KuaiShou Livestream Analysis – May-25 Trend

Created by Daniel Nguyen

Email: ***hoangfotu@gmail.com***
Linkedin: *linkedin.com/in/hoangfotu*

Content

Current Analysis

1. **Objectives**
2. **Quick overview of Data Preprocessing**

3. **Exploratory Data Analysis**

3.1 General Overview
3.2 Demographic
3.3 Streamer Activity Analysis
3.4 User Interaction Analysis
3.5 Trend Analysis
3.6 Which Categories has the highest CTR and Gift Amount?

4. User Funnel Analysis

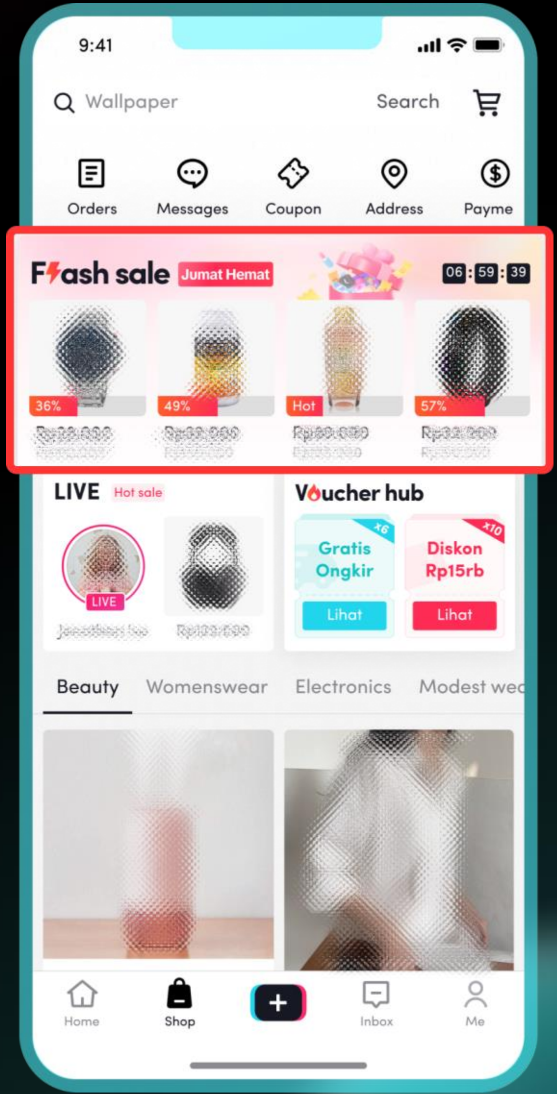
4.1 Assume User Journey: Session -> Click -> Comment -> Like -> Gift
4.2 Sankey Diagram

5. CTR Recommendation
6. Appendix

Upcoming

Applying Machine Learning techniques to predict…

1. CTR for each categories in next days
2. Content Recommendation system for specific userid



1. Objectives

Background

- KuaiShou is a leading livestreaming platforms in China, with 400 Millions DAU (daily active user), competing directly vs. TikTok and Taobao Live.
- KuaiShou's main target audiences are users in rural area of China, and main monetization strategies are user gift donations and e-commerce
- In this dataset, we have dataset records interaction logs of 23,772 users and 452,621 streamers over 21-day period in May-25

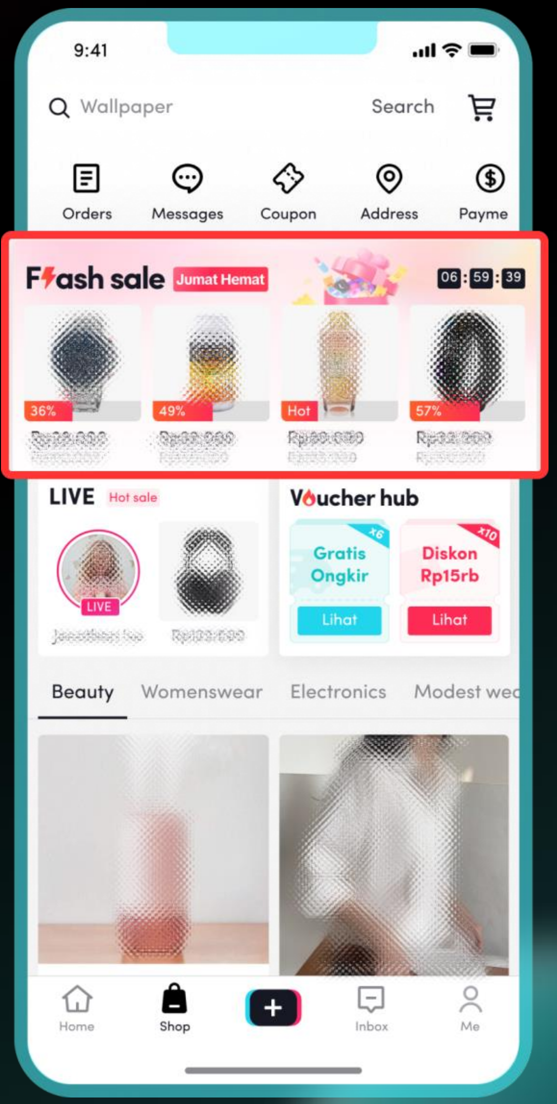
Objectives

There are 2 parts: Data Analysis and Predictive CTR / Content Recommendations

Part 1: Data Analysis

1. What are our streamers and users' characteristics overall?
2. What are the overall trend in 21 days?
3. What are the CTR (click-through-rate) overall and by each categories
4. What is the user journey? How should we do to optimize the monetization – the CTR of gift donation?

Part 2: Predictive CTR / Content Recommendations (Will implement later)



2. Quick Overview of Data Processing

Data Description

- We have 8 csv files, containing streamer, user and room information. We also have the detailed interactions negative (means user skips the content), gift, like, click, comment
- Details of the data description, please see the Appendix

```
KuaiLive
├── streamer.csv
├── room.csv
├── user.csv
├── comment.csv
├── gift.csv
├── like.csv
├── click.csv
├── negative.csv
└── title_embeddings.npy
```

Data Cleaning

- When exploring, I found some live_id has more than 1 day live duration, some even rich 20 days. There are 14,613 lives like this, contribute 0.1% of all live_id

→ I decided to remove these live_id out of all datasets, by rows level

- Transform date type into datetime format in each tables

Data Modelling

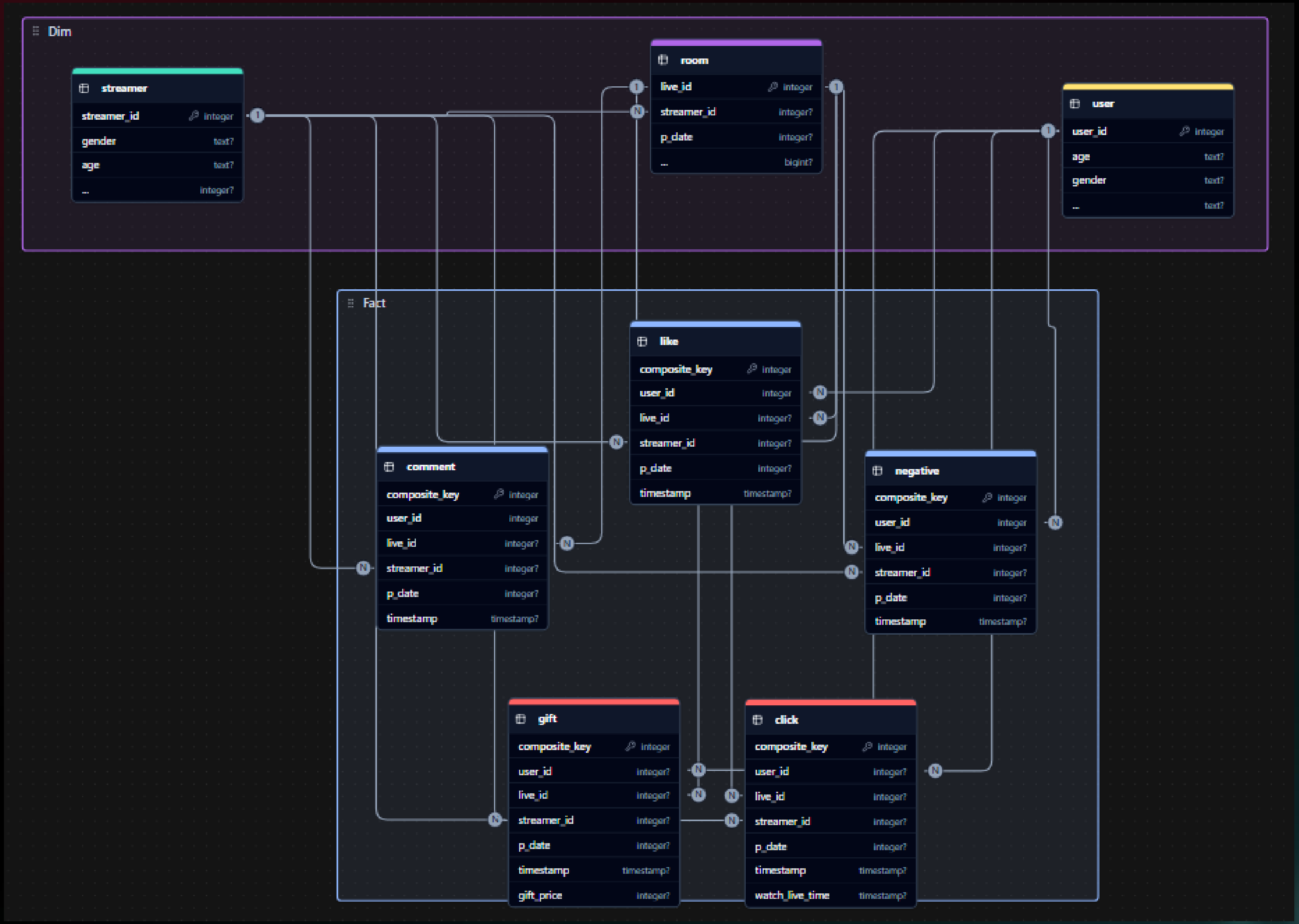
- Create the dim and fact tables, with the logic is that interaction records means fact data: dim_streamer, dim_user and dim_room, with fact_comment, fact_gift, fact_like, fact_click, fact_negative

