

SENTIMENT ANALYSIS REPORT

Prepared by Nguyen Hoan

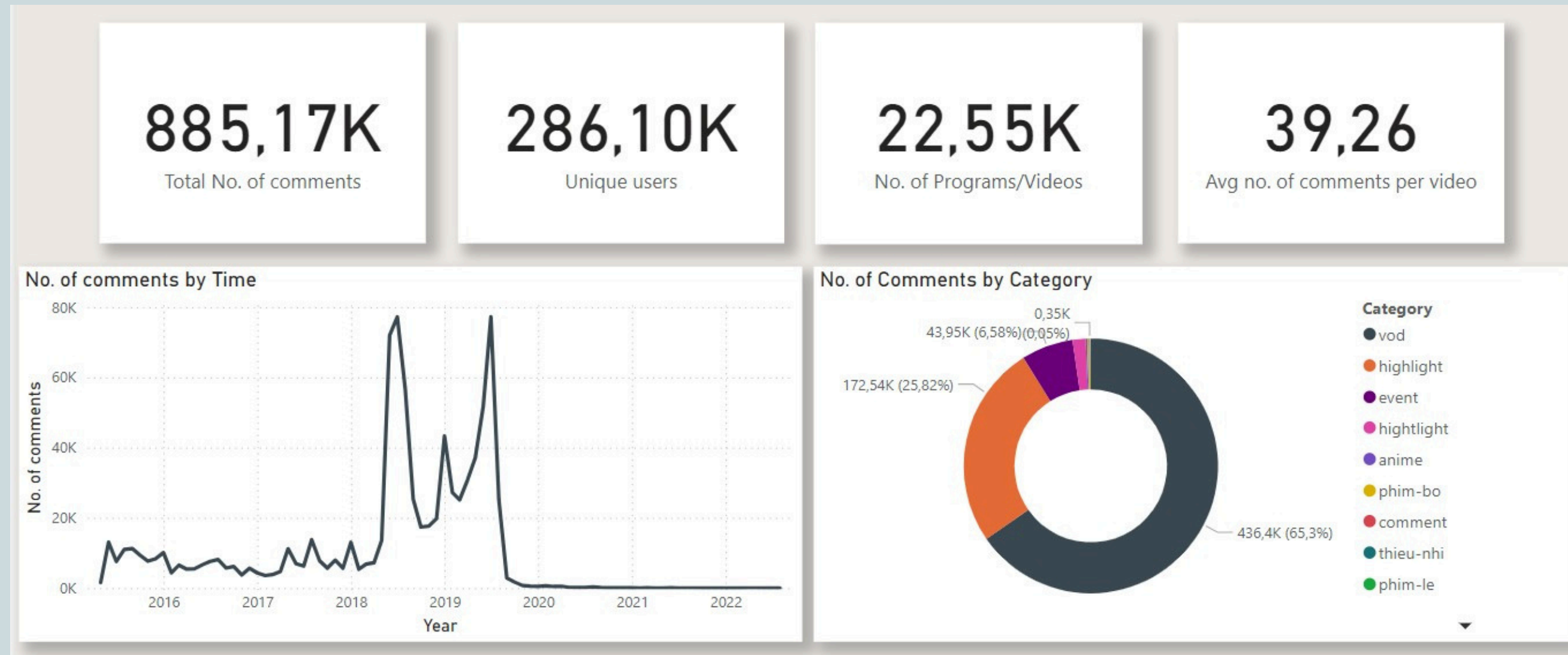


DATA OVERVIEW

Data of customer comments on FPTPlay's contents recorded between 2015 - 2022

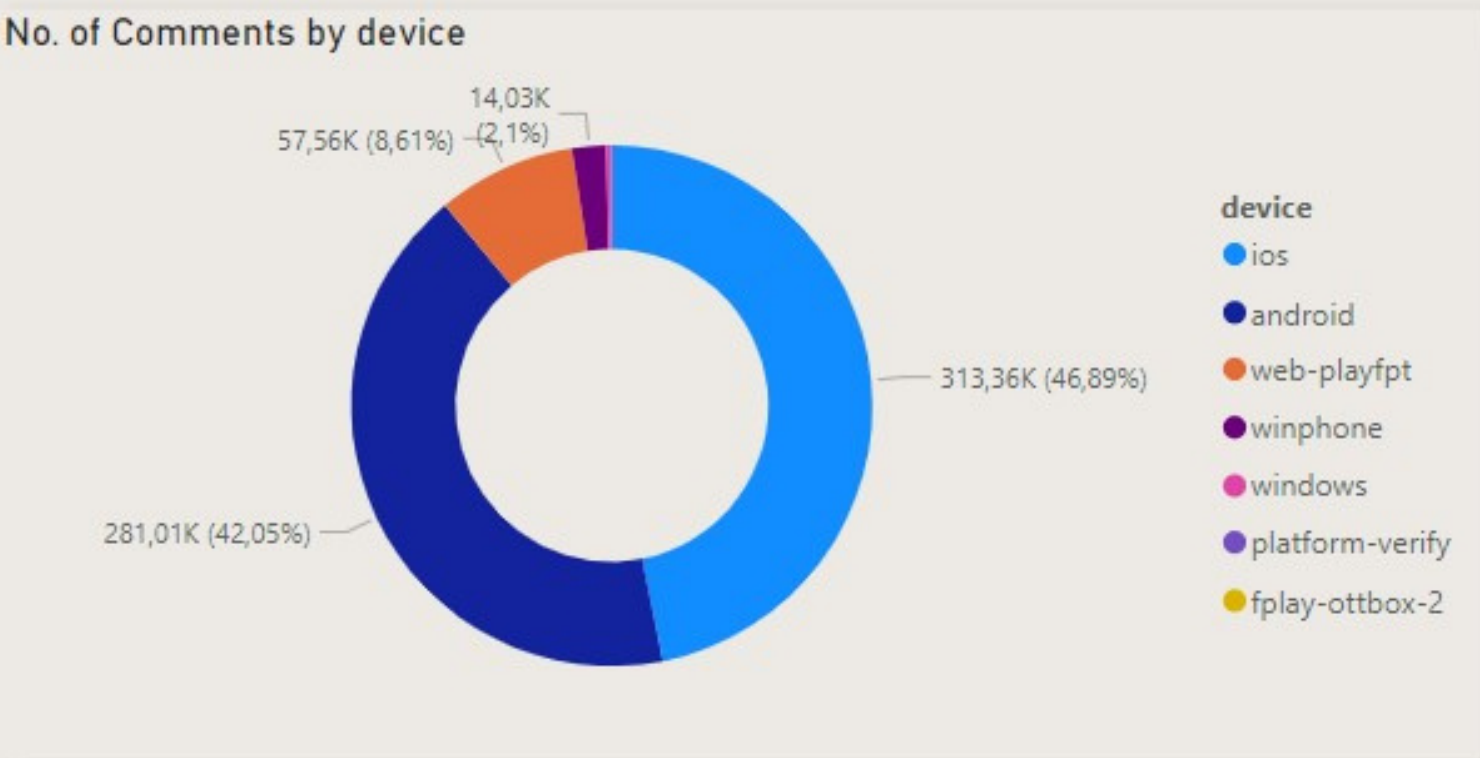
Column name	Explanation	Example
_id	Unique IDs of comments	“555612f21a77ce5d65ea0688”
user_id	Users’ IDs on the system	“4191084”
comment_on	The type of content that were commented	“vod”, “highlight”, etc.
object_id	Unique IDs of videos	“55237ea917dc1332b64c55a7”
content	what users commented	“hay quá đi thôi”
device	Type of user’s devices	“android”, “ios”, etc.
timestamp	Time that comment was posted	1431595030

