**Đề tài : Đề xuất các địa điểm được mở trong giờ nhất định.**

**1.Cơ sở lý thuyết**

Python là một ngôn ngữ lập trình bậc cao cho các mục đích lập trình đa năng, do Guido Van Rossum tạo ra và lần đầu ra mắt vào năm 1991. Python được thiết kế với ưu điểm mạnh là dễ đọc, dễ học và dễ nhớ. Python là ngôn ngữ có hình thức rất sáng sủa, cấu trúc rõ ràng, thuận tiện cho người mới học lập trình. Cấu trúc của Python còn cho phép người sử dụng viết mã lệnh với số lần gõ phím tối thiểu. Vào tháng 7 năm 2018, Van Rossum đã từ chức Leader trong cộng đồng ngôn ngữ Python sau 30 năm lãnh đạo. Sự phát triển Python đến nay có thể chia làm các giai đoạn: python 1, python 2 và python 3. [2]

Thư viện pandas trong python là một thư viện mã nguồn mở, hỗ trợ đắc lực trong thao tác dữ liệu. Đây cũng là bộ công cụ phân tích và xử lý dữ liệu mạnh mẽ của ngôn ngữ lập trình python. Thư viện này được sử dụng rộng rãi trong cả nghiên cứu lẫn phát triển các ứng dụng về khoa học dữ liệu. Thư viện này sử dụng một cấu trúc dữ liệu riêng là Dataframe. Pandas cung cấp rất nhiều chức năng xử lý và làm việc trên cấu trúc dữ liệu này. Chính sự linh hoạt và hiệu quả đã khiến cho pandas được sử dụng rộng rãi. [4]

Hàm locchấp nhận label của các row và column và trả về Chuỗi hoặc DataFrame. Bạn có thể sử dụng nó để lấy toàn bộ row hoặc column, cũng như các phần của chúng.

Mô hình n -gram mô hình hóa các chuỗi, đặc biệt là các ngôn ngữ tự nhiên, sử dụng các thuộc tính thống kê của n -gram. Mô hình n -gram hiện được sử dụng rộng rãi trong xác suất , lý thuyết giao tiếp , ngôn ngữ học tính toán (ví dụ, xử lý ngôn ngữ tự nhiên thống kê ), sinh học tính toán (ví dụ, phân tích chuỗi sinh học ) và nén dữ liệu.

Bigram hay còn gọi là digram là một dãy bao gồm hai phần tử liền kề nhau từ một chuỗi các token, chẳng hạn như các ký tự, các âm tiết hoặc các từ. Một bigram là một n-gram với n bằng 2. Phân phối tần số của mỗi bigram trong một chuỗi thường dùng cho phân tích thống kê văn bản đơn giản trong nhiều ứng dụng, bao gồm ngôn ngữ học tính toán, mật mã học, nhận dạng tiếng nói và nhiều ứng dụng khác.