Realize the needs to analyze all interactions together:

- I decided to create the **OBT (One Big Table),** that merge necessary information from dim tables and all fact tables together.
- Create column session_id from live_id and user_id to track interactions in each user session.
- Order each engagement in each user sessions by time – easier to visualize user journey.

2. Quick Overview of Data Processing

Data Model



3. Exploring Data Analysis

3.1 General Overview

Overall

total_users	total_streamers	total_users_also_streamers	sessions_per_day	lives_per_day	sessions_per_live	duration_per_live	clicks_per_day	comments_per_day	gifts_per_day	likes_per_day	skips_per_day
23,772.0	452,621.0	11,702.0	593,969.0	552,374.2	1.1	91.6	232,791.0	9,334.3	3,449.4	8,514.4	603,233.7

User Metrics

avg_active_time_per_user	sessions_per_user_per_day	lives_watch_per_user_per_day	time_spent_per_user_per_day	clicks_per_user_per_day	comments_per_user_per_day	gifts_per_user_per_day	likes_per_user_per_day	skips_per_user_per_day
1875.15	13.34	6.06	20.92	15.21	1.85	1.31	1.77	29.08

Streamer Metrics

avg_active_time_per_streamer	lives_per_streamer_per_day	live_duration_per_streamer_per_day	clicks_per_streamer_per_day	comments_per_streamer_per_day	gifts_per_streamer_per_day	likes_per_streamer_per_day	skips_per_streamer_per_day
1375.33	2.14	195.83	3.12	1.33	1.22	1.31	6.95

Overall

- After cleaning, we have 23.7k users, with 452k streamers - Quite a very imbalance number. **Streamers = x20 times users** - Is it a sign of unsustainability?
  → Kuaishou really need to focus on acquiring New Users! Especially the younger generations
- 11.7k users also streamers = 50% penetration of total users, the trend that users are converting to streamers is going quite well
- Avg. user sessions per live is 1.1. Avg duration per live is around 91 min - quite a long live. We see that skips_per_day = 3x clicks_per_day, so users skip a lot
  → We need to understand user behavior and recommend appropriate content for the users
- Conversion rate starting from the Click, **Like & Comment is around 3.6-4.0%** - which is quite a good figure. **Gift is 1.48%**, this is where we do the monetization!

Users

- User base has been active for avg 5 years (1,875 days) - quite a long time
- 1 user watch avg 6 lives per day, ~21 min time spent on the platform. In Kuaishou Q2'25 report with 126.8 minutes watch daily, it is a **BIG off**. I think they have a different way of reporting, such as measuring only loyal user

→ Therefore, we should not include any outside reports as benchmark, just focus on our current dataset

- Comments and Likes rate quite the same, even Comments (1.85) more than Likes (1.77). I think users would Click → Comments → Likes → Gifts!

Streamers

- The avg active time of streamers is 3.7 years (1,375 days). They did avg 2 lives per day, each streamers live for 3 hours+ (195 minutes).

→ The streamers were very hardworking!

- However, we see that their engagements were very low. I believe the data is highly-skewed.

3.2 Demographic

```
User:  gender
M    61.854282
F    38.145718
Name: count, dtype: float64
Streamer:  gender
F    62.828503
M    37.171497
Name: count, dtype: float64
```

Gender

Among users, 61.85% are male, while for streamers, 62.8% are female.

→ The app is more favored for Male users, who love to see female streamers

```
User:  age
50+      24.284873
31-40    19.590274
12-17    16.191318
18-23    14.874642
41-49    14.862023
24-30     9.452297
0-11      0.744573
Name: count, dtype: float64
Streamer:  age
31-40    34.727951
41-49    19.330522
50+      17.733380
24-30    13.907441
18-23    13.366591
12-17     0.893684
0-11      0.040431
Name: count, dtype: float64
```

Age

- Interesting! Majority users of this app are people from 30+ (60% of total users), especially users 50+ occupied 24%
- For streamers age, we also see the same pattern, 70% users are 30+

→ From Wikipedia, 'Compared to TikTok, Kuaishou is more popular with older users living outside China's Tier 1 cities. Its initial popularity came from videos of Chinese rural life. Kuaishou also relied more on e-commerce revenue than on advertising revenue compared to its main competitor'

→ That's why Kuaishow, the most downloaded app with 200M user base, are having more old users than TikTok

```
live_operation_tag
Chat            29.515201
Other           28.760044
E-Commerce      13.346928
Beauty           6.624748
Lifestyle        6.605526
Talent           6.356532
Education        2.579863
Relationship     2.168039
Game             2.079886
Hobbies          1.253808
Fitness          0.332508
Group            0.280146
News             0.096770
```

Streamers registered profile

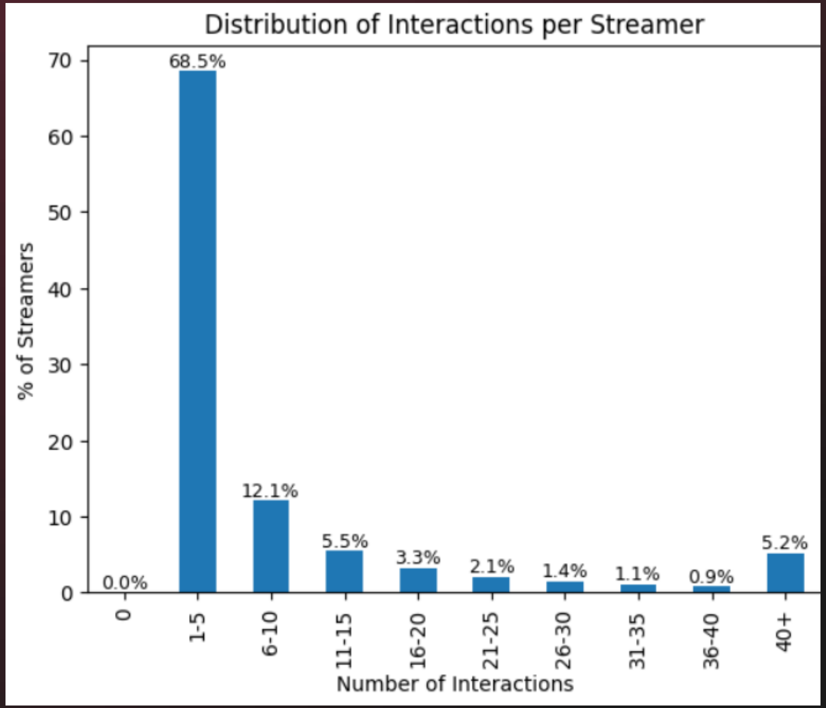
- Of course, as a social app, 29.5% of streamers must be Chat activities, 28.7% Other - we need to investigate more
- E-commerce, as in Wiki said, are major revenue stream of Kuaishow, that why 13.3% streamers has major activities in E-commerce

In general, we see that Kuaishou is a social network for rather older users, living outside China Tier 1 cities:

- That's explain why most users are older man from 40+, and streamers mostly are women also from 30 to 40+
- Most engagements are Chat and E-commerce activitiy. Main revenue sources are Gifting and E-commerce