DATA OVERVIEW

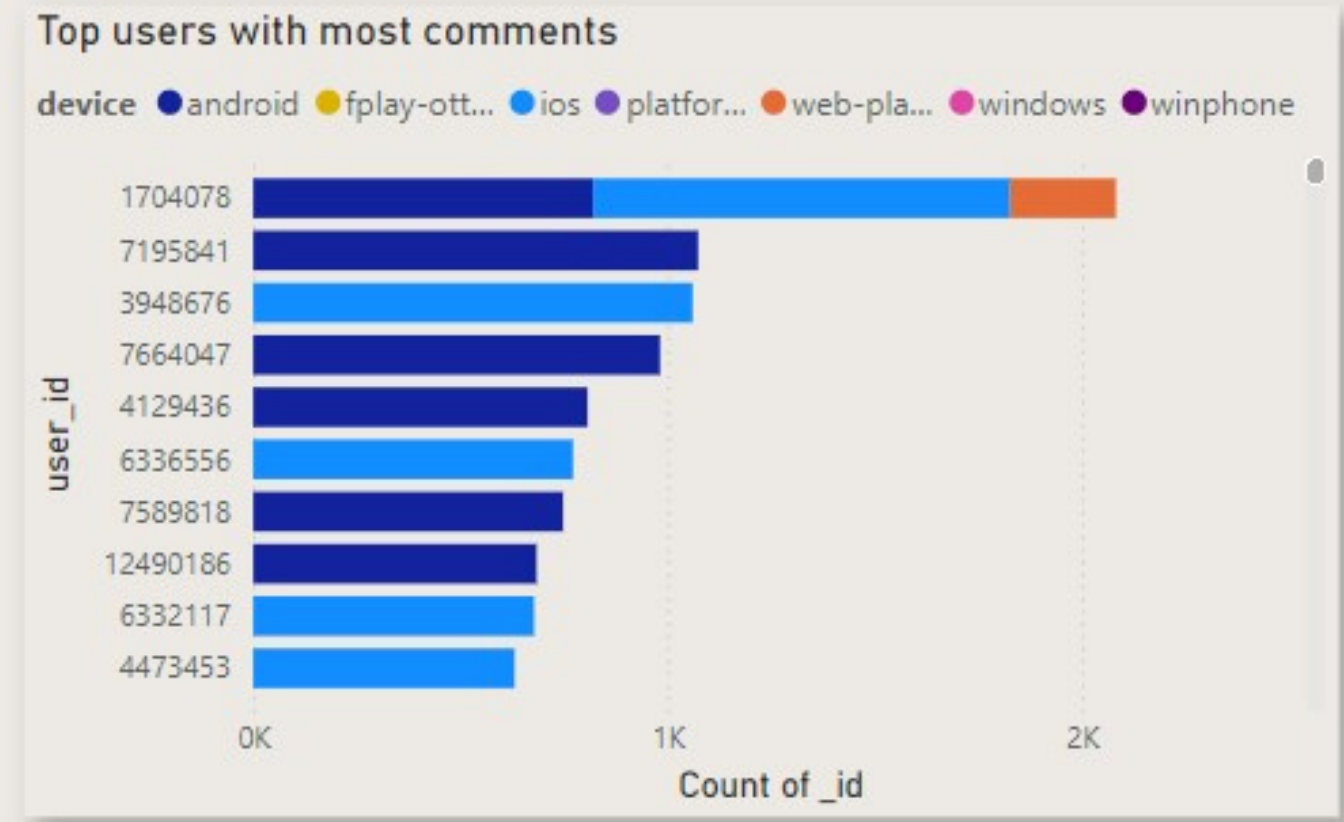
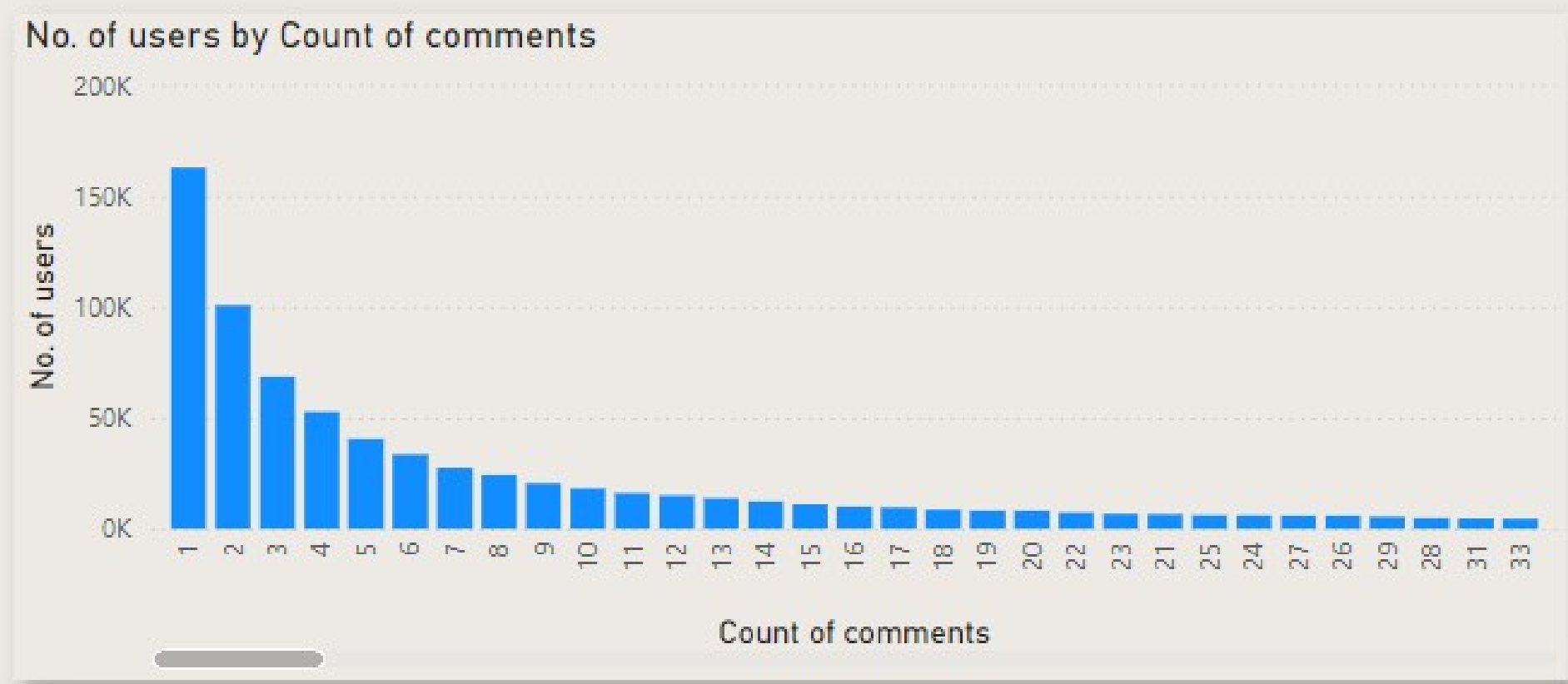


