StackExchange-Miner

Group 26

● ● ●

Aaryan Srivastava	180007	aaryans@iitk.ac.in
Ashwin Shenai	180156	ashwins@iitk.ac.in
Utkarsh Gupta	180836	utkarshg@iitk.ac.in
Varun Goyal	180850	govarun@iitk.ac.in

Introduction

- StackExchange is a network of 173 Q&A communities, where people ask, answer and learn. It has over 3.2 Million questions and over 430 Million monthly user visits.

- The data was available for dates up until the 6th of September, 2021.

- The project aims to study patterns, trends and interesting observations from data obtained from any given Stack Exchange forum.
    - For the sake of the project, data is limited to a few forums, namely: Hinduism, Ethereum, Crypto, Datascience, Space and Islam.
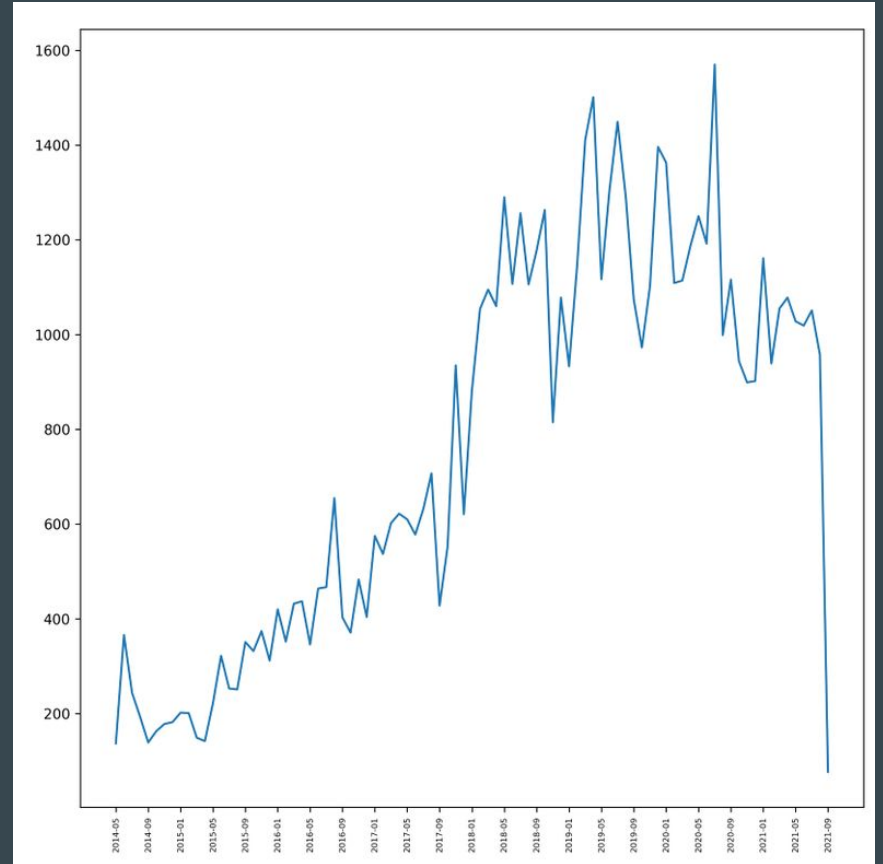
Interesting Insights

(The presentation has a few insights, kindly refer to the report and results for all insights)

Comment trends (per month)

Comments per month trend for Data Science Stack Exchange forum is shown in the adjoining image.

It clearly indicates an increasing activity for Data science forums, owing to its increasing popularity.

