Физтех-confessions, the most popular commentator of 2024

Jamclub[1]

[1]Физтех-confessions

15 декабря 2024 г.

Аннотация

Based on the number of likes each comment received, this report determines who is the most popular commentator of физтех-confessions. Comparison between the influence of the the top few commentators and the rest of the community is also made.

Содержание

1 Introduction and Motivation

2024 was a great year! Физтех-confessions was used:

1. To confess love and sins

2. As a means to procrastinate from doing work

3. To pass depression

4. To get daily dose of cringe because the brain demands it, just as our body demands 2,5L of water everyday

5. To share some brilliant memes.

We can obviously discuss the most liked posts, which will be a discussion on the created objects, but here I am more interested on the people, who keeps the community alive, so who is the most popular commentator of 2024?

2   Collection of Data

2.1   Basic statistics on the source

1. Time period = 08.05.2023 - 11.12.2024.

2. Number of posts analyzed = 5000.

3. Number of comments analyzed = 17753.

4. Number of authors of comments = 1778

An amateur code [1] was used to iterate through all posts backwards in time, starting from 11.12.2024 until 5000 posts were successfully read, and the number of likes on each comment and the author of the comment was recorded and put onto our main table. Instead of 2024, 1.5 year worth of data was used, because the физтех-confessions had change of admins and other breaks this year, and more data is always good for statistics.

2.2   Sources of errors

A python code was used to browse vk and collect data instead of vk.api. The code could not read some of the comments if the page of the post had a lot of comments. Because such comments are collapsed and the code was not smart enough to click on buttons and expand to read all comments. Since the sample size is still very large, such errors do not distort the conclusions. Vk.api would give correct data but it was not used due to task being too technically difficult in nature given the aim is to only guess the most popular commentator.

Definitions of the terms: serial: the rank when sorted based on count_likes.

rank: position based on a characteristic for a local list.

count_likes: the total number of likes an author received for the whole data.

count_comments: the total number of comments an author wrote.

density = count_likes / count_comments

h-index: the total number of comments - h, such that the number of likes in each of these comments is greater than or equal to h.

popularity = normalized_count_likes * normalized_count_comments * h-index

3   Results and Discussion

3.1   The most Liked commentator (count_likes)

Definitions:

- count_likes: the total number of likes an author received for the whole data.

- serial: the rank when the table is sorted based on count_likes.

- author: the commentator, or the author of a comment.

Таблица 1: Top-20 authors with most likes [4].

serial	author	count_likes
1	Физтех.Confessions	4116
2	Василий Андрианов	4041
3	Николай Сменилфамили	1863
4	Кафи Шаббир	1529
5	Княже Калыванович	1443
6	Кристина Юниксовна	1336
7	Адам Бушакур	1284
8	Андрей Авраменко	1187
9	Всеволод Ccfl	1130
10	Аврора Фантомхайв	1024
11	Владимир Гаврилов	899
12	Ева Астра	815
13	Даня Александров	805
14	Александр Попов	736
15	Даниил Кухмистров	717
16	Александр Губанов	711
17	Tigran Galstyan	702
18	Александр Логинов	689
19	Лариса Ретроградная	675
20	Петр Блинов	662

Таблица 2: Basic statistics on number of likes, comments and authors.

|	
Total number of likes	75 881
Total number of comments	17 753
Total number of authors	1778
Average number of likes per comment	4,27
Average number of comments per author	9,98
Number of likes received by the top-10 authors, (green rows in Table 5)	15 042
Number of comments written by the top-10 authors	2291
Average number of likes per comment for top-10 author	6,57
Percent of comments written by top-10 authors	12,9 %
Percent of likes by top-10	19,8 %
Number of likes received by the top-40 authors	35 536
Number of comments written by the top-40 authors	7846
Average number of likes per comment for top-40 author	4,5
Percent of comments written by top-40 authors	44,2 %
Percent of likes by top-40	46,8 %
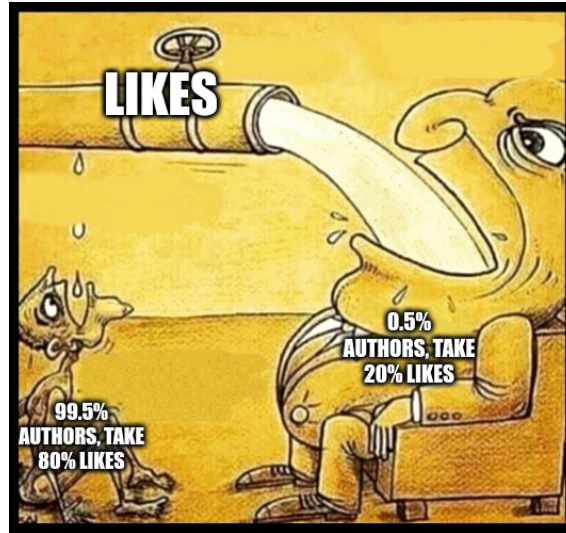
Рис. 1: A few authors take a lot of likes