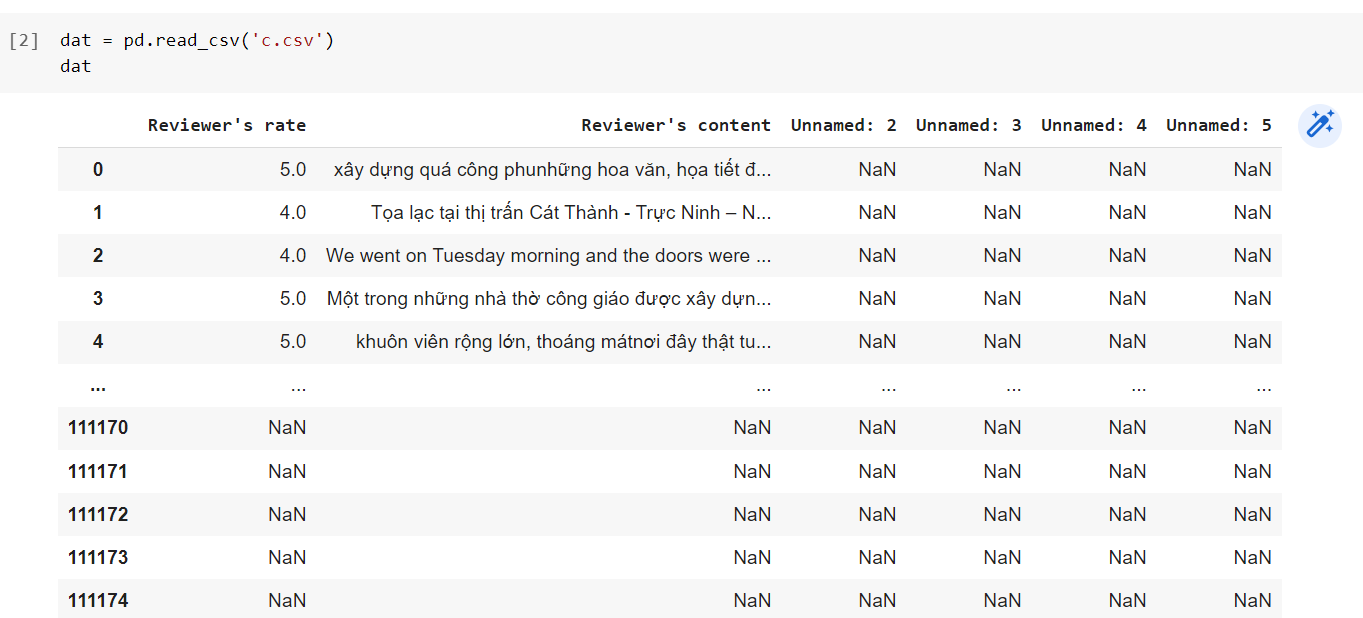
Trigram là một trường hợp đặc biệt của n-gram, với n là 3. Trigram thường được sử dụng trong xử lý ngôn ngữ tự nhiên để thực hiện phân tích thống kê văn bản và trong mật mã học để kiểm soát và sử dụng cho các cipher và mã hiệu. Trigram thường được dùng để dò lọc nội dung spam, dò lỗi chính tả trong văn bản. Người ta thường phân tích Trigram ở cấp độ từ, cấp độ ký tự, và cấp độ token.

