

**TRƯỜNG ĐẠI HỌC KINH TẾ - ĐẠI HỌC ĐÀ NẴNG**  
**KHOA THƯƠNG MẠI ĐIỆN TỬ**



**BÀI TẬP NHÓM**

**MÔN: ĐỀ ÁN THỰC HÀNH 1**

**Đề tài:**

**PHÂN TÍCH SỞ THÍCH ĂN UỐNG CỦA DU KHÁCH DỰA TRÊN  
ĐÁNH GIÁ NHÀ HÀNG: NGHIÊN CỨU TÌNH HUỐNG TẠI ĐÀ NẴNG**

<b>Giảng viên hướng dẫn</b>	<b>:</b>	GV.Lê Diên Tuấn GV.Châu Thị Dạ Hương
<b>Lớp học phần</b>	<b>:</b>	48K29.1
<b>Nhóm sinh viên thực hiện</b>	<b>:</b>	Nhóm 2 Tiêu Thanh Mai Anh Đinh Trần Công Nghĩa Nguyễn Thị Nha

*Thành phố Đà Nẵng, năm 2025*

## MỤC LỤC

CHƯƠNG 1. GIỚI THIỆU ĐỀ TÀI .....	6
1.1. Lĩnh vực nghiên cứu.....	6
1.2. Mục tiêu nghiên cứu .....	6
1.3. Câu hỏi nghiên cứu .....	6
1.4. Các bên liên quan.....	7
1.5. Quy trình nghiên cứu .....	8
CHƯƠNG 2. CƠ SỞ LÝ THUYẾT.....	9
2.1. Giới thiệu về Các Mô hình Ngôn ngữ Lớn (LLMs) .....	9
2.2. Phân tích Cảm xúc (Sentiment Analysis) và Vai trò của LLMs .....	9
2.3. Phân tích Cảm xúc Dựa trên Khía cạnh (ABSA) và tiềm năng của LLMs.....	10
2.4. Ưu điểm của LLMs trong Sentiment Analysis và ABSA .....	11
2.5. Thách thức và Giải pháp khi Sử dụng LLMs .....	11
2.6. Ứng dụng của DeepSeek-R1-Distill-Llama-70B-free.....	12
2.7. Kết luận.....	12
CHƯƠNG 3. THU THẬP DỮ LIỆU .....	12
3.1. Xác định nguồn dữ liệu.....	12
3.1.1. Bảng restaurants.....	13
3.1.2. Bảng reviews .....	14
3.1.3. Bảng features_type .....	15
3.1.4. Bảng feature_items .....	16
3.2. Quy trình cào dữ liệu .....	18
3.2.1. Thu thập dữ liệu.....	18
3.2.2. Kết quả sau khi thu thập dữ liệu .....	20
CHƯƠNG 4. TIỀN XỬ LÝ DỮ LIỆU .....	21
4.1. Xử lý giá trị thiếu.....	21

4.2. Rời rạc hoá dữ liệu.....	23
4.3. Biến đổi dữ liệu .....	25
CHƯƠNG 5. XÂY DỰNG MÔ HÌNH.....	28
5.1. Mô tả bài toán .....	28
5.2. Thu thập dữ liệu.....	30
5.3. Xử lý dữ liệu .....	31
5.4. Xây dựng mô hình .....	32
5.4.1. Lựa chọn mô hình.....	32
5.4.2. Thiết kế Prompt và Định dạng Input – Output .....	33
5.5. Ứng dụng mô hình vào dữ liệu thực tế .....	38
CHƯƠNG 6. PHÂN TÍCH DASHBOARD.....	39
6.1. Data Storytelling với dashboard .....	39
6.1.1. Du khách từ quốc gia nào đánh giá nhiều nhất?.....	40
6.1.2. Du khách thích ăn món gì vào các bữa ăn trong ngày và đâu là loại hình nhà hàng phổ biến? .....	42
6.1.3. Phân bố địa lý và chất lượng dịch vụ nhà hàng tại các quận ở Đà Nẵng như thế nào? Khu vực nào có tiềm năng phát triển nhất?.....	47
6.1.4. Phân bố đánh giá theo quốc tịch tại nhà hàng phân khúc giá cao .....	50
6.1.5. Sự hỗ trợ cho các bên liên quan.....	51
6.1.6. Từ khóa cảm xúc của du khách tại các nhà hàng Đà Nẵng .....	52
7. Các khía cạnh được du khách quan tâm nhất theo cảm xúc .....	55
8. Xu hướng cảm xúc của du khách qua thời gian tại Đà Nẵng .....	56
10. Đánh giá chi tiết các khía cạnh trải nghiệm tại nhà hàng Đà Nẵng theo cảm xúc	
58	
CHƯƠNG 7. KẾT QUẢ NGHIÊN CỨU .....	60
7.1. Khám phá xu hướng ẩm thực và định hướng phát triển nhà hàng tại Đà Nẵng	60
7.1.1. Khách hàng đến từ quốc gia nào đánh giá nhiều nhất? .....	60

7.1.2. Những món ăn nào được ưa chuộng? .....	60
7.1.3. Du khách thường ăn gì vào các bữa trong ngày? .....	61
7.1.4. Nhóm khách hàng tiềm năng nhất là ai? .....	61
7.1.5. Điểm mạnh và điểm yếu trong đánh giá chất lượng dịch vụ? .....	61
7.1.6. Cảm xúc khách hàng thay đổi như thế nào theo thời gian? .....	61
7.1.7. Khu vực nào tập trung nhiều nhà hàng nhất? .....	62
7.1.8. Loại hình nhà hàng phổ biến và có tiềm năng phát triển? .....	62
7.1.9. Nhà hàng điểm cao thường có đặc điểm gì nổi bật? .....	62
7.1.10. Mức giá phổ biến của các nhà hàng là gì? .....	62
7.2. Hỗ trợ các bên liên quan .....	62
7.2.1. Chủ sở hữu/Quản lý nhà hàng: .....	62
7.2.2. Nhà khởi nghiệp F&B: .....	63
CHƯƠNG 8. TỔNG KẾT DỰ ÁN .....	63
CHƯƠNG 9. TÀI LIỆU THAM KHẢO .....	65

## DANH MỤC HÌNH ẢNH

Hình 1: Quy trình nghiên cứu – framework .....	8
Hình 2: Quy trình thu thập dữ liệu nhà hàng và đánh giá .....	19
Hình 3: Quy trình thu thập dữ liệu loại tiện ích và tiện ích cụ thể của nhà hàng .....	19
Hình 4: Kết quả sau khi thu thập dữ liệu nhà hàng .....	20
Hình 5: Kết quả sau khi thu thập dữ liệu đánh giá .....	20
Hình 6: Kết quả sau khi thu thập loại tiện ích .....	21
Hình 7: Kết quả sau khi thu thập tiện ích cụ thể .....	21
Hình 8: Kết quả kiểm tra dữ liệu thiếu trong bảng restaurants .....	21
Hình 9: Kết quả kiểm tra dữ liệu thiếu trong bảng reviews .....	22
Hình 10: Biểu đồ heatmap ma trận tương quan giữa rating với cột <i>service_rating</i> , <i>food_rating</i> , <i>atmosphere_rating</i> .....	22
Hình 11: Dữ liệu bảng <i>features_type</i> trước khi xử lý .....	23
Hình 12: Dữ liệu bảng <i>feature_items</i> trước khi xử lý .....	23

Hình 13: Dữ liệu bảng features .....	24
Hình 14: Dữ liệu bảng restaurant_features.....	24
Hình 15: Sơ đồ quan hệ thực thể giữa các bảng trong SQL server .....	25
Hình 16: Kết quả sau khi xử lý và nhập dữ liệu thành công .....	28
Hình 17: Kết quả sau khi tách quận từ address .....	28
Hình 18: Ví dụ về bài toán trích xuất Aspect, Opinion và Sentiment từ Review .....	29
Hình 19: Dữ liệu trước khi xử lý .....	31
Hình 20: Kết quả triển khai mô hình trích xuất bộ ba cảm xúc (Aspect - Opinion - Sentiment).....	38
Hình 21: Dashboard Overview .....	39
Hình 22: Dashboard Review Analysis .....	39
Hình 23: Biểu đồ Review Distribution by Nation.....	40
Hình 24: Biểu đồ Review Distribution by Meal type.....	42
Hình 25: Biểu đồ Popular Dishes Mentioned in Reviews theo loại bữa ăn lần lượt là Dinner và Lunch .....	43
Hình 26: Biểu đồ Review Distribution by Meal type.....	44
Hình 27: Biểu đồ Popular Dishes Mentioned in Reviews của khách Hàn Quốc .....	45
Hình 28: Biểu đồ Review Distribution by Restaurant type.....	46
Hình 29: Biểu đồ Number of Restaurant and Average Rating by District.....	47
Hình 30: Bản đồ phân bố địa lý của các nhà hàng tại Đà Nẵng.....	48
Hình 31: Biểu đồ Review Distribution by Nation lần lượt ở quận Hải Châu, Sơn Trà, Ngũ Hành Sơn. ....	49
Hình 32: Biểu đồ kết hợp Phân bố đánh giá theo quốc tịch tại nhà hàng phân khúc giá cao.....	50
Hình 33: Competitive Radar Chart .....	51
Hình 34: Wordcloud for sentiment - Negative .....	52
Hình 35: Wordcloud for sentiment - Neutral .....	53
Hình 36: Wordcloud for sentiment - Positive .....	54
Hình 37: Top 10 aspects by sentiment .....	55
Hình 38: Sentiment Trends Over Time .....	56
Hình 39: Sentiment by Meal Type .....	57

Hình 40: Bảng khía cạnh theo cảm xúc tiêu cực .....	58
Hình 41: Bảng khía cạnh theo cảm xúc trung tính .....	59
Hình 42: Bảng khía cạnh theo cảm xúc tích cực .....	59

## **DANH MỤC BẢNG BIỂU**

Bảng 2: Mô tả dữ liệu bảng Restaurants .....	14
Bảng 3: Mô tả dữ liệu bảng reviews.....	15
Bảng 4: Mô tả dữ liệu bảng features_type.....	15
Bảng 5: Mô tả dữ liệu bảng feature_items .....	18

# CHƯƠNG 1. GIỚI THIỆU ĐỀ TÀI

## 1.1. Lĩnh vực nghiên cứu

Quản trị dịch vụ nhà hàng.

## 1.2. Mục tiêu nghiên cứu

Đà Nẵng là một trong những điểm đến du lịch hàng đầu Việt Nam, thu hút hàng triệu du khách mỗi năm nhờ cảnh quan biển tuyệt đẹp, nền ẩm thực phong phú và hệ thống dịch vụ du lịch phát triển. Vì thế nơi đây trở thành địa điểm nóng cho các nhà đầu tư có ý định khởi nghiệp trong lĩnh vực nhà hàng và khách sạn trong thời gian gần đây. Tuy nhiên, việc nắm bắt các hành vi và sở thích ăn uống của khách hàng lại là bài toán khó trong bối cảnh cạnh tranh ngày càng gay gắt trong ngành du lịch, nếu không thể nắm bắt được xu hướng nhu cầu thực tế sẽ rất khó để một doanh nghiệp có thể tồn tại.

Trong khi đó các nghiên cứu hiện tại về hành vi tiêu dùng ẩm thực tại Đà Nẵng chủ yếu dựa trên các phương pháp thu thập dữ liệu truyền thống như khảo sát hay bảng hỏi tốn kém, mất thời gian và chưa phản ánh đầy đủ trải nghiệm thực tế gây nhiều khó khăn cho những nhà đầu tư muốn kinh doanh ở đây. Trong nghiên cứu lần này chúng tôi tập trung phân tích các xu hướng, sở thích ẩm thực, mức độ hài lòng và các yếu tố quan trọng ảnh hưởng đến quyết định ăn uống của du khách tại các nhà hàng ở Đà Nẵng. Với việc áp dụng công nghệ phân tích dữ liệu hiện đại như mô hình ngôn ngữ lớn (LLMs - Large Language Models) cùng nguồn dữ liệu đáng tin cậy trên nền tảng Google Reviews chúng tôi tin rằng bài nghiên cứu này sẽ mang lại những thông tin có ích cho các nhà quản lý nhà hàng ở Đà Nẵng.

Kết quả nghiên cứu không chỉ nhận diện được các yếu tố ảnh hưởng đến trải nghiệm khách hàng mà còn đề xuất các giải pháp tối ưu hóa thực đơn, cải thiện dịch vụ và xây dựng chiến lược kinh doanh phù hợp với từng nhóm khách hàng, từ đó góp phần nâng cao chất lượng dịch vụ và thúc đẩy sự phát triển bền vững của ngành du lịch nhà hàng tại Đà Nẵng.

## 1.3. Câu hỏi nghiên cứu

- Sở thích ẩm thực của khách hàng?
  - Những món ăn nào được du khách từ các quốc gia khác nhau ưa chuộng tại các nhà hàng ở Đà Nẵng?
  - Du khách thích ăn món gì vào các bữa ăn khác nhau trong ngày?
- Đánh giá chất lượng dịch vụ và trải nghiệm khách hàng:
  - Phân tích điểm mạnh và điểm yếu của các nhà hàng.
  - Xác định những yếu tố khách hàng đánh giá cao.

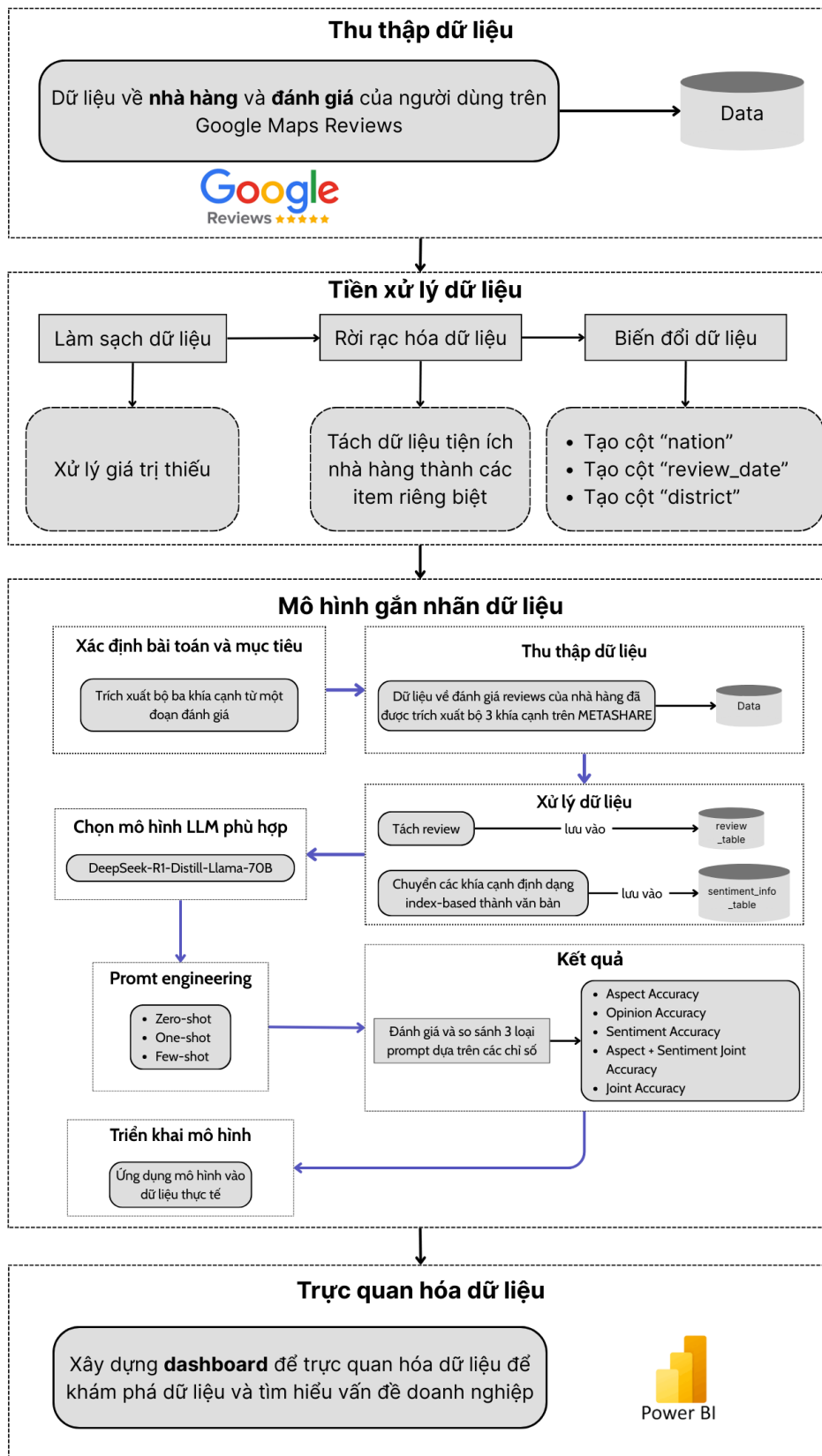
- Nhận diện những điểm cần cải thiện.
- Khu vực nào ở Đà Nẵng thường tập trung nhiều nhà hàng?
- Loại hình nhà hàng nào phổ biến, có tiềm năng phát triển ở Đà Nẵng (món âu, món á, món chay,...)?
- Những nhà hàng có điểm đánh giá trung bình cao thường sẽ có những đặc điểm nổi trội gì? (món ăn, tiện ích phục vụ, không gian, ...)
- Mức giá phổ biến của nhà hàng thường dao động ở mức nào?
- Nhóm khách hàng nào có tiềm năng khai thác nhất? (quốc tế, địa phương,...)

#### **1.4. Các bên liên quan**

- Chủ sở hữu/Quản lý nhà hàng:
  - Giúp chủ nhà hàng nhận diện được lợi thế cạnh tranh cũng như những vấn đề cần cải thiện.
  - Nắm bắt được những điểm nổi bật được khách hàng đánh giá cao để tập trung nguồn lực duy trì lợi thế cạnh tranh.
  - Xác định được đặc điểm, nhóm khách hàng thường đến sử dụng dịch vụ để điều chỉnh thực đơn món ăn và chiến lược tiếp thị nhắm đúng đối tượng khách hàng hướng đến.
  - Cải thiện dịch vụ, tối ưu hóa chiến lược kinh doanh, tăng sự hài lòng của khách hàng.
  - Nắm rõ các xu hướng ẩm thực, sở thích và hành vi ăn uống của khách hàng nhằm điều chỉnh chiến lược kinh doanh phù hợp theo từng thời điểm.
- Nhà khởi nghiệp lĩnh vực F&B
  - Xác định mô hình nhà hàng có nhu cầu cao, có tiềm năng phát triển ở Đà Nẵng.
  - Hỗ trợ chọn vị trí chiến lược đặt nhà hàng phù hợp.
  - Định hướng xây dựng mô hình nhà hàng theo phong cách nào, cần có những tiện ích gì?
  - Đối tượng khách hàng tiềm năng ở đây có những đặc điểm như thế nào?
  - Hỗ trợ xây dựng thực đơn phù hợp với từng nhóm khách hàng, định giá sản phẩm/dịch vụ phù hợp với thị trường.



## 1.5. Quy trình nghiên cứu



Hình 1: Quy trình nghiên cứu – framework

## CHƯƠNG 2. CƠ SỞ LÝ THUYẾT

### 2.1. Giới thiệu về Các Mô hình Ngôn ngữ Lớn (LLMs)

#### Định nghĩa và Kiến trúc

Các Mô hình Ngôn ngữ Lớn (LLMs) là các mô hình học sâu dựa trên kiến trúc mạng nơ-ron, thường sử dụng cấu trúc Transformer, được huấn luyện trên khối lượng dữ liệu văn bản khổng lồ để xử lý và tạo ra văn bản tương tự con người [1]. Ví dụ, mô hình DeepSeek-R1-Distill-Llama-70B-free, với 70 tỷ tham số, tận dụng kỹ thuật chưng cất tri thức (knowledge distillation) từ các mô hình lớn hơn để tối ưu hóa hiệu suất mà vẫn duy trì khả năng tính toán hiệu quả [2].

Theo Törnberg (2023), LLMs vượt trội trong các nhiệm vụ phân tích văn bản như phân loại, chú thích, và nhận diện ý nghĩa ngữ cảnh (ví dụ: mỉa mai, châm biếm) nhờ quy mô huấn luyện lớn và dữ liệu đa dạng từ internet, sách, và các nguồn khác. Điều này cho phép chúng học được các biểu diễn ngữ nghĩa phong phú, vượt xa các phương pháp truyền thống [3].

#### So sánh với Phương pháp Truyền thống

Không giống như phân tích nội dung thủ công hoặc các kỹ thuật xử lý ngôn ngữ tự nhiên (NLP) dựa trên từ điển, LLMs không yêu cầu tiền xử lý phức tạp như tokenization (phân đoạn từ) hay lemmatization (chuẩn hóa từ về dạng gốc) [3]. Chúng có thể xử lý văn bản thô, đa ngôn ngữ mà không cần chuẩn bị dữ liệu đặc biệt, khiến chúng trở thành công cụ lý tưởng cho Sentiment Analysis và ABSA.

### 2.2. Phân tích Cảm xúc (Sentiment Analysis) và Vai trò của LLMs

#### Khái niệm Sentiment Analysis

Phân tích cảm xúc là quá trình xác định và phân loại cảm xúc hoặc thái độ trong văn bản, thường được chia thành các nhãn như tích cực, tiêu cực hoặc trung lập [4]. Đây là một nhiệm vụ quan trọng trong nghiên cứu xã hội, tiếp thị, và trải nghiệm khách hàng, nhưng các phương pháp truyền thống thường gặp khó khăn với các sắc thái ngữ nghĩa phức tạp như ngữ cảnh, mỉa mai, hoặc yếu tố văn hóa [3].

#### Vai trò của LLMs

LLMs cải thiện đáng kể hiệu quả của Sentiment Analysis nhờ khả năng hiểu ngữ cảnh sâu sắc, được tích hợp từ dữ liệu huấn luyện đa dạng [5]. Ví dụ, mô hình DeepSeek-R1-Distill-Llama-70B-free phân tích cảm xúc không chỉ dựa trên từ ngữ mà còn dựa trên cấu trúc câu, ngữ cảnh, và ý định của tác giả. Điều này giúp mô hình nhận diện chính xác hơn các sắc thái cảm xúc mà các phương pháp dựa trên từ điển hoặc máy học truyền thống không thể xử lý tốt [3].

### **Ví dụ Minh họa**

Một câu như "Dịch vụ này thật tuyệt vời, nhưng tôi đùa đấy!" có thể bị các phương pháp truyền thống phân loại sai là tích cực. Tuy nhiên, LLMs, với khả năng suy luận ngữ cảnh, có thể nhận diện sự mỉa mai và phân loại đúng thành tiêu cực [5].

## **2.3. Phân tích Cảm xúc Dựa trên Khía cạnh (ABSA) và tiềm năng của LLMs**

### **Khái niệm ABSA**

ABSA là một nhánh mở rộng của Sentiment Analysis, tập trung vào việc xác định cảm xúc liên quan đến các khía cạnh cụ thể của một đối tượng trong văn bản (ví dụ: chất lượng sản phẩm, dịch vụ khách hàng) [6]. Nhiệm vụ này bao gồm hai bước chính: nhận diện khía cạnh (aspect extraction) và phân loại cảm xúc theo khía cạnh (aspect-level sentiment classification).

### **Hạn chế của Phương pháp Truyền thống**

Các phương pháp máy học giám sát truyền thống đòi hỏi dữ liệu huấn luyện được gán nhãn thủ công, gây tốn kém thời gian và không linh hoạt khi áp dụng trên các lĩnh vực hoặc ngôn ngữ khác nhau [6]. Ngược lại, LLMs mang lại giải pháp đột phá nhờ khả năng học zero-shot (không cần dữ liệu huấn luyện) và few-shot (ít dữ liệu huấn luyện) [7].

