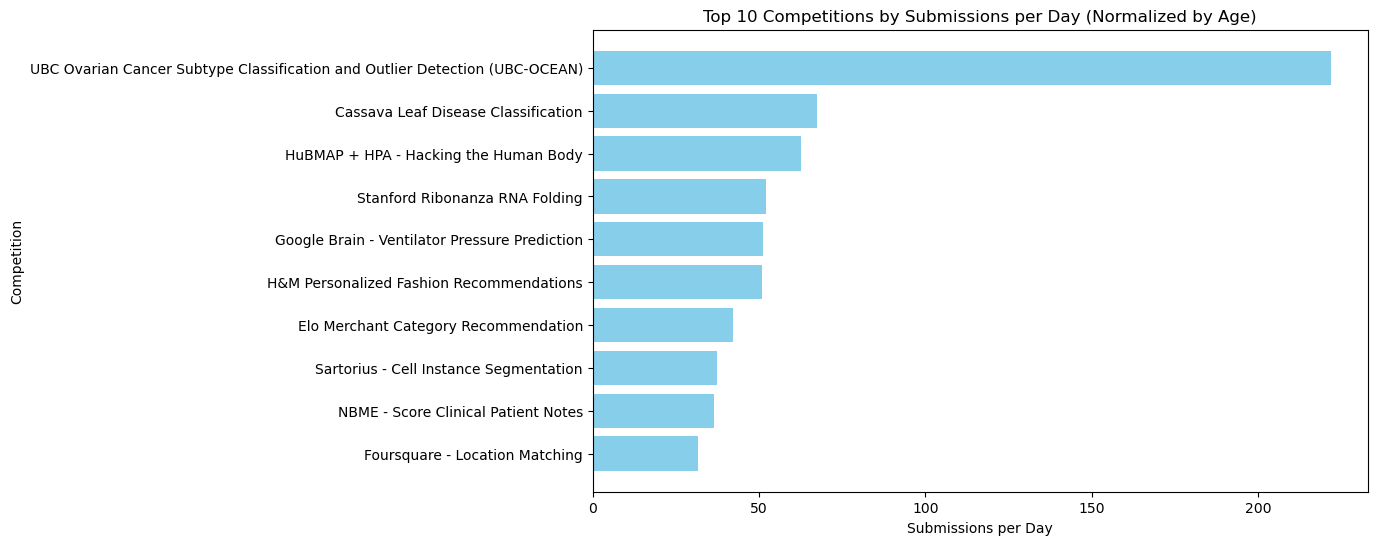
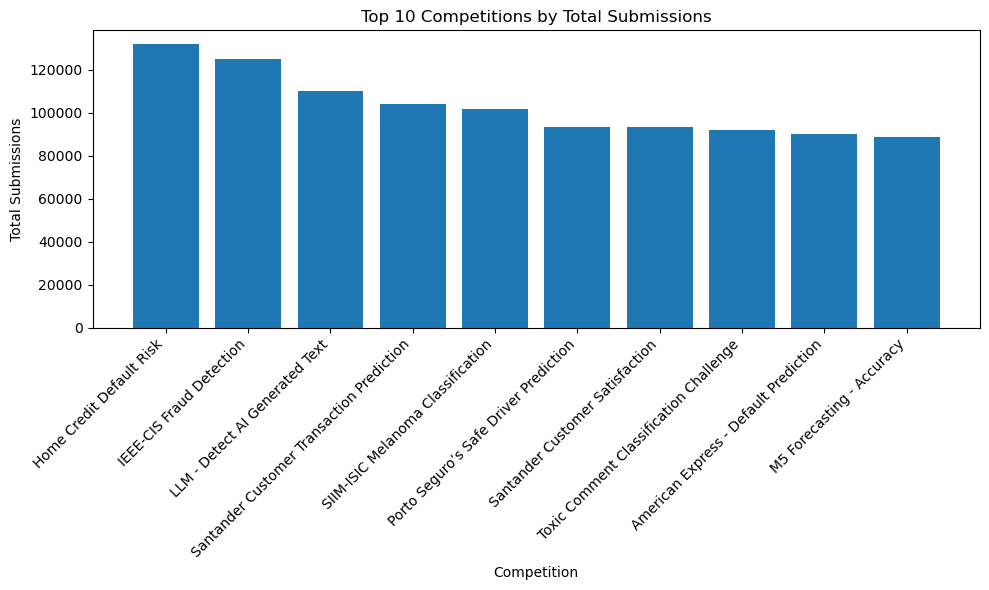
Visualization Tasks

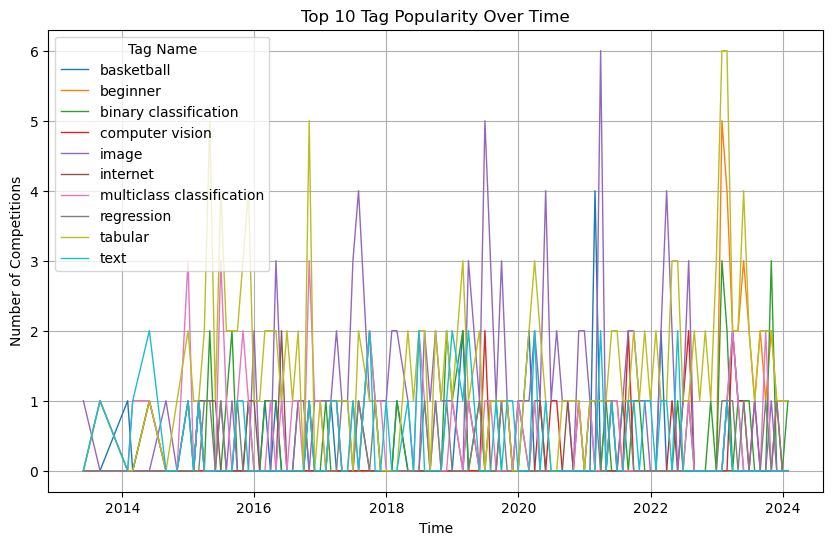
* Which competitions were most popular over time? Which competition **tags** were most popular over time?



