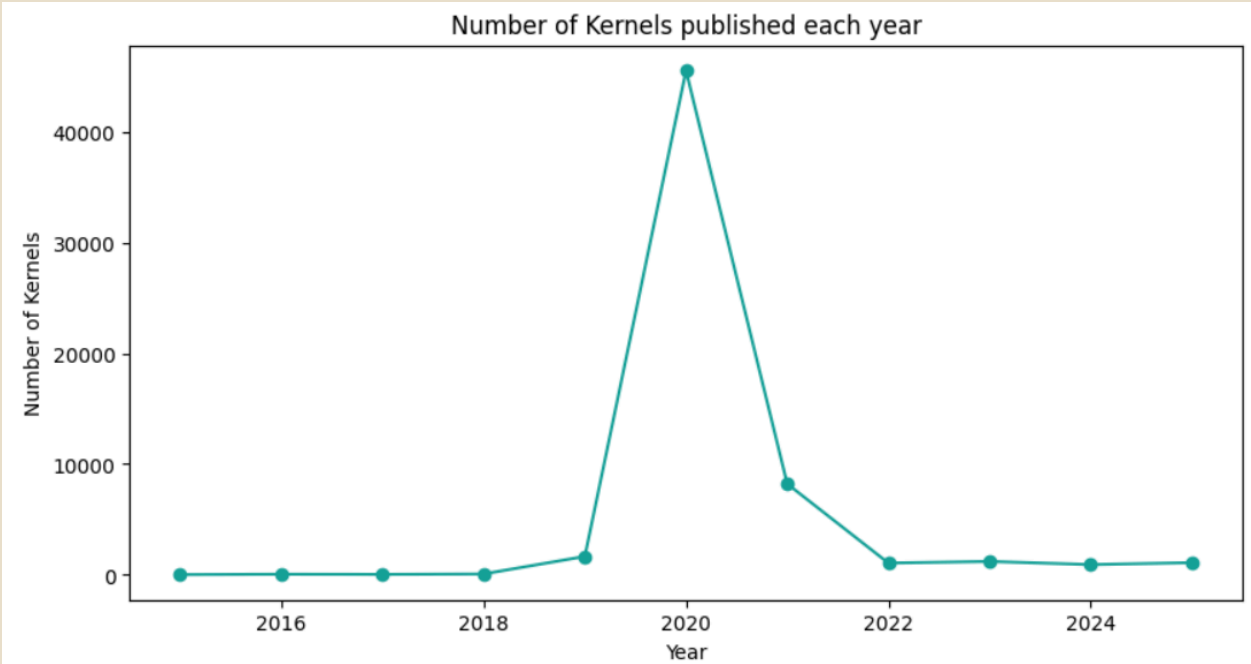
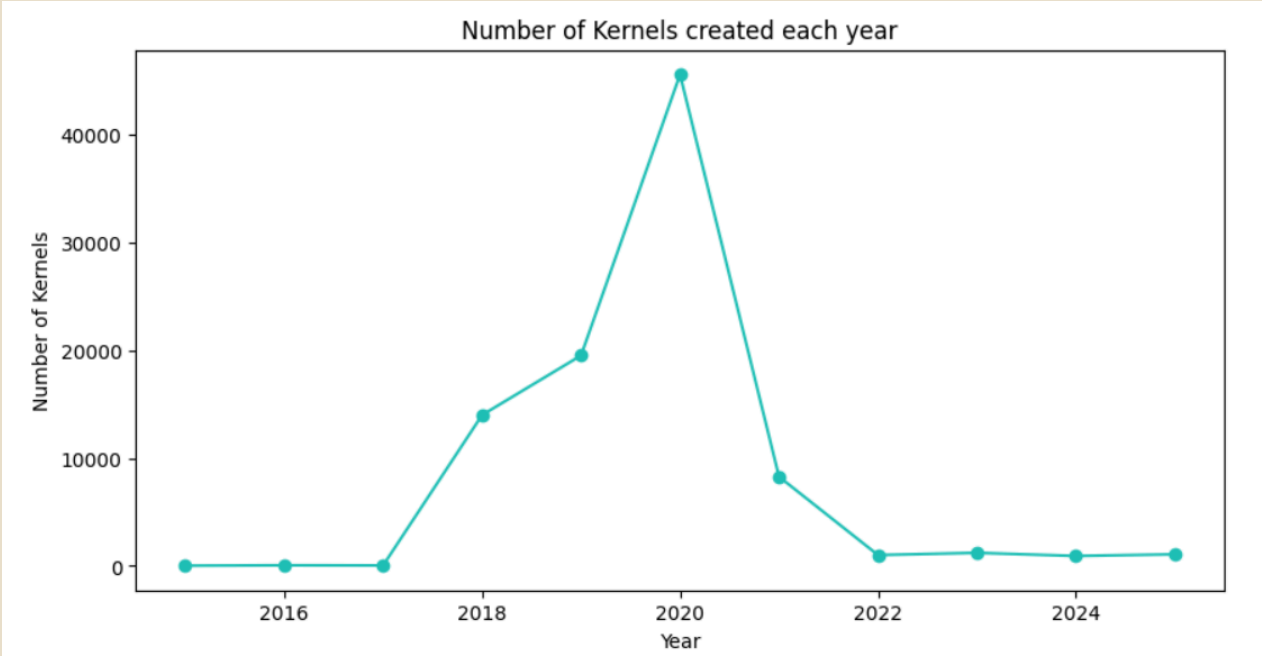


Meta Kaggle Hackathon

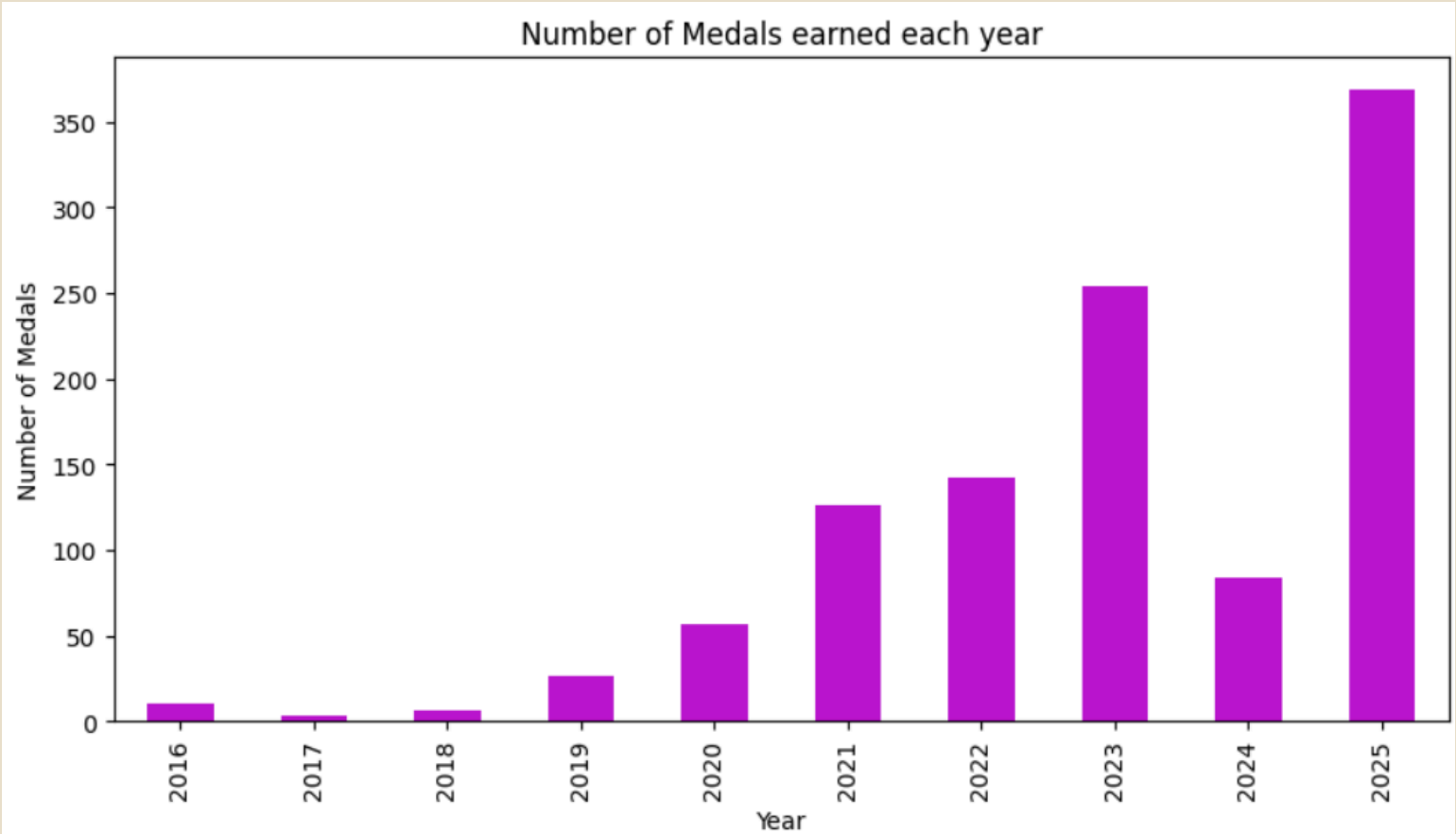
For fifteen years, Kaggle's community has majorly impacted AI and machine learning (ML) progress through Competitions. Kaggle Hackathons is inviting the community to celebrate these accomplishments in this very hackathon! Very meta. This challenge encourages participants to analyze the trends, patterns, and dynamics within this unique ecosystem. Each competition