### **Tiềm năng của LLMs trong ABSA**

Mô hình DeepSeek-R1-Distill-Llama-70B-free, với khả năng xử lý văn bản dài và suy luận ngữ cảnh, đặc biệt phù hợp cho ABSA. Nó có thể nhận diện các khía cạnh trong văn bản và đánh giá cảm xúc liên quan mà không cần dữ liệu gán nhãn lớn, mang lại độ chính xác cao trên các tập dữ liệu đa dạng [3]. Ví dụ, trong câu "Máy tính này có

màn hình đẹp nhưng pin kém", mô hình có thể tách biệt cảm xúc tích cực về "màn hình" và tiêu cực về "pin".

## 2.4. Ưu điểm của LLMs trong Sentiment Analysis và ABSA

LLMs mang lại nhiều lợi thế vượt trội so với các phương pháp truyền thống, bao gồm:

- **Tính linh hoạt:** Có thể áp dụng trên nhiều loại văn bản (bài viết, bình luận, đánh giá) và ngôn ngữ mà không cần tiền xử lý phức tạp [3].
- **Khả năng hiểu ngữ cảnh:** Nhận diện các sắc thái cảm xúc phức tạp như mỉa mai, châm biếm, hoặc ẩn ý nhờ dữ liệu huấn luyện phong phú [5].
- **Hiệu quả chi phí và thời gian:** Giảm đáng kể chi phí và thời gian so với phân tích thủ công hoặc xây dựng mô hình máy học từ đầu [3].
- **Khả năng mở rộng:** Xử lý khối lượng dữ liệu lớn, phù hợp cho nghiên cứu quy mô lớn hoặc phân tích thời gian thực [6].

## 2.5. Thách thức và Giải pháp khi Sử dụng LLMs

### Thách thức

- Dù mạnh mẽ, LLMs vẫn đối mặt với một số thách thức:
- **Độ chính xác và không thiên vị:** Kết quả có thể bị ảnh hưởng bởi thiên kiến trong dữ liệu huấn luyện [3].
- **Vấn đề đạo đức:** Sử dụng dữ liệu công khai mà không có sự đồng ý rõ ràng của tác giả gây tranh cãi [8].

### Giải pháp

- **Xác thực kết quả:** Sử dụng các phương pháp như Krippendorff's alpha để đo mức độ tương đồng giữa kết quả của LLM và dữ liệu tham chiếu [3].
- **Tối ưu hóa lời nhắc (Prompt Engineering):** Thử nghiệm lặp lại để thiết kế lời nhắc phù hợp với nhiệm vụ phân tích [3].
- **Tuân thủ đạo đức:** Áp dụng các hướng dẫn như của British Sociological Association (2017) để đảm bảo quyền riêng tư và ý định của người tạo nội dung được tôn trọng [9].

## 2.6. Ứng dụng của DeepSeek-R1-Distill-Llama-70B-free

### Kiến trúc và Hiệu suất

Mô hình DeepSeek-R1-Distill-Llama-70B-free, phát triển bởi DeepSeek AI, dựa trên kiến trúc Llama và được tối ưu hóa với 70 tỷ tham số. Kỹ thuật chưng cất tri thức giúp mô hình duy trì hiệu suất cao trong khi giảm chi phí tính toán [2]. Các kết quả benchmark bao gồm:

- MATH-500: 94.5% pass@1 [10].
- AIME 2024: 86.7 cons@64 [Hugging Face].
- GPQA Diamond: 65.2% [10].
- LiveCodeBench: 57.5% [10].
- CodeForces: Rating 1633 [10].

### Ứng dụng trong Sentiment Analysis và ABSA

Với tốc độ 120.2 token/giây, thời gian đến token đầu tiên (TTFT) 0.42 giây, và chi phí \$0.60/triệu token [11], mô hình này lý tưởng cho các nhiệm vụ phân tích văn bản phức tạp. Nó vượt trội trong việc nhận diện cảm xúc tổng thể và chi tiết theo khía cạnh, đặc biệt khi kết hợp với lời nhắc được thiết kế cẩn thận.

## 2.7. Kết luận

LLMs, đặc biệt là DeepSeek-R1-Distill-Llama-70B-free, mang lại bước tiến lớn trong Sentiment Analysis và ABSA nhờ tính linh hoạt, khả năng hiểu ngữ cảnh, và hiệu quả chi phí. Tuy nhiên, để đảm bảo độ tin cậy, cần chú trọng xác thực kết quả, tối ưu hóa lời nhắc, và tuân thủ đạo đức. Những yếu tố này giúp LLMs trở thành công cụ mạnh mẽ và đáng tin cậy cho nghiên cứu phân tích văn bản hiện đại.

## CHƯƠNG 3. THU THẬP DỮ LIỆU

### 3.1. Xác định nguồn dữ liệu

Dữ liệu được thu thập từ nền tảng Google Reviews bao gồm các thông tin chi tiết của một nhà hàng tại Đà Nẵng cùng với các đánh giá của khách hàng về sản phẩm/dịch vụ của nhà hàng đó. Dữ liệu sau khi thu thập sẽ được tổ chức thành các bảng với cấu trúc và mô tả chi tiết như sau:

### 3.1.1. Bảng restaurants

- Bảng restaurants chứa thông tin cơ bản của từng nhà hàng đang hoạt động tại Đà Nẵng. Đây là bảng trung tâm, kết nối với các bảng còn lại thông qua khóa chính là restaurant\_id. Gồm 12 cột như sau:

STT	Fields	Data types	Definition	Value
1	Restaurant_id	Int	Mã định danh nhà hàng	
2	Restaurant_name	Text	Tên nhà hàng	
3	url	Text	Đường dẫn cụ thể để lấy thông tin của từng nhà hàng	
4	Restaurant_type	Text	Loại nhà hàng	Vietnamese restaurant, Restaurant, Fine dining restaurant, Steak house, Indian restaurant,...
5	Rating_average	Float	Điểm đánh giá trung bình	
6	Num_of_reviews	Int	Tổng số lượt đánh giá	
7	Price_level	Text	Mức giá	₫100–200K, ₫1,000,000+,....
8	Address	Text	Địa chỉ	
9	Latitude	Float	Vĩ độ của nhà hàng trên bản đồ	
10	Longitude	Float	Kinh độ của nhà hàng trên bản đồ	

<b>11</b>	Phone	Text	Số điện thoại	
<b>12</b>	Created_at	Date	Ngày thu thập dữ liệu	

*Bảng 1: Mô tả dữ liệu bảng Restaurants*

### 3.1.2. Bảng reviews

- Bảng reviews chứa thông tin về các đánh giá của từng nhà hàng, bao gồm thông tin của người đánh giá, loại dịch vụ họ sử dụng và cảm nhận của họ khi trải nghiệm dịch vụ ăn uống tại nhà hàng. Mỗi bản ghi thể hiện cảm nhận của một người dùng đối với một nhà hàng cụ thể. Gồm 14 cột như sau:

STT	Fields	Data Type	Definition	Value
<b>1</b>	Review_id	Int	Mã định danh cho mỗi review	
<b>2</b>	Reviewer_name	Text	Tên người đánh giá	
<b>3</b>	Reviewer_info	Text	Thông tin về người đánh giá	Gồm số lượng đánh giá đã viết và ảnh đã đăng trên gg review của người đánh giá
<b>4</b>	Rating	Float	Điểm đánh giá	1 → 5
<b>5</b>	Review_time	Text	Thời điểm viết đánh giá	Khoảng cách thời gian từ lúc đánh giá được viết cho đến thời điểm dữ liệu được cào về
<b>6</b>	Review_Text	Text	Đánh giá của khách hàng về nhà hàng	

7	Service_rating	Float	Điểm đánh giá dịch vụ	
8	Food_rating	Float	Điểm đánh giá món ăn	
9	Atmosphere_rating	Float	Điểm đánh giá không gian	
10	Service_type	Text	Loại dịch vụ	Dine in
11	Meal_type	Text	Bữa ăn	Breakfast, Lunch, Dinner,...
12	Language	Text	Ngôn ngữ người đánh giá viết	
13	Created_at	Datetime	Ngày thu thập dữ liệu	
14	Restaurant_id	Int	Khoá ngoại tham chiếu đến mã định danh nhà hàng tương ứng	

*Bảng 2: Mô tả dữ liệu bảng reviews*

### 3.1.3. Bảng features\_type

- Bảng features\_type sẽ chứa các danh mục tiện ích dịch vụ mà mỗi nhà hàng có thể cung cấp cho khách hàng của mình.

STT	Fields	Data Type	Definition	Value
1	Restaurant_name	Text	Tên nhà hàng	
2	Feature_type	Text	Danh mục các tiện ích	

*Bảng 3: Mô tả dữ liệu bảng features\_type*



### 3.1.4. Bảng feature\_items

- Bảng feature\_items lưu trữ các thông tin chi tiết và cụ thể về từng tiện ích dịch vụ mà nhà hàng có trong bảng features\_type. Mỗi cột sẽ mô tả chi tiết các dịch vụ mà nhà hàng có thể cung cấp cho khách hàng.

STT	Fields	Data Type	Definition	Value
1	Restaurant_name	Text	Tên nhà hàng	
2	Offerings	Text	Các sản phẩm hoặc dịch vụ đặc biệt mà nhà hàng cung cấp	Đồ uống có cồn, thức ăn nhanh, thực phẩm hữu cơ, món chay,...
3	Dining_options	Text	Mô tả các tùy chọn ăn uống mà nhà hàng cung cấp	Bữa sáng, bữa brunch, bữa trưa, bữa tối, món tráng miệng,...
4	Crowd	Text	Mô tả nhóm đối tượng khách hàng mà nhà hàng hướng đến	Phù hợp cho gia đình, nhóm đông người, cộng đồng LGBT, khách du lịch, học sinh - sinh viên,...
5	Pets	Text	Chính sách của nhà hàng đối với thú cưng	Cho phép mang thú cưng vào, chỉ cho phép mang thú cưng vào khu vực ngoài trời,...
6	From_the_business	Text	Cam kết từ nhà hàng	Identifies as women-owned
7	Popular_for	Text	Loại bữa ăn mà nhà hàng thường xuyên phục vụ	Bữa sáng, bữa trưa, bữa tối,....

8	Highlights	Text	Các đặc điểm nổi bật của nhà hàng	Có loại bia, cà phê, món tráng miệng, trà, cocktail, rượu, ... được đánh giá là tốt nhất/nổi bật trên thị trường
9	Children	Text	Dịch vụ mà nhà hàng có thể cung cấp dành cho trẻ em	Loại nhà hàng phù hợp cho trẻ em, có thực đơn riêng cho trẻ em, có tổ chức sinh nhật cho trẻ,...
10	Parking	Text	Mô tả các dịch vụ về đậu đỗ xe	Đỗ xe miễn phí trên đường, có bãi đỗ xe miễn phí, khó tìm chỗ đỗ xe, ...
11	Service_options	Text	Mô tả các dịch vụ mà nhà hàng cung cấp	Dịch vụ mang đi, ăn tại chỗ, ngồi ngoài trời, giao hàng,....
12	Accessibility	Text	Dịch vụ cung cấp cho người khuyết tật	Có/Không bãi đỗ xe, nhà vệ sinh, lối đi riêng, chỗ ngồi dành cho người khuyết tật
13	Amenities	Text	Các tiện ích có sẵn	Có nhà vệ sinh, Wifi, quầy bar,...
14	Atmosphere	Text	Không gian nhà hàng	Môi trường thoải mái, ấm cúng, thời thượng, hiện đại, yên tĩnh, lãng mạn,...
15	Planning	Text	Chính sách đặt bàn ở nhà hàng	Có đặt bàn trước, không cho đặt bàn trước, thường xuyên chờ đợi do không đủ bàn, ...

16	Payments	Text	Phương thức thanh toán	Tiền mặt, thẻ tín dụng, thẻ ghi nợ, ...
----	----------	------	------------------------	---

*Bảng 4: Mô tả dữ liệu bảng feature\_items*

## 3.2. Quy trình cào dữ liệu

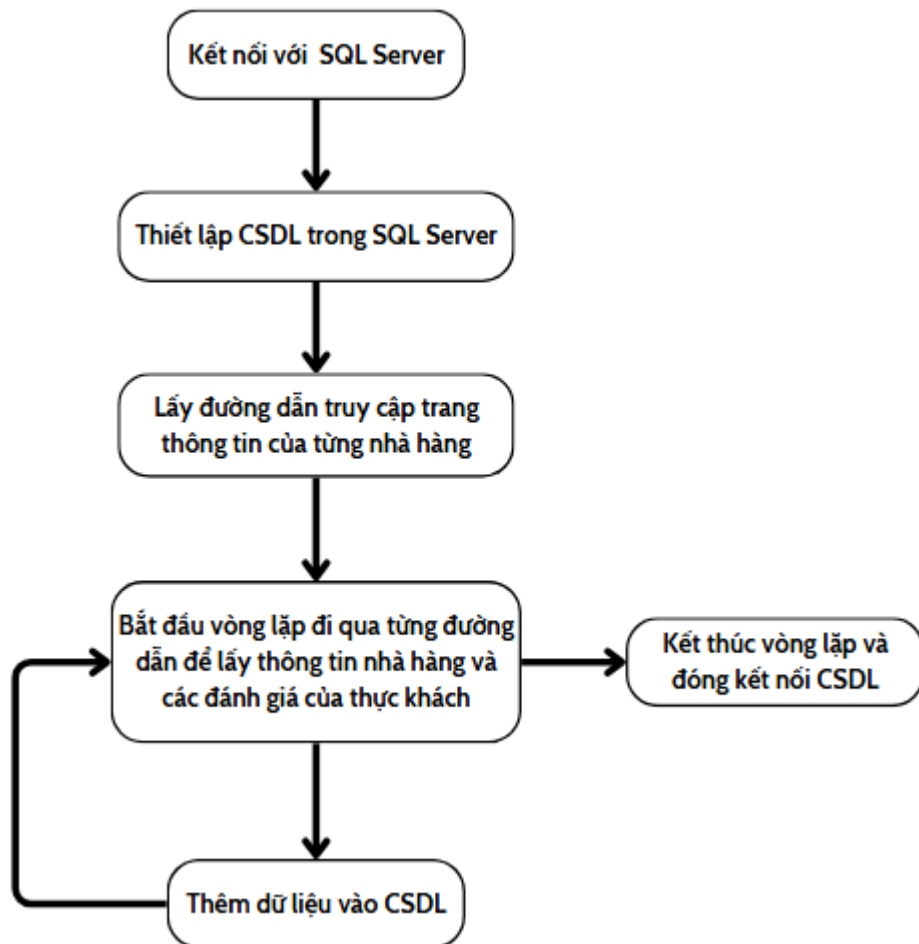
### 3.2.1. Thu thập dữ liệu

Thu thập dữ liệu là bước đầu tiên khi ta thực hiện bất kì một nghiên cứu, phân tích nào đó. Dữ liệu có thể được lấy từ nhiều nguồn khác nhau nhằm mục đích phục vụ cho việc lưu trữ, phân tích và sử dụng cho các mục tiêu như nghiên cứu, kinh doanh hoặc phát triển ứng dụng, phần mềm nào đó. Trong dự án nghiên cứu này nhóm thực hiện thu thập dữ liệu từ google review.

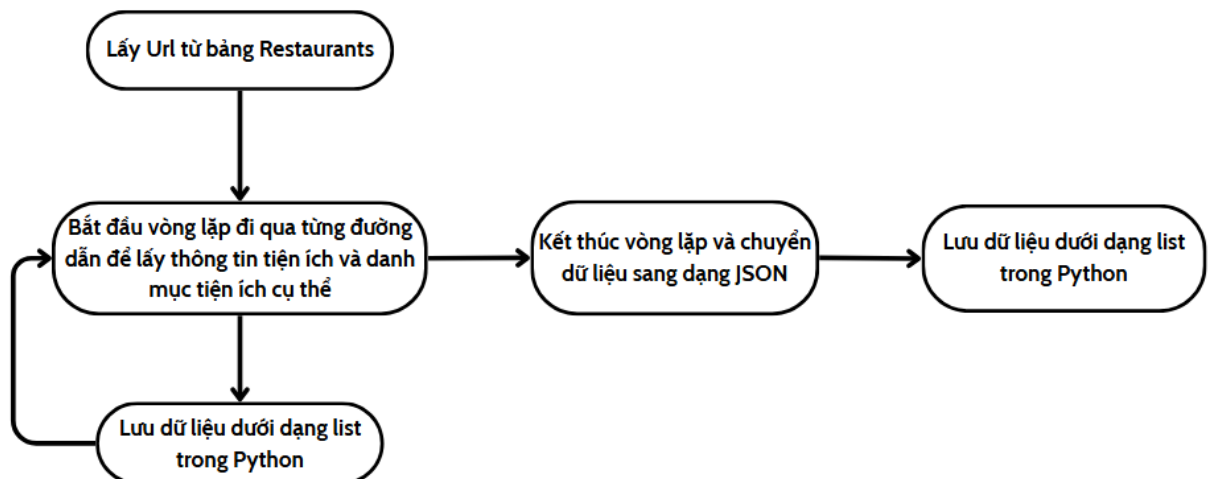
**Phương pháp thực hiện:** sử dụng ngôn ngữ lập trình Python để lấy thông tin của các nhà hàng tại khu vực Thành phố Đà Nẵng và đánh giá của thực khách về nhà hàng đó.

**Website:** <https://www.google.com/maps/search/Restaurants+in+Da+Nang>

**Quy trình thực hiện:**



Hình 2: Quy trình thu thập dữ liệu nhà hàng và đánh giá



Hình 3: Quy trình thu thập dữ liệu loại tiện ích và tiện ích cụ thể của nhà hàng

**Bước 1:** Import các thư viện cần dùng để kết nối CSDL, thiết lập trình duyệt, hỗ trợ xác định các phân tử cần thu thập

**Bước 2:** Tạo file nhật ký để ghi lại quá trình Crawl nhằm hỗ trợ cho việc phát hiện và xử lý lỗi (nếu có) trong quá trình thực thi đoạn code. Cứ mỗi bước trong quá

trình Crawl sẽ đều được ghi lại thành công hay thất bại (có kèm thông báo lỗi) trong file nhật ký.

**Bước 3:** Kết nối với SQL Server và tạo CSDL

**Bước 4:** Thiết lập các hàm hỗ trợ trong quá trình Crawl

**Bước 5:** Thiết lập hàm thu thập dữ liệu

**Bước 6:** Thực thi đoạn mã Crawl

**Bước 7:** Thu thập dữ liệu tiện ích nhà hàng

**Bước 8:** Đóng kết nối với CSDL

3.2.2. Kết quả sau khi thu thập dữ liệu

- Bảng Restaurants:** Gồm 68 dòng và 1 cột

restaurant_id	restaurant_name	url	restaurant_type	rating_average	num_of_reviews	price_level	address	latitude	longitude	phone	created_at
1	Nhà hàng NHÀ BẾP CHO HÂN	https://www.google.com/maps/place/Nhà+Bếp+Cho+Hân	Vietnamese restaurant	4.7	4406	\$100-200K	22 Hùng Vương, Hải Châu 1, Hải Châu, Đà Nẵng 550000	10.6886874	108.2243118	02363986268	2025-03-20 23:52:16.330
2	Thị GI Restaurant Đà Nẵng	https://www.google.com/maps/place/Thị+Gi+Restaurant	Vietnamese restaurant	4.8	6339	\$100-200K	53 Phạm Thiệu Duyệt, Bắc Mỹ Phú, Ngũ Hành Sơn, Đà	10.6553996	108.2415819	02363986065	2025-03-21 00:31:56.500
3	Bếp Culin Đà Nẵng	https://www.google.com/maps/place/Bếp+Culin+Đà+N	Vietnamese restaurant	4.8	7430	\$100-200K	31-33 Trần Bạch Đằng, Phước Mỹ, Sơn Trà, Đà Nẵng 550	10.6579376	108.2461059	0702689869	2025-03-21 01:14:41.647
4	Cà Ba Phở bô	https://www.google.com/maps/place/Cà+Ba+Phở+bô	Restaurant	4.8	5432	\$100-200K	154 Bạch Đằng, Hải Châu 1, Hải Châu, Đà Nẵng 550000	10.6669384	108.2248857	02363986268	2025-03-21 02:05:42.390
5	Nhà Hàng Nhà Bếp Khuê Mỹ	https://www.google.com/maps/place/Nhà+Hàng+Nhà+B	Vietnamese restaurant	4.6	4290	\$100-200K	416 Vô Nguyên Giáp, Bắc Mỹ An, Ngũ Hành Sơn, Đà N	10.6419369	108.2483139	02363986000	2025-03-21 02:44:52.247
6	Sky View Restaurant	https://www.google.com/maps/place/Sky+View+Restaur	Fine dining restaurant	4.6	955	Unknown	216 Vô Nguyên Giáp, Phước Mỹ, Sơn Trà, Đà Nẵng 550	10.6675621	108.244812	02363986268	2025-03-21 03:26:35.977
7	Ấn Thai Restaurant	https://www.google.com/maps/place/Ấn+Thai+Restaur	Vietnamese restaurant	4.6	2636	\$100-200K	114 Bạch Đằng, Hải Châu 1, Hải Châu, Đà Nẵng 550000	10.6681877	108.2248154	0807624265	2025-03-21 04:17:22.507
8	Red Sky Steakhouse	https://www.google.com/maps/place/Red+Sky+Steakhou	Steak house	4.6	820	\$\$\$	248 Trần Phú, Phước Ninh, Hải Châu, Đà Nẵng 550000	10.6628128	108.2231903	0905727576	2025-03-21 04:56:36.617
9	MAAZI Da Nang	https://www.google.com/maps/place/MAAZI+Da+Nang	Indian restaurant	4.8	2683	\$100-200K	264 Trần Phú, Phước Ninh, Hải Châu, Đà Nẵng 700000	10.6623131	108.2231903	02363522211	2025-03-21 06:01:59.347
10	Redorab Danang	https://www.google.com/maps/place/Redorab+Danang	Traditional restaurant	4.8	1319	Unknown	138 Phạm Văn Đồng, sweet, Sơn Trà, Đà Nẵng 550000	10.6705624	108.235939	0394724919	2025-03-21 15:09:43.947
11	Nhà Bếp Xua Restaurant	https://www.google.com/maps/place/Nhà+Bếp+Xua+R	Vietnamese restaurant	4.8	5819	\$100-200K	648 Hà Bổng, Phước Mỹ, Sơn Trà, Đà Nẵng 550000	10.6690629	108.2443161	0906123685	2025-03-21 15:30:05.413
12	Luca Restaurant - Danang seafood restaurant - Լուկ	https://www.google.com/maps/place/Luca+Restauran	Restaurant	4.8	2288	\$\$\$	15 D. Lê Duẩn, P. Hải Châu, Đà Nẵng 550000	10.6715618	108.2241866	0904495785	2025-03-21 16:09:03.457
13	Nhà Hàng Sĩ Dining	https://www.google.com/maps/place/Nhà+Hàng+Sĩ+D	Italian restaurant	4.9	461	\$1,000,000+	1 Giang Châu 2 Phường Khuê Mỹ, Quận Ngũ Hành Sơn	10.6301876	108.23806	0774464441	2025-03-21 20:04:39.537
14	Nhà hàng Lân Giô	https://www.google.com/maps/place/Nhà+Hàng+Lân+G	Vietnamese restaurant	4.9	1066	Unknown	169 Trần Phú, Hải Châu 1, Hải Châu, Đà Nẵng 550000	10.6668666	108.2238159	0973972074	2025-03-21 22:30:40.427
15	Nhà hàng Thiên Kim	https://www.google.com/maps/place/Nhà+Hàng+Thiên+K	Restaurant	4.8	6136	\$100-300K	166 Bạch Đằng, Hải Châu 1, Hải Châu, Đà Nẵng 550000	10.6668666	108.2248857	0773568882	2025-03-21 22:59:30.207
16	MaChin Bistro Danang	https://www.google.com/maps/place/MaChin+Bistro+D	Restaurant	4.6	303	Unknown	44 Đào Xánh 4, Khu Bật Thử, Đào Xánh, Hải Châu, Đà	10.6479574	108.227592	0905263891	2025-03-22 12:26:23.987
17	SIX ON SIX CAFE - BRUNCH & RESTAURANT	https://www.google.com/maps/place/SIX+ON+SIX+CAF	Restaurant	4.8	1664	\$100-200K	64 Bà Huyện Thanh Quan, Bắc Mỹ An, Ngũ Hành Sơn	10.6431081	108.2441984	0946114967	2025-03-22 12:58:49.923
18	Mỹ Hương Cà Sầu	https://www.google.com/maps/place/Mỹ+Hương+Cà+S	Restaurant	4.7	880	\$1-100,000	397 B. Trần Hưng Đạo, An Hải Trung, Sơn Trà, Đà N	10.6661889	108.2299066	0963862634	2025-03-22 13:44:00.450
19	EMOY'S HOMECOOKED VIETNAMESE CUISINE	https://www.google.com/maps/place/EMOY'S+HOME	Vietnamese restaurant	4.8	2232	\$100-200K	Đường Mỹ Khê 4, Phước Mỹ, Sơn Trà, Đà Nẵng 550000	10.6581875	108.2454376	0934044968	2025-03-22 14:30:42.140
20	Restaurant PHO HOUSE	https://www.google.com/maps/place/Restaurant+PHO	Restaurant	4.7	413	\$100-200K	111 Ngä Nam 8, Hoà Cường Bắc, Hải Châu, Đà Nẵng	10.6348129	108.2299066	0708352141	2025-03-22 15:12:30.183
21	Nhà Hàng Chay Ngự Chi - Vegetarian Restaurant M	https://www.google.com/maps/place/Nhà+Hàng+Chay	Vegetarian restaurant	4.6	971	\$1-100,000	202 Hoàng Diệu, Nam Dương, Hải Châu, Đà Nẵng 550000	10.6619373	108.2174377	0935085068	2025-03-22 15:31:34.960
22	Anchor Beer Club	https://www.google.com/maps/place/Anchor+Beer+Cl	Restaurant	4.4	342	\$\$\$	29 Quang Dũng, Vĩnh Trung, Thanh Khê, Đà Nẵng 5500	10.6634364	108.2116149	0379172285	2025-03-22 16:26:13.847
23	Nhà Hàng Đà Nẵng - Ho Xanh Restaurant	https://www.google.com/maps/place/Nhà+Hàng+Đà+N	Restaurant	4	177	\$\$\$	K 1 Đỗ Quang, Vĩnh Trung, Thanh Khê, Đà Nẵng 550000	10.6628689	108.2128143	0905719477	2025-03-22 21:11:12.577
24	Benares Central Da Nang	https://www.google.com/maps/place/Benares+Central	Restaurant	4.1	202	Unknown	207 Nguyễn Văn Thoại, Anh Hải Đông, Sơn Trà, Đà N	10.655887	108.2435608	0961888597	2025-03-22 22:19:32.750
25	Talaysai Harashi Da Nang	https://www.google.com/maps/place/Talaysai+Harash	Talaysai restaurant	4.9	64	\$1-100,000	118 Phan Thanh, Thạc Gián, Thanh Khê, Đà Nẵng 550000	10.6618134	108.2299067	0971012369	2025-03-22 23:16:43.770
26	Le Rendez Vous	https://www.google.com/maps/place/Le+Rendez+Vous	French restaurant	4.3	589	\$1,000,000+	28 Nguyễn Cao Luyến, An Hải Bắc, Sơn Trà, Đà Nẵng 5	10.6683113	108.2436981	0905641966	2025-03-23 06:25:25.310
27	Bún Bò Bà Rôi	https://www.google.com/maps/place/Bún+Bò+Bà+Rôi	Noodle shop	4.5	199	\$1-100,000	5 Phan Thuận Tài, Hòa Thuận Đông, Hải Châu, Đà N	10.655688	108.2214355	0704194497	2025-03-23 09:27:47.143