3.3 Streamer Activity Analysis

68.5% Streamers has below 5 interactions

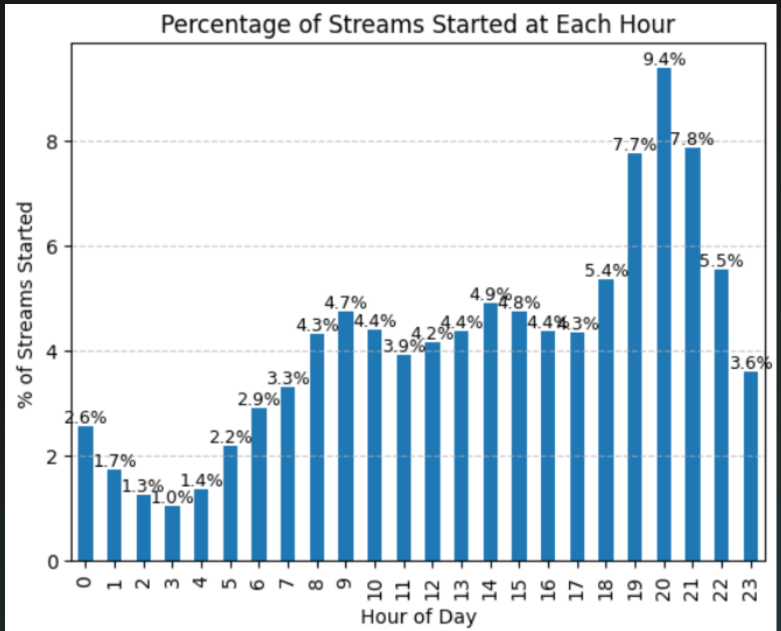


- 68.5% streamers receive less than 5 interactions, while only 5.2% receive more than 40+
- I believe that is the Coldstart problem, where new streamers often get less views and interactions than older streamers
- Therefore, the tagging and recommendation activities are very important to help new streamers get starts and receive more engagements from the users

Streamers with more than 1-year tenure has 8.5+ interactions

```
active_group
0-30         0.572833
31-60        0.678349
61-90        0.440412
91-180       0.756203
181-365      0.829593
365+         8.530669
Name: interactions_per_streamer,
```

- Again, it is very unfair for streamer < 1 year, while Streamer 365+ days has avg. 8.5 interactions, while streamers < 30 days only have 0.57 interactions avg.
- We need a solution for our streamers / creators' development, so we can have a younger streamer generation and also younger user as well!

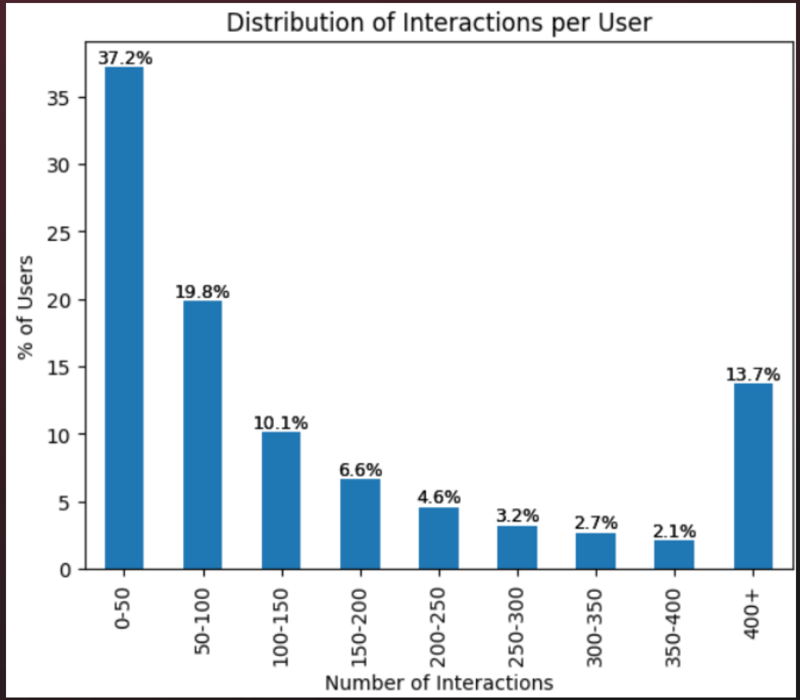


Streamers active mostly at lunch time and afterwork

- Most live time start after 7pm, after work time, and decreasing from 11pm till morning
- Streamers are likely to schedule the time that most users (which are rural 30+ people) often online, to get the most views and interactions
- The recommender system need to take this patterns to the model, which better recommend appropriate content to the users and help streamer perform effectively

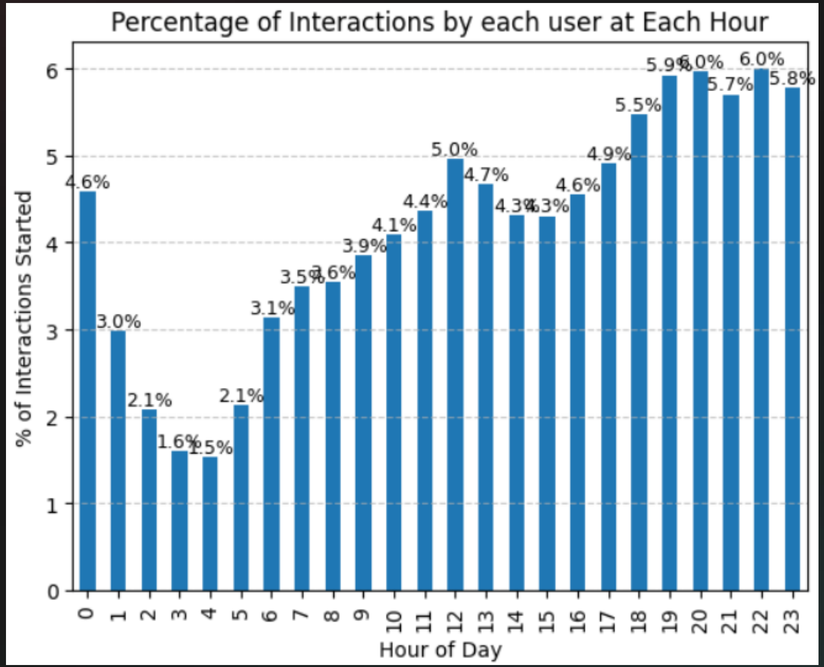
3.4 User Interaction Analysis

37% Users has below 50 engagements, create challenges for recommendation



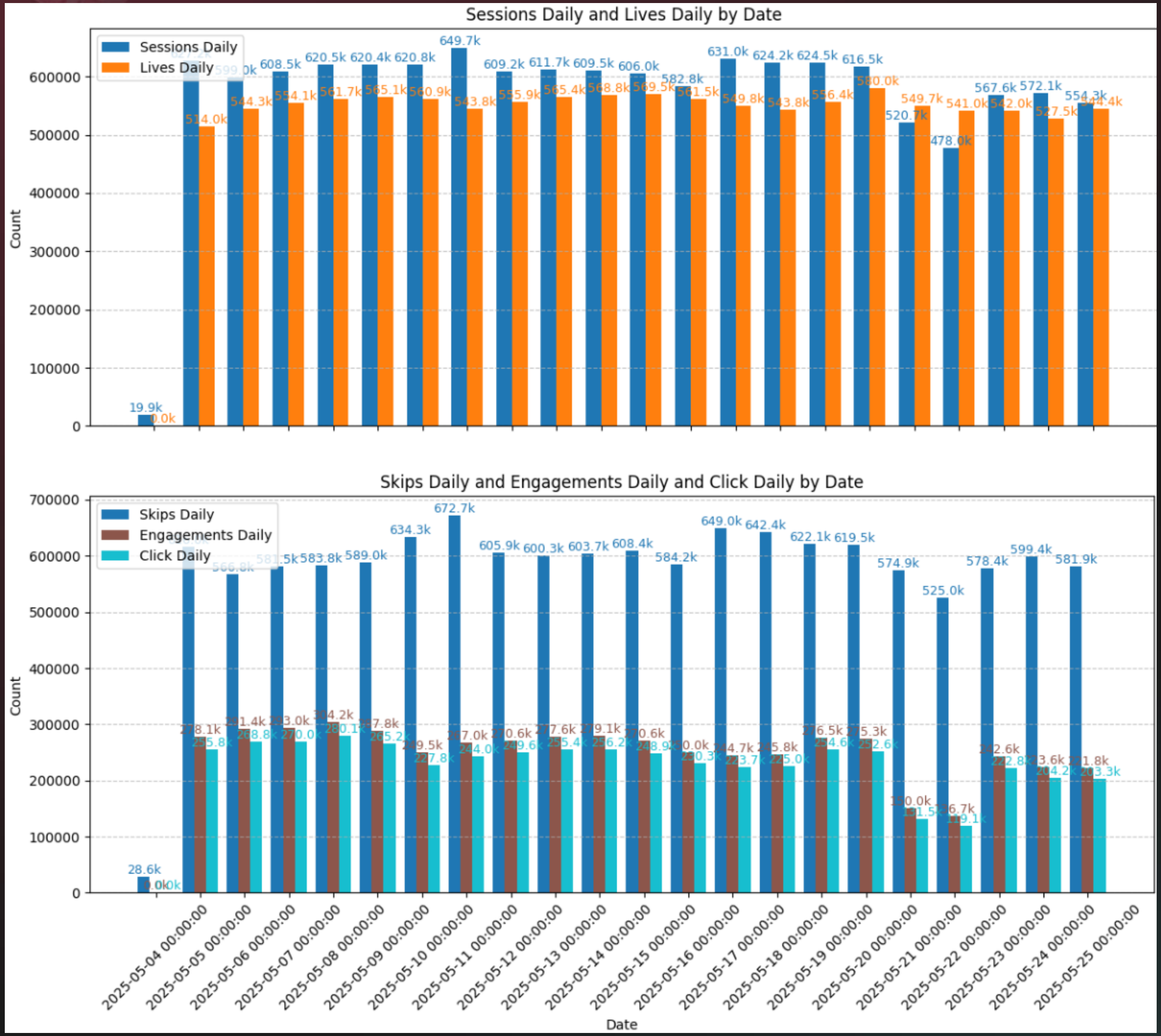
- We can see that 37% of user have less than 50 engagements, while 13.7% have 400+ interactions
- It pose great challenges for recommender system, where we don't have enough informations for recommend to users with less interactions. But for heavy-usage users with diversified content, it's also difficult to recommend CTA content (like gifts) to these users

