

Precog Task-3

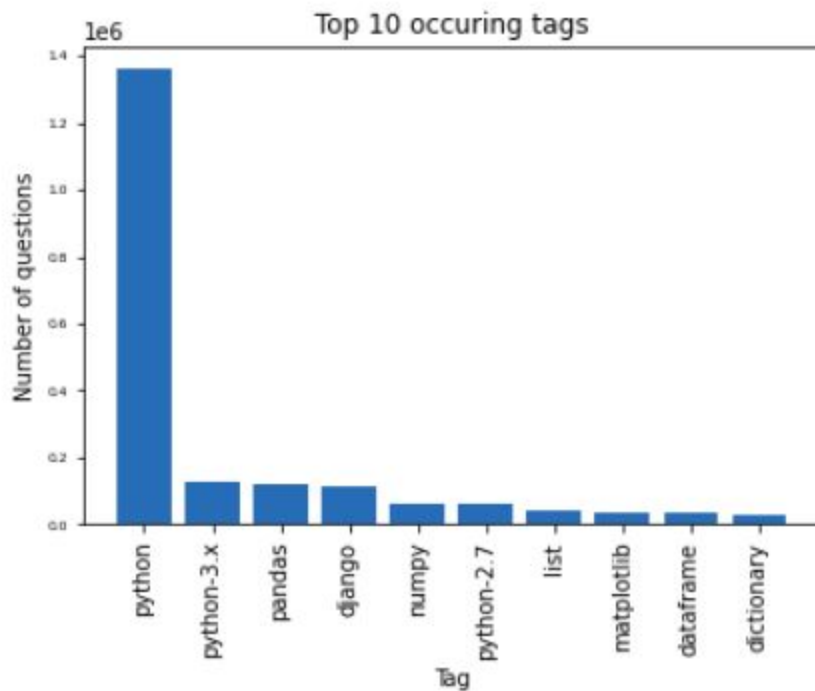
Sub-Task B

Exploratory Data Analysis

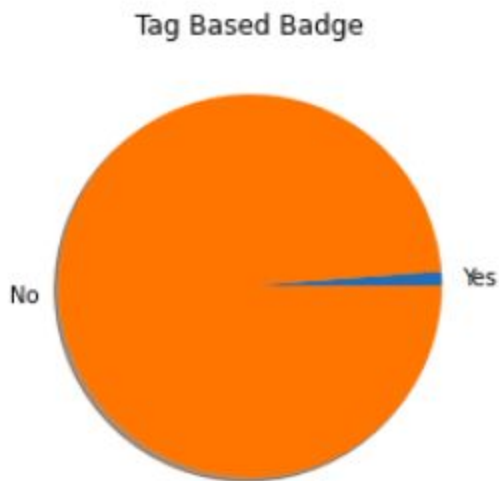
- Criteria used for subsampling: Time period 2008-2020
- Word cloud has been created for TagNames in the Tag XML file.



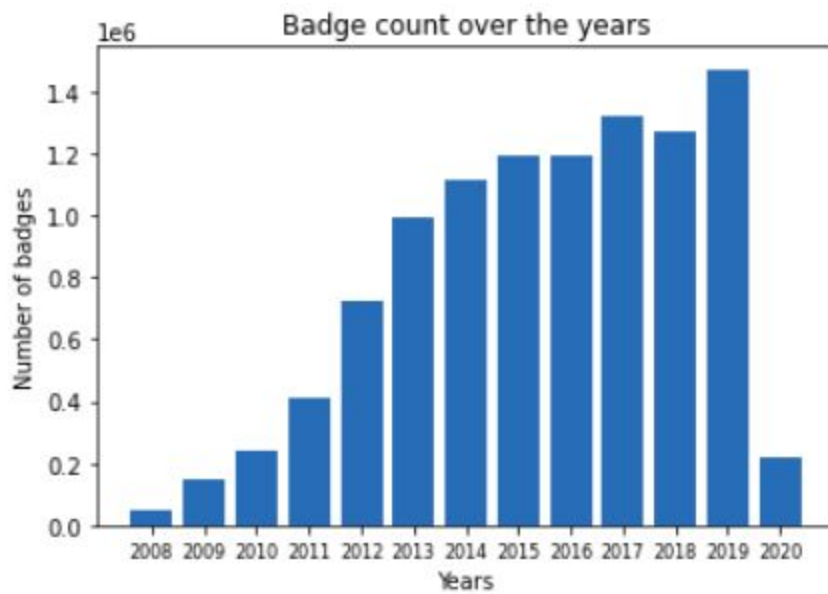
- Word cloud has also been created for locations in the User XML file.



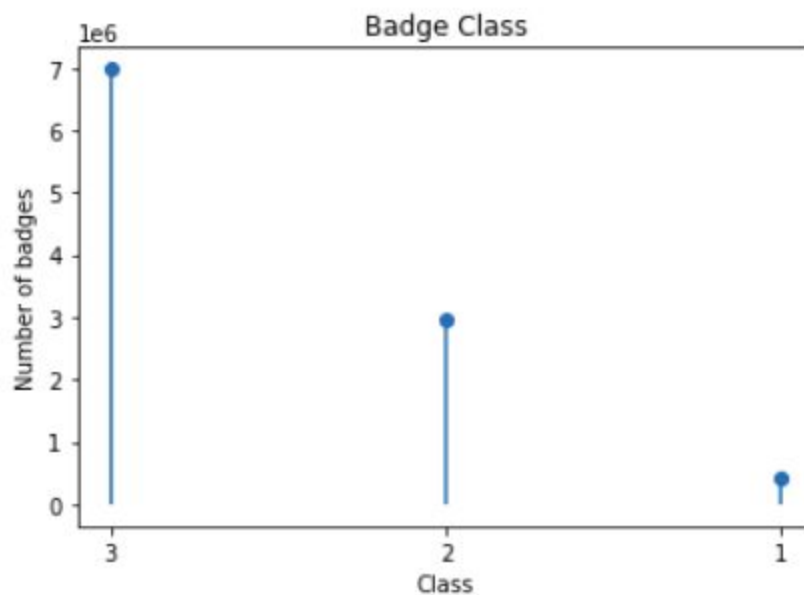
- Most of the badges were not tag based.