- The total number of comments saw a sharp increase in June and July 2018, during the World Cup. Most of the comments during this period were posted on World Cup highlight videos.
- A large amount of the comments were posted on movies (VOD) and sports highlight videos (65,3% and 25,8% respectively)

USER ANALYSIS



- A majority of users utilize either iOS devices, or Android devices, including phones and Smart TVs.
- On average, each user comments 3 times, however, the distribution is highly positively skewed, with the majority of users commenting only once or twice.

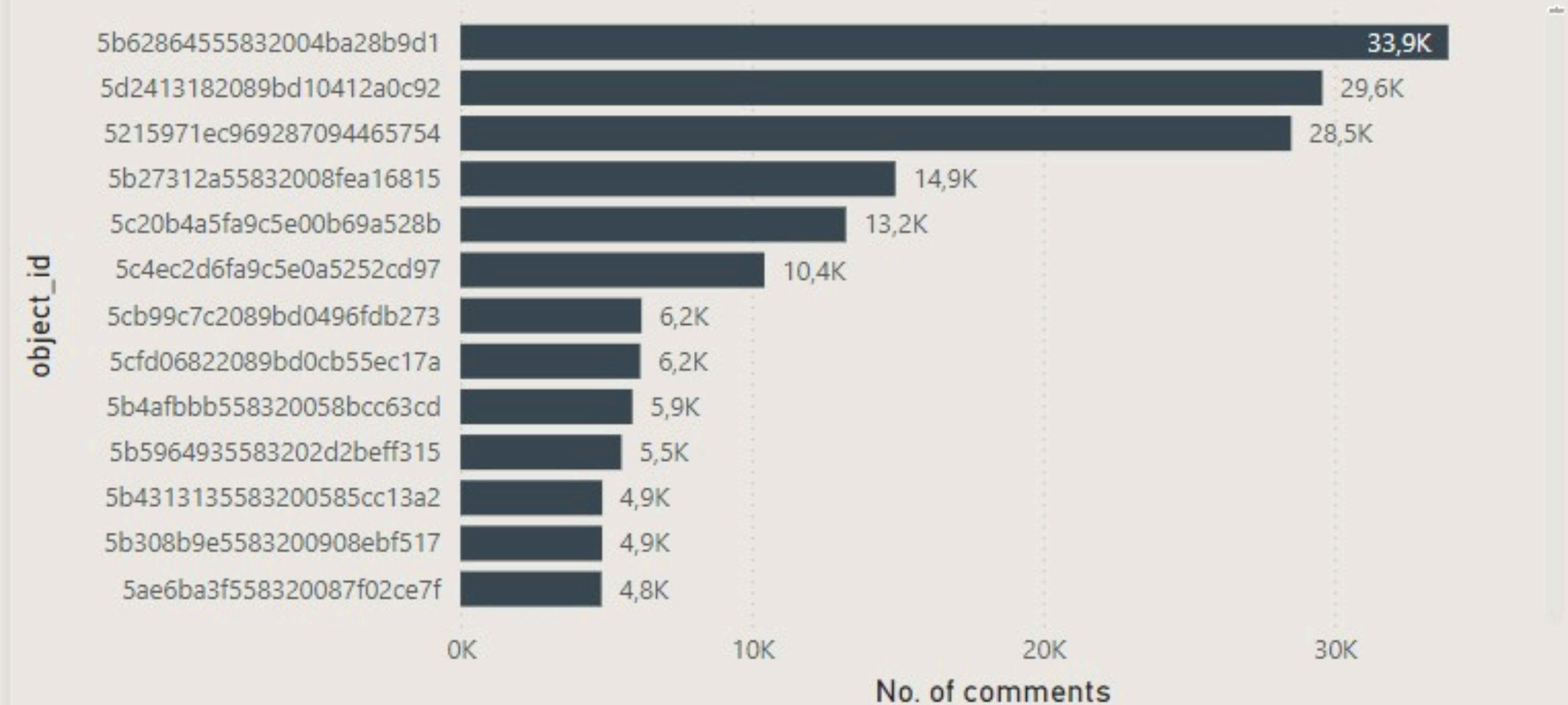


CONTENT ANALYSIS

- 9 out of the top 10 most-commented videos are vod (video on demand), which are **movie-related content**.
- Similarly, the **"vod"** tag has the **highest total number of comments**.
- In addition, the **"event"** tag also has a very high number of comments per video, indicating a **high level of interaction for events**.
- The reason why the remaining tags have a low number of comments is that **many comments do not have any tags**, while some other comments on the **same video have different tags**.

Tag	No. of Comments	Comments per video
vod	436402	30,42
	216878	39,87
highlight	172543	32,07
event	43945	36,29
hightlight	11012	9,43
anime	1437	2,70
phim-bo	1210	2,42
comment	576	1,88
thieu-nhi	352	2,32
phim-le	270	1,29
TV show	171	1,49
the thao	80	1,40
Total	885169	39,26

