DATA REPORT · NOV 2026

17 YEARS · STORIES · COMMENTS · TAGS

The Shahvani Corpus

A data portrait of 37,837 stories, 536,919 comments and 710 tags: what people wrote, what they read, and how they talked back, from 2009 to 2026.



Minor-related words omitted. Size follows frequency.

37,837
stories

536,919
comments

74.1M
words analysed

2.8B
story views

CONTENTS

What is in this report

Sections follow the data: stories first, then language, comments, the other tables, deeper analyses and method.

Section	Title	Page
01	Data anatomy	4
02	Growth	5
03	Attention	6
04	Rhythm	9
05	Tags	10
06	Story anatomy	14
07	Language	15
08	Language over time	18
09	Comments	22
10	Accounts	25
11	Forum	29
12	Photos	32
13	Listings	34
14	Story lifetime	35
15	Retention	37
16	Commenter network	38
17	Series	39
18	Style	40
19	Tag profiles	41
20	Then and now	45
21	Predicting attention	46
22	Method	48

00 · SUMMARY

Executive summary

A data portrait of 37,837 stories and 536,919 comments, from 2009-10-21 to 2026-11-16. Numbers are computed from the database snapshot; none are typed in.

37,837	536,919	2.8B	74.2M
stories	comments	story views	Persian words

Key findings

1. **Attention is concentrated, but less than in many media.** The top 1% of stories hold 9.1% of all views and the top 10% hold 38.8% (Gini 0.53). The median story has 43,701 views.
2. **Readers are a small, loyal core.** The top 1% of 65,665 commenting accounts wrote 34% of comments and the top 10% wrote 71%, while 45% of accounts commented only once.
3. **Conversation is quick.** Half of all comments arrive within 15.4 hours of the story, 64% within a day and 86% within a week.
4. **Quality is not what makes a story popular.** Story length barely tracks view percentile (Spearman 0.05); likes and dislikes both track it far more (0.32 and 0.40), because both are signs of being seen.
5. **Publishing peaked recently.** The busiest month for new stories was Aug 2026 (386 stories). The record also has a 15-month silent stretch, Jul 2019 to Sep 2020, which is a gap in the data rather than a quiet period.
6. **Accounts are mostly men.** Of 234,077 profiles, 89% state male and 8.7% female. The largest province holds 38% of province-stated profiles against 17% of the population; see section 10.
7. **The forum is a place to ask, and to name a city.** 303,977 topics in 30 boards from 2009 to 2026. On the request boards, 46% of 39,557 topic titles name a place (section 11).
8. **Some fields do not mean what their names suggest.** 100% of the stories with a later last-activity stamp share one single day (2019-06-19), so the stamp cannot measure how long a story stays alive. Views per day is used instead (section 14).
9. **The language shifts with the platform.** Comment style changes over the years (emoticons, emoji, Latin script, laughter), and vocabulary trends track cultural events; see sections 8 and 9.
10. **Most commenters visit once.** 53% of 65,496 accounts never comment on a second day, and 14% are still commenting a year after their first comment (section 15).
11. **Stories come in parts, and attention thins out.** 1,939 series were found from part numbers in titles (14% of stories). The median part 2 has 79% of the views of part 1 (section 17).
12. **Writing style is stable, with small shifts.** 37,294 stories were measured for sentence length, pronouns, spoken forms and dialogue; none of these tracks popularity strongly (largest |Spearman| 0.12, section 18).
13. **Voting changed character, so vote counts do not span the whole record.** Likes per story rise from 1.7 in 2015 to 7.3 in 2016, and the share of stories with no vote falls from 41% to 1% (section 19).
14. **Tags differ in length and in how fast readers answer.** Among tags with enough stories, the longest median story is under انتقام (10,953 characters) and the shortest under جلق (3,464); the fastest first comment comes under مادرزن (median 0.3 hours), section 19. Section 20 compares the first and last years of the record.
15. **Tags predict attention; little else does.** A gradient-boosting model explains 51% of the variance in view percentile on unseen stories, and tags alone give 44%. Body length, hour and weekday add almost nothing (section 21).

Read with care. 57% of commenting accounts have no profile row, 0.5% of comments were flagged as spam or duplicates (kept in volume counts, left out of word statistics), and counts such as views are as of the snapshot date.

01 · DATA ANATOMY

What the database holds

A single snapshot of the site, read as is. 37,837 stories and 536,919 comments carry the core analysis; profiles, forum topics, photos and listings follow in sections 10 to 18.

37,837
stories

536,919
comments

65,665
distinct commenters

74.2M
Persian words in stories and comments

Table	Rows	First day	Last day
comments	536,919	2010-04-02	2026-11-16
forum topics	303,977	2009-11-04	2026-11-16
profiles	234,077	2015-12-01	2026-11-16
photos	63,885	2010-09-07	2026-11-15
story tags	59,766	n/a	n/a
stories	37,837	2009-10-21	2026-11-15
tags	710	n/a	n/a
sex worker listings	77	n/a	n/a

Known quirks

Finding	Count
Stories with a negative like count (clipped to 0)	142
Stories with a negative dislike count (clipped to 0)	19
Comments with a negative like count (clipped to 0)	8
Comments dated before their story (excluded from lag figures)	254
Comments with an empty body	23
Stories without any tag	1,873
Stories without an audio file	674

Read with care. 57% of commenters (227,163 comments) have no profile row, so profile-based figures cover only part of the commenters. Profile columns that are empty in this snapshot: about me, birth year, height cm, is featured, is vip, looking for, marital status, profile view count, sexual orientation, weight kg. Counts of views, likes and dislikes are as of the scrape. Hour-of-day figures use one fixed Tehran offset.

02 · GROWTH

Stories and comments per month

The corpus spans 2009 to 2026. Stories peaked in Aug 2026 (386); comments in Feb 2013 (6,919).

386
stories in the busiest month (Aug 2026)

6,919
comments in the busiest month (Feb 2013)

2023
busiest year for stories (3,287)

Stories published per month



Numbered lines: events listed in section 2.

Comments posted per month



Comments are counted by their own publication date. All comments are included, spam too. Numbered lines: events listed in section 2.

Outside events marked on the timelines

#	Date	Event
1	2018-04-30	Telegram blocked in Iran (judiciary order)
2	2018-07-17	Yahoo Messenger shut down
3	2019-11-16	Nationwide internet shutdown during the November protests
4	2020-02-19	First confirmed Covid-19 cases in Iran
5	2022-09-16	Death of Mahsa Amini and the protests that followed
6	2024-12-24	WhatsApp and Google Play unblocked
7	2025-06-13	Israel-Iran war begins (internet restrictions)

From the public record, not from the database; the dates are approximate to the day and worth checking before the report is shared.

Note. The newest month (Nov 2026) can be incomplete because the snapshot ends on 2026-11-16. The longest silent stretch is 15 months, Jul 2019 to Sep 2020, with no stories and no comments at all; that is a gap in the record, not a quiet period.

03 · ATTENTION

How views are distributed

Views are heavily skewed: the top 10% of stories hold 38.8% of all views, and the Gini coefficient is 0.53.

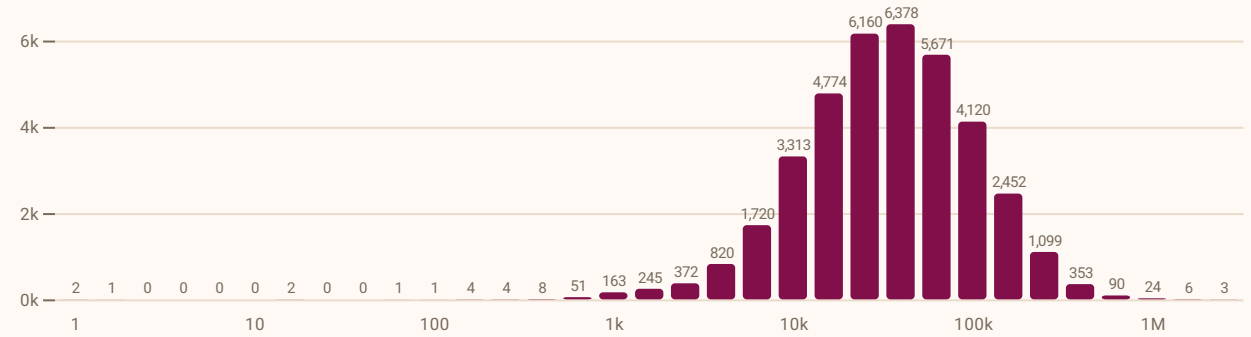
43,701
median views per story

72,823
mean views per story

3,894,401
most views on one story

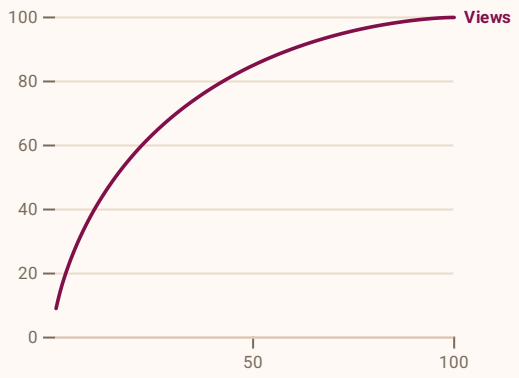
9.1%
views held by the top 1%

Stories by number of views (log scale)



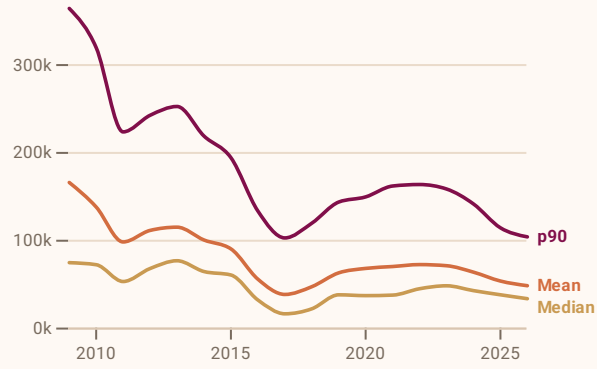
Each bar covers a fifth of a power of ten, labelled by its lower edge.

Concentration of views



Horizontal: top % of stories. Vertical: % of all views they hold. A straight diagonal would mean equal views.

Views per story by publication year



Older stories have had longer to collect views.

Reading it. The median story from 2009 has 75,035 views against 34,001 for 2026. Percentile figures later in the report rank stories within their own year to remove this age effect.

03 · ATTENTION

Hall of fame

The most viewed, most liked and most divisive stories. Minor-adjacent titles are hidden from every list.

Most viewed

#	Title	Year	Views	Likes
1	همه برای مهدیس	2021	3,894,401	250
2	بهترین سکس خانوادگی ما	2012	3,306,114	30
3	اولین سکس من - زن صاحب خانه	2011	2,639,994	7
4	سکس با خاله جون در حمام	2012	2,356,405	26
5	کس دادن زنم به پسر همسایه	2009	1,893,317	9
6	سکس من و پدرشوهرم	2009	1,891,338	43
7	بیا با هم بازی کنیم	2021	1,839,001	251
8	سکس با خواهرم تو حموم	2011	1,832,533	19
9	سکس مهیج با مامان	2013	1,699,729	29
10	کون مامانم و خواهرجونم	2015	1,477,988	32

Most liked

#	Title	Year	Views	Likes
1	میشه خاله صدات کنم؟	2022	512,801	635
2	مامان پری و پارسا (۳)	2023	679,201	484
3	مامان پری و پارسا (۴)	2023	573,201	451
4	سکس یواشکی منو خواهرم (۱)	2024	669,801	443
5	تو جنده‌ی خودمی	2021	590,201	413
6	مامانم میشی؟ (۲)	2023	279,801	382
7	سکس یواشکی منو خواهرم (2)	2024	645,601	381
8	راز مشترک من و برادرشوهرم	2021	522,501	343
9	مامانم میشی؟ (۳)	2023	243,501	330
10	مامانم میشی؟ (۱)	2023	386,701	328

Most divisive

#	Title	Year	Views	Like share
1	بدهی شوهرم و منو دوتا دختر دسته گلم (۱)	2023	483,601	58%
2	پسر صاحبخونه پردمو زد و جندم کرد (۱)	2025	131,901	39%
3	ضربدری تو کون خواهر زن و باجناق	2017	409,227	37%
4	من و خواهر ورزشکارم	2023	453,501	69%
5	فرشته زن جنده من و سفر به روستای پدری	2023	292,801	73%
6	به شاگردم دادم	2024	458,001	76%
7	خیانت یهویی ولی شیرین (1)	2024	321,501	77%
8	محرم، حلال شد	2026	306,501	78%
9	زن عمویی که برای گاییده شدن خوب تربیت شد	2025	255,901	79%
10	ماجراهای مامانم و کون دادن من (۴)	2020	381,301	81%

Stories with at least 200 votes, ranked by how close the like share is to 50%.

03 · ATTENTION

Top story of each year

The most viewed story among those published in each year.

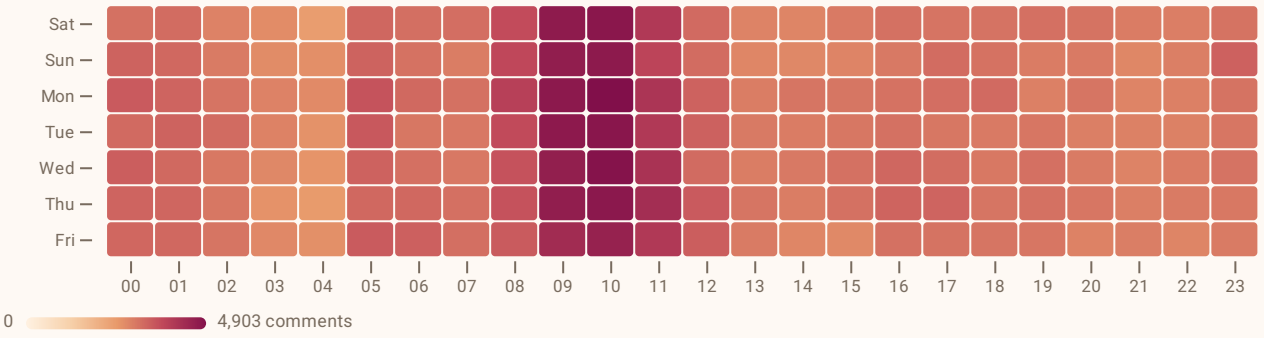
Year	Title	Views	Likes
2009	کس دادن زنم به پسر همسایه	1,893,317	9
2010	آریا و مامان و خاله جون	1,377,464	70
2011	اولین سکس من - زن صاحب خانه	2,639,994	7
2012	بهترین سکس خانوادگی ما	3,306,114	30
2013	سکس مهیج با مامان	1,699,729	29
2014	سکس با پسرم	1,301,405	42
2015	کون مامانم و خواهرجونم	1,477,988	32
2016	داداشم و گائیدن کسم	563,038	29
2017	اجازه دادم زنم جلو خودم کوس بده	723,526	58
2018	منو خواهرم زهرا در خانه	1,021,049	41
2019	مامان و خاله مهناز، شب عروسی پسردایی	1,078,587	26
2020	واقعا مامانمو کرد	538,101	88
2021	همه برای مهدیس	3,894,401	250
2022	پدر	1,036,701	172
2023	مامان پری و پارسا (۳)	679,201	484
2024	سکس یواشکی منو خواهرم (۱)	669,801	443
2025	آرش و مامان سهیلا	465,301	246
2026	کارای ناجور توی حموم	425,701	127

04 · RHYTHM

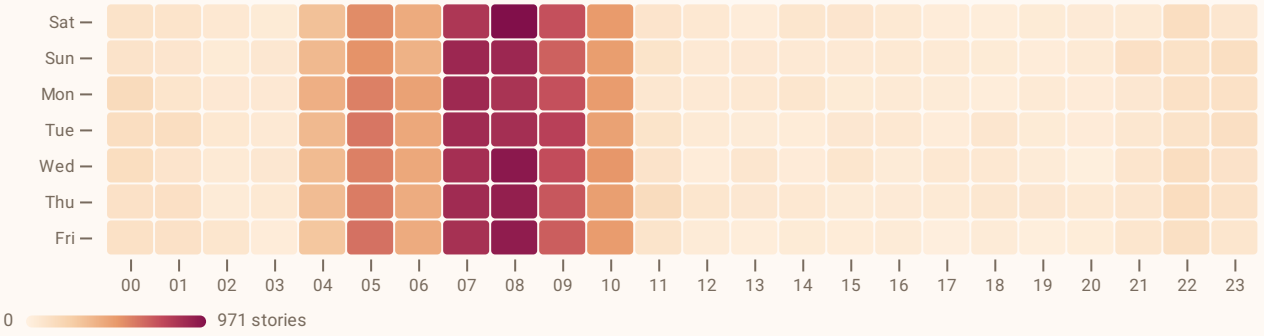
When the site is active

Comments peak at 10:00 (6.2% of all comments); stories at 08:00 (16.5%).

Comments by weekday and hour



Stories by weekday and hour



Method. Hours are in approximate Tehran time: the stored timestamps are converted from UTC with a fixed +03:30 offset, so daylight saving is ignored. The busiest weekday for comments is Mon and for stories Thu; the busiest day has 1.04 times the comments of the quietest, and 1.07 times the stories.

