

Sentiment

Analysis Report



Prepared by Huong Nguyen

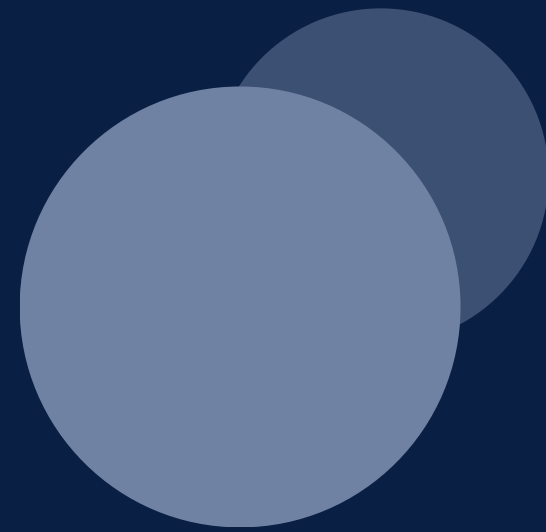


INTRODUCTION



The data on customer comments on videos from the digital viewing platform was recorded over the period from 2015 to 2022.

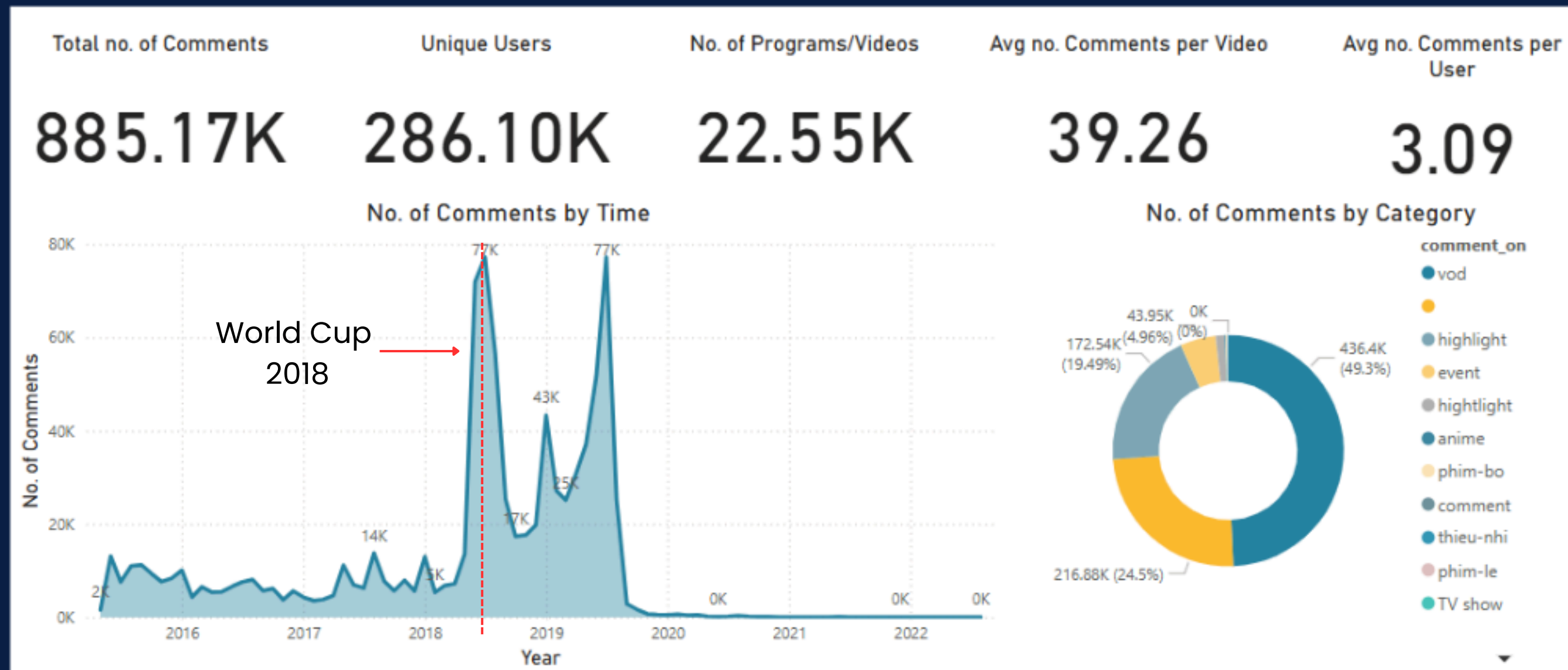
Column Name	Meaning	Example
user_id	Unique customer code	"4191084"
comment_on	The genre of the commented video	"vod", "highlight"
object_id	Unique video code	"55237ea917dc1332b64c55a7"
content	The content of the comment	"hay quá đi thôi"
device	The type of device	"android", "ios"
timestamp	The time (in Unix timestamp format) when the comment was posted	1431595030