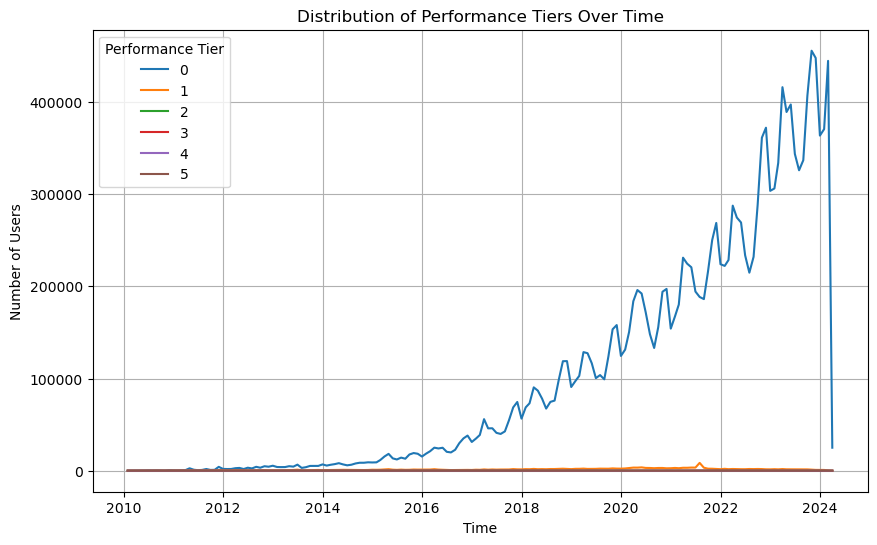
**Figure 1 : Bar plot showing the top 10 competitions ranked by number of total submissions divided by age of competition**



**Figure 2 : A Bar plot of the top 10 competitions ranked by total all-time submissions**

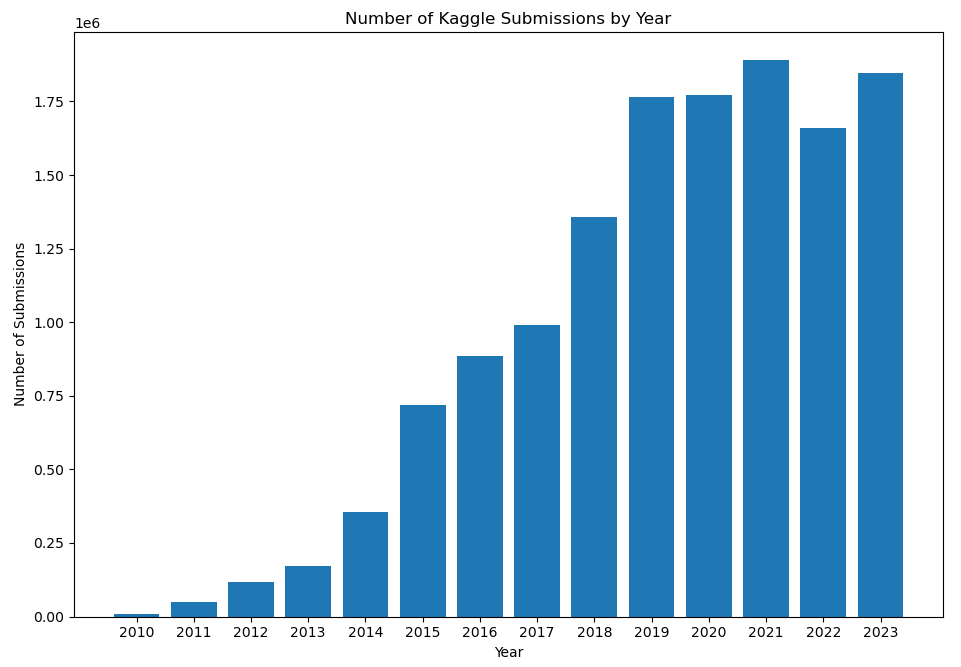


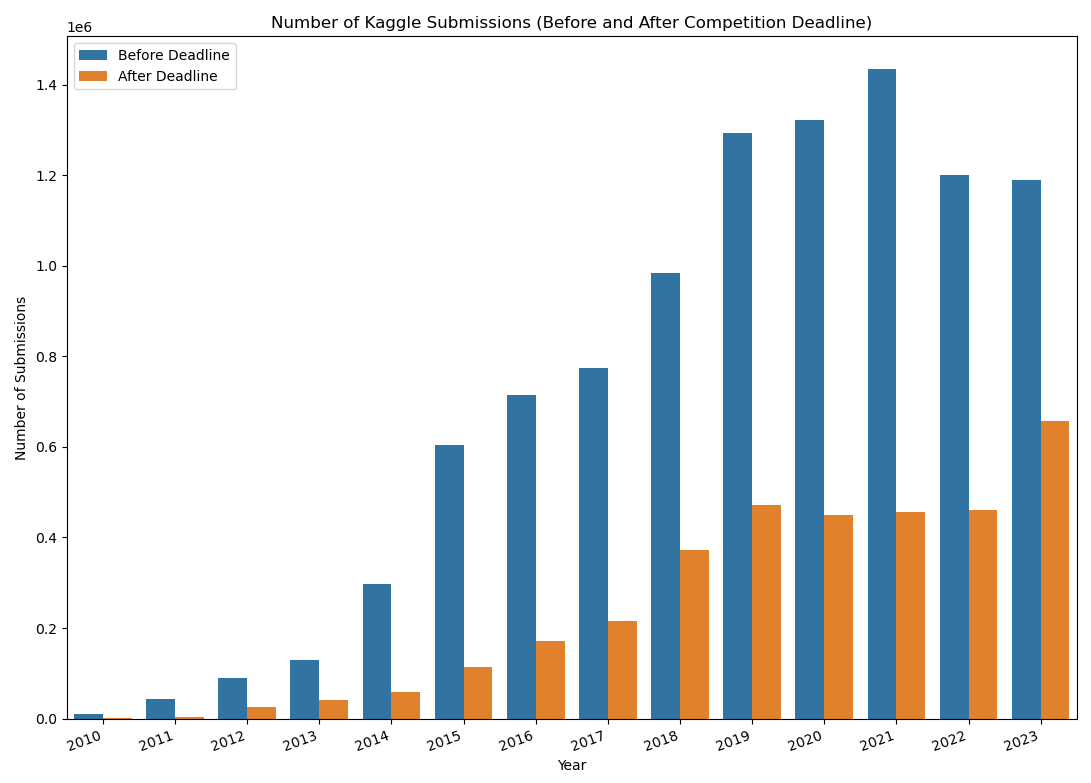
**Figure 3 : Line plot visualizing the popularity of top 10 competition tags over time**