Though the last month shows a steep decline, it is due to incomplete data, as the data available for uptill the date 6th September 2021.
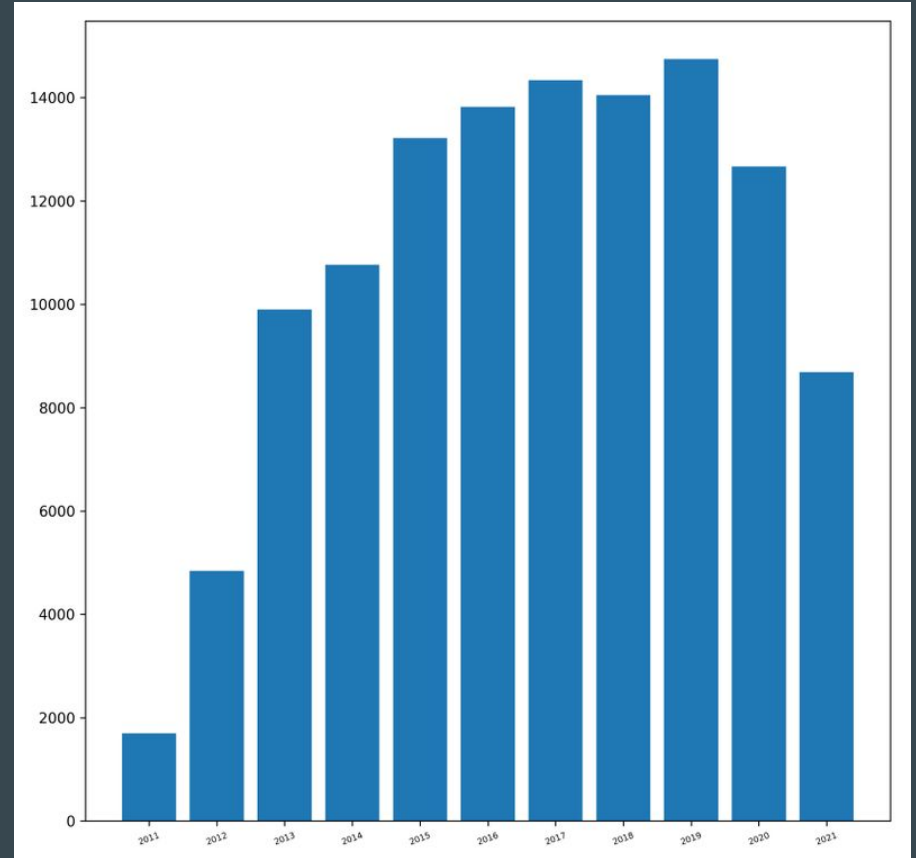
Comment trends (per year)

Comments per year pattern for Crypto Stack Exchange forum is shown in the adjoining image.

Surprisingly, the activity was appreciably high ever since 2015, even though there has been massive spike in popularity since 2020, after the COVID-19 was announced a global pandemic.

Though the last year shows a decline, it is due to incomplete data, as the data available for uptill the date 6th September 2021, only.
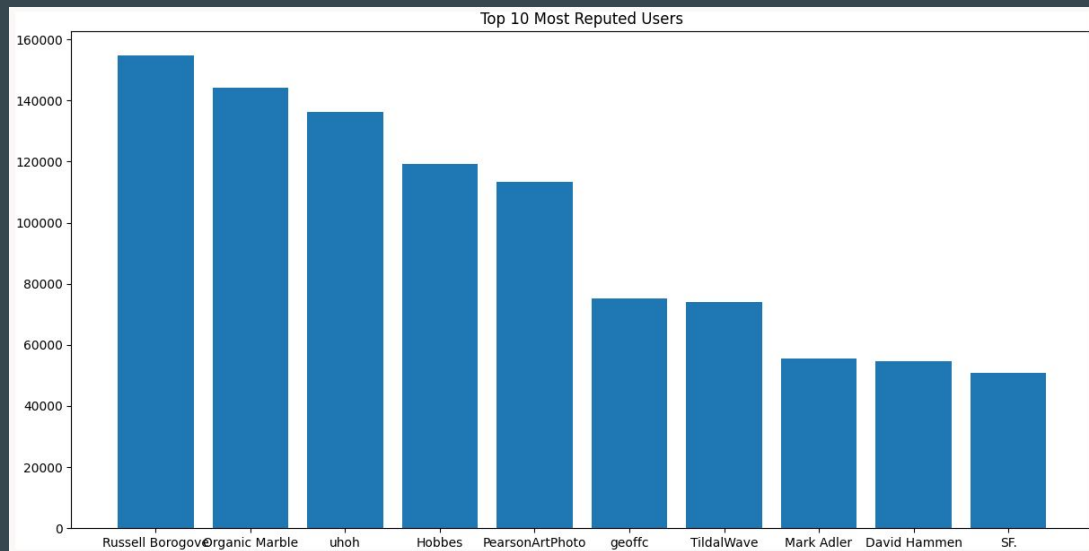
Top 10 users, ranked by reputation

Users can gain reputation by getting up-voted by other users based on their helpful posts, answers and comments.