GENERAL REPORT



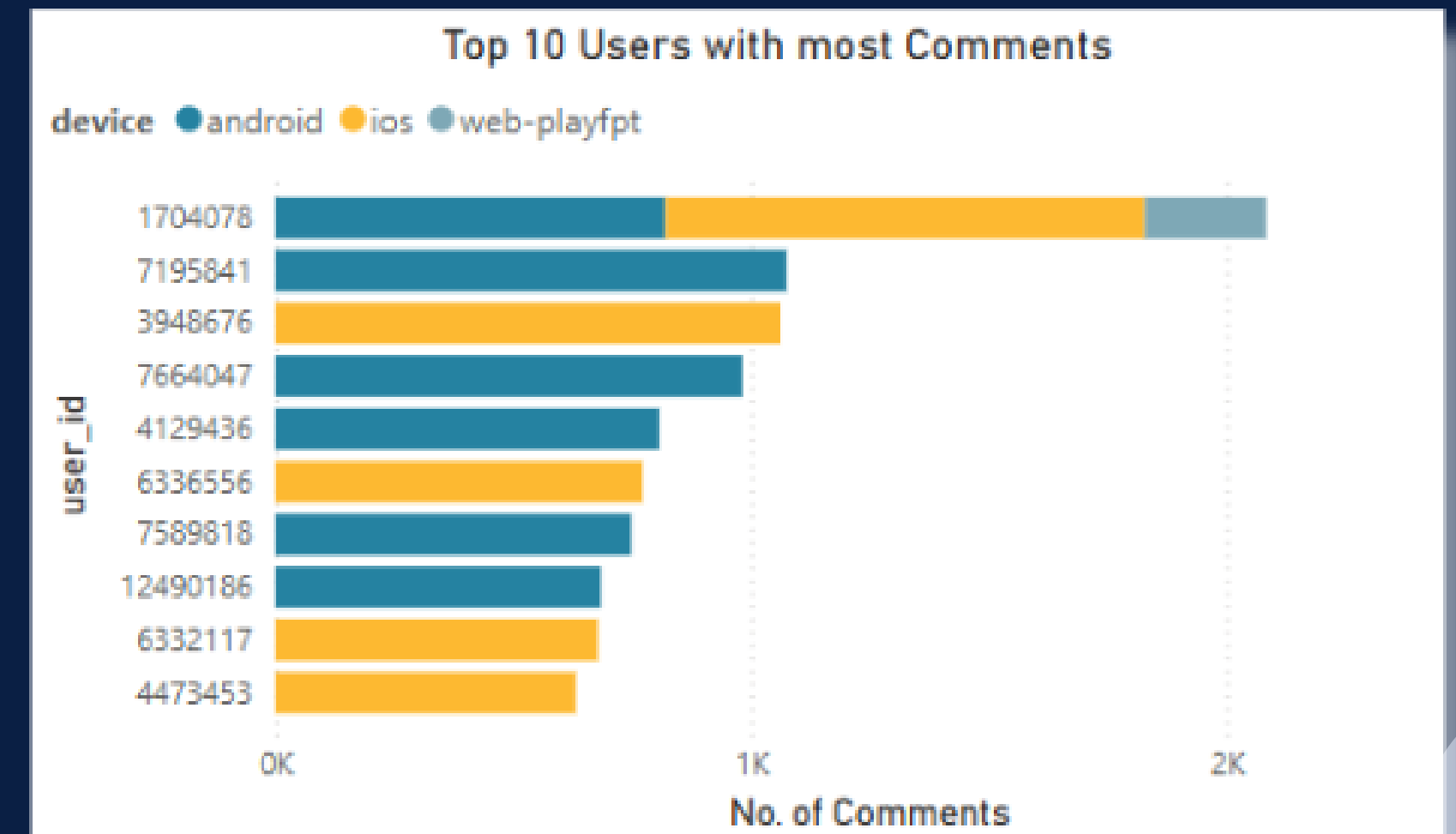