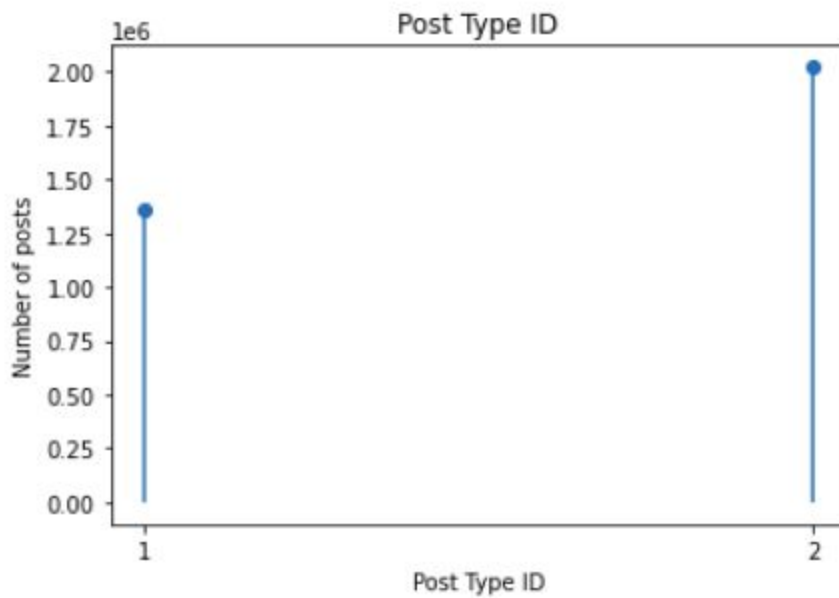
- Badge count has overall been increasing but there was a sudden fall in 2020. This could be because the entire data of the year 2020 is not present.



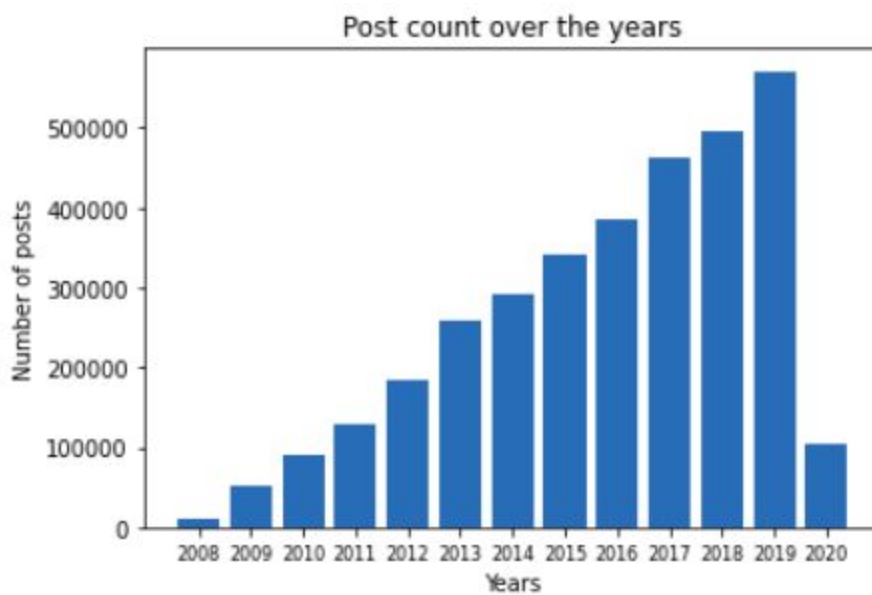
- Three badge classes are present in the data. Class 3 has the highest frequency.



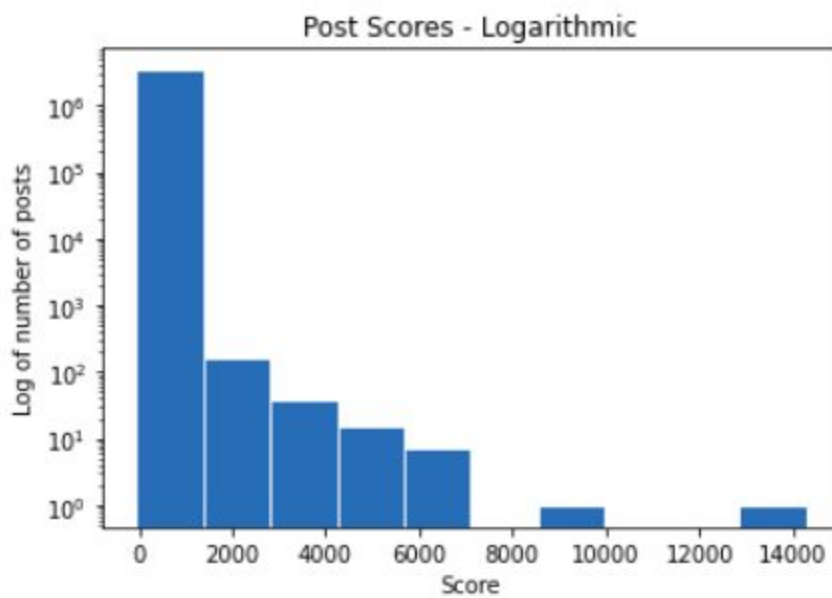
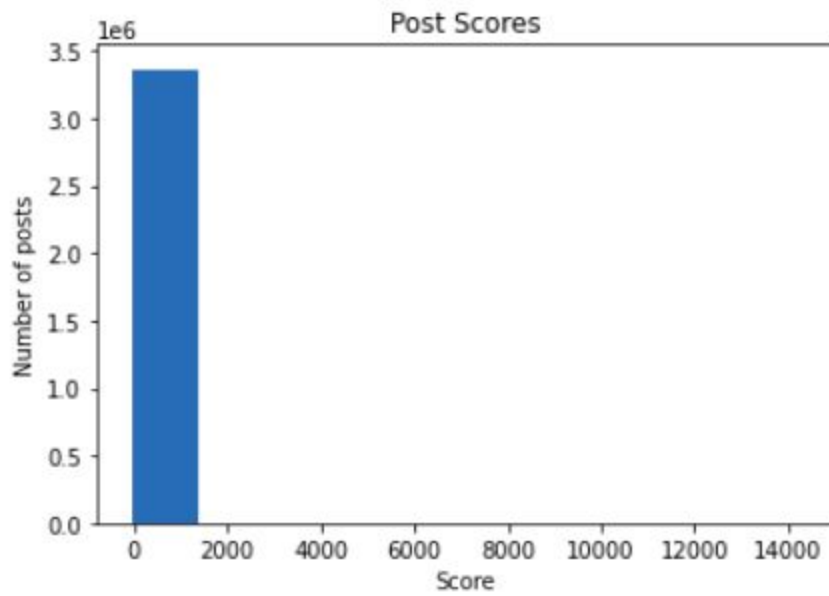
- Two types of post IDs are present. ID 2 has the highest frequency.



