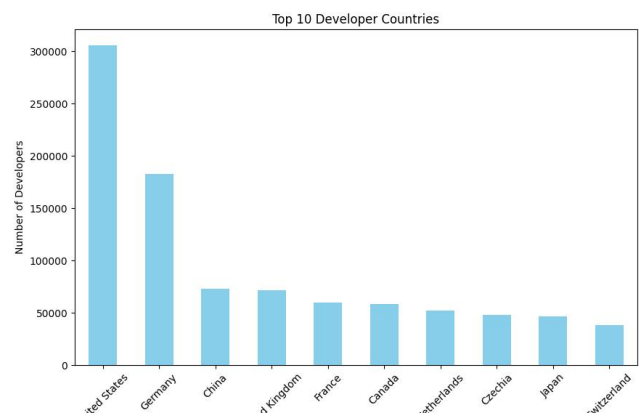


图 1

美国的开发者数量遥遥领先，超过 300,000 人，是排名第二的德国的两倍以上。

德国位居第二，开发者数量超过 150,000 人。

中国、英国、法国、加拿大的开发者数量相对接近，介于 50,000 到 100,000 人之间。

荷兰、捷克、日本和瑞士的开发者数量相对较少，大约在 50,000 人以下。

美国的开发者数量明显占据全球首位。欧洲国家如德国、英国、法国、荷兰和捷克显示出较强的开发者聚集，说明这些国家可能是技术行业的重要组成部分。

2.分析主要城市的开发者密度，发现技术热点区域。

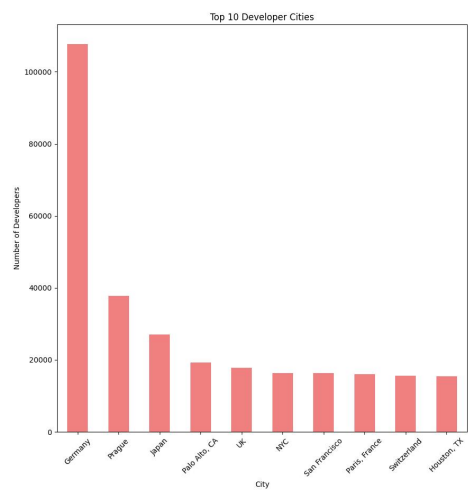


图 2

此图显示了开发者人数最多的前 10 个城市或国家。（由于数据集并没有严格的全部显示城市）采用 `location` 作为参考。德国明显领先于其他地区，开发者数量超过 10 万，表明德国可能拥有发达的技术产业和较高的开发者集中度。相比之下，其他城市如布拉格、日本和美国的技术中心（例如硅谷和纽约）开发者数量分布相对均匀。

3.分析不同地区用户的协作时间模式

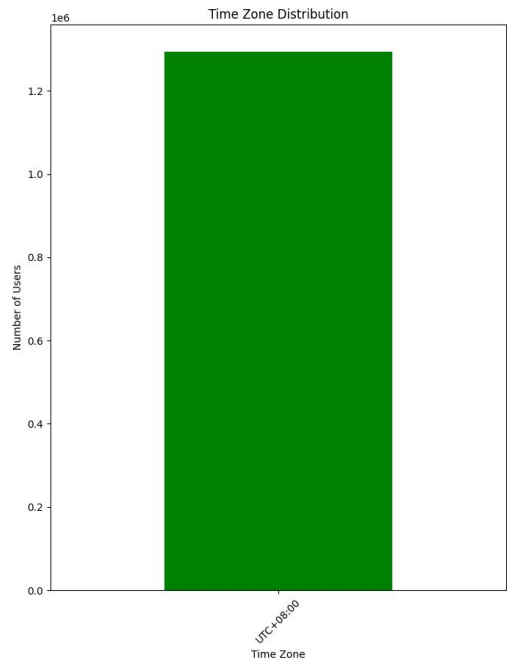


图 3

图表展示了特定时区（UTC+08:00）的用户数量，人数超过 120 万。这可能反映出该时区（覆盖中国等东亚地区）在全球开发者或用户中的重要性。

二、协作行为分析

高活跃用户（前 10 名）：

	user_id	push_count
0	11146458	35993
1	158862	14460
2	28706372	9160
3	8188402	8273
4	40306929	7836
5	43724913	4985
6	50149701	4617
7	1580956	4494
8	140370	4118
9	30802565	3885