Hình 4: Kết quả sau khi thu thập dữ liệu nhà hàng

- Bảng review:** Gồm 20420 dòng và 14 cột

review_id	reviewer_name	reviewer_info	rating	review_time	review_text	service_rating	food_rating	atmosphere_rating	service_type	meal_type	language	nation	created_at	restaurant_id
1	Jojo Gneo	Local Guide · 46 reviews · 28 photos	5	3 weeks ago	As soon as the food arrived I knew I'd made the rig...	5	5	5	Unknown	Unknown	Unknown	Unknown	2025-03-20 23:57:54.067	1
2	Annie Song	Local Guide · 172 reviews · 684 photos	5	3 weeks ago	As Vietnamese living overseas, we loved the flavo...	5	5	5	Dine in	Lunch	Unknown	Unknown	2025-03-20 23:57:43.930	1
3	Jiwon Kim	4 reviews · 11 photos	5	2 months ago	It was a huge and cozy place to try Vietnam stoo f...	5	5	5	Dine in	Lunch	Unknown	Unknown	2025-03-20 23:58:04.227	1
4	kattise tang	Local Guide · 11 reviews · 14 photos	3	5 months ago	Had the beef pho and fried rice. Flavors not too au...	3	3	4	Dine in	Lunch	Unknown	Unknown	2025-03-20 23:58:24.627	1
5	Rina Runnel	Local Guide · 12 reviews · 22 photos	5	5 months ago	I would like to give more stars! I'm a big fan of thi...	5	5	5	Unknown	Unknown	Unknown	Unknown	2025-03-21 00:00:16.657	1
6	Richard Hinman	Local Guide · 83 reviews · 313 photos	3	2 months ago	Very Koreanized version of Vietnamese food. We ca...	3	1	5	Unknown	Unknown	Unknown	Unknown	2025-03-20 23:58:34.717	1
7	Eik T.	Local Guide · 19 reviews · 11 photos	4	3 months ago	Tourist-friendly place to eat. Clean and accommodat...	5	4	5	Dine in	Lunch	Unknown	Unknown	2025-03-20 23:58:44.907	1
8	Emily Lee	Local Guide · 47 reviews · 204 photos	4	3 months ago	Service is good. Staff are friendly and attentive. The...	4	4	4	Dine in	Lunch	Unknown	Unknown	2025-03-20 23:59:15.510	1
9	Erin Mphumuboon	Local Guide · 11 reviews · 25 photos	5	2 years ago	Food is nice and the portion is generous. We ord...	NULL	NULL	NULL	Dine in	Dinner	Unknown	Unknown	2025-03-21 00:02:48.523	1
10	ThayNha H	Local Guide · 123 reviews · 1,222 photos	3	a month ago	Located across from Chợ Hàn, very busy at lunch ti...	4	3	4	Dine in	Lunch	Unknown	Unknown	2025-03-20 23:59:05.350	1
11	Boon Keng	Local Guide · 130 reviews · 229 photos	5	3 months ago	Restaurant was huge and it was still full during wee...	5	5	5	Dine in	Dinner	Unknown	Unknown	2025-03-20 23:59:25.777	1
12	Elanore A.	Local Guide · 49 reviews · 301 photos	4	4 months ago	We took the 9 dishes menu set for 2 at 210,000 don...	4	4	4	Dine in	Dinner	Unknown	Unknown	2025-03-21 00:00:06.557	1
13	Bernard Lee	1 review	5	2 months ago	Large variety of choices, from set menus to alla cart...	5	5	4	Dine in	Lunch	Unknown	Unknown	2025-03-21 00:08:03.887	1
14	Leslie Peady	2 reviews · 1 photo	5	2 months ago	Very busy, but so efficiently run, we didn't have to w...	5	5	5	Unknown	Unknown	Unknown	Unknown	2025-03-21 00:00:57.317	1
15	Gloria LEE	2 reviews · 8 photos	4	2 weeks ago	Nice environment, good air-conditioning and good s...	4	4	4	Dine in	Lunch	Unknown	Unknown	2025-03-21 00:01:38.030	1
16	chrisina lea ekawaty	Local Guide · 148 reviews · 105 photos	5	a year ago	Nice place to try authentic vietnamese food. Good a...	5	5	5	Dine in	Lunch	Unknown	Unknown	2025-03-21 00:01:27.840	1
17	ZB Liang	7 reviews · 3 photos	5	3 months ago	We bumped into this place after a visit to Han Mak...	5	5	5	Dine in	Lunch	Unknown	Unknown	2025-03-21 00:09:25.010	1
18	Justin Phup	Local Guide · 55 reviews · 6 photos	5	2 months ago	The pho featured a clear, flavorful broth and tender...	5	5	5	Dine in	Dinner	Unknown	Unknown	2025-03-21 00:01:07.460	1
19	Jack Lim	Local Guide · 218 reviews · 274 photos	5	5 months ago	Great place, good food, and fair price. We enjoy ma...	5	4	5	Dine in	Lunch	Unknown	Unknown	2025-03-21 00:01:58.483	1
20	Kelly	Local Guide · 141 reviews · 685 photos	5	11 months ago	One of my favorite restaurants in DaNang. The Hoi...	5	5	5	Dine in	Lunch	Unknown	Unknown	2025-03-21 00:01:42.963	1
21	Francis Au	Local Guide · 4 reviews · 43 photos	5	a year ago	traditional Vietnamese cuisine , clear menu with lots...	5	5	5	Dine in	Dinner	Unknown	Unknown	2025-03-21 00:09:55.333	1
22	Hup Lee Tan	Local Guide · 14 reviews · 13 photos	5	3 months ago	Good food and great service. One of the staff even...	5	5	5	Dine in	Lunch	Unknown	Unknown	2025-03-21 00:01:48.283	1
23	Boy Trung	14 reviews · 9 photos	1	5 months ago	Due to the high reviews, our expectations were set...	3	3	4	Unknown	Unknown	Unknown	Unknown	2025-03-21 00:03:09.571	1
24	Lydia See	2 reviews · 6 photos	5	3 weeks ago	Great food and affordable price! One of the best sta...	4	4	5	Unknown	Dinner	Unknown	Unknown	2025-03-21 00:02:59.710	1
25	Michael Huang	1 review	5	2 weeks ago	good food and free GREAT mango smoothie ask for...	NULL	NULL	NULL	Unknown	Unknown	Unknown	Unknown	2025-03-21 00:11:26.580	1
26	Leang	Local Guide · 124 reviews · 417 photos	5	a year ago	Another restaurant popular with Korean tourists. Ch...	5	5	4	Dine in	Lunch	Unknown	Unknown	2025-03-21 00:03:40.293	1
27	Zach Tan	3 reviews	2	3 months ago	Honestly very nice place, staff is proficien in English...	5	2	5	Unknown	Unknown	Unknown	Unknown	2025-03-21 00:03:30.130	1
28	Celia Hu	6 reviews	5	2 months ago	Great place to grab lunch after shopping the Han M...	5	5	5	Dine in	Lunch	Unknown	Unknown	2025-03-21 00:04:10.807	1
29	-Mingpei-	1 review	5	a month ago	So good place to eat, price reasonable, food is so ri...	5	5	5	Unknown	Unknown	Unknown	Unknown	2025-03-21 00:15:19.750	1
30	Fion CC	Local Guide · 20 reviews · 10 photos	5	3 months ago	Was looking for a lunch place while standing opposi...	5	5	5	Dine in	Lunch	Unknown	Unknown	2025-03-21 00:03:50.453	1
31	Tom G.	Local Guide · 102 reviews · 156 photos	1	a year ago	Korean tourist trap! Reviews are bought and incoi...	1	1	2	Unknown	Unknown	Unknown	Unknown	2025-03-21 00:05:11.753	1

Hình 5: Kết quả sau khi thu thập dữ liệu đánh giá

- Bảng features\_type:** Gồm 68 dòng và 2 cột

restaurant_name	feature_type
The View - Yacht Restaurant Da Nang	["Service options", "Highlights", "Popular for", "Offerings", "Dining options", "Amenities", "Atmosphere", "Crowd", "Planning", "Payments", "Children", "
Bếp Quê	["Accessibility", "Service options", "Highlights", "Popular for", "Offerings", "Dining options", "Amenities", "Atmosphere", "Crowd", "Planning", "Paymen
Riverside Terrace Restaurant (formerly L'Italiano Riverside DaNang)	["Accessibility", "Service options", "Highlights", "Popular for", "Offerings", "Dining options", "Amenities", "Atmosphere", "Crowd", "Planning", "Childrer
O'Kia - Seafood Restaurant	["Accessibility", "From the business", "Service options", "Highlights", "Popular for", "Offerings", "Dining options", "Amenities", "Atmosphere", "Crowd", "
Roots Plant-based Cafe - Healthy Vegan restaurant	["Accessibility", "Service options", "Highlights", "Popular for", "Offerings", "Dining options", "Amenities", "Atmosphere", "Crowd", "Planning", "Paymen
Bikini Bottom Express	["Accessibility", "Service options", "Highlights", "Popular for", "Offerings", "Dining options", "Amenities", "Atmosphere", "Crowd", "Planning", "Paymen

Hình 6: Kết quả sau khi thu thập loại tiện ích

- **Bảng feature\_items:** Gồm 68 dòng và 16 cột

Restaurant_name	Offerings	Dining_op	Crowd	Pets	From_the	Popular_fr	Highlights	Children	Parking	Service_of	Accessibil	Amenities	Atmosphe	Planning	Payments
The View - Yacht Restaurant Da Nang	["Serves al	["Serves bi	["Family fr	["]	["]	["Popular	["Has grea	["Good for	["Free of c	["Has outc	["]	["Has bar	["Cosy", "C	["Accepts	["Accepts
Bếp Quê	["Serves al	["Serves bi	["Good for	["]	["]	["Popular	["Has grea	["Good for	["Difficult	["Serves di	["No whee	["Has toile	["Casual",	["Accepts	["Accepts
Riverside Terrace Restaurant (formerly L'Italiano Riverside DaNang)	["Serves al	["Serves bi	["Good for	["]	["]	["Popular	["Has grea	["Good for	["Free par	["Has outc	["Has whe	["Has toile	["Casual",	["Accepts	["]
O'Kia - Seafood Restaurant	["Serves al	["Serves bi	["Family fr	["]	["Identifier	["Popular	["Has grea	["Good for	["Free of c	["Has outc	["Has whe	["Has geni	["Quiet",	["Accepts	["Accepts
Roots Plant-based Cafe - Healthy Vegan restaurant	["Serves al	["Serves bi	["Family fr	["Dogs ali	["]	["Popular	["Serves gi	["Good for	["Free of c	["Has outc	["Has whe	["Has toile	["Casual",	["Accepts	["Accepts
Bikini Bottom Express	["Serves al	["Serves bi	["Family fr	["Dogs ali	["]	["Popular	["Has grea	["Good for	["Free of c	["Has outc	["Has whe	["Has bar	["Casual",	["Accepts	["Accepts

Hình 7: Kết quả sau khi thu thập tiện ích cụ thể

## CHƯƠNG 4. TIỀN XỬ LÝ DỮ LIỆU

### 4.1. Xử lý giá trị thiếu

- **Bảng Restaurants:**

- Cột “price\_level” chứa 14 giá trị thiếu. Cột này bị thiếu vì không có dữ liệu trên trang thông tin của nhà hàng, do đó không thể điền giá trị thay thế được → Thay thế bằng giá trị “Unknown” tức là không có giá trị.

	Column_Name	Null_Count
1	restaurant_name	0
2	url	0
3	restaurant_type	0
4	rating_average	0
5	num_of_reviews	0
6	price_level	14
7	address	0
8	latitude	0
9	longitude	0
10	phone	0

Hình 8: Kết quả kiểm tra dữ liệu thiếu trong bảng restaurants

- **Bảng Reviews:**

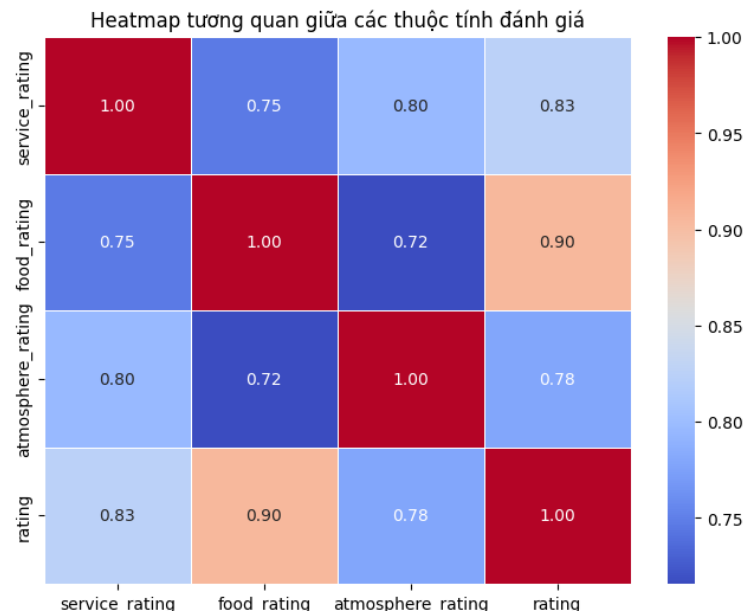
- Kết quả sau khi rà soát có các cột sau bị thiếu dữ liệu:
  - Cột “reviewer\_info” thiếu 126 dòng.
  - Cột “service\_rating” thiếu 3757 dòng.
  - Cột “food\_rating” thiếu 3791 dòng.
  - Cột “atmosphere\_rating” thiếu 3807 dòng.
  - Cột “service\_type” thiếu 10008 dòng.
  - Cột “meal\_type” thiếu 9958 dòng.
  - Cột “language” thiếu 9741 dòng.

	Column_Name	Null_Count
1	reviewer_name	0
2	reviewer_info	126
3	rating	0
4	review_time	0
5	review_text	0
6	service_rating	3757
7	food_rating	3791
8	atmosphere_rating	3807
9	service_type	10008
10	meal_type	9958
11	language	9741
12	nation	9741

Hình 9: Kết quả kiểm tra dữ liệu thiếu trong bảng reviews

○ Cách xử lý:

- Tương tự với bảng Restaurants, các cột reviewer\_info, service\_type, meal\_type, language bị thiếu vì không có thông tin trên trang web, do đó thay thế chúng bằng giá trị “Unknown”.
- Đối với các cột service\_rating, food\_rating, atmosphere\_rating chúng tôi thực hiện đánh giá mức độ tương quan của chúng với cột rating.



Hình 10: Biểu đồ heatmap ma trận tương quan giữa rating với cột service\_rating, food\_rating, atmosphere\_rating

- Kết quả cho thấy chúng có sự tương quan cao với nhau. Vì thế chúng tôi sử dụng mô hình hồi quy để dự đoán và thay thế các giá trị thiếu trong bộ dữ liệu.

## 4.2. Rời rạc hoá dữ liệu

- **Bảng features\_type và feature\_items:**

- Dữ liệu cào về có dạng list trong python nên sẽ rất khó để có thể trực quan hoá cũng như cho các phân tích sau này.

- **Dữ liệu bảng features\_type:**

	restaurant_name	features_type
1	Nhà hàng NHÀ BẾP CHO HÀN	[Accessibility, 'Service options', 'Highlights', 'Popular for', 'Offerings', 'Dining options', 'Amenities', 'Atmosphere', 'Crowd', 'Planning', 'Payments', 'Children', 'Pets']
2	Thịa Gổ Restaurant Da Nang	[Accessibility, 'Service options', 'Highlights', 'Popular for', 'Offerings', 'Dining options', 'Amenities', 'Atmosphere', 'Crowd', 'Planning', 'Payments', 'Children', 'Parking']
3	Bếp Cuốn Đà Nẵng	[Accessibility, 'Service options', 'Highlights', 'Popular for', 'Offerings', 'Dining options', 'Amenities', 'Atmosphere', 'Crowd', 'Planning', 'Payments', 'Children', 'Parking']
4	Cố Ba Phở bò	[Accessibility, 'Service options', 'Highlights', 'Popular for', 'Offerings', 'Dining options', 'Amenities', 'Atmosphere', 'Crowd', 'Planning', 'Payments', 'Children', 'Parking']
5	Nhà Hàng Nhà Bếp Khuê Mỹ	[Accessibility, 'From the business', 'Service options', 'Highlights', 'Popular for', 'Offerings', 'Dining options', 'Amenities', 'Atmosphere', 'Crowd', 'Planning', 'Payments', 'Children', '...']
6	Sky View Restaurant	[Accessibility, 'Service options', 'Highlights', 'Popular for', 'Offerings', 'Dining options', 'Amenities', 'Atmosphere', 'Crowd', 'Planning', 'Payments', 'Children', 'Parking']
7	Ẩn Thới Restaurant	[Accessibility, 'Service options', 'Highlights', 'Popular for', 'Offerings', 'Dining options', 'Amenities', 'Atmosphere', 'Crowd', 'Planning', 'Payments', 'Children', 'Parking']
8	Red Sky Steakhouse	[Accessibility, 'Service options', 'Highlights', 'Popular for', 'Offerings', 'Dining options', 'Amenities', 'Atmosphere', 'Crowd', 'Planning', 'Payments', 'Children', 'Parking', 'Pets']
9	MAAZI Da Nang	[Accessibility, 'Service options', 'Highlights', 'Popular for', 'Offerings', 'Dining options', 'Amenities', 'Atmosphere', 'Crowd', 'Planning', 'Payments', 'Children', 'Parking']
10	Redcrab Danang	[Accessibility, 'Service options', 'Highlights', 'Popular for', 'Offerings', 'Dining options', 'Amenities', 'Atmosphere', 'Crowd', 'Planning', 'Payments', 'Children', 'Parking', 'Pets']
11	Nhà Bếp Xùa Restaurant	[Accessibility, 'Service options', 'Highlights', 'Popular for', 'Offerings', 'Dining options', 'Amenities', 'Atmosphere', 'Crowd', 'Planning', 'Payments', 'Children', 'Parking']
12	LoCo Restaurant - Danang seafood restaurant - 다낭 ...	[Accessibility, 'Service options', 'Highlights', 'Popular for', 'Offerings', 'Dining options', 'Amenities', 'Atmosphere', 'Crowd', 'Planning', 'Payments', 'Children', 'Parking']
13	Nhà Hàng Sĩ Dining	[Accessibility, 'Service options', 'Highlights', 'Popular for', 'Offerings', 'Dining options', 'Amenities', 'Atmosphere', 'Crowd', 'Planning', 'Payments', 'Children', 'Parking']
14	Nhà hàng Lân Gió	[Service options, 'Popular for', 'Offerings', 'Dining options', 'Amenities', 'Atmosphere', 'Crowd', 'Planning', 'Payments', 'Children', 'Parking']
15	Nhà hàng Thiên Kim	[Accessibility, 'Service options', 'Highlights', 'Popular for', 'Offerings', 'Dining options', 'Amenities', 'Atmosphere', 'Crowd', 'Planning', 'Payments', 'Children', 'Parking']
16	MeliOh Bistro Danang	[Accessibility, 'From the business', 'Service options', 'Highlights', 'Popular for', 'Offerings', 'Dining options', 'Amenities', 'Atmosphere', 'Crowd', 'Planning', 'Payments', 'Children', '...']
17	SIX ON SIX CAFE - BRUNCH & RESTAURANT	[Accessibility, 'From the business', 'Service options', 'Highlights', 'Popular for', 'Offerings', 'Dining options', 'Amenities', 'Atmosphere', 'Crowd', 'Planning', 'Payments', 'Children', '...']
18	Mỹ Quảng Cổ Sầu	[Accessibility, 'From the business', 'Service options', 'Popular for', 'Offerings', 'Dining options', 'Amenities', 'Atmosphere', 'Crowd', 'Planning', 'Payments', 'Children', 'Parking', 'P...