Users active mostly the same time as streamers live activity



- Same pattern as streamer activity, where most user interactions happen after 6pm till 12am
- The peak in 12pm, maybe it is the resting time of the day for most users, having lunch, playing some games, etc.
- Based on these timing and behavior, we can recommend appropriate content, such as food during meal, or entertaining content like games, movies after works

3.5 Trend Analysis



Sessions vs. Lives Trend from May 5th to May 25th

- Highly consistent: 600k sessions and 540k-570k+ lives per day
- Drop and Recovery: There is a prominent dip on 2025-05-22, where sessions fall sharply to 478k (from previous days consistently above 540k), and lives also drop but less drastically. Both metrics recover quickly in subsequent days, returning to prior stable levels.
- Outliers: on 2025-05-04, I believe it is the data slide cut, so we no need to care about it

Skips vs. Engagement Trend

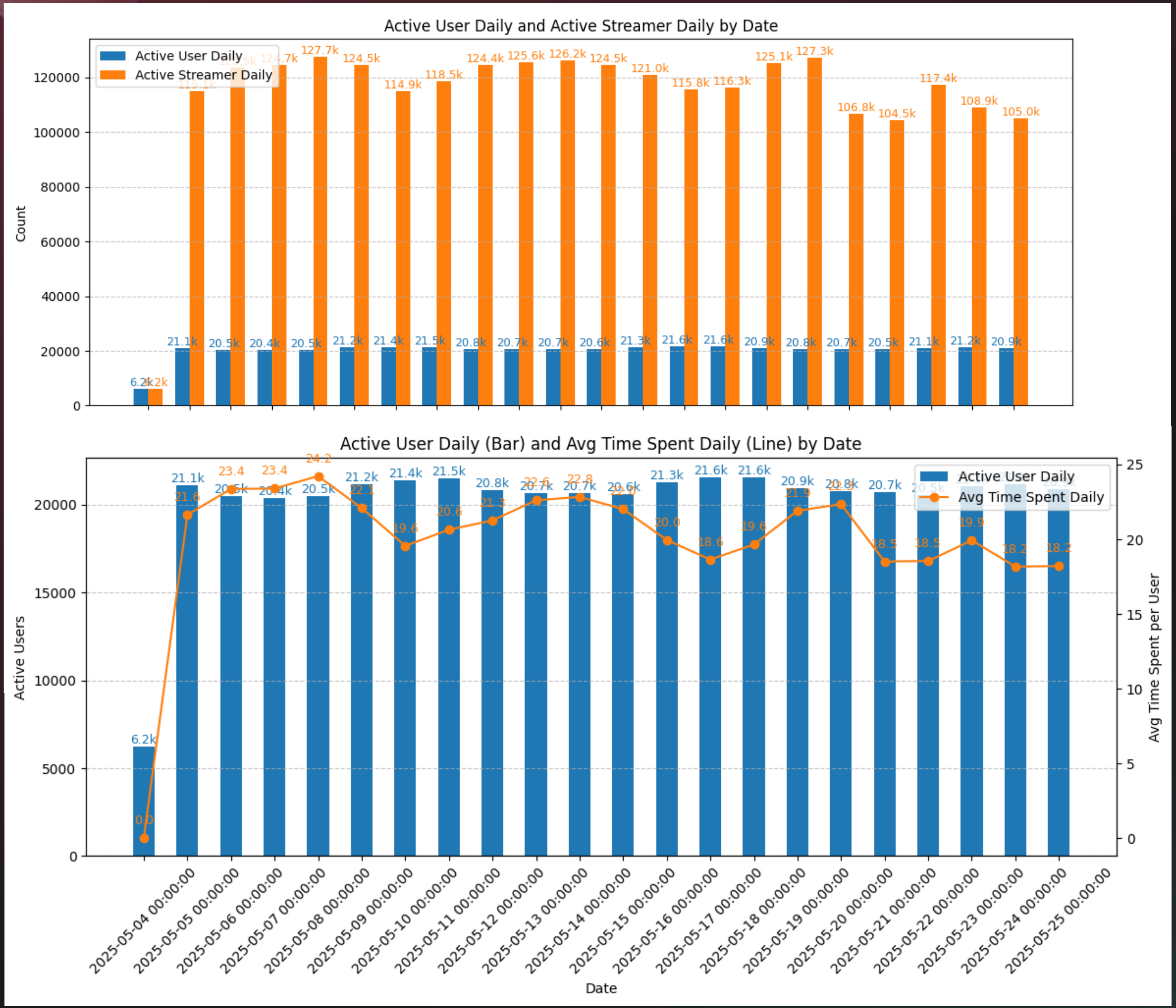
Here the trend is also very aligned with Sessions and Livestreams:

- **The skips are very high**, around 570-670k per day, x3 total engagements.

→ This means majority of user interactions are just sliding past lives without actually watching them. We really need to take care of this!

- Total engagements trend is also the same, 230-275k daily, and a big drop in 2025-05-21 and 2025-05-22.
- Clicks trend follow the total engagement trend, contribute 90% of total engagement

3.5 Trend Analysis



DAU (Daily Active Users) vs. DAS (Daily Active Streamers) Trend from May 5th to May 25th

Nothing very specially as well

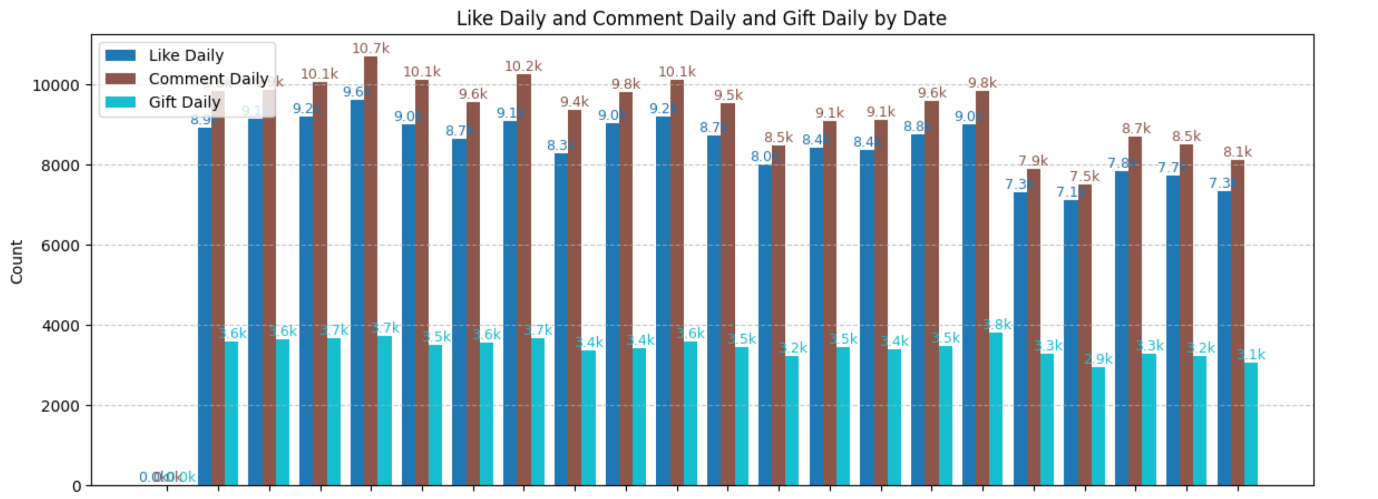
- Active users are very consistent, around 20-21k active users per day
- Active streamers trend also the same, with a drop in May-21st and May-22nd