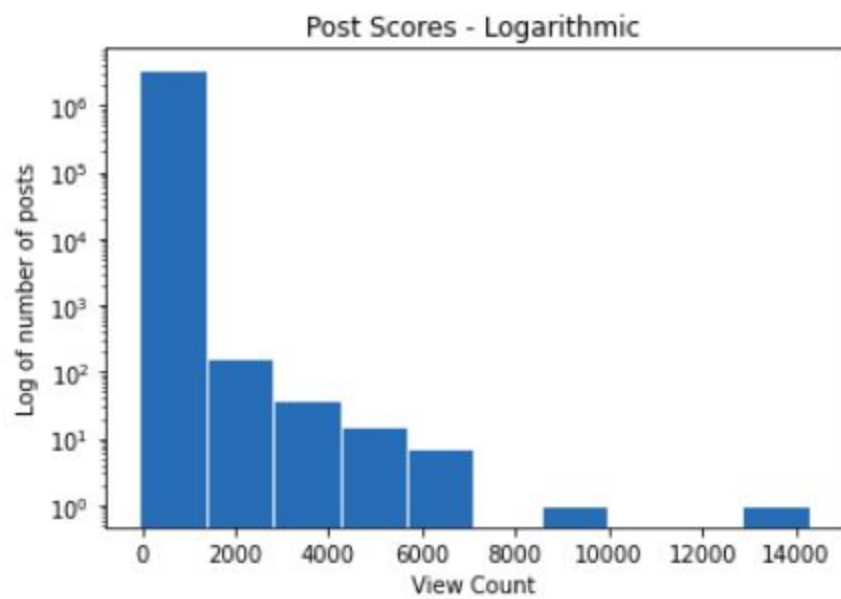
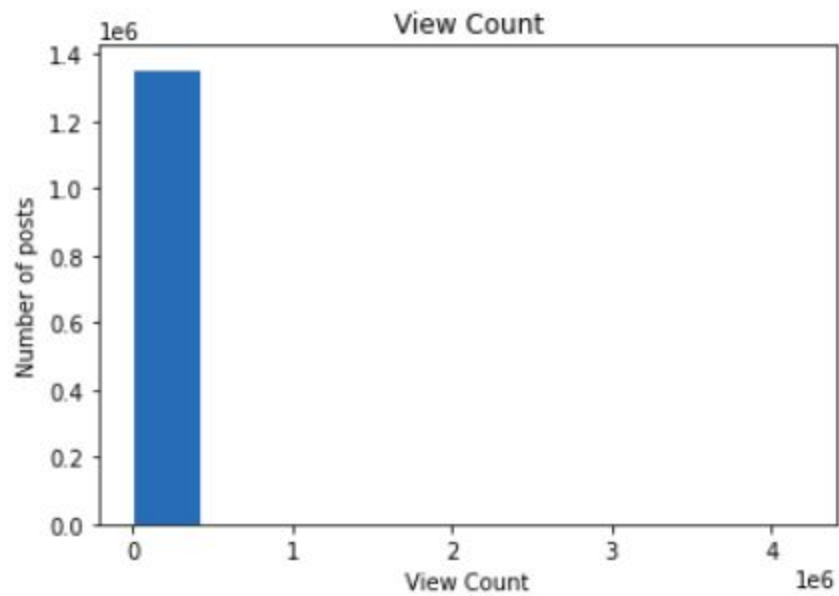
- Post count has been increasing each year.



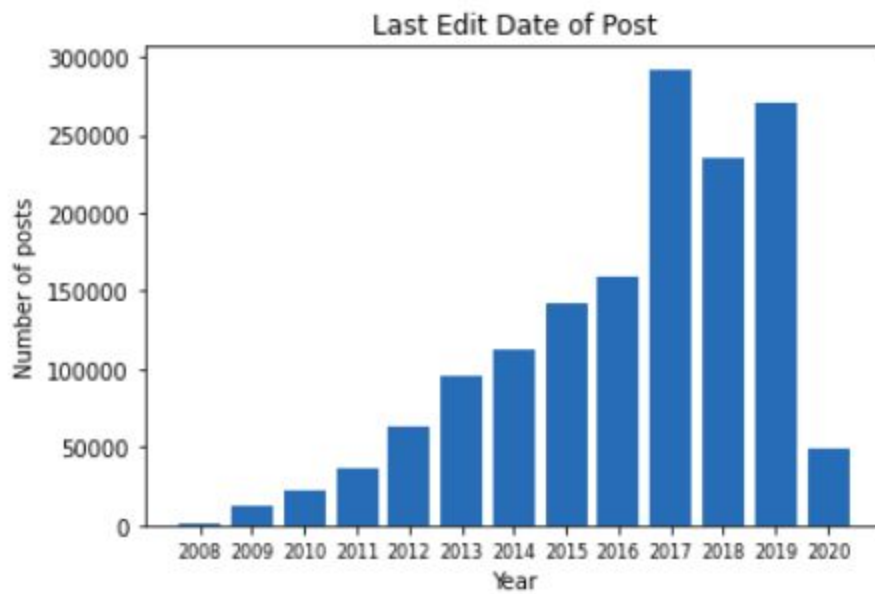
- Most post scores fall within the range 0-1000



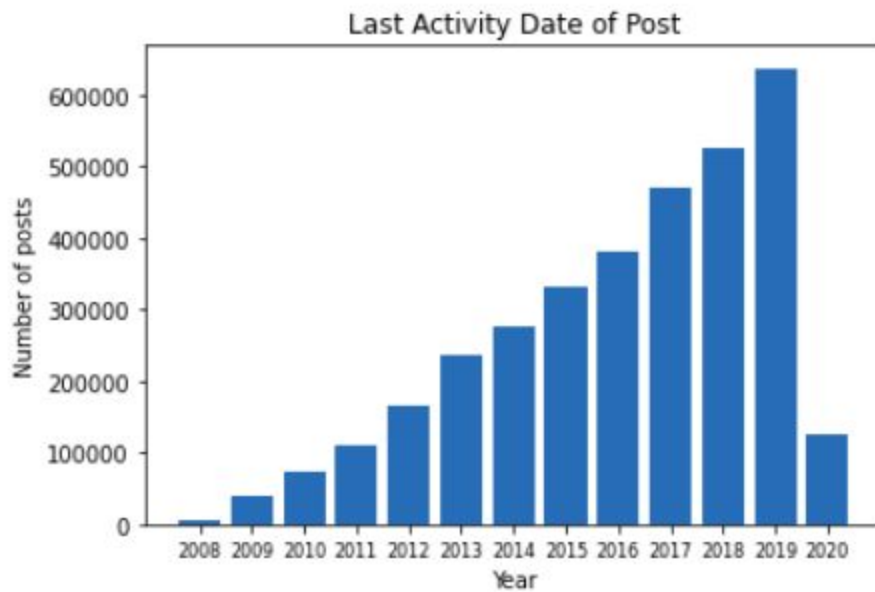
- Most view counts fall within the range 0-1500



