

Final Project

▼ Giới thiệu

Xuyên suốt lịch sử trong lĩnh vực trò chơi điện tử có nhiều tựa game nổi tiếng qua các thời kì, được đông đảo cộng đồng người chơi ưa thích ví dụ như các thể loại game:

- Bắn súng: **Call of Duty: Modern Warfare, Valorant, ...**
- Chiến thuật: **Dota 2, Liên quân mobile, ...**
- Sinh tồn: **Minecraft, PUBG, ...**

Và còn rất nhiều tựa game khác vẫn đang được phát triển, cập nhật, sửa lỗi cho đến ngày nay

Mỗi trò chơi đều có một Fanpage, Website cụ thể riêng để thông báo và đưa những tin tức về sự thay đổi trong phiên bản mới sắp tới, những dự án đang phát triển,... cho cộng đồng đều có thể biết.

Vấn đề là: Có phải Fanpage chỉ nói về game mà không liên quan đến những thứ khác hay không? Có phải mọi người trong Fanpage đều quan tâm đến thông tin đến các bài đăng hay không? Mọi người có thường xuyên tương tác, cập nhật thông tin hay không?

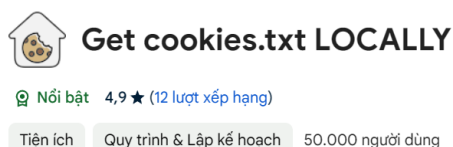
Bài báo cáo này giải đáp những thắc mắc và phân tích những suy nghĩ, cảm nhận, đánh giá của những người quan tâm đến tựa game tiêu biểu đó là game **League of Legends** trên một trang cộng đồng trên nền tảng **Facebook**.

▼ Nội dung

▼ Dữ liệu

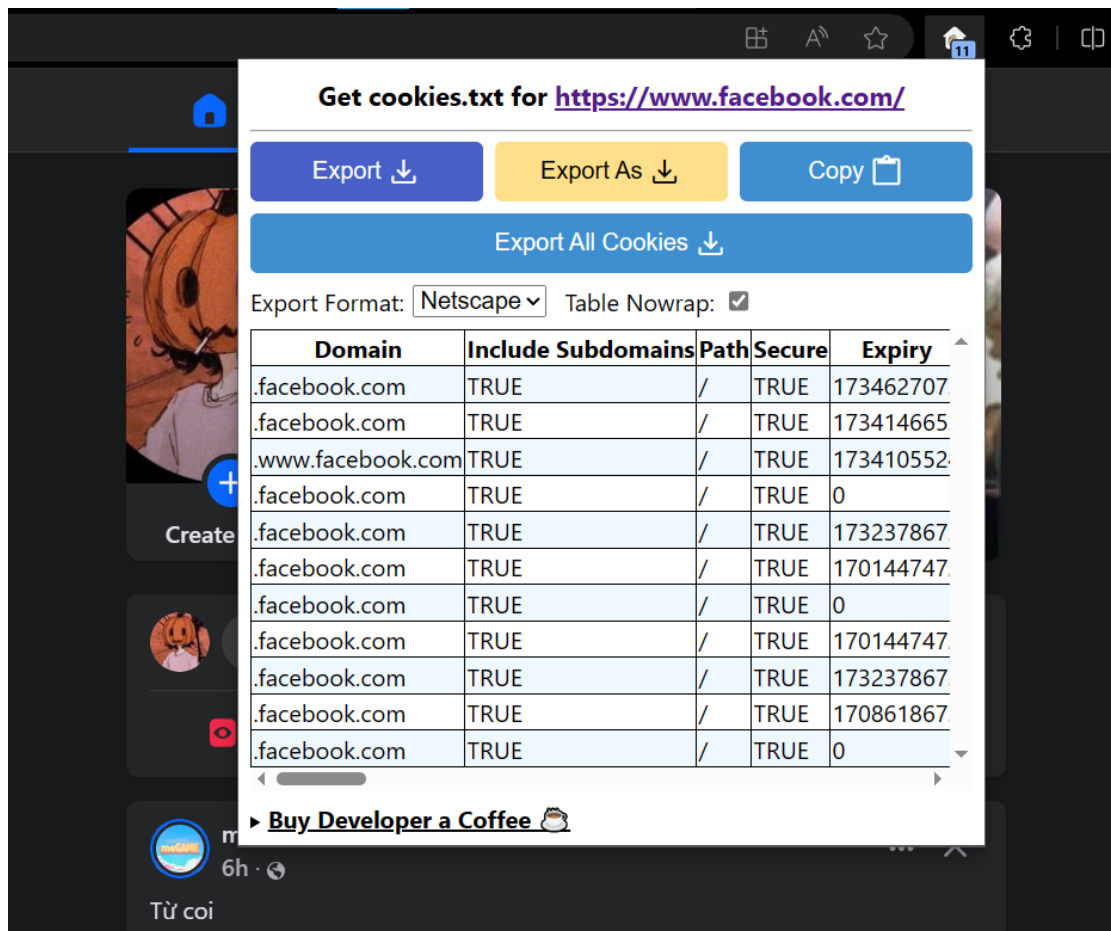
Để hỗ trợ cho việc thu thập dữ liệu thì đầu tiên:

- Bước 1: Cần tải **Get cookies.txt LOCALLY** từ cửa hàng **Chrome**.

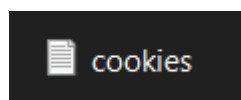


Tải về

- Bước 2: Sau đó vào trang cá nhân trên **Facebook** rồi mở tiện ích vừa cài đặt và chọn **Export** để tải tập về máy.



Khi tải về hoàn tất chúng ta nhận được file **cookies.txt**



Việc sử dụng **Cookies** cho việc thu thập dữ liệu dùng để xác thực người dùng và cấp quyền truy cập vào các trang web hoặc tài nguyên cụ thể, tránh phải đăng nhập nhiều lần vào cùng 1 trang web. Ngoài ra còn để lưu trữ dữ liệu của người dùng.

- Bước 3: Tiếp theo tải và thêm thư viện **facebook_scraper** và thêm hàm **get_posts()** vào đoạn code để lấy dữ liệu từ **Facebook** và trích xuất thông tin từ nó.


```
%pip install facebook_scraper
from facebook_scraper import get_posts
```

Hàm **get_posts()** là một hàm trong thư viện này, được sử dụng để lấy các bài đăng từ một trang **Facebook**, nhóm hoặc hồ sơ.

Đây cũng là một công cụ được sử dụng để thu thập dữ liệu, cho phép các nhà phát triển truy cập dữ liệu từ **Facebook** một cách hợp pháp và tôn trọng quyền riêng tư của người dùng.

- Bước 4: Sau khi thu thập dữ liệu thành công, cài đặt code để tạo một khung dữ liệu Pandas từ danh sách bài đăng, lưu khung dữ liệu vào tệp CSV.
- Bước 5: Tiếp theo là lọc và tổ chức lại dữ liệu thô vừa thu thập được.

Dữ liệu thô này có thể chứa nhiều trường không có giá trị nào hoặc những trường này ta không sử dụng phục vụ cho việc phân tích.


 Data.csv

Có thể xóa các cột có tất cả các giá trị là NaN, các cột không sử dụng và lọc một số hàng có chứa giá trị NaN.

Còn có trường chứa những trường nhỏ hơn như **reactions**, **time**. Trong trường **reactions** có **angry**, **care**, **haha**, **like**, **love**, **sad**, **wow**. Trong trường **time** có chứa **day**, **month**, **year**, **hour**.

Việc tách các trường lớn thành các trường nhỏ có thể dễ dàng thống kê hơn, cải thiện hiệu quả cho việc phân tích, đáp ứng các yêu cầu cho hệ thống dữ liệu cụ thể.