**2.Code**

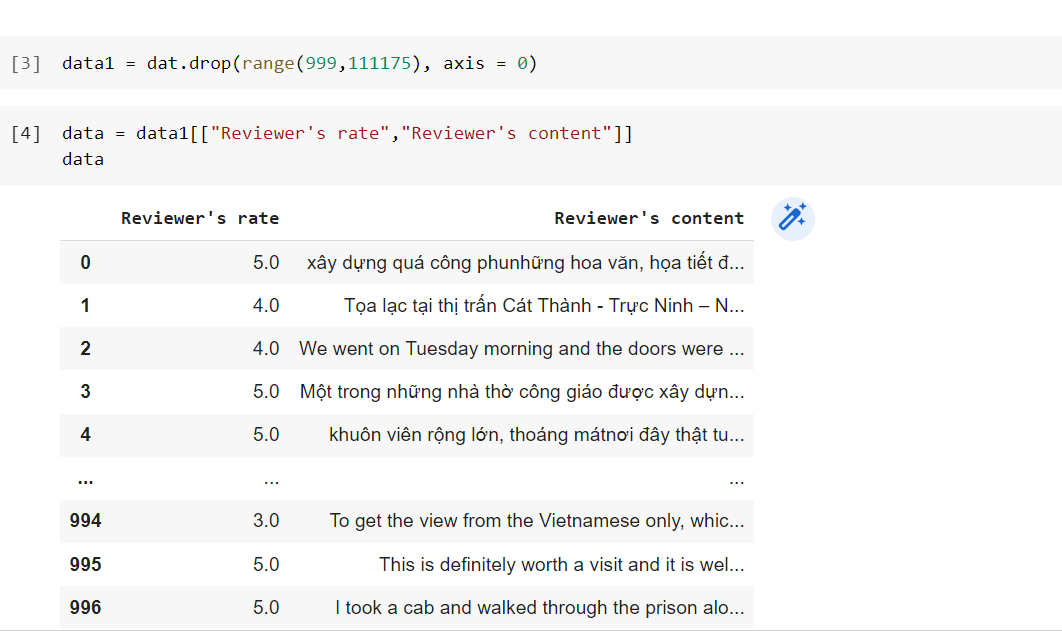
Các thư viện được sử dụng



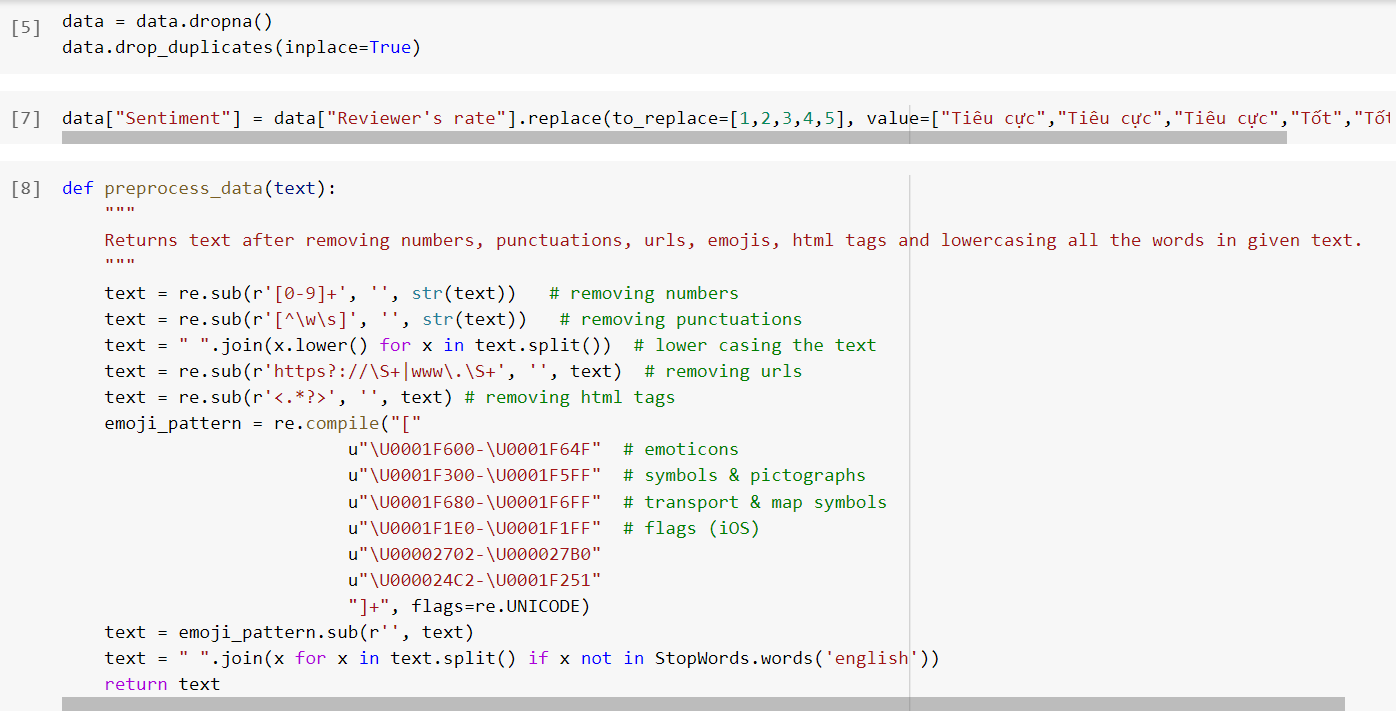
Data sử dụng



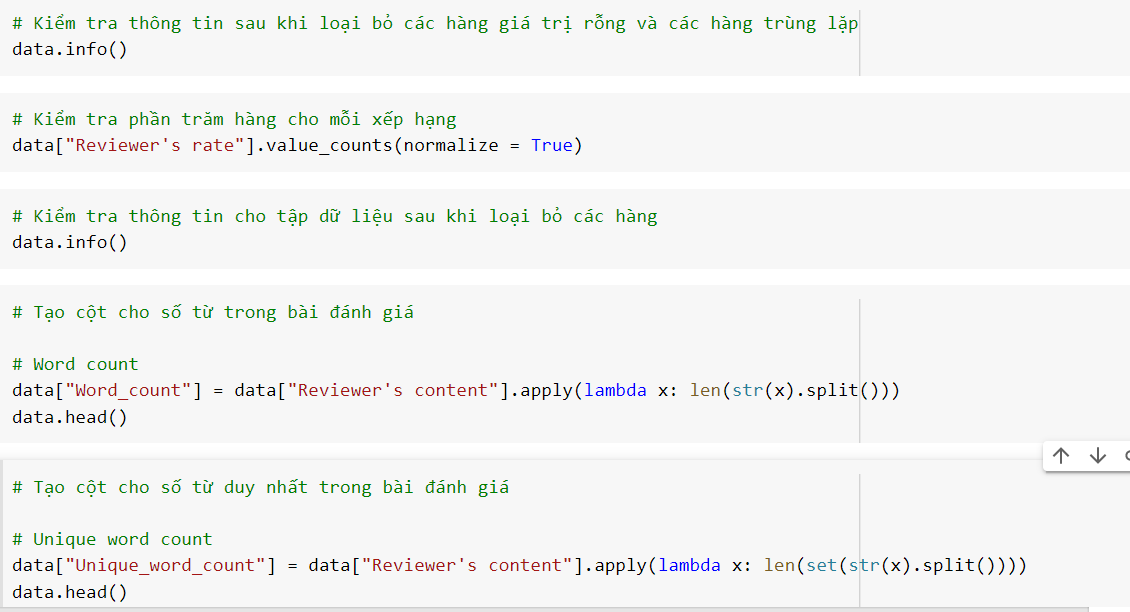