Hình 11: Dữ liệu bảng features\_type trước khi xử lý

- **Dữ liệu bảng feature\_items:**

	restaurant_name	Offerings	Dining_options	Crowd	Pets	From_the...	Popular_for	Highlights	Children	Parking	Service_options
1	Nhà hàng NHÀ BẾP CHO HÀN	['Serves alcoh...	['Serves breakfast', 'S...	['Family friendly', 'Goo...	['Dogs all...	0	['Popular for br...	['Has great beer ...	['Good for kids', 'Hig...	0	['Offers takeaway
2	Thịa Gổ Restaurant Da Nang	['Serves alcoh...	['Serves breakfast', 'S...	['Family friendly', 'Goo...	0	0	['Popular for lu...	['Has great beer ...	['Good for kids', 'Hig...	['Free of charge...	['Has outdoor se
3	Bếp Cuốn Đà Nẵng	['Serves alcoh...	['Serves breakfast', 'S...	['Good for groups', 'P...	0	0	['Popular for lu...	['Has great beer ...	['Good for kids', 'Hig...	['Free of charge...	['Has outdoor se
4	Cố Ba Phở bò	['Serves alcoh...	['Serves breakfast', 'S...	['Good for groups', 'P...	0	0	['Popular for br...	['Has great beer ...	['Good for kids', 'Hig...	['Free of charge...	['Has no-contact
5	Nhà Hàng Nhà Bếp Khuê Mỹ	['Serves alcoh...	['Serves breakfast', 'S...	['Good for groups', 'P...	0	['Identifie...	['Popular for lu...	['Has great beer ...	['Good for kids', 'Hig...	['Free of charge...	['Has no-contact
6	Sky View Restaurant	['Serves alcoh...	['Serves breakfast', 'S...	['Family friendly', 'Goo...	0	0	['Popular for di...	['Has great beer ...	['Good for kids', 'Hig...	['Free parking l...	['Has outdoor se
7	Takoyaki Hatachi Da Nang	['Quick bite']	['Has catering']	['Good for groups', 'P...	0	0	0	0	['Good for kids']	['Free of charge...	['Offers delivery',
8	Ẩn Thới Restaurant	['Serves alcoh...	['Serves breakfast', 'S...	['Good for groups', 'P...	0	0	['Popular for br...	['Has great beer ...	['Good for kids', 'Hig...	['Free parking l...	['Has no-contact
9	Red Sky Steakhouse	['Serves alcoh...	['Serves breakfast', 'S...	['Family friendly', 'Goo...	['Dogs all...	0	['Popular for lu...	['Serves great co...	['Good for kids', 'Hig...	['Free of charge...	['Has no-contact
10	MAAZI Da Nang	['Serves alcoh...	['Serves breakfast', 'S...	['Family friendly', 'Goo...	0	0	['Popular for lu...	['Has great beer ...	['Good for kids', 'Hig...	['Free of charge...	['Offers kerbside
11	Redcrab Danang	['Serves alcoh...	['Serves lunch', 'Serve...	['Family friendly', 'Goo...	['Dogs all...	0	['Popular for di...	['Has live music']	['Good for kids', 'Hig...	['Free parking l...	['Offers kerbside
12	Nhà Bếp Xùa Restaurant	['Serves alcoh...	['Serves breakfast', 'S...	['Family friendly', 'Goo...	0	0	['Popular for lu...	['Has great beer ...	['Good for kids', 'Hig...	['Free parking l...	['Has outdoor se
13	LoCo Restaurant - Danang seafood restaurant - 다낭 ...	['Serves alcoh...	['Serves brunch', 'Ser...	['Family friendly', 'Goo...	0	0	['Popular for lu...	['Has great beer ...	['Good for kids', 'Has...	['Free of charge...	['Has outdoor se
14	Nhà Hàng Sĩ Dining	['Serves alcoh...	['Serves dinner', 'Has ...	['Family friendly', 'Goo...	0	0	['Popular for di...	['Has great beer ...	['Good for kids', 'Hig...	['Free of charge...	['Has outdoor se
15	Nhà hàng Lân Gió	['Serves alcoh...	['Serves lunch', 'Serve...	['Family friendly', 'Goo...	0	0	['Popular for lu...	0	['Good for kids', 'Hig...	['Free parking l...	['Offers takeaway
16	Nhà hàng Thiên Kim	['Serves alcoh...	['Serves breakfast', 'S...	['Good for groups', 'P...	0	0	['Popular for br...	['Has great beer ...	['Good for kids', 'Hig...	['Free of charge...	['Has outdoor se
17	MeliOh Bistro Danang	['Serves alcoh...	['Serves breakfast', 'S...	['Good for groups', 'L...	0	['Identifie...	['Popular for di...	['Serves great co...	['High chairs availabl...	['Free of charge...	['Has outdoor se

Hình 12: Dữ liệu bảng feature\_items trước khi xử lý

- Do đó chúng tôi quyết định tạo bảng features để chứa các loại tiện ích và các **mục tiện ích cụ thể riêng biệt**, gán mã id duy nhất cho từng danh mục.
- Dữ liệu sau khi thêm từng loại tiện ích vào bảng với feature\_id là mã định danh duy nhất cho **từng loại tiện ích**. Giá trị cột parent\_id để trống các dòng đầu tiên có nghĩa đây là các danh mục tiện ích cha. Dùng làm khoá ngoại để các danh mục con (tiện ích cụ thể) tham chiếu.



	feature_id	features_name	parent_id				
1	1	Accessibility	NULL	16	16	Accepts cheques	11
2	2	Amenities	NULL	17	17	Accepts credit cards	11
3	3	Atmosphere	NULL	18	18	Accepts debit cards	11
4	4	Children	NULL	19	19	Accepts NFC mobi...	11
5	5	Crowd	NULL	20	20	Accepts reservatio...	13
6	6	Dining options	NULL	21	21	Cash only	11
7	7	From the business	NULL	22	22	Casual	3
8	8	Highlights	NULL	23	23	Cosy	3
9	9	Offerings	NULL	24	24	Credit cards	11
10	10	Parking	NULL	25	25	Difficult to find a s...	10
11	11	Payments	NULL	26	26	Does not have ass...	1
12	12	Pets	NULL	27	27	Dogs allowed	12
13	13	Planning	NULL	28	28	Dogs allowed insi...	12
14	14	Popular for	NULL	29	29	Dogs allowed out...	12
15	15	Service options	NULL	29	29	Dogs allowed out...	12

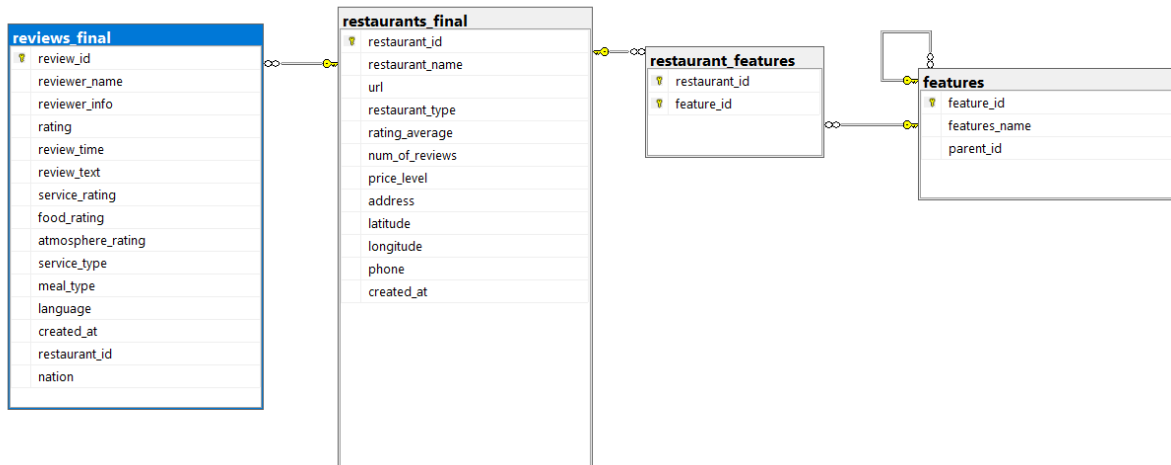
Hình 13: Dữ liệu bảng features

- Vì mối quan hệ giữa nhà hàng và tiện ích là mối quan hệ nhiều - nhiều nên cần tạo bảng trung gian là restaurant\_features để biểu diễn mối quan hệ này. Bảng trung gian có khoá ngoại tham chiếu đến cột restaurant\_id của từng nhà hàng và khoá ngoại tham chiếu đến feature\_id của từng danh mục tiện ích con.

	restaurant_id	feature_id	21	1	63	42	2	18
1	1	17	22	1	66	43	2	19
2	1	20	23	1	67	44	2	20
3	1	22	24	1	69	45	2	21
4	1	23	25	1	72	46	2	22
5	1	27	26	1	79	47	2	23
6	1	28	27	1	86	48	2	24
7	1	29	28	1	91	49	2	26
8	1	30	29	1	97	50	2	30
9	1	33	30	1	100	51	2	31
10	1	34	31	1	101	52	2	32
11	1	35	32	1	102	53	2	33
12	1	41	33	1	103	54	2	34
13	1	48	34	1	105	55	2	35
14	1	53	35	1	106	56	2	40
15	1	54	36	1	107	57	2	41
16	1	56	37	1	108	58	2	42
17	1	57	38	1	109	59	2	48
18	1	58	39	1	110	60	2	51
19	1	61	40	1	111	61	2	53
20	1	62	41	2	17	62	2	54
						63	2	55

Hình 14: Dữ liệu bảng restaurant\_features

- Sơ đồ quan hệ thực thể:



Hình 15: Sơ đồ quan hệ thực thể giữa các bảng trong SQL server

### 4.3. Biến đổi dữ liệu

#### a. Tạo cột “nation”:

- Tạo cột nation (quốc gia) và cập nhật dữ liệu: mỗi ngôn ngữ sẽ thuộc 1 quốc gia cụ thể, thứ tự cập nhật dựa trên giá trị cột “language” như sau:
  - **Unknown** → **Unknown**: Dữ liệu không xác định, vì vậy giữ nguyên là "Unknown".
  - **Korean** → **Korea**: Ngôn ngữ Korean là ngôn ngữ chính thức của Korea, do đó gán vào "Korea".
  - **Vietnamese** → **Vietnam**: Ngôn ngữ Vietnamese là ngôn ngữ chính thức của Vietnam, do đó gán vào "Vietnam".
  - **Chinese** → **China**: Chinese là ngôn ngữ chính thức của China, mặc dù có các phương ngữ khác nhau (như Cantonese, Mandarin), nhưng chung lại, ngôn ngữ chính của quốc gia này là Chinese.
  - **Japanese** → **Japan**: Japanese là ngôn ngữ chính thức của Japan, do đó gán vào "Japan".
  - **Russian** → **Russia**: Russian là ngôn ngữ chính thức của Russia, do đó gán vào "Russia".
  - **German** → **Germany**: German là ngôn ngữ chính thức của Germany, do đó gán vào "Germany".
  - **French** → **France**: French là ngôn ngữ chính thức của France, do đó gán vào "France".
  - **Thai** → **Thailand**: Thai là ngôn ngữ chính thức của Thailand, do đó gán vào "Thailand".

- **Spanish → Spain:** Spanish (hoặc Castilian) là ngôn ngữ chính thức của Spain, do đó gán vào "Spain". Mặc dù có các phương ngữ khác trong khu vực, nhưng Spanish là ngôn ngữ quốc gia.
- **Hindi, Gujarati, Marathi, Telugu, Kannada → India:** Các ngôn ngữ Hindi, Gujarati, Marathi, Telugu, và Kannada đều là ngôn ngữ chính của India. Mặc dù mỗi ngôn ngữ có thể liên quan đến một số vùng đặc biệt trong Ấn Độ, nhưng tất cả đều là các ngôn ngữ chính thức trong quốc gia này, vì vậy gán vào "India".
- **Italian → Italy:** Italian là ngôn ngữ chính thức của Italy, do đó gán vào "Italy".
- **Dutch → Netherlands:** Dutch là ngôn ngữ chính thức của Netherlands, do đó gán vào "Netherlands".
- **Arabic → Arab countries:** Arabic là ngôn ngữ chính thức ở nhiều quốc gia Trung Đông và Bắc Phi. Do đó, tôi sử dụng "Arab countries" để bao quát tất cả các quốc gia này.
- **Danish → Denmark:** Danish là ngôn ngữ chính thức của Denmark, do đó gán vào "Denmark".
- **Polish → Poland:** Polish là ngôn ngữ chính thức của Poland, do đó gán vào "Poland".
- **Hebrew → Israel:** Hebrew là ngôn ngữ chính thức của Israel, do đó gán vào "Israel".
- **Norwegian → Norway:** Norwegian là ngôn ngữ chính thức của Norway, do đó gán vào "Norway".
- **Portuguese → Portugal:** Portuguese là ngôn ngữ chính thức của Portugal, do đó gán vào "Portugal".
- **Swedish → Sweden:** Swedish là ngôn ngữ chính thức của Sweden, do đó gán vào "Sweden".
- **Turkish → Turkey:** Turkish là ngôn ngữ chính thức của Turkey, do đó gán vào "Turkey".
- **Indonesian → Indonesia:** Indonesian là ngôn ngữ chính thức của Indonesia, do đó gán vào "Indonesia".
- **Ukrainian → Ukraine:** Ukrainian là ngôn ngữ chính thức của Ukraine, do đó gán vào "Ukraine".
- **Filipino → Philippines:** Filipino là ngôn ngữ chính thức của Philippines, do đó gán vào "Philippines".
- **Czech → Czech Republic:** Czech là ngôn ngữ chính thức của Czech Republic, do đó gán vào "Czech Republic".

- **Slovak → Slovakia:** Slovak là ngôn ngữ chính thức của Slovakia, do đó gán vào "Slovakia".
- **Catalan → Spain:** Catalan là một trong những ngôn ngữ chính trong Spain, đặc biệt ở khu vực Catalunya. Tuy nhiên, vì Catalan không phải là ngôn ngữ chính thức của toàn quốc, tôi đã gán vào Spain.
- **Igbo → Nigeria:** Igbo là một trong các ngôn ngữ chính ở Nigeria, do đó gán vào "Nigeria".
- **Cebuano → Philippines:** Cebuano là một trong các ngôn ngữ chính ở Philippines, do đó gán vào "Philippines".
- **Northern Sotho → South Africa:** Northern Sotho là một trong các ngôn ngữ chính ở South Africa, do đó gán vào "South Africa".
- **Finnish → Finland:** Finnish là ngôn ngữ chính thức của Finland, do đó gán vào "Finland".
- **Krio → Sierra Leone:** Krio là ngôn ngữ chính ở Sierra Leone, do đó gán vào "Sierra Leone".
- **Croatian → Croatia:** Croatian là ngôn ngữ chính thức của Croatia, do đó gán vào "Croatia".
- **Oromo → Ethiopia:** Oromo là ngôn ngữ chính trong Ethiopia, do đó gán vào "Ethiopia".
- **Hungarian → Hungary:** Hungarian là ngôn ngữ chính thức của Hungary, do đó gán vào "Hungary".
- **Bulgarian → Bulgaria:** Bulgarian là ngôn ngữ chính ở Bulgaria, do đó gán vào "Bulgaria".
- **Kazakh → Kazakhstan:** Kazakh là ngôn ngữ chính thức của Kazakhstan, do đó gán vào "Kazakhstan".
- **Greek → Greece :** Greek là ngôn ngữ chính trong Greece, do đó gán vào "Greece".

*b. Tạo cột "review\_date":*

- **Xử lý cột "created\_at":** Chuyển đổi kiểu dữ liệu Object về datetime bằng Python

```
# Chuyển đổi cột `created_at` sang datetime
df["created_at"] = pd.to_datetime(df["created_at"])
```

- Tạo cột "review\_date" và cập nhật dữ liệu:
  - Cột review\_date được tính bằng giá trị ngày thu thập dữ liệu lùi số ngày tương ứng trong cột review\_time.

- Ví dụ: review\_time là 1 tuần thì lấy thời điểm ngày thu thập dữ liệu là ngày 20/03/2025 lùi lại 7 ngày sẽ ra kết quả ngày viết đánh giá là 27/02/2025.

		created_at	review_time	review_date
0	2025-03-20	23:57:54.067	3 weeks ago	2025-02-27
1	2025-03-20	23:57:43.930	3 weeks ago	2025-02-27
2	2025-03-20	23:58:04.227	2 months ago	2025-01-19
3	2025-03-20	23:58:24.627	5 months ago	2024-10-21
4	2025-03-21	00:00:16.657	5 months ago	2024-10-22
...	...	...	...	...
20414	2025-05-02	13:09:35.527	7 months ago	2024-10-04
20415	2025-05-02	13:09:35.527	a year ago	2024-05-02
20416	2025-05-02	13:09:35.527	a year ago	2024-05-02
20417	2025-05-02	13:09:35.527	3 months ago	2025-02-01
20418	2025-05-02	13:09:35.527	a year ago	2024-05-02

Hình 16: Kết quả sau khi xử lý và nhập dữ liệu thành công

### c. Tạo cột “district”:

Cột “Address” nội dung được ghi là địa chỉ đầy đủ bao gồm số nhà, tên đường, tên phường, quận và thành phố. Các mô hình và tools không thể nhận biết rõ cụ thể thông tin địa chỉ được ghi chép chi tiết như thế này, vì thế nhóm tách giá trị quận từ địa chỉ nhà hàng bằng Python. Kết quả sau khi xử lý:

restaurant_id		restaurant_name	address	district
0	1	Nhà hàng NHÀ BẾP CHỢ HÀN	22 Hùng Vương, Hải Châu 1, Hải Châu, Đà Nẵng 5...	Hải Châu
1	2	Thìa Gỏi Restaurant Da Nang	53 Phan Thúc Duyệt, Bắc Mỹ Phú, Ngũ Hành Sơn, ...	Ngũ Hành Sơn
2	3	Bếp Cuốn Đà Nẵng	31-33 Trần Bạch Đằng, Phước Mỹ, Sơn Trà, Đà Nẵ...	Sơn Trà
3	4	Cô Ba Phở bò	154 Bạch Đằng, Hải Châu 1, Hải Châu, Đà Nẵng 5...	Hải Châu
4	5	Nhà Hàng Nhà Bếp Khuê Mỹ	416 Võ Nguyên Giáp, Bắc Mỹ An, Ngũ Hành Sơn, Đ...	Ngũ Hành Sơn
...	...	...	...	...
63	67	Bếp Quê	187 Trần Phú, Phước Ninh, Hải Châu, Đà Nẵng 50000	Hải Châu
64	68	Riverside Terrace Restaurant (formerly L'Itali...	493 Đ. Trần Hưng Đạo, An Hải Trung, Sơn Trà, Đ...	Sơn Trà
65	69	Ơi Kia - Seafood Restaurant	7 Trần Phú, Thạch Thang, Hải Châu, Đà Nẵng 550000	Hải Châu
66	70	Roots Plant-based Cafe - Healthy Vegan restaurant	26-27 Trần Bạch Đằng, Bắc Mỹ Phú, Ngũ Hành Sơn...	Ngũ Hành Sơn
67	71	Bikini Bottom Express	45-47 An Thượng 2, Bắc Mỹ An, Ngũ Hành Sơn, Đà...	Ngũ Hành Sơn

Hình 17: Kết quả sau khi tách quận từ address

## CHƯƠNG 5. XÂY DỰNG MÔ HÌNH

### 5.1. Mô tả bài toán

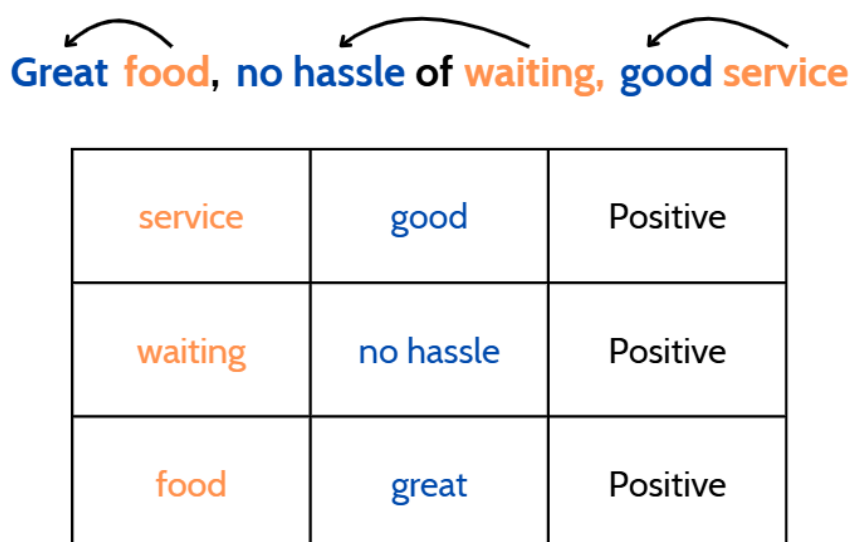
Bài toán được đặt ra là **trích xuất bộ ba khía cạnh** từ một đoạn đánh giá (review) văn bản do người dùng viết, bao gồm:

- **Aspect:** Khía cạnh cụ thể mà người dùng đề cập tới (ví dụ: "dịch vụ", "giá cả", "nhân viên").

- **Opinion:** Cụm từ thể hiện quan điểm hoặc nhận định của người dùng về khía cạnh đó (ví dụ: "tận tình", "quá cao").
- **Sentiment:** Sắc thái cảm xúc tương ứng với quan điểm, được phân loại thành *Tích cực (Positive)*, *Tiêu cực (Negative)*, hoặc *Trung tính (Neutral)*.

Mục tiêu là xây dựng một hệ thống có khả năng tự động phân tích các đánh giá này bằng cách sử dụng mô hình ngôn ngữ lớn (LLM). Dữ liệu đầu vào là các câu đánh giá của khách hàng, đầu ra là một danh sách các bộ ba (*aspect, opinion, sentiment*) được trích xuất.

**Ví dụ:**



Hình 18: Ví dụ về bài toán trích xuất Aspect, Opinion và Sentiment từ Review

Ví dụ trong **Hình 18** mô tả một câu đánh giá:

**"Great food, no hassle of waiting, good service."**

Từ câu này, ta có thể trích xuất ra ba bộ ba (triplet) biểu thị cảm xúc liên quan đến các khía cạnh (aspect) khác nhau:

- ("food", "great", Positive)
- ("waiting", "no hassle", Positive)
- ("service", "good", Positive)

Đây là ví dụ điển hình cho bài toán Aspect Sentiment Triplet Extraction (ASTE), trong đó hệ thống cần trích xuất đồng thời:

- Khía cạnh (aspect),

- Cụm từ thể hiện quan điểm (opinion term),
- Cảm xúc liên quan (sentiment polarity: tích cực, tiêu cực, trung tính).

## 5.2. Thu thập dữ liệu

Bộ dữ liệu **16res** được thu thập từ [GitHub repository của bài báo Learning Span-Level Interactions for Aspect Sentiment Triplet Extraction](#) do tác giả công bố tại hội nghị ACL 2021. Dữ liệu này thuộc phiên bản **ASTE-Data-V2**, được tinh chỉnh lại từ phiên bản ASTE-Data-V1 để loại bỏ các triplet xung đột về mặt cảm xúc và định dạng thẻ không đầy đủ.

Dữ liệu được chia thành 3 phần: tập huấn luyện (*train*), tập kiểm tra (*test*) và tập xác thực (*dev*), được lưu trữ dưới dạng file *.txt*. Mỗi dòng trong file bao gồm một câu đầu vào và danh sách các bộ ba (aspect, opinion, sentiment) tương ứng. Cấu trúc cụ thể của một dòng như sau:

**sentence##### [(aspect\_span, opinion\_span, sentiment), ...]**

**Trong đó:**

- **aspect\_span** và **opinion\_span** là các chỉ số từ (theo token) thể hiện vị trí của từ hoặc cụm từ đại diện cho chủ thể (aspect) và nhận xét (opinion).
- **sentiment** là nhãn cảm xúc tương ứng: 'POS' (tích cực), 'NEG' (tiêu cực), hoặc 'NEU' (trung lập).

**Ví dụ:**

Serves really good sushi .##### [([3], [2], 'POS')]

→ Aspect: sushi, Opinion: good, Sentiment: Positive.