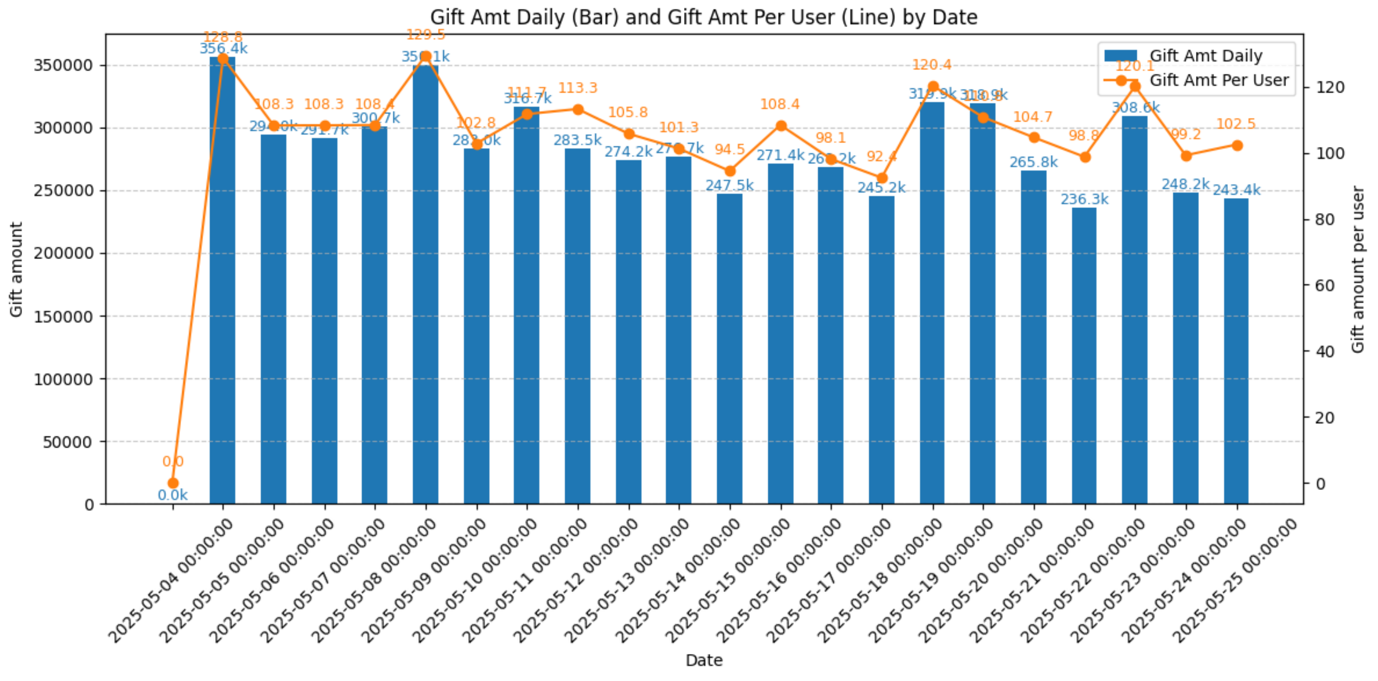
DAU vs. User Time Spent Trend

- The trend continue to be the same, but 1 big problem is that Avg Time Spent of user are in decreasing trend, from 23.4 min on May 5th, to 18.2 min on May 25th. While Active users are still the same!

→ That's a big alarm: The streamers are still active and produce contents, but the content may not catch the user's attention!

3.5 Trend Analysis





Like, Comment and Gift Trend from May 5th to May 25th

Overall, we can see all engagement metrics are in decreasing trend as well (with also a big dip on May-21st and May-22nd)!

Gift Amount vs. Gift Amoung per user Trend

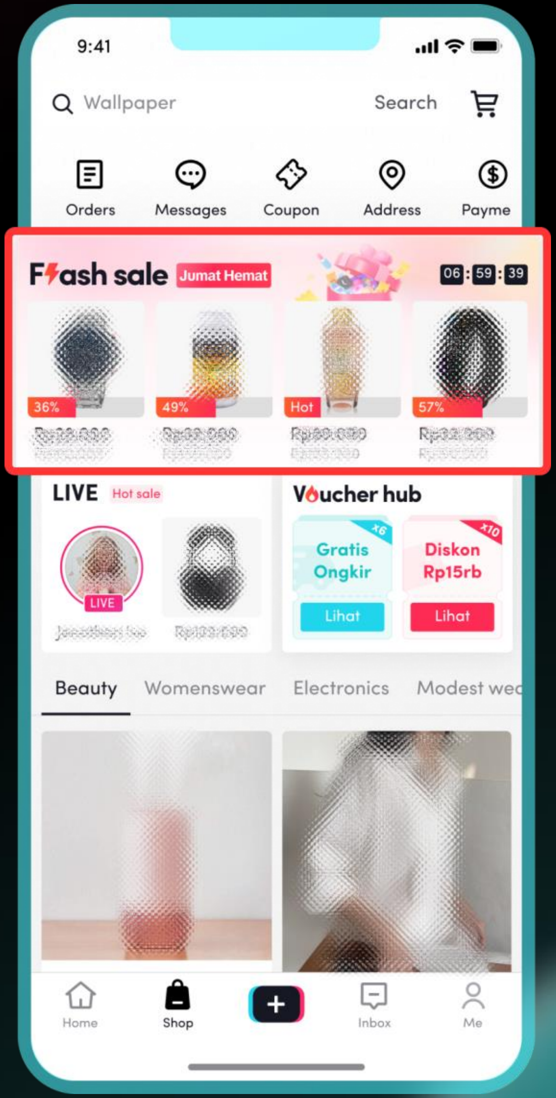
- Regarding the Gift Amount (which is 1 of Kuaishou's main revenue source), it also has a decreasing trend, averagely from 110k to 100k CNY per day!
- We also see some peak days: May 5th (Monday after the Labor holiday), May 9th (Friday, maybe Payday, weekend effect), May 18th (Sunday) and May 20th (Chinese Valentine day)

→ Next, we should investigate more and find out what categories has most Gifts, and which cat has decreasing trend so need to take care!

3.5 Trend Analysis

In general, we see that the trend is quite consistent, but in decreasing trend:

- **Skips still x2 vs. Clicks**, which raised a quite big concern. The numbers of **streamers outweight the users**, too many content for small user groups, as well as the users are aging, causing challenges for the platform to acquire new users
- Sessions, numbers of lives are consistent, with a little dip on May-21st and 22nd (after the Chinese Valentine day). We don't know, but maybe it is a break after the big festival
- While streaming activities are consistent, the active users and avg. time spent are decreasing, especially time spent from 24 minutes to 18 minutes per day per user → Need to raise an alarm
- About the revenue streams from gifting, we also see the decreasing trend. Let's look upon CTR and Monetization!



3.6 Which Categories have the highest CTR?

Content Categories	sessions_cat_cont	lives_cat_cont	avg_time_spent_cat	engagement_skip_ratio
Chat	31.8	36.1	104.9	41.2
Beauty	14.8	11.3	33.5	28.6
Other	14.7	16.2	36.7	49.6
E-Commerce	7.9	11.4	28.9	71.5
Talent	6.9	7.3	23	54.1
Lifestyle	6.4	6.1	30.1	61.6
Game	5.3	3.1	19.2	17.8
Group	3.9	0.7	48.3	33.6
Hobbies	3.3	1.6	10.2	23.2
Relationship	2.3	3.1	41.4	55.7
Education	2.2	2.6	29.6	77
Fitness	0.3	0.3	5.4	41.3
News	0.3	0.1	11.7	28.6

Content Categories	click_to_like	click_to_comment	click_to_gift	gift_amt_cat	gift_amt_per_user_cat
Chat	3.9	4.0	1.4	1,991,979	202.4
Beauty	4.0	4.1	1.9	1,781,647	373.7
Other	3.4	3.9	1.3	593,439	102.9
E-Commerce	2.4	3.7	1.0	54,705	21.9
Talent	3.7	3.7	1.6	699,894	251.9
Lifestyle	2.2	2.4	0.7	50,127	29.3
Game	6.2	8.3	3.6	242,591	89.5
Group	6.0	5.9	2.2	387,495	167.6
Hobbies	3.9	4.5	2.3	70,417	54.4
Relationship	4.3	4.9	1.5	100,065	89.9
Education	3.5	3.8	1.5	22,487	18.2
Fitness	3.2	3.3	1.1	4,536	42.4
News	2.8	2.8	0.6	967	17.6

Top Performers:

1. Chat: Dominance in All Key Metrics

- Largest Sessions, Longest avg. time spent: 104.9 minutes
- Gift amount (total): ¥1,9M (biggest) & Gift per donating user: Massive, ¥202.4
- Engagement/Skip ratio: 41.2% (medium) & Strong Conversions from click (like/comment/gift)

→ Chat is the clear powerhouse—it leads in session share, engagement, retention, and monetization, both total and per user. This should remain a focus for growth and campaign investments, especially for live gifting.