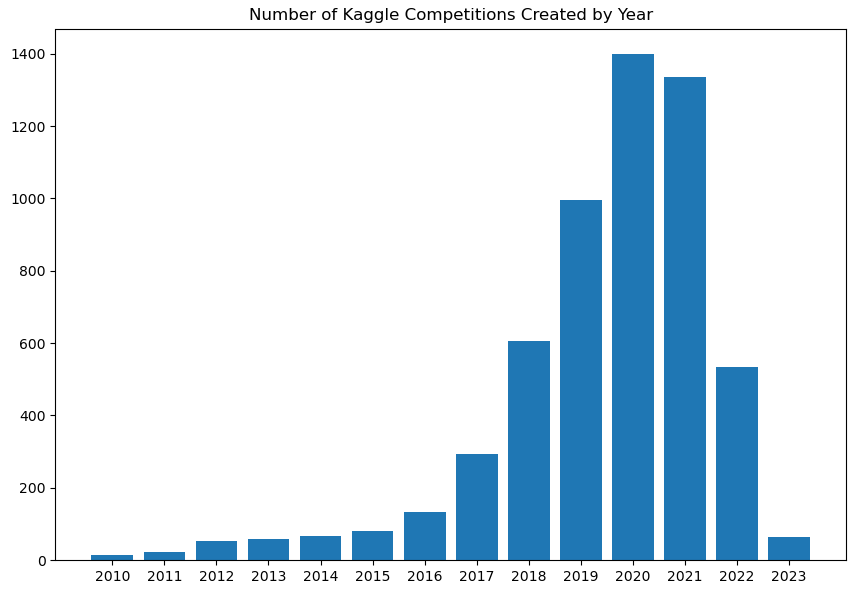
* How did the distribution of performance tiers (see Users.csv) change over time? Show this as a table of percentages.

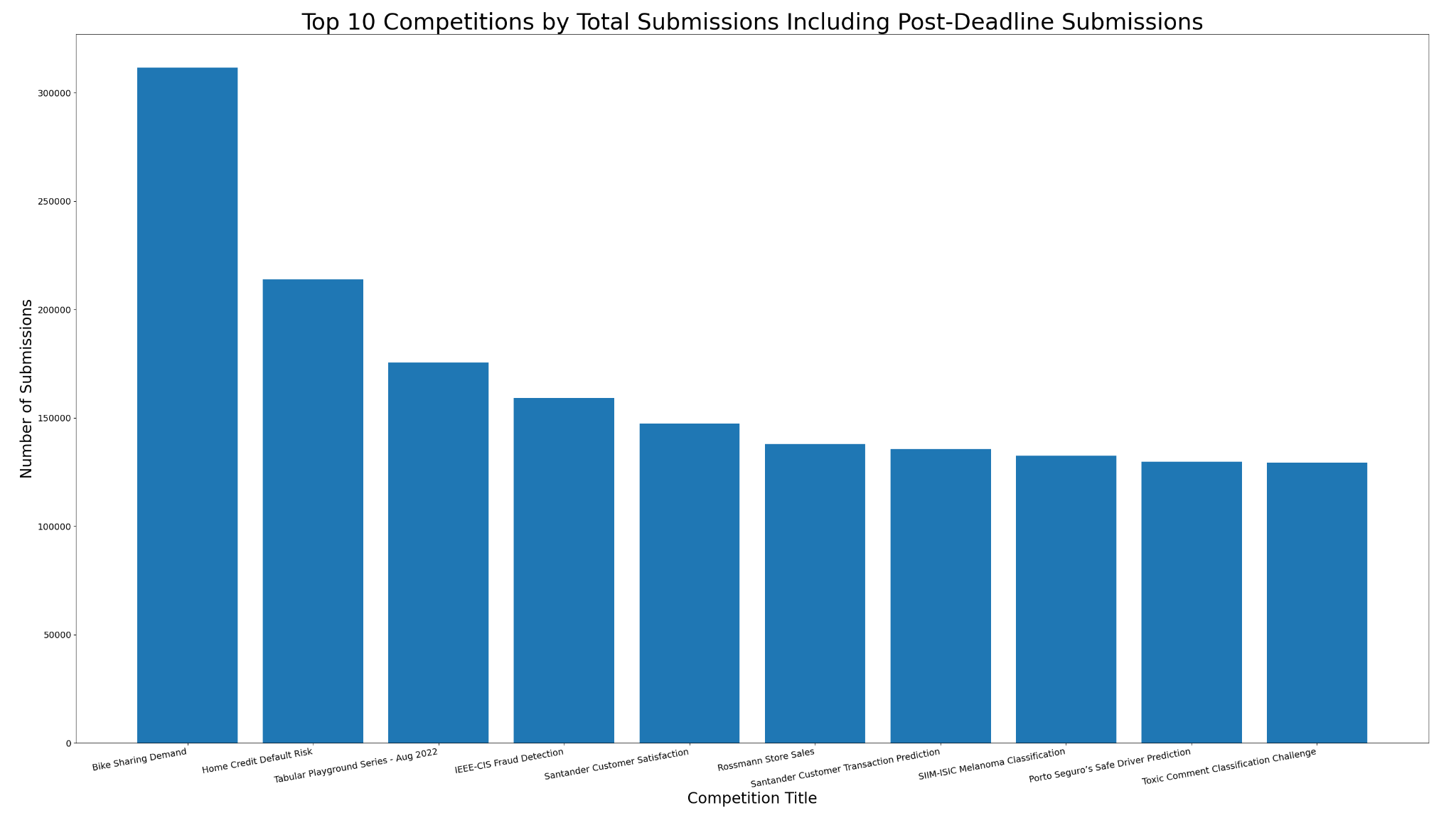
**Figure 4 : This figure visualizes the distribution of user performance tiers over time using a line plot**