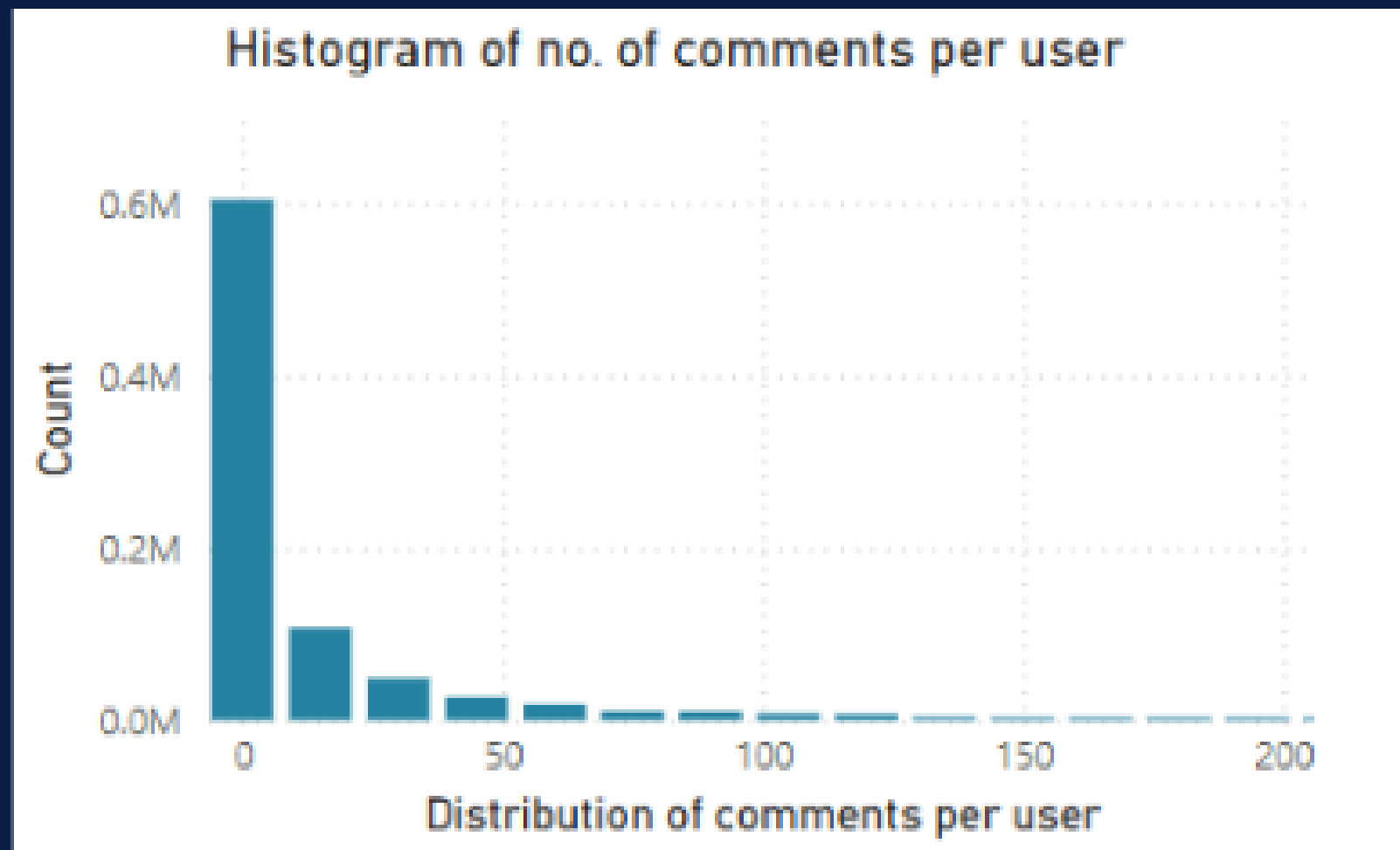
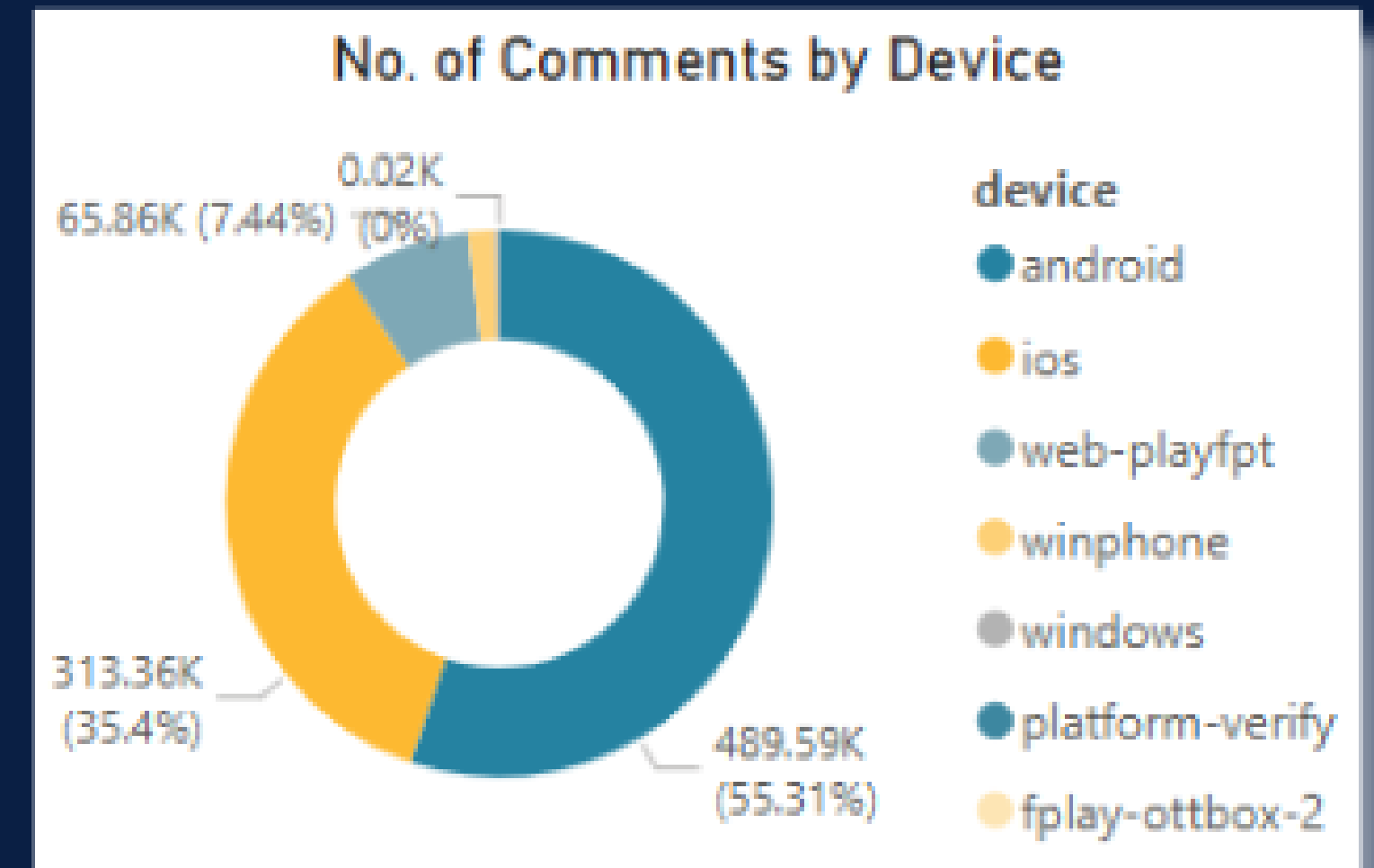
DATA OVERVIEW