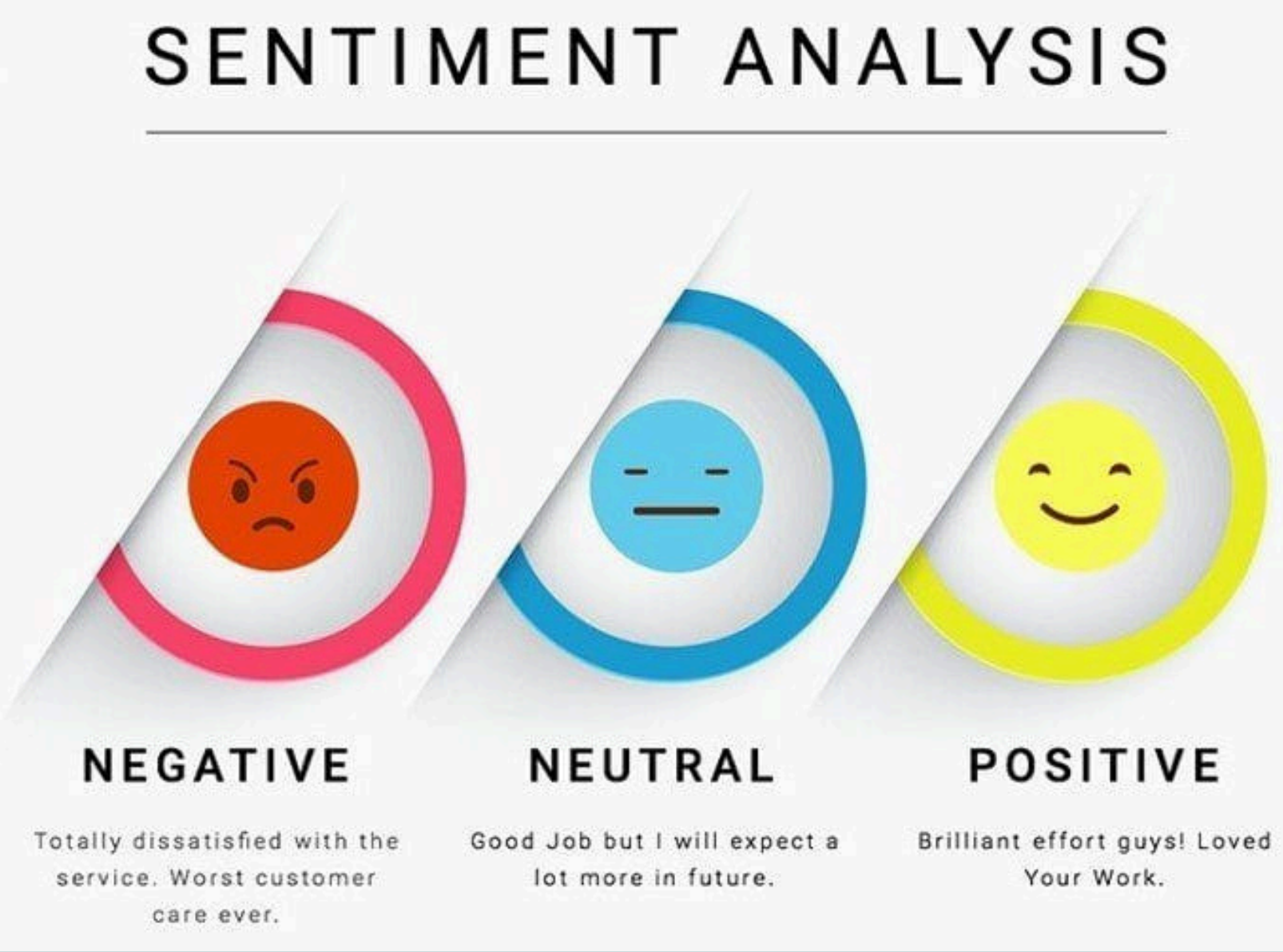