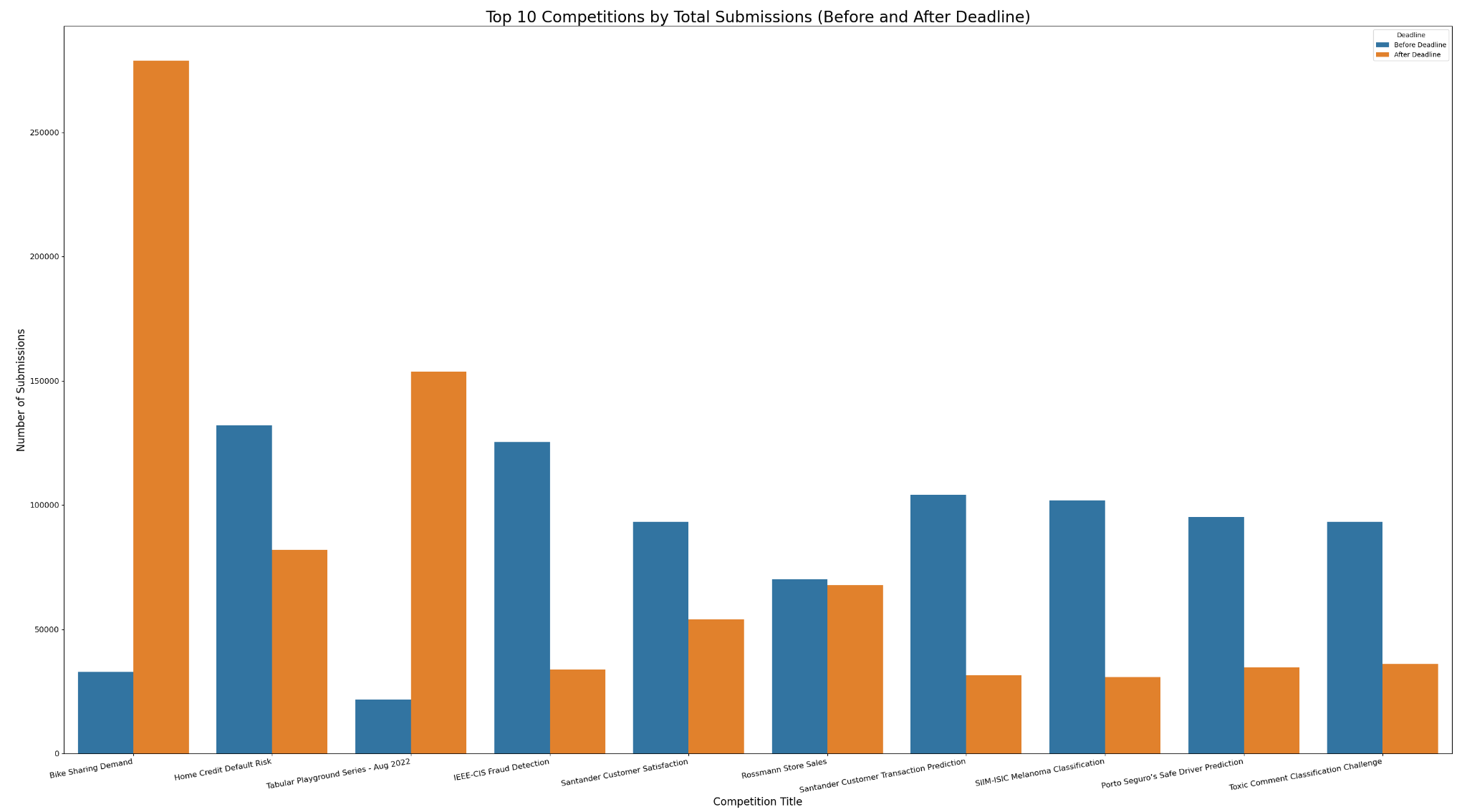
* Which data sets were most popular over time? Which data set **tags** were most popular over time? Measure popularity by number of downloads, number of kernels, number of forum posts, … anything else? (Monisha)
* How does collaborating on competitions (team size) impact the likelihood of winning or achieving top rankings? (Monisha)
* How do the number of new users change per year? In addition to the current plot of number of new competitions, create a line graph that plots the number of new competitions created every year for the top X competition tags (Kabir)
* How has the popularity of Python and R changed over time? (Kabir)
* Does the size of competition prizes correlate with higher participation rates and submission volumes? (Ishika)
* How does active participation in Kaggle forums (ex: number of posts, number of votes in forum, etc.) correlate with competition performance and ranking improvements? (YSC)
* Plot relationships between how many votes users received in forums, how many posts they made, how many submissions they made, competitions rankings, upvotes/post ratio, …. How do these relationships change over time? (YSC)  
    
  The plots suggested above turned out to be not very informative or insightful, so I came up with other plots:

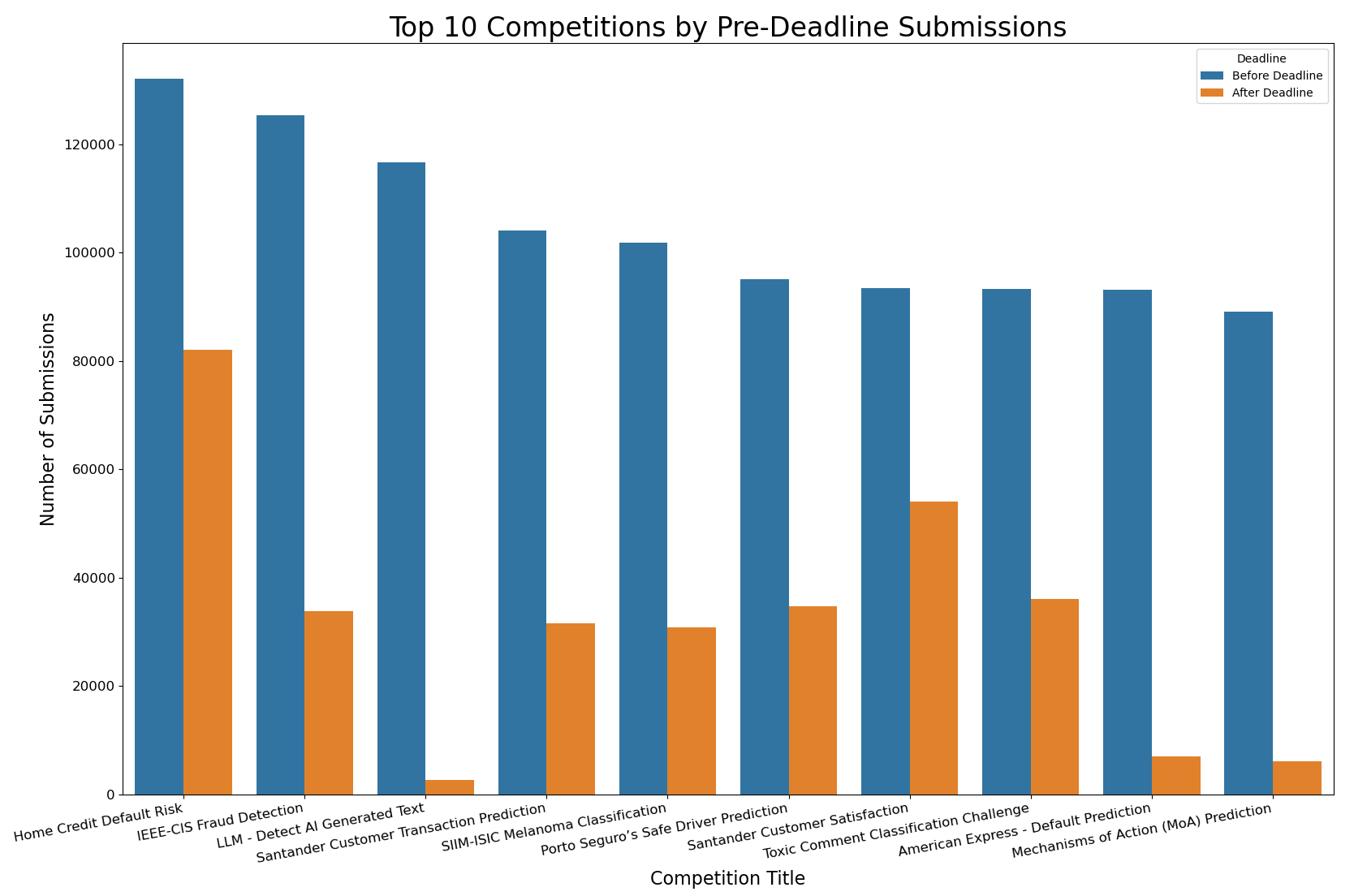


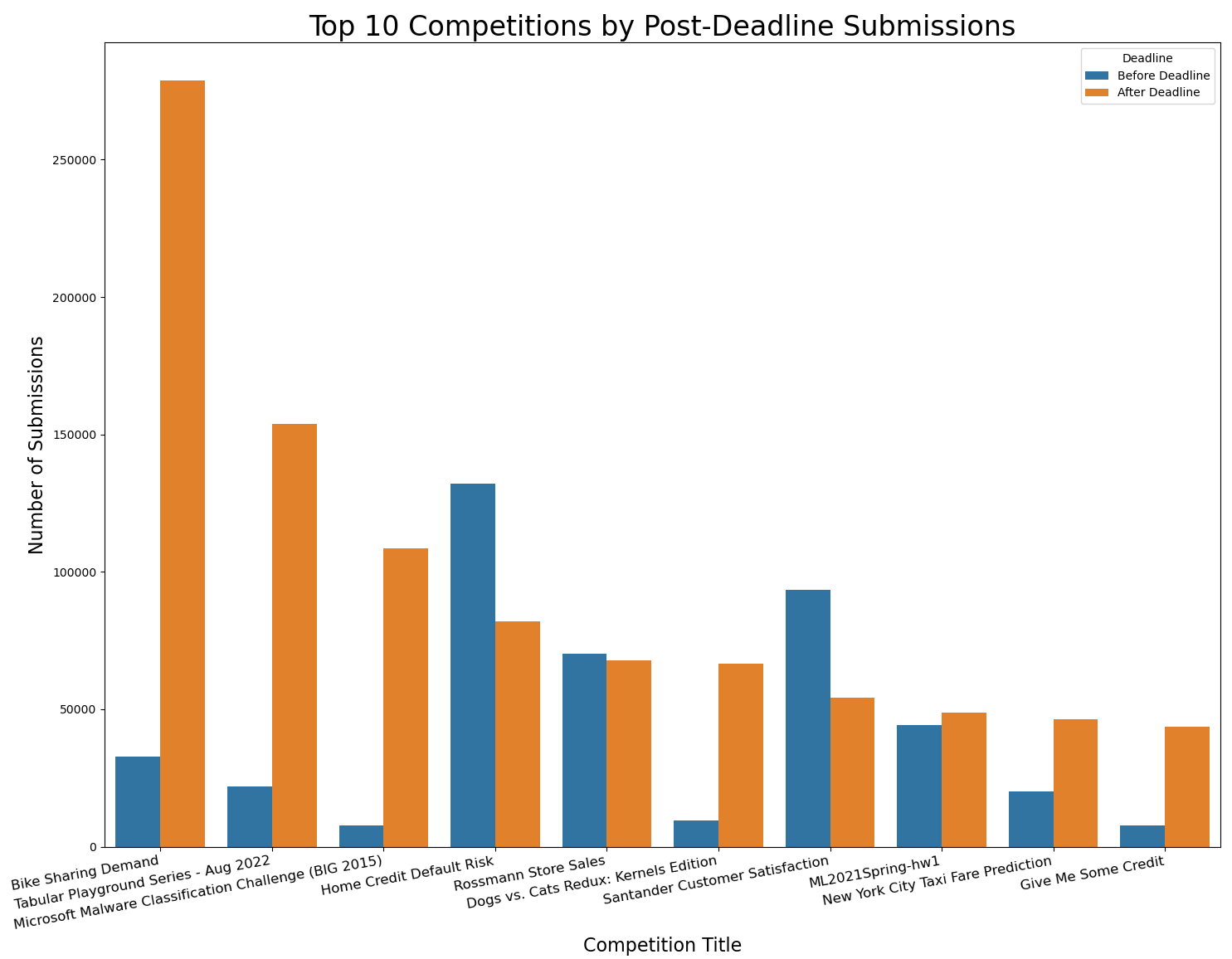


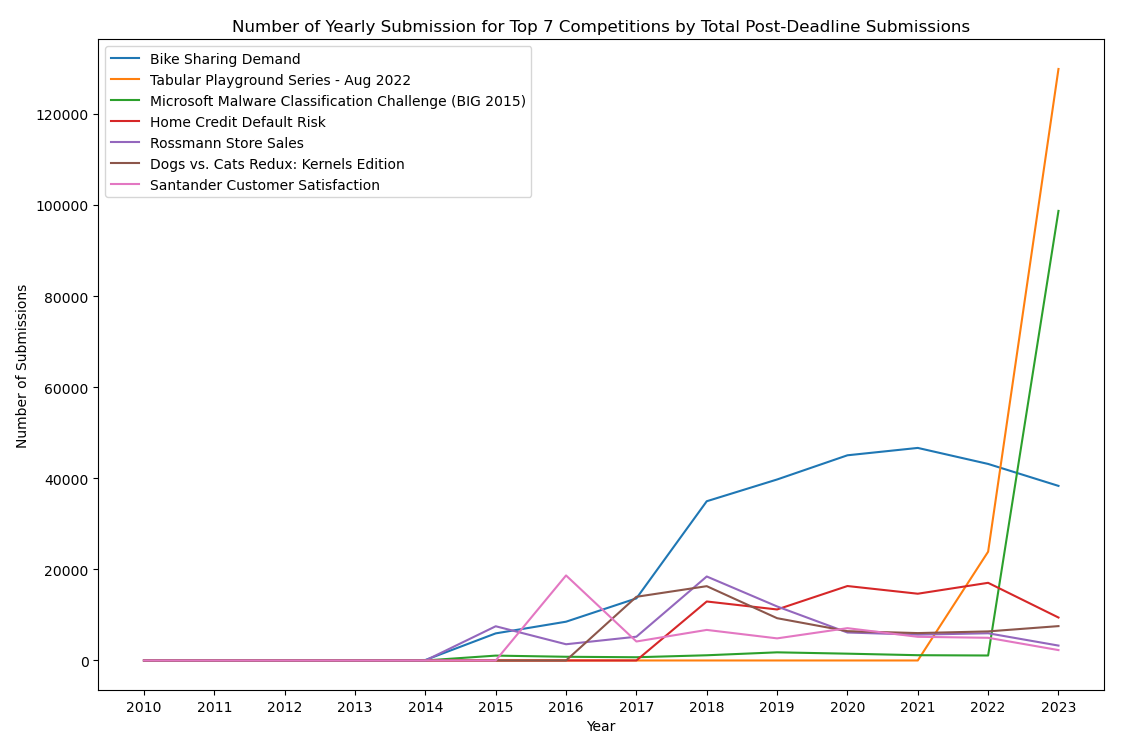


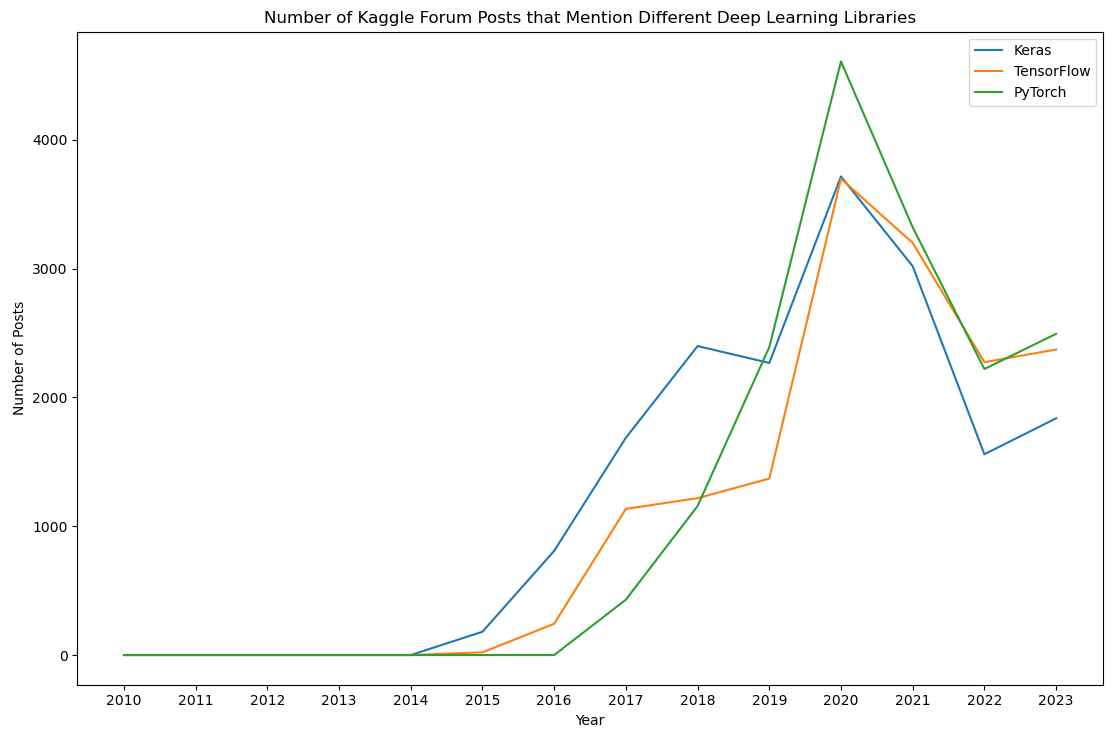




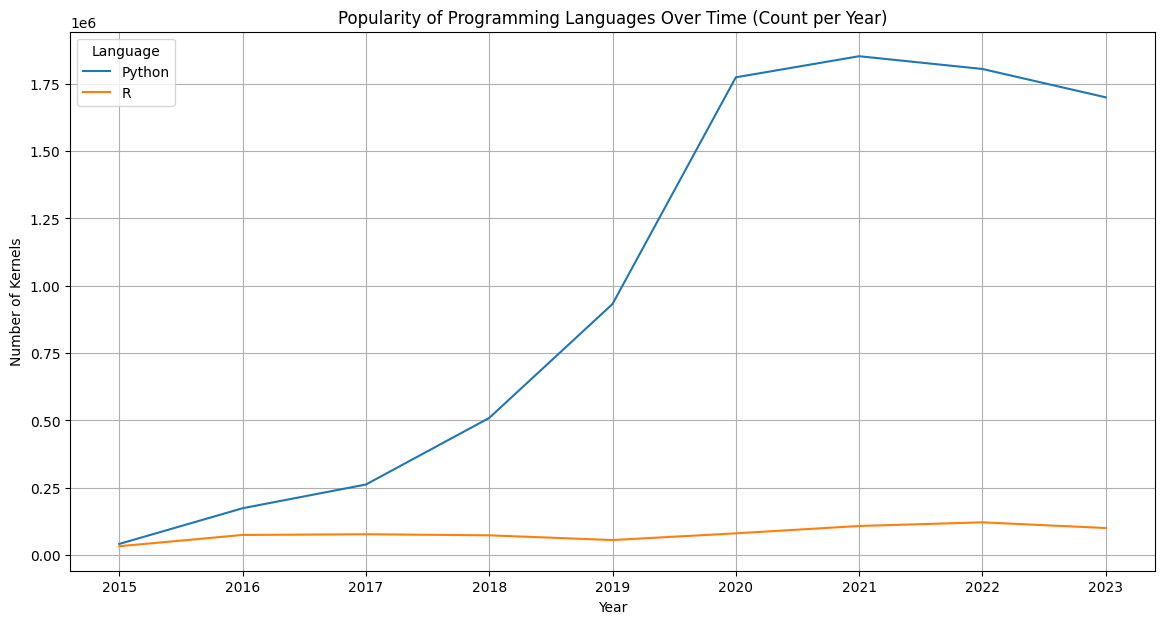








* **How has the popularity of Python and R changed over time?**



* **How does collaborating on competitions (team size) impact the likelihood of winning or achieving top rankings?**