Top 10 most reputed users for Space Stack Exchange forum is shown in the adjoining image.

The most reputed user is a certain <u>Russell Borgove</u>, who has 1250+ answers and has reached a whopping 3.2 Million people through his contributions.
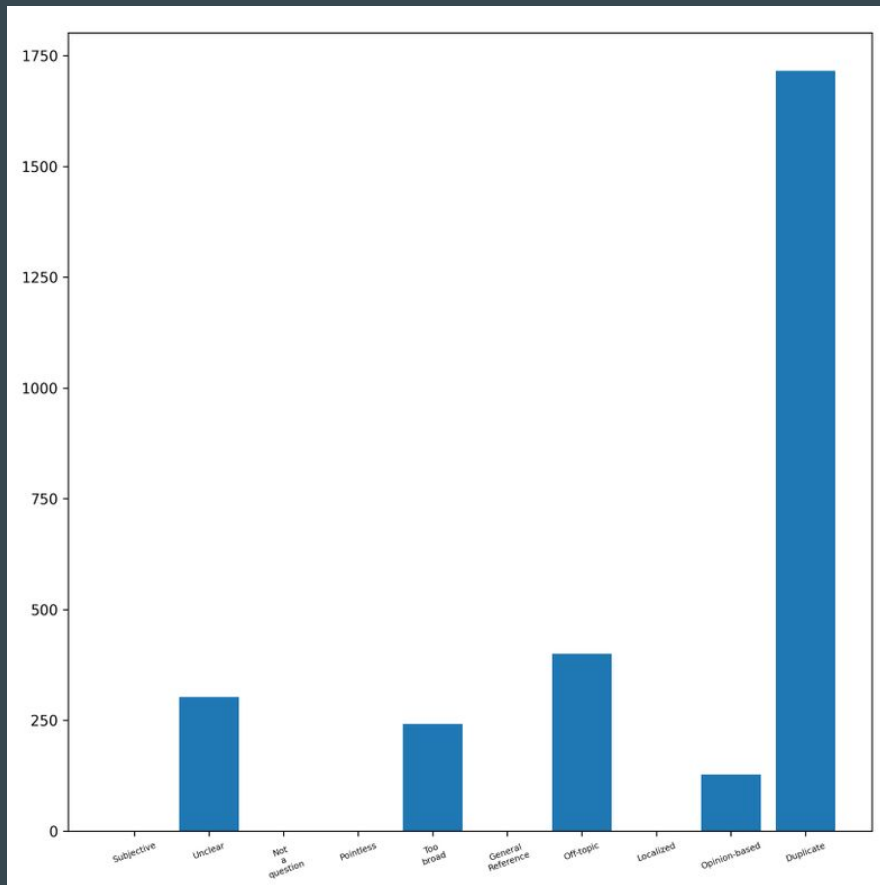
Why are posts closed?

Posts can be marked closed if they are not fit or relevant to the forum, for instance they may be possibly 'Duplicate' or 'Off-topic'.

The adjoining figure shows the frequency of closed posts corresponding to each reason for the Ethereum Stack Exchange.