We can also say that half of the comment activity comes from just 40 authors. Now for simplicity, let us consider only the top-500 authors, and will have a look at the distribution curve on how the likes are distributed amount these 500 authors. It is reasonable to take the top-500 instead of all 1778 because, the 500th author when sorted on count_likes had only 18 likes in total. We can assume that 501st author and on wards do not have much desire to get likes, therefore it will not be fair to add them to the distribution curve and conclude that the like distribution among the rich and poor is very large. By rich, of course we mean, those who took a large share of the total likes, and by poor who took a small share of the total likes. Also, the top-500 contribute to 93 % of the total likes, so completely removing the others does not leave a big proportion of the data out.
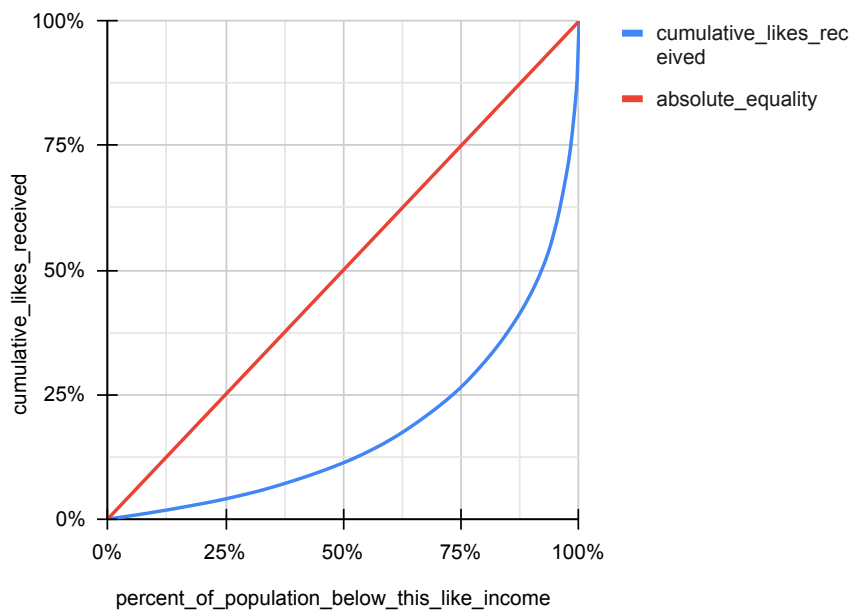

Рис. 2: Determination of gini-coefficient [5]

We have used 500 authors to form the gini-distribution, so the top 10 are the top 2 % wealthy men and the top 40 are the top 8 % wealthy men in физтех-confessions. We see that their like contribution is 26,7 % and 50,1 % respectively. This is considering the top-500 authors, when we compare this with table 2, which includes all the 1778 authors, the percent of contribution of likes is not very different [2].

We determine our inequality by the gini-coefficient [5], which is calculated from figure 2, here the gini coefficient is 0,64 which makes it similar to South Africa, which has the highest income inequality in the world. In other words in South Afria, the top 2 % richest men take 26,7 % of the country's annual wealth, and the top 8 % of the the richest men takes 50,1 % of the country's annual wealth, just as it is the case with физтех-confessions, except luckily it is not food, water and iron, but just the amount of adrenaline rush from receiving notifications of a comment being liked. For comparison, we can look at other countries:

Таблица 3: Gini-coefficients of various countries

rank	country	gini-coefficient
1	Физтех-confessions	0,64
2	South Africa	0,63
3	USA	0,39
4	Russia	0,36
5	Sweden	0,29
6	Norway	0,23

However note that we talk about South Africa based on our curve of физтех-confessions since we have almost the same gini-coefficient, but the exact distribution for the top 2 % and the top 8 % can be slightly different because gini-coefficient is due to the area between the red and the blue curve, this coefficient can be same due to the same area but actual blue curves can be slightly different.

3.2   The most Depressed commentator (count_comments)

Definitions:

- rank: position based on a characteristic for a local table, while the serial is always the rank when the table was sorted according to count_likes.

- count_comments: the total number of comments an author wrote.

- sr-difference = serial - rank is the difference between the the positions in the table of count_likes and int this table. We would want to know if we change the measure based on which we sort the table how much does it differ from the table which was sorted based on the number of likes.