2. Beauty

- Sessions and Engagements just after Chat
- Gifts total: ¥1,781,647 (high) & Gift per user: ¥373.7. Click-to-gift conversion: 1.9 (highest after chat)

→ Beauty is also highly successful both overall and per user. Engagement/skip is moderate (28.6%), but gifting efficiency is high, making this a strong monetization category.

Promising categories:

3. Game:

- Per-user gift: ¥424 (superior to Beauty, lower total due to smaller user base)

- Best click-to-comment (8.3): suggests viewers are deeply involved, but very low engagement / skips (17.8%)
→ We have to recommend right content for Game users, because when they love it, they are hooked on it

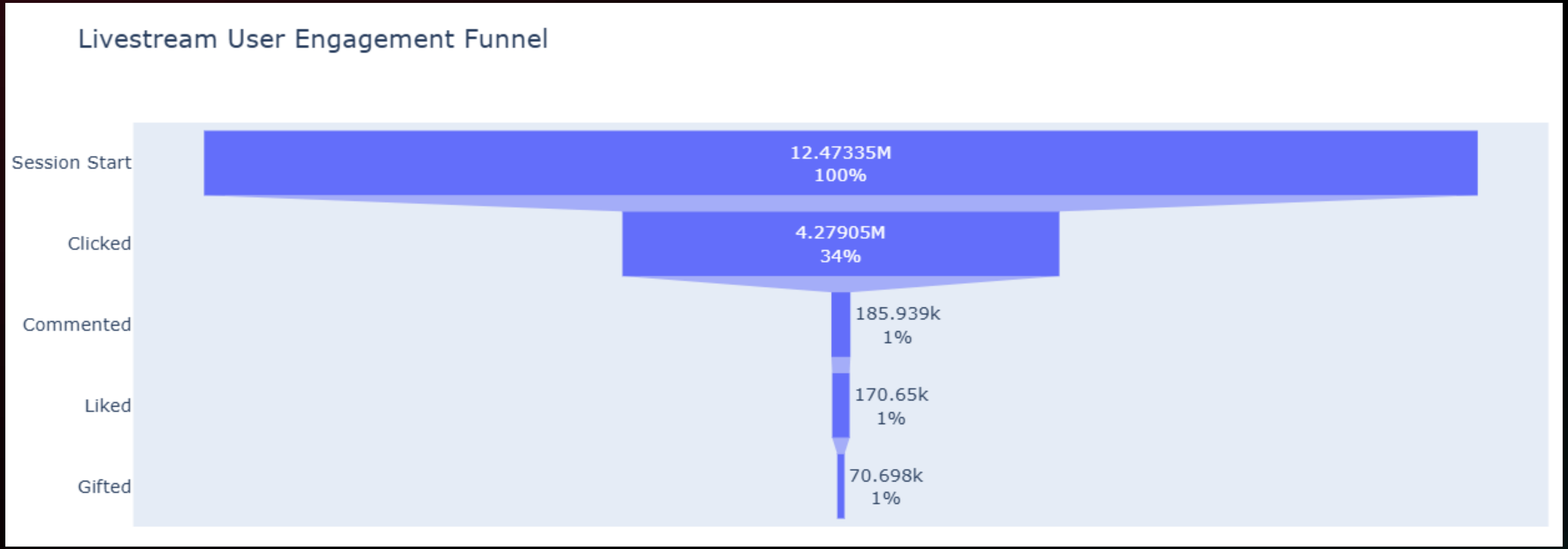
→ Comment: Game punch above their weight in per-user monetization, making them strong candidates to target for premium gift promotions or creator investment.

4. User Funnel Analysis

4.1 User Journey: Session Start -> Click -> Comment -> Like -> Gift

OK, now we have a quick visual for our assumed flow:

- 1 interesting fact is that we even have more Comments than Likes - which reflect a different user behavior vs. other platform (Comment first then like)
- We can see that Kuaishou engagement (Session-To-Click) is quite excellent (34%) vs. industry standard (20-30%). Click-to-Gift conversion 1.65% (70.6k Gifts / 4.27M Clicks) is around industry average (1-2%) for live shopping
- Areas to improve: While we have good Session-To-Click, Comments and Like is very small (around 4%).



Look at the benchmark table, I have some recommendations below:

- Boost Like/Comment Engagement: Train hosts to actively request likes and comments, use polls, Q&A sessions
- Gamification: Implement comment-to-win contests and interactive features
- Content Quality: Focus on more engaging, interactive content to increase depth metri

Metric	Your Performance	Industry Average	Top Performers	Status
Session-to-Click	34%	20-30%	30-35%	✅ Excellent
Click-to-Like	4.0%	6-12%	10-15%	⚠️ Below Average
Click-to-Comment	4.3%	5-8%	8-12%	⚠️ Below Average
Click-to-Gift	1.65%	1-2%	2-3%	✅ Good

4.2 Sankey Diagram: users have more different paths

In reality, users have many paths rather than the assumed flow above. Overall behaviors:

1. Browse-Heavy Behavior - users predominantly browse through multiple streams rather than deeply engaging with single content, suggesting:

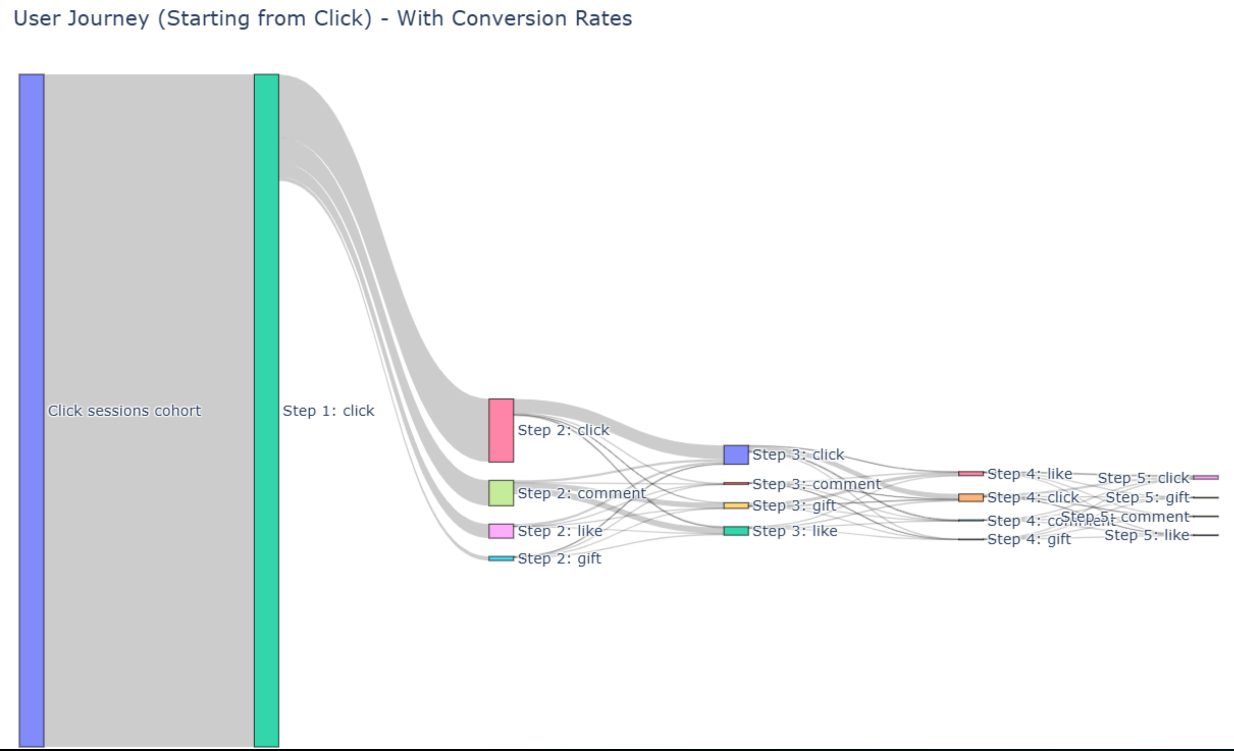
- Content discovery mode (users searching for preferred content)
- Short attention spans per stream
- Algorithm showing diverse content that doesn't immediately hook viewers

2. Low Initial Engagement Conversion

- Very thin flows from Step 1 to comments/likes/gifts indicate weak initial engagement
- Most users who click don't immediately engage - they continue browsing instead

3. Positive: Sequential Engagement

- Users who do engage (comment/like/gift) often continue engaging further (Steps 3-5)
- This shows that once hooked, users become highly engaged
- Gift users appear to be particularly loyal, continuing to multiple steps



Strategic Implications

1. Opportunities:
- **Reduce Browse-and-Leave**: Implement features to capture attention faster in first few seconds of click
- **Improve Content Matching**: Better algorithmic recommendations to reduce excessive browsing
- **Engagement Triggers**: Add interactive elements (polls, Q&A) to convert browsers to engagers sooner

2. Strengths to Leverage:

- Strong Sequential Engagement: Once users engage, they stay engaged - optimize for this conversion moment
- Gift User Loyalty: Gift users show strong continued engagement - create VIP experiences for them

4.2 Sankey Diagram: users have more different paths

Now, I am thinking about: **What actions will lead users to the Gifting**? So we can do better steps before the monetization stage!