that Kaggle hosts represents a challenging problem valuable to the larger scientific community. The progress made in these competitions, as visualized by climbing leaderboards, and the vast amounts of metadata surrounding them, offers a unique opportunity to understand how the industry and community have evolved.

This analysis focuses on analyzing two types of data from the numerous Meta Kaggle datasets. The first data is kernels data which refers to the code notebooks or script that runs in Kaggle's cloud environment. There are multiple datasets for kernels thus, only important datasets were chosen to be merged and analyzed to get helpful insights. In addition, to know the changes that occurred for competitions throughout the years, the Competitions dataset was deeply analyzed to explore the interesting changes that happened in competitions in the previous years.

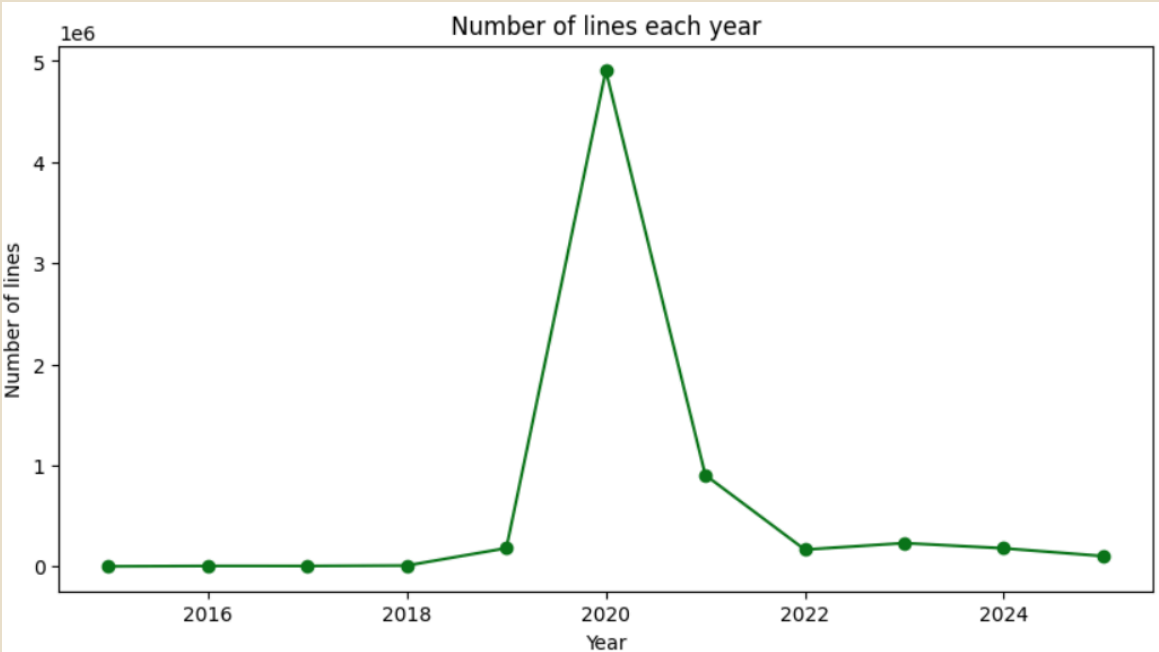
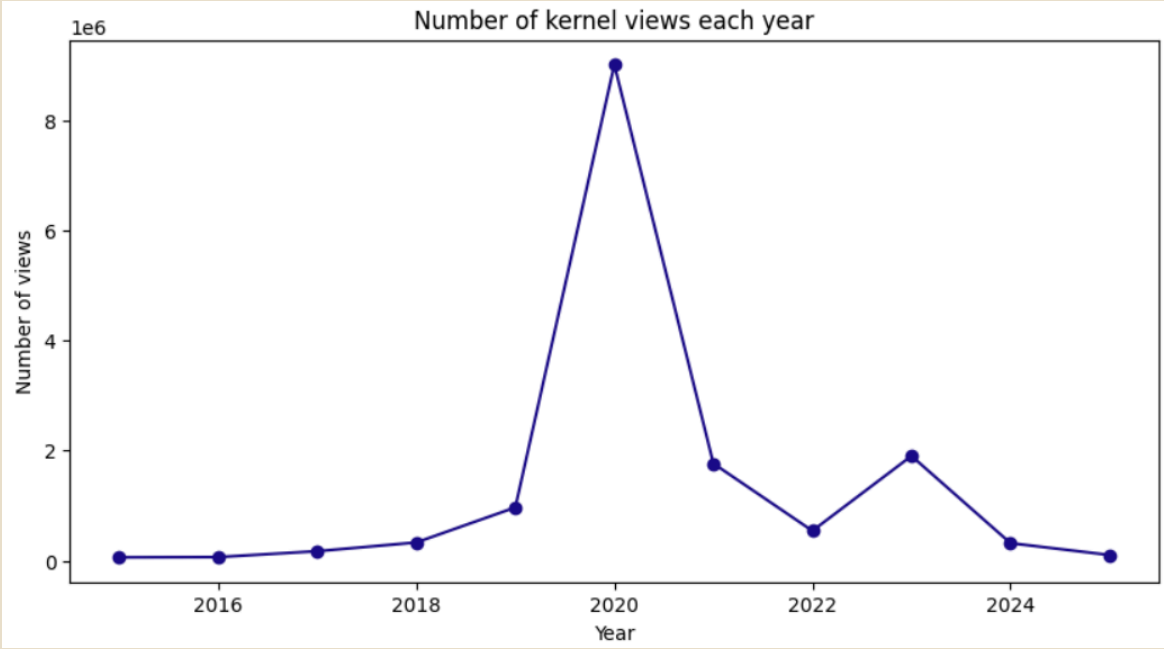