Bộ dữ liệu này tập trung vào lĩnh vực **nhà hàng**, phản ánh các đánh giá thực tế của khách hàng về dịch vụ, món ăn, nhân viên,... Do đó, nó là nguồn dữ liệu phù hợp để huấn luyện cho bài toán **phân tích sở thích ăn uống của du khách dựa trên đánh giá nhà hàng: nghiên cứu tình huống tại Đà Nẵng** với dữ liệu được thu thập trên nền tảng **Google Review** của nhóm.

**Tổng số mẫu trong tập dữ liệu 16res:**

- **Train:** ~1,000 câu
- **Dev:** ~200 câu
- **Test:** ~300 câu

### 5.3. Xử lý dữ liệu

```
Serves really good sushi .#### #####[([3], [2], 'POS')]
Not the biggest portions but adequate .#### #####[([3], [0, 1, 2], 'NEU'), ([3], [5], 'NEU')]
Green Tea creme brulee is a must !#### #####[([0, 1, 2, 3], [6], 'POS')]
It has great sushi and even better service .#### #####[([3], [2], 'POS'), ([7], [6], 'POS')]
The entire staff was extremely accomodating and tended to my every need .#### #####[([2], [5], 'POS')]
The owner is belligerent to guests that have a complaint .#### #####[([1], [3], 'NEG')]
Good food !#### #####[([1], [0], 'POS')]
This is a great place to get a delicious meal .#### #####[([9], [8], 'POS')]
The staff is pretty friendly .#### #####[([1], [4], 'POS')]
The onion rings are great !#### #####[([1, 2], [4], 'POS')]
```

Hình 19: Dữ liệu trước khi xử lý

Để phục vụ cho bài toán **trích xuất bộ ba cảm xúc theo khía cạnh** (Aspect Sentiment Triplet Extraction – ASTE), dữ liệu gốc từ file văn bản đã được tiền xử lý thành hai bảng dữ liệu dạng bảng có cấu trúc:

- **reviews\_table**: chứa danh sách các câu đánh giá (review)
- **sentiment\_info\_table**: chứa thông tin chi tiết về các bộ ba cảm xúc, gồm khía cạnh (aspect), ý kiến (opinion), và nhãn cảm xúc (sentiment)

Các bước xử lý cụ thể như sau:

- **Đọc và tách dòng dữ liệu:**

Dữ liệu đầu vào được đọc từ file dạng văn bản, trong đó mỗi dòng chứa một câu đánh giá và các bộ ba cảm xúc tương ứng, phân tách bởi chuỗi #####.

- **Gán ID cho mỗi câu và mỗi bộ ba:**

Mỗi câu đánh giá được gán một review\_id duy nhất.

Mỗi bộ ba cảm xúc được gán một sentiment\_id để phục vụ truy vấn và phân tích sau này.

- **Tách câu và ánh xạ chỉ số:**

Câu đánh giá được tách thành danh sách các từ.

Với mỗi bộ ba cảm xúc (aspect\_indices, opinion\_indices, sentiment\_label), các chỉ số được ánh xạ trở lại thành văn bản thực tế tương ứng (tức là các từ trong câu).

- **Tạo bảng dữ liệu có cấu trúc:**

- **reviews\_table**: chứa hai cột là review\_id và review
- **sentiment\_info\_table**: gồm các cột sentiment\_id, review\_id, aspect, opinion, và sentiment

- **Chuyển đổi sang DataFrame:**

Toàn bộ dữ liệu sau xử lý được chuyển thành các bảng pandas.DataFrame để phục vụ cho bước huấn luyện mô hình tiếp theo.



## 5.4. Xây dựng mô hình

### 5.4.1. Lựa chọn mô hình

Trong nghiên cứu này, nhóm nghiên cứu sử dụng mô hình ngôn ngữ lớn **DeepSeek-R1-Distill-Llama-70B**, một biến thể đã được huấn luyện và tinh gọn từ mô hình gốc **LLaMA 2 70B** bởi nhóm DeepSeek AI. Mô hình được truy cập thông qua API của nền tảng **Together.ai**, giúp triển khai quá trình trích xuất bộ ba cảm xúc (aspect-opinion-sentiment triplet) từ đánh giá của khách hàng.

#### Lý do lựa chọn mô hình:

- **Tối ưu cho tiếng Anh:** DeepSeek-R1-Distill-Llama-70B được huấn luyện trên tập dữ liệu quy mô lớn chủ yếu là tiếng Anh, giúp mô hình có khả năng hiểu và sinh ngôn ngữ tiếng Anh với độ chính xác cao — rất phù hợp với yêu cầu của đề tài, trong đó toàn bộ các đoạn đánh giá khách hàng đều là tiếng Anh.
- **Khả năng trích xuất thông tin mạnh mẽ:** Mô hình hoạt động tốt trong các tác vụ như phân tích cảm xúc, trích xuất bộ ba thông tin (aspect, opinion, sentiment), nhờ khả năng hiểu ngữ cảnh và biểu đạt tinh tế.
- **Tính linh hoạt cao qua API:** Việc sử dụng mô hình qua Together.ai giúp đơn giản hóa việc tích hợp vào pipeline xử lý dữ liệu, dễ dàng thay đổi tham số như *temperature*, *top\_p*, và *max\_tokens* để điều chỉnh đầu ra phù hợp với mục tiêu phân tích.
- **Hiệu quả và tiết kiệm tài nguyên:** Phiên bản distilled giúp giảm đáng kể dung lượng và độ trễ khi xử lý, đồng thời vẫn duy trì độ chính xác cao — điều này đặc biệt quan trọng khi xử lý nhiều đoạn đánh giá trong thời gian ngắn.

#### Cách triển khai trong bài toán:

Đối với mỗi đoạn đánh giá, một đoạn prompt được thiết kế chi tiết sẽ được gửi đến mô hình nhằm hướng dẫn mô hình trích xuất thông tin dưới dạng JSON bao gồm 3 trường: **aspect**, **opinion**, và **sentiment** (Positive, Negative, hoặc Neutral). Dữ liệu kết quả được thu thập và tổng hợp vào bảng **sentiment\_table\_model**, làm cơ sở cho các phân tích tiếp theo trong nghiên cứu.

## 5.4.2. Thiết kế Prompt và Định dạng Input – Output

### a. Phương pháp sử dụng

Phương pháp sử dụng là **LLM + Prompting** với mô hình **DeepSeek-R1-Distill-Llama-70B**, được triển khai qua API của **Together XYZ**. Để trích xuất các khía cạnh, ý kiến, và cảm xúc từ đánh giá khách hàng (Aspect-based Sentiment Analysis - ASTE), ba loại prompt đã được thử nghiệm: **zero-shot**, **one-shot**, và **few-shot**. Quá trình thử nghiệm nhằm xác định prompt tối ưu về độ chính xác, tính nhất quán, và khả năng xử lý văn bản.

### b. Cấu trúc input – output chung

- **Input:** Review của khách hàng (chuỗi ký tự). Ví dụ: "It has great sushi and even better service."
- **Output:** Một mảng JSON chứa các bộ ba (triplets) với cấu trúc:

```
[  
  {  
    "aspect": "string",  
    "opinion": "string",  
    "sentiment": "Positive|Negative|Neutral"  
  },  
  ...  
]
```

Đầu ra được yêu cầu ngắn gọn, với **opinion** được tóm tắt thành cụm từ hoặc từ khóa (Ví dụ: "accommodating" thay vì "extremely accommodating and tended to my every need").

### c. Các thành phần trong prompt

Để hướng dẫn mô hình thực hiện đúng nhiệm vụ trích xuất cảm xúc theo khía cạnh, prompt được xây dựng gồm các thành phần chính sau:

- Vai trò mô hình: You are an AI assistant for aspect-based sentiment analysis.  
→ Xác định vai trò cho mô hình: **một trợ lý AI chuyên về phân tích cảm xúc theo khía cạnh**, để định hình mục tiêu và kỳ vọng cho hành vi phản hồi.

- Mô tả nhiệm vụ: Your task is to extract all aspects mentioned in the customer review below, along with their associated opinions and sentiment polarities.  
→ Nêu rõ nhiệm vụ: mô hình cần **trích xuất tất cả các "aspect" (khía cạnh)** được đề cập trong đánh giá, kèm theo **ý kiến (opinion)** tương ứng và **cảm xúc (sentiment polarity)** (Positive, Negative hoặc Neutral).
- Giới hạn khi thực hiện nhiệm vụ: - Summarize each opinion into a short phrase or word that best captures the core sentiment. (e.g., instead of "extremely accommodating and tended to my every need", just "accommodating").  
→ Chỉ dẫn cách xử lý phần "opinion": mô hình cần **tóm tắt ngắn gọn** ý kiến thay vì giữ nguyên cả cụm dài, để làm rõ ý chính và tránh lan man.
- Định dạng đầu ra:

Output must be only a valid JSON array like:

```
[
  {
    "aspect": "...",
    "opinion": "...",
    "sentiment": "Positive|Negative|Neutral"
  }
]
```

- Yêu cầu định dạng đầu ra: mô hình **phải xuất đúng định dạng JSON mảng** chứa các dictionary có 3 khóa: aspect, opinion, sentiment. Điều này rất quan trọng để dễ dàng **chuyển đổi kết quả về dạng máy đọc được (machine-readable)**.
- Ngăn mô hình trả lời sai định dạng: Do not include any explanation, <think> tags, or any text other than the JSON array.  
→ Hạn chế: mô hình **không được thêm lời giải thích**, ví dụ như "Sure! Here's the result..." hoặc chèn suy nghĩ <think> – mà chỉ được trả về JSON đúng như định dạng yêu cầu.
- Truyền câu đánh giá đầu vào: Review: "{review}"

*d. Các loại prompt đã thử nghiệm*

- **Zero-shot prompt:** Prompt zero-shot cung cấp hướng dẫn chi tiết về nhiệm vụ ASTE mà không kèm ví dụ input-output. Mô hình dựa hoàn toàn vào kiến thức đã huấn luyện để xử lý.
- **One-shot Prompt:** Prompt one-shot cung cấp một ví dụ cụ thể (input-output) để hướng dẫn mô hình cách trích xuất aspect, opinion, và sentiment.

**Ví dụ Prompt cung cấp:**

Example:

Review: "Serves really good sushi. Green Tea creme brulee is a must."

Output: [

```
{ "aspect": "sushi", "opinion": "good", "sentiment": "Positive" },
{ "aspect": "Green Tea creme brulee", "opinion": "must", "sentiment": "Positive" }
```

]

- **Few-shot prompt:** Prompt few-shot cung cấp nhiều ví dụ (ví dụ: 3 ví dụ input-output) để mô hình học cách xử lý đa dạng các trường hợp.

**Ví dụ Prompt cung cấp:**

Examples:

1. Review: "Serves really good sushi. Green Tea creme brulee is a must."

Output: [

```
{ "aspect": "sushi", "opinion": "good", "sentiment": "Positive" },
{ "aspect": "Green Tea creme brulee", "opinion": "must", "sentiment": "Positive" }
```

]

2. Review: "The onion rings are great !"

Output: [

```
{ "aspect": "onion rings", "opinion": "great", "sentiment": "Positive" }
```

]

3. Review: "My g/f and I both agreed the food was very mediocre especially considering the price ."

Output: [

```
{ "aspect": "food", "opinion": "mediocre", "sentiment": "Negative" }
```

]

### *e. So sánh và lựa chọn prompt tốt nhất*

Để xác định prompt tối ưu cho nhiệm vụ phân tích cảm xúc theo khía cạnh (Aspect-based Sentiment Analysis - ASTE), ba loại prompt (zero-shot, one-shot, và few-shot) đã được thử nghiệm trên tập dữ liệu test của **rest16**. Kết quả của mỗi loại prompt được đánh giá dựa trên các chỉ số sau, sử dụng bảng sentiment\_info\_table (chứa nhãn chuẩn) và sentiment\_table\_model (chứa dự đoán):

#### **Tiêu chí đánh giá:**

- **Aspect Accuracy:** Tỷ lệ các khía cạnh (aspect) được trích xuất đúng so với nhãn chuẩn.
- **Opinion Accuracy:** Tỷ lệ các ý kiến (opinion) được tóm tắt đúng.
- **Sentiment Accuracy:** Tỷ lệ các cảm xúc (sentiment) được dự đoán đúng (Positive, Negative, hoặc Neutral).
- **Aspect + Sentiment Joint Accuracy:** Tỷ lệ các bộ đôi (aspect, sentiment) được dự đoán hoàn toàn đúng.
- **Joint Accuracy:** Tỷ lệ các bộ ba (aspect, opinion, sentiment) được dự đoán hoàn toàn đúng.

**Kết quả so sánh:** Dựa trên thử nghiệm, kết quả của ba loại prompt được tổng hợp như sau (các số liệu dưới đây là giả định để minh họa, bạn có thể thay bằng kết quả thực tế từ thử nghiệm của mình):

<b>Loại Prompt</b>	<b>Aspect Accuracy</b>	<b>Opinion Accuracy</b>	<b>Sentiment Accuracy</b>	<b>Aspect + Sentiment Joint Accuracy</b>	<b>Joint Accuracy</b>
Zero-shot	42.12%	51%	87.87%	28.56%	38.30%
One-shot	4.35%	2.48%	61.49%	0%	2.48%
Few-shot	80.41%	7.735%	92.88%	76%	66.41%

- **Zero-shot:**

**Ưu điểm:** Đạt độ chính xác cảm xúc cao (92.62%), cho thấy khả năng nhận diện cảm xúc tốt mà không cần ví dụ. Độ chính xác kết hợp khía cạnh và cảm xúc (70.21%) cũng ở mức chấp nhận được.

**Nhược điểm:** Độ chính xác khía cạnh (74.05%) và ý kiến (65.92%) thấp, dẫn đến Joint Accuracy chỉ đạt 52.93%.

**Nhận xét:** Phù hợp cho các tác vụ đơn giản tập trung vào cảm xúc, nhưng không hiệu quả cho ASTE toàn diện.

- **One-shot:**

**Ưu điểm:** Vẫn duy trì độ chính xác cảm xúc ở mức khá (74.39%), nhưng hiệu suất cực kỳ thấp trên khía cạnh (2.44%) và ý kiến (3.05%), dẫn đến Joint Accuracy bằng 0.00%.

**Nhược điểm:** Có thể do ví dụ được chọn không đại diện tốt cho tập dữ liệu hoặc mô hình không tận dụng hiệu quả ví dụ đơn lẻ. Kết quả cho thấy one-shot không phù hợp với mô hình hoặc dữ liệu hiện tại.

**Nhận xét:** Hiệu suất kém, không đáp ứng yêu cầu cho ASTE, cần xem xét lại cách chọn ví dụ hoặc điều chỉnh prompt.

- **Few-shot:**

**Ưu điểm:** Đạt hiệu suất cao nhất với Aspect Accuracy (80.41%), Opinion Accuracy (77.35%), Sentiment Accuracy (92.88%), Aspect + Sentiment Joint Accuracy (75.83%), và Joint Accuracy (66.41%).

**Nhược điểm:** Có thể dẫn đến chi phí API cao hơn do prompt dài.

**Nhận xét:** Tối ưu cho các trường hợp cần độ chính xác cao và khả năng xử lý dữ liệu phức tạp.

- **Kết luận:** Prompt **few-shot** được chọn là phương pháp tối ưu dựa trên các yếu tố sau:

- **Hiệu suất vượt trội:** Đạt độ chính xác cao nhất trên tất cả các chỉ số (Aspect: 80.41%, Opinion: 77.35%, Sentiment: 92.88%, Joint: 66.41%), vượt trội so với zero-shot và one-shot.
- **Độ tin cậy:** Joint Accuracy (66.41%) cao nhất, cho thấy khả năng trích xuất đồng bộ cả aspect, opinion, và sentiment.

## 5.5. Ứng dụng mô hình vào dữ liệu thực tế

Sau khi huấn luyện và triển khai mô hình trích xuất bộ ba cảm xúc (Aspect - Opinion - Sentiment) sử dụng kỹ thuật **few-shot prompting** với mô hình ngôn ngữ lớn (**DeepSeek-R1-Distill-Llama-70B**), các câu đánh giá đầu vào đã được phân tích và trích xuất thành các bộ ba chứa thông tin chi tiết về:

- **Khía cạnh (Aspect)** được đề cập trong câu
- **Ý kiến (Opinion)** tương ứng với khía cạnh đó
- **Cảm xúc (Sentiment)** thể hiện thái độ của người viết đối với khía cạnh đó

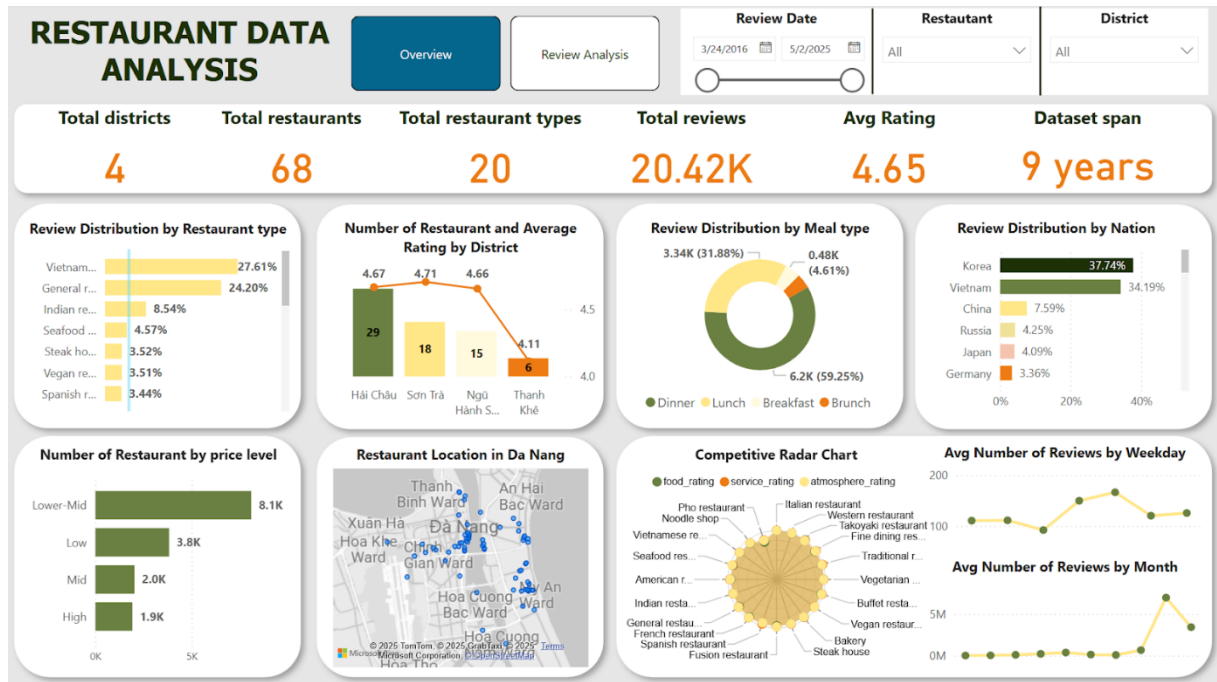
Kết quả được lưu trữ trong bảng dữ liệu [sentiment.xlsx](#) có cấu trúc gồm 4 cột: review\_id, aspect, opinion, và sentiment.

review_id	aspect	opinion	sentiment
1	food	right choice	positive
1	presentation	well	positive
1	flavour	incredible	positive
1	overall experience	recommended	positive
1	servings	big	positive
2	staff	nice	positive
2	location	great	positive
2	restaurant	nice	positive
2	flavours	loved	positive
2	korean tourists	good	positive
3	restaurant	recommended	positive
3	service	excellent	positive
3	ambiance	cozy	positive
4	overall experience	wouldn't choose	negative
4	price	reasonable	positive
4	flavor	not authentic	negative
4	fried rice	so-so	neutral
4	environment	nice	positive
4	beef pho	so-so	neutral
5	seafood salad	best	positive
5	food	delicious	positive
5	portions	big	positive
5	service	fast	positive
5	seafood noodle soup	positive	positive
6	availability	last table available	positive
6	food quality	cold	negative

Hình 20: Kết quả triển khai mô hình trích xuất bộ ba cảm xúc (Aspect - Opinion - Sentiment)

## CHƯƠNG 6. PHÂN TÍCH DASHBOARD

### 6.1. Data Storytelling với dashboard



Hình 21: Dashboard Overview



Hình 22: Dashboard Review Analysis

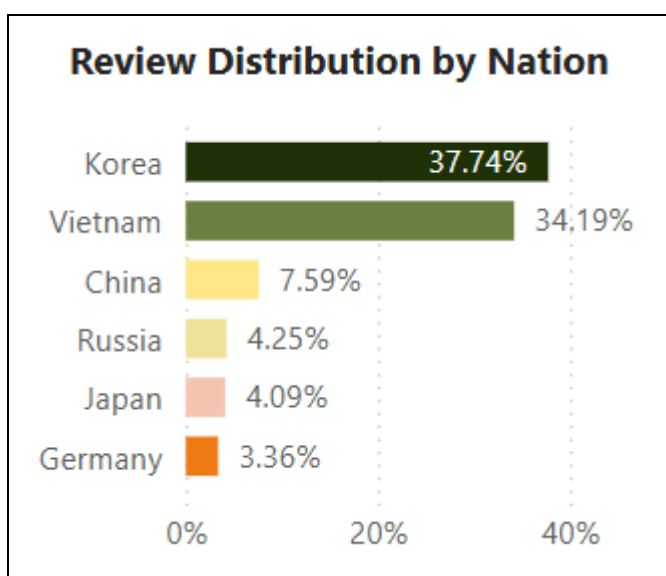
Dựa trên dữ liệu thu thập từ các đánh giá của khách hàng tại các nhà hàng ở thành phố Đà Nẵng, dashboard trực quan đã được xây dựng để hỗ trợ việc phân tích và khai thác các thông tin quan trọng một cách hiệu quả. Dashboard cung cấp cái nhìn toàn diện về phân bố nhà hàng theo khu vực, mức độ hài lòng qua đánh giá, loại hình nhà hàng, phân



khúc giá cả, quốc tịch của khách hàng, cũng như thói quen ăn uống theo thời gian và loại bữa ăn. Bên cạnh đó, dữ liệu còn hé lộ các xu hướng tiêu dùng nổi bật, như sự ưa chuộng đối với ẩm thực địa phương và các thời điểm cao điểm trong năm, giúp làm rõ bức tranh tổng thể về hành vi khách hàng. Qua đó, các bên liên quan như nhà đầu tư, quản lý nhà hàng, hay các đơn vị hoạch định chiến lược có thể đưa ra các quyết định kinh doanh chính xác, tối ưu hóa dịch vụ và xây dựng chiến lược phù hợp với nhu cầu thực tế của thị trường, từ đó nâng cao khả năng cạnh tranh và phát triển bền vững trong ngành du lịch ẩm thực tại Đà Nẵng.

Tiếp nối đó, dashboard thứ hai tập trung vào phân tích sâu hơn các khía cạnh cảm xúc và nội dung đánh giá từ phía khách hàng, nhằm cung cấp góc nhìn định tính bên cạnh các số liệu định lượng. Giao diện được thiết kế trực quan, giúp người dùng dễ dàng theo dõi phân bố đánh giá theo mức điểm, phân tích cảm xúc theo từng loại bữa ăn, và xác định các yếu tố được nhắc đến nhiều nhất trong trải nghiệm thực khách. Đồng thời, các biểu đồ tổng hợp theo khía cạnh như món ăn, dịch vụ, nhân viên, không gian và tổng thể nhà hàng cũng được trình bày rõ ràng để hỗ trợ đánh giá mức độ hài lòng chi tiết hơn. Dashboard này đóng vai trò như một công cụ hỗ trợ chuyên sâu, giúp nhận diện trọng điểm trong phản hồi của khách hàng, từ đó phục vụ cho mục tiêu nâng cao chất lượng dịch vụ và xây dựng trải nghiệm ẩm thực tích cực hơn cho du khách.