This table above show that:
- Comment → Gift: 15.6-21.2% (your best converter!)

- Click → Gift: Only 0.43-1.12% (your weakest)

- Like → Gift: 0.07-0.10% (very low)

- Gift → Gift: 0.04-0.22% (repeat gifting is rare)

from_step	from_action	to_action	count	conversion_rate_%
2	comment	gift	34,239	21.24
4	comment	gift	1,038	16.22
3	comment	gift	1,688	15.61
4	click	gift	528	1.12
3	click	gift	930	0.77
1	click	gift	25,760	0.6
2	click	gift	1,732	0.43
4	gift	gift	6	0.22
2	gift	gift	28	0.11
2	like	gift	92	0.1
3	like	gift	38	0.07
4	like	gift	19	0.07
3	gift	gift	15	0.04

Insights:

- In whatever step 2, 3 or 4. As long as user comments, highly chances they will come to the gifting steps

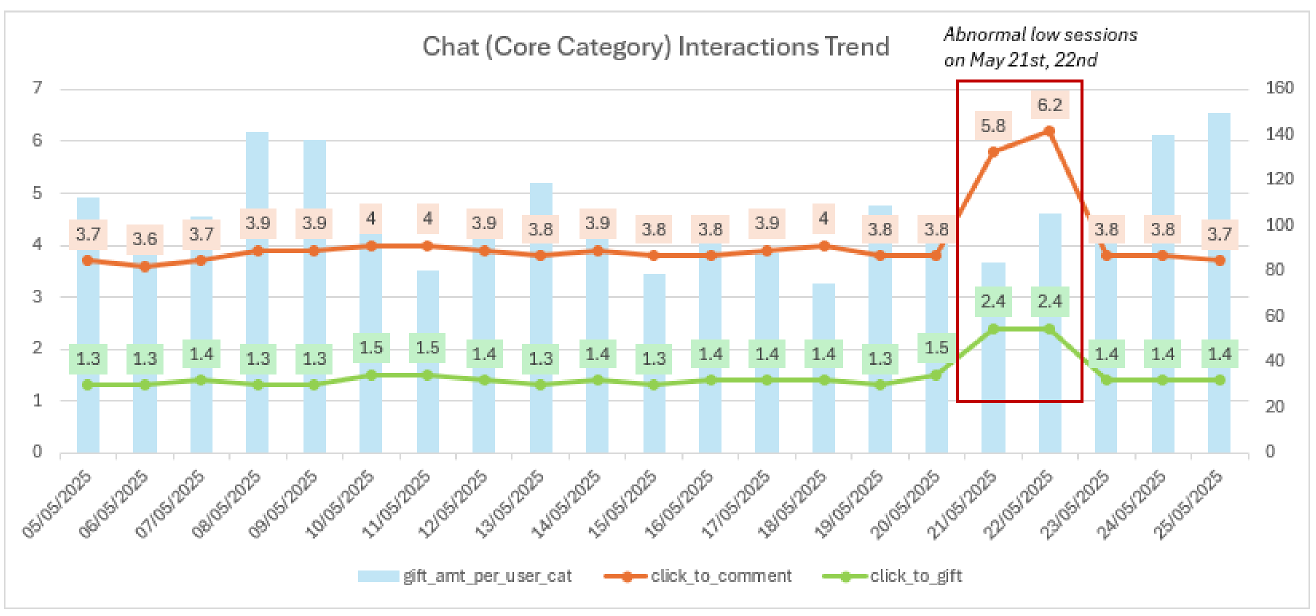
- Users who comment are 20-50x more likely to gift than users who just click or like. This suggests that active participation/interaction is the strongest predictor of monetization.

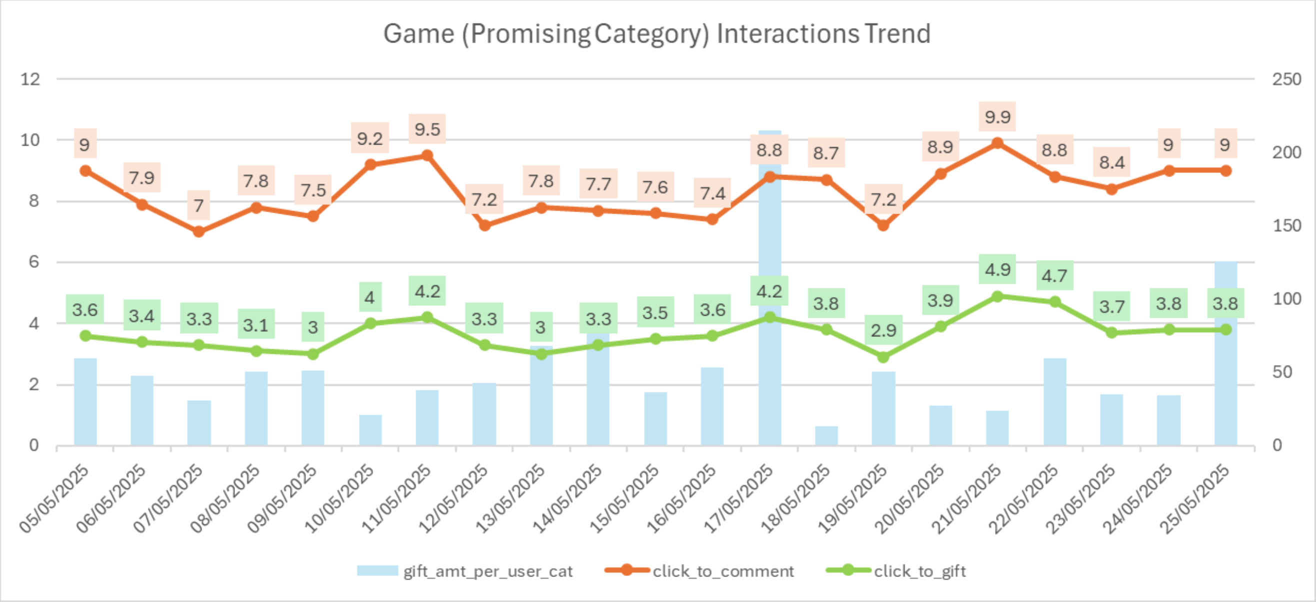
Recommendation: Prioritize Comment Generation (Immediate Impact), improving the Click-to-Comment, so we have more Gift afterwards

- Host prompts: Train streamers to actively ask questions, polls, "type 1 if you agree"
- Comment incentives: "First 10 commenters get special shoutouts"
- Interactive content: Q&A sessions, story-telling with audience participation
- Comment contests: "Best comment wins a prize"

5. CTR Target Recommendation

5.1 CTR Target Recommendation





Strategy:

- Like what we discussed before, **Optimize Click-to-comment** because Users who comment are high chance to donate gift

- We also care about Click-to-gift, as they are the most direct shortcut to monetization.

- We focus on Chat as Core Category (has highest sessions and engagements), and Game as Promising Category (has highest Gift rate!)

- We won't take May 21st, 22nd because they are exceptionally low sessions

Chat – Core Category:

- For Chat, avg Click-to-comment is very stable around 3.8 – 3.9. Beside the abnormal days, we should aim at **4.1-4.2 as target Click-to-comment**. I believe the chat activity should be more interactive in terms of chat/ 2-way conversation

- Click-to-gift is also stable around 1.4-1.5. It is very difficult for gifting, so better focus on Click-to-comment as the Top OKRs for Chat categories, thereby improving the Gift conversion rate and Gift Amount

Games – Promising Category:

- For games, avg Click-to-comment is very fluctuate and exceptionally high, even reach 9.5 on May 11st. As shown in User Journey Funnel, We should aim **9.2 Click-to-comment** as the target.

- As commented in previous analysis, once users start to click and engage, very high chance users will donate the gift. So we should invest on activities to make user gift once they click the videos. Setting **4.0 as Click-to-comment target** will be a good start!