05 · TAGS

The shape of the tag universe

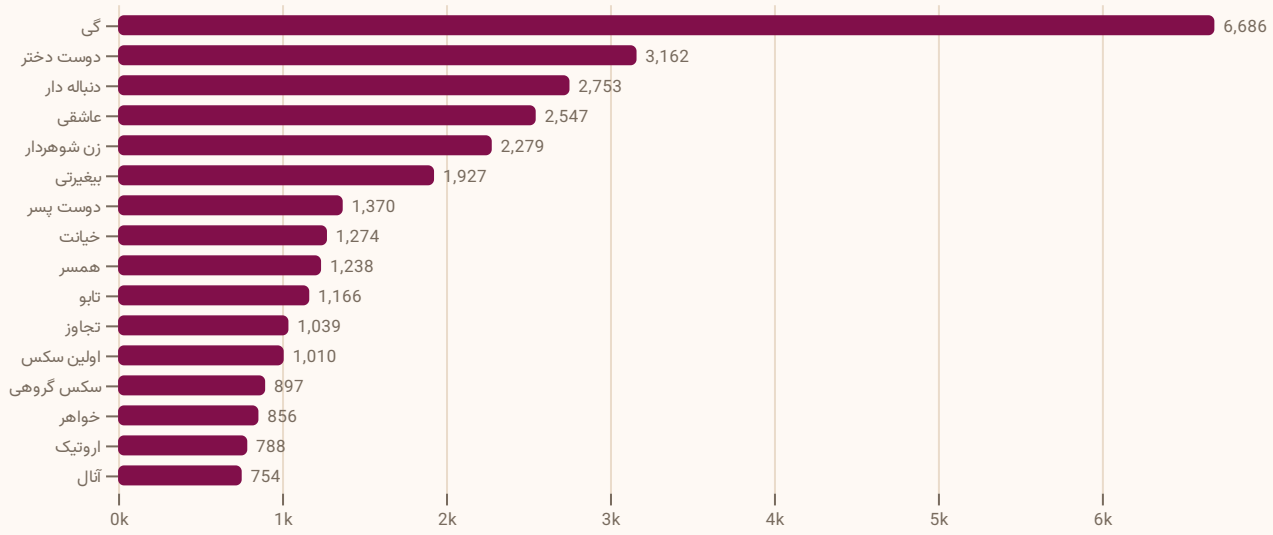
710 tags label 35,964 of 37,837 stories. The ten most used tags account for 41% of all 59,766 story-tag links.

710
tags in use

1.66
tags per tagged story

67
tags with at least 150 stories

Most used tags



Tag cloud



Reading it

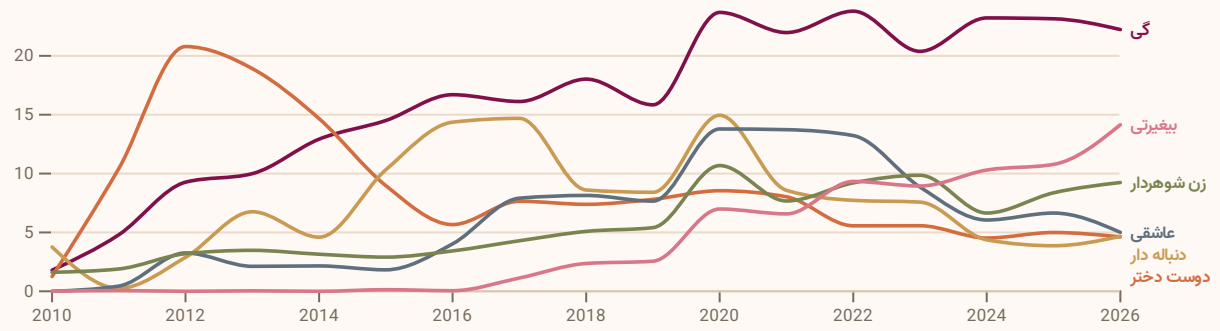
Tag names are category labels, not corpus vocabulary. Tags that touch minors are excluded everywhere in this report, so counts here can be lower than the site's.

05 · TAGS

Tags over time: risers and fallers

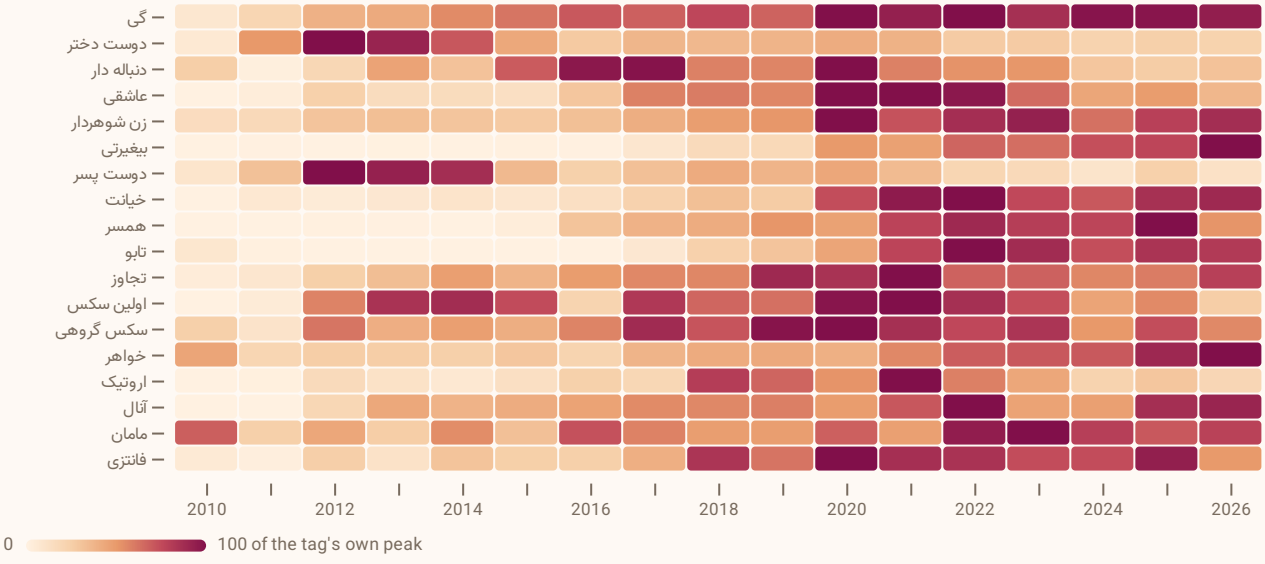
Share of stories carrying each tag, comparing 2010 to 2012 with 2024 to 2026.

Share of stories per year, 6 most used tags



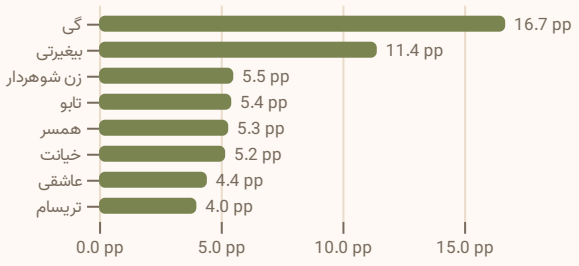
Vertical: % of that year's stories. Years before 2010 have fewer than 300 stories and are left out.

When each tag peaked: share of the year's stories, 18 most used tags

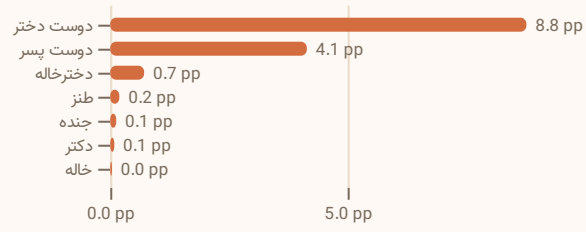


Each row is scaled to its own peak, so the pattern shows when a tag peaked, not how big it is.

Risers (percentage points gained)



Fallers (percentage points lost)

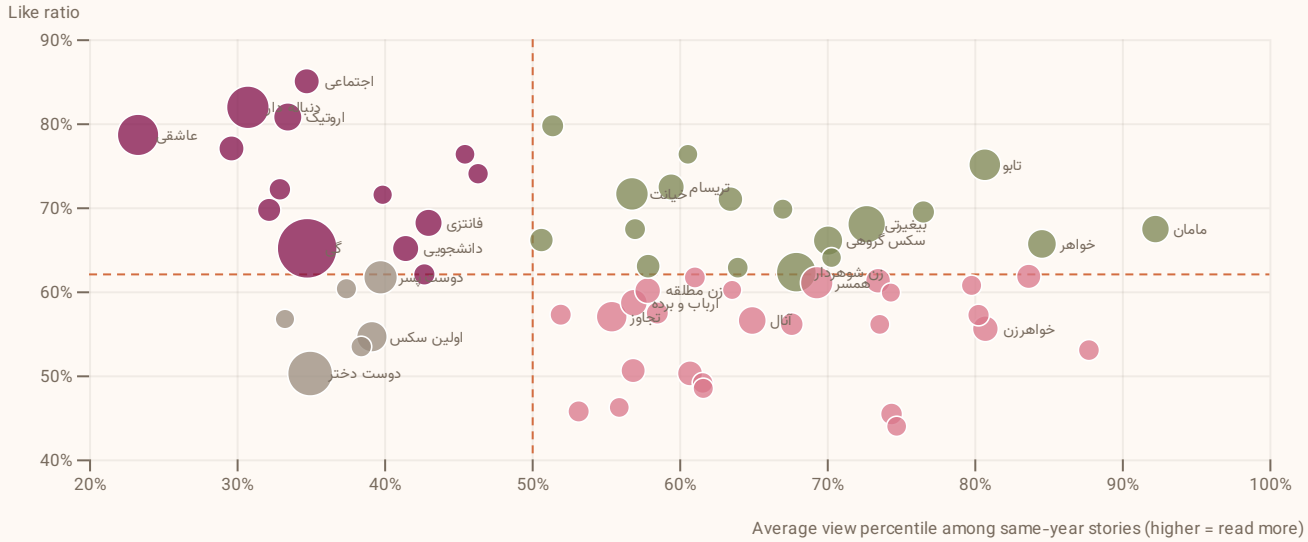


05 · TAGS

Read versus loved

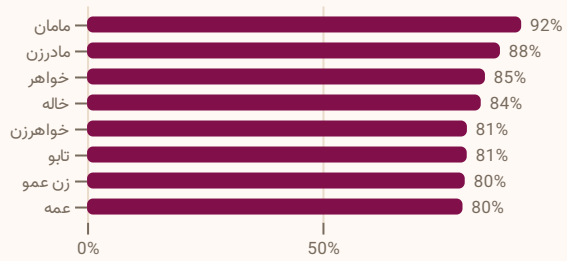
Tags with at least 150 stories. Read: average view percentile within the publication year (50% is typical). Loved: likes as a share of likes plus dislikes.

Reach against approval, by tag

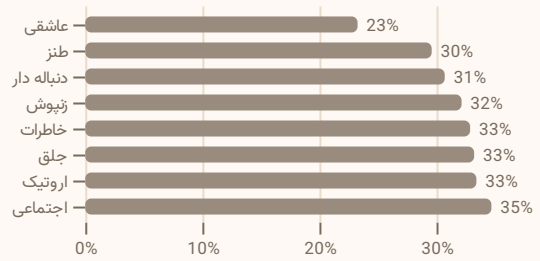


Bubble size is the number of stories. Dashed lines: the typical view percentile (50%) and the median tag's like ratio. Colour marks the quadrant.

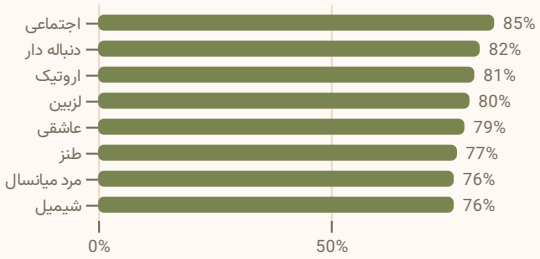
Most read tags



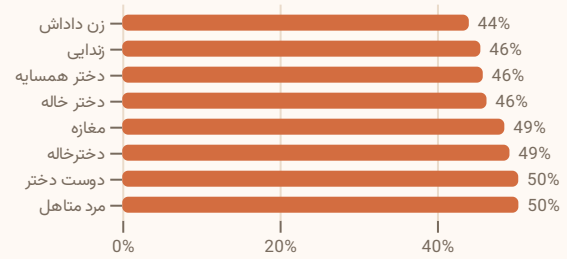
Least read tags



Most loved tags



Least loved tags



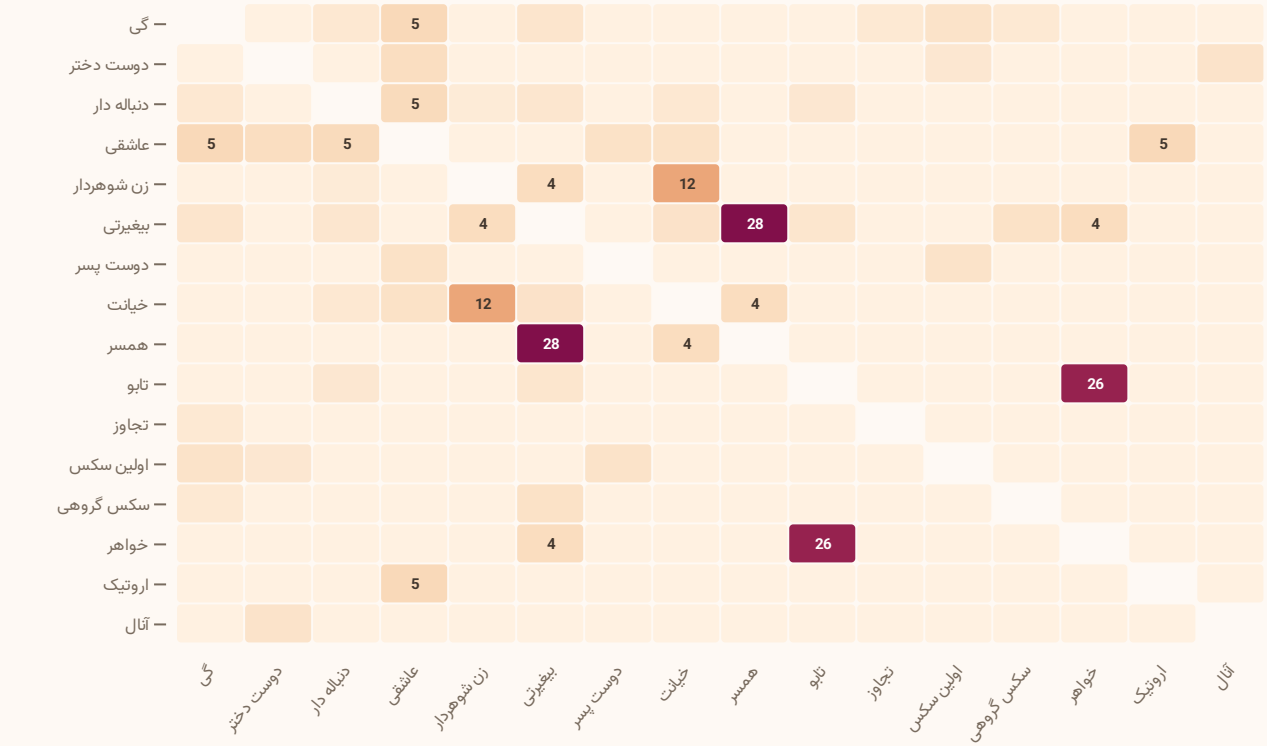
Caution. Tags describe how a story is filed, not why it was read. A high view percentile means readers chose these stories more often than other stories of the same year.

05 · TAGS

Tags that go together

Pairs of tags that share at least 60 stories. Jaccard is shared stories over stories with either tag; lift compares the pair with what chance would give.

Jaccard overlap, %, the 16 biggest tags



Cell: shared stories divided by stories with either tag. Pairs under 60 shared stories are not recorded and show as empty; values from 4% are printed.

Strongest overlap (Jaccard)

Tag pair	Stories	Jaccard
همسر + بیغیرتی	695	0.281
خواهر + تابو	415	0.258
مامان + تابو	211	0.124
خیانت + زن شوهردار	388	0.123
فتیش + ارباب و برده	117	0.119
فوت فتیش + ارباب و برده	86	0.097
بی دی اس ام + ارباب و برده	76	0.094
مامان + بیغیرتی	152	0.060

Strongest affinity (lift)

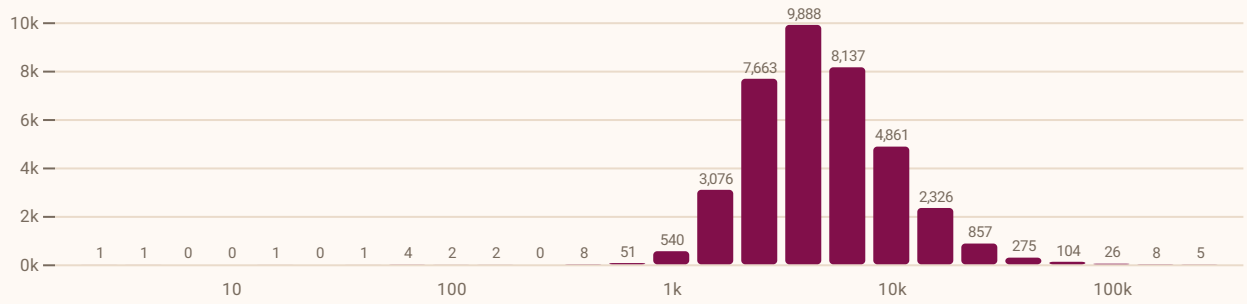
Tag pair	Stories	Lift
بی دی اس ام + ارباب و برده	76	19.1
خواهر + تابو	415	15.7
فوت فتیش + ارباب و برده	86	15.7
فتیش + ارباب و برده	117	15.1
همسر + بیغیرتی	695	11.0
مامان + تابو	211	9.3
خیانت + زن شوهردار	388	5.1
تریسام + بیغیرتی	133	4.3

06 · STORY ANATOMY

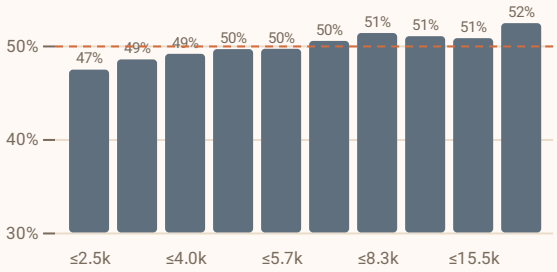
Length, titles and attention

The median story has 5,664 characters and the median title 19. Longer stories are only slightly more read (Spearman 0.05 against view percentile).

