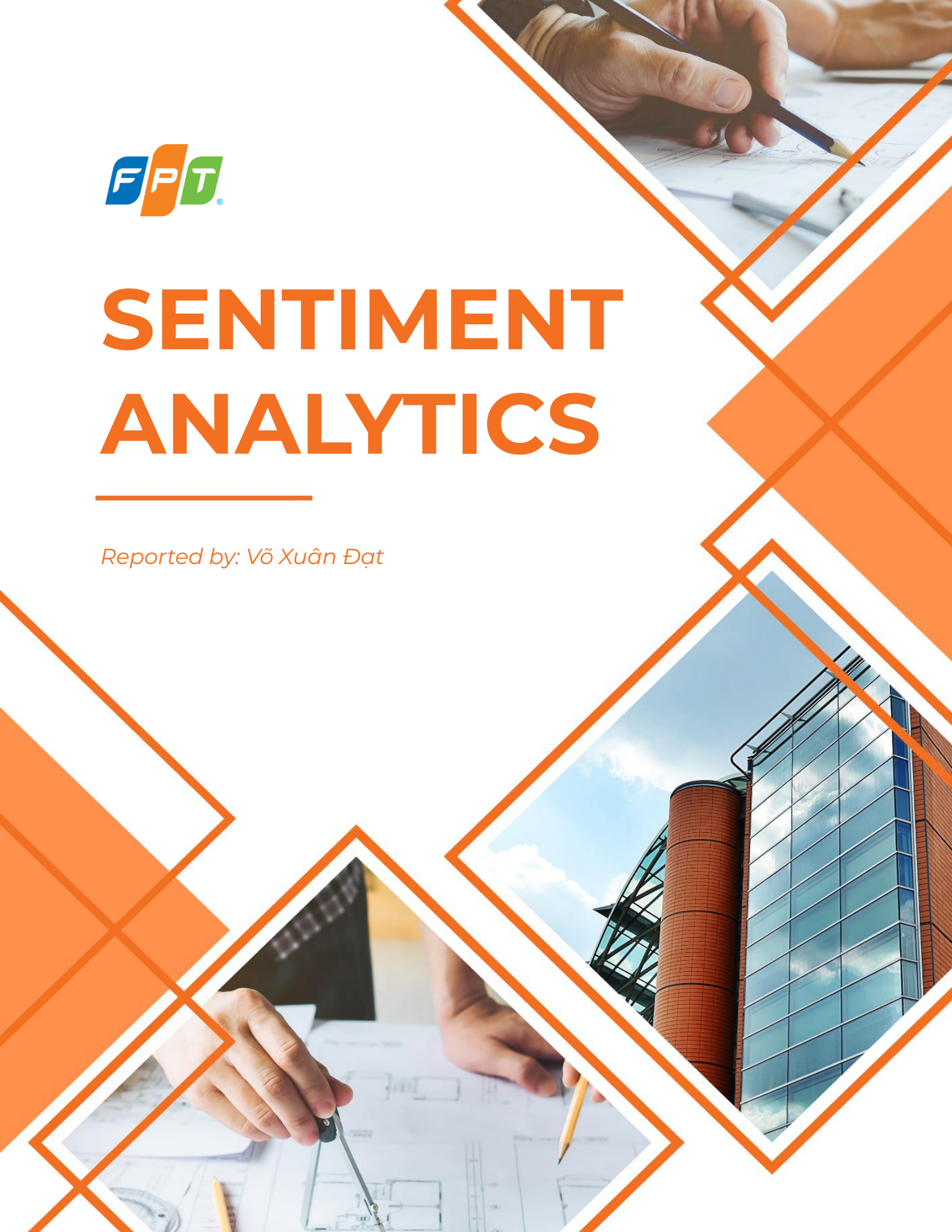




SENTIMENT ANALYTICS

Reported by: Võ Xuân Đạt



MỤC LỤC

I.	GIỚI THIỆU.....	1
II.	PHƯƠNG PHÁP SỬ DỤNG	2
1.	<i>Sentiment Analysis</i>	2
2.	<i>Mô hình phân tích ngôn ngữ tiếng Việt _ PhoBERT</i>	2
III.	CÁC BƯỚC THỰC HIỆN.....	3
1.	<i>Sử dụng Python kết nối MongoDB (NoSQL Database) để lấy dữ liệu</i>	3
2.	<i>Sử dụng Python để clean & transform data</i>	3
3.	<i>Ứng dụng model PhoBERT Base Vietnamese Sentiment</i>	3
IV.	TRỰC QUAN HÓA VÀ PHÂN TÍCH DỮ LIỆU.....	5
V.	KẾT LUẬN	9

I. GIỚI THIỆU

Có rất nhiều bình luận hằng ngày trên ứng dụng **FPT Play**, tuy nhiên, hầu hết đều không được thu thập để phân tích.