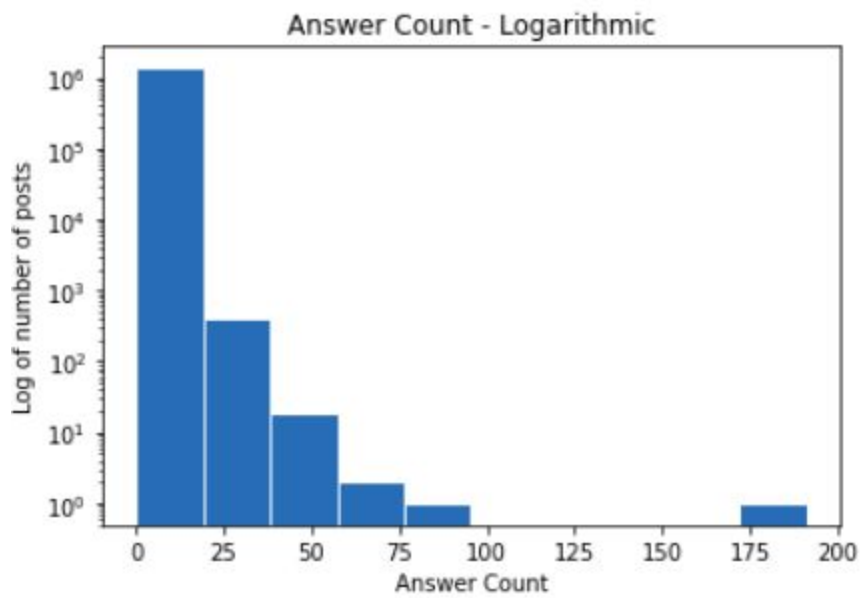
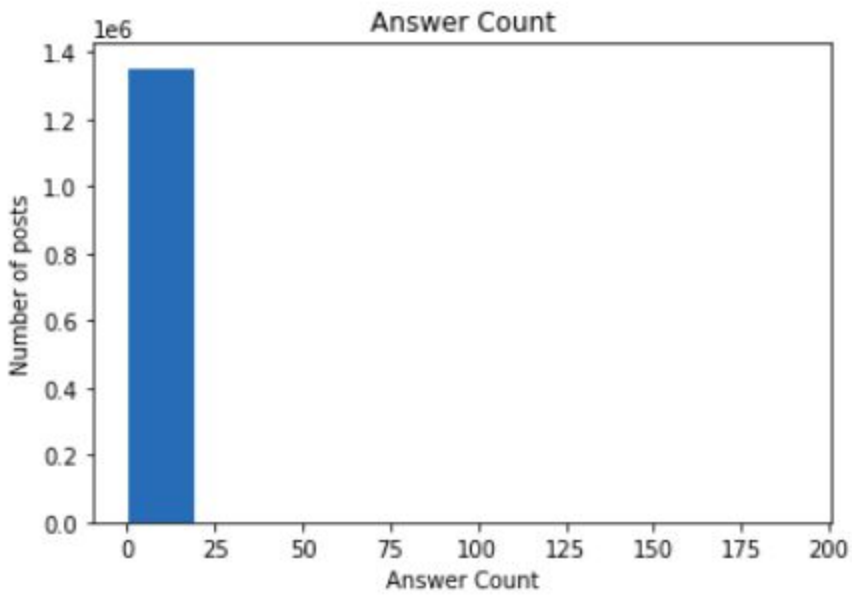
- Maximum number of posts have been edited in the year 2017.



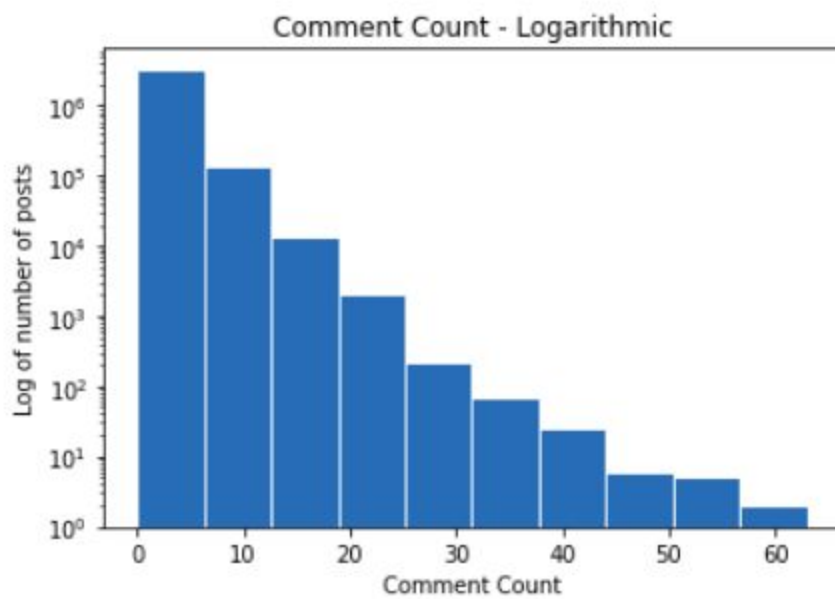
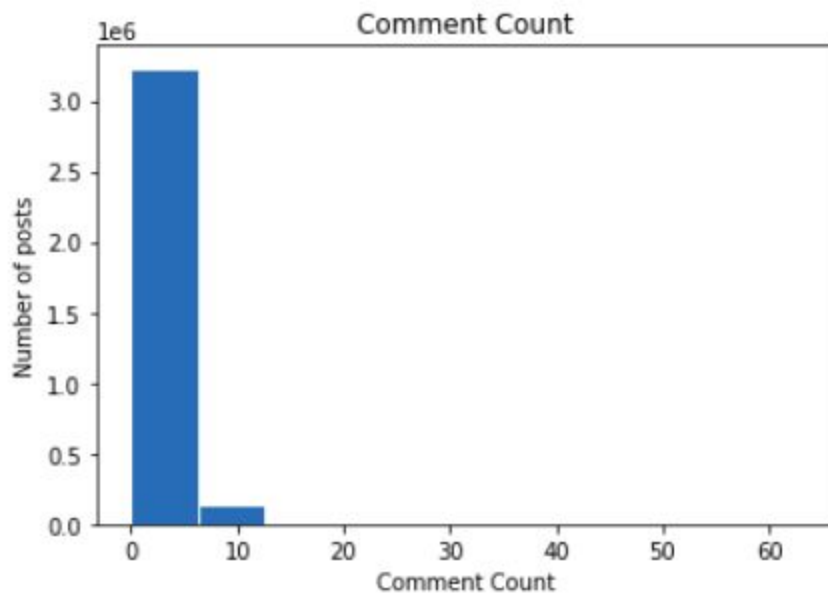
- Last activity date of posts