Stories by length in characters (log scale)

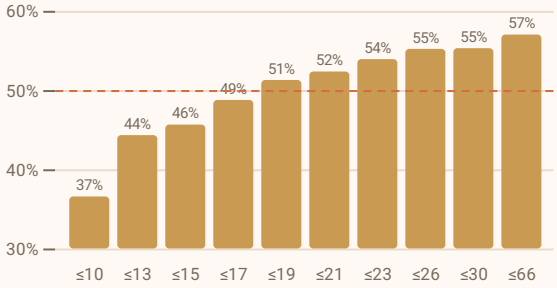


View percentile by story length



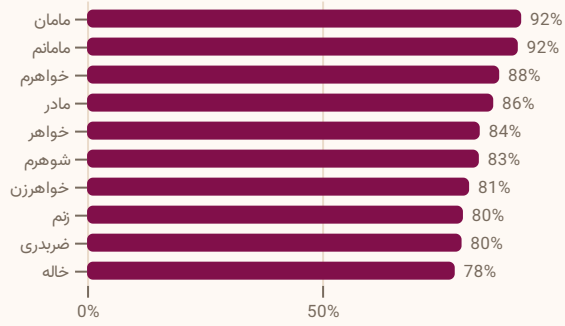
Ten equal-sized groups; labels are each group's longest story.

View percentile by title length



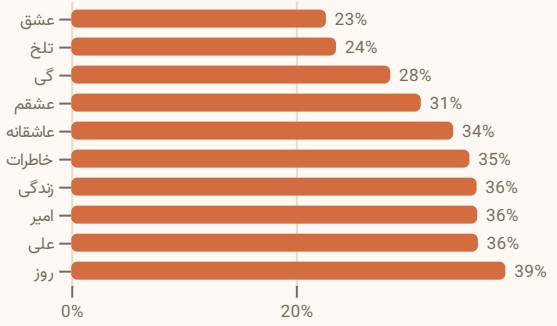
Title length in characters.

Title words linked to more views



Minor-related words omitted.

Title words linked to fewer views



Minor-related words omitted.

Average view percentile of stories whose title contains the word, for words in at least 150 titles. An association, not a cause.

07 · LANGUAGE

The vocabulary

60.9M Persian words in stories and 13.2M in clean comments, in 348k and 281k distinct spellings. Word lists skip function words; words tied to minors are withheld.

60.9M	348k	13.2M	281k
Persian words in stories	distinct spellings in stories	Persian words in clean comments	distinct spellings in comments

Most frequent content words in stories



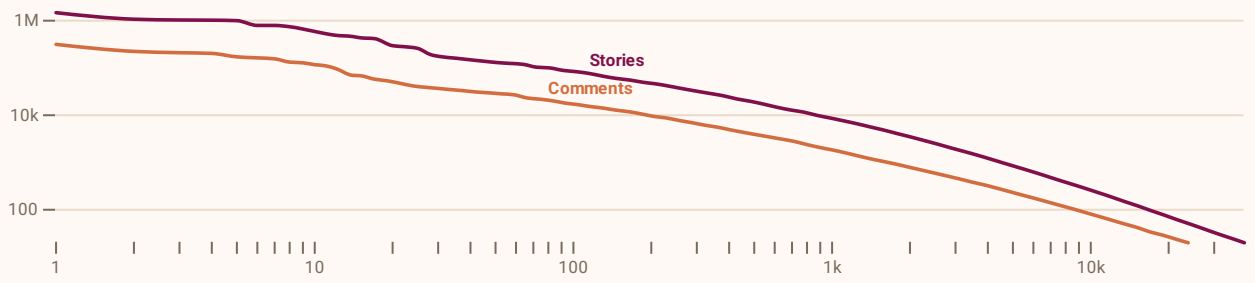
Minor-related words omitted. Size follows frequency; 1 minor-related words are left out.

Most frequent content words in comments



Minor-related words omitted. 1 minor-related words are left out.

Rank against frequency (both axes logarithmic)



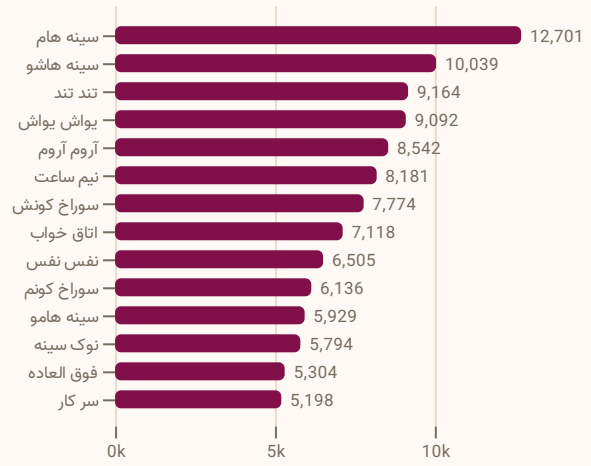
Every spelling used at least 20 times, function words included. A straight line is Zipf's law; fitted slope over ranks 10 to 1,000: -0.94 for stories, -0.91 for comments.

07 · LANGUAGE

Phrases and signature words

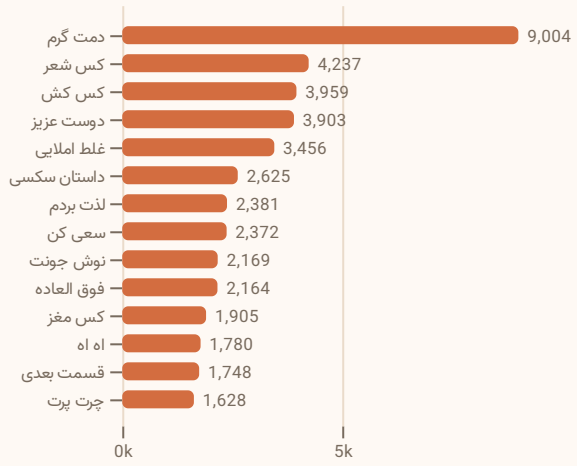
Pairs of adjacent content words show the formulas of each corpus. Signature words are the ones a tag uses far more than the corpus as a whole.

Most common word pairs in stories



Minor-related words omitted.

Most common word pairs in comments



Minor-related words omitted.

Signature words of the largest tags

Tag	Stories	Signature words
گی	6,686	سوراخت · گی · سوراخمو · سوراخم · فاعل · مفعول
دوست دختر	3,162	شمارتو · تاپشو · زدیمو · رسوندمش · مانتوشو · پردشو
دنباله دار	2,753	پرونده · مجموعه · سیاوش · میشنوم · نفوذ · دخترک
عاشقی	2,547	عشقت · عاشقی · دلتنگی · نقاشی · علیرضا · دورت
زن شوهردار	2,279	متاهلم · متاهل · شوهرش · بینهایت · شوهرشم · شوهرم
بیغیرتی	1,927	زنت · مامانمو · زنم · غیرتی · زنتو · زنمو
دوست پسر	1,370	چوچولم · کسمو · مانتومو · سوتینمو · چوچولمو · تاپمو
خیانت	1,274	شوهرم · پیمان · مهریه · هاله · سهیل · خیانت
همسر	1,238	غیرتی · زنت · خانومم · زنم · زنتو · زنمو
تابو	1,166	خواهرت · داداشی · متین · مامان · سیاوش · باران
تجاوز	1,039	گریم · ولم · التماسش · شکایت · چاقو · تجاوز
اولین سکس	1,010	بم · استرس · سوراخمو · بارم · اولم · سکسم
سکس گروهی	897	کیرشون · عسل · ستاره · ضربدری · زوج · گروهی
خواهر	856	خواهرت · سیاوش · آبجی · داداشی · خواهرم · خواهرمو

Minor-related words omitted. Words ranked by how much more often the tag uses them than all stories do (share inside the tag divided by share in the corpus), for words used at least 100 times in the tag and 200 times overall, in at least 400 stories. Tags that touch minors are not listed.

07 · LANGUAGE

Signature words as clouds

The same signature words, drawn for the largest tags with more words each. Size follows how much more the tag uses a word than the corpus does.

گی



Minor-related words omitted. 6,686 stories.

دوست دختر



Minor-related words omitted. 3,162 stories.

دنباله دار



Minor-related words omitted. 2,753 stories.

عاشقی



Minor-related words omitted. 2,547 stories.

زن شوهردار



Minor-related words omitted. 2,279 stories.

بیغیرتی



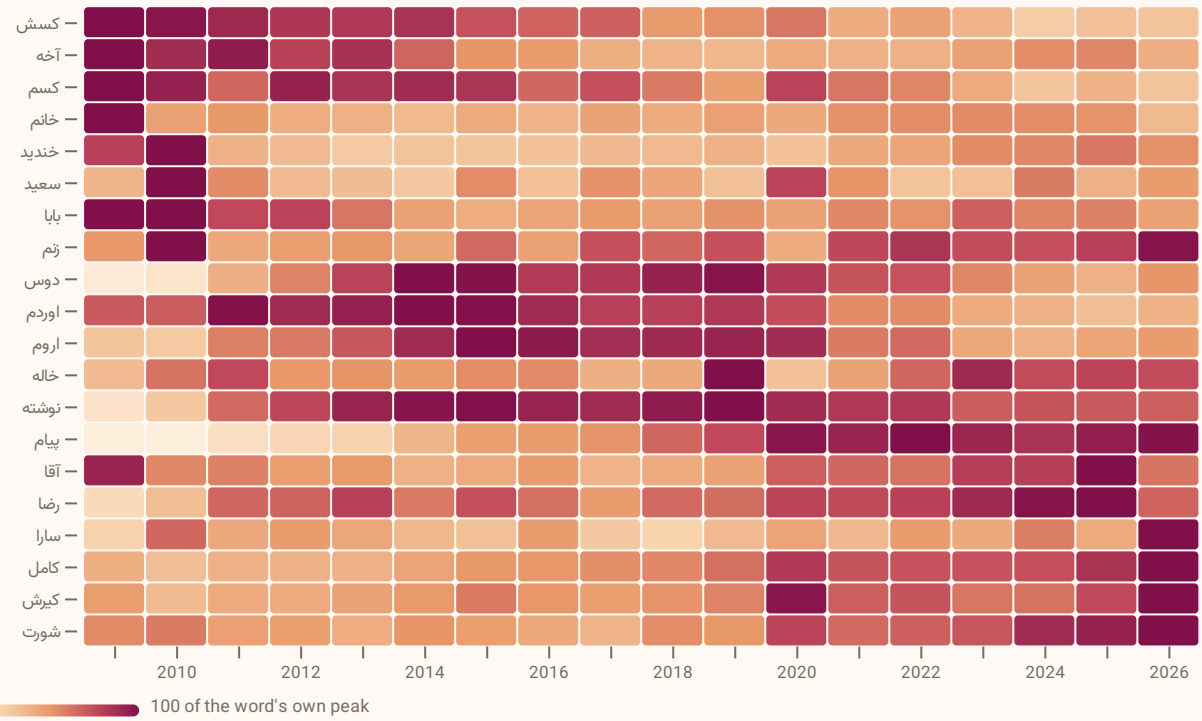
Minor-related words omitted. 1,927 stories.

08 · LANGUAGE OVER TIME

A corpus with a calendar

Words per million Persian words, by year of publication. Shown: the 20 words, among the 300 most frequent, whose use varies most over time, ordered by the year they peaked.

Top story words by year (each row scaled to its own peak)



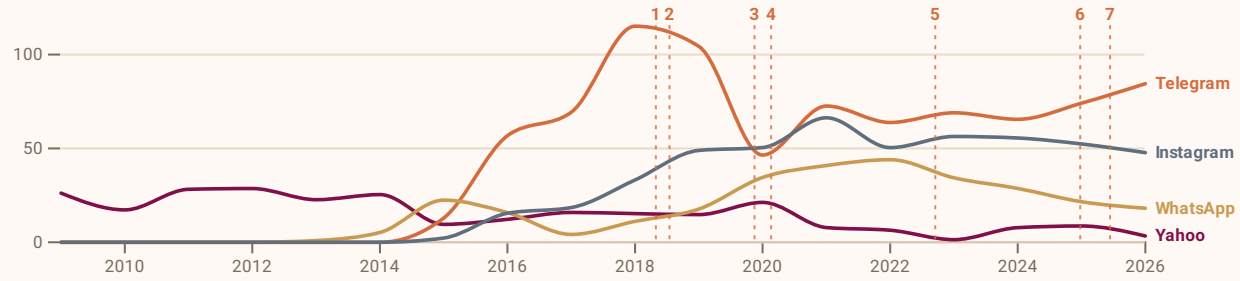
Minor-related words omitted. Years with fewer than 100,000 Persian words are left out.

08 · LANGUAGE OVER TIME

Cultural fingerprints

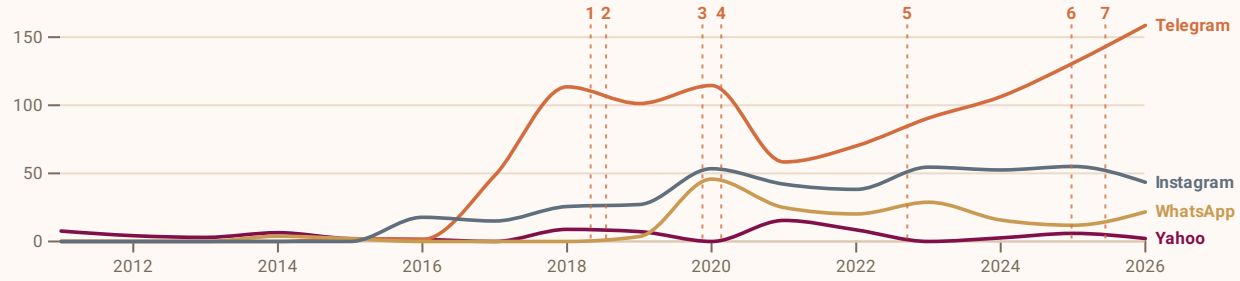
Rates per million words remove the effect of the corpus growing. Years with fewer than 100,000 Persian words are not plotted.

Messengers and platforms in stories



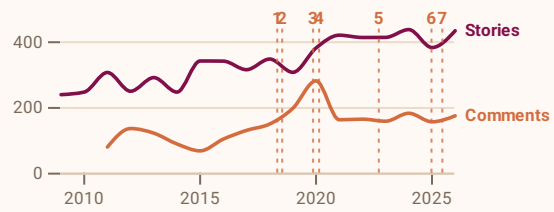
Per million Persian words. Words: Yahoo: یاهو, yahoo, مسنجر; Telegram: تلگرام, telegram; WhatsApp: واتساپ, واتس, whatsapp; Instagram: اینستاگرام, اینستا, instagram. Numbered lines: events listed in section 2.

Messengers and platforms in comments



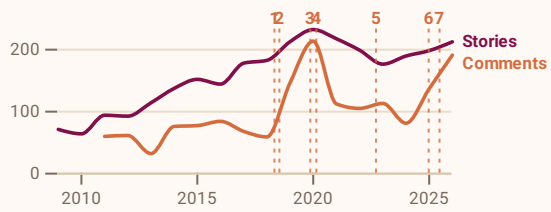
Per million Persian words. Words: Yahoo: یاهو, yahoo, مسنجر; Telegram: تلگرام, telegram; WhatsApp: واتساپ, واتس, whatsapp; Instagram: اینستاگرام, اینستا, instagram. Numbered lines: events listed in section 2.

Mobile phones



Per million Persian words. Words: گوشی, موبایل, آیفون, اندروید. Numbered lines: events listed in section 2.

Chat



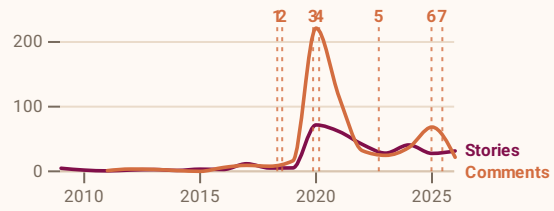
Per million Persian words. Words: چت. Numbered lines: events listed in section 2.

08 · LANGUAGE OVER TIME

Events and habits

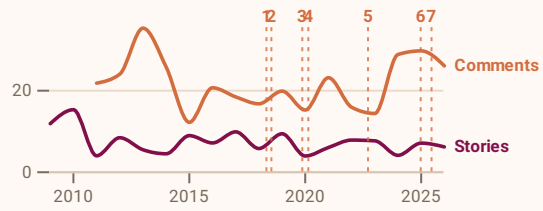
Rates of words tied to events, platforms and habits, in stories and in comments.

Corona



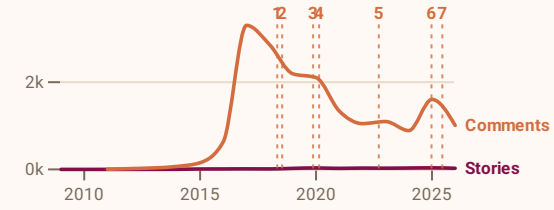
Per million Persian words. Words: کرونا, قرنطینه, ماسک. Numbered lines: events listed in section 2.

Web filtering



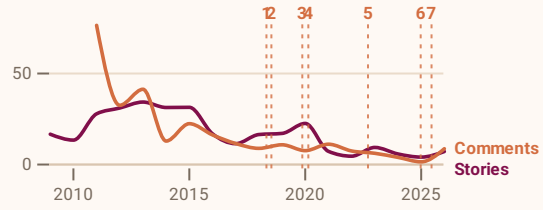
Per million Persian words. Words: فیلتر, فیلترشکن. Numbered lines: events listed in section 2.

Likes



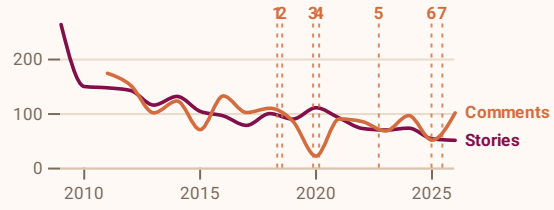
Per million Persian words. Words: لایک. Numbered lines: events listed in section 2.

Email and text messages



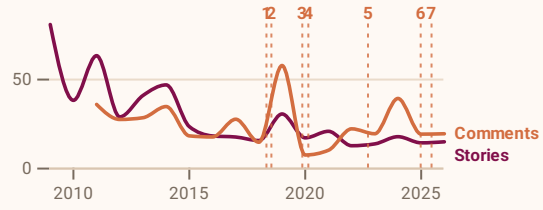
Per million Persian words. Words: ایمیل, اسمس. Numbered lines: events listed in section 2.