Sau đó dữ liệu thô đã được làm sạch lưu sang một file CSV khác để tránh làm hỏng dữ liệu gốc, làm cho dữ liệu dễ dàng truy cập và sử dụng hơn, tạo điều kiện cho việc phân tích dữ liệu.

 Data_Cleaned.csv

▼ Phương pháp

Bài báo cáo này phân tích dựa vào những phương pháp sau:

- **Phân tích cụm (Cluster analysis)**

Một kỹ thuật phân tích dữ liệu được sử dụng để gom nhóm các đối tượng trong một tập dữ liệu dựa trên sự tương đồng giữa chúng. Mục tiêu của phân tích cụm là tìm ra các cụm (cluster) sao cho các đối tượng trong cùng một cụm có tính giống nhau nhất, trong khi các đối tượng thuộc các cụm khác nhau có tính khác nhau nhất.

- **Phân tích theo nhóm (Cohort analysis)**

Một kỹ thuật phân tích dữ liệu được sử dụng để theo dõi hành vi của một nhóm người dùng hoặc khách hàng có chung một đặc điểm nào đó. Các nhóm này, thường được gọi là "cohort", chia sẻ những đặc điểm hoặc trải nghiệm tương tự trong một khoảng thời gian cụ thể.

- **Phân tích chuỗi thời gian (Time series analysis)**

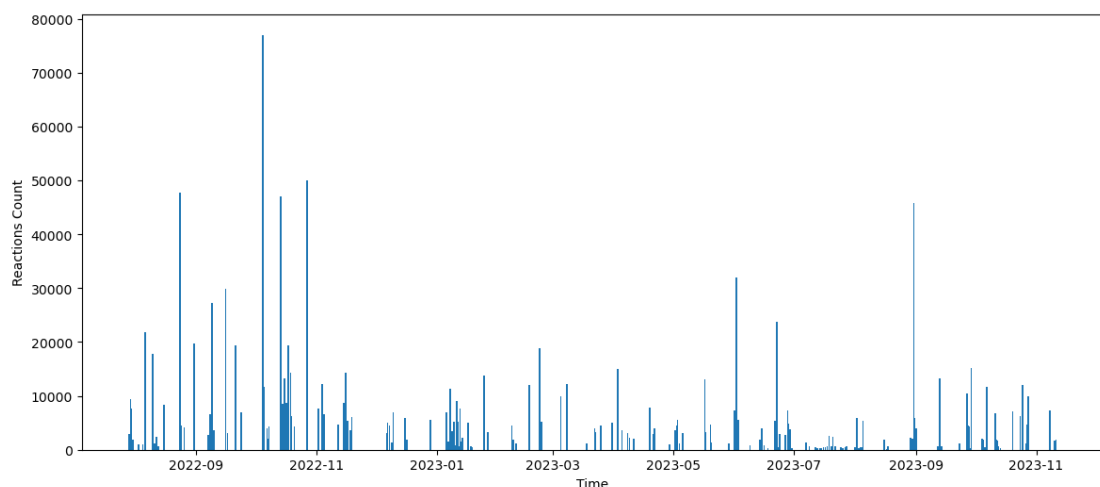
Một kỹ thuật thống kê được sử dụng để nghiên cứu các dữ liệu được thu thập theo thời gian. Mục tiêu của phân tích chuỗi thời gian là tìm hiểu các đặc điểm của dữ liệu, chẳng hạn như xu hướng, chu kỳ, và biến động ngẫu nhiên.

- **Phân tích định tính:** Sử dụng các kỹ thuật phân tích định tính để hiểu ý nghĩa của dữ liệu.