Nếu có thể phân tích được dữ liệu này và rút ra các insights thì chúng ta có thể hiểu thêm rất nhiều về khách hàng cũng như trải nghiệm khi sử dụng dịch vụ của họ, từ đó đội ngũ có thể vạch ra những chiến lược để chăm sóc khách hàng tốt hơn và cải thiện chất lượng dịch vụ đáp ứng nhu cầu của khách hàng.

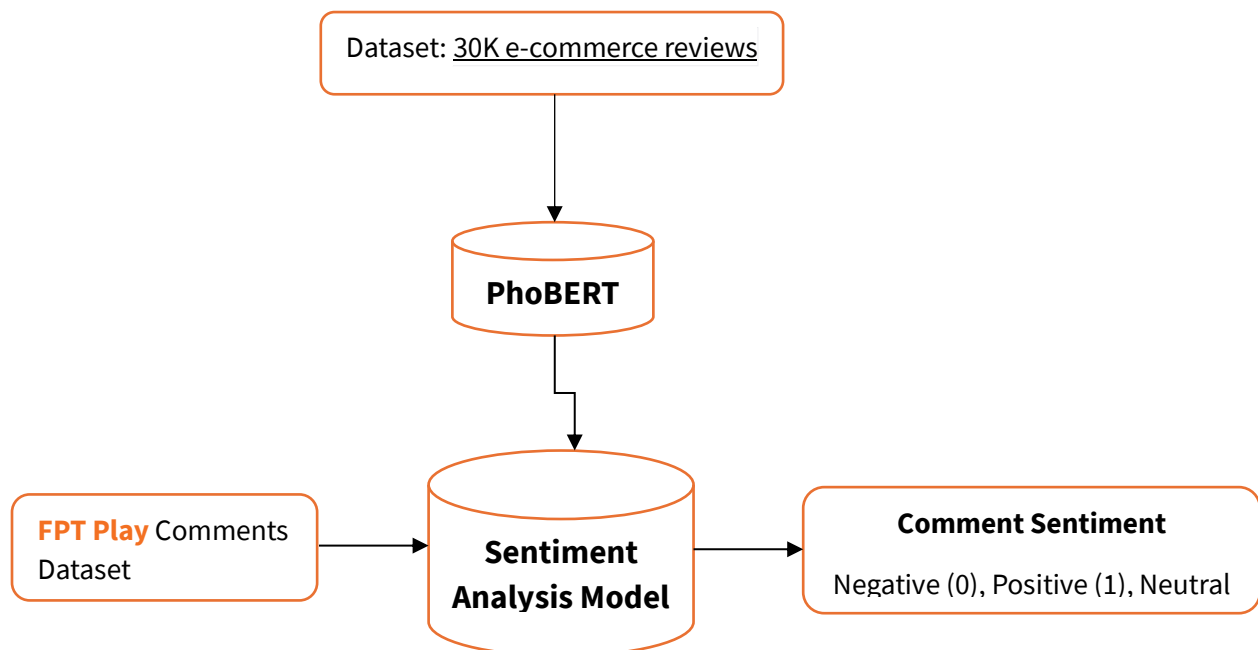
II. PHƯƠNG PHÁP SỬ DỤNG

1. Sentiment Analysis

- **Sentiment Analysis** là quy trình xử lý dữ liệu dạng văn bản để xác định cảm xúc con người bên trong dữ liệu. Cảm xúc có thể là tích cực/ đồng ý, tiêu cực/phản đối hay trung tính.
- **Sentiment Analysis** được thực hiện bằng cách áp dụng **Machine Learning** hoặc **Deep Learning** vào **NLP (Natural Language Processing)**

2. Mô hình phân tích ngôn ngữ tiếng Việt _ PhoBERT

- **PhoBERT** là một model phân tích ngôn ngữ tiếng Việt phát triển bởi VinAI, được sử dụng vào nhiều mục đích như xác định loại từ, tự động hoàn thiện câu, v.v.
- Tham khảo thêm về **PhoBERT: Pre-trained language models for Vietnamese** : [link](#)
- Để thực hiện Sentiment Analysis, chúng ta sẽ sử dụng 1 model đã được dựng sẵn, được train dựa vào model PhoBERT và dataset chứa 30,000 comments review bằng tiếng Việt trên một trang thương mại điện tử.
- Quy trình xử lý để tạo ra kết quả của Sentiment Analysis được áp dụng trong project này:



- Tham khảo thêm về **PhoBERT Base Vietnamese Sentiment**: [link](#)

III. CÁC BƯỚC THỰC HIỆN

- Xem chi tiết: [fpt_sentiment_analysis.ipynb](#)

1. Sử dụng Python kết nối MongoDB (NoSQL Database) để lấy dữ liệu

- Dữ liệu mô tả thông tin về các bình luận của người dùng trên nền tảng **FPT Play** được ghi lại trong khoảng **2015-2022**
- Có tất cả 22 cột nhưng lọc ra 8 cột cần thiết để phân tích:

#	COLUMN NAME	DATA TYPE	DATA SAMPLE	MEANING
1	_id	object	{'\$oid': '5553269a1a77ce5d65ea0604'}	Mã riêng của từng comment
2	user_id	object	464817	Mã riêng của từng user
3	comment_on	object	event	Thể loại của nội dung được comment
4	ip	object	42.116.8.164	Địa chỉ IP
5	object_id	object	54a53e4817dc131729ede7c8	Mã riêng của video
6	content	object	chuong trinh hay qua	Nội dung comment
7	timestamp	object	5/13/2015 10:32:00 AM	Thời gian comment được đăng
8	device	object	android	Thiết bị sử dụng để comment

2. Sử dụng Python để clean & transform data

- Biến đổi cột `_id` để lấy mã
- Từ cột `timestamp` đưa về đúng định dạng và thêm các cột `date` và `hour` phục vụ phân tích
- Từ cột `comment_on` tạo thêm cột `category` phân nhóm dữ liệu theo 'VOD', 'Sports', 'TV Show'

3. Ứng dụng model PhoBERT Base Vietnamese Sentiment

Từ cột `content` chứa nội dung comment của users, chạy PhoBERT Base Vietnamese Sentiment để phân nhóm các comment theo:

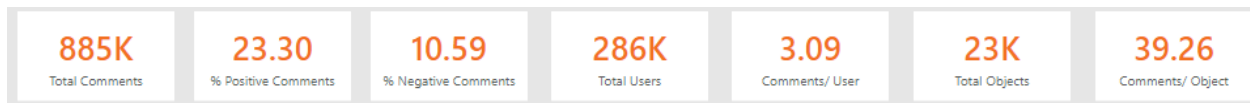
- 0: Negative (Tiêu cực)
- 1: Positive (Tích cực)
- 2: Neutral (Trung tính)

IV. TRỰC QUAN HÓA VÀ PHÂN TÍCH DỮ LIỆU

Sử dụng **PowerBI**: [link dashboard](#)

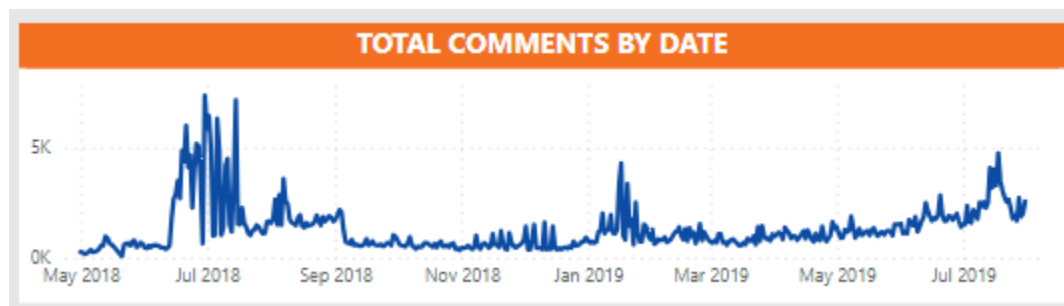
➤ Phân tích tổng quan

Dữ liệu về **885K lượt bình luận** của người dùng **FPT Play** từ 13/05/2015 đến 30/08/2022, chỉ ra trung bình mỗi user thực hiện 3.09 lượt comments và trung bình 39.26 lượt comments / 1 chương trình từ 3 loại chính là VOD, Sport và TV Show