Internet and computers



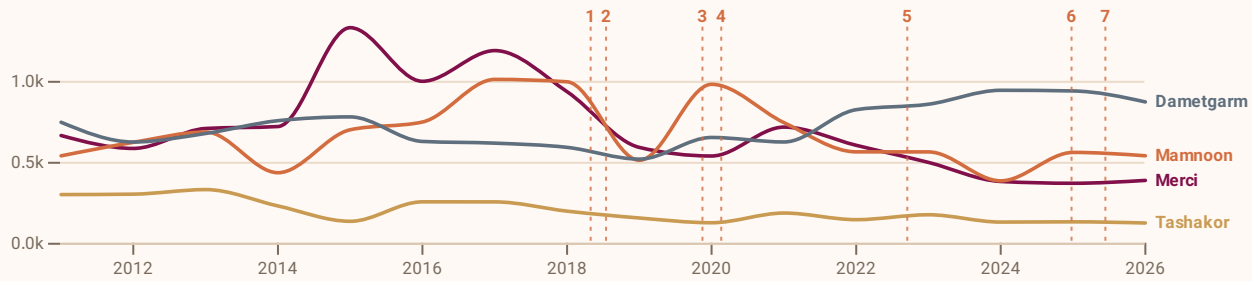
Per million Persian words. Words: آنلاین, لپتاپ, کامپیوتر, اینترنت. Numbered lines: events listed in section 2.

Satellite TV



Per million Persian words. Words: ماهواره. Numbered lines: events listed in section 2.

Ways of saying thanks in comments



Per million Persian words. Words: Merci: مرسی; Mamnoon: ممنون; Tashakor: تشکر; Dametgarm: دمت. Numbered lines: events listed in section 2.

08 · LANGUAGE OVER TIME

Risers and fallers

Content words whose share changed most between the first and the last years of each corpus. Minor-related words omitted.

Rising in stories

Word	2009–2011	2024–2026	Change
کصش	0.00	397	×398
کصم	0.00	218	×219
کص	0.00	201	×202
داگی	1.43	224	×93
تلگرام	0.00	73.6	×75
کوصش	0.00	52.0	×53
اسنپ	0.00	51.9	×53
کصشو	0.00	50.1	×51
کوص	0.00	43.8	×45
اینستا	0.00	38.7	×40

Falling in stories

Word	2009–2011	2024–2026	Change
اطاق	113	7.95	×0.08
دفه	73.0	8.01	×0.12
شیوا	215	26.3	×0.13
اخمی	52.2	6.46	×0.14
ارش	55.8	8.26	×0.16
مثه	94.0	15.1	×0.17
شرو	65.1	11.2	×0.19
شیرین	675	124	×0.19
جلق	134	25.2	×0.19
پرستو	127	24.7	×0.20

Rates are occurrences per million Persian words in 2009–2011 and 2024–2026; change is the ratio of the two, each rate raised by 1 to damp rare words. Words need at least 300 occurrences in the two windows together and 150 documents overall.

Rising in comments

Word	2011–2013	2024–2026	Change
کص	1.37	1248	×528
دیسلایک	0.00	208	×209
کوص	0.27	263	×208
لایک	20.2	1173	×55
دیس	2.73	167	×45
فاعل	18.9	212	×11
روون	12.6	139	×10
مفعول	28.4	290	×9.9
زوج	16.7	163	×9.3
احمد	23.8	223	×9.1

Falling in comments

Word	2011–2013	2024–2026	Change
پریچهر	169	0.00	×0.01
آریزونا	146	0.00	×0.01
تکاور	138	0.00	×0.01
سیلور	125	0.00	×0.01
فریجاب	94.0	0.00	×0.01
آمیرزا	91.8	0.00	×0.01
پروازی	124	0.53	×0.01
ساینا	181	2.64	×0.02
هه	282	6.34	×0.03
هیوا	134	3.17	×0.03

Rates are occurrences per million Persian words in 2011–2013 and 2024–2026; change is the ratio of the two, each rate raised by 1 to damp rare words. Words need at least 300 occurrences in the two windows together and 150 documents overall.

09 · COMMENTS

The conversation

536,919 comments on 37,556 stories. Half of all comments arrive within 15 h of the story.

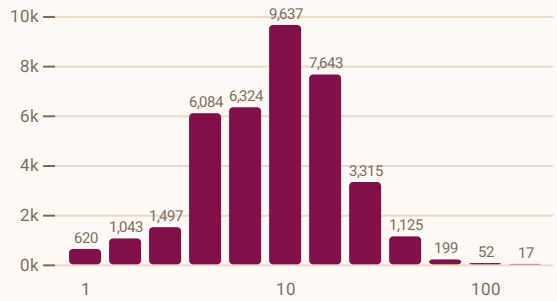
99.3%
of stories have at least one comment

11
median comments per commented story

210
most comments on one story

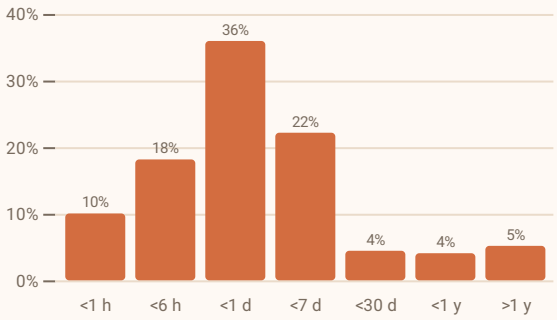
15.4 h
median delay before a comment

Comments per commented story (log scale)



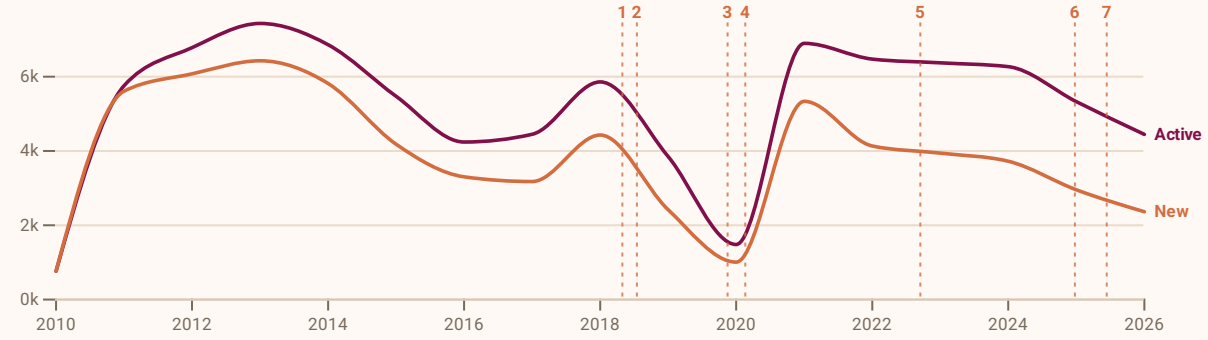
Each bar covers a fifth of a power of ten, labelled by its lower edge.

How long after the story a comment arrives



Share of comments by delay; each bar holds the delays between the previous label and its own. Comments dated before their story (254) are left out.

Accounts commenting each year



Active accounts wrote at least one comment that year; new ones wrote their first. The rest are returning accounts. Spam is included, as these are all comments. Numbered lines: events listed in section 2.

Most commented stories

#	Title	Year	Comments	Views
1	بالاخره کردمش	2013	210	54,780
2	بدون مرز	2022	190	289,901
3	سحر	2021	187	222,401
4	تو برادر منی	2021	181	571,601
5	زن حشری در تاکسی	2013	177	301,872
6	مفت از دستش دادم	2013	175	42,989
7	مامانم میشی؟ (۵ و پایانی)	2023	175	197,501

Minor-adjacent titles are hidden.

09 · COMMENTS

Who comments

The top 1% of 65,665 accounts wrote 34% of all comments, and the top 10% wrote 71%.

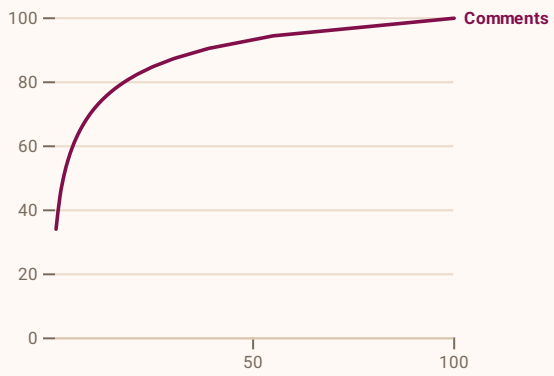
65,665
accounts have commented

45%
commented only once

34.1%
of comments from the top 1%

71.0%
of comments from the top 10%

Concentration of comments



Horizontal: top % of accounts. Vertical: % of all comments they wrote.

Accounts by comments written

Comments	Accounts	% of comments
1	29,553	5.5%
2–4	19,717	9.7%
5–19	11,836	19.8%
20–99	3,775	28.3%
100–499	710	25.0%
500+	74	11.6%

All comments, spam included.

Top commenters

#	User	Comments	Stories	Likes received	Likes / comment	Active	Flagged
1	شاه ایکس	2,818	2,428	14,279	5.1	2017–2026	11
2	Fael shiraz	2,073	1,835	1,650	0.8	2024–2025	70
3	aliabadan	2,051	2,043	1,956	1.0	2011–2026	0
4	takavarjoon	1,809	1,662	58	0.0	2011–2016	16
5	shadow69	1,673	1,266	1,982	1.2	2017–2019	6
6	milad1ma	1,567	1,551	471	0.3	2013–2023	0
7	sepideh58	1,555	1,139	2,514	1.6	2011–2025	2
8	ESF_ROMANCE	1,374	1,327	1,021	0.7	2021–2026	0
9	داریوشم	1,290	1,264	1,219	0.9	2015–2026	3
10	saman.32	1,243	1,227	496	0.4	2015–2024	1
11	MS.TEACHER	1,132	1,018	31	0.0	2012–2014	0
12	farirad	1,078	1,064	887	0.8	2020–2026	0
13	kiredivoone	1,078	1,059	883	0.8	2018–2022	0
14	david1975	1,019	969	763	0.7	2021–2026	0
15	Ahmad007ir	985	955	1,079	1.1	2021–2026	0

User names are shown as they appear on the site. Flagged comments are duplicates or contain a link (the spam rule of the data-quality step).

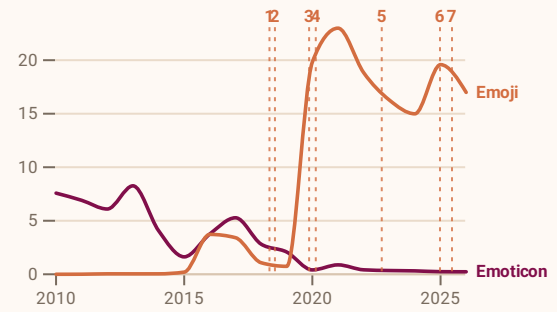
Reading it. The most-liked regular among these is شاه ایکس at 5.1 likes per comment against 1.3 for the group. 45% of accounts wrote a single comment, yet together they account for only 5.5% of all comments; the 74 accounts with 500+ comments wrote 11.6%.

09 · COMMENTS

How comments changed

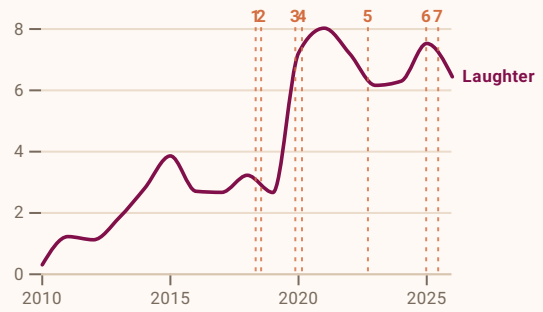
Shares of clean comments per year, for years with at least 1,000 comments.

Emoticons and emoji



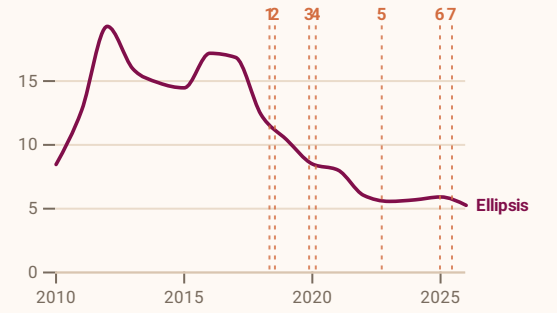
% of comments. Emoticons: :), :D, :(, ;), :P, :–), :–(, (:, ^_^. Emoji by Unicode block. Numbered lines: events listed in section 2.

Laughter



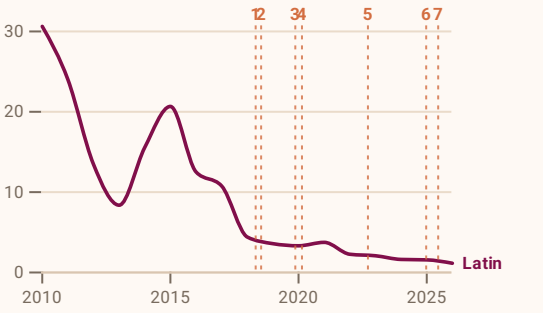
% of comments containing خخخ, هههه, هاها, lol and 2 laughing emoji. Numbered lines: events listed in section 2.

Ellipsis



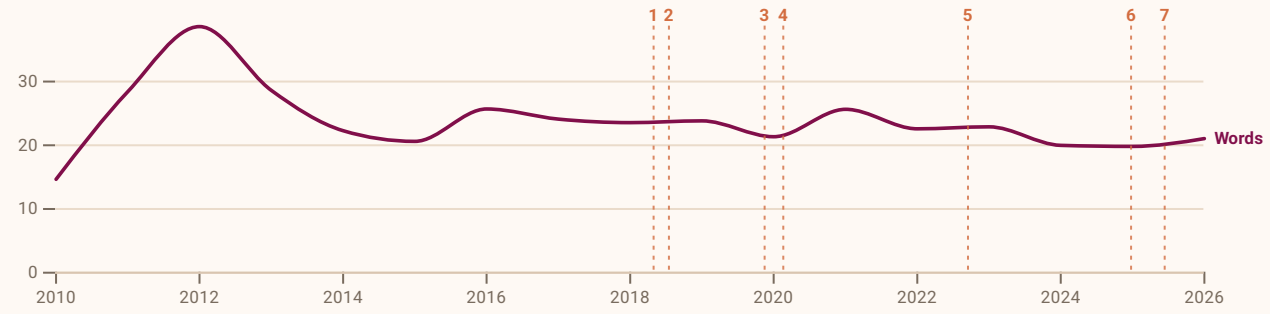
% of comments containing ..., ... Numbered lines: events listed in section 2.

Latin letters



% of comments with a Latin word of three letters or more, mostly Persian typed in Latin script and English words. Numbered lines: events listed in section 2.

Words per comment



Average Persian words per clean comment. Numbered lines: events listed in section 2.

Reading it. Emoticons peaked in 2013 at 8.3% of comments and stood at 0.2% in 2026. Emoji peaked in 2021 at 23.0% (17.0% in 2026). Latin letters appeared in 30.6% of comments at the peak (2010) against 1.1% in 2026. Years differ in size, so read each line as a shape, not as an exact count.

10 · ACCOUNTS

Who the accounts are

234,077 profiles. Gender is stated by 99%; the rest is a blank or an unlisted value.

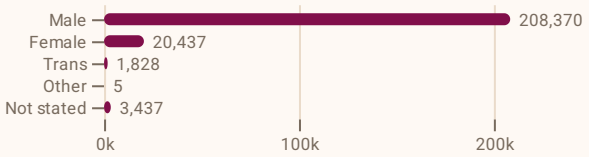
234,077
profiles

89.0%
stated male

8.7%
stated female

0.8%
stated trans

Accounts by stated gender

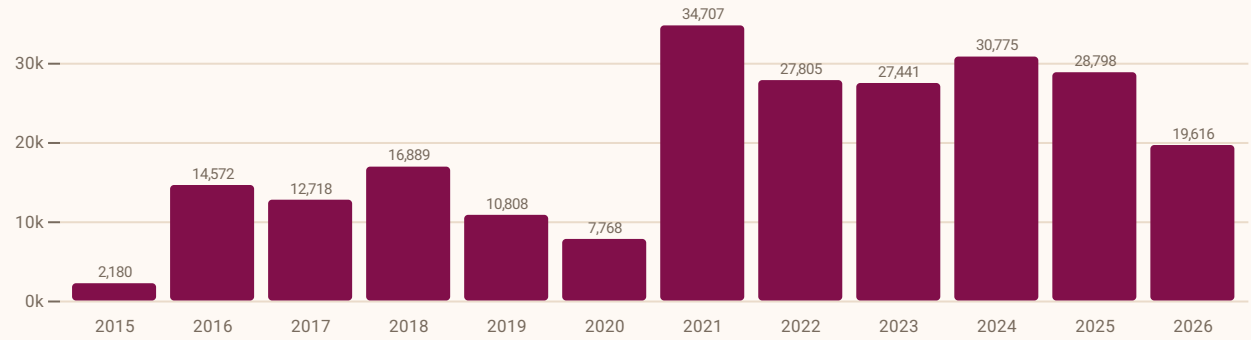


"Other" collects the few free-text values that are not one of the three listed options.

How much of each profile is filled in

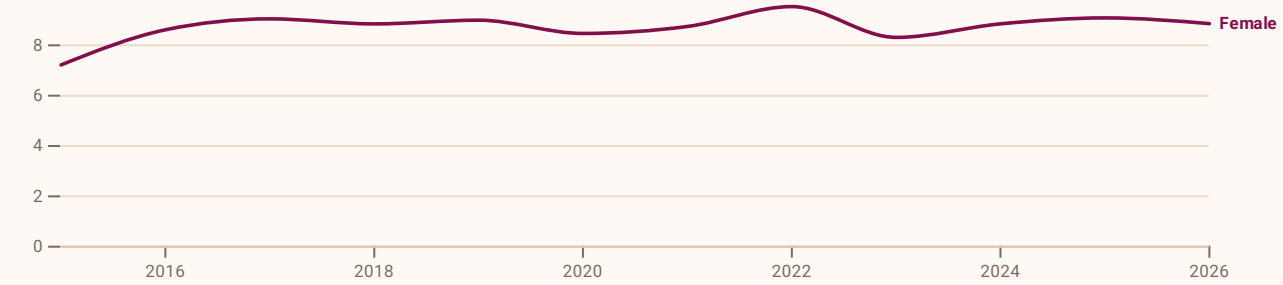
Field	Filled	Share
Gender	230,640	98.5%
Bio	54,883	23.4%
City	123,706	52.8%
Country	225,883	96.5%
Province	160,725	68.7%

Accounts created per sign-up year



Only accounts with a sign-up date are counted; the earliest date in the table is 2015, so older accounts are undated rather than absent. 2026 is a partial year.