Rút gọn data về những data dùng được



Loại bỏ các data trùng nhau, xử lý dữ liệu.



Kiểm tra dữ liệu, mở rộng dữ liệu.



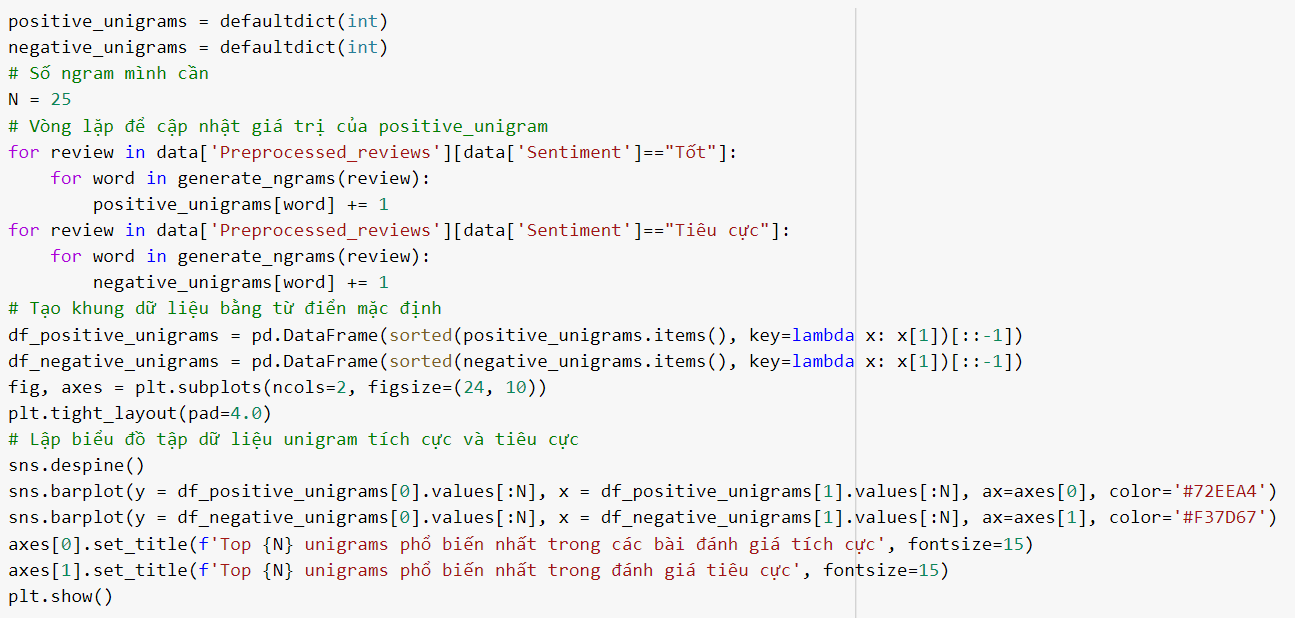
Biểu đồ cột thể hiện tỉ lệ rating review.