As expected, the most often reason is them being duplicate, since users usually don't do their due diligence before asking a similar question
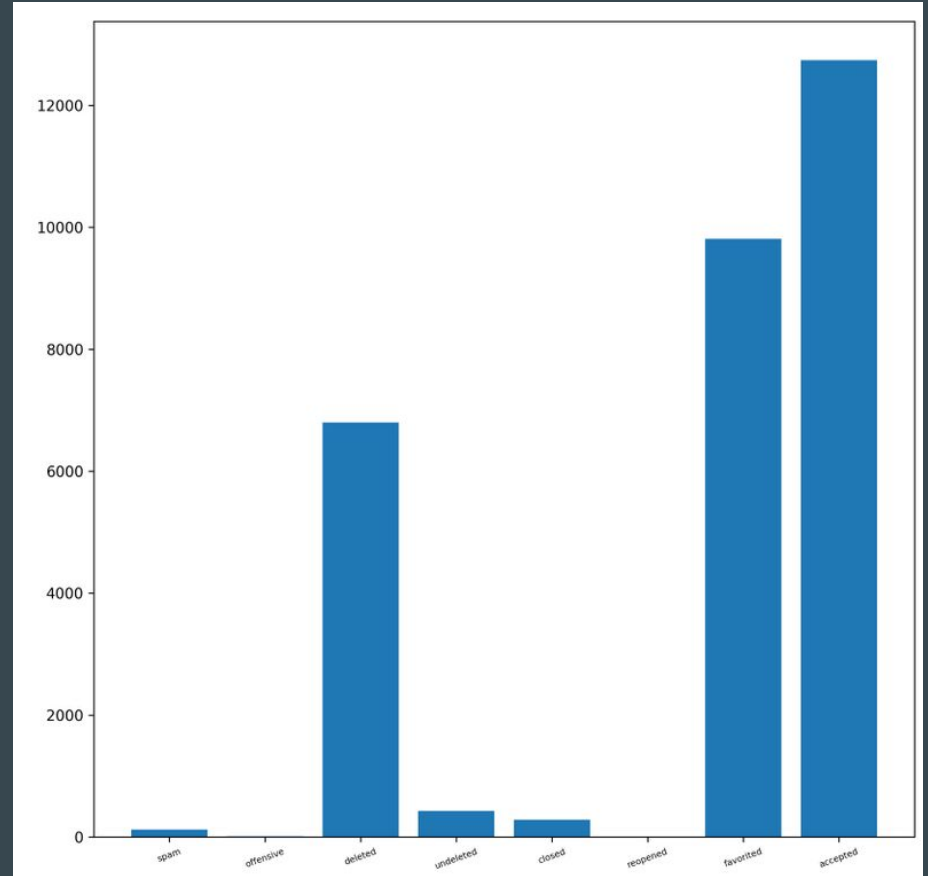
Various types of posts

Various posts are marked as accepted, spam, offensive, deleted etc. by the community.

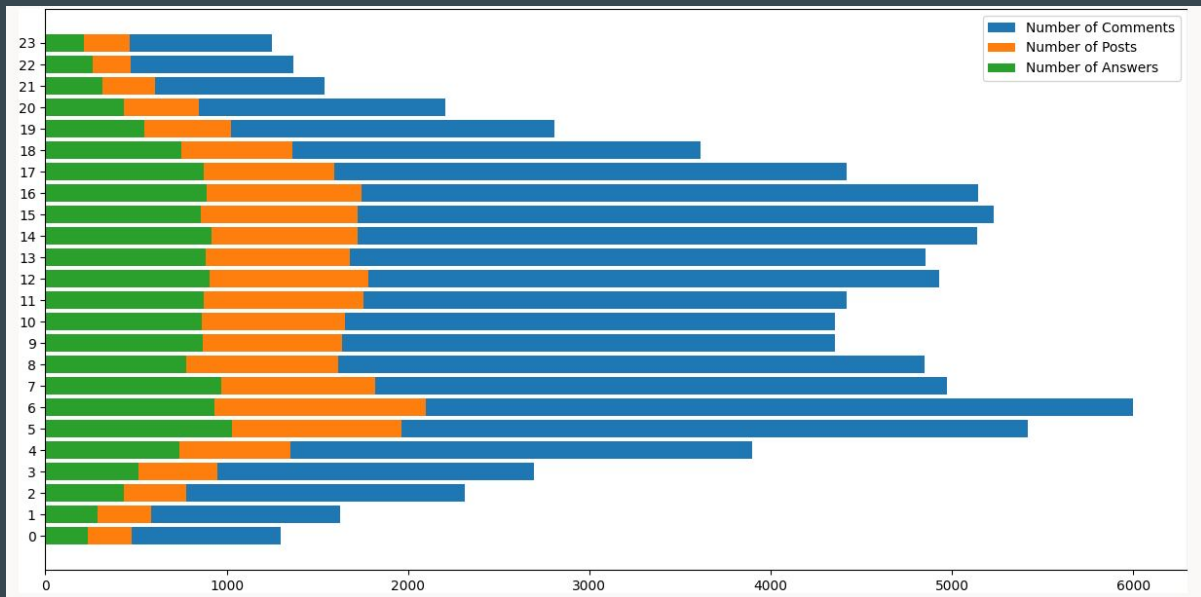The adjoining figure shows the frequency of posts corresponding to each category for the Islam Stack Exchange forum.