- The total number of comments experienced a **significant spike** during **June–July 2018**, coinciding with the **World Cup** period. Most of the comments during this time were posted on World Cup highlight videos.
- The majority of comments were posted on **movies (VOD)** and **sports highlight** videos.

USERS ANALYSIS

- On average, each user posted 3 comments; however, the distribution is **highly positively skewed**, with the majority of users commenting only once.
- Most users accessed the platform via **Android** devices (including smartphones and Smart TVs) or **iOS** devices.



NO. OF COMMENTS BY VIDEO AND TAG

- 9/10 videos with the highest number of comments are **movies (VOD)**.
- Similarly, the **“vod”** tag has the highest total number of comments. Additionally, the **“event”** tag also shows a high number of comments per video → **strong user interaction during events**.
- The reason for the lower comment volume associated with other tags is that many comments are not tagged, while some others on the same video are tagged differently.

Top 10 tags with most comments

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
Total	884625	39.44

Top 10 videos with most comments

Video ID	Tag	No. of Comments
5b62864555832004ba28b9d1	vod	33909
5d2413182089bd10412a0c92		29587
5215971ec969287094465754	vod	28517
5b27312a55832008fea16815	vod	14932
5c20b4a5fa9c5e00b69a528b	vod	13233
5c4ec2d6fa9c5e0a5252cd97		10435
5cb99c7c2089bd0496fdb273		6213
5cfd06822089bd0cb55ec17a		6174
5b4afbbb558320058bcc63cd	highlight	5904
5b5964935583202d2beff315	vod	5528
Total		154432