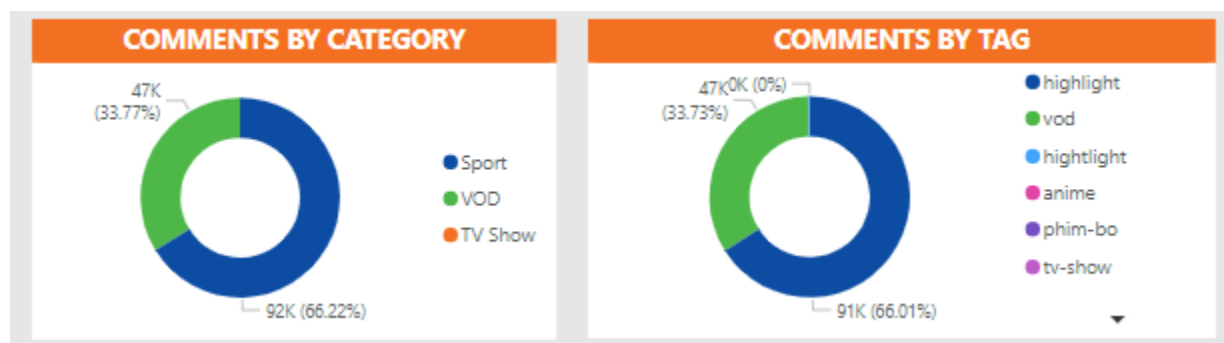


⇒ Ngoài lượng lớn người sử dụng thì không chỉ xem nội dung mà người dùng cũng có sự tương tác tốt với FPT Play

Trong khoảng thời gian các tháng 6,7 năm 2018 **chứng kiến tăng mạnh lượt comments** khi đây là lúc World Cup 2018 diễn ra. Theo sau là các tháng 6,7 năm 2019 khi lúc này diễn ra Vòng chung kết giải vô địch bóng đá Nam Mỹ.

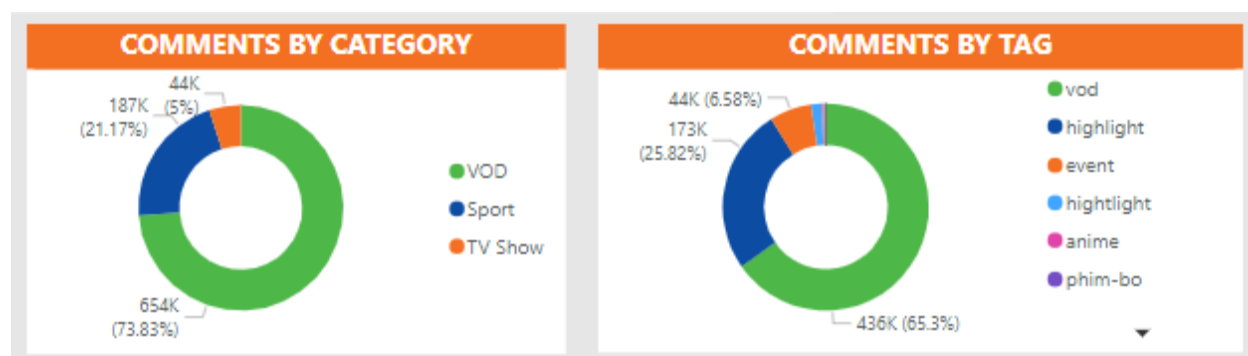


Highlight (Sport) liên quan đến bóng đá nói chung trong khoảng thời gian này và World Cup 2018, Copa America 2019 nói riêng **thu hút được nhiều lượt tương tác nhất**.



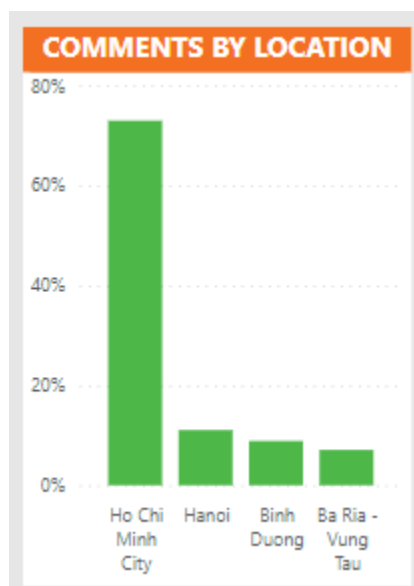
⇒ Cho thấy nội dung của FPT Play quan tâm, đáp ứng kịp thời và chính xác thị hiếu của người dùng vào những thời điểm nhất định.

Thông thường **VOD sẽ chứng kiến lượt bình luận nhiều nhất**, sau đó là Highlight và Event. Trong khi đó các nội dung khác như anime, thiếu nhi, chương trình Tết lượng bình luận không cao khi chỉ được xem và tương tác bởi 1 số nhóm người sử dụng nhất định.