A large number of deleted posts show that there have been quite a few instances of unfit questions, which are then deleted by users or the community.

What is the best hour to ask a question?

The plot depicts the relative number of posts, comments and answers during different hours of the day for the Hinduism Stack Exchange forum (the hours are based on the GMT timezone).

We notice there is a high peak around 11:30 AM IST (6 AM GMT), and then there is a continuous active streak up until the night, around 10:30 PM IST (5 PM GMT).
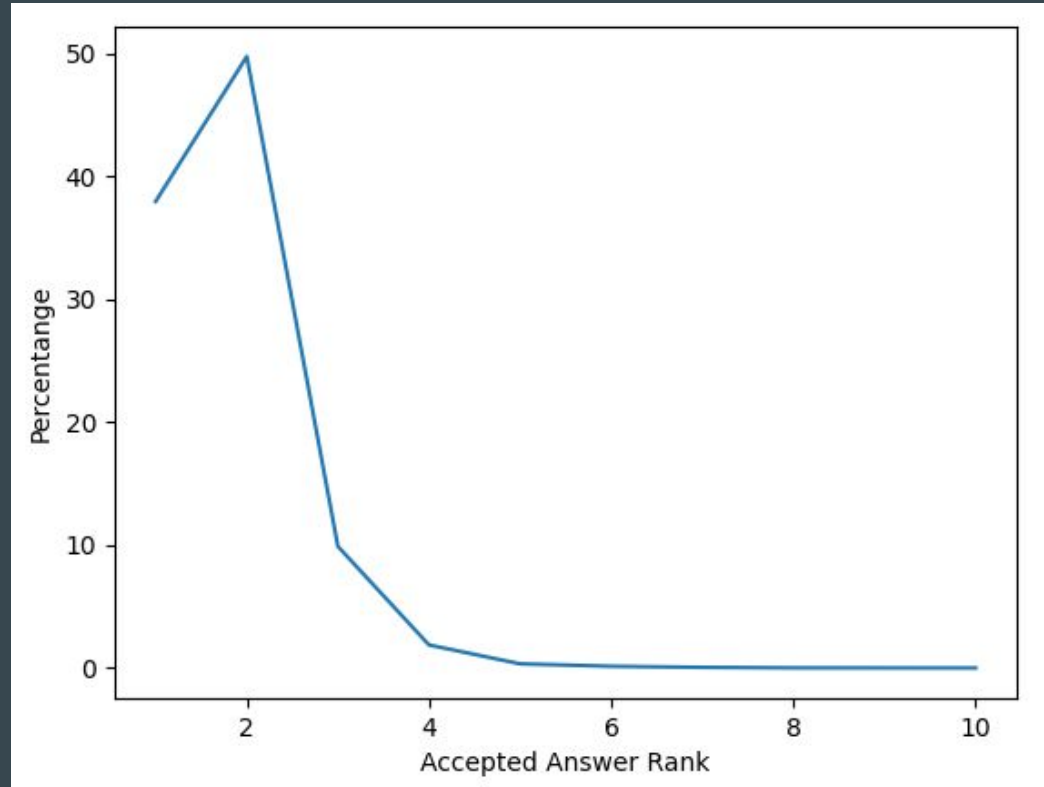
Fastest Gun in the West

It is noticed that most accepted answers are generally the earlier answer for the respective questions.