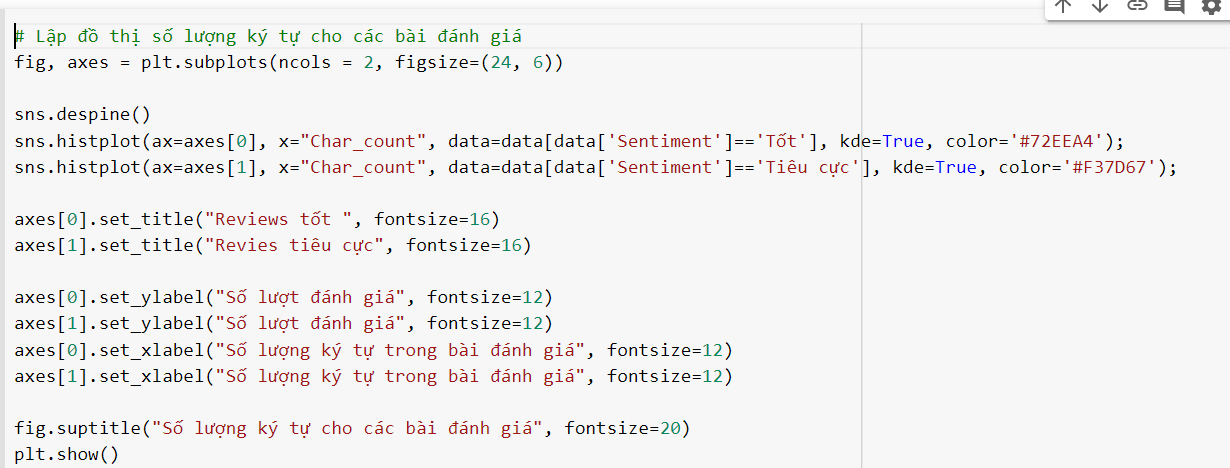
Áp dụng data processing



Vẽ biểu đồ Unigrams



Phân tích dữ liệu thăm dò

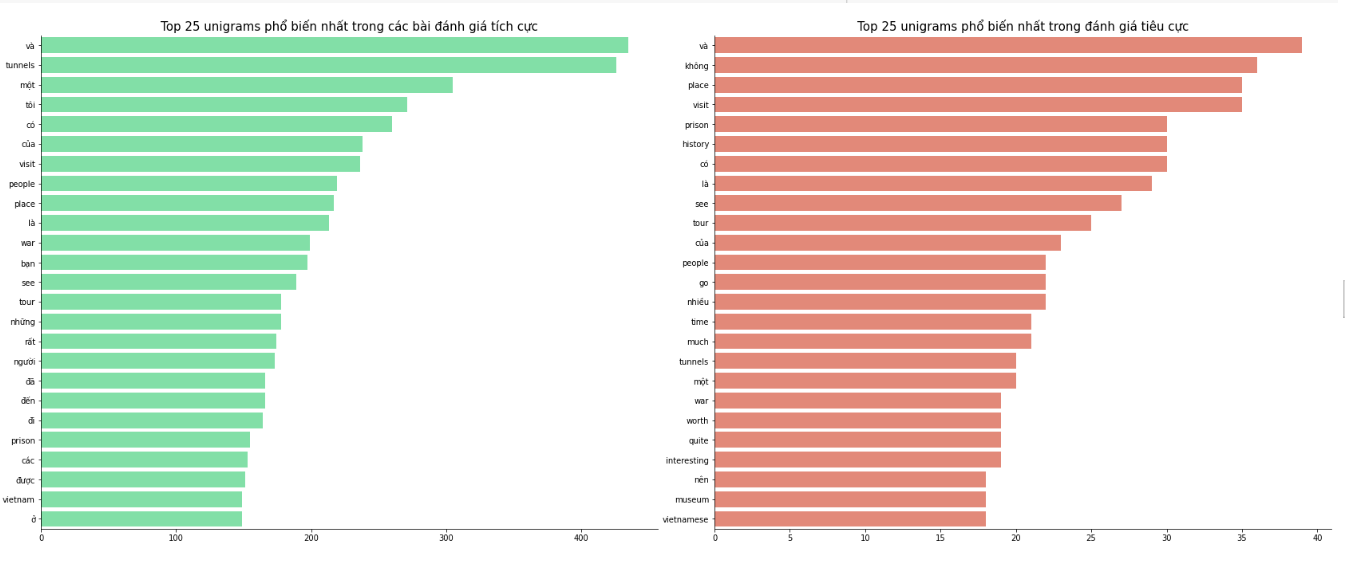


Hàm phân tích và vẽ biểu đồ phân biệt Review tích cực và Review tiêu cực.