The number of created kernels (notebooks in Kaggle's cloud) is at its peak in 2020. This may be due to the quarantine during Covid-19 when people had more free time to spend on coding. Another insight is the difference between the number of created and public kernels in 2018 and 2019. The significant difference for those years in the two graphs means that many created kernels stayed private.



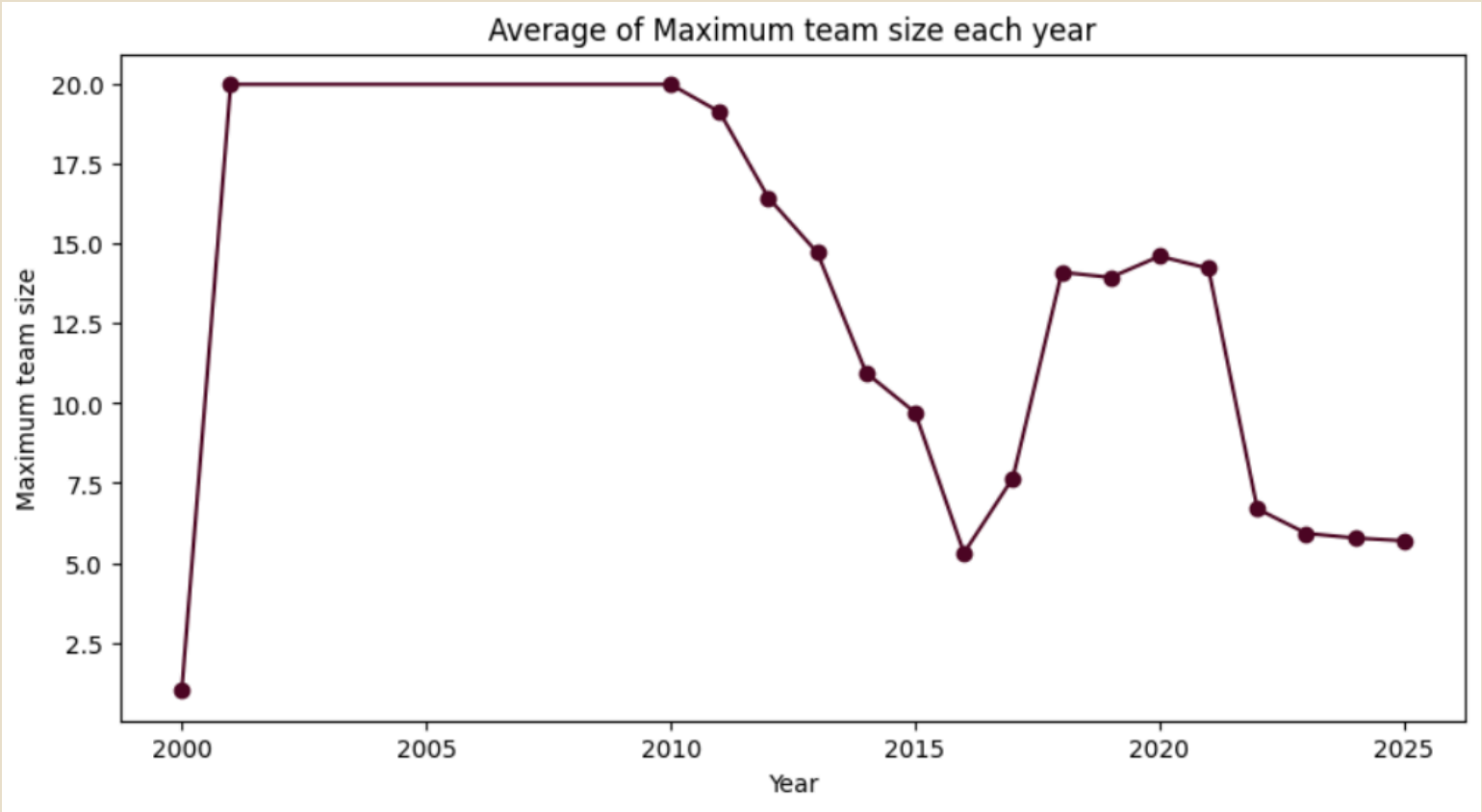
The highest number of medals earned is in 2025. It is surprising that 2020 has one of the lowest number of earned medals despite being the