**GUIDELINE LABEL DỮ LIỆU CHO BÀI TOÁN PHÂN TÍCH, DỰ BÁO CÁC PHẢN HỒI VỀ DU LỊCH QUY NHƠN**

1. **Tổng quan bài toán:**

* Input: Các review, bình luận về du lịch ở Quy Nhơn.
* Output: Có 2 loại output chính: Khía cạnh và mức độ biểu cảm.
  + **Khía cạnh**: được đề cập trong bình luận được chia thành 6 loại sau:
    - Dịch vụ vui chơi giải trí: (các trò chơi, tour du lịch, view, cảnh đẹp).
* Chất lượng của hướng dẫn viên du lịch?
* Cảnh quan của các địa điểm thăm quan như thế nào ?
* Chất lượng của các loại hình giải trí như thế nào (công viên, bảo tàng, tắm biển, massage, bar, sở thú)
* Mức độ an toàn của các dịch vụ vui chơi (bảo vệ, cứu hộ, ..)
  + - Dịch vụ lưu trú (khách sạn, nhà nghỉ).
* Địa điểm lưu trú như thế nào? (gần quốc lộ, biển, bến xe, …)
* Thái độ phục vụ của lễ tân, bảo vệ?
* Giá cả hợp lý không?
* Trang thiết bị, nội thất có đầy đủ không (máy lạnh, camera, …)? Chất lượng như thế nào?

* + - Hệ thống nhà hàng phục vụ khách du lịch: (nhà hàng, thái độ nhân viên).
* Nếu người đọc biết reviews nhắc đến nhà hàng (keywords: )
* Thái độ phục vụ của nhân viên và thái độ của bảo vệ?
* Không gian nhà hàng như thế nào?
* Trang thiết bị của nhà hàng như thế nào?
* Chất lượng thực phẩm ở nhà hàng tốt không? Số lượng món ăn có đa dạng không? Giá cả hợp lý hay không?
* Vị trí nhà hàng có thuận tiện di chuyển tới các địa điểm khác hay không (gần biển, khách sạn, trung tâm mua sắm, gần đường lớn, ….)
  + - Dịch vụ ăn uống (đồ ăn, thức uống).
* Chất lượng món ăn? Ngon hay dở?
* Giá cả thế nào? Đồ ăn có phù hợp với giá tiền?
* Ăn không hết có được mang về không?
* Đồ ăn có phong phú không?
* Đồ ăn có chất lượng hay không?
* Thái độ của chủ quán, nhân viên, phục vụ có tốt không?
* đồ ăn bản địa?

* + - Dịch vụ di chuyển :
* Giá thuê xe có hợp lý không?
* Xe thuê có chất lượng không?
* Địa điểm thuê xe xa hay gần trung tâm?
* Thái độ người lái xe ( giới thiệu, trò chuyện, am hiểu du lịch) ?
* Trang thiết bị cho xe có chất lượng không, xe mới hay cũ?
* Những thứ liên quan về tiện ích đi lại của khách du lịch?
* Dịch vụ di chuyển có bị delay hay không?
* Thời điểm khởi hành có hợp lý không?
* Di chuyển có đảm bảo an toàn không?
* Trang thiết bị bảo hộ không đầy đủ không ?
  + - Dịch vụ mua sắm (hàng hóa).
* Liên quan tới hàng hóa (quần áo, quà lưu niệm, hoa…).
* Chất lượng hàng hóa như thế nào?
* Thái độ nhân viên bán hàng.
* Các loại mặt hàng có đa dạng không?
* Giá cả như thế nào?
* Trung tâm mua sắm có dễ dàng tìm thấy và di chuyển k ?
* Một số lưu ý về trường hợp nhập nhằng:
  + Nếu review về nhà hàng nhưng đồng thời có bình luận về ăn uống (Nhà hàng X bày trí rất đẹp, thức ăn rất ngon, …) thì gán đồng thời cả khía cạnh Hệ thống nhà hàng + Dịch vụ ăn uống.
  + Nếu chỉ nhắc đến nhà hàng (không gian, thái độ nhân viên, giá cả), không bình luận thêm về ăn uống thì chỉ gán mỗi khía cạnh Hệ thống nhà hàng.
  + Tương tự, đối với các trường hợp khác, nếu xuất hiện nhiều khía cạnh cùng lúc thì gán nhiều khía cạnh, các trường hợp không rõ ràng hoặc một vài khía cạnh xuất hiện rất ít thì chỉ gán khía cạnh nổi bật và là nội dung chính của review.
  + **Mức độ biểu cảm**: 1 -> 5 lần lượt ứng với: Rất không hài lòng / Không hài lòng / Bình thường / Hài lòng / Rất hài lòng.

1. **Hướng dẫn gán nhãn:**

* Mỗi đoạn review có thể gồm một hoặc nhiều câu, độ dài khác nhau.
* Mỗi đoạn review có thể có nhiều khía cạnh khác nhau cùng xuất hiện, ứng với mỗi khía cạnh xuất hiện thì gán mức độ biểu cảm tương ứng với khía cạnh đó.
* Data được lưu dưới dạng bảng (file csv), gồm n + 1 hàng (n ứng với số đoạn review ở file đó) và 8 cột lần lượt:
  + ID: số thứ tự tăng dần từ 1.
  + Review: đoạn bình luận / review cần gán nhãn.
  + DV vui chơi giải trí.
  + DV lưu trú.
  + Hệ thống nhà hàng.
  + DV ăn uống.
  + DV di chuyển.
  + Dịch vụ mua sắm.
* Mỗi hàng ứng với một đoạn review, nếu khía cạnh nào xuất hiện thì nhập thêm giá trị mức độ biểu cảm vào cột tương ứng (1-> 5), các khía cạnh còn lại để trống hoặc nhập 0 (ứng với không có).

1. **Một số ví dụ:**

[Example labelling](https://docs.google.com/spreadsheets/d/1dD1wsMN7LWozjMScp0b8J4_IM2BKknopnLZoBqf704k/edit#gid=1210691631)