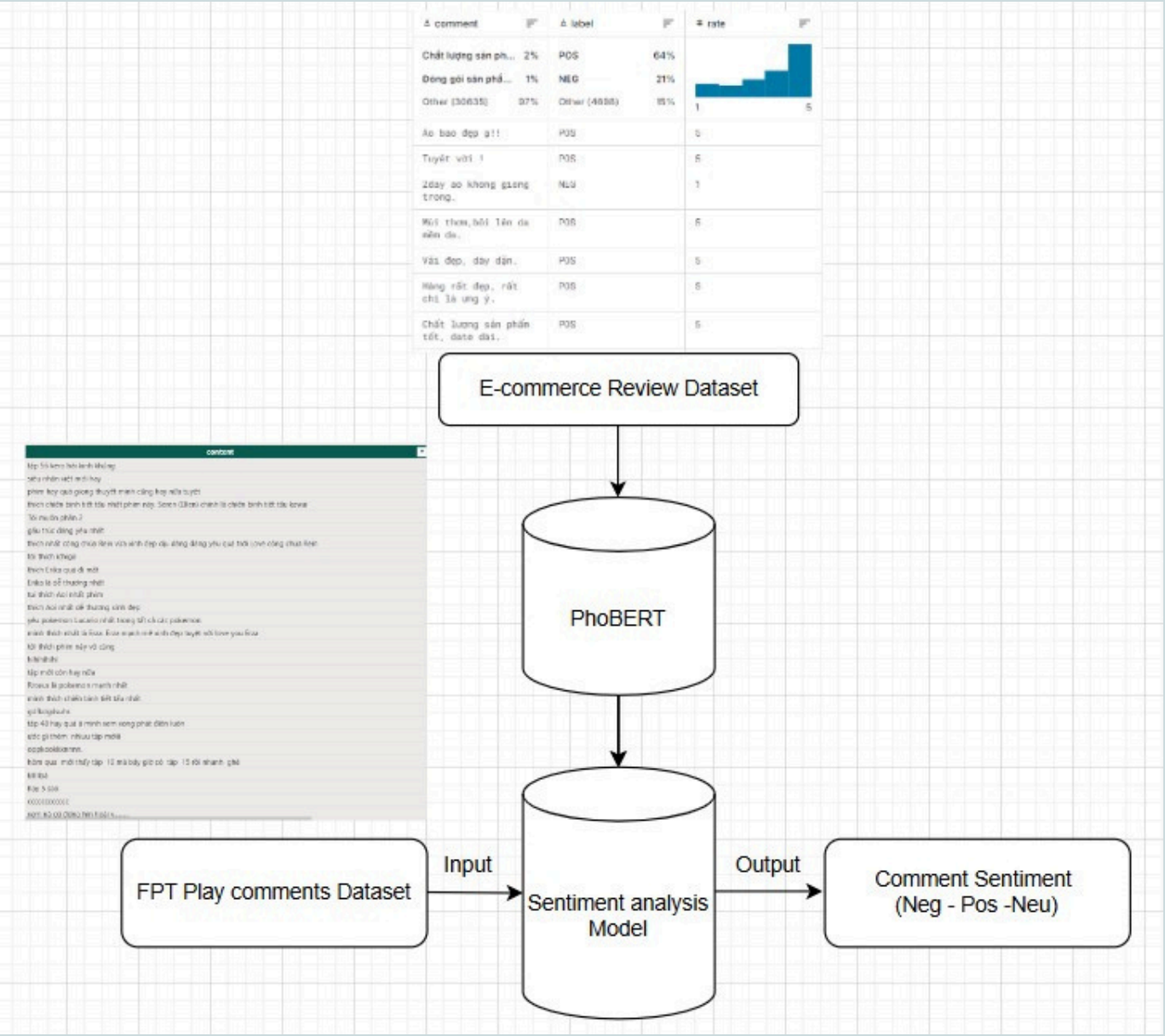
No. of comments by object_id