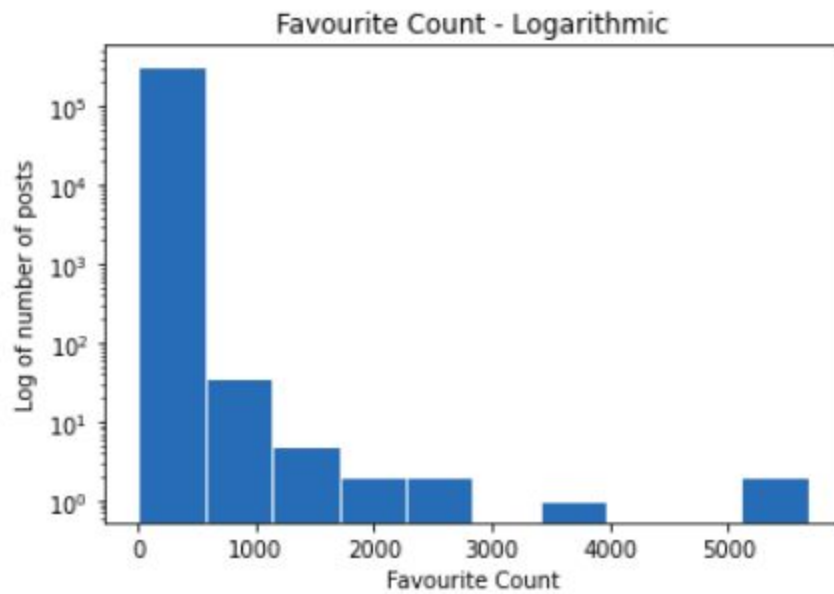
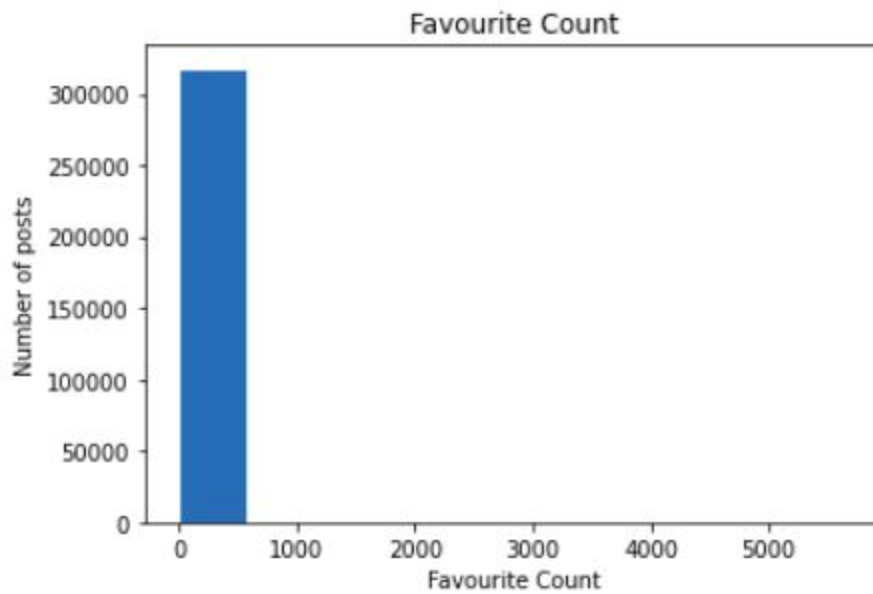
- Most answer counts fall within the range 0-20



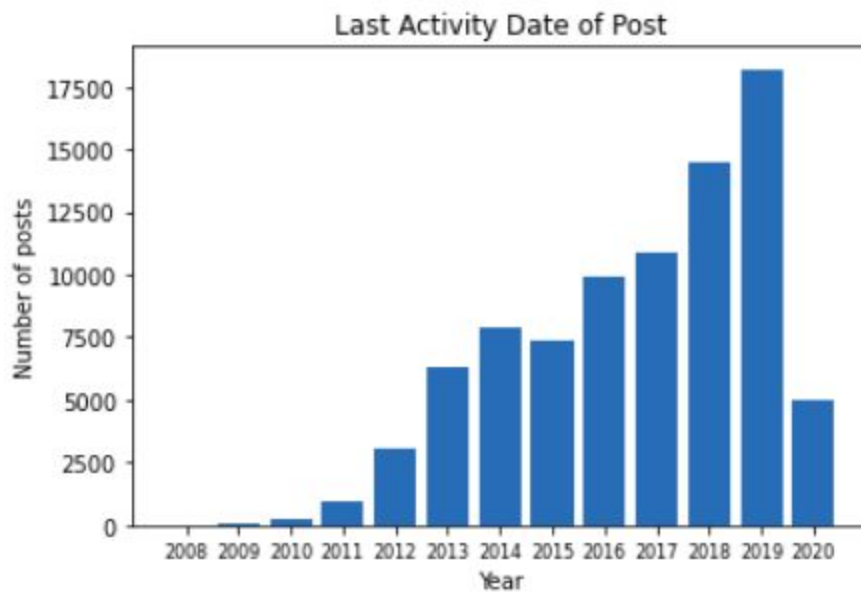
- Most comment counts fall within the range 0-10



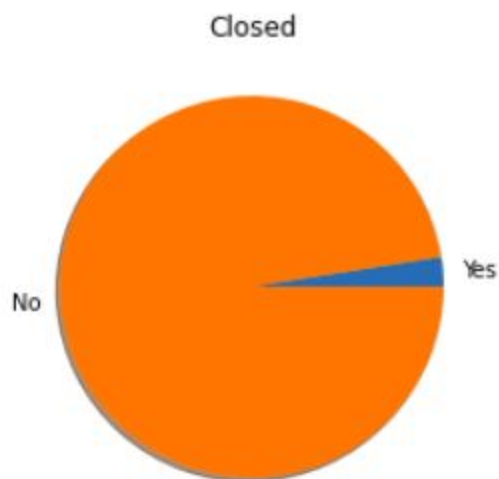
- Most favourite counts fall within the range 0-500