Women's share of accounts that state a gender, by sign-up year



Percent of accounts created that year with a stated gender.

Reading it. The busiest sign-up year is 2021 (34,707 accounts). Men are 89% of all profiles; the second-largest stated group is women at 8.7%.

10 · ACCOUNTS

Where accounts are

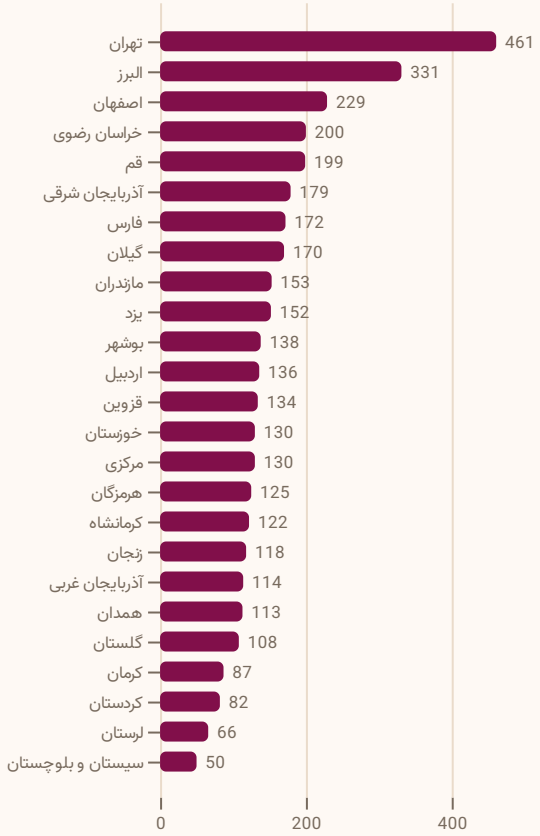
160,725 profiles name a province (69%); 160,523 of them an Iranian province, 202 a place outside the provincial list. 97% of profiles that state a country state Iran.

68.6%
of profiles in an Iranian province

201
profiles per 100,000 inhabitants (stated provinces)

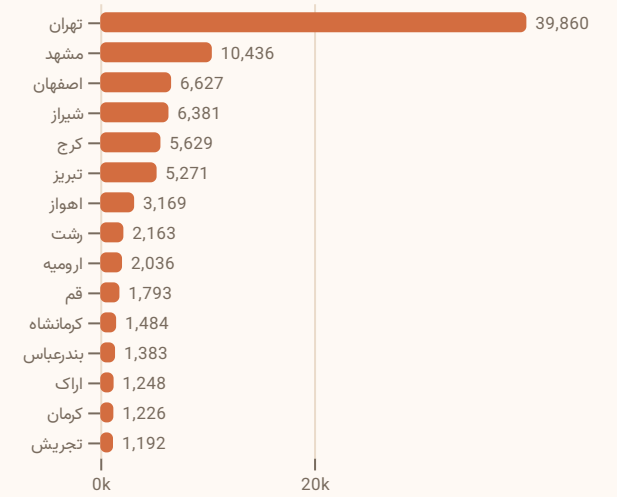
38.1%
of province-stated profiles are in the largest province, against 16.6% of the population

Profiles per 100,000 inhabitants (provinces with 1,000+ profiles)



Per-capita rates use Statistical Centre of Iran, 2016 census (rounded).

Most named cities



Cities are free text, so spellings split counts; the 15 most common spellings are shown.

Reading it. تهران has the highest rate (461 per 100,000) and سیستان و بلوچستان the lowest among provinces with 1,000+ profiles (50). Rates compare profiles to residents, not to internet users, and people may state a province where they do not live. The abroad list is led by آلمان (987), آمریکا (719), ترکیه (590).

10 · ACCOUNTS

What bios say

54,883 profiles (23%) have a bio, most of them short. Only aggregated word counts are shown.

54,883
profiles with a bio

28 / 33
median bio length in characters, men / women

23% / 27%
with a bio, men / women

Most used bio words



Minor-related words omitted.

Words that set women's bios apart from men's, and the reverse

More in women's bios	More in men's bios
لزبین	حاشیه
لز	کلفت
واتساپ	قد
تصویری	حرفه
خیس	بلند
نمیدم	فیس
تماس	عشق
بلاک	تمیز
وب	مشهد
جواب	میگردم
درخواست	خوش
سینه	پسر
هماهنگی	کون
وایف	سن
اینستاگرام	مهربون
هزینه	ساک

Ranked by the ratio of a word's share in one group's bios to its share in the other's; words need 20+ uses in the group and 40+ overall.

10 · COMMENTER PROFILES

Commenters seen through profiles

Only 43% of commenting accounts (58% of comments) have a profile row, so everything here describes that matched part. A heavy commenter wrote 100+ comments.

28,378
commenters with a profile

37,287
commenters without one

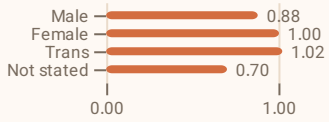
531
heavy commenters among them

Comments per account



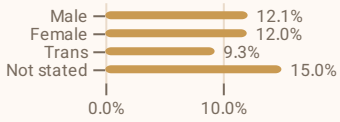
Matched commenting accounts.

Likes per comment



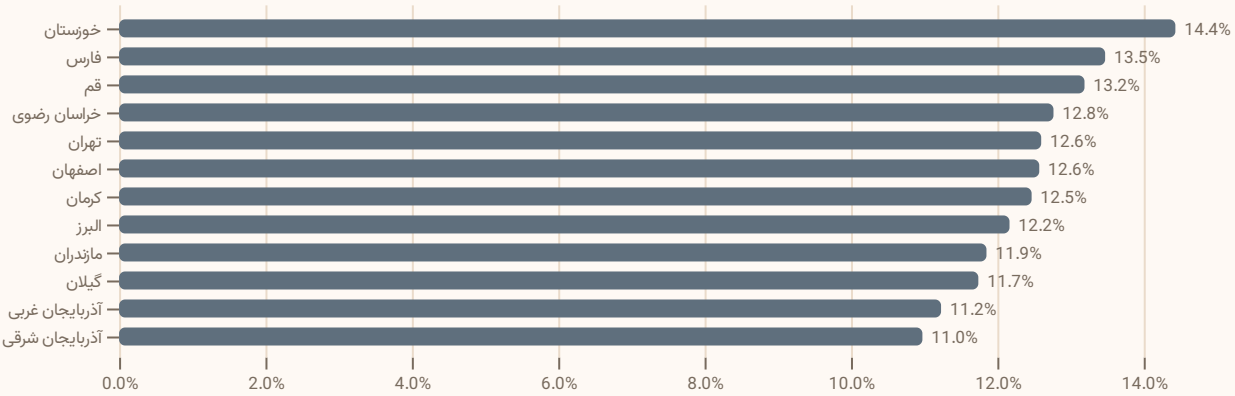
Likes on comments by matched accounts.

Accounts that comment



Share of profiles that have commented.

Share of profiles that have commented, in the 12 provinces with the most profiles



Where heavy commenters live

Province	Heavy commenters	Share of heavy	Share of profiles	Ratio
تهران	143	43.1%	38.1%	1.13
فارس	26	7.8%	5.2%	1.51
اصفهان	22	6.6%	7.3%	0.91
خراسان رضوی	19	5.7%	8.0%	0.71
البرز	14	4.2%	5.6%	0.75
خوزستان	13	3.9%	3.8%	1.02
مازندران	12	3.6%	3.1%	1.15
گیلان	12	3.6%	2.7%	1.35
آذربایجان شرقی	8	2.4%	4.4%	0.55
کرمان	8	2.4%	1.7%	1.41

Ratio above 1 means more heavy commenters than the province's share of profiles would suggest. Heavy commenters with a stated Iranian province: 332; 9.0% of all heavy commenters are women, against 8.7% of profiles.

11 · FORUM

Forum boards

303,977 topics in 30 boards, from 2009-11-04 to 2026-11-16. The table has no author column: it counts topics, not people.

303,977
topics

11.0M
replies

3.7B
topic views

36.0
replies per topic

Boards by number of topics

Board	Topics	Replies per topic	Median replies	Median views	No reply
گی	68,417	19.0	5	620	16%
فتیش	33,522	16.8	5	1,340	22%
سکس گروهی	26,244	18.8	4	590	25%
درخواست دوست دختر	25,885	5.2	0	220	54%
فیلم سکسی	22,406	27.8	16	26,145	4%
عکس‌های سکسی	20,422	21.8	9	6,540	12%
عکس سکسی آماتوری	19,312	35.1	17	9,740	5%
کافه شهوانی	18,509	34.3	12	2,000	8%
گفتگو درباره ی سکس و پورن	15,949	23.5	10	2,900	10%
سکس حضوری و صیغه	9,246	16.0	1	520	35%
بحث های اجتماعی/سیاسی	6,221	22.7	11	1,530	8%
طنز و خنده	5,921	24.9	9	3,370	8%
مجله و داستان مصور سکسی	5,073	13.0	8	66,410	9%
درخواست دوست پسر	4,426	21.9	3	800	28%
ستاره پورن / مدل های آس	4,355	15.3	8	9,170	7%
هنر	3,208	34.3	9	1,280	9%
خاطرات و داستان های سکسی	1,798	16.7	7	4,678	13%
درباره سایت و انجمن شهوانی	1,724	18.0	6	1,327	12%
سکسولوژی	1,543	12.7	7	7,070	18%
دین	1,480	35.0	15	2,370	9%
لزبین	1,436	9.5	4	1,000	18%
معرفی سایت و وبلاگ سکسی	1,186	16.5	3	10,785	24%
سکس تصویری	1,134	4140.1	3	2,415	23%
موسیقی	1,129	14.9	5	1,206	11%
علم و تکنولوژی	1,122	16.2	7.5	1,530	12%
انیمه، هنتای و هوش مصنوعی	776	12.0	7	9,210	8%
سینما	754	15.5	7	3,130	10%
داستان های آرشیوی	471	20.9	8	2,570	10%
کتاب	270	18.8	11	1,490	7%
داستان های صوتی	38	23.3	13.5	54,230	3%

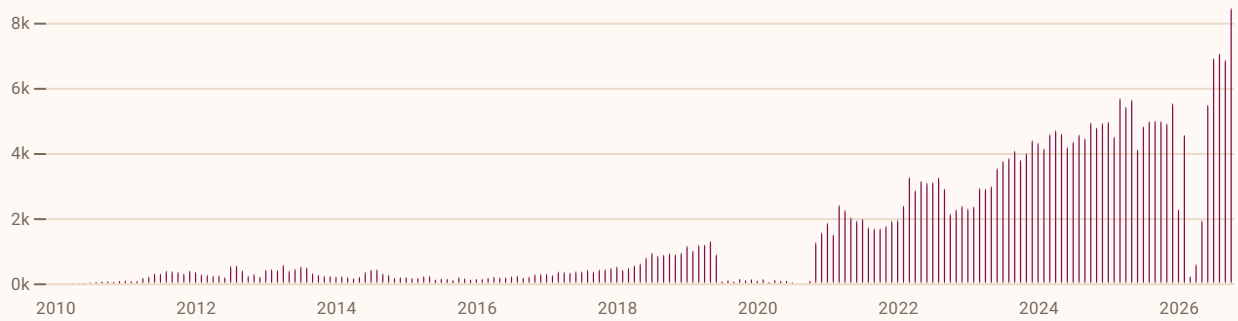
Medians are per topic; views and replies are cumulative at collection time. Replies per topic are more than 20 times the typical board's on سکس تصویری; the median shows a normal board, so a few topics carry most of the replies.

11 · FORUM

Over time and in words

The forum record covers 205 calendar months. Its busiest full month is Oct 2026 (8,445 topics).

Topics per month



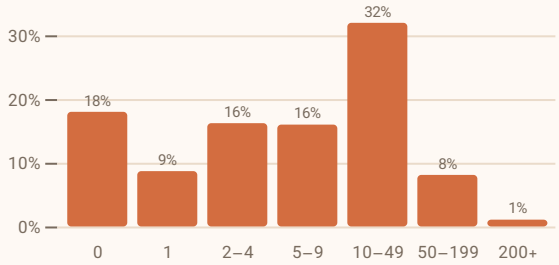
The first and last months are partial and left out. Jul 2019 to Oct 2020 holds under a quarter of the surrounding months, which looks like a gap in the record rather than a quiet period.

Median replies per topic, by start month



Median replies of the topics started that month. Replies are cumulative at collection time, so older months have had longer to collect them. Numbered lines: events listed in section 2.

How many replies a topic gets



Share of all topics by number of replies.

Signature title words per board

Board	Topics	Words used here more than across the forum
گی	68,417	مکاندار · تاپم · ساکرم · سافت · ورس
فتیش	33,522	ناموست · مامانتو · بیغیرتا · ناموسم · ناموسشو
سکس گروهی	26,244	نفرسومم · رسمی · سومم · ضرب · موازی
فیلم سکسی	22,406	دختره · مستقیم · کلیپ · تیک · تاک
عکس‌های سکسی	20,422	انیمه · سریالی · گالری · کلکسیون · متحرک
عکس سکسی آماتوری	19,312	ایرونی · دامن · سلفی · تعداد · سوگل

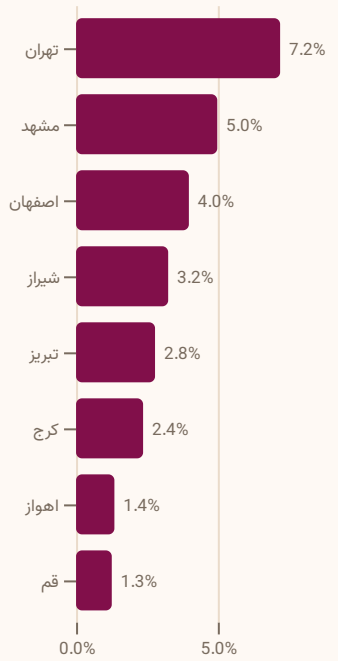
Ranked by lift of a word's share in the board's titles over its share in all titles, with 60+ uses in the board. The request boards have their own page.

11 · FORUM

Request boards

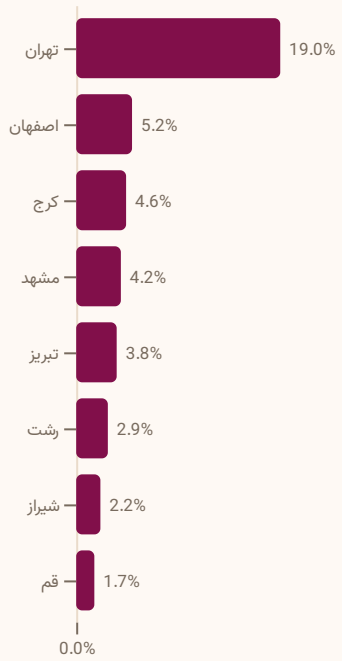
Boards where people ask for company, matched by name. Places are the cities and provinces that profiles name, found as words in topic titles; the adjective form with a trailing letter counts too.

درخواست دوست دختر



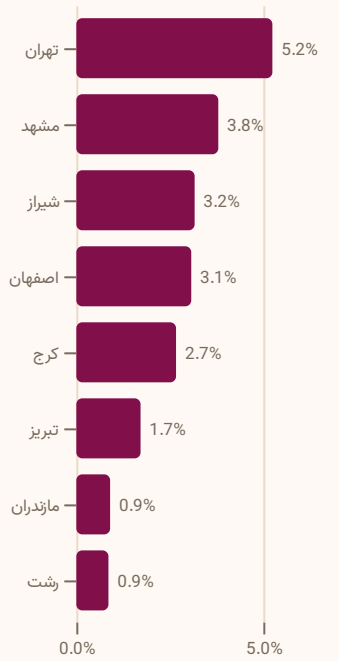
44% of 25,885 topic titles name a place. Bars: share of all topics on the board.

سکس حضوری و صیغه



58% of 9,246 topic titles name a place. Bars: share of all topics on the board.

درخواست دوست پسر



33% of 4,426 topic titles name a place. Bars: share of all topics on the board.

Wording that marks each request board

Board	Title words used more here than across the forum
درخواست دوست دختر	خانوم · دوستی · دختر · مجرد · بیوه · پایدار · مامی · شوگر · همخونه · مطلقه · شوگرمامی · بنفیت
سکس حضوری و صیغه	کانال · میپذیریم · حضوری · صیغه · نده · ساعتی · کیس · تلگ · حضوریم · پرداخت · هسم · فیکی
درخواست دوست پسر	سلام · هات · دختر · دنبال · دوستی · پارتنر · میخام · دوست · مرد · مامی · پسر · شوگر

Place names are removed from this list. Same lift measure as the previous page, 60+ uses in the board.

Reading it. Topics are compared with the profiles that name the same place. Board درخواست دوست دختر: of the 8 most named places, most topics per 1,000 profiles in تبریز (136.8), fewest in تهران (29.9). Board سکس حضوری و صیغه: of the 8 most named places, most topics per 1,000 profiles in رشت (125.8), fewest in تهران (28.1). Board درخواست دوست پسر: of the 8 most named places, most topics per 1,000 profiles in شیراز (21.9), fewest in تهران (3.7).

12 · PHOTOS

The photo gallery

63,885 photos from 2010-09-07 to 2026-11-15. Descriptive only: counts of uploads, categories and views.