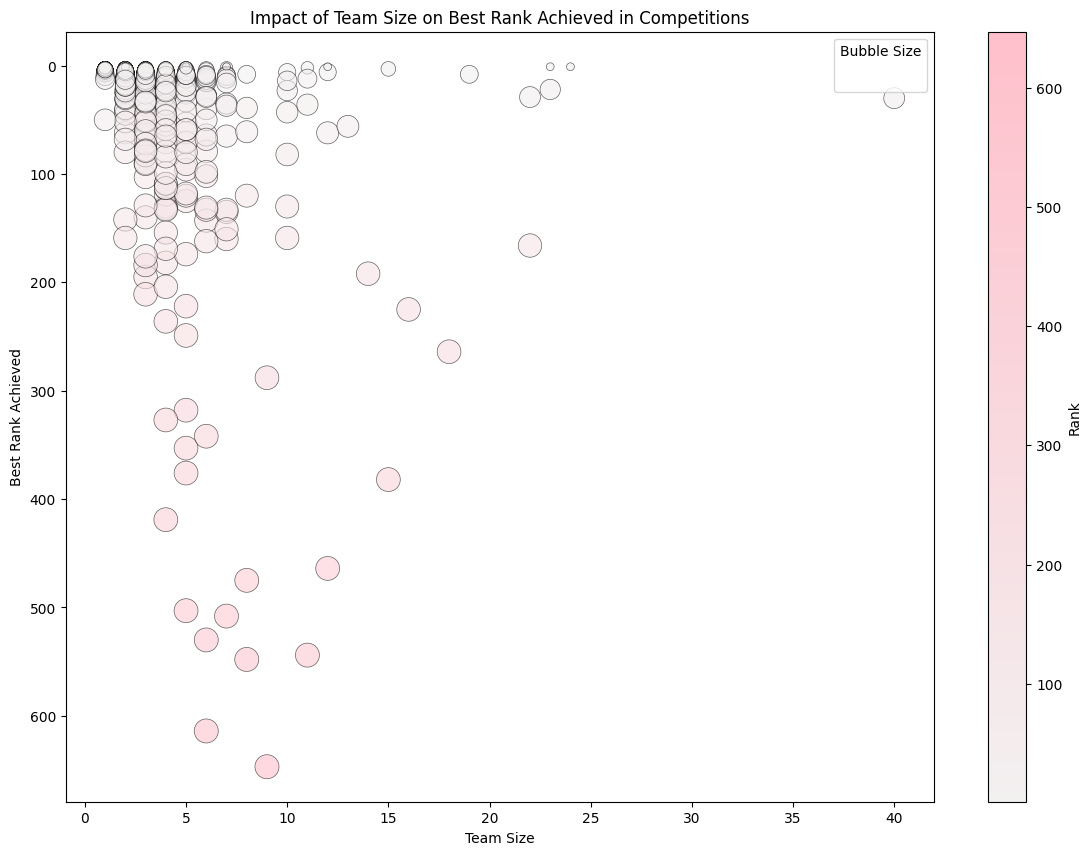


Fig. . Correlation Between Team Size and Peak Performance in Competitive Settings

* **Do datasets owned by organizations receive more views/downloads compared to those owned by individuals?**

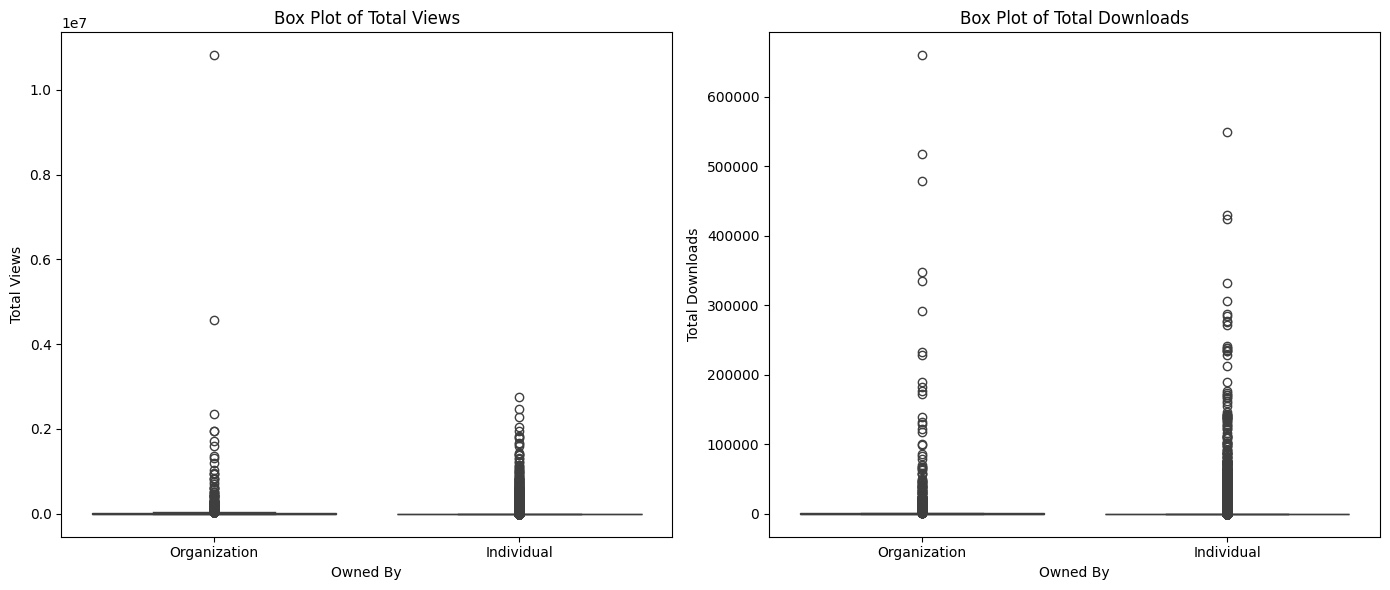


Fig . . Content Engagement:Organizational vs. Individual Impact on Views and Downloads

* **How has the popularity of datasets changed over time? (measured by views, downloads and votes)?**