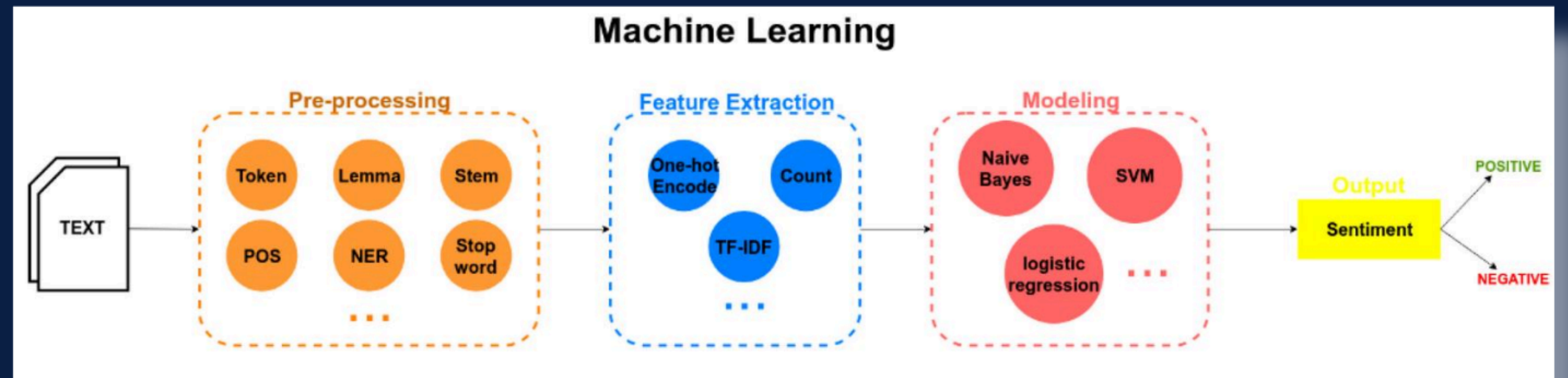
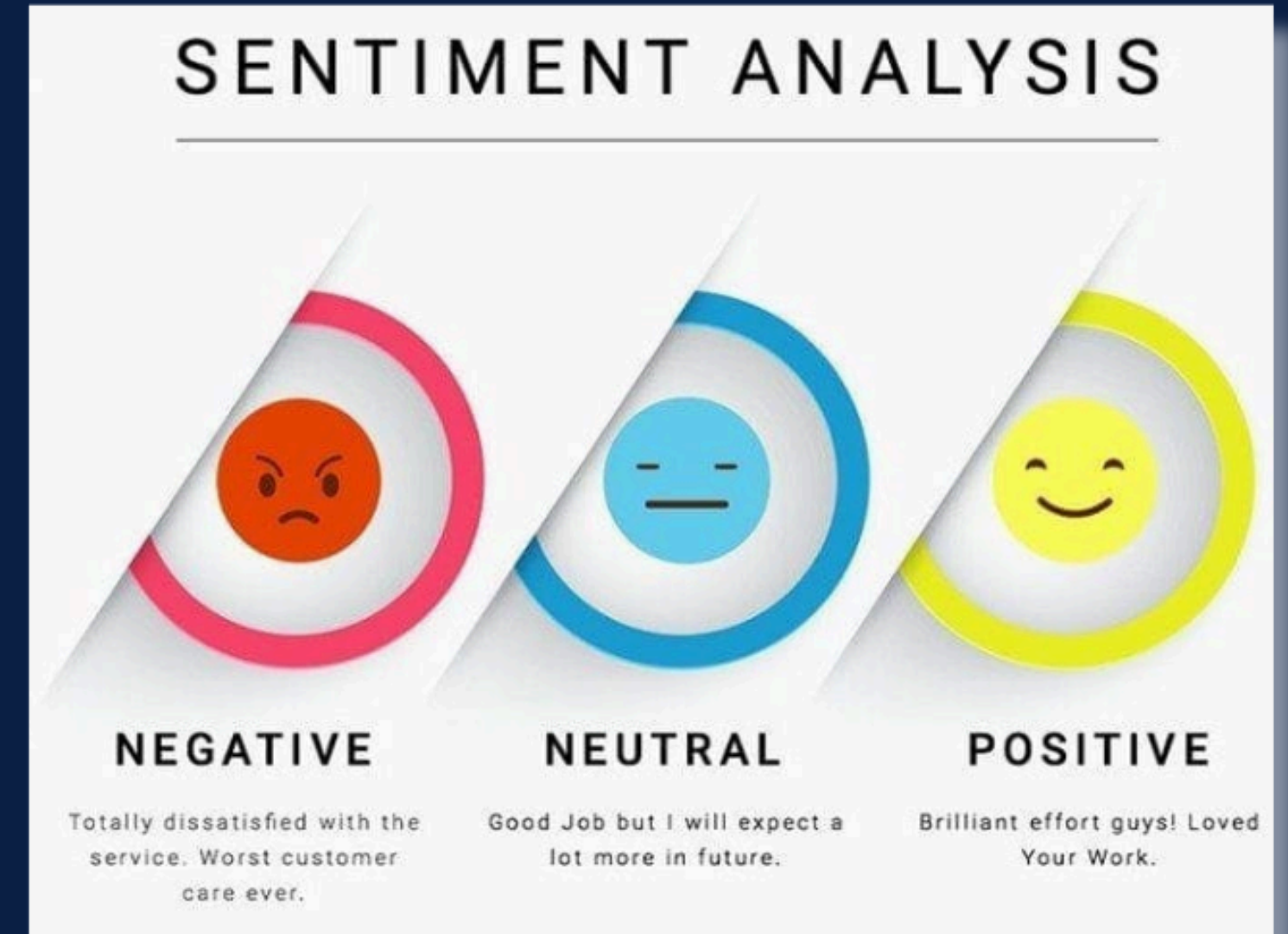


