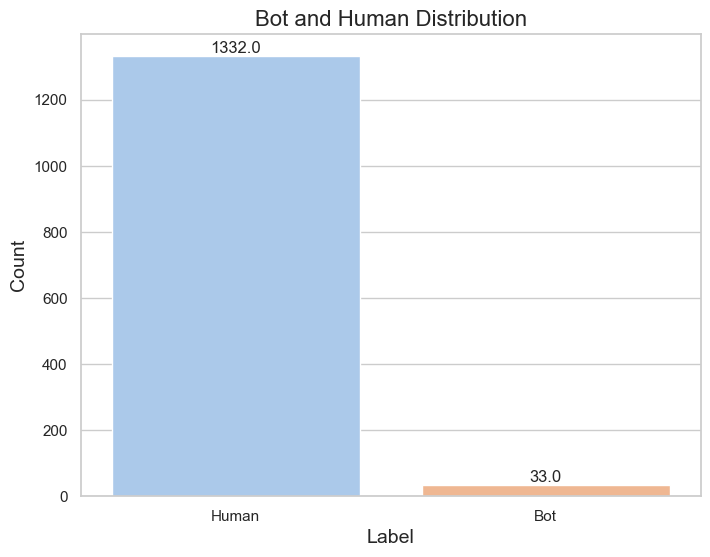
分析报告

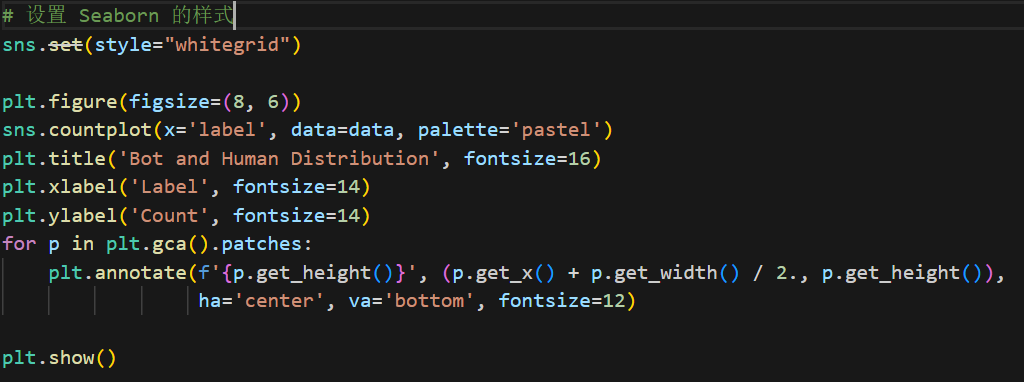
经过数据缺失值处理、数据类型转换、重复数据删除等操作，我们将原有的（20358,34）条数据，修剪为（1365，33）条有效数据。以下是数据可视化结果分析。

4.1：可视化bot和human类型的情况

图表选择：考虑到label字段的类型较少（只有bot和human两种类型），若使用条形图，其每一类的数据大小可以通过条形长度直观比较，清晰地展示了绝对数量，比较合适。