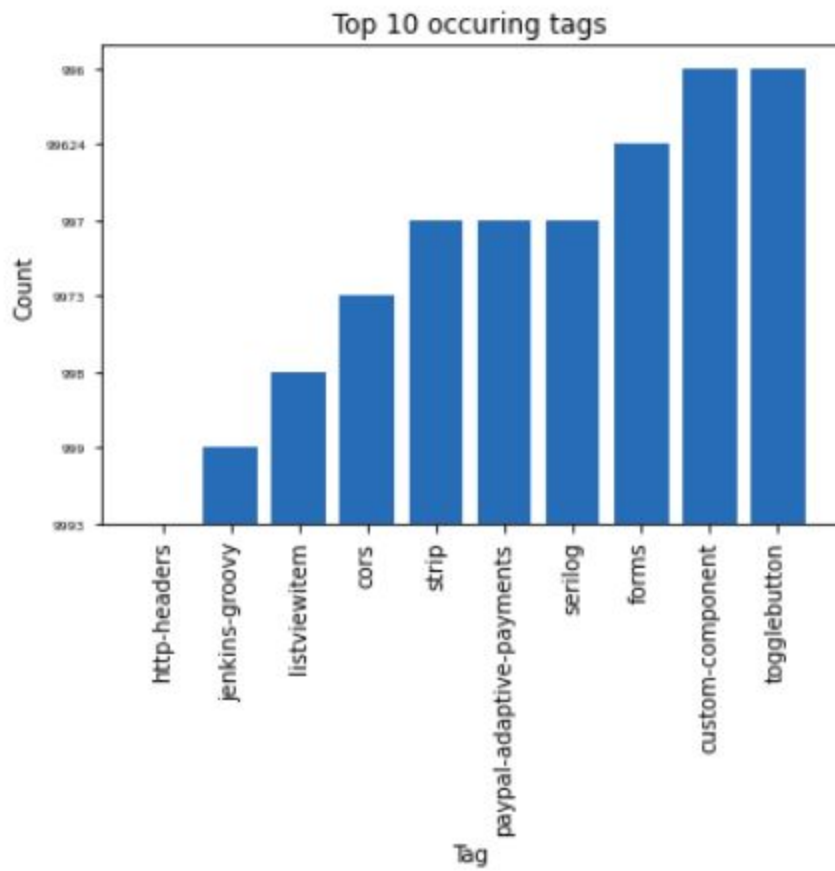
- Last activity date of post



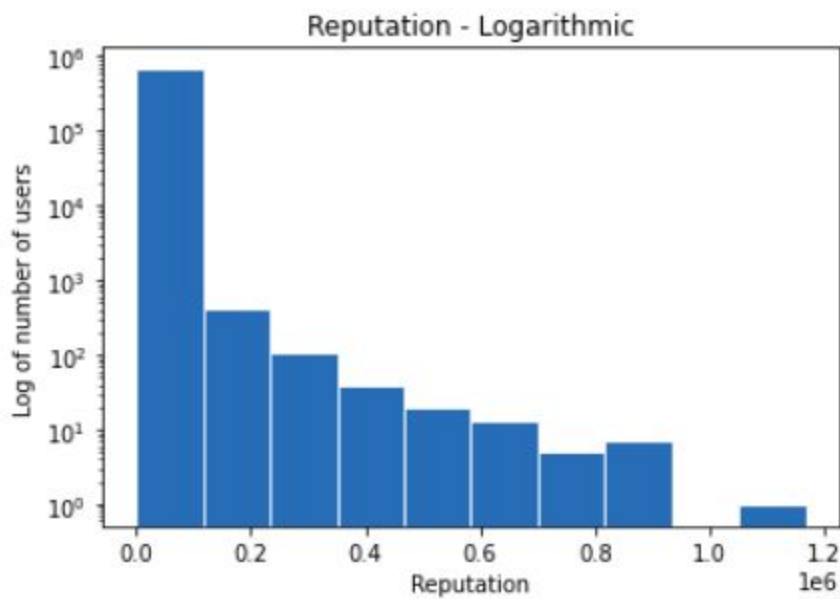
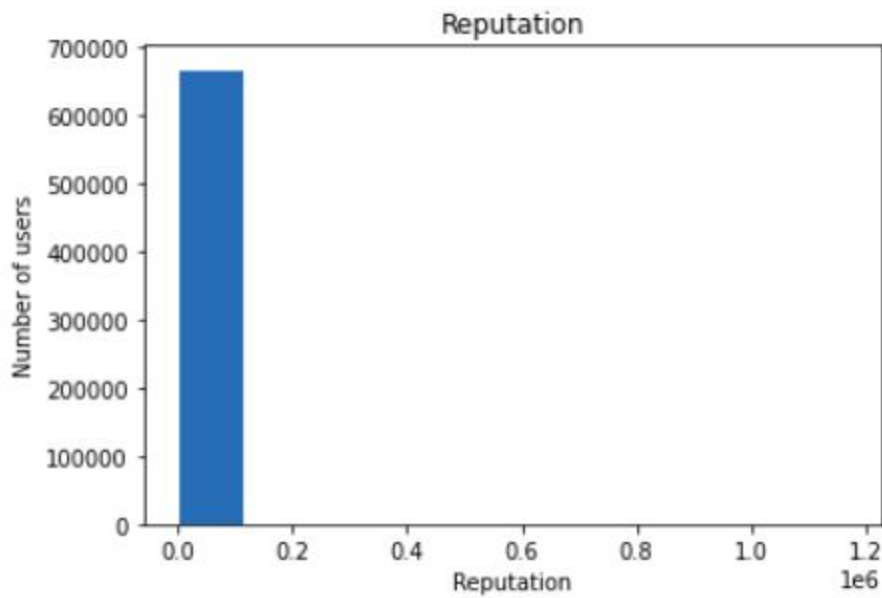
- Most of the posts are open



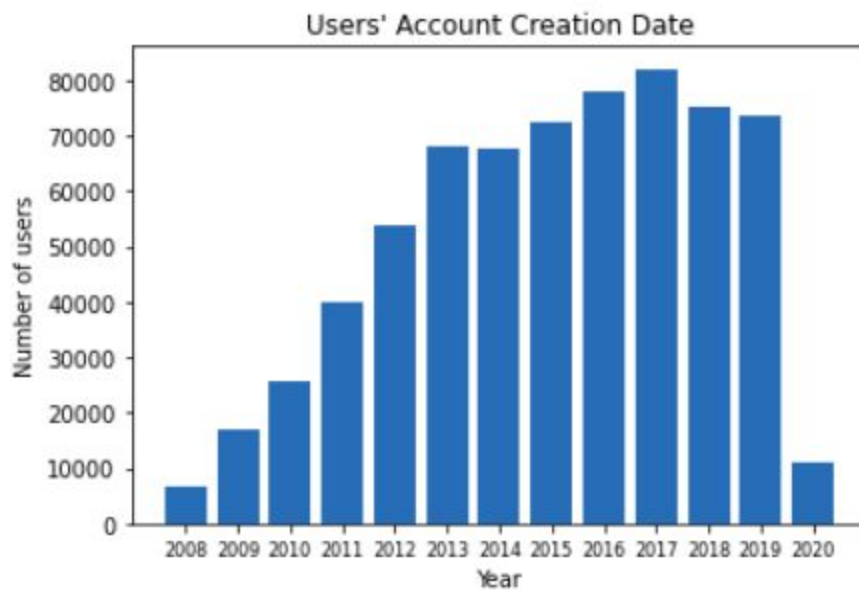
- Top 10 occurring tags in Tags XML and their counts are shown. The highest occurring tag is togglebutton.