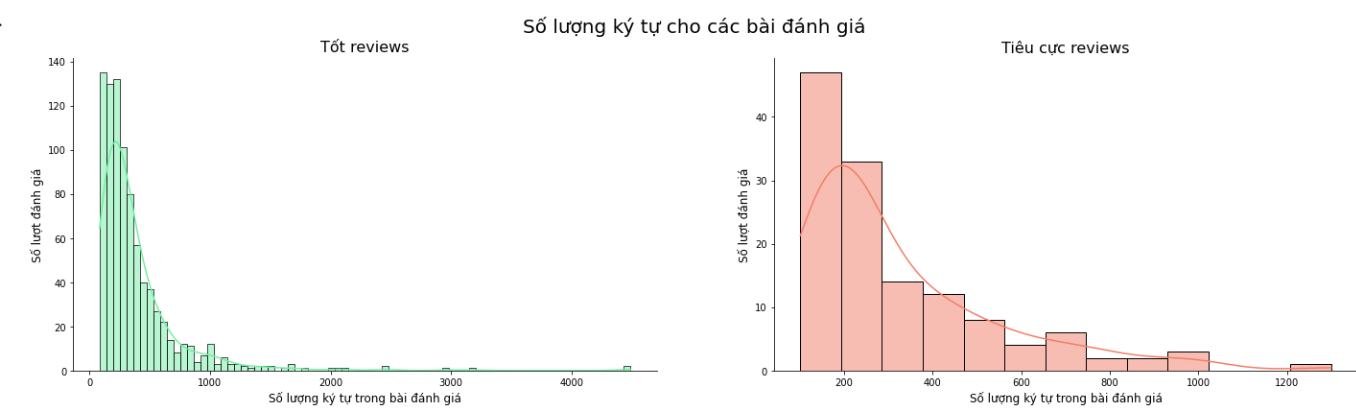


3.Kết quả.

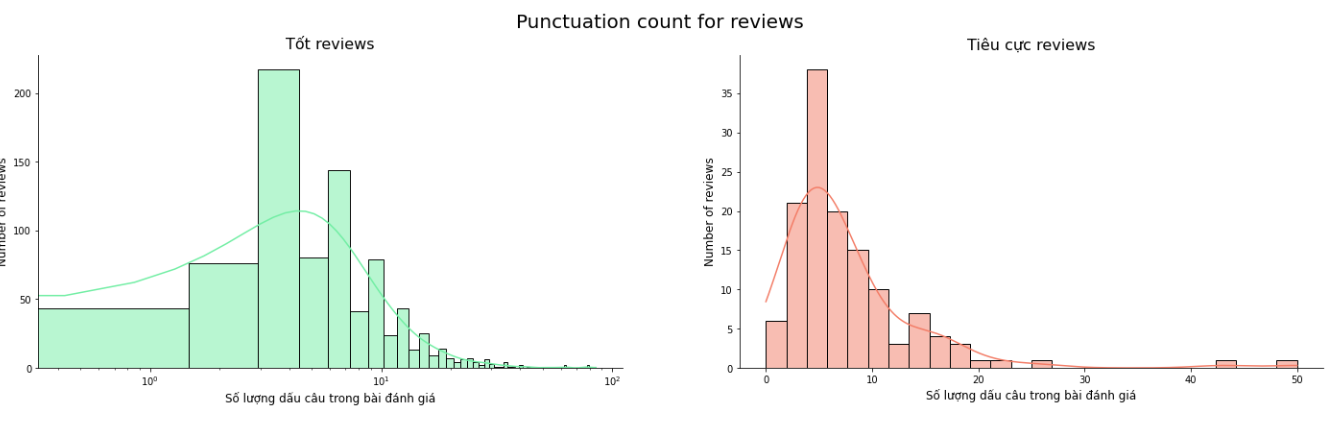
Biểu đồ Unigram thể hiện số từ xuất hiện nhiều nhất