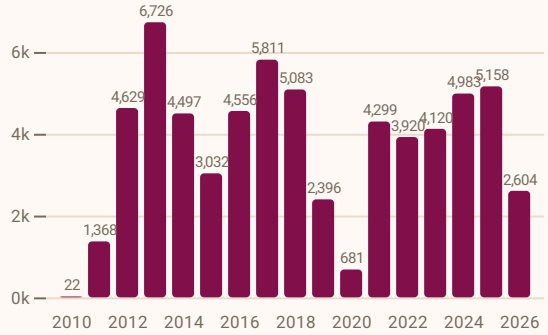
63,885
photos

4.5B
photo views

73
named uploaders

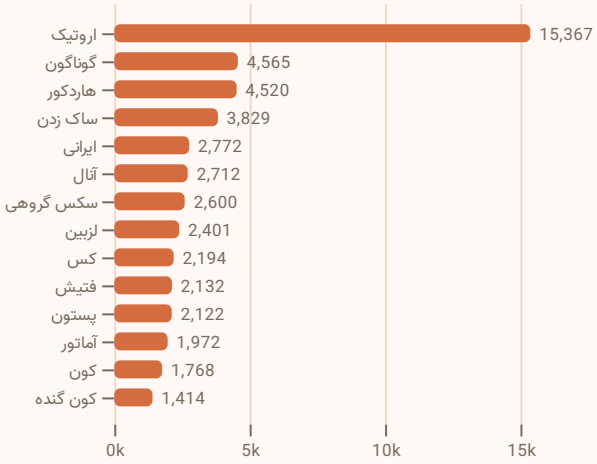
31%
photos without a named uploader

Photos per year



The last year is partial.

Largest categories



Category labels that are excluded as minor-related or hidden (1 labels, 449 photos) are not listed. 41 labels remain.

Top uploaders

#	User	Photos	Share of named uploads	Views
1	pinaco	29,124	66.4%	1,956,656,513
2	PouyaShahvani	10,747	24.5%	1,492,898,812
3	pouyapoya2012	2,275	5.2%	67,208,821
4	sj0087	659	1.5%	98,437,300
5	sukhoi	171	0.4%	4,848,317
6	zhinaless	143	0.3%	4,016,409
7	Lady_Ronaak	111	0.3%	6,142,159
8	mr_deraz	89	0.2%	5,356,372
9	ShadowLust	76	0.2%	272,500
10	Hooman.Afjeh	67	0.2%	4,206,900
11	amir__hosein	60	0.1%	2,730,991
12	Egalite	49	0.1%	1,594,401

User names are shown as on the site.

Reading it. The top uploader accounts for 66% of named uploads and the top 5 for 98%; the median photo has 42,300 views.

12 · PHOTOS

How photos are titled

Most used content words in photo titles (function words dropped), at least 30 uses.

Title words



Minor-related words omitted.

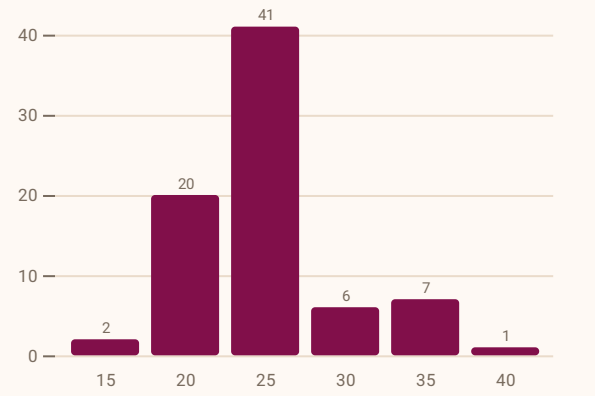
Captions are not analysed. Only 1,075 of 63,885 photos (2%) have a caption with Persian text once image tags are removed; the rest are image markup alone.

13 · LISTINGS

Listings

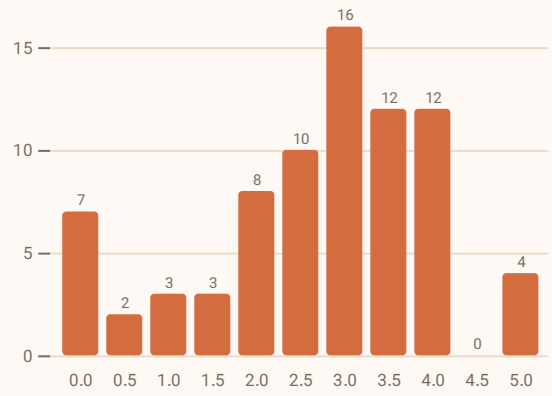
77 service listings. With so few rows this is a descriptive table, not a statistical sample: no inference is drawn, and user names, links and bios are not shown.

Ages



Listings per 5-year band, labelled by its lower age, as stated.

Average rating



Listings per 0.5-point band of the site's average rating, labelled by its lower edge.

At a glance

Measure	Value
Listings	77
Marked active	77 (100%)
With golden badge	40 (52%)
Age: median, range	25 (18–40)
Rating: median, mean	3.09, 2.89
Mean rating, golden badge	3.28
Mean rating, no badge	2.46
Naming a province	45 (58%)
Naming a city	29 (38%)
With a bio	43 (56%)
Credit values	0: 7, 10: 70

Provinces named

Province	Listings
تهران	36
مازندران	2
آذربایجان شرقی	1
اصفهان	1
خوزستان	1
فارس	1

Cities named

City	Listings
تهران	17
پرند	2
آمل	1
اصفهان	1
بابلسر	1
تبریز	1

Read with care. Province and city are filled for only part of the rows (45 and 29 of 77), and the rating's scale and meaning are the site's own.

14 · STORY LIFETIME

Views per day

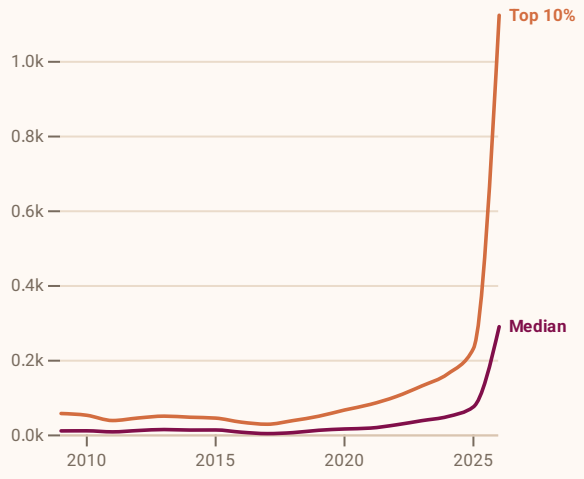
Total views divided by the days since publication, for 37,567 stories at least 30 days old. The record ends on 2026-11-16.

20.6
median views per day

0.71
Spearman, views per day against total views

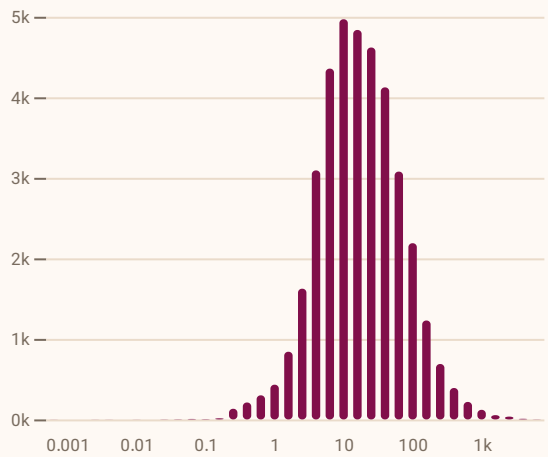
-0.54
Spearman, views per day against story age

Views per day, by publication year



Median and 90th percentile of stories published that year.

Stories by views per day (log scale)



Each bar covers a fifth of a power of ten, labelled by its lower edge.

Highest views per day

#	Title	Year	Views	Per day	Rank by views
1	خاله سهیلای خوشگل من (۱)	2026	281,501	6,467	1,184
2	محرم، حلال شد	2026	306,501	6,314	955
3	طاهره یا مرضیه؟ (۱)	2026	373,701	4,879	578
4	مهسا ، آغاز ماجراهای من	2026	191,501	4,718	2,834
5	شورت مهرنوش (۲)	2026	318,301	4,575	867
6	طاهره یا مرضیه (۵)	2026	147,701	4,281	4,527
7	مادرزن مذهبی و سفیدم	2026	278,401	4,241	1,220
8	کون خانم دکتر (۱)	2026	160,501	4,160	3,931

Minor-adjacent titles are hidden. Rank by views is among the stories old enough to be rated.

Reading it. The median story from 2009 draws 12.1 views a day, against 291.0 for 2026, so recent stories are read far more per day than old ones, and this shapes any ranking by views per day. Of the 100 stories with the highest views per day, 1 are also among the 100 with the most total views: the two rankings answer different questions.

14 · STORY LIFETIME

The activity stamp and audio coverage

A story's lifetime would need a trustworthy "last activity" date. This page checks whether the stamp in the data can provide one, then covers audio.

51%
of all stories carry the same last-activity day, 2019-06-19

19,400
stories whose stamp is later than publication, on 4 distinct days

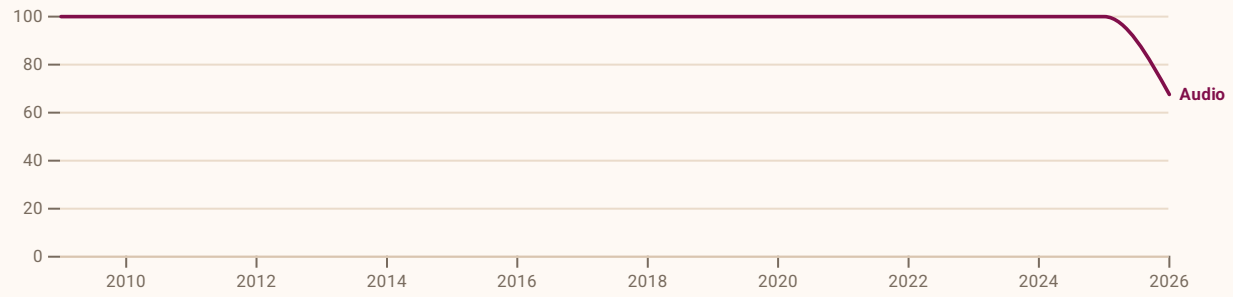
37,163
stories with an audio entry

Last-activity stamp by publication year

Published	Stories	Stamp later than publication	Distinct stamp days
2009	210	210	1
2010	558	558	3
2011	1,583	1,583	1
2012	1,597	1,597	1
2013	2,501	2,501	1
2014	2,414	2,414	1
2015	1,482	1,482	1
2016	2,012	2,012	1
2017	2,655	2,655	1
2018	3,046	3,046	1
2019	1,332	1,332	1
2020	515	0	0
2021	3,103	0	0
2022	3,236	0	0
2023	3,287	0	0
2024	3,204	0	0
2025	3,024	0	0
2026	2,078	10	1

99.9% of the stories with a later stamp share the single day 2019-06-19, which looks like a bulk update (for example a site migration) rather than activity; the stamp does not match the time of the last comment either. It is therefore not used to measure lifetime. Re-read this against the final database.

Stories with an audio entry, by publication year



% of stories published that year. The stored value is the site's standard marker (na) for 37,158 stories, a file name for 5, and empty for 674, all published in 2026. What the marker means on the site is not documented in the data, so coverage means "has an entry". Audio files are not downloaded.

15 · RETENTION

Do commenters come back?

65,496 accounts with a dated clean comment, followed from their first comment. 53% never comment on a second day.

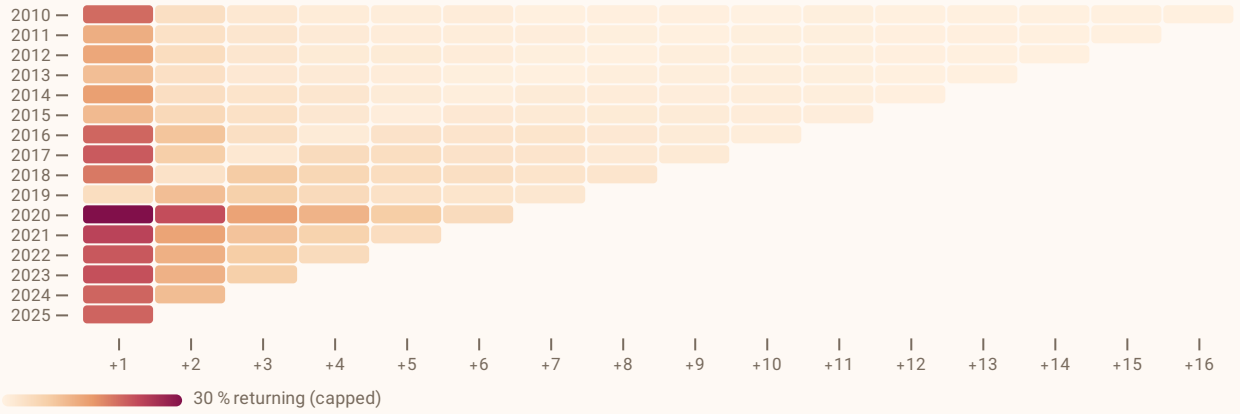
65,496
accounts followed

53%
active for under 1 d

14%
still commenting 1 y after the first

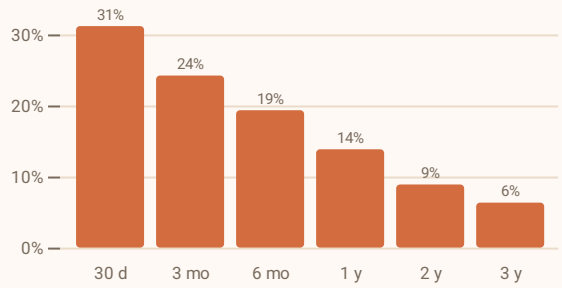
5.4%
active for 3 y or more

Share of each first-year cohort that comments again



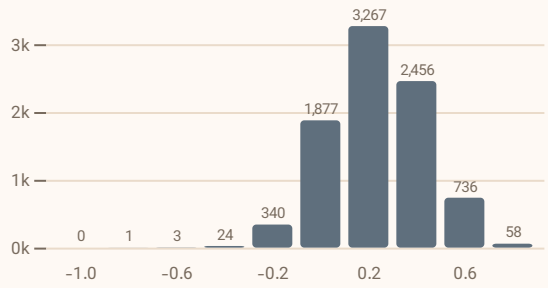
Row: year of the account's first comment. Column: years later. Cell: share of the cohort with at least one comment that year. Shades are capped at 30% so a few high cells do not wash out the rest. Cohorts under 200 accounts are left out, and 2026 is a partial year.

Still active after a given age



Share of accounts whose last comment is at least this long after their first. Only accounts that old by the end of the record are counted.

Rhythm of regular commenters



Burstiness of the gaps between comments, for the 8,762 accounts with 10+ comments: -1 is clockwork, 0 is random, +1 is bursts. 96% are above 0.

How long accounts stay active

Time between first and last comment	Accounts	% of accounts	% of comments
Under 1 d	35,014	53.5%	8.2%
1 d to 30 d	10,201	15.6%	10.2%
30 d to 1 y	11,589	17.7%	28.0%
1 y to 3 y	5,147	7.9%	21.8%
3 y or more	3,545	5.4%	31.8%

Clean comments only (spam and duplicates removed). An account that comments once is active for under a day.

16 · COMMENTER NETWORK

Who comments where

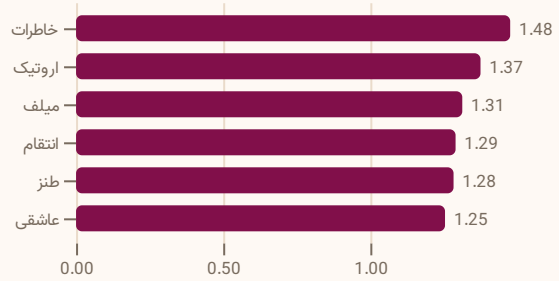
8,762 accounts wrote at least 10 clean comments (76% of all of them). Tags are compared by who comments on them.

8,762
accounts in the network

779
heavy commenters (100+ comments)

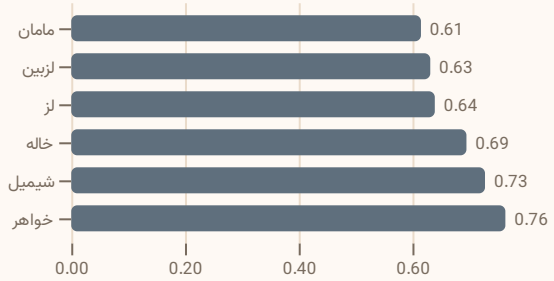
67
tags compared

Tags heavy commenters favour



Ratio of a tag's share of heavy accounts' comments to its share of everyone else's. Above 1 means heavy commenters are over-represented.

Tags heavy commenters avoid



Same ratio, lowest first. Tags need 2,000 comments to be ranked.

Tags that share an audience

Tag	Tag	Shared accounts	Jaccard	Lift
سفر	ماساژ	100	0.29	18.1
خاطرات	میلف	104	0.29	16.7
زن بیوه	میلف	100	0.27	15.8
زندایی	دختر خاله	101	0.23	14.8
زن عمو	زن میانسال	103	0.25	14.1
ارباب و برده	فوت فتیش	113	0.23	13.8
زندایی	عمه	110	0.24	13.4
مادرزن	عمه	101	0.23	13.1
اقوام	عمه	113	0.24	13.1
زن عمو	زن بیوه	110	0.25	13.1
همکار	زن بیوه	148	0.29	12.9
زن عمو	زن داداش	114	0.26	12.8

An account is in a tag's audience when it commented at least 3 times on stories with that tag. Lift compares the shared audience with what chance would give; pairs need 100 shared accounts. Tags with sensitive names are left out.

Why there is no reply graph. The database has no parent-comment id, so replies could only be rebuilt from "@name" mentions. Just 747 of 536,919 comments (0.14%) contain one, 174 start with one, and only 68 name the user name or display name of an earlier commenter on the same story (median 12 h later). The other 91% are mostly contact handles in advertisements. That is too thin to draw threads, so the analysis stops at this count.

17 · SERIES

Stories in parts

1,939 series found from part numbers at the end of titles, holding 5,165 stories (13.7% of all).