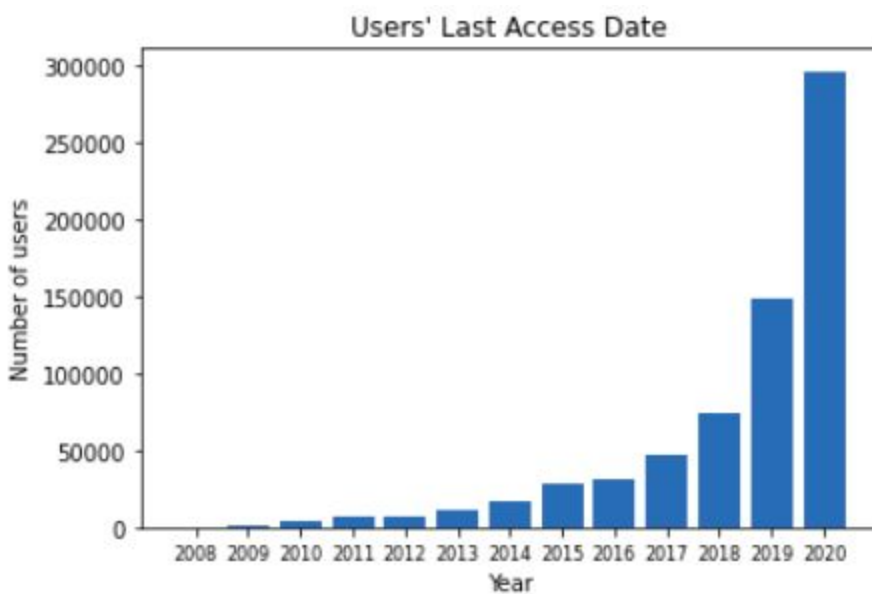
- Reputation of users mostly falls in the range 0-100000.



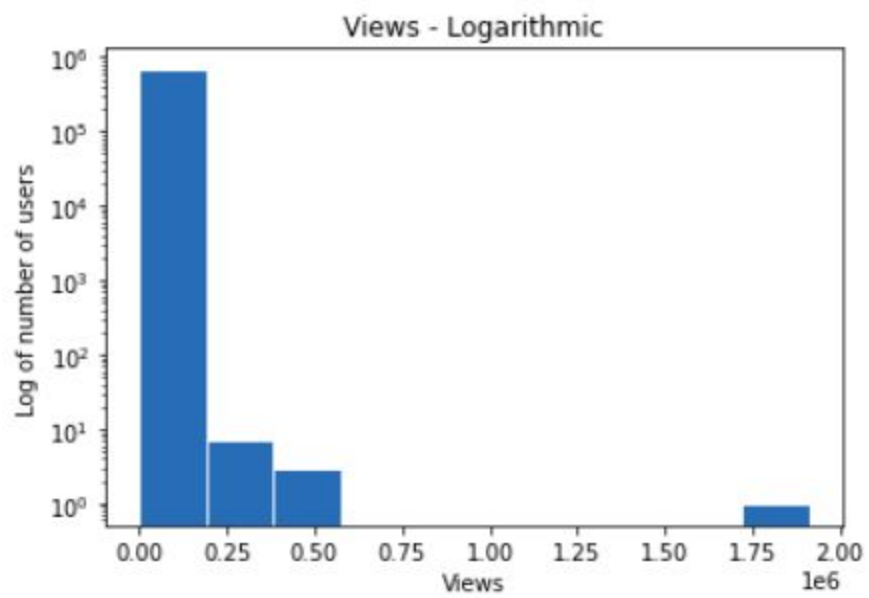
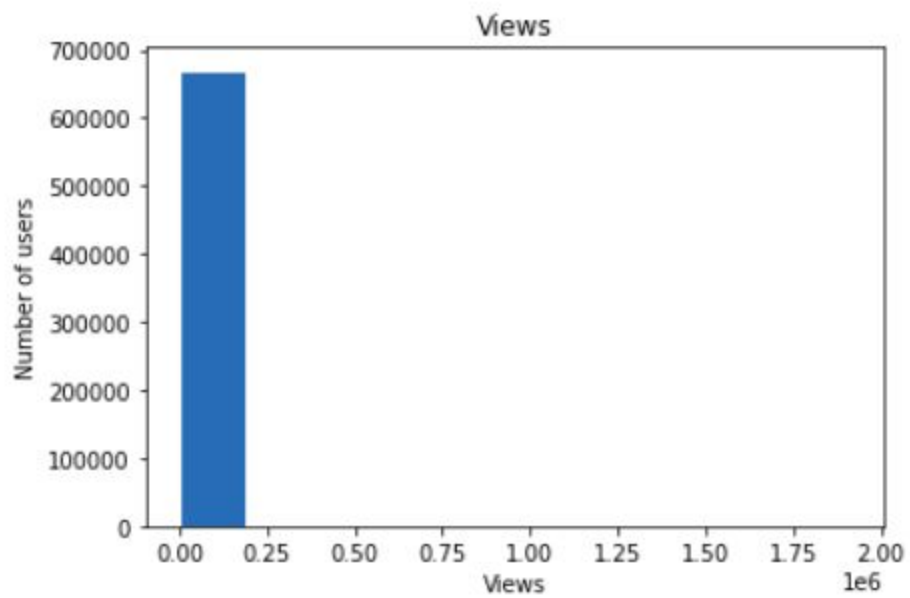
- User account creation date



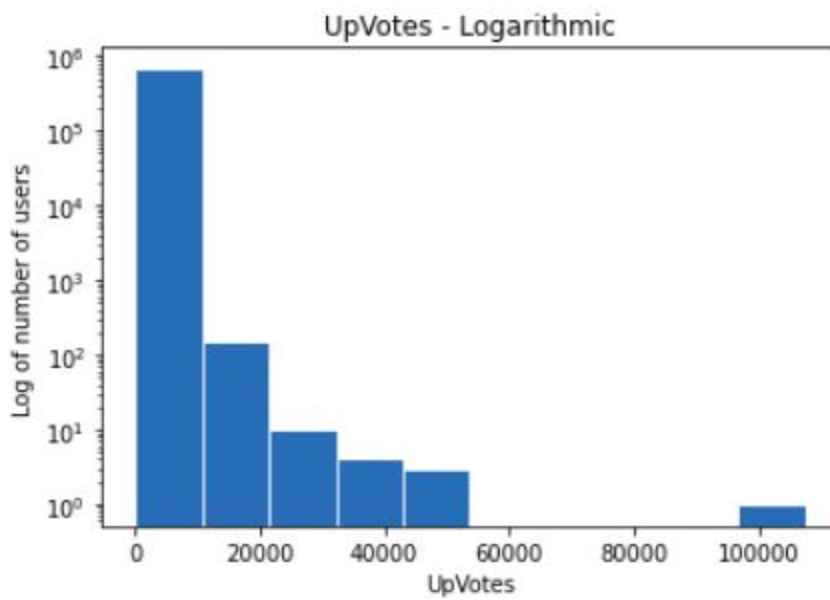
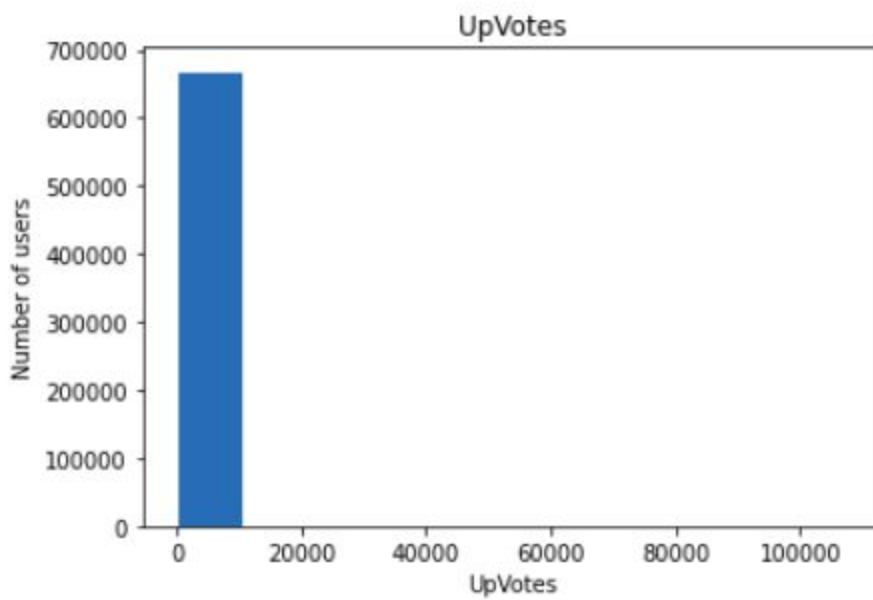
- User last access date