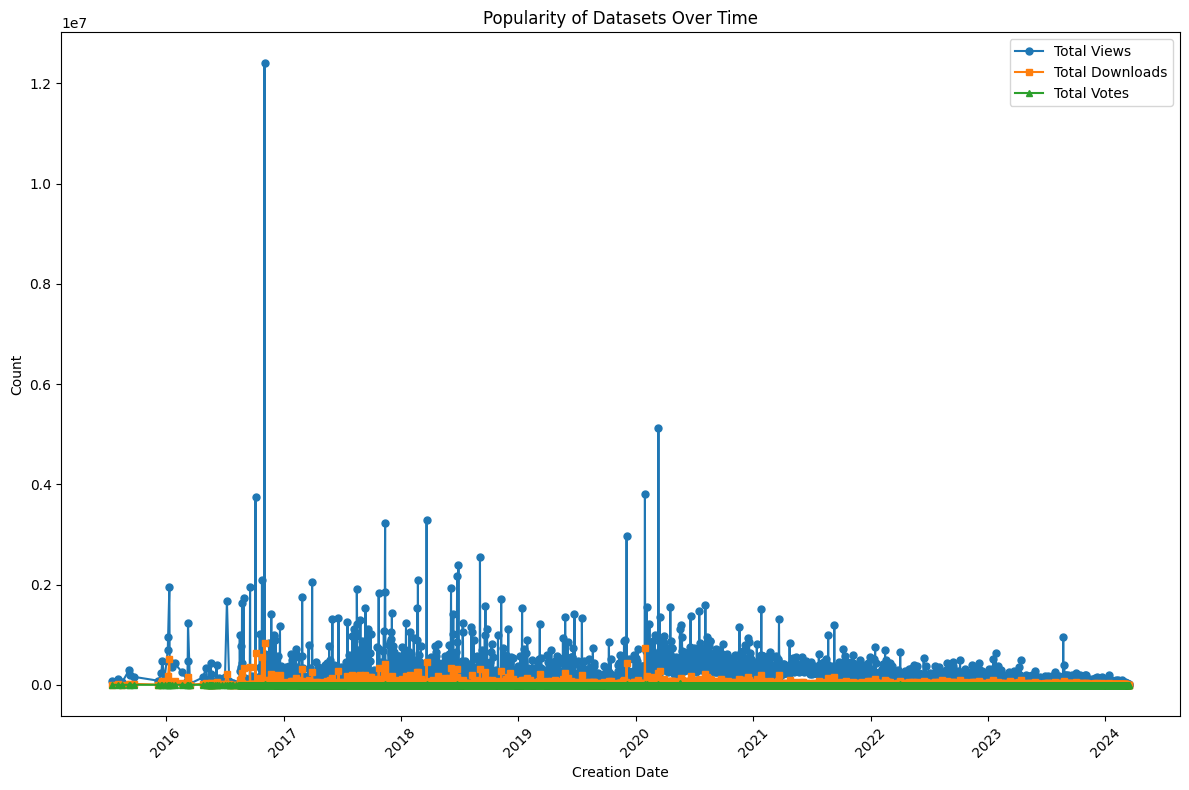


Fig . . Trends in Dataset Popularity: A Time Series Analysis of Views, Downloads, and Votes

* **What are the top 10 datasets in terms of total downloads and average monthly downloads?**

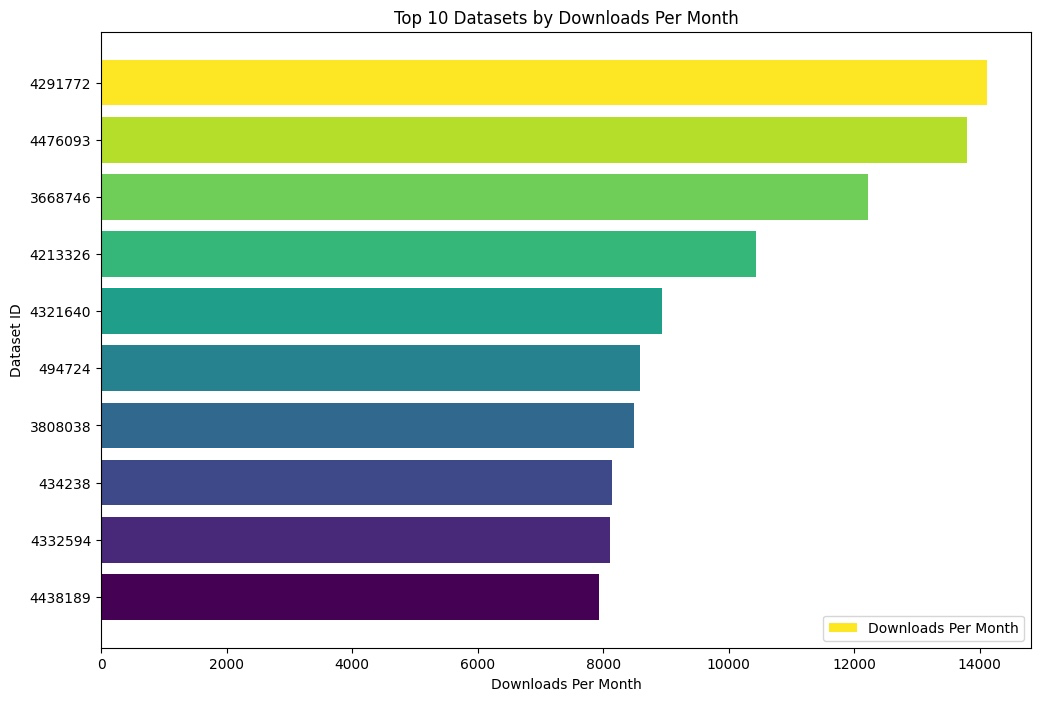
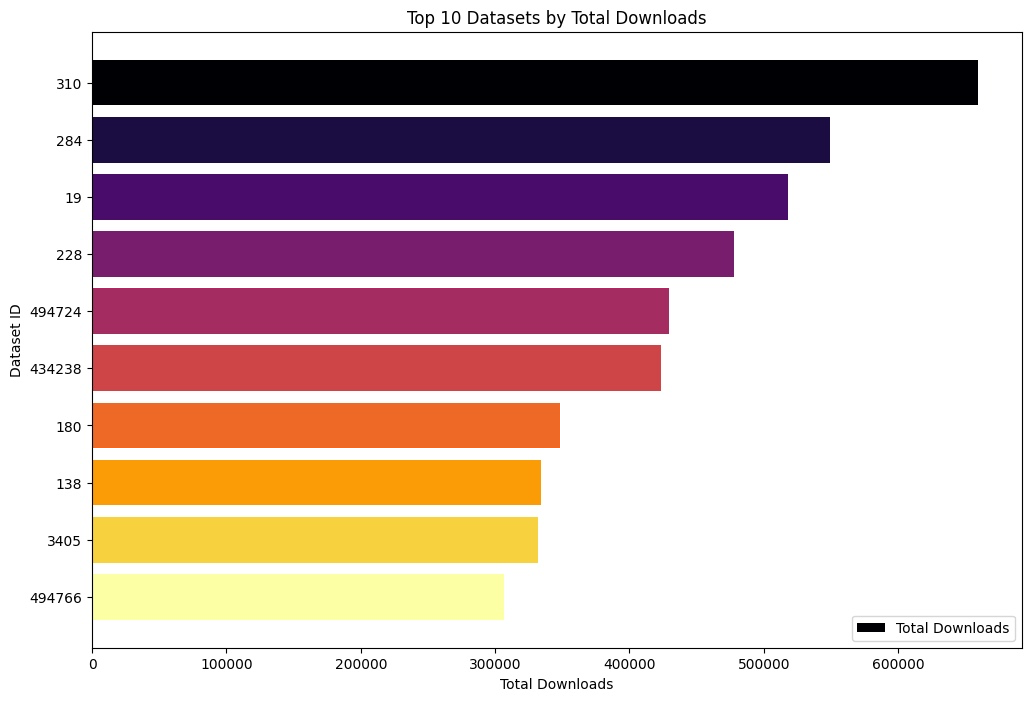


Fig. . Comparative Analysis of Top 10 Datasets by Total and Monthly Downloads