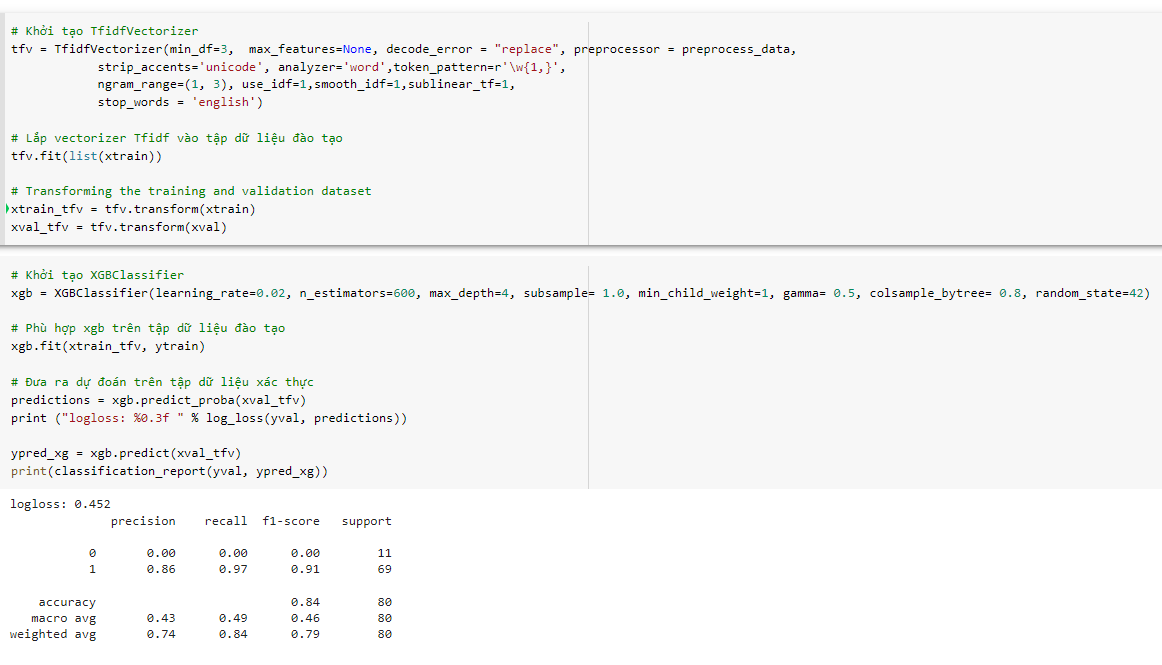
Biểu đồ phân tích số liệu thăm dò



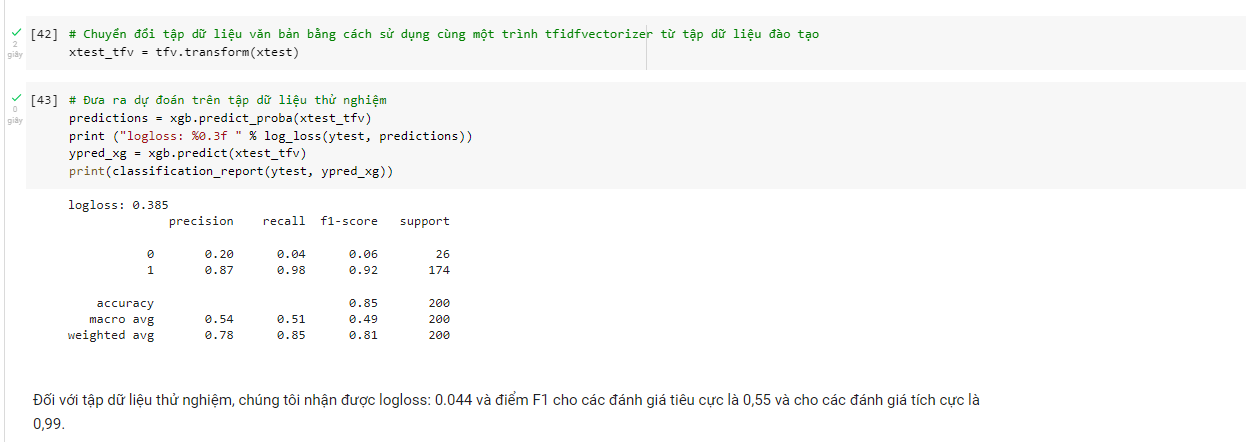
Biểu đồ phân biệt Review tích cực và Review tiêu cực.



Mô hình training



Mô hình đánh giá



4. Kết quả mô hình

-Đưa ra các biểu đồ phân biệt được reviews tích cực hay tiêu cực, đưa ra các biểu đồ khác liên quan phục vụ cho việc nhận xét địa điểm được review.

-Các biểu đồ tổng hợp khác thể hiện các mặt khía cạnh của những bài review.