The adjoining figure is plot of the percentage of accepted answers versus their rank (by time) for the Space Stack Exchange forum.

Contrary to expectations, the second answer for a post is more accepted compared to the first answer for the Space Stack Exchange.

Association Rule Mining

Badges/Tags from each User/Post are encoded as a one-hot vector.

We detect frequent itemsets to find the groups of badges/tags that frequently occur together.

Association Rules are mined to see if presence of certain Badges/Tags imply the presence of others.

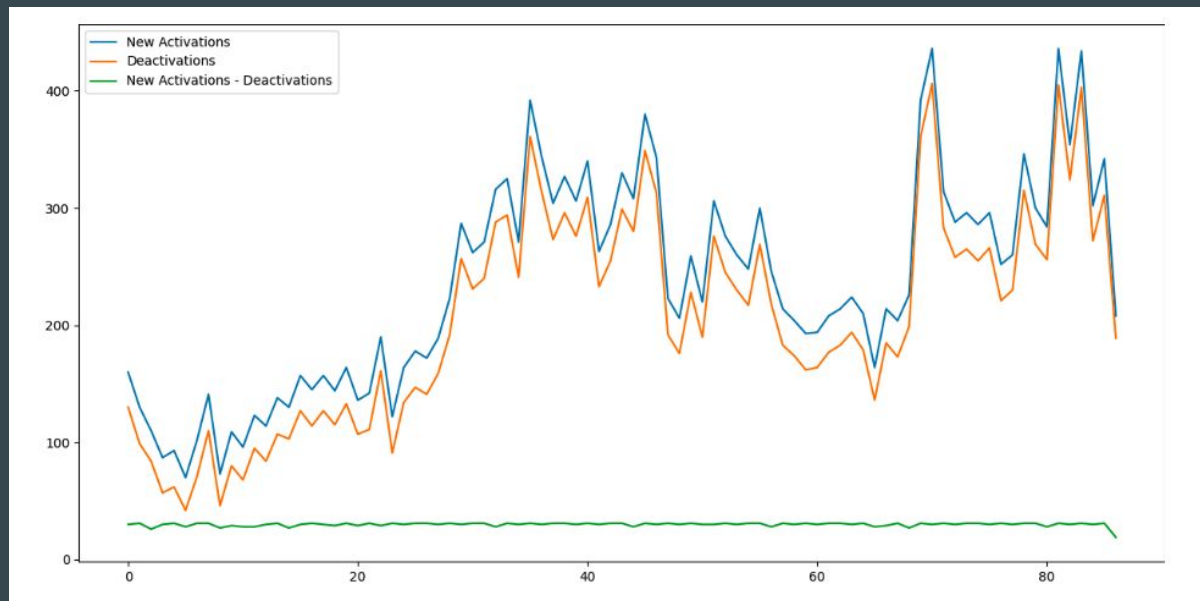| id	LHS	RHS	support	confidence	coverage	lift	count
360	{title,exaltation}	{non-attachment}	0.0012	0.3393	0.0035	33.4963	38
361	{teaching,prisoners}	{nimbarkacharya}	0.001	0.2667	0.0038	16.0056	32
362	{teaching,nimbarkacharya}	{prisoners}	0.001	0.4571	0.0022	32.7027	32
363	{nimbarkacharya,prisoners}	{teaching}	0.001	0.25	0.004	20.4722	32
364	{appayya-dikshitar,devtas}	{kumarila-bhatta}	0.0013	0.2642	0.005	5.0581	42
365	{appayya-dikshitar,kumarila-bhatta}	{devtas}	0.0013	0.25	0.0053	5.6031	42
366	{kumarila-bhatta,devtas}	{appayya-dikshitar}	0.0013	0.1826	0.0073	6.163	42
367	{lalita-sahasranama,devtas}	{kumarila-bhatta}	0.001	0.3626	0.0029	6.944	33
368	{kumarila-bhatta,lalita-sahasranama}	{devtas}	0.001	0.2171	0.0048	4.8658	33
369	{kumarila-bhatta,devtas}	{lalita-sahasranama}	0.001	0.1435	0.0073	3.9573	33