year that has the most created and published kernels. Therefore, this may suggest that having a high number of kernels in a specific year does not necessarily mean that more competitions or medals are offered. The kernels may be for different purposes other than competition submissions.



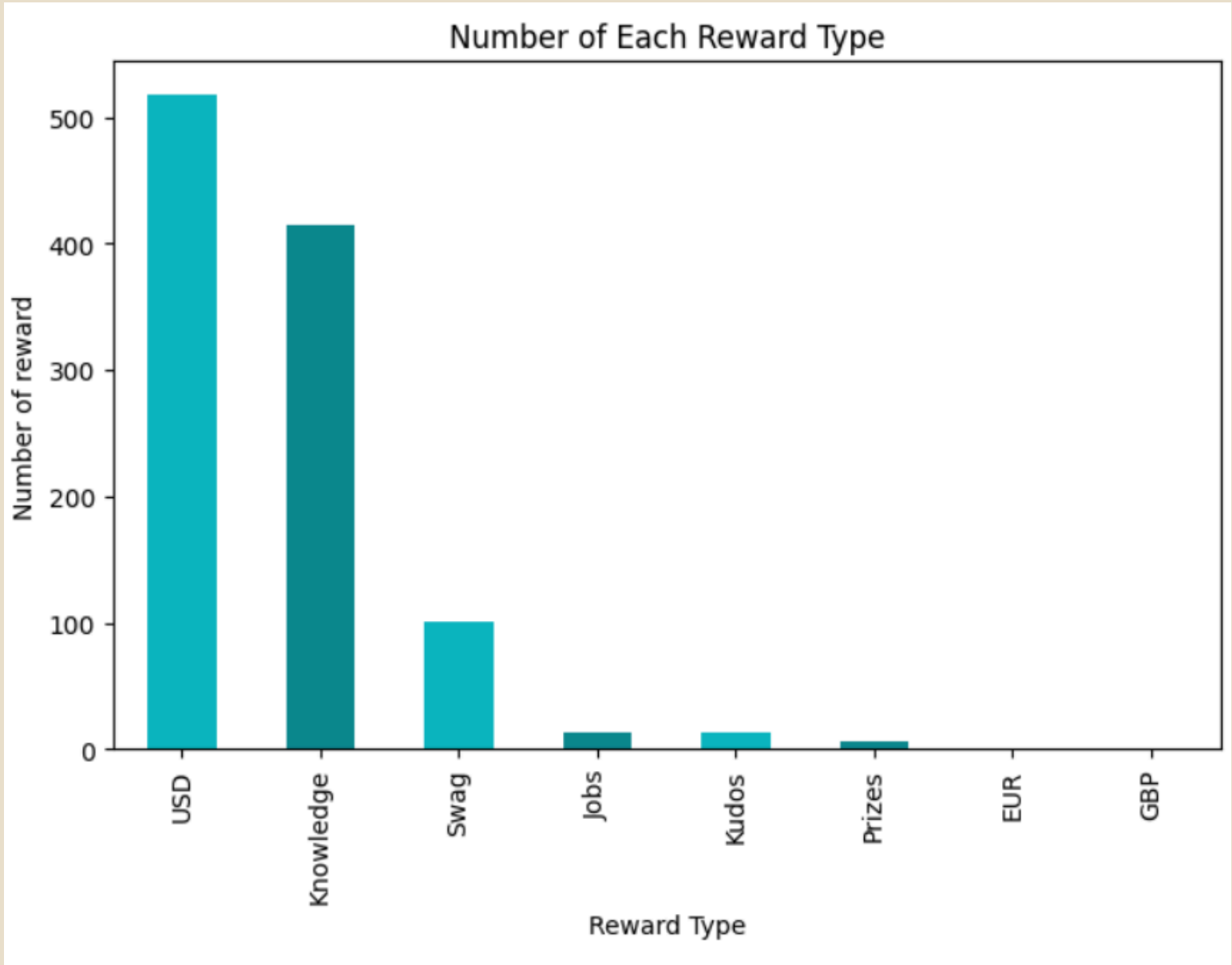
The kernel views and kernel lines graphs support the number of kernels published graph insight as they show that 2020 has the highest views number and the highest number of kernel