SENTIMENT ANALYSIS

Sentiment analysis is the process of processing textual data to extract human emotions embedded within the data. Sentiments can be expressed as **positive/neutral/negative**.

MODEL USED



PhoBERT is a Vietnamese language model developed by VinAI, used for various purposes such as part-of-speech tagging, sentence completion, etc

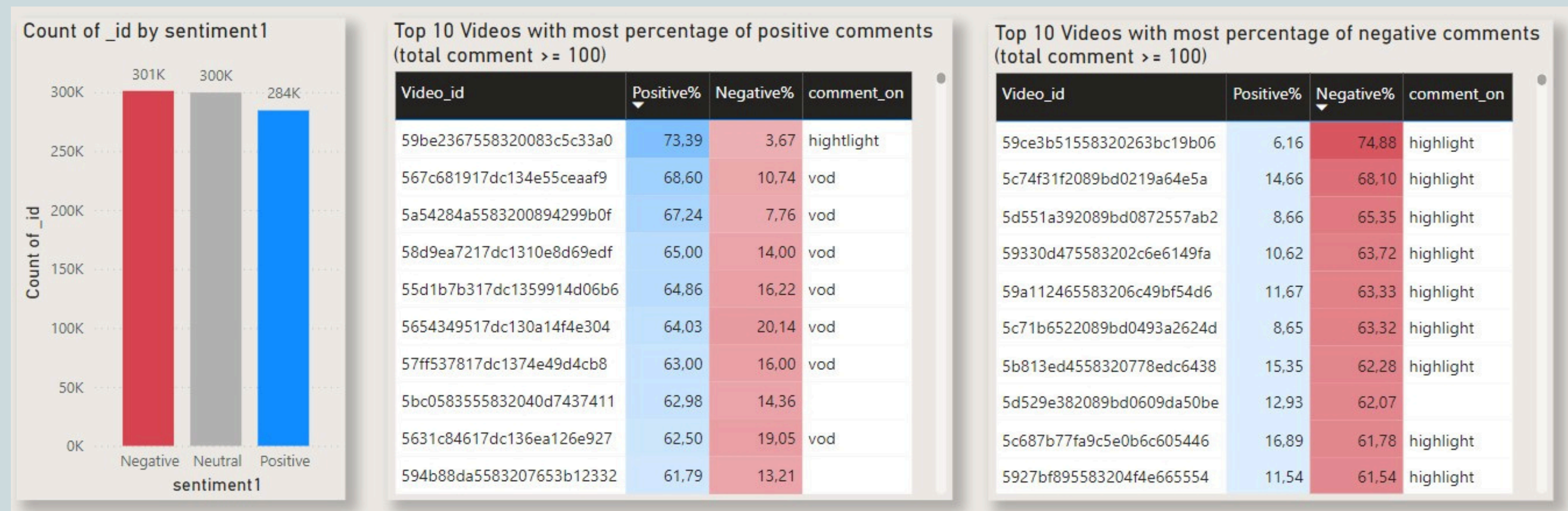
To perform sentiment analysis, we will utilize a pre-trained model, fine-tuned on the **PhoBERT model** and a dataset of 30,000 Vietnamese product reviews from an e-commerce platform