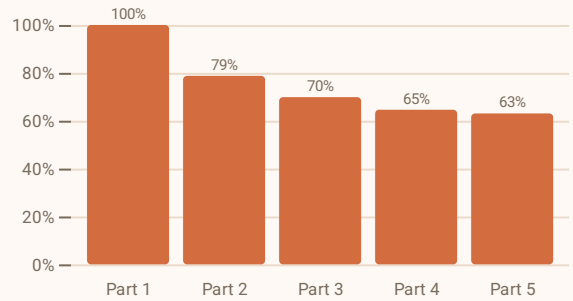
1,939
series of two or more parts

14%
of stories are in one

43%
of series have three or more parts

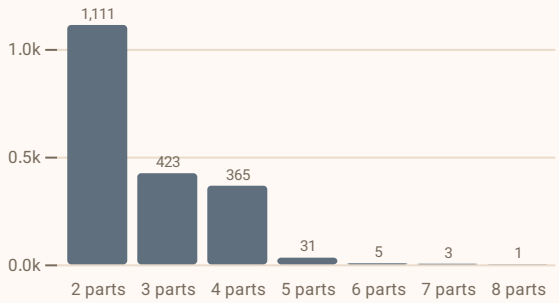
79%
views of part 2 against part 1 (median)

Views of each part against part 1



Median across series that have both parts. Later parts are also newer, so this mixes interest with age.

Series by number of parts



Series with at least 2 parts; a part is counted once even if the part numbers have gaps.

Most viewed series

#	Title	Parts	First year	Total views	Last part vs first
1	سکس با مامان	4	2022	2,680,604	65%
2	مامان پری و پارسا	4	2023	2,256,004	101%
3	حموم با مامان	4	2022	2,072,204	65%
4	سکس یواشکی منو خواهرم	3	2024	1,719,403	60%
5	ماجرا های مامانم و کون دادن من	4	2020	1,490,504	83%
6	مامان شکوفه	4	2021	1,452,704	65%
7	مامان لیلا	4	2022	1,356,204	93%
8	حسام و مادر دوستش	4	2014	1,322,599	15%

Minor-adjacent titles are hidden; the stem is the title without its part number.

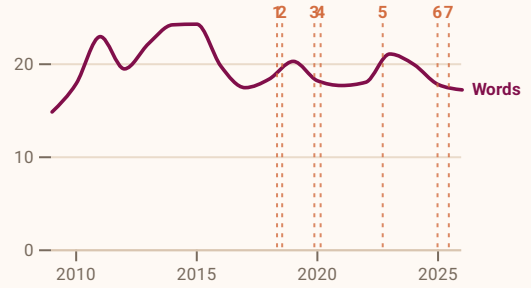
Reading it. The median story in a series ranks at the 46th percentile of its publication year on views, against the 51th for standalone stories. 29% of series stories carry the serial tag, which covers 2,889 stories in all, so 49% of tagged stories show no part number in the title. 92 stems were skipped because two stories shared a part number (different stories with the same title).

18 · STYLE

How stories are written

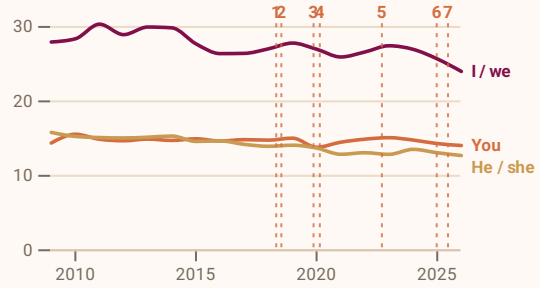
37,294 of 37,837 stories have at least 300 words and are measured. Yearly lines use years with 100+ such stories.

Sentence length



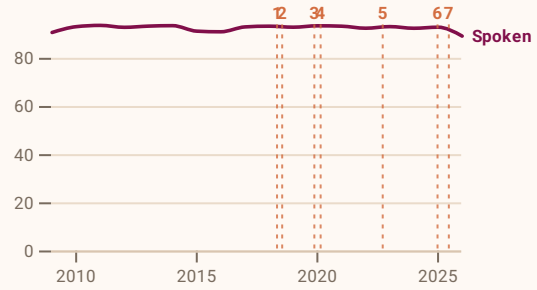
Median words per sentence; sentences end at a full stop, question or exclamation mark, ellipsis or line break. Numbered lines: events listed in section 2.

Who is speaking



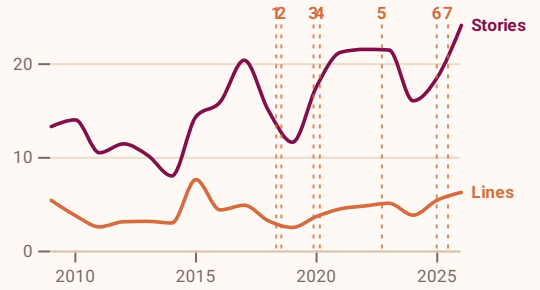
Pronouns per 1,000 words, from the lists in config.toml (including spoken forms). Numbered lines: events listed in section 2.

Spoken or formal



Spoken forms as % of spoken plus formal markers (for example the spoken and formal object marker). Numbered lines: events listed in section 2.

Dialogue



Stories: % of stories with a quoted or dashed line. Lines: mean % of a story's lines that are such lines. Numbered lines: events listed in section 2.

Style against popularity

Feature	Spearman with view percentile
Story length	0.05
Sentence length	0.12
Word length	-0.02
Vocabulary variety	-0.08
Share of dialogue lines	-0.12
First-person pronouns	0.03
Third-person pronouns	-0.05
Colloquial share	0.01

37,294 stories; view percentile is the rank of views within the publication year.

Tags at the ends of the range

Most dialogue	% of lines
اروتیک	13%
اجتماعی	13%
دنباله دار	13%
بی دی اس ام	11%
انتقام	9%

Most formal	% spoken
طنز	97%
ارباب و برده	98%
دکتر	98%
بی دی اس ام	99%
خاطرات	99%

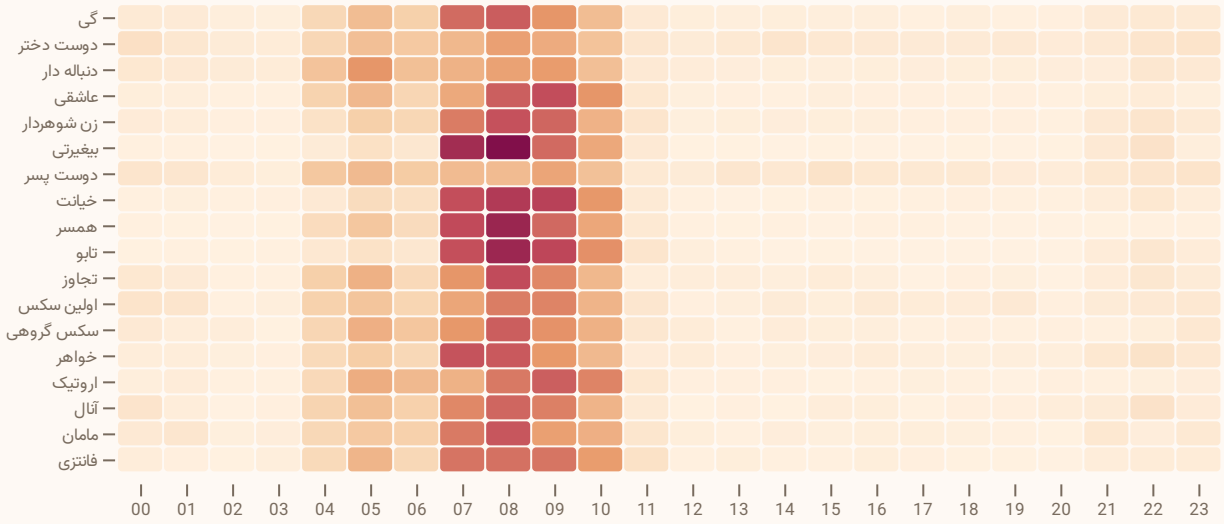
Tags with 100+ measured stories; medians per story.

19 · TAG PROFILES

When each tag is published

Hour of day and day of the week for the 18 most used tags. Each row is that tag's own distribution, so rows show habit, not volume.

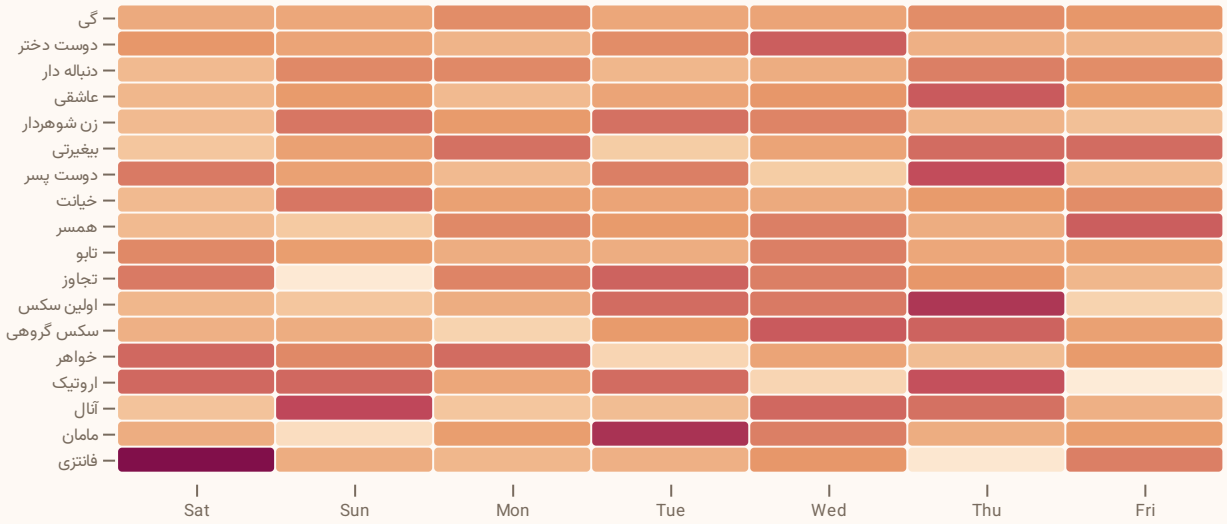
By hour of day



0 — 29 % of the tag's stories

Tehran time at a fixed offset; one cell is the share of the tag's stories published in that hour.

By day of the week



11 — 17.9 % of the tag's stories

The Iranian week starts on Saturday. An even spread would be 14.3% a day; the colour scale starts at the lowest value.

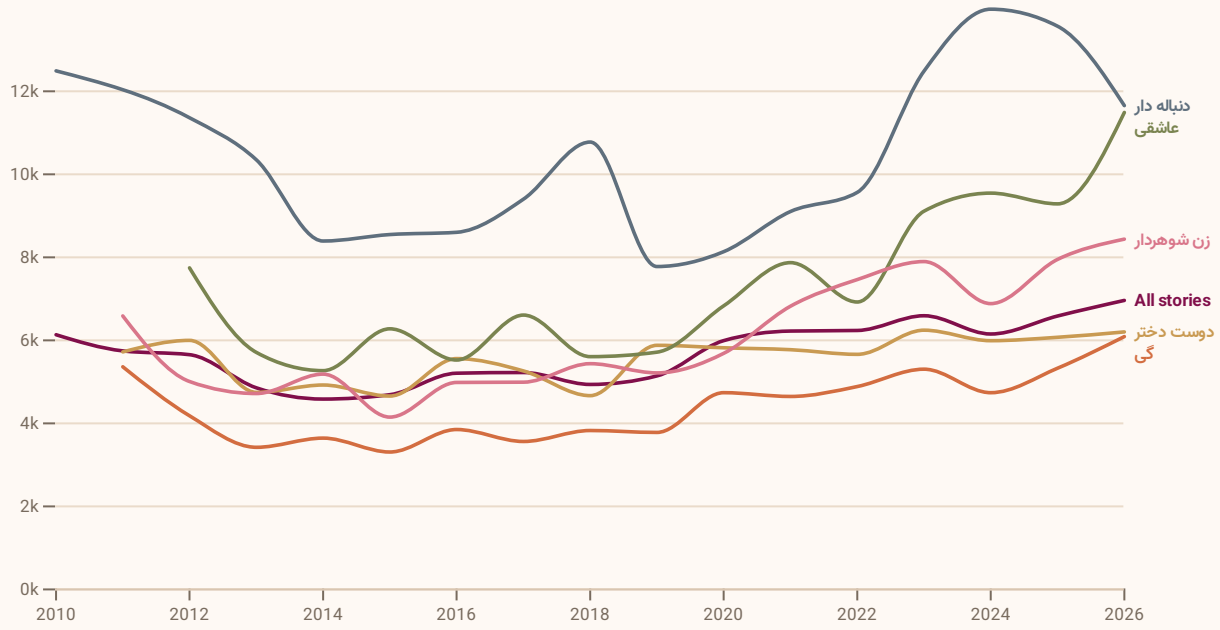
Reading it. دنباله دار stories peak earliest in the day (05:00) and اروتیک stories latest (09:00).

19 · TAG PROFILES

Story length by tag and year

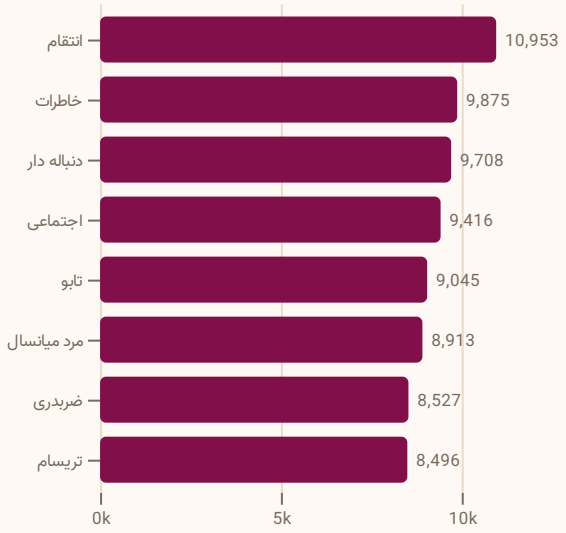
Median story length in characters (HTML removed). Tags with at least 150 stories; a tag's yearly median needs 20 stories in that year.

Median length per year

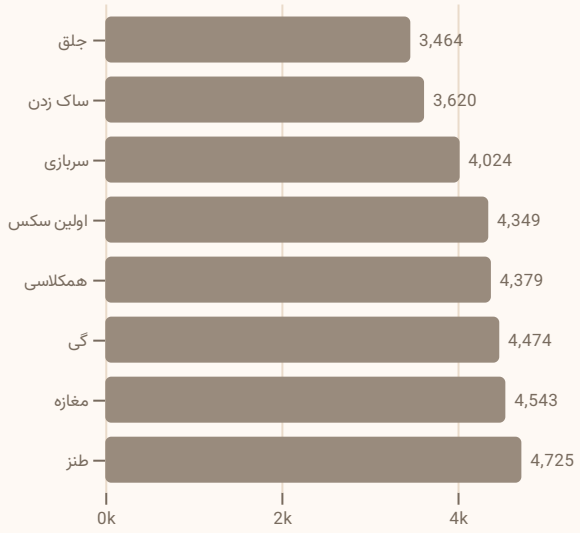


Characters. Years with fewer than 300 stories are left out. The average story-weighted yearly median is 5,760.

Longest stories by tag



Shortest stories by tag

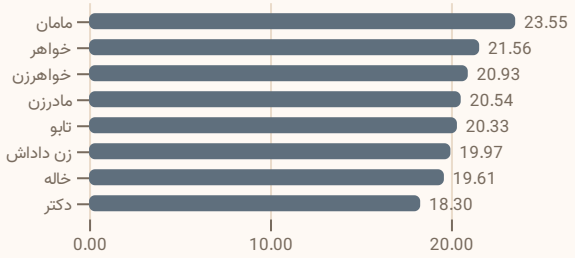


19 · TAG PROFILES

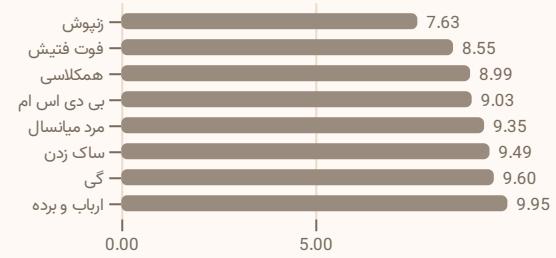
Discussion and first replies by tag

How much each tag is talked about, and how fast the first comment arrives. Comments of every kind are counted here, spam included.

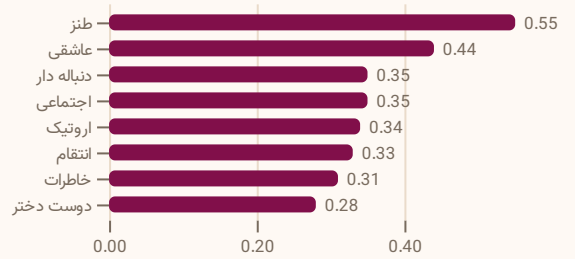
Most comments per story



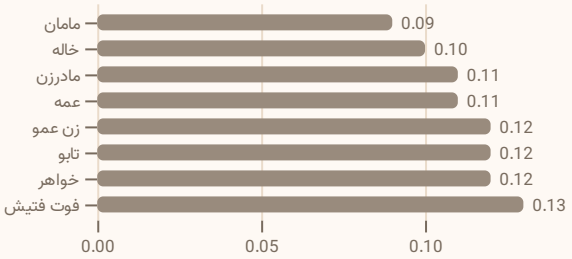
Fewest comments per story



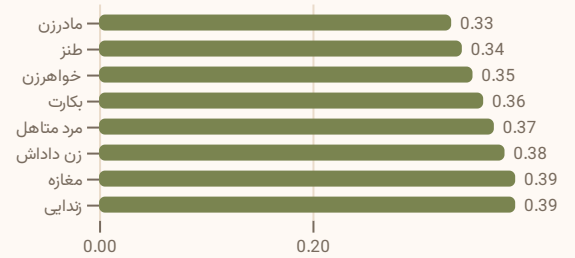
Most discussed per 1,000 views



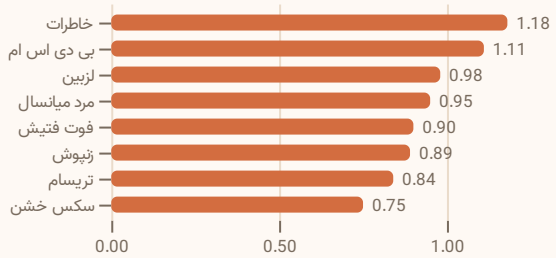
Least discussed per 1,000 views



Fastest first comment (median hours)



Slowest first comment (median hours)



Across all tagged stories there are 13.5 comments per story on average. Delay figures use stories with a comment dated after publication and at least 50 such stories per tag; a story with several tags counts once per tag.