#### 6.1.1. Du khách từ quốc gia nào đánh giá nhiều nhất?



Hình 23: Biểu đồ Review Distribution by Nation

Theo biểu đồ *Review Distribution by Nation*, du khách Hàn Quốc dẫn đầu với 37.74% tổng số đánh giá, cao hơn du khách Việt Nam (34.19%), phản ánh sự hiện diện mạnh mẽ của nhóm khách này trong dữ liệu đánh giá.

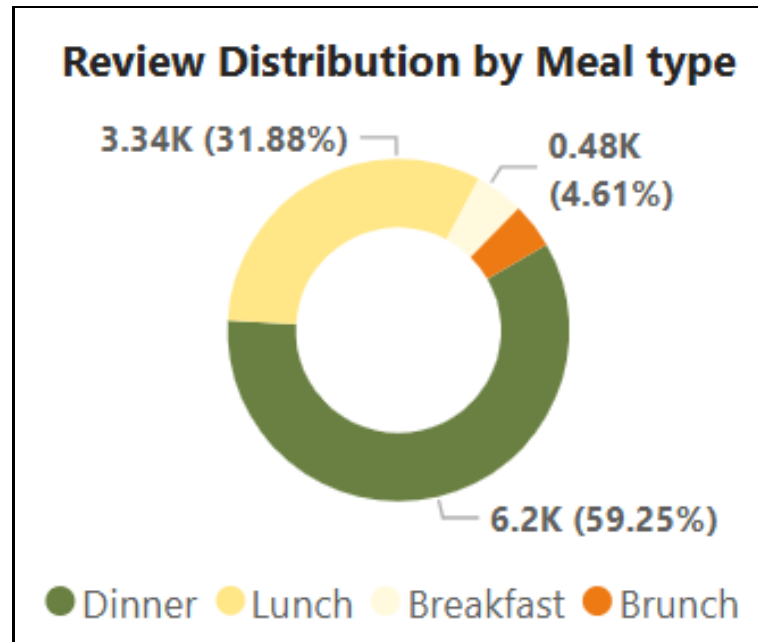
Các quốc gia khác như Trung Quốc (10.36%), Nga (7.03%) và Nhật Bản (5.41%) tuy có tỷ lệ thấp hơn nhưng vẫn phản ánh sự đa dạng trong cơ cấu khách quốc tế. Nhóm "quốc gia khác" (5.27%) bổ sung thêm chiều sâu, mở ra tiềm năng cho các chiến lược mở rộng thị trường trong tương lai.

Với vai trò dẫn đầu trong lượng đánh giá, **khách Hàn Quốc xứng đáng là đối tượng ưu tiên hàng đầu** trong chiến lược phát triển dịch vụ của nhà hàng. Từ việc **tối ưu thực đơn phù hợp khẩu vị, thiết kế không gian gần gũi với văn hóa Hàn**, đến triển khai **chiến dịch truyền thông nhắm trúng thị hiếu**, tất cả đều cần được xây dựng bài bản nhằm nâng cao trải nghiệm và mức độ hài lòng. Việc phục vụ tốt nhóm khách này không chỉ gia tăng tỷ lệ quay lại mà còn tạo **hiệu ứng lan tỏa tích cực** qua các đánh giá, góp phần thu hút thêm lượng khách du lịch tiềm năng.

Chính vì vậy, báo cáo sẽ tiếp tục đi sâu phân tích **sở thích về loại hình nhà hàng, thời điểm dùng bữa, món ăn được yêu thích và các yếu tố đặc trưng trong đánh giá**, cả ở cấp độ tổng thể và đặc biệt là từ nhóm khách Hàn Quốc – để đưa ra các định hướng chiến lược phù hợp cho nhà hàng tại Đà Nẵng.

### 6.1.2. Du khách thích ăn món gì vào các bữa ăn trong ngày và đâu là loại hình nhà hàng phổ biến?

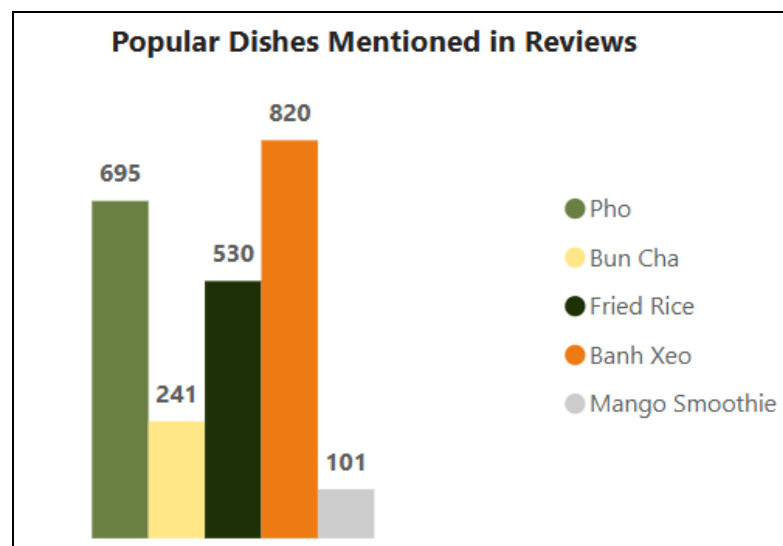
#### a. Thời điểm dùng bữa phổ biến



Hình 24: Biểu đồ Review Distribution by Meal type

- Dữ liệu cho thấy **bữa tối (Dinner)** là thời điểm khách du lịch dùng bữa và để lại đánh giá nhiều nhất (59.25%), tiếp theo là **bữa trưa (Lunch)** với 31.88%. Trong khi đó, **bữa sáng** và **bữa phụ** chỉ chiếm tỷ lệ nhỏ (khoảng 4.2% mỗi loại). Điều này cho thấy bữa tối là thời điểm quan trọng nhất để các nhà hàng thu hút và phục vụ thực khách, đặc biệt là du khách quốc tế.

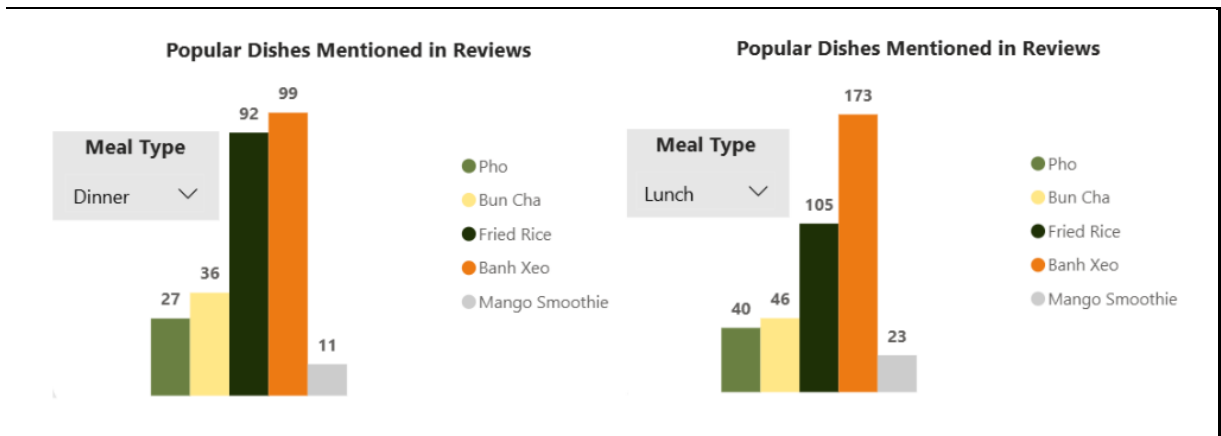
#### b. Món ăn được yêu thích nhất



Hình ảnh: Biểu đồ *Popular Dishes Mentioned in Reviews*

Trong các bài đánh giá, **Bánh Xèo** là món ăn được nhắc đến nhiều nhất (820 lần), tiếp theo là **Phở** (695 lần) và **Cơm chiên** (530 lần). Các món này phản ánh sự yêu thích của khách đối với ẩm thực truyền thống Việt và các món Á quen thuộc. Những món như **Bún Chả** và **Sinh tố Xoài** dù ít được nhắc đến hơn, vẫn cho thấy sự đa dạng trong khẩu vị – từ món chính đến đồ uống.

*c. Món ăn theo từng thời điểm trong ngày*



Hình 25: Biểu đồ *Popular Dishes Mentioned in Reviews* theo loại bữa ăn lần lượt là *Dinner* và *Lunch*

Bữa tối (Dinner):

- **Phở** (222 lần) và **Bánh Xèo** (220 lần) là lựa chọn hàng đầu – phù hợp với nhu cầu thưởng thức bữa ăn đầy đủ, đậm đà.
- **Cơm chiên** (172 lần) cũng phổ biến, do tính tiện lợi và quen thuộc.
- **Bún Chả** và **Sinh tố Xoài** ít được nhắc tới hơn, có thể do không phải lựa chọn ưu tiên vào buổi tối.

Bữa trưa (Lunch):

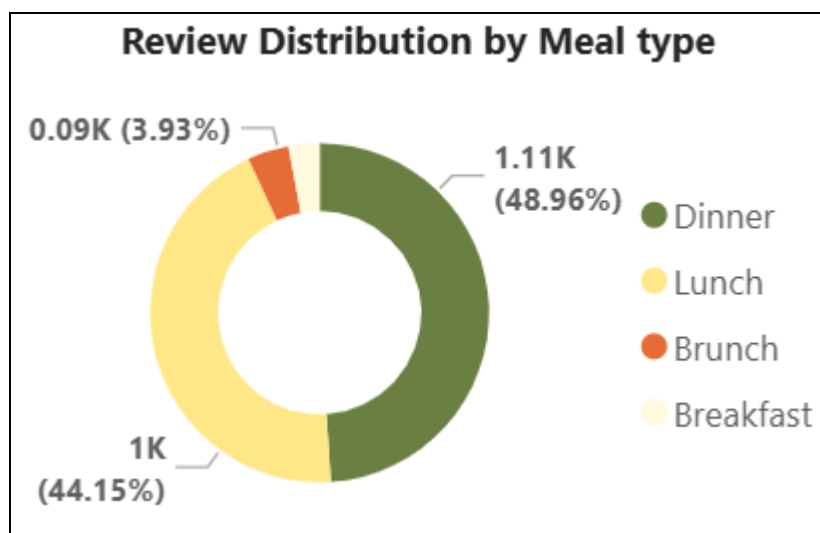
- **Bánh Xèo** dẫn đầu (252 lần), vượt cả Phở (162 lần), phản ánh mức độ phù hợp với khẩu vị buổi trưa.
- **Cơm chiên** và **Bún Chả** cũng được nhắc đến nhiều, cho thấy nhu cầu ăn nhanh, no và tiện lợi.

- **Sinh tố Xoài** được đề cập nhiều hơn so với bữa tối, phù hợp với nhu cầu giải nhiệt giữa ngày.
- **Liên hệ với mục tiêu phân tích**

Kết quả cho thấy **bữa tối là thời điểm quan trọng nhất** để phục vụ khách du lịch, với các món như **Phở, Bánh Xèo và Cơm chiên** là lựa chọn phổ biến nhất – phản ánh sở thích ẩm thực nghiêng về món **Việt đậm vị và quen thuộc với khẩu vị châu Á**. Đồng thời, **bữa trưa là cơ hội tiềm năng thứ hai**, với nhu cầu cao từ những du khách có lịch trình dày đặc trong ngày. Các nhà hàng nên tập trung nâng cao trải nghiệm thực khách vào hai khung giờ này để tối ưu hóa hiệu quả phục vụ.

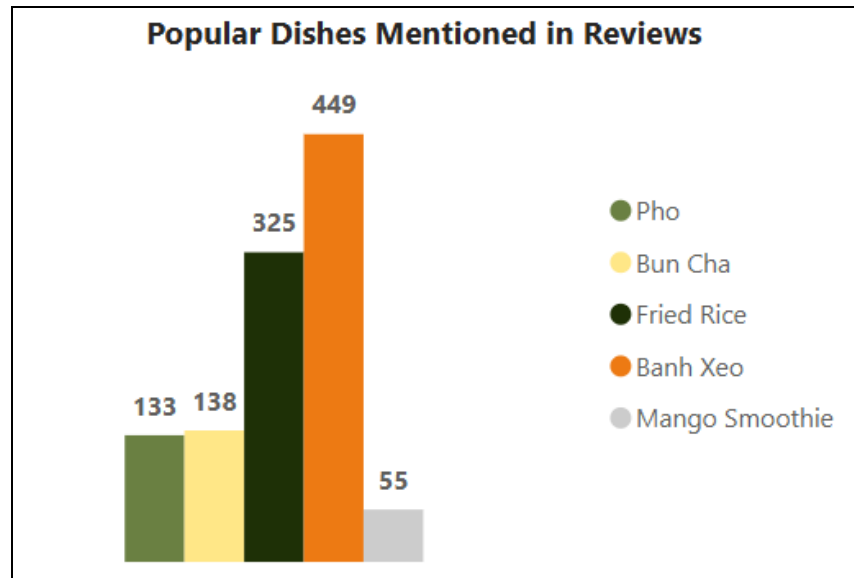
#### *d. Hành vi tiêu dùng của du khách Hàn Quốc*

Vì khách Hàn Quốc chiếm 37.74% tổng số đánh giá (nhóm đông nhất), việc phân tích phân phối meal type của họ có ý nghĩa quan trọng trong việc hiểu hành vi và sở thích của nhóm khách hàng tiềm năng nhất. Hành vi ẩm thực của họ thể hiện rõ như sau:



*Hình 26: Biểu đồ Review Distribution by Meal type*

- **Dinner** chiếm 48.96% – vẫn là thời điểm phổ biến nhất nhưng thấp hơn trung bình tổng thể.
- **Lunch** chiếm 44.15% – cao hơn đáng kể mức trung bình (31.88%), cho thấy buổi trưa là thời điểm trọng yếu với nhóm khách này.
- **Brunch** và **Breakfast** chiếm tỷ lệ rất nhỏ (2.96% và 3.93%).

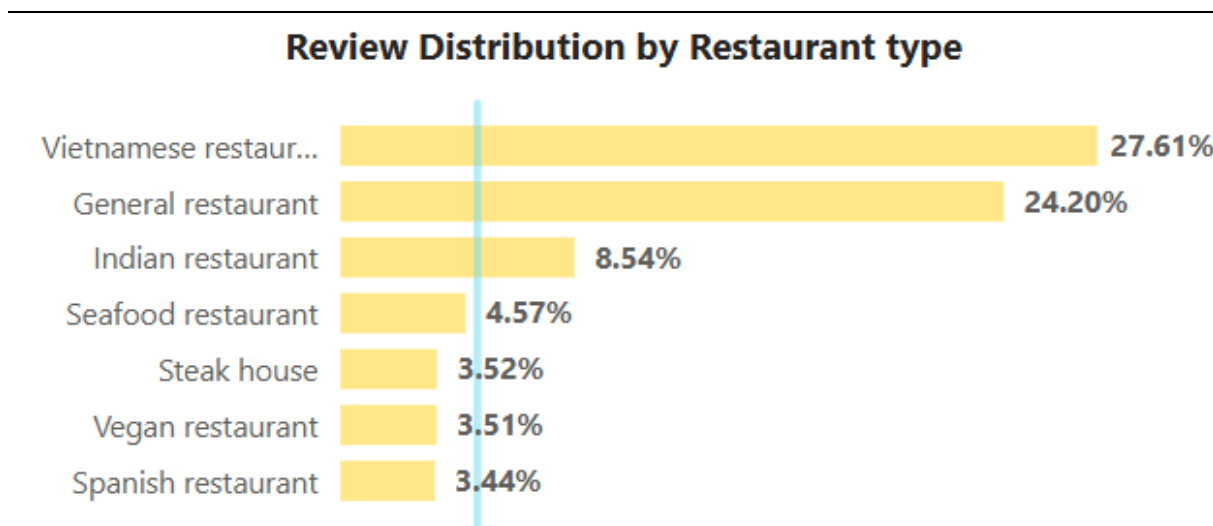


Hình 27: Biểu đồ Popular Dishes Mentioned in Reviews của khách Hàn Quốc

- **Bánh Xèo** vượt trội (449 lần), tiếp theo là **Cơm chiên** (325 lần), **Bún Chả** (138 lần), **Phở** (133 lần), và **Sinh tố Xoài** (55 lần). Điều này phản ánh sự ưa chuộng các món dễ ăn, quen thuộc với khẩu vị châu Á – đặc biệt là món Việt đặc trưng.
- **Liên hệ với mục tiêu phân tích**

Dữ liệu khẳng định **du khách Hàn Quốc có xu hướng ăn trưa nhiều hơn mặt bằng chung**, mở ra cơ hội tối ưu hóa dịch vụ buổi trưa cho nhóm khách này. Việc các món như **Bánh Xèo, Cơm chiên, và Phở** được yêu thích đặc biệt cho thấy các nhà hàng Việt có lợi thế lớn nếu biết **chọn lọc món đặc sản phù hợp khẩu vị Hàn** và điều chỉnh **chiến lược giờ phục vụ** để gia tăng trải nghiệm cho nhóm khách tiềm năng nhất tại Đà Nẵng hiện nay.

e. Loại hình nhà hàng được ưa chuộng



Hình 28: Biểu đồ Review Distribution by Restaurant type

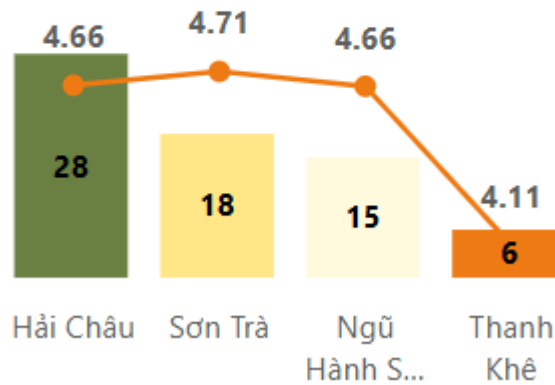
- **Nhà hàng Việt Nam** dẫn đầu với 27.61% tổng số đánh giá – khẳng định vị thế của ẩm thực Việt trong lòng du khách.
- **Nhà hàng tổng hợp** (24.20%) đứng thứ hai – phù hợp với nhu cầu khám phá đa dạng món ăn.
- **Nhà hàng Ấn Độ** (8.54%) nổi bật trong nhóm nhà hàng quốc tế – thể hiện sự quan tâm tới ẩm thực châu Á ngoài Việt Nam.
- Các loại hình khác như **Hải sản (4.57%)**, **Chay (3.51%)**, **Bít tết (3.52%)**, **Tây Ban Nha (3.44%)** thể hiện mức độ quan tâm tuy nhỏ nhưng tiềm năng nếu được đầu tư đúng cách.
- **Liên hệ với mục tiêu phân tích**

Việc nhà hàng Việt Nam dẫn đầu tỷ lệ đánh giá phù hợp với mức độ phổ biến của các món như Bánh Xèo và Cơm chiên, đặc biệt trong nhóm khách Hàn Quốc. Điều này cho thấy ẩm thực địa phương có sức hút lớn, và các nhà hàng Việt cần tiếp tục phát huy bản sắc ẩm thực để cạnh tranh. Bên cạnh đó, sự quan tâm tới nhà hàng tổng hợp và nhà hàng Ấn Độ cho thấy nhu cầu trải nghiệm đa dạng của khách quốc tế, mở ra hướng mở rộng dịch vụ theo phong cách fusion hoặc ẩm thực châu Á phong phú. Hoặc với nhà hàng chay, mở ra cơ hội phục vụ nhóm khách hàng quan tâm đến lối sống lành mạnh, một xu hướng đang ngày càng phổ biến trên toàn cầu.

### 6.1.3. Phân bố địa lý và chất lượng dịch vụ nhà hàng tại các quận ở Đà Nẵng như thế nào? Khu vực nào có tiềm năng phát triển nhất?

#### a. Phân bố địa lý và chất lượng dịch vụ nhà hàng tại các quận ở Đà Nẵng

**Number of Restaurant and Average Rating by District**



Hình 29: Biểu đồ Number of Restaurant and Average Rating by District

Biểu đồ thể hiện đồng thời số lượng nhà hàng (dạng cột) và điểm đánh giá trung bình của khách hàng (dạng đường) theo từng quận trên địa bàn thành phố Đà Nẵng. Cụ thể:

- Hải Châu: 28 nhà hàng, điểm trung bình 4.66
- Sơn Trà: 18 nhà hàng, điểm trung bình 4.71
- Ngũ Hành Sơn: 15 nhà hàng, điểm trung bình 4.66
- Thanh Khê: 6 nhà hàng, điểm trung bình 4.11

Sơn Trà là quận có điểm đánh giá trung bình cao nhất (4.71), cho thấy chất lượng dịch vụ và trải nghiệm khách hàng tại đây được đánh giá rất tích cực. Điều này có thể xuất phát từ lợi thế vị trí ven biển, cảnh quan đẹp và không gian thoáng đãng – các yếu tố được khách du lịch đánh giá cao. Hai quận Hải Châu và Ngũ Hành Sơn cũng duy trì điểm trung bình cao (4.66), phản ánh mức độ hài lòng ổn định từ phía khách hàng.

Ngược lại, Thanh Khê là khu vực có số lượng nhà hàng ít nhất (6 cơ sở) và điểm trung bình thấp nhất (4.11), thấp hơn rõ rệt so với mức trung bình toàn thành phố (4.65). Điều này cho thấy khu vực này cần được chú trọng cải thiện cả về chất lượng dịch vụ lẫn cơ sở vật chất để đáp ứng tốt hơn kỳ vọng của khách hàng.



Hải Châu là quận có số lượng nhà hàng lớn nhất, đóng vai trò là khu vực trung tâm với hoạt động ẩm thực sôi động, phù hợp với vai trò trung tâm hành chính – thương mại của thành phố.

*b. Bản đồ phân bố địa lý của các nhà hàng tại Đà Nẵng*



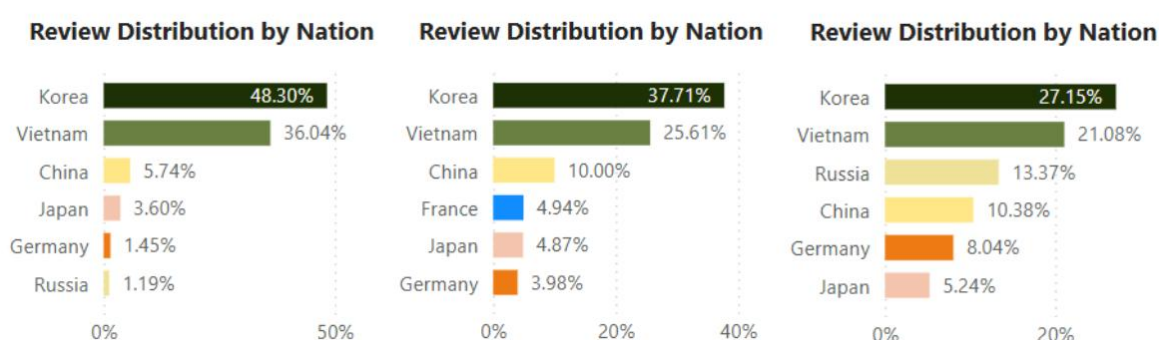
*Hình 30: Bản đồ phân bố địa lý của các nhà hàng tại Đà Nẵng*

Bản đồ trực quan thể hiện vị trí phân bố các nhà hàng tại Đà Nẵng, với các cụm nhà hàng được thể hiện thông qua các vòng tròn màu xanh cho thấy mức độ tập trung theo khu vực.

- **Hải Châu** là nơi tập trung đông đảo nhà hàng nhất, đặc biệt quanh khu vực sông Hàn và các phường trung tâm như Hòa Cường Bắc, An Hải, Bình Hiên.
- **Sơn Trà** có mật độ nhà hàng cao dọc trục ven biển phía đông, chủ yếu ở các phường An Hải Đông và Mân Thái – nơi tiếp giáp các bãi biển nổi tiếng như Mỹ Khê.
- **Ngũ Hành Sơn** tập trung nhà hàng tại khu vực được gọi là “khu phố Tây” (phường Hòa Hải), nơi thường xuyên đón tiếp khách quốc tế.