RESULT

THE OUTPUT IS THAT EACH COMMENT IS EVALUATED BY THE MODEL AND ASSIGNED ONE OF THREE SENTIMENT LABELS:
NEGATIVE (0), POSITIVE (1), OR NEUTRAL (2).

content_std	sentiment
tập 56 kero hài kinh khủng	1
siêu nhân việt mới hay	1
phim hay quá giọng thuyết minh cũng hay nữa tuyệt	1
thích chiến binh tiết tấu nhất phim này seren (ellen) chính là chiến binh tiết tấu kawai	1
tôi muốn phần 2	1
gấu trúc đáng yêu nhất	1
thích nhất công chúa rein vừa xinh đẹp dịu dàng đáng yêu quá trời love công chúa rein	1
tôi thích ichigo	1
thích erika quá đi mất	1
erika là dễ thương nhất	1
tui thích aoi nhất phim	1
thích aoi nhất dễ thương xinh đẹp	1
yêu pokemon lucario nhất trong tất cả các pokemon	1
mình thích nhất là erza erza mạnh mẽ xinh đẹp tuyệt vời love you erza	1
tôi thích phim này vô cùng	1
hihihihihi	1
tập mới còn hay nữa	1
rrceus là pokemon mạnh nhất	1
mình thích chiến binh tiết tấu nhất	1

SENTIMENT BY VIDEOS



- The total number of comments with positive, negative, and neutral sentiments is **fairly even**.
- Top 10 video with the highest percentage of **positive** comments, including tags like **idols tv-show**, **anime** and **love story film**.
- 9/10 of the top videos with the highest percentage of **negative** comments are about **sports highlights**

Top 10 Tags with most percentage of positive comments (total comment > 10)

Tag	Positive%	Negative%	Comments per video
fadio	47,37	26,32	1,46
phim-le	47,04	27,41	1,29
livetv	46,43	25,00	1,08
tvshow	45,61	24,56	1,36
the-thao	45,00	36,25	1,40
tv-show	41,86	25,58	1,59
thieu-nhi	41,19	27,27	2,32
anime	36,95	28,18	2,70
comment	36,28	30,38	1,88
phim-bo	36,03	29,09	2,42

Top 10 Tags with most percentage of negative comments (total comment > 10)

Tag	Negative%	Positive%	Comments per video
hightlight	44,28	22,60	9,43
event	41,75	25,25	36,29
ngoai-hang-anh	40,91	13,64	1,16
highlight	39,99	25,18	32,07
the-thao	36,25	45,00	1,40
vod	32,08	35,07	30,42
fifa-u20-world-cup	31,25	18,75	1,23
	31,15	33,54	39,87
comment	30,38	36,28	1,88
phim-bo	29,09	36,03	2,42

SENTIMENT BY TAGS

- Top tags with the **highest percentage of positive** comments generally have a **low number of comments per video**, while top tags with a **high percentage of negative comments** tend to have a **relatively high number of comments per video**.

=> Highly niche content is often enjoyed more by customers than mass-market content

- Many tags in the negative top are sports and football tags.

=> It's easy to understand why sports generate a lot of emotions. The Vietnamese sports market is quite pessimistic, sometimes posts comment that launches personal attacks against players/coaches.