lines. It is logical as increasing the number of kernels leads to increasing views even if each kernel has few views, the total number of views from all kernels in that year will result in a high number.



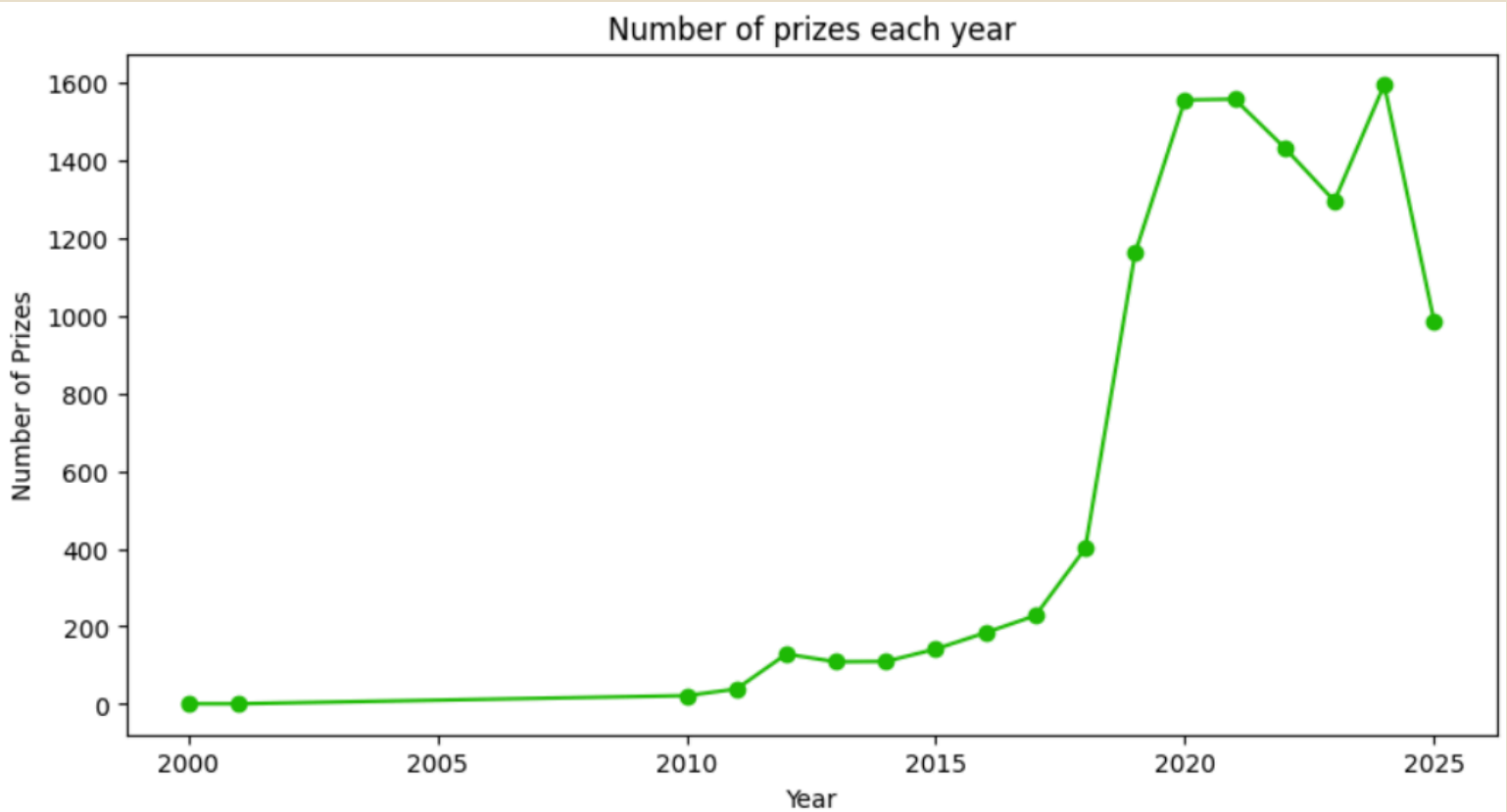
Competitions set a maximum number of members per team so, from this graph we conclude that there is no constant number of team members as each competition differs in skills required, duration, and level thus, the maximum number of team members is dependent on those