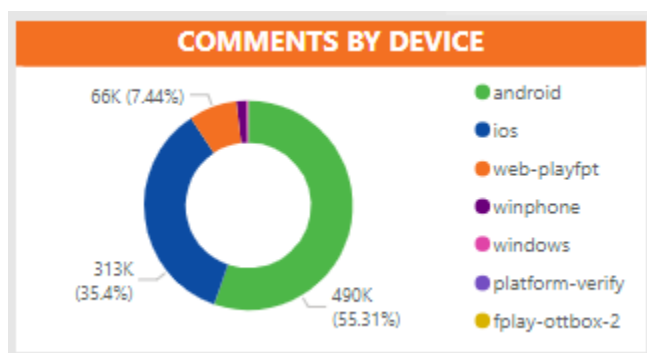
⇒ FPT Play cũng khá thành công trong xác định và quan tâm đến sở thích đa dạng của người sử dụng

Lượt comments **chủ yếu từ người sử dụng ở Tp. Hồ Chí Minh (73%)**, khá bất ngờ khi thủ đô Hà Nội lại ít tiếp cận với FPT Play hơn.



Tỉ lệ người sử dụng **bình luận trên các thiết bị Android phổ biến nhất**, chiếm hơn 1 nửa so với các thiết bị còn lại khi mà ngoài điện thoại thì các TV thông minh đều có hỗ trợ cài

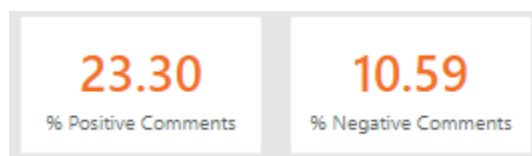
đặt app FPT Play từ Google Play Store. Tiếp sau đó là các thiết bị IOS như Iphone hay Ipad. (35.4%) Trong khi đó website chỉ chứng kiến 7.44% tổng số lượt bình luận.



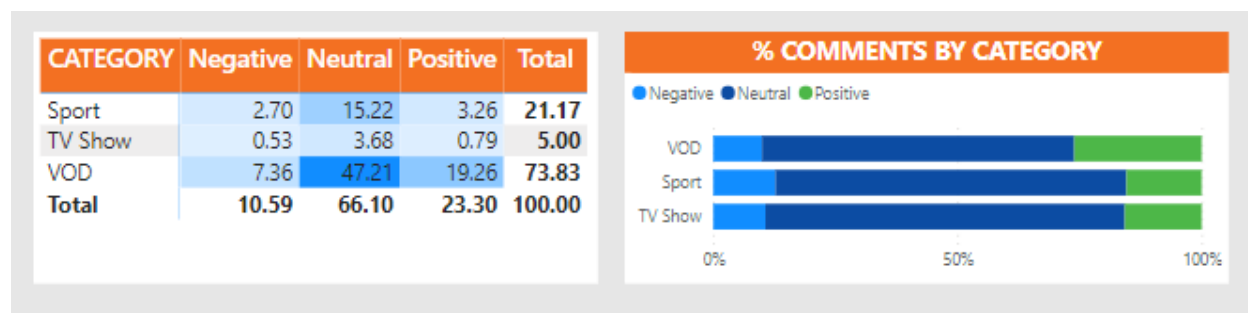
⇒ Cho thấy được sự tiện lợi và gần gũi của FPT Play App đối với người sử dụng, tuy nhiên cần cải tiến trải nghiệm ở Website nhiều hơn và FPT Play cũng nên tập trung phân khúc khách hàng ở các tỉnh/ thành phố lớn khác ngoài Tp. Hồ Chí Minh