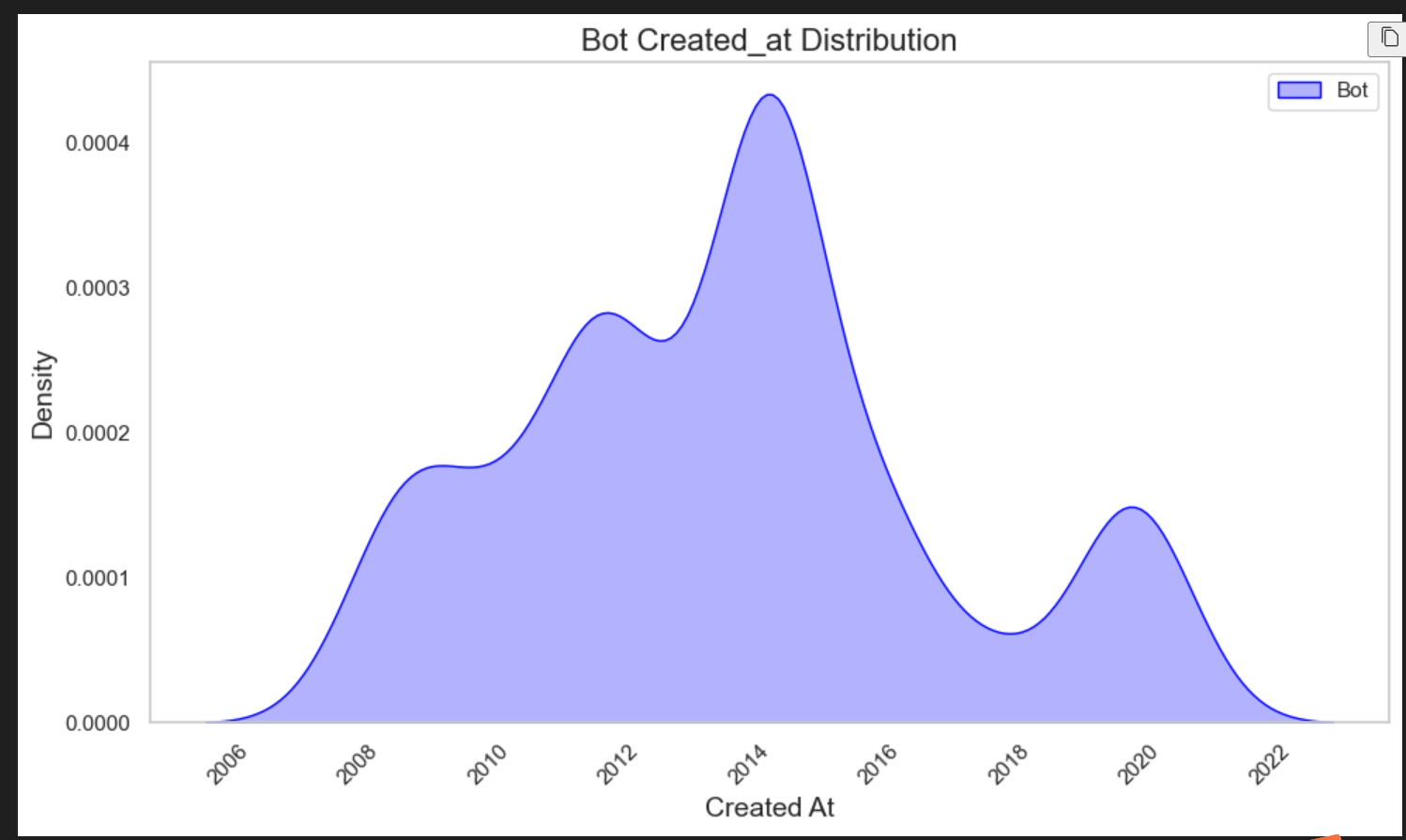
结果分析：Human数量显著高于Bot，结果显示该数据集以人类用户为主

数据洞察：为维护GitHub的良好生态，应该鼓励人类用户构成社区。如果设置Bot数目过于庞大，可能在issue中造成太多噪音，不利于交流与实际应用。人类用户数量远高于bot数量也反映出人类用户往往在代码质量以及项目协作中，仍然起着关键性作用，体现出人类创造力在开源社区的重要价值。



4.2：可视化bot类型账号的created\_at情况

图表选择：选择核密度估计（KDE）曲线图，很好地展示了不同时间段Bot用户的创建情况，从而在数据量较大的情况下，识别出活跃期与低谷期。并且由于Bot用户可能在整个时间段内分布不均，密度曲线能够将这些数据点整合为一个连续的曲线，更容易观察到时间上的变化趋势。

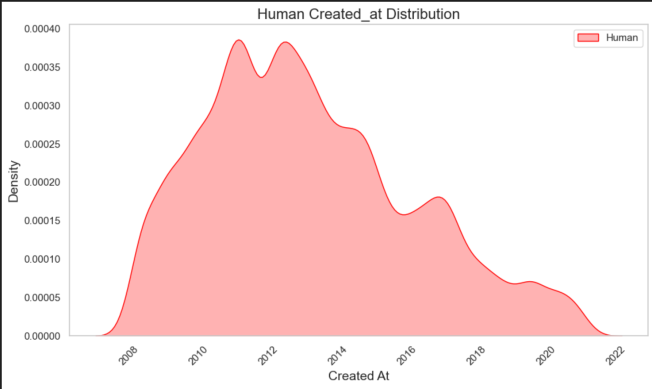
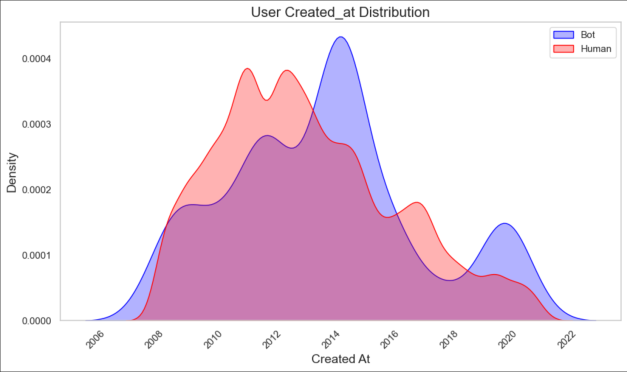