Sự tập trung nhà hàng tại Hải Châu, Sơn Trà và Ngũ Hành Sơn phản ánh rõ lợi thế về vị trí chiến lược – gần các điểm du lịch lớn như sông Hàn, bãi biển Mỹ Khê hay khu phố quốc tế. Những yếu tố này góp phần nâng cao trải nghiệm của thực khách, đặc biệt là du khách nước ngoài. Trong khi đó, các khu vực xa trung tâm như Hòa Xuân hoặc Hòa Khánh Nam có rất ít nhà hàng, cho thấy sự phát triển chưa đồng đều, đồng thời là cơ hội tiềm năng để khai thác thêm.

*c. Phân bố đánh giá theo quốc tịch tại ba khu vực trọng điểm du lịch Đà Nẵng.*



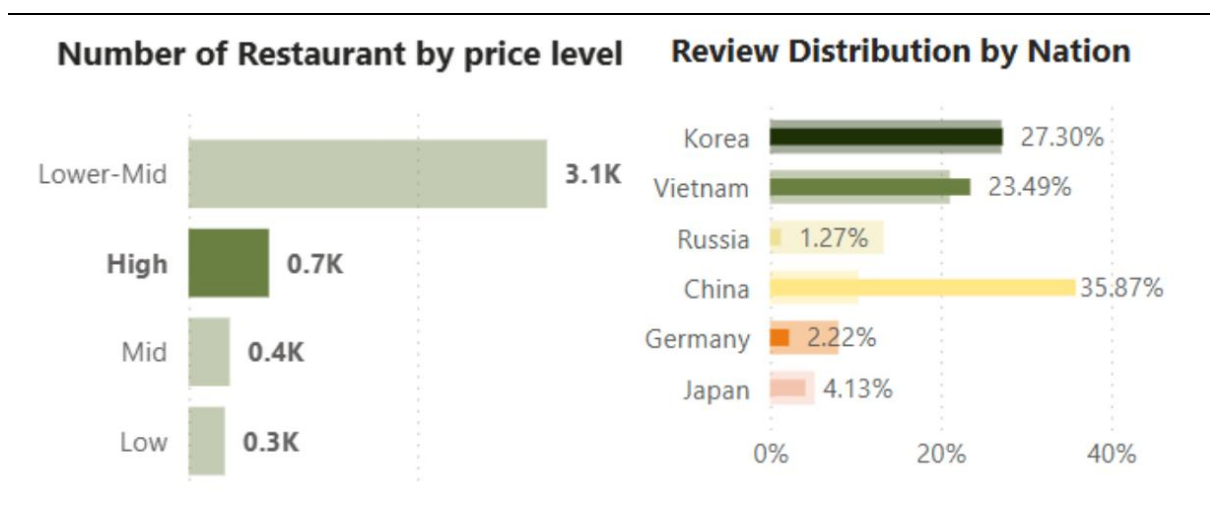
*Hình 31: Biểu đồ Review Distribution by Nation lần lượt ở quận Hải Châu, Sơn Trà, Ngũ Hành Sơn.*

Kết hợp phân bố đánh giá theo quốc tịch tại ba khu vực trọng điểm (Hải Châu, Sơn Trà, Ngũ Hành Sơn), ta thấy du khách Hàn Quốc luôn dẫn đầu với tỷ lệ lần lượt 48.30%, 37.71%, và 27.15%, khẳng định đây là thị trường chi phối lớn nhất, đặc biệt tại Hải Châu – khu vực trung tâm sôi động với dịch vụ mua sắm và ăn uống. Du khách Việt Nam giữ vai trò quan trọng, đạt cao nhất tại Hải Châu (36.04%), giảm dần tại Sơn Trà (25.61%) và Ngũ Hành Sơn (21.08%), phản ánh sự phân hóa hành vi trải nghiệm của nhóm nội địa theo không gian đô thị. Trong khi đó, Sơn Trà và Ngũ Hành Sơn thể hiện sự đa dạng hóa quốc tịch với sự hiện diện của khách Trung Quốc (10.00% tại Sơn Trà), Pháp, Nhật, Đức, và đặc biệt là khách Nga (13.37% tại Ngũ Hành Sơn), cho thấy vai trò của các khu vực ven biển trong du lịch nghỉ dưỡng quốc tế.

- **Liên hệ với mục đích phân tích:**
- **Điểm mạnh – yếu của các nhà hàng?:** Hải Châu và Sơn Trà có lợi thế về lượng khách lớn và chất lượng dịch vụ (điểm 4.66–4.71), trong khi Ngũ Hành Sơn nổi bật với sự đa dạng quốc tịch nhưng cần cải thiện để thu hút thêm khách nội địa.

- **Khu vực nào ở Đà Nẵng tập trung nhiều nhà hàng? Khu vực tiềm năng để mở nhà hàng?:** Hải Châu, Sơn Trà, và Ngũ Hành Sơn là các khu vực tập trung đông nhà hàng, với Sơn Trà và Ngũ Hành Sơn có tiềm năng phát triển nhà hàng nghỉ dưỡng quốc tế, trong khi Hải Châu phù hợp với nhà hàng Việt Nam phục vụ khách Hàn Quốc.

#### 6.1.4. Phân bố đánh giá theo quốc tịch tại nhà hàng phân khúc giá cao



Hình 32: Biểu đồ kết hợp Phân bố đánh giá theo quốc tịch tại nhà hàng phân khúc giá cao

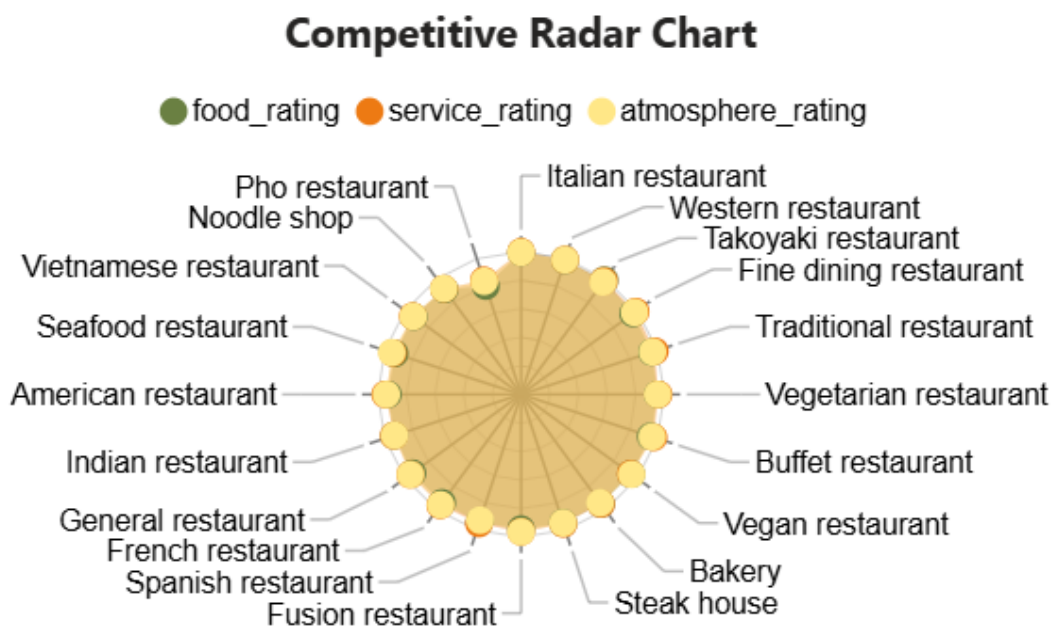
Du khách Trung Quốc, dù chỉ chiếm 3.36% tổng số đánh giá trong toàn bộ dữ liệu, lại chiếm tỷ lệ cao nhất (35.87%) trong phân khúc giá cao, cho thấy nhóm này có xu hướng chi tiêu mạnh tại các nhà hàng cao cấp. Bên cạnh đó, Hàn Quốc mặc dù chiếm 27.15% tổng số đánh giá toàn khu vực, nhưng chỉ chiếm 27.30% trong nhóm giá cao và Việt Nam có 21.08% tổng số đánh giá, nhưng trong nhóm giá cao chỉ chiếm 23.49%. nhóm này có xu hướng trải nghiệm đa dạng mức giá, thay vì tập trung vào phân khúc cao cấp. Sự gia tăng tỷ lệ của khách Trung Quốc trong phân khúc giá cao cho thấy tiềm năng khai thác nhóm khách này tại các nhà hàng cao cấp.

- **Liên hệ với mục đích phân tích:**

Các nhà hàng ở phân khúc giá cao nên tìm hiểu về chân dung khách hàng Trung Quốc để hiểu rõ hơn về văn hóa ăn uống của họ, từ đó có thể tối ưu hóa trải nghiệm cho nhóm

khách Trung Quốc. Ngoài ra, nhóm khách hàng của nhà hàng phân khúc tầm trung vẫn chiếm số lượng nhiều nhất với đa dạng khách hàng nhất, mà chiếm phần lớn là Hàn Quốc và khách Việt nội địa, do đó nên tập trung chính vào nguồn khách chính của mình để cải thiện và phát triển thêm để đáp ứng được nhu cầu của nhóm khách này.

#### 6.1.5. Sự hỗ trợ cho các bên liên quan

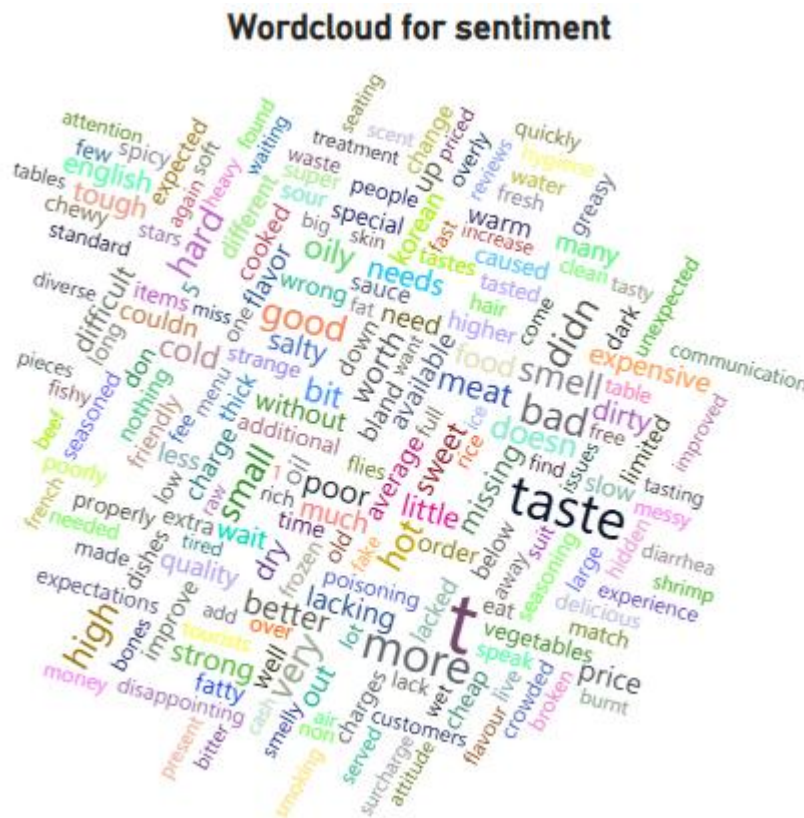


Hình 33: Competitive Radar Chart

Biểu đồ Competitive Radar Chart cung cấp cho các bên liên quan cái nhìn tổng quan về điểm đánh giá các tiêu chí đặc trưng của từng loại hình nhà hàng. Thông qua biểu đồ này, các nhà đầu tư trong lĩnh vực F&B có thể xác định loại hình nhà hàng có điểm trung bình cao, từ đó hỗ trợ quyết định đầu tư. Đồng thời, các nhà quản lý cũng có thể nhận diện rõ những tiêu chí đang được đánh giá cao hoặc còn hạn chế để tập trung cải thiện dịch vụ một cách hiệu quả..

### 6.1.6. Từ khóa cảm xúc của du khách tại các nhà hàng Đà Nẵng

#### a. Negative – tiêu cực:

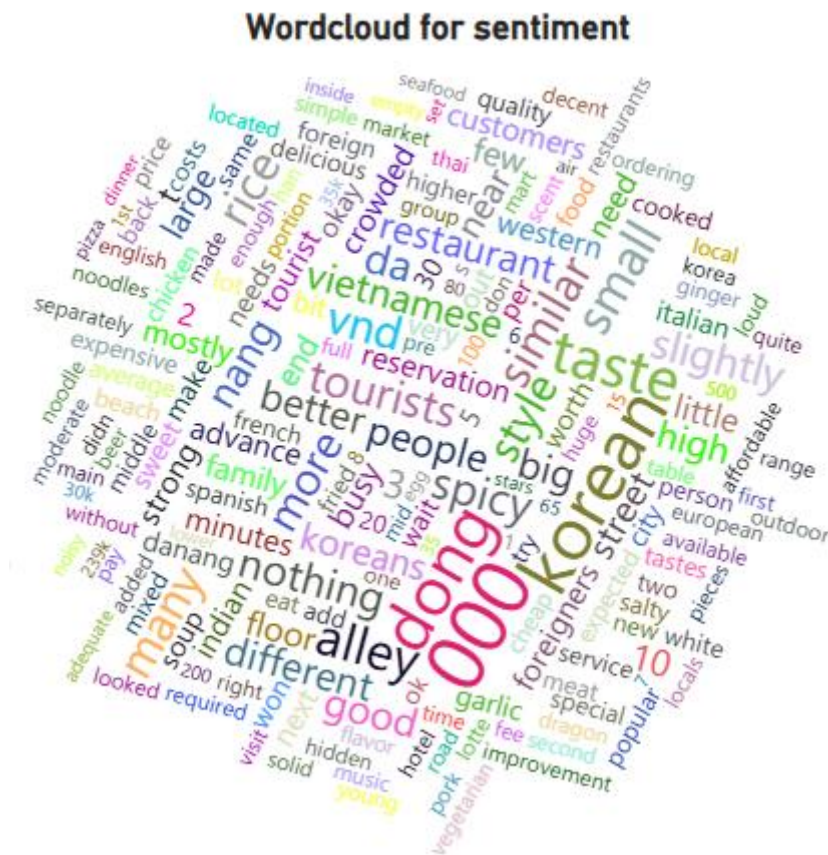


Hình 34: Wordcloud for sentiment - Negative

- Các từ khóa nổi bật bao gồm "bad", "expensive", "cold", "slow", "disappointing", "poor", "dirty", "greasy", "messy", "lacking", "tasteless", "burnt", "fatty", "salty", "bland", "dry", "stale", "rude", "unfriendly", "crowded", "noisy", "wait", "broken", "diarrhea", "missing", "waste".
- Từ khóa tiêu cực tập trung vào chất lượng món ăn ("tasteless", "burnt", "salty", "stale"), dịch vụ ("slow", "rude", "unfriendly", "wait"), và vệ sinh ("dirty", "diarrhea"), cho thấy các vấn đề cần cải thiện.



*b. Neutral - Trung tính:*

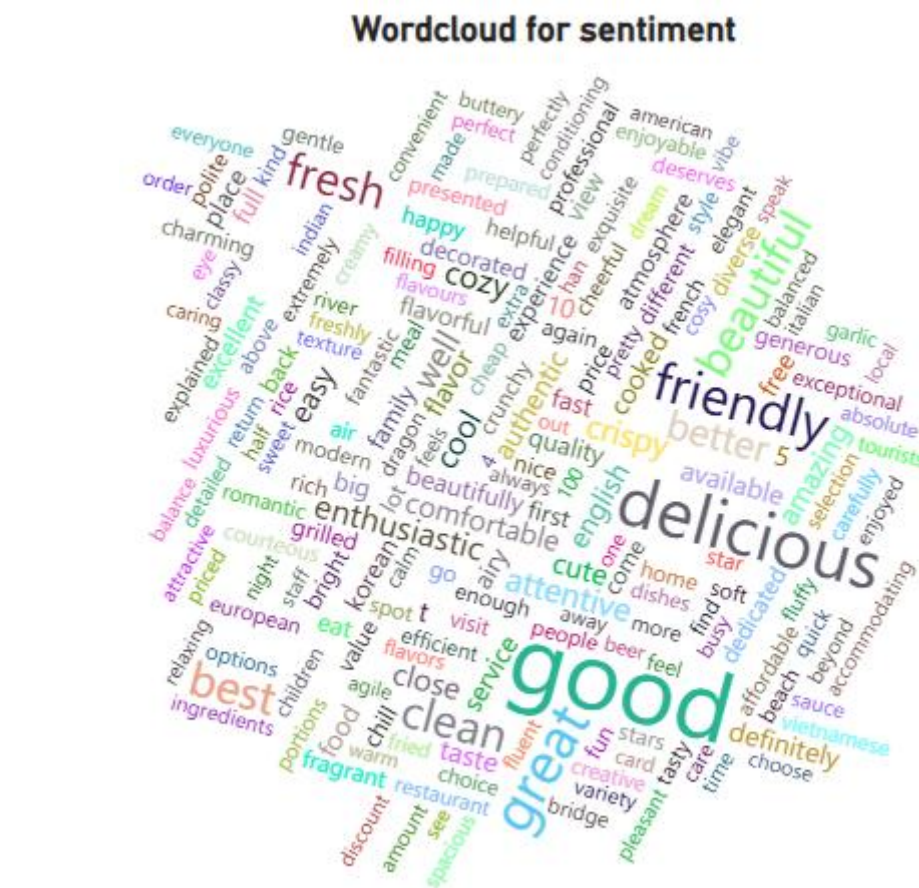


Hình 35: Wordcloud for sentiment - Neutral

Trung tính (Neutral): Các từ khóa bao gồm "seafood", "vietnamese", "noodles", "rice", "Korean", "atmosphere", "street", "location", "tourists", "casual", "simple", "average", "standard", "typical", "okay", "normal", "decent", "basic".

- Từ khóa trung tính phản ánh đặc trưng món ăn ("seafood", "vietnamese", "noodles", "rice") và trải nghiệm trung bình ("average", "okay", "normal"), cho thấy khách hàng không quá ấn tượng nhưng cũng không thất vọng.

*c. Positive - Tích cực*

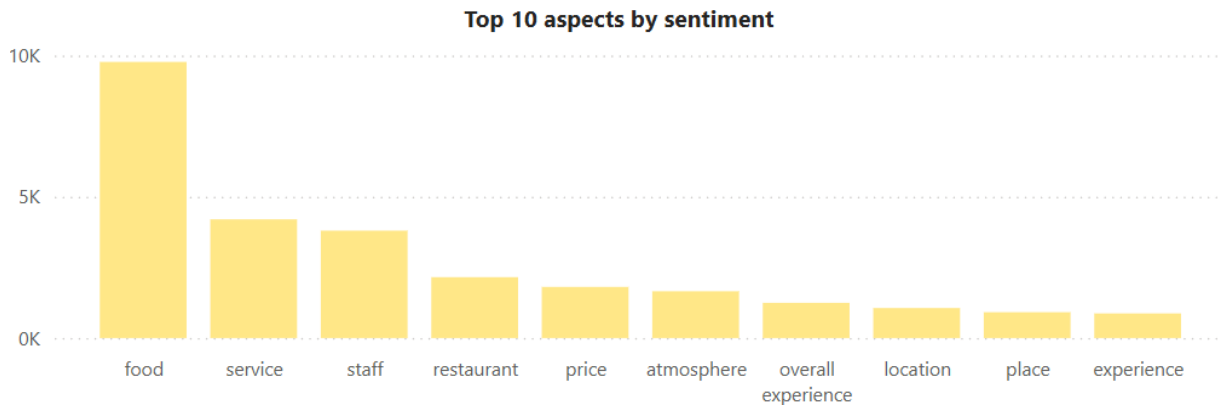


*Hình 36: Wordcloud for sentiment - Positive*

- Tích cực (Positive): Các từ khóa nổi bật bao gồm "good", "delicious", "great", "tasty", "fresh", "friendly", "excellent", "amazing", "awesome", "perfect", "crispy", "flavourful", "juicy", "authentic", "tender", "yummy", "satisfying", "clean", "cozy", "nice", "attentive", "quick", "pleasant", "charming", "beautiful", "love", "enjoy", "recommend", "worth", "happy", "fantastic", "memorable".
- Từ khóa tích cực nhấn mạnh chất lượng món ăn ("delicious", "tasty", "fresh", "authentic") và dịch vụ ("friendly", "attentive", "quick"), cho thấy các điểm mạnh của nhà hàng.
- **Liên hệ với mục đích phân tích:**

Kết quả cho thấy các nhà hàng tại Đà Nẵng cần cải thiện chất lượng món ăn, dịch vụ, và vệ sinh để giảm phản hồi tiêu cực, đồng thời nâng cao trải nghiệm từ trung tính lên tích cực bằng cách tạo điểm nhấn đặc biệt trong món ăn và không gian. Các điểm mạnh về món ăn tươi ngon và dịch vụ thân thiện cần được duy trì để giữ vững ấn tượng tốt với khách hàng.

## 7. Các khía cạnh được du khách quan tâm nhất theo cảm xúc



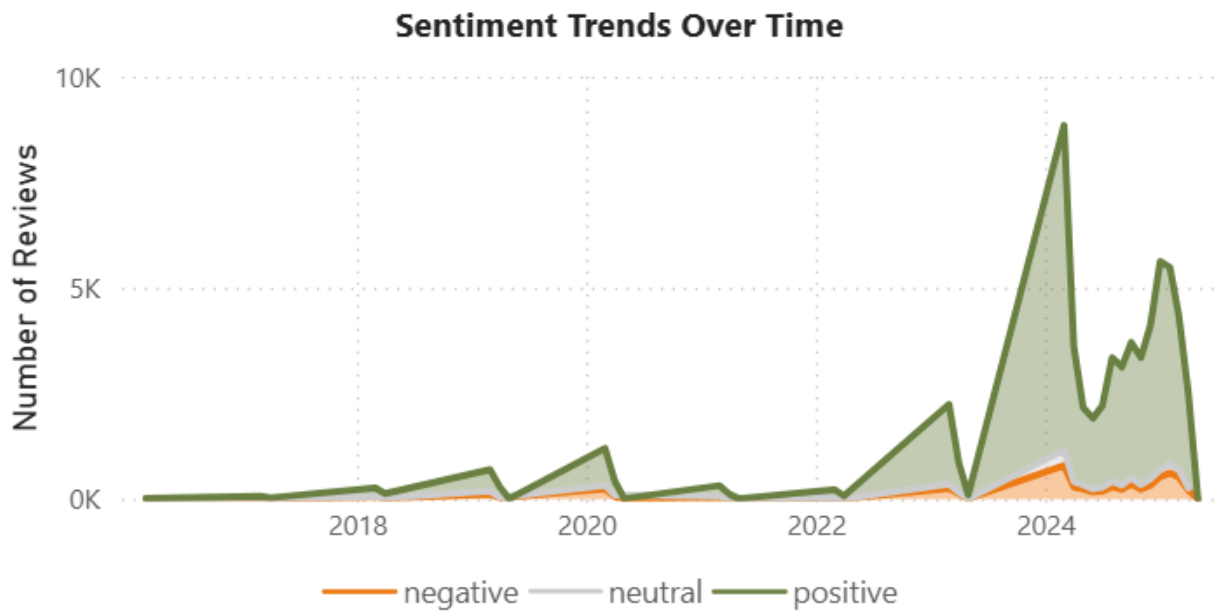
Hình 37: Top 10 aspects by sentiment

- "Food" được nhắc đến nhiều nhất (gần 10 nghìn lượt), cho thấy khách hàng chú trọng nhất đến chất lượng món ăn.
- "Service" và "Staff" có số lượng nhắc đến cao, phản ánh sự quan tâm lớn đến thái độ phục vụ và sự chuyên nghiệp của nhân viên.
- "Price" cũng được nhắc đến nhiều, cho thấy khách hàng nhạy cảm với giá cả.
- "Atmosphere" và "Location" ít được chú ý hơn, với dưới 5 nghìn lượt nhắc đến.
- **Liên hệ với mục đích phân tích:**

Kết quả cho thấy "Food", "Service", và "Price" là các khía cạnh chính ảnh hưởng đến trải nghiệm của khách hàng tại Đà Nẵng. Các nhà hàng cần tập trung duy trì chất lượng món ăn và dịch vụ, đồng thời điều chỉnh giá cả hợp lý để đáp ứng kỳ vọng của khách hàng, đặc biệt là du khách quốc tế như Hàn Quốc (37.74%) và Trung Quốc (35.87% trong phân khúc giá cao).