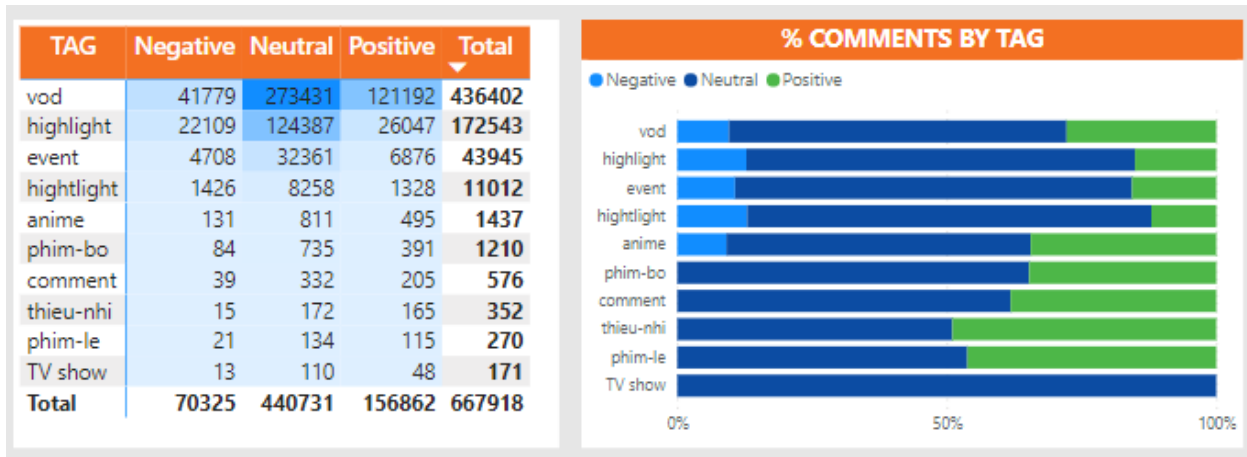
➤ Phân tích về cảm xúc (sentiment) của người dùng

Trong số 885K lượt comments trên các nền tảng của FPT Play thì **23.3% là những comments tích cực, 10.6% là những comments tiêu cực**



Nhìn chung thì các nội dung phổ biến như vod, highlight hay event **lượng tương tác cao đi đôi với số comments tiêu cực cũng cao theo**. Ngoại trừ anime khi lượng comments không nhiều nhưng comments tiêu cực cũng chiếm kha khá với 10.71%.





⇒ Cũng khá dễ hiểu khi anime thường được những người sử dụng độ tuổi thanh thiếu niên quan tâm

Các nội dung liên quan đến phim và đặc biệt cho thiếu nhi **không chứng kiến comments tiêu cực nào**

Comments tiêu cực ở đây cả về liên quan đến nội dung như phim ra chậm, kết phim không đúng ý, đội đã thua, quyết định thiên vị từ trọng tài,... từ những users phản ứng thái quá và về dịch vụ, trải nghiệm người dùng như mạng chậm, lag, mất kết nối, quá nhiều quảng cáo, âm thanh chất lượng lỏng lẻo tiếng phim,...

⇒ 10.6% là những comments tiêu cực cũng là 1 điểm tốt đáng khen cho đội ngũ FPT Play nhưng cần kiểm soát hơn những comment tiêu cực về nội dung để văn hoá văn minh hơn và cải thiện cơ sở hạ tầng mang đến trải nghiệm tốt hơn cho người sử dụng

V. KẾT LUẬN

- **FPT Play** đã thể hiện được sự quan tâm, đáp ứng nhu cầu, thị hiếu của người dùng với đa dạng các nội dung VOD, Sports, TV Shows, đặc biệt vào các dịp cao điểm trong các tháng hè.
- **FPT Play** cần cải tiến để tăng trải nghiệm trên nền tảng website nhiều hơn khi các thiết bị android hay ios chiếm tỉ lệ quá vượt trội cũng như cần tập trung vào các khách hàng ở các tỉnh/ thành phố khác ngoài Tp. Hồ Chí Minh.
- Các nội dung phổ biến xem nhiều đi đôi với lượng comments tiêu cực cao như vod, highlight, events cần giải quyết ở 2 khía cạnh: Nội dung và dịch vụ

Góp ý:

Các giai đoạn đầu năm và các tháng hè 6,7 cũng như các khung giờ 13h-17h sẽ chứng kiến lượng người xem nhiều hơn đi đôi với lượt tương tác, comments nhiều hơn vì đây là các khoảng thời gian nghỉ Tết, học sinh sinh viên được nghỉ hè, người đi làm không mắc khối lượng công việc lớn như cuối năm

⇒ FPT ngoài việc cân đối nội dung để đáp ứng thị hiếu người dùng thì cần cân nhắc thêm về các dịch vụ đi kèm như tích điểm đổi thưởng cho những users có lượt tương tác nhiều nhất, khuyến mãi combo sử dụng trong 2,3 tháng,... ; cân đối quảng cáo, tiếp thị; tăng băng thông mạng những lúc cao điểm kể trên; dồn nguồn lực xử lý, lọc, nhắc nhở với những comment tiêu cực liên tục => Trải nghiệm người dùng tốt hơn và tối ưu chi phí, tối đa doanh thu.