SENTIMENT ANALYSIS BASED ON COMMENT CONTENT



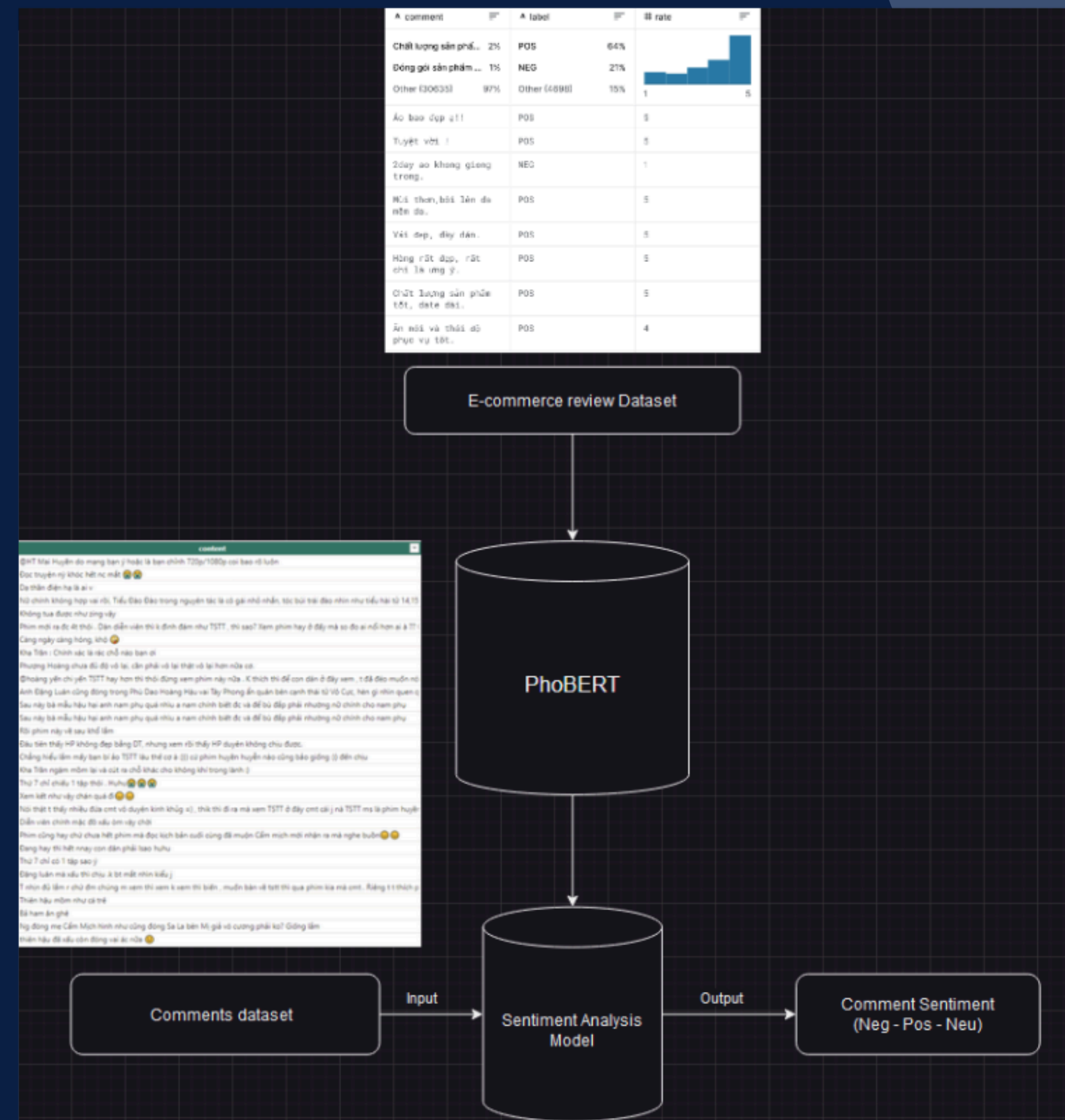
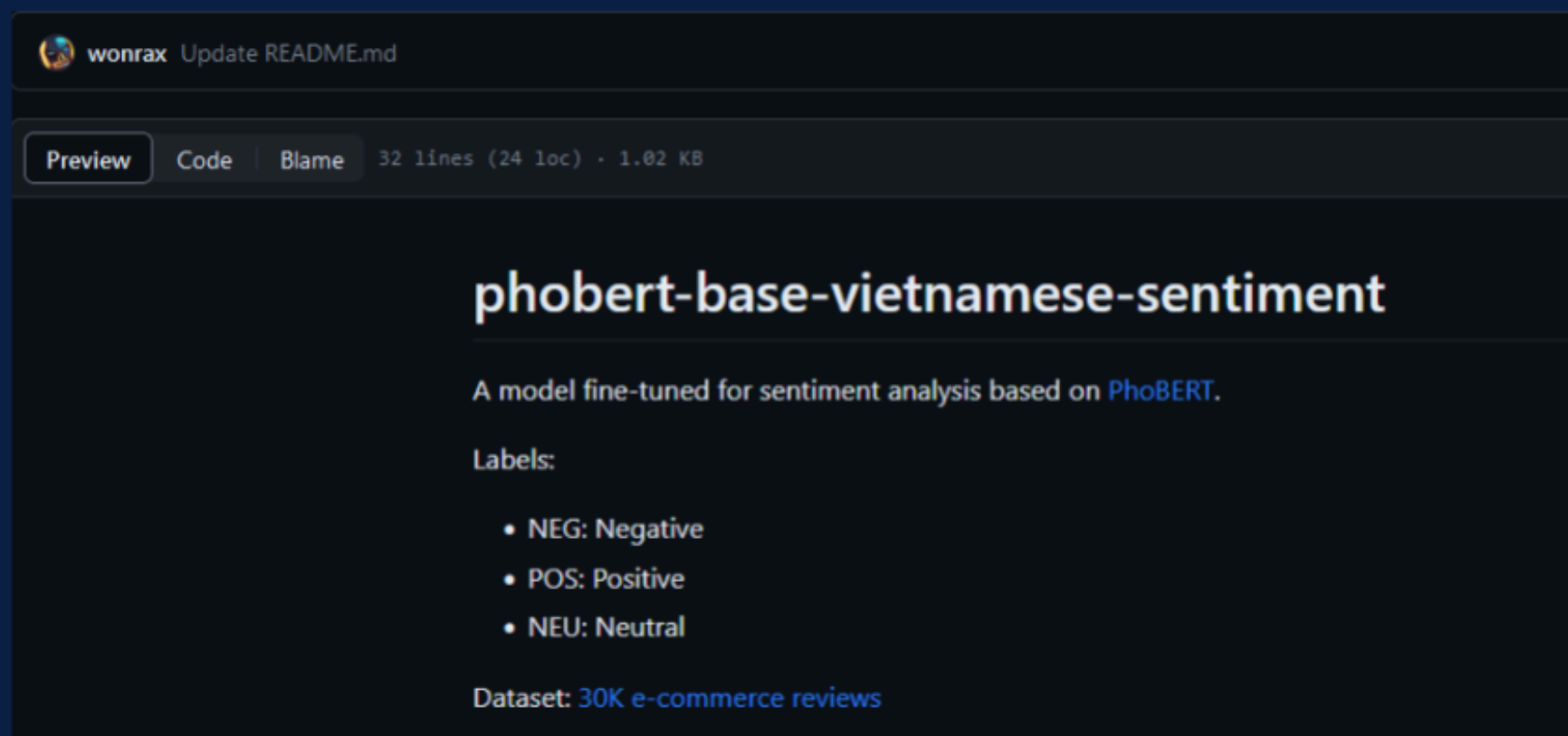
SENTIMENT ANALYSIS OVERVIEW