▼ Phân tích

Mục tiêu của phần phân tích này đưa ra cái nhìn trực quan hơn về một trang cộng đồng trên **Facebook** liên quan đến game **League of Legends** nói riêng và các thể loại game khác nói chung

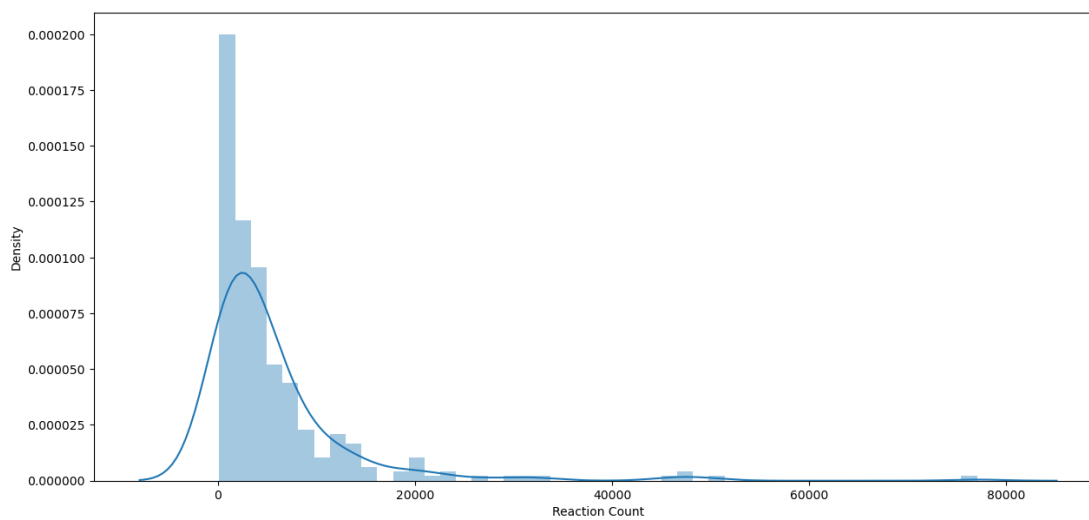
Biểu đồ đầu tiên nói số **reactions** của các bài đăng từ ngày 29 tháng 7 năm 2022 đến ngày 11 tháng 10 năm 2023.



Fanpage có tổng số 16 triệu lượt thích tính đến thời điểm tháng 11 năm 2023. Biểu đồ trên thống kê khoảng 300 tổng số bài đăng gần nhất và số lượt **reactions** của mỗi bài đăng.

Bài có nhiều **reactions** nhất vào ngày 2022-10-05 với xấp xỉ 77000 lượt thả biểu tượng cảm xúc.

Tiếp theo là dạng biểu đồ phân phối tần suất của số lượng tương tác trên các bài đăng.