19 · TAG PROFILES

Likes and dislikes over time

Votes on stories by the year the story was published, for years with at least 300 stories. Older stories have had longer to collect votes.

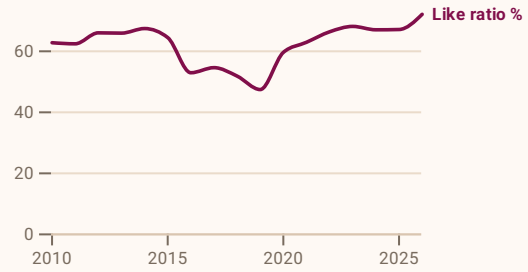
63%
like ratio, 2010

72%
like ratio, 2026

1.7 to 7.3
likes per story, 2015 to 2016

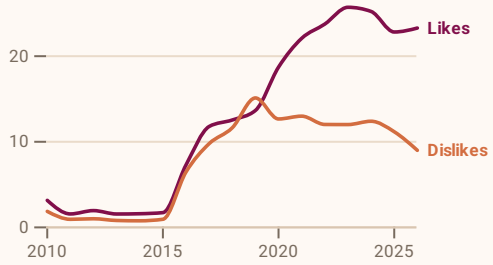
A break in the record. Likes per story jump from 1.7 to 7.3 between 2015 and 2016, and stories with no vote fall from 41% to 1%. That points to a change in how votes were recorded, not to readers suddenly voting more, so compare votes within the same era.

Like ratio

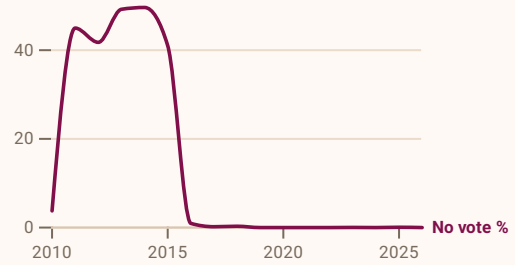


Likes as a share of likes plus dislikes, in %.

Votes per story

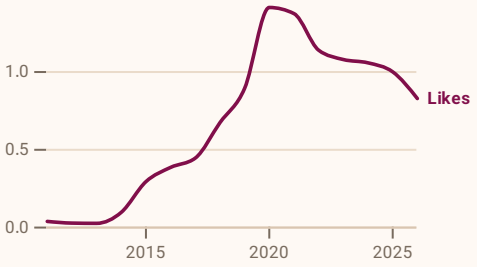


Stories nobody voted on



Share of the year's stories with neither a like nor a dislike, in %.

Likes per comment



Comments are grouped by the year they were written; years with at least 5,000 comments.

20 · THEN AND NOW

What changed

The first and the last 3 years with enough stories, side by side: 2010–2012 against 2024–2026.

×2.2
stories published per year

×1.13
median story length

×0.52
comments per story

The same measures, then and now

Measure	2010–2012	2024–2026	Change
New stories per year	1,246	2,769	×2.22
Comments written per year	23,844	31,622	×1.33
Comments per story (all time)	20.9	10.7	×0.52
Median story length (characters)	5,765	6,514	×1.13
Median title length (characters)	16.1	19.7	×1.22
Tags per story	0.56	1.53	×2.76
Like ratio	64.0%	68.4%	+4.3 pp
Stories with no vote	37%	0%	-37.4 pp
Median hours to first comment	2.3	1.1	×0.49
Likes per comment	1.72	0.96	×0.56

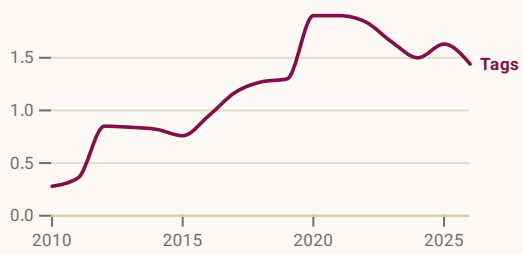
Top tags, 2010–2012

Tag	Share of stories
دوست دختر	13.5%
گی	6.3%
دوست پسر	5.6%
زن شوهردار	2.4%
دنباله دار	1.9%
عاشقی	1.6%
اقوام	1.3%
سکس گروهی	1.3%

Top tags, 2024–2026

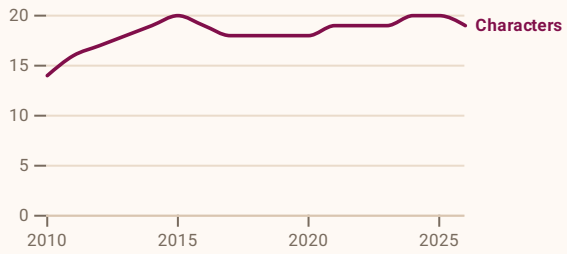
Tag	Share of stories
گی	22.9%
بیغیرتی	11.4%
زن شوهردار	7.9%
عاشقی	6.0%
تابو	5.5%
خیانت	5.5%
همسر	5.3%
دوست دختر	4.7%

Tags per story, by year



Average number of tags on a story published that year.

Median title length, by year



Characters in the median title of stories published that year.

Vote measures (like ratio, no-vote share, likes per comment) are not comparable across the 2015 to 2016 break in how votes were recorded; see the page on likes and dislikes. Per-story figures average the yearly values weighted by stories; medians are averages of yearly medians. Comments per story and votes grow with a story's age, so recent stories are at a disadvantage on those measures.

21 · PREDICTING ATTENTION

How well can attention be predicted?

A gradient-boosting model reads what is known when a story is published (tags, title length, body length, hour, weekday, year and month) and predicts its view percentile within its publication year. It is tested on 7,556 stories it never saw.

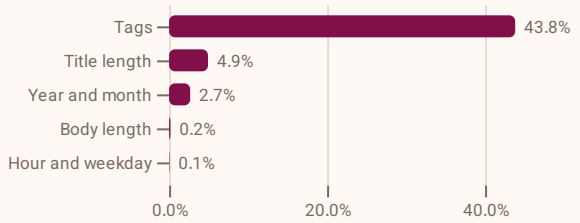
51%
of the variance explained (test R²)

0.71
rank correlation, predicted vs real

59%
of predicted top 20% really are (chance 20%)

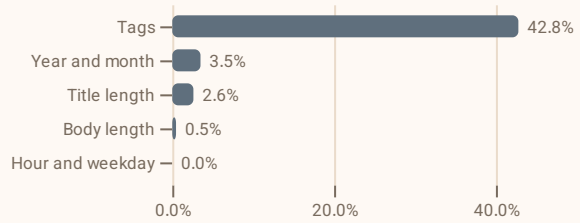
Setup. 37,837 dated stories are split by a hash of the story id into 24,721 for training, 5,560 for early stopping and 7,556 for testing. The model uses 75 inputs (67 are tag flags) and stopped after 899 trees. Error is 0.202 against 0.288 for always guessing the average. This describes association, not cause.

Each group on its own



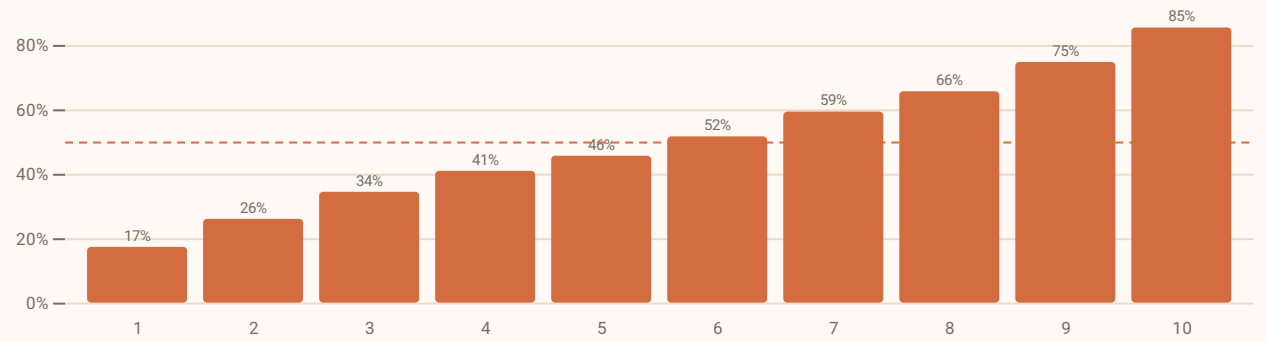
Test R² of a model that sees only this group.

Lost when the group is removed



Full-model R² minus the R² of a model without the group. Negative values are shown as zero.

Predicted decile against what happened



Test stories sorted by predicted percentile and cut into ten equal groups; each bar is the real mean view percentile. The dashed line is 50%.

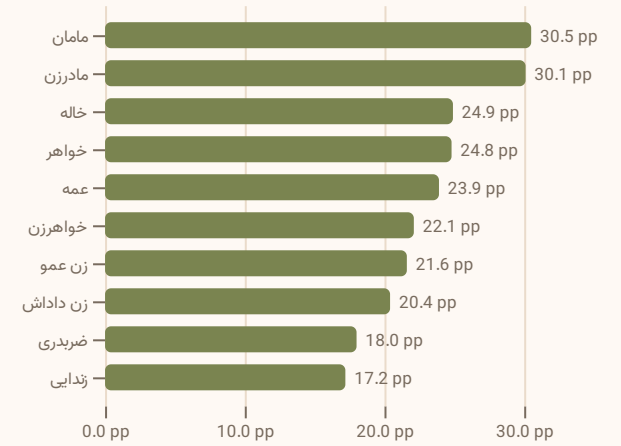
The strongest group, tags, reaches 43.8% alone; the whole model reaches 51.0%. Only the tags with at least 150 stories enter the model, so the tag count counts those only.

21 · PREDICTING ATTENTION

What moves the prediction

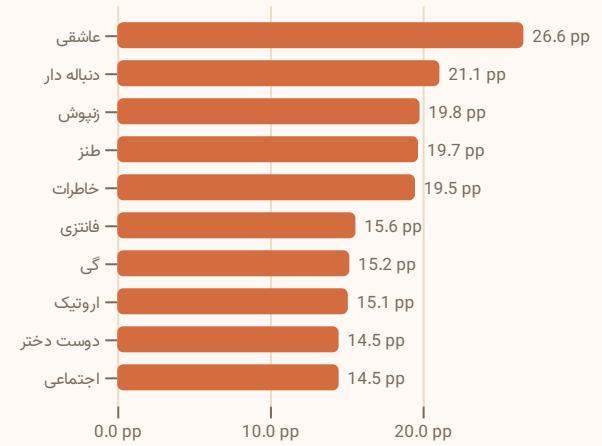
Effects are averaged over a fixed sample of stories: change one input, keep the rest, and see how the predicted view percentile moves. They show what the model learned, and tags overlap, so read them as hints.

Tags that push views up



Change in predicted percentile (points) when the tag is added.

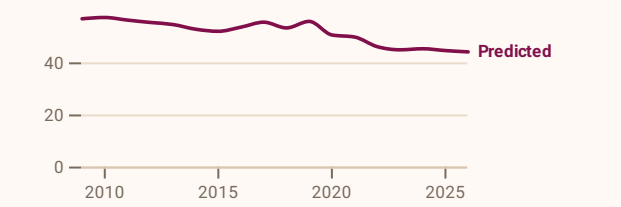
Tags that push views down



Points of predicted percentile lost when the tag is added.

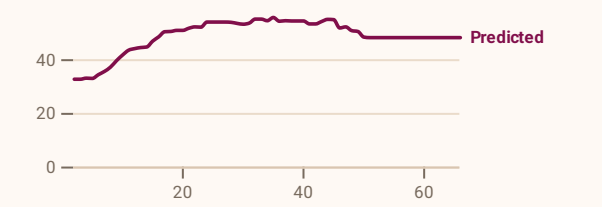
Reading it. Adding مامان moves the predicted percentile by 30.5 points and adding عاشقی by -26.6. Tag choice is by far the strongest signal; hour, weekday, month barely move it.

Year



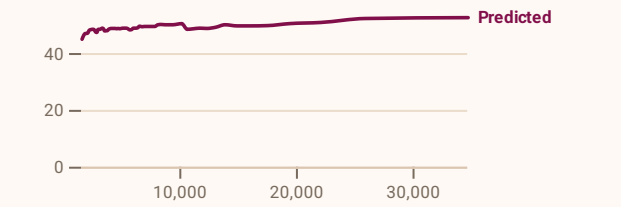
Average predicted percentile (0 to 100) by publication year.

Title length



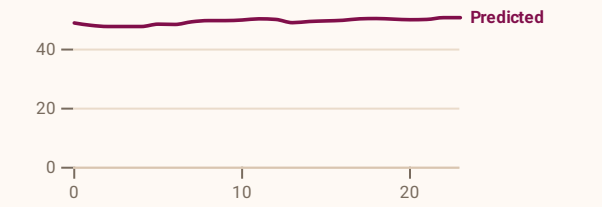
By characters in the title.

Body length



By characters in the body, shown up to 50,000 (longer stories are rare and the curve stays flat).

Hour of day



Tehran time at a fixed offset.

Percentile is within publication year, so the year curve shows which eras held more of the stories that went on to be read, not time itself. The model is a fixed-size gradient-boosted ensemble; its settings are in config.toml.

22 · METHOD

Method and caveats

How the numbers were made, and what they cannot tell you.

Method

Topic	What was done
Source	One SQLite snapshot (checksum e647be3c3fbb…). The collection process is not documented, so absence of a row is not evidence of deletion on the site. Counts such as views are as of collection.
Cleaning	Negative like counts are clipped to zero; comments dated before their story (254) are excluded from delay figures only; HTML is stripped from bodies before tokenising.
Tokens	Persian letters only, with Arabic ی ك ي folded to ک ی respectively, zero-width joiners treated as spaces, diacritics and tatweel dropped and runs of 3+ repeated letters collapsed. Words under 2 letters are ignored. Latin tokens of 3+ letters are counted apart.
Spam	A comment is flagged when its normalised text repeats at least 5 times with at least 6 words, or contains a link marker. 2,818 comments were flagged.
Hours	Hour of day uses one fixed offset of UTC+3:30 for every year.
Thresholds	Tag statistics need 150 stories per tag; tag pairs need 60 shared stories; title words need 150 titles.
Profiles	Gender, province and city are taken as stated. Provincial rates divide profile counts by Statistical Centre of Iran, 2016 census (rounded); this is the only external figure used.
Forum places	Places are the cities named by at least 150 profiles plus every province, matched as words in topic titles; a trailing ی counts as a match.
Views per day	Age is measured to the end of the record; stories under 30 days old are not rated. The last-activity stamp is not used (see section 14).
Retention	Accounts are followed through clean comments from their first to their last. Survival counts only accounts old enough to be observed for the age asked. Burstiness needs 10+ comments.
Series	A series is a set of stories whose titles share a stem and end in part numbers (a bracketed number, a marker word and a number, or an ordinal word). Stems where two stories share a part number are skipped. Authors are unknown, so two unrelated stories with the same stem and different numbers can be joined by mistake.
Style	Style features use stories with at least 300 Persian words. Pronoun, spoken and formal markers are exact-word lists in config.toml, so they are indicators, not a grammar analysis.
Outside events	Seven dated events (messenger bans and shutdowns, Covid-19, protests, the 2025 war) are marked on date-axis charts and listed in section 2. They come from the public record, not from the database, and are set in config.toml. They show timing only, not cause.
Reply mentions	There is no parent-comment id, and only 68 comments name an earlier commenter with an "@" mention, so no reply graph was built (section 16).
Sensitive content	Explicit terms are shown as they appear in the corpus (private report). Titles that mention minors are hidden from every listing and tags of that kind are excluded.

Caveats

- This is one snapshot, and the database is still growing: figures will change when it is replaced.
- 37,287 of 65,665 commenting accounts have no profile row, so profile-based comparisons cover only the matched part.
- 10 profile columns are empty in this snapshot and are not analysed.
- Views, likes and dislikes are cumulative totals at collection time, so recent stories have had less time to gather them; percentiles within publication year reduce, not remove, this bias.
- The forum table has no author column, so forum figures count topics, not people. Photo uploads are dominated by very few accounts.
- Correlations describe association only. Popular stories draw both likes and dislikes because both need a viewer.
- Word rankings are shaped by spelling variants, names and series; the function-word list is hand-built and imperfect.

22 · GLOSSARY

Glossary

Terms used in the report.

Term	Meaning
View percentile	Rank of a story's views among stories published in the same year (0 to 100), which removes the advantage of age.
Like ratio	Likes divided by likes plus dislikes, with negative counts clipped to zero.
Gini coefficient	Inequality of views across stories: 0 means equal, 1 means one story has everything.
Spearman correlation	Rank correlation between two measures; 0 means unrelated, 1 means the same order.
Lift	How much more often something occurs inside a group than in the whole corpus; 1 means no difference.
Jaccard	Shared stories divided by stories carrying either tag, for a pair of tags.
Per million words	Word count divided by the year's Persian word total, times one million, so years of different size compare.
Function words	Pronouns, particles, clitics and auxiliary verb forms removed before ranking vocabulary.
Clean comments	Comments left after removing repeated copy-paste texts and link spam; used for word and style statistics.
Heavy commenter	A commenting account with at least the configured number of comments (see section 10).
Views per day	A story's total views divided by the days between publication and the end of the record; only stories above a minimum age are rated.
Burstiness	Measure of how uneven the gaps between one account's comments are: -1 is perfectly regular, 0 random, +1 comes in bursts.
Type-token ratio	Distinct words divided by words in a fixed-length stretch of text (here the first words of each story); higher means more varied vocabulary.
Zipf slope	Slope of word frequency against rank on log-log axes; about -1 is typical of natural language.