低活跃用户（前 10 名）：

	user_id	push_count
493	814283	2
492	985347	5
491	19261257	8
490	19796701	35
489	3162115	39
488	2223172	93
487	7165771	118
486	17899797	123
485	16256802	124
484	1133777	125

三、补充维度洞察

1.Github 使用者最青睐的技术分析

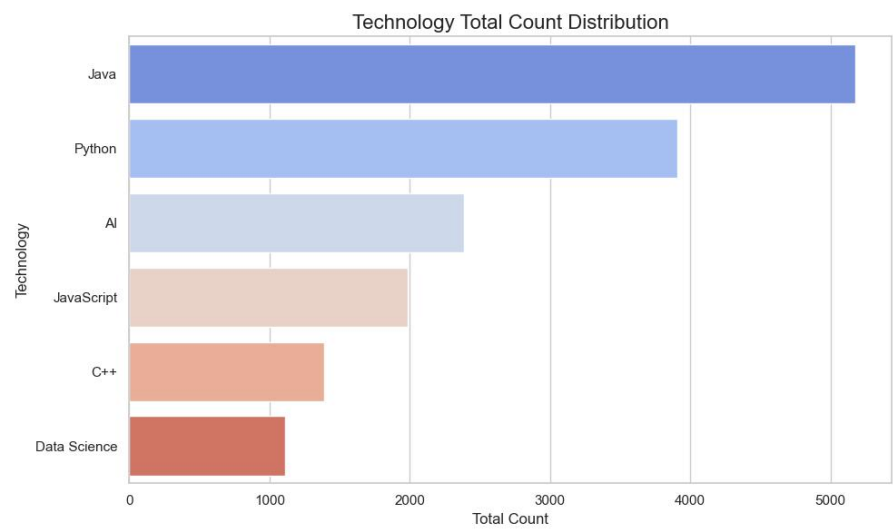


图 4

从图中可以看出，Java 和 Python 是最受欢迎的技术，开发者的使用数量显著高于其他技术，尤其是数据科学、C++和 JavaScript。这反映了编程语言和技术方向在当前行业

2.GitHub 中各公司开发着数量排名

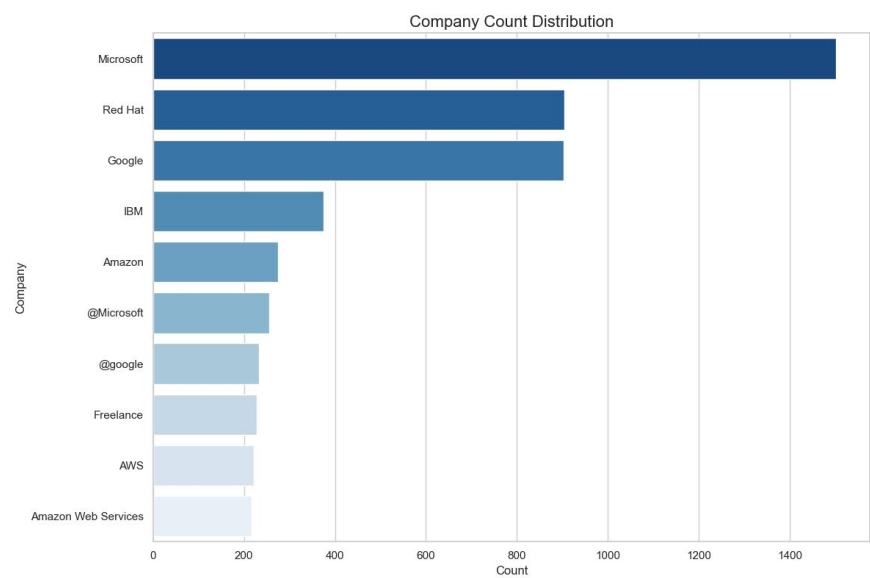


图 5

Microsoft、Red Hat 和 Google 是开发者数量最多的公司，这表明这些公司可能在技术社区中具有较高的吸引力和影响力。同时，也显示了自由职业者和云计算相关公司（如 AWS）在开发者中的受欢迎程度。