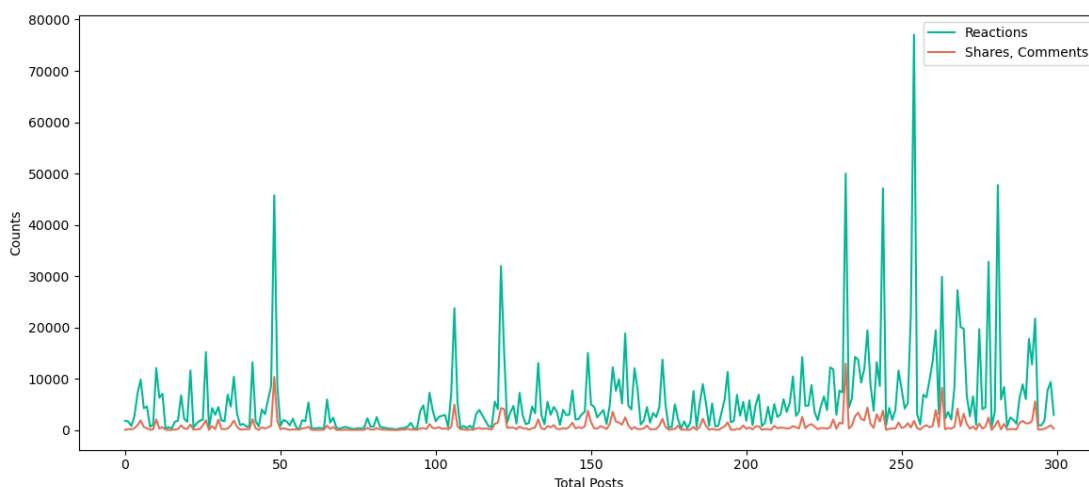


Trong khoảng 300 bài gần đây, tần số của số lượt tương tác nhiều nhất rơi vào khoảng 0 đến 10000 lượt tương tác mỗi bài. Khoảng 10000 đến 40000 lượt cũng khá phổ biến và rất ít bài có tương tác trên 40000.

Biểu đồ kế tiếp thể hiện tần suất các từ xuất hiện nhiều nhất trong nội dung tiêu đề của tổng 300 bài đăng.

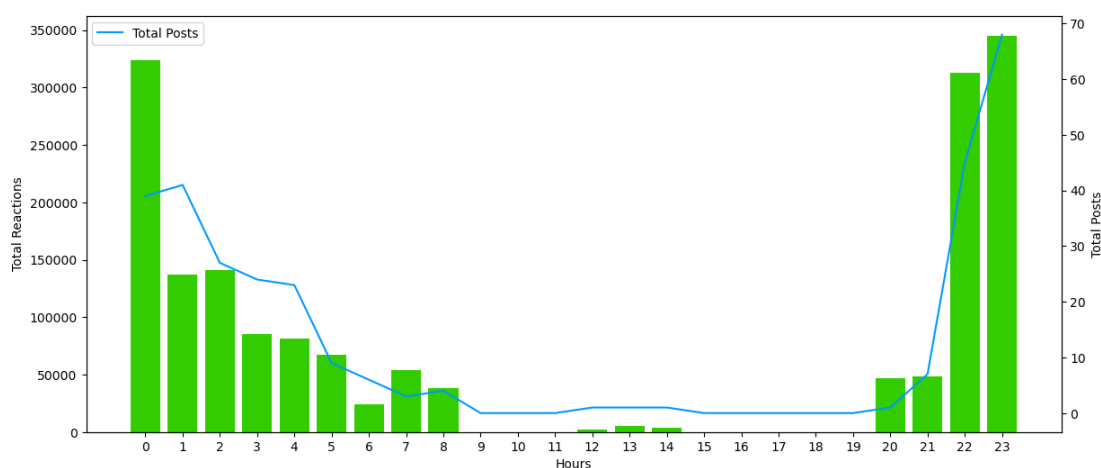
Dữ liệu được lọc lấy cột **post_text** và xử lý các từ xuất hiện trong 300 bài.

Biểu đồ sau đây thể hiện sự tương quan giữa việc thả biểu cảm tiện lợi, nhanh chóng hơn với việc bày tỏ bằng lời nói, bình luận cũng nói lên được bài viết đó có nội dung như thế nào.



Tiếp đến là 1 dạng biểu đồ kết hợp giữa biểu đồ đường và biểu đồ cột thể hiện sự tương quan giữa số lượt tương tác và số bài đăng qua các mốc thời gian cụ thể là tổng số 24 giờ từ 0 giờ đến 23 giờ.

Dữ liệu trích xuất tổng số lượt tương tác và tổng số bài đăng theo nhóm giờ.