5.2 Summarized Analysis

Users & Streamer Analysis:

- Kuaishou is the platform for rural users in China, with 60% are male users, who are interested in 60% female streamers. Looking at the age, we see the problems – 60% of users are from 30+, and the streamers outnumber users (x20 times)

→ The platform face the problem of attracting new users, especially younger ages. Users are overloaded by lots of content!

- High left-skewed interactions: 68.5% streamers receive less than 5 interactions, while only 5.2% receive more than 40+. Streamers more than 1 year tenure has 8.5 interactions, while others below 1 year only have around 0.5.

→ This is the Cold-start problem, where new streamers are not favored by the platform. Most users are attracted to streamers who have high tenure, create challenges to attract new streamers to the platform

- 37% of user have less than 50 engagements, while 13.7% have 400+ interactions

→ Challenging for recommender system to provide content to less-interactive users, and also most favorable content to super heavy users

- Though the DAU are steady, the avg time spent are decreasing, showing that users are losing interests in the content being offered!

Funnel Analysis:

- Skips (12.6M) x3 Clicks (4.8M): it is a big warning, where users skips most of the contents

- The popular journey is: Click -> Comment -> Like -> Gift

- However, users have many paths (go back to Sankey diagram). Users often click, then click as next steps → They are in content discovery mode.

- Especially, users who do engage (comment/like/gift) often continue engaging further (Steps 3-5)

→ We need solutions to keep users stay and engage rather than continuous click and swipe. Once they are engaged, higher chance they make more valuable engagement steps (like gifting)

- We also find out that **Comment-to-Gift has the highest conversion rate** overall (26%)

→ This should be focused as strategic initiative to increase Gifting conversion rate and Gift amount

Categories Analysis: We find out 2 category groups

- Core (Chat & Beauty) Categories which hold most sessions and gift amount of platform

- Promising Categories (Game) which have high CTR, especially the Click-to-gift and Gift amount per user

5.2 Summarized Analysis

Strategic Recommendations:

1. Focus on New User Acquisition:

As streamers outweigh – x20 users, we need to focus on new user recruitment, especially the younger age and in urban areas, who have higher power of purchase

2. Coldstart program for New Streamers:

As we are facing the problems that streamers with 1 year tenure attract most interactions, we should have education program to train new streamers on creating hooked content, and also budget policy to sponsor promising new rising streamers

3. Keep users stay with the content and engage:

As Skips x3 Clicks, and Users Journey as content discovery (Click-to-Click), we need solutions to keep users stay and engage rather than continuous click and swipe. Once they are engaged, higher chance they make more valuable engagement steps (like gifting). Improving the recommender system for suggest appropriate content, as well as education program for streamers are some good approaches

4. Strategies for Core and Promising categories to improve CTR, especially Gifting conversion:

- As we analysed in the user funnel, Comment-to-gift has the highest conversion rate! Therefore, we should focus on improving Click-to-Comment as a base to improve the Gifting conversion rate!

- For core categories like Chat and Beauty, especially focus on improving the Click-to-Comment as the top OKRs

- For promising categories like Games, once users are engaged, higher chances they will donate, gifting. So we need to educate more streamers in game content, increase the sessions contribution and also Gifting metrics (Click-to-gift and Gift amount per user)

Next Steps:

5. Analyse User Retention and Recommendation to improve

6. CTR predictive modeling & Content Recommendation system for users

6. Appendix

Data Description

The file streamer.csv contains comprehensive information about all streamers, including streamer_id and associated side information, such as age, country, and a set of binary features.

Field Name	Description	Type	Example
streamer_id	The ID of the streamer.	int64	56006
gender	The gender of the streamer.	string	M
age	The age range of the streamer.	string (range)	24-30
country	The country where the streamer resides.	string	China
device_brand	The brand of the device used by the streamer.	string	APPLE
device_price	The price range of the device.	string (range)	5000+
live_operation_tag	The operational category of the streamer.	string	Relationship
fans_user_num	The number of users who have followed the streamer.	string (range)	10000-100000
fans_group_fans_num	The number of fans from the streamer's fans group.	string (range)	0-10
follow_user_num	The number of users followed by the streamer.	string (range)	1000-10000
first_live_timestamp	The date when the streamer started their first live room.	timestamp	04/02/2018
accu_live_cnt	The total number of live rooms the streamer has hosted.	string (range)	100-500
accu_live_duration	The cumulative duration of all live rooms, in milliseconds.	string(range)	500000000-1000000000
accu_play_cnt	The total number of times the streamer's live rooms have been viewed.	string (range)	100000-500000
accu_play_duration	The cumulative duration of all live rooms, in milliseconds.	string (range)	50000000000-100000000000
reg_timestamp	The date when the streamer registered on the platform.	timestamp	03/07/2014

The file room.csv contains detailed information about all live rooms, including basic fields such as the date, room id, the corresponding streamer, as well as temporal data and categorical descriptors.

Field Name	**Description**	**Type**	**Example**
p_date	The date of the live room.	int64	20250525
live_id	The id of the live room.	int64	7336601
streamer_id	The id of the streamer.	int64	252634
live_type	The type the live room, represented as a categorical code.	int64	1
start_timestamp	The start time of the live room.	int64	1748131180878
end_timestamp	The end time of the live room.	int64	1748275200000
live_content_category	The content category of the live room.	string	shop
live_name_id	The id associated with the live room title, used to index the encoded title embeddings.	int64	0

Data Description

The file user.csv contains comprehensive information about all users, including the user_id and side information such as age, country, and a set of binary features

Field Name	Description	Type	Example
user_id	The ID of the user.	int64	22733
age	The age range of the user.	string (range)	18-23
gender	The gender of the user.	string	M
country	The country where the user resides.	string	China
device_brand	The brand of the device used by the user.	string	DESKTOP
device_price	The price range of the device.	string (range)	0
reg_timestamp	The date when the user registered on the platform.	timestamp	2023-05-03
fans_num	The number of fans who have followed the user.	string (range)	0-10
follow_num	The number of users followed by the user.	string (range)	10-100
first_watch_live_timestamp	The date when the user first watched a live room.	timestamp	2023-05-03
accu_watch_live_cnt	The total number of live rooms the user has watched.	string (range)	0-50000
accu_watch_live_duration	The cumulative duration of all live rooms the user has watched, in milliseconds.	string (range)	0-1000000000
is_live_streamer	A binary indicator showing whether the user is a live streamer (1 = yes, 0 = no).	int64	0
is_photo_author	A binary indicator showing whether the user is a photo content author (1 = yes, 0 = no).	int64	0

The file comment.csv records user interaction in the form of comments. Each record corresponds to a single commenting event and includes the associated user, live room, streamer, and timestamp.

Field Name	Description	Type	Example
user_id	The id of the user.	int64	23154
live_id	The id of the live room.	int64	7865151
streamer_id	The id of the streamer.	int64	433825
timestamp	The timestamp when this interaction occurred.	int64	1746374400819

The file gift.csv records user interactions in the form of sending gifts. Each record corresponds to a single gifting event and includes the associated user, live room, streamer, timestamp, and the price of the gift.

Field Name	Description	Type	Example
user_id	The id of the user.	int64	11504
live_id	The id of the live room.	int64	6086847
streamer_id	The id of the streamer.	int64	114419
timestamp	The timestamp when this interaction occurred.	int64	1746374441260
gift_price	The total price of gifts sent during this interaction.	int64	2

Data Description

The file like.csv records user interaction in the form of liking. Each record corresponds to a single liking event and includes the associated user, live room, streamer, and timestamp.

Field Name	Description	Type	Example
user_id	The id of the user.	int64	5222
live_id	The id of the live room.	int64	541927
streamer_id	The id of the streamer.	int64	244121
timestamp	The timestamp when this interaction occurred.	int64	1746374414059

The file negative.csv contains records of all exposures that were presented to users but not clicked. Each record includes the corresponding user, live room, streamer, and the timestamp of the exposure.

Field Name	Description	Type	Example
user_id	The id of the user.	int64	9810
live_id	The id of the live room.	int64	10816308
streamer_id	The id of the streamer.	int64	17452
timestamp	The timestamp when this interaction occurred.	int64	1746525926498

The file click.csv contains records of user interactions in the form of click-to-watch behavior. Each record corresponds to a single click-to-watching event and includes the user id, the associated live room and streamer, the timestamp of the interaction, and the watch time.

Field Name	Description	Type	Example
user_id	The id of the user.	int64	8505
live_id	The id of the live room.	int64	9342705
streamer_id	The id of the streamer.	int64	392199
timestamp	The timestamp when this interaction occurred.	int64	1746374400022
watch_live_time	The user’s watch time for this interaction, in milliseconds.	int64	2852

Thank You!

Created by Daniel Nguyen

Email: ***hoangfotu@gmail.com***
Linkedin: *linkedin.com/in/hoangfotu*