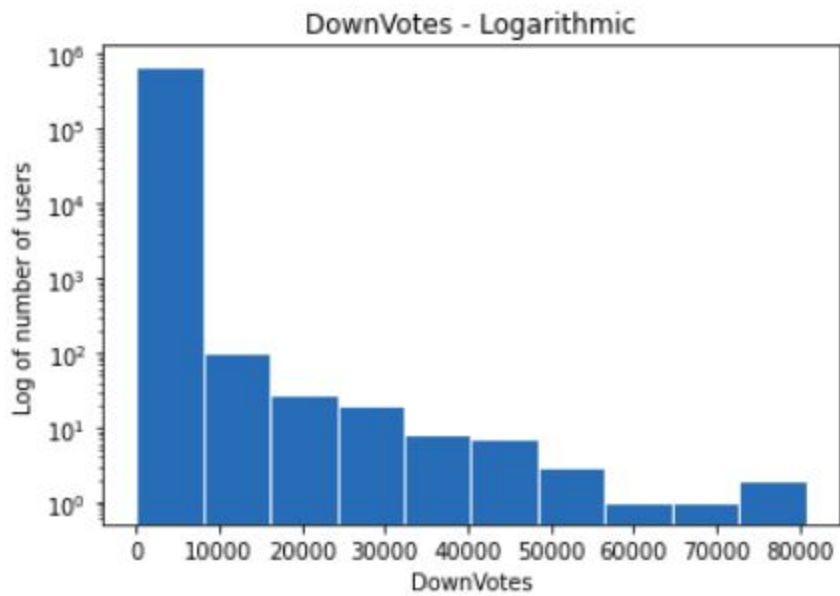
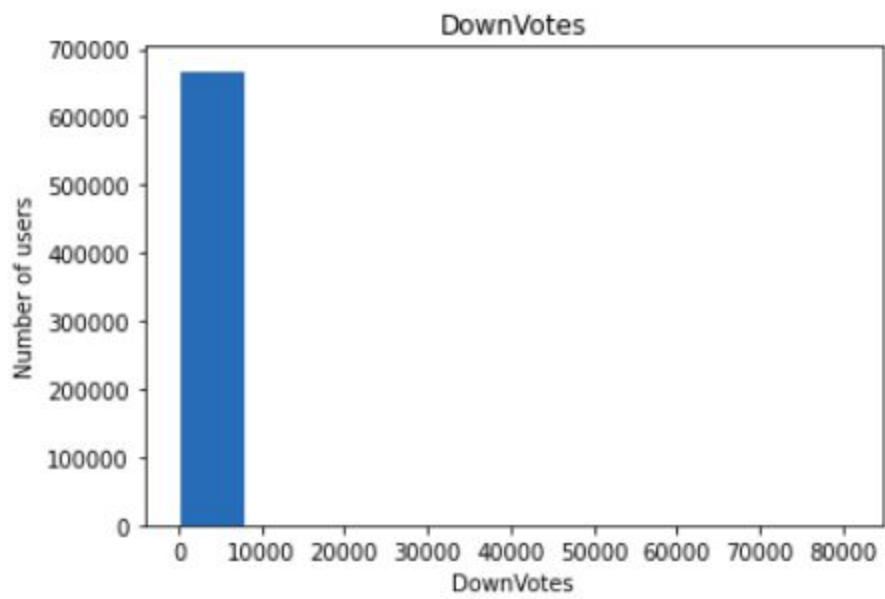
- Most user views fall in the range 0-200000



- Most upvotes fall in the range 0-10000



- Most downvotes fall in the range 0-8000

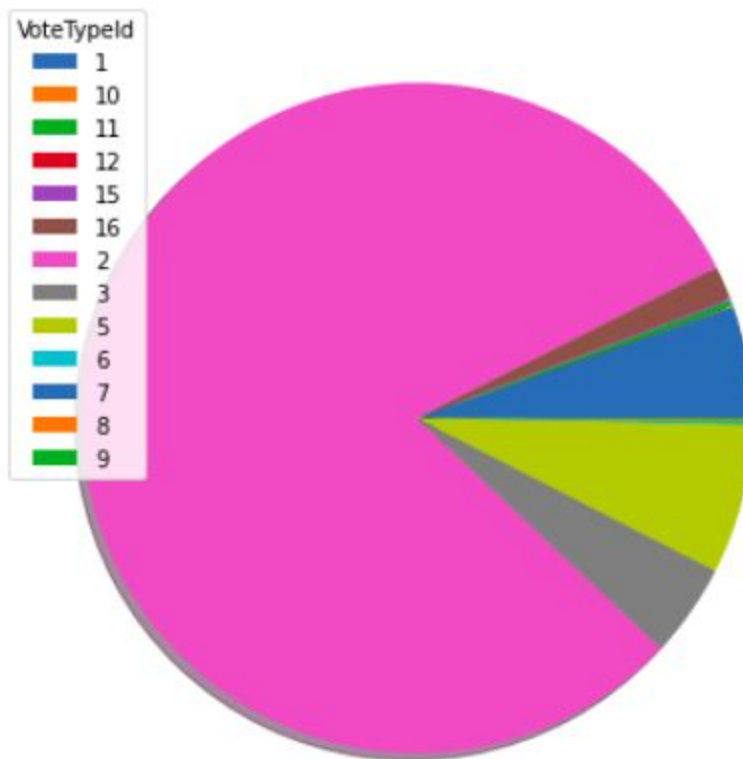


- Most users have a profile image.

Profile Image Present



- There are 16 type of vote type IDs and 2 is the most common.



- Vote creation date