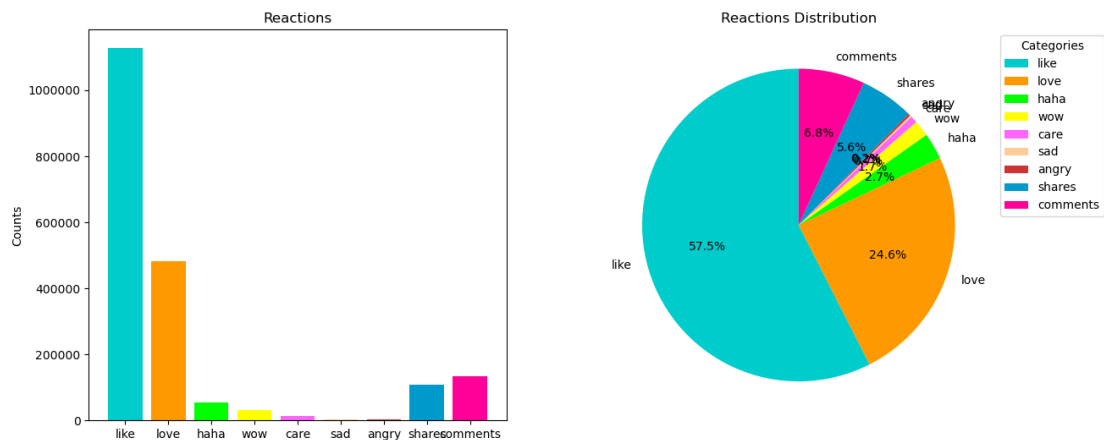


Fanpage hoạt động với tổng số bài đăng và lượt tương tác phần lớn rơi vào khoảng 22 giờ tối đến 1 giờ sáng tính theo giờ Việt Nam. Khung giờ khá phổ biến là 2 giờ đến 8h sáng cộng thêm 20, 21 giờ tối. Từ 9 giờ cho đến 19 giờ, con số đăng bài không đáng kể, nên số tương tác bị kéo theo.

Kế tiếp là biểu đồ đường chỉ ra độ chênh lệch giữa số lượt **like, love, haha, wow, care, sad, angry** với số lượt **shares, comments**.

Biểu đồ cột và biểu đồ tròn sau đây thống kê các số liệu về các cảm xúc, mức độ quan tâm, đóng góp ý kiến, tương tác, chia sẻ của người dùng **Facebook** với tổng số biểu cảm **like, love, haha, wow, care, sad, angry, shares, comments** và tỉ lệ giữa chúng.

Dữ liệu lọc các cột tương tác và cộng tất cả lượt tương tác đó từ 300 bài đăng.