- **Sentiment Analysis** is the process of analyzing textual data to extract the underlying human emotions conveyed within the content. Sentiments can reflect agreement/disagreement or be classified as positive/negative.
- Sentiment Analysis is typically conducted by applying **Machine Learning** or **Deep Learning** techniques within the field of **Natural Language Processing (NLP)**.



MODEL USED

- **PhoBERT** is a Vietnamese language model developed by VinAI, designed for various natural language processing tasks such as part-of-speech tagging, sentence completion, and more.
- To perform **Sentiment Analysis**, we utilize a **pretrained model** built upon **PhoBERT** and fine-tuned on a dataset of 30,000 Vietnamese review comments collected from an e-commerce platform.



RESULT

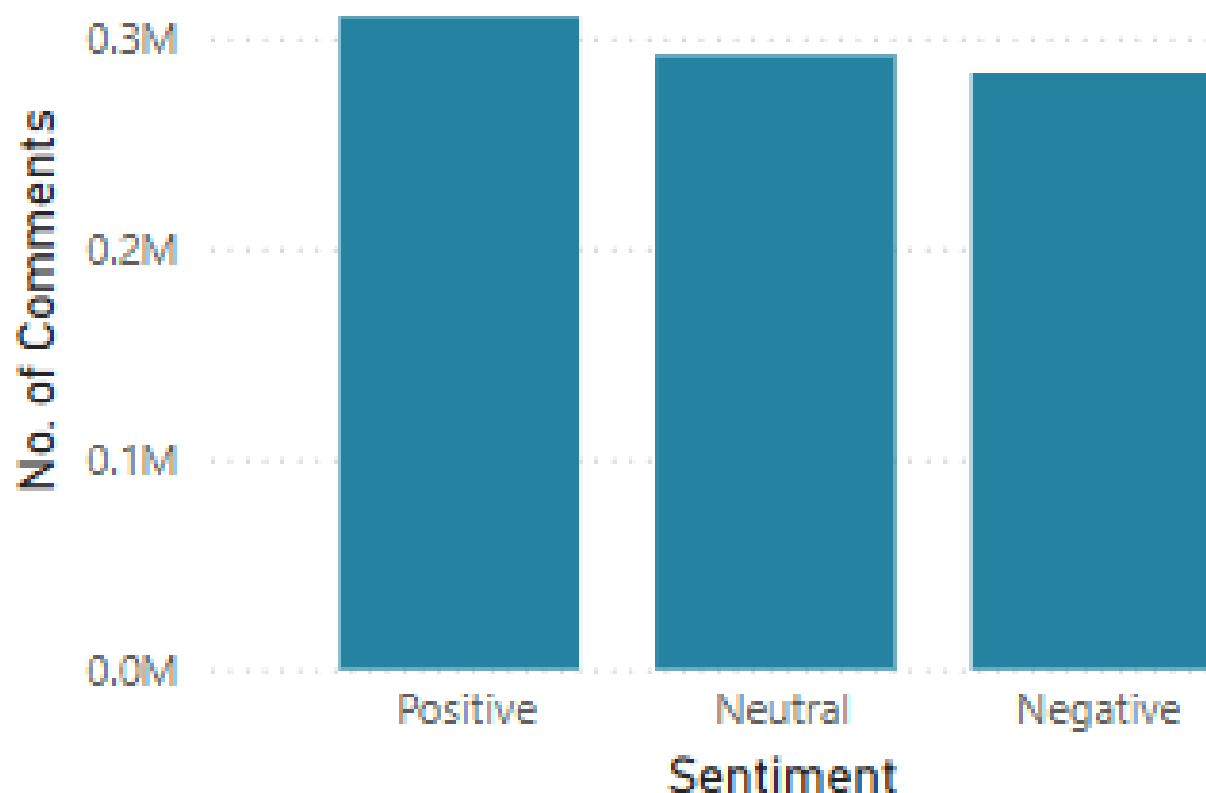
The resulting output assigns each comment a sentiment label predicted by the model: **Negative (0)**, **Positive (1)**, or **Neutral (2)**.