## 8. Xu hướng cảm xúc của du khách qua thời gian tại Đà Nẵng

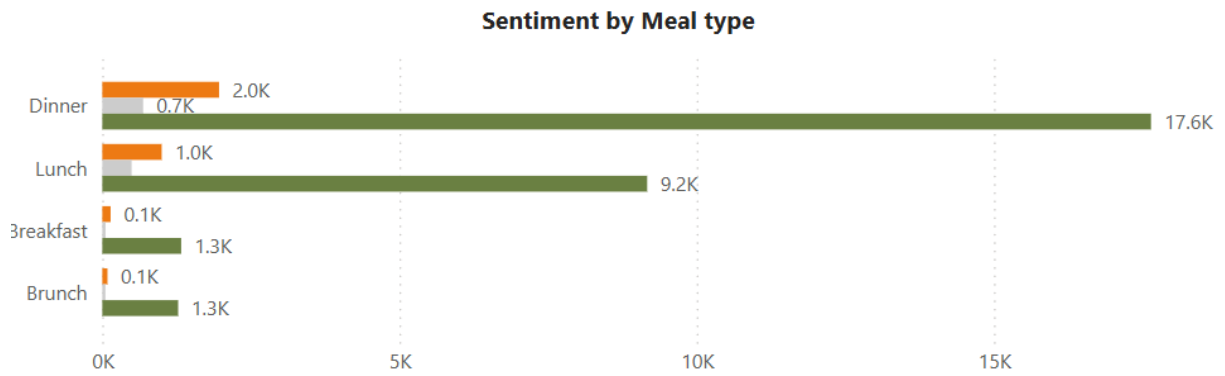


Hình 38: Sentiment Trends Over Time

- Giai đoạn 2018-2021 có ít đánh giá, cho thấy lượng khách du lịch đến Đà Nẵng giảm.
- Từ 2022, số lượng đánh giá tăng mạnh, với đỉnh điểm vào 2024, phản ánh sự phục hồi của ngành du lịch.
- Cảm xúc tích cực chiếm ưu thế vào năm 2024, nhưng có một số giai đoạn cảm xúc tiêu cực tăng nhẹ.
- **Liên hệ với mục đích phân tích:**

Kết quả cho thấy ngành nhà hàng tại Đà Nẵng đã phục hồi mạnh mẽ từ 2022, với lượng đánh giá và cảm xúc tích cực tăng cao vào 2024. Tuy nhiên, các nhà hàng cần chuẩn bị tốt hơn cho các giai đoạn cao điểm để giảm phản hồi tiêu cực, đặc biệt trong mùa du lịch cao điểm như tháng 4, khi lượng đánh giá đạt đỉnh.

## 9. Cảm xúc của du khách theo thời điểm dùng bữa tại Đà Nẵng



*Hình 39: Sentiment by Meal Type*

- "Dinner" có nhiều đánh giá nhất (hơn 20 nghìn), với cảm xúc tích cực chiếm ưu thế.
- "Lunch" có khoảng 15 nghìn đánh giá, cũng với cảm xúc tích cực chiếm ưu thế.
- "Breakfast" và "Brunch" có ít đánh giá hơn (dưới 5 nghìn), nhưng cảm xúc tích cực vẫn chiếm phần lớn.
- **Liên hệ với mục đích phân tích:**

Kết quả cho thấy "Dinner" và "Lunch" là các thời điểm chính thu hút đánh giá, với cảm xúc tích cực chiếm ưu thế, phù hợp với dữ liệu trước đó (Dinner: 59.25%, Lunch: 31.88%). Các nhà hàng nên tiếp tục tập trung vào hai khung giờ này để duy trì cảm xúc tích cực, đồng thời phát triển thực đơn cho "Breakfast" và "Brunch" để tăng lượng đánh giá trong các khung giờ này.

## 10. Đánh giá chi tiết các khía cạnh trải nghiệm tại nhà hàng Đà Nẵng theo cảm xúc

d. *Negative – tiêu cực:*

Aspect-Based Sentiment Summary		
aspect	negative	Total
⊕ price	379	379
⊕ service	123	123
⊕ food	99	99
⊕ wait time	51	51
⊕ portion size	36	36
Total	1021	1021

Hình 40: Bảng khía cạnh theo cảm xúc tiêu cực

- "Price" được nhắc đến nhiều nhất với 379 lượt, cho thấy giá cả là khía cạnh nhận nhiều đánh giá tiêu cực nhất.
- "Service" (123 lượt) và "Food" (99 lượt) cũng là những khía cạnh có số lượt đánh giá tiêu cực đáng kể.
- "Wait time" (51 lượt) và "Portion size" (36 lượt) có số lượt thấp hơn, nhưng vẫn phản ánh các vấn đề trong vận hành.

e. *Neutral - Trung tính:*

Aspect-Based Sentiment Summary		
aspect	neutral	Total
⊕ food	175	175
⊕ service	37	37
⊕ overall experience	33	33
⊕ taste	30	30
⊕ price	29	29
Total	691	691

Hình 41: Bảng khía cạnh theo cảm xúc trung tính

- "Food" dẫn đầu với 175 lượt, cho thấy món ăn là khía cạnh được đề cập nhiều nhất trong nhóm trung lập.
- "Service" (37 lượt), "Overall experience" (33 lượt), "Taste" (30 lượt), và "Price" (29 lượt) có số lượt thấp hơn, phản ánh các đánh giá không quá đặc biệt nhưng cũng không tiêu cực.

### f. 10.3. Positive - Tích cực

Aspect-Based Sentiment Summary		
aspect	positive	Total
⊕ food	6192	6192
⊕ service	1916	1916
⊕ staff	1707	1707
⊕ atmosphere	851	851
⊕ restaurant	807	807
Total	13531	13531

Hình 42: Bảng khía cạnh theo cảm xúc tích cực

- "Food" được nhắc đến nhiều nhất với 6,192 lượt, cho thấy món ăn là khía cạnh nhận được nhiều đánh giá tích cực nhất.
- "Service" với 1,916 lượt và "Staff" với 1,707 lượt cho thấy dịch vụ và nhân viên cũng là những điểm được khách hàng đánh giá cao.
- "Atmosphere" (851 lượt) và "Restaurant" (807 lượt) có số lượt thấp hơn, nhưng vẫn phản ánh trải nghiệm tích cực về không gian và tổng thể nhà hàng.
- **Liên hệ với mục đích phân tích**

**Điểm mạnh – yếu của các nhà hàng?:** Điểm mạnh là "Food" (6,192 lượt tích cực) và "Service" (1,916 lượt tích cực), trong khi điểm yếu là "Price" (379 lượt tiêu cực), "Service" (123 lượt tiêu cực), và "Food" (99 lượt tiêu cực).

**Yếu tố khách hàng đánh giá cao và cần cải thiện?:** Khách hàng đánh giá cao "Food" và "Service", nhưng cần cải thiện "Price", "Wait time", và "Portion size" để giảm phản hồi tiêu cực.

## **CHƯƠNG 7. KẾT QUẢ NGHIÊN CỨU VÀ ĐỀ XUẤT**

### **7.1. Khám phá xu hướng ẩm thực và định hướng phát triển nhà hàng tại Đà Nẵng**

#### **7.1.1. Khách hàng đến từ quốc gia nào đánh giá nhiều nhất?**

Khách Hàn Quốc chiếm tỷ lệ lớn nhất với 37.74% tổng số đánh giá, cho thấy đây là nhóm khách hàng quốc tế trọng điểm. Khách Việt Nam (34.19%) giữ vai trò quan trọng ở thị trường nội địa, đặc biệt tại Hải Châu. Khách Trung Quốc chiếm ưu thế ở phân khúc giá cao (35.87%), mở ra cơ hội cho các mô hình nhà hàng cao cấp tại các khu vực như Sơn Trà và Ngũ Hành Sơn.

#### **7.1.2. Những món ăn nào được ưa chuộng?**

Các món Việt truyền thống như Bánh Xèo (820 lượt), Phở (695 lượt) và Cơm chiên (530 lượt) được nhắc đến nhiều nhất. Riêng khách Hàn Quốc đặc biệt yêu thích Bánh Xèo, Cơm chiên và Bún Chả, phản ánh sự ưa chuộng các món Việt quen thuộc. Dữ liệu cũng cho thấy khách Trung Quốc có xu hướng chọn các món ăn cao cấp như hải sản, tuy thông tin chi tiết còn hạn chế.

### 7.1.3. Du khách thường ăn gì vào các bữa trong ngày?

- Bữa tối (59.25%): Phở, Bánh Xèo và Cơm chiên là những món được nhắc đến nhiều nhất.
- Bữa trưa (31.88%): Bánh Xèo và Phở vẫn giữ vị trí cao, cùng với các món nhẹ như sinh tố xoài.
- Bữa sáng và brunch (4.2% mỗi loại): Dữ liệu chưa đủ để xác định món ăn đặc trưng.

### 7.1.4. Nhóm khách hàng tiềm năng nhất là ai?

- Khách Hàn Quốc (37.74%) là nhóm tiềm năng nhờ mức độ tương tác cao và yêu thích món Việt.
- Khách Trung Quốc nổi bật trong phân khúc cao cấp (35.87%).
- Khách nội địa (34.19%) quan trọng ở khu vực trung tâm như Hải Châu.

### 7.1.5. Điểm mạnh và điểm yếu trong đánh giá chất lượng dịch vụ?

- Điểm mạnh:
  - Chất lượng món ăn (“delicious”, “tasty”, “fresh”) với 6,192 lượt tích cực.
  - Dịch vụ và nhân viên (“friendly”, “attentive”) với 1,916 và 1,707 lượt tích cực.
  - Không gian và bầu không khí (“clean”, “cozy”) với 851 lượt tích cực.
- Điểm yếu:
  - Giá cả (“expensive” – 379 lượt tiêu cực).
  - Dịch vụ chậm, chất lượng món ăn kém (“slow”, “tasteless”, “diarrhea”).
  - Thời gian chờ (51 lượt) và khẩu phần nhỏ (36 lượt).

### 7.1.6. Cảm xúc khách hàng thay đổi như thế nào theo thời gian?

Số lượng đánh giá tăng mạnh từ 2022, đạt đỉnh năm 2024 với tỉ lệ cảm xúc tích cực chiếm ưu thế. Tuy nhiên, giai đoạn này cũng ghi nhận sự gia tăng nhẹ của các đánh giá tiêu cực, phản ánh yêu cầu ngày càng cao từ khách hàng.

### **7.1.7. Khu vực nào tập trung nhiều nhà hàng nhất?**

- Hải Châu dẫn đầu với 28 nhà hàng, theo sau là Sơn Trà (18) và Ngũ Hành Sơn (15).
- Thanh Khê có ít nhà hàng nhất (6), phản ánh tiềm năng phát triển còn hạn chế.

### **7.1.8. Loại hình nhà hàng phổ biến và có tiềm năng phát triển?**

- Phổ biến nhất là nhà hàng Việt Nam (27.61%) và tổng hợp (24.20%).
- Tiềm năng phát triển dành cho các mô hình hải sản (4.57%), chay (3.51%), Tây Ban Nha (3.44%), và cao cấp, đặc biệt khi nhắm đến khách Trung Quốc và quốc tế.

### **7.1.9. Nhà hàng điểm cao thường có đặc điểm gì nổi bật?**

- Tập trung ở Sơn Trà (điểm trung bình 4.71), nổi bật với:
  - Món ăn tươi ngon (“fresh”, “tasty”).
  - Không gian ven biển đẹp, yên tĩnh (“beautiful”, “pleasant”).
  - Dịch vụ tốt, phục vụ nhanh và thân thiện (“friendly”, “quick”).

### **7.1.10. Mức giá phổ biến của các nhà hàng là gì?**

- Phân khúc Lower-Mid là phổ biến nhất (3.1K nhà hàng).
- Phân khúc cao cấp còn ít (0.7K nhà hàng) nhưng có tiềm năng phát triển, đặc biệt cho nhóm khách Trung Quốc.
- Tuy nhiên, giá cả vẫn là yếu tố nhạy cảm, với 379 lượt đánh giá tiêu cực liên quan đến giá cao.

## **7.2. Hỗ trợ các bên liên quan**

### **7.2.1. Chủ sở hữu/Quản lý nhà hàng:**

- Lợi thế cạnh tranh: Chất lượng món ăn (6,192 lượt tích cực), dịch vụ (1,916 lượt tích cực), và sức hút của nhà hàng Việt Nam (27.61%).
- Vấn đề cần cải thiện: Giá cả (379 lượt tiêu cực), thời gian phục vụ (51 lượt tiêu cực), và vệ sinh (“diarrhea”).

- Điểm nổi bật được đánh giá cao: "Food" ("delicious", "fresh") và "Service" ("friendly", "attentive").
- Nhóm khách hàng trọng điểm: Khách Hàn Quốc (37.74%, thích Bánh Xèo, Com chiên), Trung Quốc (35.87% trong phân khúc giá cao), và nội địa (34.19%).
- Chiến lược đề xuất: Tối ưu hóa thực đơn cho khách Hàn Quốc (Bánh Xèo, Phở), phát triển nhà hàng cao cấp cho khách Trung Quốc, cải thiện giá cả và vận hành.

### 7.2.2. Nhà khởi nghiệp F&B:

- Mô hình nhà hàng tiềm năng: Nhà hàng Việt Nam (27.61%), tổng hợp (24.20%), và cao cấp (phục vụ khách Trung Quốc).
- Vị trí chiến lược: Sơn Trà và Ngũ Hành Sơn (điểm cao, khách quốc tế đa dạng), Hải Châu (khách nội địa và Hàn Quốc).
- Phong cách và tiện ích: Nhà hàng ven biển tại Sơn Trà nên tập trung vào không gian ("beautiful", "cozy") và món hải sản; nhà hàng tại Hải Châu nên nhấn mạnh dịch vụ nhanh và thực đơn Việt Nam.
- Đối tượng khách hàng tiềm năng: Khách Hàn Quốc (37.74%, thích món Việt), Trung Quốc (phân khúc giá cao), và nội địa (Hải Châu).
- Xây dựng thực đơn và định giá: Thực đơn nên ưu tiên Bánh Xèo, Phở, Com chiên; định giá cần tránh mức cao gây phản hồi tiêu cực (379 lượt "expensive").

## CHƯƠNG 8. TỔNG KẾT DỰ ÁN

**Đề tài "Phân tích Sở thích Ăn uống của Du khách dựa trên Đánh giá Nhà hàng: Nghiên cứu Tình huống tại Đà Nẵng"** được thực hiện với mục tiêu chính là khám phá sâu sắc các xu hướng ẩm thực, hành vi tiêu dùng và mức độ hài lòng của du khách tại các nhà hàng ở Đà Nẵng, từ đó xác định các yếu tố then chốt ảnh hưởng đến trải nghiệm khách hàng và đề xuất các giải pháp chiến lược cho các doanh nghiệp và nhà đầu tư trong lĩnh vực F&B. Để đạt được các mục tiêu này, nhóm nghiên cứu đã triển khai một quy trình phân tích dữ liệu bài bản, áp dụng hiệu quả các kiến thức và kỹ thuật đã học: Tổ chức và Thực thi Quy trình Phân tích Dữ liệu Chuyên sâu: Xác định vấn đề và mục tiêu nghiên cứu rõ ràng: Ngay từ đầu, đề tài đã định hình các câu hỏi nghiên cứu cốt lõi (ví dụ: "Sở thích ẩm thực của khách hàng?", "Đánh giá chất lượng dịch vụ và trải nghiệm khách hàng?", "Khu vực nào ở Đà Nẵng tập trung nhiều nhà hàng?") và mục tiêu phân



tích cụ thể cho từng giai đoạn. Thu thập và Tiền xử lý dữ liệu toàn diện: Đã tiến hành thu thập dữ liệu từ nền tảng Google Reviews thông qua kỹ thuật cào dữ liệu (web scraping) bằng Python. Tiếp đó, áp dụng một loạt các kỹ thuật tiền xử lý như làm sạch dữ liệu, xử lý giá trị thiếu (thay thế bằng "Unknown" hoặc dự đoán bằng mô hình hồi quy), rời rạc hóa dữ liệu (tạo bảng features và restaurant\_features), và biến đổi dữ liệu (tạo các cột mới như nation, review\_date, district) để chuẩn bị dữ liệu cho phân tích. Xây dựng và Ứng dụng Mô hình Ngôn ngữ Lớn (LLM): Triển khai mô hình DeepSeek-R1-Distill-Llama-70B-free thông qua API của Together.ai để thực hiện Phân tích Cảm xúc Dựa trên Khía cạnh (Aspect-Based Sentiment Analysis - ABSA). Đã thử nghiệm và lựa chọn prompt few-shot để đạt độ chính xác cao nhất trong việc trích xuất bộ ba (aspect, opinion, sentiment), với Joint Accuracy đạt 66.41%. Phân tích Khám phá Dữ liệu (EDA) và Trực quan hóa Dữ liệu Tương tác: Sử dụng các phương pháp thống kê mô tả và xây dựng hai dashboards trực quan trên Tableau (Dashboard Overview và Dashboard Review Analysis) để phân tích và trình bày các kết quả một cách trực quan, dễ hiểu, giúp khai thác sâu các thông tin quan trọng về phân bố nhà hàng, mức độ hài lòng, loại hình, giá cả, quốc tịch khách hàng, thói quen ăn uống và các khía cạnh cảm xúc. Các Kết quả và Đóng góp Chính của Đề tài: Xây dựng được công cụ phân tích và trực quan hóa dữ liệu hiệu quả: Hai dashboards trên Tableau cung cấp cái nhìn toàn diện và chi tiết về thị trường nhà hàng tại Đà Nẵng, làm rõ các xu hướng tiêu dùng nổi bật (như sự ưa chuộng ẩm thực địa phương của khách Hàn Quốc với 37.74% lượng đánh giá, xu hướng chi tiêu cao của khách Trung Quốc với 35.87% đánh giá trong phân khúc giá cao), các thời điểm cao điểm (bữa tối chiếm 59.25% đánh giá), phân bố địa lý (Hải Châu có 28 nhà hàng, Sơn Trà có điểm đánh giá trung bình cao nhất 4.71) và chất lượng dịch vụ theo quận, cũng như các yếu tố ảnh hưởng đến sự hài lòng của thực khách. Ứng dụng thành công Mô hình Ngôn ngữ Lớn trong phân tích định tính: Mô hình DeepSeek-R1-Distill-Llama-70B-free với prompt few-shot đã chứng minh hiệu quả trong việc tự động trích xuất các khía cạnh ("Food" được nhắc đến nhiều nhất với gần 10 nghìn lượt), ý kiến và cảm xúc từ hàng ngàn đánh giá của khách hàng, cung cấp nguồn dữ liệu định tính phong phú và sâu sắc, vượt trội so với các phương pháp phân tích thủ công truyền thống. Nhận diện sâu sắc các yếu tố ảnh hưởng đến trải nghiệm khách hàng và thị trường: Xác định rõ các nhóm khách hàng chủ đạo (Hàn Quốc, Việt Nam, Trung Quốc) và đặc điểm tiêu dùng riêng biệt của từng nhóm. Làm nổi bật các món ăn được yêu thích (Bánh

Xèo – 820 lượt nhắc, Phở – 695 lượt, Cơm chiên – 530 lượt) và thời điểm dùng bữa phổ biến. Đánh giá chi tiết điểm mạnh (chất lượng món ăn với 6,192 lượt tích cực, dịch vụ/nhân viên với hơn 1,900 lượt tích cực) và điểm yếu (giá cả với 379 lượt tiêu cực, thời gian phục vụ, vệ sinh). Phân tích được sự tập trung nhà hàng theo khu vực và tiềm năng của từng khu vực. Đưa ra được các bộ đề xuất chiến lược cụ thể và khả thi cho các chủ thể kinh doanh: Dựa trên những hiểu biết sâu sắc từ dữ liệu, đề tài đã xây dựng các đề xuất mang tính thực tiễn cho chủ sở hữu/quản lý nhà hàng và nhà khởi nghiệp trong lĩnh vực F&B, nhằm tối ưu hóa hoạt động kinh doanh, nâng cao chất lượng dịch vụ và thúc đẩy sự phát triển bền vững của ngành du lịch ẩm thực tại Đà Nẵng. Tóm lại, đề tài không chỉ hoàn thành các mục tiêu nghiên cứu đã đề ra mà còn thể hiện được năng lực áp dụng một cách có hệ thống các kỹ thuật thu thập, tiền xử lý, phân tích dữ liệu và ứng dụng mô hình ngôn ngữ lớn để giải quyết vấn đề thực tiễn trong lĩnh vực quản trị dịch vụ nhà hàng. Các kết quả và đề xuất của nghiên cứu cung cấp những thông tin và giải pháp giá trị, góp phần nâng cao khả năng cạnh tranh cho các doanh nghiệp và sự phát triển chung của ngành F&B tại Đà Nẵng.

## CHƯƠNG 9. TÀI LIỆU THAM KHẢO

- [1] B. M. N. R. Tom B. Brown, “Language Models are Few-Shot Learners,” 2020. [Trực tuyến]. Available: [https://www.researchgate.net/publication/341724146\\_Language\\_Models\\_are\\_Few-Shot\\_Learners](https://www.researchgate.net/publication/341724146_Language_Models_are_Few-Shot_Learners).
- [2] deepseek, “Incentivizing Reasoning Capability in LLMs via,” Hugging Face, 2024. [Trực tuyến]. Available: <https://huggingface.co/deepseek-ai/DeepSeek-R1-Distill-Llama-70B>.
- [3] P. Törnberg, “How to use LLMs for Text Analysis,” arXiv Is Hiring a DevOps Engineer, 24 07 2023. [Trực tuyến]. Available: <https://arxiv.org/abs/2307.13106>.
- [4] W. v. Atteveldt, “The Validity of Sentiment Analysis: Comparing Manual Annotation, Crowd-Coding, Dictionary Approaches, and Machine Learning Algorithms,” *Communication Methods and Measures*, số 2, pp. 121 - 140, 28 01 2021.

- [5] M. A. M. K. Fabrizio Gilardi, “ChatGPT outperforms crowd-workers for text-annotation tasks.,” 19 07 2023. [Trực tuyến]. Available: <https://arxiv.org/pdf/2303.15056>.
- [6] B. C. S. Kirk A. HawkinsORCID, “Textual analysis: Big data approaches,” *Taylor & Francis Group*, p. 22, 2018.
- [7] J. Wei, “Emergent Abilities of Large Language Models,” arXiv, 26 10 2022. [Trực tuyến]. Available: <https://arxiv.org/abs/2206.07682>.
- [8] A. B. M. Z. C. M. E. Aline Shakti Franzke, “Internet Research: Ethical Guidelines 3.0. Association of Internet Researchers,” 06 10 2019. [Trực tuyến]. Available: <https://aoir.org/reports/ethics3.pdf>.
- [9] BSA, “British sociological association statement of ethical practice.,” 2017. [Trực tuyến]. Available: [https://www.britsoc.co.uk/media/24310/bsa\\_statement\\_of\\_ethical\\_practice.pdf](https://www.britsoc.co.uk/media/24310/bsa_statement_of_ethical_practice.pdf).
- [10] Groq, “Groq Console,” Groq, 2025. [Trực tuyến]. Available: <https://console.groq.com/home>.
- [11] T. AI, “AI Acceleration Cloud,” Together AI, 2025. [Trực tuyến]. Available: <https://www.together.ai/>.