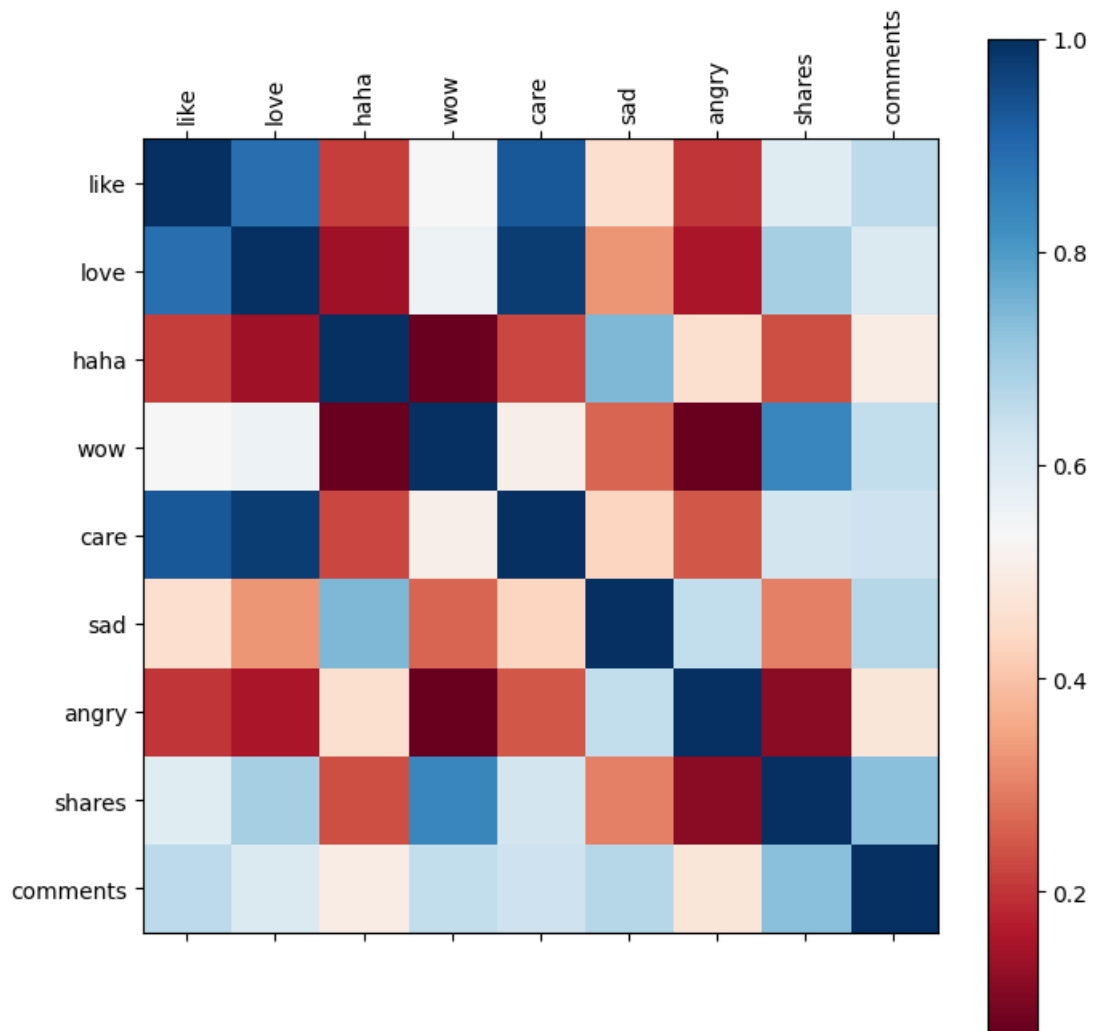


Like chiếm tỉ lệ 57,5% với hơn 1 triệu lượt thả, **love** chiếm gần 1/4 tổng số lượt tương tác với gần 50000 lượt thả, còn lại các biểu cảm khác chiếm không đáng kể xấp xỉ 3% trở xuống. Số lượt **shares** và **comments** khá phổ biến với tỉ số lần lượt là 5,6% và 6,8%.

Cuối cùng là biểu đồ tương quan thể hiện mối quan hệ tương quan giữa các biến số trong dữ liệu cụ thể là số các biểu cảm, bình luận, chia sẻ. Mỗi ô trong biểu đồ thể hiện hệ số tương quan giữa hai biến số tương ứng. Hệ số tương quan có giá trị từ -1 đến 1, trong đó:

- **Hệ số tương quan bằng -1:** Biến số có mối quan hệ nghịch đảo hoàn hảo, nghĩa là khi một biến số tăng thì biến số kia giảm và ngược lại.
- **Hệ số tương quan bằng 0:** Biến số không có mối quan hệ tương quan, nghĩa là sự thay đổi của một biến số không ảnh hưởng đến sự thay đổi của biến số kia.
- **Hệ số tương quan bằng 1:** Biến số có mối quan hệ thuận nghịch hoàn hảo, nghĩa là khi một biến số tăng thì biến số kia cũng tăng và ngược lại.

Màu sắc của các ô trong biểu đồ thể hiện mức độ tương quan giữa các biến số. Màu đỏ thể hiện mối quan hệ thuận nghịch mạnh, màu xanh thể hiện mối quan hệ nghịch đảo mạnh, và màu trắng thể hiện không có mối quan hệ tương quan.



▼ Kết quả

Kết quả của việc phân tích dữ liệu các biểu đồ trên chỉ ra rằng:

Fanpage không thường xuyên đăng bài. Có ngày đăng 2-3 bài, có ngày không đăng bài nào. Cho thấy đây là một trang **Facebook** không có xu hướng tương tác cố định, số lượt tương tác không đều. Không đăng bài với mục đích để tăng số lượt tương tác.