factors. However, the maximum number of members tends to decrease since 2022.



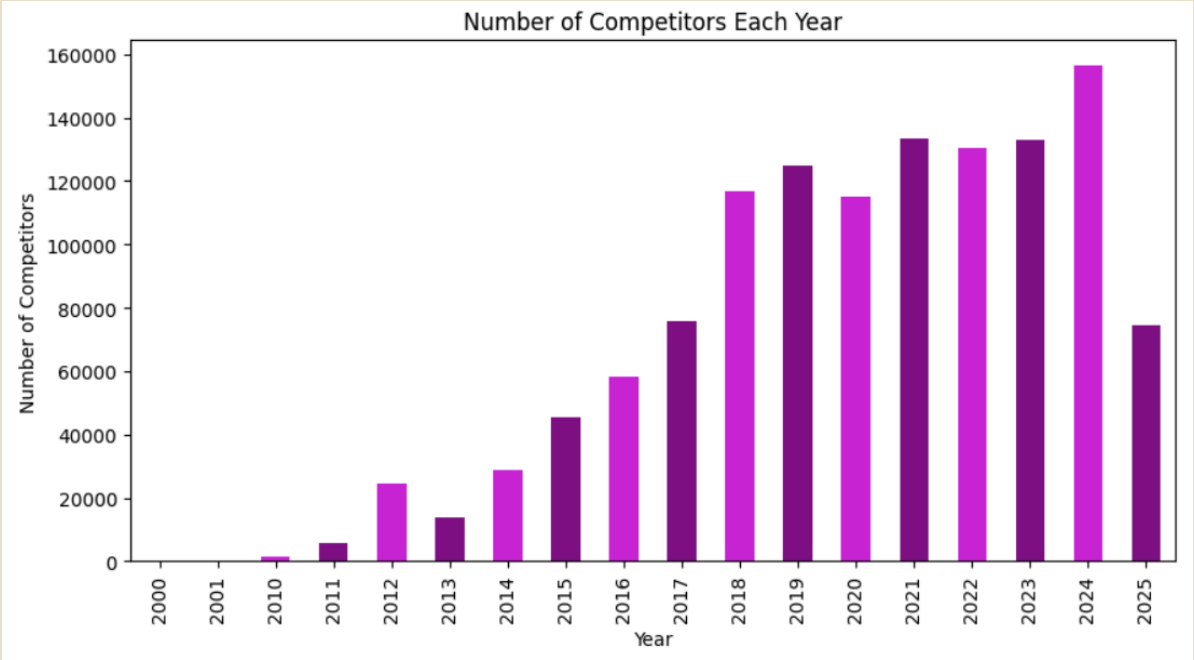
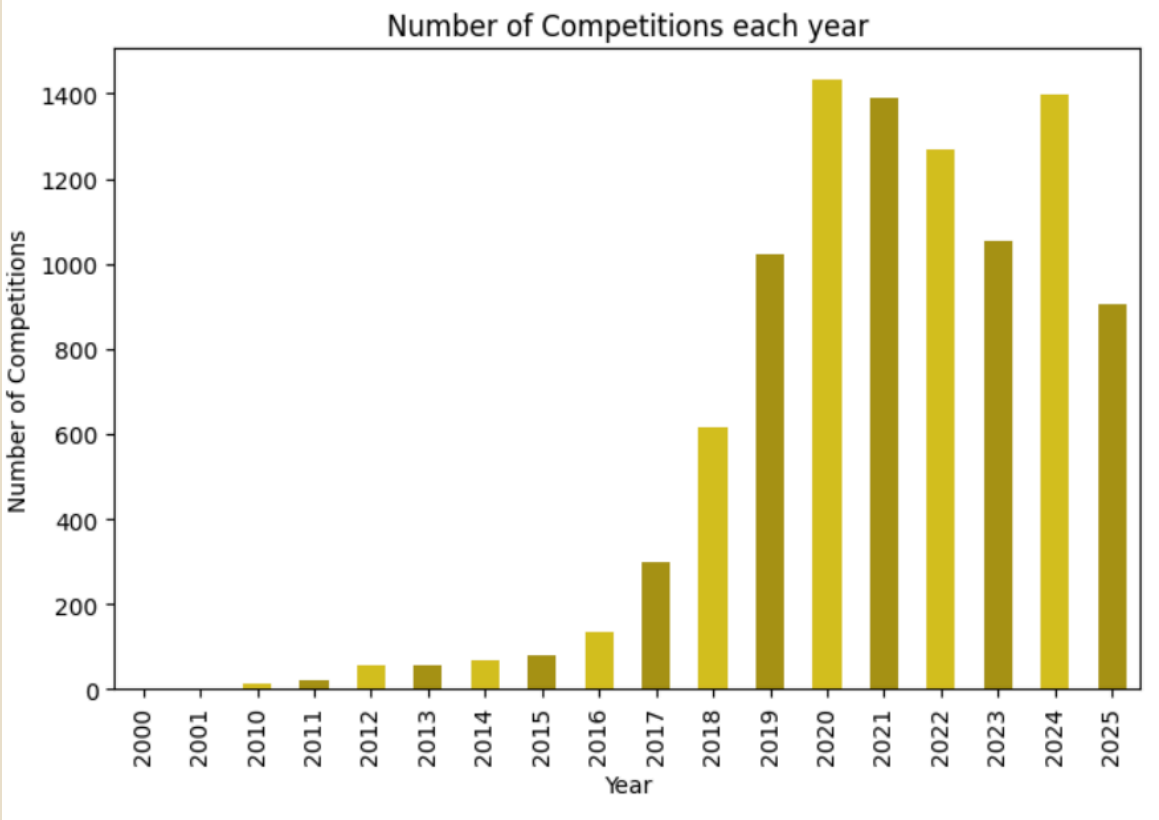
One of the main motivations for participation in competitions is the reward/ outcome after finishing it. Therefore, it is important to have appropriate rewards that interest the target audience. The graph reveals that USD and Knowledge are the most common rewards in Kaggle competitions.



The graph shows that 2024 had the highest number of prizes for kaggle competitions which may indicate that this year had the highest number of competitions offered.



The graph indicates that 2020 had the highest number of competitions (1434) while 2024 had the second highest number of competitions (1397). And the second graph shows that 2024 had the highest number of competitors.



The highest number of teams and highest number of submissions are both found in 2024 which may indicate a correlation as number of teams increases, the number of submissions increases too