结果分析：观察已有数据可以估计，Bot用户创建在2014-2016年间达到高峰。在2010年之前，2020年之后处于低谷期。其中2020年出现了小高峰期，2018年处于小低谷期。2006-2014年间bot用户创建基本处于增长状态。

数据洞察：2014年前bot用户创建的增长趋势，反映了自动化工具的普及使用。而其在2014-2016年间达到高峰，可能表明当时的开源项目有大量的机器人用户参与。而在2016年之后bot用户数量呈下降趋势，可能与GitHub政策变动有关，开源项目兴起、开发流行框架推出，使得人类用户在开源项目中的贡献度远远高于bot用户，并且也反映出GitHub构建的良好生态，即人们更愿意参与到项目贡献中，发挥了人类的贡献质量及创新能力。而2020年出现的“小高峰”也许与bot参与的特定项目有关，这些因素直接影响了bot的创建行为。



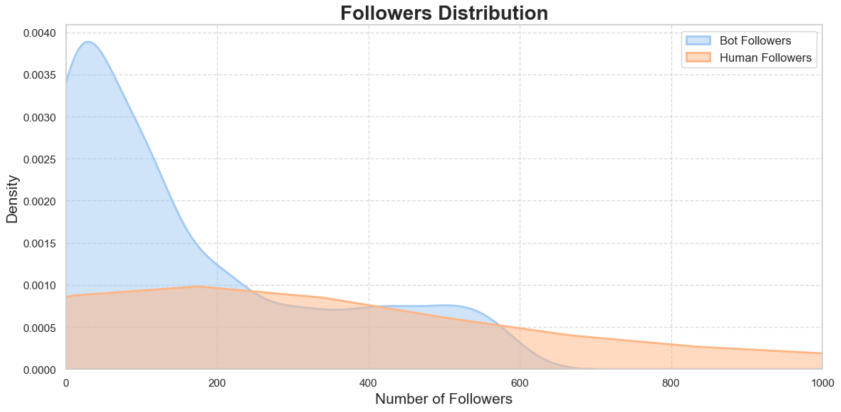
4.3：可视化human类型账号的created\_at情况

图表选择：选择核密度估计（KDE）曲线图，理由同上。最后我们希望将两曲线整合，以对比bot和human的创建时间分布情况。

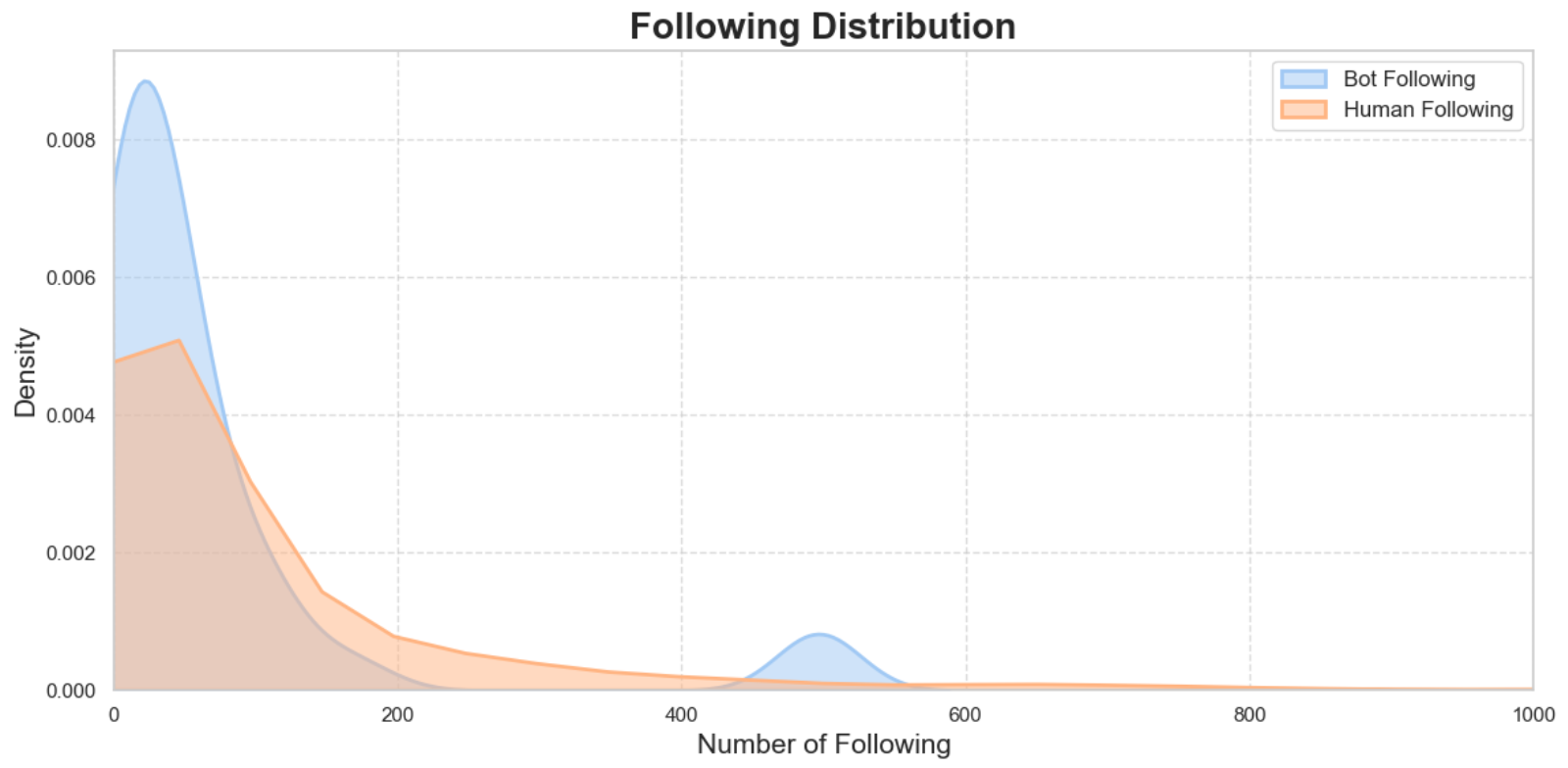
结果分析：人类用户创建在2010-2014年间达到高峰，总体来说用户注册存在长期趋势。但是由于人类用户注册总体基数大于bot，意味着人类仍然还是更占优势的平台用户。而大量自动化注册，更有可能会影响平台的生态环境。由图可知GitHub一直维护着较好的开源生态。