Bài có nhiều tương tác nhất vào ngày 2022-10-05 với xấp xỉ 77000 lượt tương tác. Nội dung của bài này không phải nói về bản cập nhật, lỗi chơi hay bất kì thứ gì liên quan đến game **League of Legends** mà đây chỉ là một bài đăng thông báo thay đổi ảnh đại diện của trang

(<https://facebook.com/leagueoflegends/posts/10166328789730556>).

Bài viết tiếp theo có gần 50000 lượt tương tác nhiều thứ 2 nói về một sự kiện có thông điệp từ tổng thống với sự hợp tác của **Riot Game** -nhà phát hành trò

chơi điện tử trong đó có cả **League of Legends**- và ca sĩ nổi tiếng **Lil Nas X** (<https://facebook.com/leagueoflegends/posts/10167124921715556>).

Hai bài viết chứa nội dung ngoài lề của trò chơi có nhiều lượt reactions nhất cho thấy Fanpage không chỉ đăng dự án bản cập nhật của trò chơi mà còn đăng những thông tin hữu ích khác cho cộng đồng yêu thích game **League of Legends** như thông điệp từ những người nổi tiếng, giải đấu thế giới, ca sĩ khách mời cho sự kiện lớn...

Bên cạnh đó nội dung của tiêu đề các bài đăng chứa nhiều từ **new** nhất trong số 300 bài.

Có thể thấy những bài đăng phổ biến đều là những thông tin về bản cập nhật mới, trang phục mới, tướng mới, mùa giải và chế độ chơi mới,... Nhưng với 2 bài tương tác cao ở trên thấy được rằng cộng đồng quan tâm nhiều hơn đến những thông tin độc lạ hơn là những bản cập nhật, sửa lỗi thường xuyên.

Sự quan tâm của cộng đồng với bài viết nếu chỉ ấn một nút như thả cảm xúc không bằng những dòng bình luận thể hiện cảm xúc sâu sắc hơn. Nội dung của hầu hết các bình luận xoay quay từ **skin, game**. Rõ ràng mọi người rất quan tâm đến những dòng trang phục mới đẹp mắt, huyền ảo, độc đáo của những vị tướng.

Ngoài từ **Nguyễn** cũng là từ xuất hiện phổ biến. Từ đó cũng biết được rằng đông đảo người Việt Nam khá chú ý thể loại game này trên một trang quốc tế nước ngoài.

Lượt tương tác **like, love** trong hầu hết 300 bài viết chiếm phần lớn, các bài đăng đều có tính chất là thông báo, thông tin đến với cộng đồng. Trong khi đó còn có một số bài đăng mang tính chất giải trí với lượt thả **haha** nhiều hơn **like** (<https://facebook.com/leagueoflegends/posts/851594899670301>, <https://facebook.com/leagueoflegends/posts/836362397860218>). Cho thấy nội dung bài đăng tác động lớn đến suy nghĩ của người dùng. Nếu nội dung không mạch lạc, rõ ràng có thể khiến người đọc, người nghe hiểu sai lệch, gây ảnh hưởng đến nhiều vấn đề như doanh thu, danh dự, lòng tin của mọi người,...

▼ Kết luận

Tóm lại bài báo cáo giải đáp được rằng trang **Facebook** của game **League of Legends** không chỉ nói về các dự án, bản cập nhật của game mà còn chia sẻ vô số thông tin độc đáo khác nữa. Mọi người quan tâm rất nhiều đến những thông tin ngoài lề của trò chơi như sự kiện và giải đấu thế giới,... Group có lượng người

tham gia lớn mà không phải group nào cũng có nhưng chỉ 1 phần nhỏ trong số đó tương tác, quan tâm nhiều hơn đến những thông tin về game và cả về những sự kiện trong tương lai.