Some association rules for Hinduism SE tags
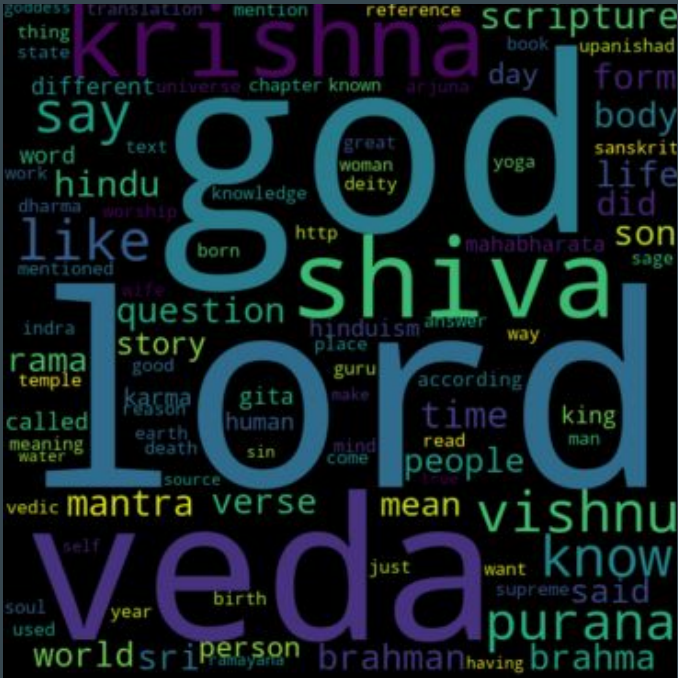
User Activations over time

The activations and deactivations keep varying over the months.

The adjoining figure shows new user activations, deactivations and net changes for the Hinduism StackExchange.
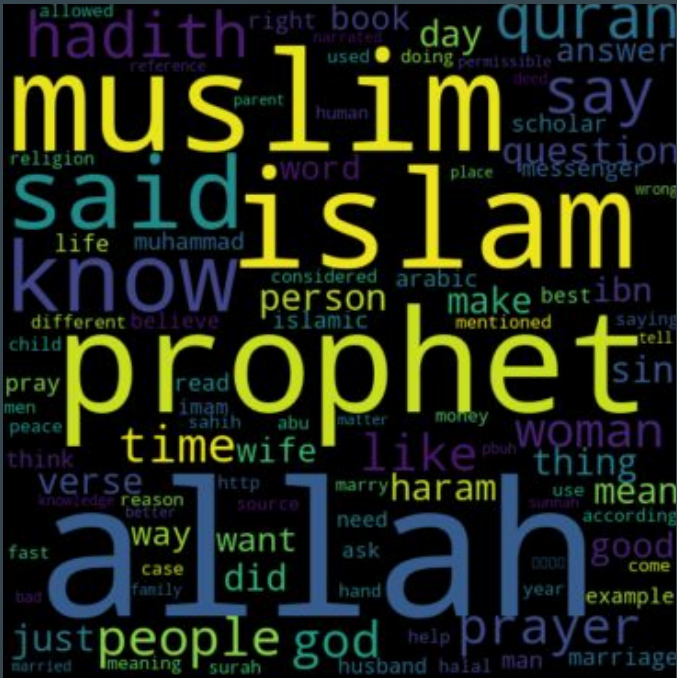
But we see an interesting trend in the number of users increasing (i.e. the green curve in the figure) is almost constant.

Few obtained world clouds

Word cloud for tags from the Crypto forum

Word cloud for tags from the Data Science forum

Few obtained world clouds (contd.)

Post body word cloud from Hinduism SE forum

Post body word cloud from Islam SE forum

Thank you!
...