content_std	sentiment
hay quá à	1
phim hay toẹt vời	1
nao co tap 19 zay	1
phim hay quá đúng không mọi người 🤔🤔🤔🤔	1
ước gì có ông anh trai như eric nhỉ 🤔🤔	1
phim hay 7 5 diem	1
nữ 9 vs giang thần xuyên 1 đôi ik	1
hay 🤔🤔🤔🤔🤔	1
đáng yêu ghê 🤔🤔🤔	1
tuyệt voi	1
hay quá đi rất xúc động 🤔🤔🤔	1
phim hay quá à 🤔	1

SENTIMENT BY VIDEO

- The total number of comments classified as positive, negative, and neutral is relatively evenly distributed.
- Videos with the highest percentage of **positive comments** are primarily **VOD content**.
- While those with the highest percentage of **negative comments** are mostly **sports highlights**.

No. of Comments by Sentiment



Top Videos with most percentage of positive comments

Video ID	Tag	Positive %
56f0c8e517dc132d37183ed3	vod	81.90%
59be2367558320083c5c33a0	highlight	73.65%
567c681917dc134e55ceaaf9	vod	73.39%
5654349517dc130a14f4e304	vod	72.33%
57ff537817dc1374e49d4cb8	vod	72.22%
5acc7c4a558320088052fb4f	vod	70.25%
5631c84617dc136ea126e927	vod	70.24%
533a7c80c96928187bbcb181	vod	68.91%
Total		34.09%

Top Videos with most percentage of negative comments

Video ID	Tag	Negative %
59ce3b51558320263bc19b06	highlight	70.43%
59a112465583206c49bf54d6	highlight	66.03%
5927bf895583204f4e665554	highlight	64.43%
59bc8b90558320083a5c227e	highlight	63.31%
5d551a392089bd0872557ab2	highlight	62.57%
5c74f31f2089bd0219a64e5a	highlight	62.50%
5d529e382089bd0609da50be		60.34%
5b813ed4558320778edc6438	highlight	60.23%
5a981c27558320686157b935	vod	60.19%
Total		32.38%

Top tags with most percentage of positive comments			
Tag	Positive %	Negative %	Comments Per Video
phim-le	49.63%	27.04%	1.29
thieu-nhi	44.03%	24.72%	2.32
comment	41.15%	25.35%	1.88
phim-bo	39.42%	26.78%	2.42
anime	38.62%	26.10%	2.70
vod	37.71%	30.98%	30.42
TV show	36.84%	26.32%	1.49
Total	35.03%	32.01%	39.40

Top tags with most percentage of negative comments			
Tag	Negative %	Positive %	Comments Per Video
highlight	41.05%	26.34%	9.43
event	37.09%	30.11%	36.29
highlight	34.75%	29.31%	32.07
vod	30.98%	37.71%	30.42
	30.50%	35.51%	39.87
phim-le	27.04%	49.63%	1.29
phim-bo	26.78%	39.42%	2.42
Total	32.01%	35.03%	39.40

SENTIMENT BY TAG

- The tags with the highest percentage of **positive comments** generally have a **low total number of comments**, whereas the tags with the highest percentage of **negative comments** tend to have a **relatively large volume of comments**.
 - Content with a high degree of **niche appeal** tends to be **more appreciated by users** compared to more common content.
- Many of the tags with the highest proportion of negative comments are related to **sports** and **event** content.
 - This is understandable, as **sports** and **event-related** content often evoke strong emotions among viewers.



Thank You