Таблица 4: Top-20 authors with most comments, [3].

rank	sr-difference	author	serial	count_likes	count_comments
1	0	Физтех.Confessions	1	4116	2128
2	0	Василий Андрианов	2	4041	649
3	+14	Tigran Galstyan	17	702	481
4	+15	Лариса Ретроградная	19	675	392
5	-1	Кафи Шаббир	4	1529	355
6	+4	Аврора Фантомхайв	10	1024	287
7	+1	Андрей Авраменко	8	1187	230
8	+21	Ярослав Аммосов	29	499	202
9	+9	Александр Логинов	18	689	173
10	-5	Княже Калыванович	5	1443	172
11	-8	Николай Сменилфамили	3	1863	172
12	-1	Владимир Гаврилов	11	899	169
13	-1	Ева Астра	12	815	167
14	+6	Петр Блинов	20	662	152
15	-2	Даня Александров	13	805	147
16	+10	Станислав Шушкевич	26	528	125
17	-2	Даниил Кухмистров	15	717	112
18	-12	Кристина Юниксовна	6	1336	111
19	+6	Алина Куринная	25	528	103
20	+12	Камиль Калиновский	32	420	103

Certainly the most depressed student in физтех writes the most amount of comments. And now by the legendary sr-difference we see that Tigran Galstyan has moved up 14 places as compared to Table 5. Лариса Ретроградная has also moved up a massive amount of 15 places it is because her comments are self contradictory, trolling or just toxic. Ярослав Аммосов has moved up many places it is because his comments are usually emojis or compliments appreciating the post or a comment of another author. Кафи Шаббир has moved down one place which indicates that his ratio of quality over quantity if not neural slightly good for the community. Николай Сменилфамилиюнавзрослуюсерьёзную has gone down by 8 places which means his quality over quantity is pretty high.

3.3 The most Sigma commentator (density)

Таблица 5: Top-20 authors with most likes. Full table here [4].

serial	author	count_likes
1	Физтех.Confessions	4116
2	Василий Андрианов	4041
3	Николай Сменилфамили	1863
4	Кафи Шаббир	1529
5	Княже Калыванович	1443
6	Кристина Юниксовна	1336
7	Адам Бушакур	1284
8	Андрей Авраменко	1187
9	Всеволод Ccfl	1130
10	Аврора Фантомхайв	1024
11	Владимир Гаврилов	899
12	Ева Астра	815
13	Даня Александров	805
14	Александр Попов	736
15	Даниил Кухмистров	717
16	Александр Губанов	711
17	Tigran Galstyan	702
18	Александр Логинов	689
19	Лариса Ретроградная	675
20	Петр Блинов	662

The number of likes and the number of comments certainly does not give us all the information. Possibly there are sigma's who keep silent, but when they speak something, it caries a lot of value.

How let us look at the distribution of the number of comments and likes.

Here Физтех.Confessions and Василий Андрианов were excluded to keep the scales reasonable of axes reasonable. We can see that the number of likes and the number of comments deviates significantly from the trend line.

3.4 The most cringe commentator (reverse density)

We can possibly divide the count like and count comment graph into 4 zones.

3.5 The best Female commentator

3.6 The most Productive Commentator

H-index depends on the number of likes and the comment density

4 Conclusion

This report was prepared without the consent to publish name or profile pictures of the commentators. If you have a problem with it, you can cry about it.

Список литературы

[1] scrape.py. https://drive.google.com/file/d/1x_XWGbxKHCPxmJo7_cfyQRLmpgFyUDie/view?usp=drive_link.

[2] Table calculation of gini coefficient. https://docs.google.com/spreadsheets/d/1m4j6RICIfq6DIdJ1OtQPb95STsEhHjRoEMEKSS9QAVE/edit?gid=1472074717#gid=1472074717.

[3] Table count of comments. https://docs.google.com/spreadsheets/d/1m4j6RICIfq6DIdJ1OtQPb95STsEhHjRoEMEKSS9QAVE/edit?gid=1821382722#gid=1821382722.

[4] Table count of likes of top authors. https://docs.google.com/spreadsheets/d/1m4j6RICIfq6DIdJ1OtQPb95STsEhHjRoEMEKSS9QAVE/edit?gid=1373352226#gid=1373352226.

[5] Коэффициент Джини — Википедия — ru.wikipedia.org. https://ru.wikipedia.org/wiki/%D0%9A%D0%BE%D1%8D%D1%84%D1%84%D0%B8%D1%86%D0%B8%D0%B5%D0%BD%D1%82_%D0%94%D0%B6%D0%B8%D0%BD%D0%B8. [Accessed 14-12-2024].