数据洞察：GitHub是一个主要面向开发人员和开源项目的协作平台，人类用户通常有明确的需求，这驱动了他们的注册和活跃度。而问题跟踪、代码审查等机制需要人类用户的真实互动，Bot用户常常不具备这样的需求。GitHub的生态系统建立在开源文化之上，许多开发者愿意共享他们的代码和项目，这种文化吸引了大量的人类用户。相比之下，Bot 用户主要用于自动抓取数据或执行特定任务，缺乏参与开源项目的动力。

4.4：可视化bot类型账号的followers和following情况

选用图表：核密度估计（KDE）图，识别Bot用户的活跃程度，可以直观发现其不同followers和following的分布比例，并且发现其高峰、低谷。并且其便于在同一图表中同时展示Bot和Human用户的分布，从而进行对比分析。

结果分析： 其大部分Followers与Following数量都很少（集中在0-200区间内）。Followers数量有部分处于400-600间，但是几乎没有超过600的了。



数据洞察：Bot用户通常是自动化脚本，其目的是数据采集、自动回复等，因此他们不太会关注其他用户并且积累粉丝。他们缺乏人类用户的社交行为，因此不会有大量关注者和粉丝。并且平台严格监控Bot用户行为，防止进行恶意活动。总之Bot用户存在的意义，也许不是社交，而是为了提高自动化效率，获取信息。

4.5可视化bot类型账号的followers和following情况

选用图表：核密度估计（KDE）图，理由同上（由于两者同时分析，所以用的也是上图）。

结果分析：多数人类的Following集中在0-200区间内，相比之下Followers的分布较为集中。对比bot与人类的分布图，可见绝大部分bot的数据都集中在0-200区间内，分布极不平均。而人类数据分布较为分散，这可能与数据量大小有关。

数据洞察：人类用户在社交媒体上寻求与他人的连接以及获取相关信息，因此他们会主动关注感兴趣的主题、用户或群体。这种目标驱动的行为导致他们的关注和粉丝量更大。社交平台常常通过各种激励措施（如推荐、算法推送等）促进用户关注和互动，这对于人类用户尤其有效。平台希望增强用户间的互动，从而提高活跃度和用户粘性。

这些差异帮助平台理解用户的社交行为和互动模式，优化算法和推荐系统，以提升用户体验。而理解了人类用户的社交动机和行为，可以帮助平台更好地构建内容生态，促进健康的社交环境和社区发展。并且通过识别人类和Bot在关注及粉丝数量的差异，便于平台识别出异常行为。